

TESIS

**PENERAPAN VGG19, RESNET50 DAN RESMASKNET PADA DETEKSI
EKSPRESI WAJAH**



Disusun oleh:

Nama : Fahma Inti Ilmawati
NIM : 21.51.2115
Konsentrasi : Intelligence Animation

PROGRAM STUDI S2 INFORMATIKA
PROGRAM PASCASARJANA UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA

2024

TESIS

**PENERAPAN VGG19, RESNET50 DAN RESMASKNET PADA DETEKSI
EKSPRESI WAJAH**

**APPLICATION OF VGG19, RESNET50 AND RESMASKNET TO
FACIAL EXPRESSION DETECTION**

Diajukan untuk memenuhi salah satu syarat memperoleh derajat Magister



Disusun oleh:

Nama : Fahma Inti Ilmawati
NIM : 21.51.2115
Konsentrasi : Intelllgence Animation

PROGRAM STUDI S2 INFORMATIKA
PROGRAM PASCASARJANA UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA

2024

HALAMAN PENGESAHAN

**PENERAPAN VGG19, RESNET50 DAN RESMASKNET PADA DETEKSI
EKSPRESI WAJAH**

**APPLICATION OF VGG19, RESNET50 AND RESMASKNET TO FACIAL
EXPRESSION DETECTION**

Dipersiapkan dan Disusun oleh

Fahma Inti Ilmawati

21.51.2115

Telah Diujikan dan Dipertahankan dalam Sidang Ujian Tesis
Program Studi S2 Informatika
Program Pascasarjana Universitas AMIKOM Yogyakarta
pada hari Senin, 3 Juni 2024

Tesis ini telah diterima sebagai salah satu persyaratan
untuk memperoleh gelar Magister Komputer

Yogyakarta, 3 Juni 2024

Rektor

Prof. Dr. M. Suyanto, M.M.
NIK. 190302001

HALAMAN PERSETUJUAN

**PENERAPAN VGG19, RESNET50 DAN RESMASKNET PADA DETEKSI
EKSPRESI WAJAH**

**APPLICATION OF VGG19, RESNET50 AND RESMASKNET TO FACIAL
EXPRESSION DETECTION**

Dipersiapkan dan Disusun oleh

Fahma Inti Ilmawati

21.51.2115

Telah Diujikan dan Dipertahankan dalam Sidang Ujian Tesis
Program Studi S2 Informatika
Program Pascasarjana Universitas AMIKOM Yogyakarta
pada hari Senin, 3 Juni 2024

Pembimbing Utama

Anggota Tim Penguji

Prof. Dr. Kusriani, M.Kom
NIK. 190302106

M. Hanafi, S.Kom., M.Eng., Ph.D.
NIK. 190302024

Pembimbing Pendamping

Dr. Kumara Ari Yuana, S.T., M.T.
NIK. 190302575

Tonny Hidayat, M.Kom., Ph.D
NIK. 190302182

Prof. Dr. Kusriani, M.Kom
NIK. 190302106

Tesis ini telah diterima sebagai salah satu persyaratan
untuk memperoleh gelar Magister Komputer

Yogyakarta, 3 Juni 2024

Direktur Program Pascasarjana

Prof. Dr. Kusriani, M.Kom.
NIK. 190302106

HALAMAN PERNYATAAN KEASLIAN TESIS

Yang bertandatangan di bawah ini,

Nama mahasiswa : Fahma Inti Ilmawati
NIM : 21.51.2115
Konsentrasi : Intelligence Animation

Menyatakan bahwa Tesis dengan judul berikut:

PENERAPAN VGG19, RESNET50 DAN RESMASKNET PADA DETEKSI EKSPRESI WAJAH

Dosen Pembimbing Utama : Prof. Dr. Kusriani, M.Kom.
Dosen Pembimbing Pendamping : Tonny Hidayat, M.Kom., Ph.D

1. Karya tulis ini adalah benar-benar ASLI dan BELUM PERNAH diajukan untuk mendapatkan gelar akademik, baik di Universitas AMIKOM Yogyakarta maupun di Perguruan Tinggi lainnya
2. Karya tulis ini merupakan gagasan, rumusan dan penelitian SAYA sendiri, tanpa bantuan pihak lain kecuali arahan dari Tim Dosen Pembimbing
3. Dalam karya tulis ini tidak terdapat karya atau pendapat orang lain, kecuali secara tertulis dengan jelas dicantumkan sebagai acuan dalam naskah dengan disebutkan nama pengarang dan disebutkan dalam Daftar Pustaka pada karya tulis ini
4. Perangkat lunak yang digunakan dalam penelitian ini sepenuhnya menjadi tanggung jawab SAYA, bukan tanggung jawab Universitas AMIKOM Yogyakarta
5. Pernyataan ini SAYA buat dengan sesungguhnya, apabila di kemudian hari terdapat penyimpangan dan ketidakbenaran dalam pernyataan ini, maka SAYA bersedia menerima SANKSI AKADEMIK dengan pencabutan gelar yang sudah diperoleh, serta sanksi lainnya sesuai dengan norma yang berlaku di Perguruan Tinggi

Yogyakarta, 3 Juni 2024
Yang Menyatakan,



10000
MELAKUKAKAN
TEMPEL

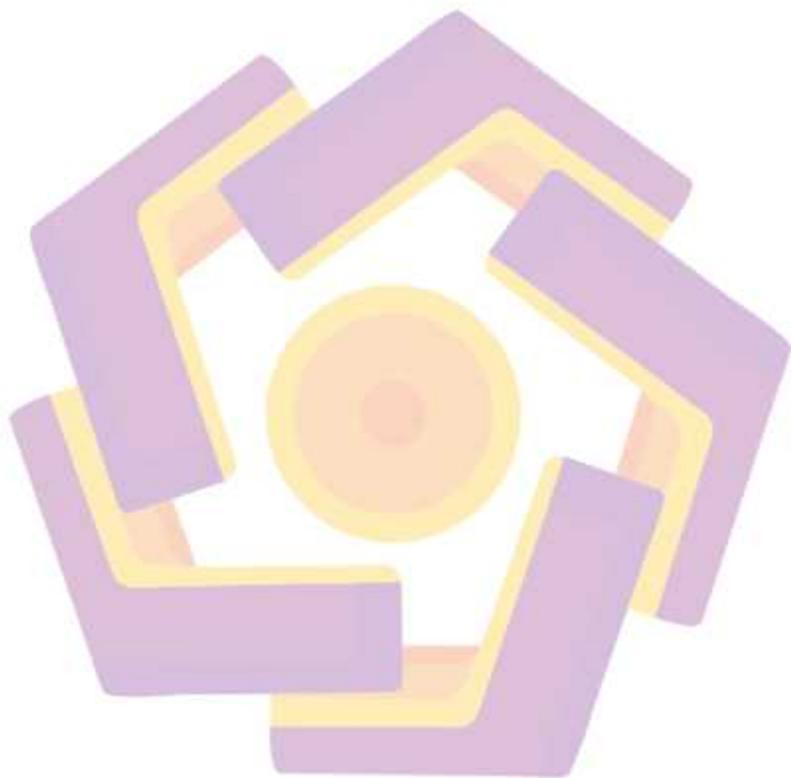
Fahma Inti Ilmawati

HALAMAN PERSEMBAHAN

Dengan rasa syukur yang mendalam, dengan telah diselesaikannya tesis ini penulis mempersembahkannya kepada:

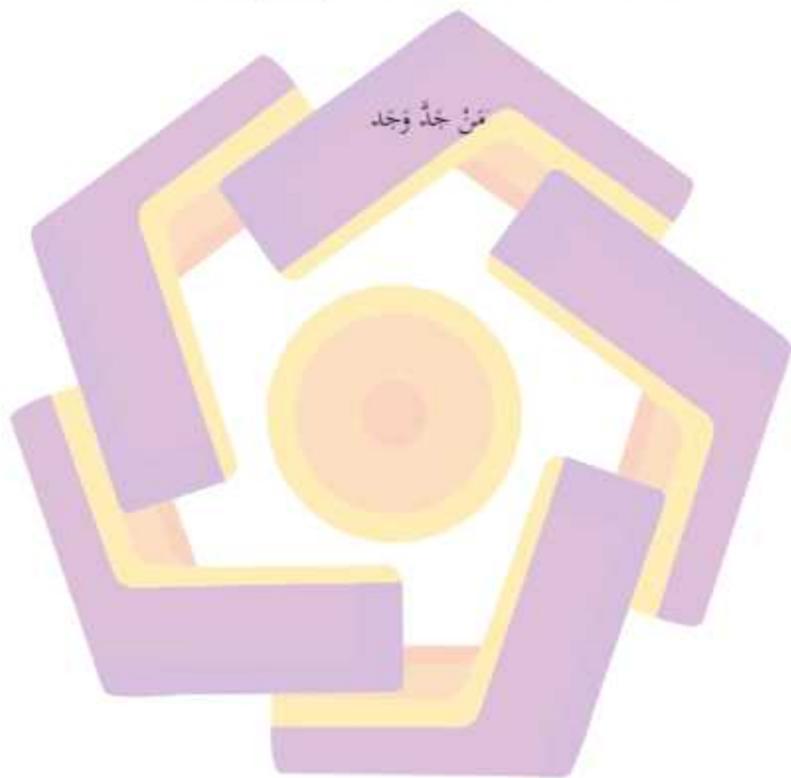
1. Allah SWT yang telah melimpahkan rahmat dan hidayah-Nya sehingga penulis dapat menyelesaikan tesis ini dengan baik dan tepat waktu.
2. Kedua orang tua saya yang memberi "Kehidupan", Ibuku tersayang Sarjiah dan Bapakku tersayang Muhtar yang telah memberikan dukungan moril maupun material serta doa yang tiada henti untuk saya.
3. Kepada diri saya sendiri yang mampu bertahan, berusaha, dan terus berjuang tanpa henti serta tidak menyerah pada kehidupan ini.
4. Ibu Prof. Dr. Kusriani, M. Kom., selaku Dosen Pembimbing 1 yang telah memberikan bimbingan, saran, kritik, dan motivasi kepada penulis sehingga tesis ini dapat terselesaikan.
5. Bapak Tonny Hidayat, S.Kom., M.Kom., Ph.D., selaku Dosen Pembimbing 2 yang telah memberikan arahan dan masukan selama penulisan tesis ini.
6. Muhamad Sulton, sang kekasih hati, suami tercinta, yang menjadi alasan saya untuk segera menyelesaikan tesis ini.
7. Kakakku tersayang Hafid Khairudin, Hamdan Fauzi, Nisa Farissa, dan Citra Ayu yang telah memberikan dukungan serta motivasi selama penyusunan tesis ini.
8. Teman-teman MTI Angkatan 27 Kelas A yang telah memberikan semangat, saran, masukan dan pengalaman yang tak ternilai, khususnya Maie.

9. Semua orang yang memberi daya hidup, teruslah hidup dalam cinta dan kebaikan.



HALAMAN MOTTO

مَنْ عَرَفَ اللَّهَ أَزَالَ الثُّمَمَةَ # وَقَالَ كُلُّ فَعْلِهِ بِالْحِكْمَةِ



KATA PENGANTAR

Assalamu'alaikum Wr.Wb.

Alhamdulillahirabbil'Alamin. Segala puji bagi Allah SWT yang telah memberikan rahmat, hidayah, dan ridha-Nya, sehingga penulis dapat menyelesaikan tesis yang berjudul **"PENERAPAN VGG19, RESNET50 DAN RESMASKNET PADA DETEKSI EKSPRESI WAJAH"**. Penulis menyadari masih terdapat banyak kekurangan selama penyusunan tesis, dan tesis ini dapat diselesaikan karena doa, dukungan, bantuan serta bimbingan dari berbagai pihak. Dalam kesempatan ini, dengan segala kerendahan hati, penulis ingin menyampaikan ucapan terimakasih yang sebesar-besarnya kepada:

1. Bapak Prof. Dr. M. Suyanto, M.M., selaku Rektor Universitas AMIKOM Yogyakarta
2. Ibu Prof. Dr. Kusriani, M.Kom., selaku Direktur Program Pascasarjana Universitas AMIKOM Yogyakarta
3. Bapak Tonny Hidayat, S.Kom., M.Kom., Ph.D. selaku dosen pembimbing yang dengan sabar memberikan bimbingan, arahan, masukan, dan motivasi selama proses penulisan naskah tesis ini.
4. Segenap civitas akademika Pascasarjana, terutama seluruh dosen, yang telah memberikan ilmu dan bimbingannya.
5. Kedua orang tua yang tak pernah lelah memberikan dukungan dan doa selama ini.

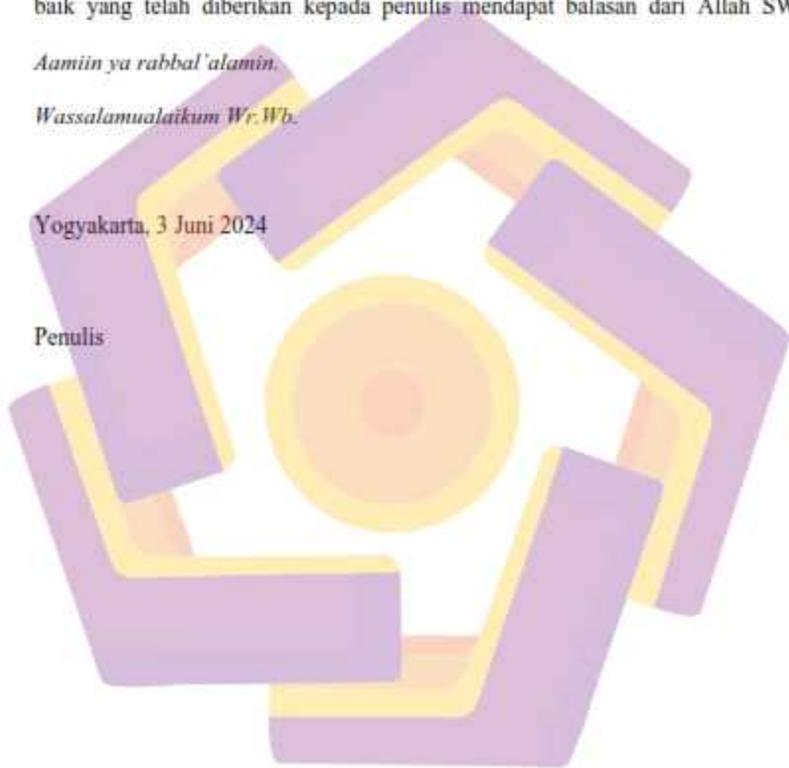
Penulis menyadari bahwa keterbatasan pengetahuan dan pengalaman penulis yang masih jauh dari harapan, untuk itu penulis sangat mengharapkan kritik, saran, dan masukan yang bersifat membangun ke arah perbaikan dan penyempurnaan tesis ini. Penulis berharap tesis ini dapat bermanfaat bagi seluruh pihak, dan semoga amal baik yang telah diberikan kepada penulis mendapat balasan dari Allah SWT.

Aamin ya rabbal'amin.

Wassalamualaikum Wr.Wb.

Yogyakarta, 3 Juni 2024

Penulis



DAFTAR ISI

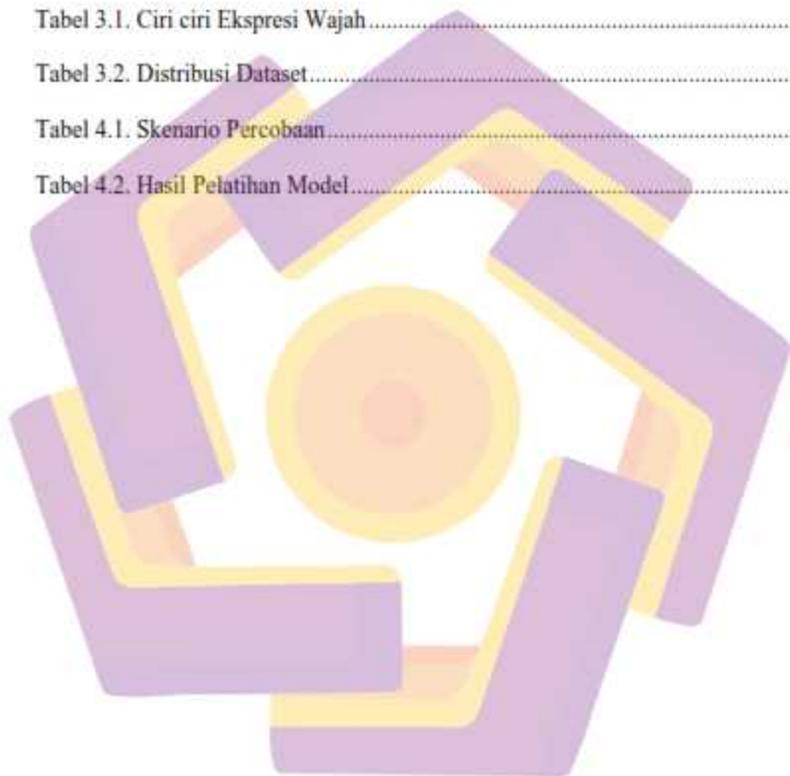
HALAMAN JUDUL.....	ii
HALAMAN PENGESAHAN.....	iii
HALAMAN PERSETUJUAN.....	iv
HALAMAN PERNYATAAN KEASLIAN TESIS.....	v
HALAMAN PERSEMBAHAN.....	vi
HALAMAN MOTTO.....	viii
KATA PENGANTAR.....	ix
DAFTAR ISI.....	xi
DAFTAR TABEL.....	xiv
DAFTAR GAMBAR.....	xv
INTISARI.....	xvi
<i>ABSTRACT</i>	xvii
BAB I PENDAHULUAN.....	1
1.1. Latar Belakang Masalah.....	1
1.2. Rumusan Masalah.....	4
1.3. Batasan Masalah.....	5
1.4. Tujuan Penelitian.....	5
1.5. Manfaat Penelitian.....	5
BAB II TINJAUAN PUSTAKA.....	7
2.1. Tinjauan Pustaka.....	7
2.2. Keaslian Penelitian.....	11

2.3. Landasan Teori.....	16
2.3.1. Arsitektur CNN	16
2.3.2. VGG19	17
2.3.3. ResNet-50.....	18
2.3.4. Residual Masking Network.....	20
2.3.5. One Sided Selection	21
2.3.6. Deteksi.....	21
BAB III METODE PENELITIAN.....	24
3.1. Jenis, Sifat, dan Pendekatan Penelitian.....	24
3.2. Metode Pengumpulan Data.....	25
3.3. Alur Penelitian	30
BAB IV HASIL PENELITIAN DAN PEMBAHASAN.....	35
4.1. Pengumpulan Data.....	35
4.2. Skenario Percobaan.....	36
4.3. Preprocessing Data.....	36
4.4.1. Augmentasi Data	37
4.4.2. Reshape.....	38
4.4.3. <i>Balancing</i>	39
4.4.4. Split Data.....	41
4.4. Arsitektur Model.....	41
4.4.1. Arsitektur VGG19	42

4.4.2. Arsitektur ResNet50	44
4.4.3. ResMaskingNet	45
4.5. Hasil Pelatihan dan Evaluasi Model	48
4.5.1. Hasil Pelatihan Model	48
4.5.2. Hasil Pengujian Model	56
4.6. Evaluasi Model	61
4.7. Perbandingan dengan Penelitian Sebelumnya	66
BAB V PENUTUP	72
5.1. Kesimpulan	72
5.2. Saran	73
DAFTAR PUSTAKA	76

DAFTAR TABEL

Tabel 2.1. Matriks Literature Review	11
Tabel 3.1. Ciri ciri Ekspresi Wajah	26
Tabel 3.2. Distribusi Dataset	28
Tabel 4.1. Skenario Percobaan	36
Tabel 4.2. Hasil Pelatihan Model	49



DAFTAR GAMBAR

Gambar 2.1. Proses CNN.....	17
Gambar 2.2. Arsitektur VGG-19.....	17
Gambar 2.3. Arsitektur ResNet-50	19
Gambar 2.4. ResMaskNet.....	21
Gambar 3.1. Ekspresi Wajah Dalam Dataset FEREC.....	27
Gambar 3.2. Alur Penelitian.....	30
Gambar 3.3. Proses Balancing	33
Gambar 4.1. Distribusi Dataset FEREC	35
Gambar 4.2. Grafik Akurasi dan Loss data Unbalance, pada VGG19 (a), ResNet50 (b), dan ResMaskNet (c).....	52
Gambar 4.3. Grafik Akurasi dan Loss data Balanced, pada VGG19 (a), ResNet50 (b), dan ResMaskNet (c).....	54
Gambar 4.4. Time Consumption Pada VGG19, ResNet50, ResMaskNet.....	55
Gambar 4.5. Confusion Matrix Skenario 1 pada (a) VGG19, (b) ResNet50, (c)ResMaskNet, dan Skenario 2 pada (d) VGG19, (e) ResNet50, (f) ResMaskNet.....	58
Gambar 4.6. Evaluasi Metric pada Skenario 1.....	61
Gambar 4.7. Perbandingan Akurasi	66

INTISARI

Pada bidang pengenalan ekspresi wajah (FER), ketersediaan kumpulan data yang seimbang dan representatif adalah kunci keberhasilan dalam melatih model yang akurat. Namun, kumpulan data Tantangan Pengenalan Ekspresi Wajah (FERC) sering kali menghadapi tantangan ketidakseimbangan kelas, di mana beberapa ekspresi wajah memiliki jumlah sampel yang jauh lebih kecil dibandingkan ekspresi wajah lainnya. Masalah ini dapat mengakibatkan performa model menjadi bias dan tidak memuaskan, terutama dalam mengenali ekspresi wajah yang kurang umum.

Teknik augmentasi data menjadi strategi penting karena dapat memperluas kumpulan data dengan menciptakan variasi baru dari sampel yang ada, sehingga meningkatkan variasi dan keragaman data. Augmentasi data dapat digunakan untuk meningkatkan jumlah sampel untuk kelas ekspresi wajah yang kurang umum, sehingga meningkatkan kemampuan model untuk mengenali dan memahami beragam ekspresi wajah. Hal ini dikarenakan jumlah sampel dalam data FERC tidak seimbang, sehingga dapat menyebabkan model memiliki kecenderungan dalam mempelajari kelas mayoritas dengan baik sehingga mengabaikan kelas minoritas. Pada penelitian ini, untuk mengatasi hal tersebut dilakukan proses oversampling menggunakan metode "OneSideSelection".

Dalam penelitian ini, VGG19, Resnet50 dan ResMakingNet digunakan untuk mendukung performa model yang lebih baik. Hal ini akan memberikan pedoman berharga untuk mengoptimalkan model CNN yang lebih canggih di masa depan dan dapat mendorong penelitian lebih lanjut dalam menciptakan teknik augmentasi yang lebih inovatif.

Kata kunci: Augmentasi, Convolutional Neural Network, Resnet50, ResMakingNet, VGG19

ABSTRACT

In the field of facial expression recognition (FER), the availability of a balanced and representative dataset is key to success in training accurate models. However, the Facial Expression Recognition Challenge (FERC) dataset often faces the challenge of class imbalance, where some facial expressions have a much smaller number of samples than other facial expressions. This problem can result in biased and unsatisfactory model performance, especially in recognizing less common facial expressions. Data augmentation techniques are an important strategy as they can expand the data set by creating new variations of existing samples, thus increasing the variety and diversity of the data.

Data augmentation can be used to increase the number of samples for less common facial expression classes, thereby improving the model's ability to recognize and understand diverse facial expressions. This is because the number of samples in the FERC data is unbalanced, which may cause the model to have a tendency to learn the majority class well while ignoring the minority class. In this study, to overcome this, an oversampling process was carried out using the "OneSideSelection" method.

In this study, VGG19, Resnet50 and ResMakingNet were used to support better model performance. This will provide valuable guidelines for optimizing more sophisticated CNN models in the future and may encourage researchers to develop more sophisticated CNN models.

Keyword: Augmentation, Convolutional Neural Network, Resnet50, ResMakingNet, VGG19.

BAB I

PENDAHULUAN

1.1. Latar Belakang Masalah

Ekspresi wajah adalah salah satu bentuk komunikasi non-verbal yang memegang peran penting dalam interaksi manusia. Sebagian besar komunikasi adalah secara nonverbal, yaitu dengan menggunakan gerak tubuh, mengubah suara, dan menampakkan ekspresi wajah tertentu untuk menunjukkan perasaan kita terhadap sesuatu. Ekspresi wajah secara akurat merepresentasikan kondisi emosional seseorang. Dengan ekspresi wajah, dapat mengungkapkan informasi tambahan yang sering kali tidak ketara. Melalui ekspresi wajah, kondisi emosional manusia secara lahiriah dapat disampaikan dengan mengubah tatapan mata, tersenyum ataupun mengubah ekspresi ketika merasakan suatu emosi. Meskipun ada emosi yang rumit yang disengaja diproyeksikan oleh manusia, emosi-emosi utama ini masih dapat dikenali secara luas berdasarkan definisi yang sederhana.

Penelitian oleh (Ekman) mengklasifikasikan tujuh ekspresi wajah dasar yang ditampilkan secara universal, yaitu marah, takut, jijik, sedih, bahagia, terkejut, dan menghina. Ketika seseorang tersenyum, itu mencerminkan kebahagiaan mereka dan terlihat dari mata yang membentuk lengkungan. Ekspresi kesedihan mencerminkan perasaan sedih, yang sering ditunjukkan oleh angkatannya alis, dan kerutan di wajah. Kemarahan, terkait dengan situasi yang mengganggu dan menjengkelkan. Ekspresi kemarahan tercermin dalam mata

yang sempit, alis yang naik, dan lipatan yang memanjang. Ekspresi rasa jijik diekspresikan dengan alis yang menurun dan hidung yang berkerut. Kemudian ekspresi kaget terjadi ketika ada situasi yang tak terduga. Ini biasanya terlihat dalam bentuk mata yang melebar dan mulut yang terbuka, dan merupakan ekspresi yang mudah dikenali. Sedangkan, ekspresi ketakutan berhubungan dengan ekspresi terkejut, yang sering ditandai dengan terangkatnya alis secara tajam (Revina I, Emmanuel W, 2021).

Dalam hal ekspresi wajah, perlu dicatat bahwa sebagian besar isyarat berasal dari beberapa daerah wajah seperti mata, mulut. Sebaliknya, daerah lain memiliki kontribusi kecil untuk output, misalnya rambut, rahang. Seperti pada penelitian (Li S, Deng W, 2018), berfokus pada wilayah penting tersebut dengan menggunakan fitur perantara landmark wajah. Sedangkan penelitian yang dilakukan oleh (Pham L et al, 2020) fokus mengidentifikasi enam ekspresi emosional wajah, dan netral dalam gambar statis, tanpa mempertimbangkan faktor waktu. Ini menunjukkan bahwa, metode deteksi ekspresi wajah yang telah diimplementasikan secara tradisional, hasilnya membutuhkan waktu yang sangat lama untuk melakukan ekstrak fitur dalam gambar (An & Liu, 2020; Suma, Lakshminarayana, & Tech, 2018).

Belakangan ini, penggunaan teknologi deteksi wajah semakin banyak digunakan dalam berbagai aplikasi, seperti untuk keamanan, deteksi identitas, dan kemudahan untuk mengakses teknologi (He, 2023). Sistem ini memberikan kemudahan akan kebutuhan akses keamanan tradisional seperti kartu, token, atau

kunci, serta meningkatkan produktivitas bisnis. Seperti halnya dalam konteks Google foto, deteksi wajah juga digunakan untuk pengaturan foto, untuk memetakan dan menganalisis, dan mengonfirmasi wajah yang ada dalam foto. Selain itu juga digunakan dalam penegakan hukum untuk menghasilkan petunjuk dalam penyelidikan criminal, memantau, mengawasi lokasi, dan juga mengendalikan akses gedung maupun fasilitas umum.

Penelitian lainnya (Pham et al., n.d.) menggunakan metode deteksi ekspresi wajah berbasis "Residual Masking Network" untuk menganalisis data wajah. ResMaskingNet merupakan teknik deteksi wajah berbasis deep learning yang efektif dalam mengidentifikasi dan mengklasifikasikan ekspresi wajah. Penelitian yang dilakukan oleh (Eastern Washington University et al., n.d.) mengenai deteksi emosi wajah menggunakan CNN dengan augmentasi data. Dengan menggabungkan beberapa kumpulan data, hasil akurasi yang diperoleh cukup tinggi, dengan augmentasi data sebesar 96,24%. Dari banyaknya emosi, penelitian yang dilakukan hanya menggunakan empat kelas emosi saja. Peluang untuk mengembangkan lebih lanjut penelitian tersebut yaitu pemanfaatan jumlah data dan kelas emosi wajah yang digunakan, dan mengatasi masalah ketidakseimbangan data yang ada.

Terdapat banyak dataset mengenai emosi wajah, namun pada penelitian ini akan menggunakan dataset FEREC. Data FEREC berisikan kumpulan data dengan 7 emosi yang dibuat pada 2014 yang bertujuan membuat model CNN (Dewan, Likhith Ashwin, Likhith, Allabhaneni, & Janardhana, 2022). Dataset ini terdiri dari

35.888 gambar dari emosi marah, netral, jijik, takut, bahagia, sedih, dan terkejut. Model pembelajaran mendalam, terutama Convolutional Neural Networks (CNN), telah menunjukkan hasil yang menjanjikan dalam deteksi emosi wajah (Rangaswamy, n.d.; Tong, Cao, Sun, & Chen, 2021). Namun deteksi emosi wajah yang akurat masih menjadi tantangan karena heterogenitas wajah manusia dan variasi gambar, seperti pose wajah dan pencahayaan yang berbeda (Khairuddin & Chen, 2021). Selain itu, dataset tersebut memiliki kekurangan yaitu adanya ketidakseimbangan pada setiap kelasnya. Salah satu masalah yang sering muncul pada klasifikasi yaitu adanya ketidakseimbangan kelas yang sangat tinggi. Ketidakseimbangan pada kelas tersebut dapat berdampak negatif pada kinerja klasifikasi model CNN (Ding et al., 2020). Oleh karena itu, maka pada penelitian ini akan melakukan klasifikasi emosi wajah dengan menerapkan teknik undersampling One Side Selection untuk mengatasi ketidakseimbangan kelas dan membandingkan arsitektur VGG19, ResNet50, dan ResMasking Net menggunakan dataset FEREC.

1.2. Rumusan Masalah

Rumusan masalah penelitian ini dapat dijabarkan sebagai berikut:

- a. Berapa tingkat akurasi deteksi ekspresi wajah dengan VGG19, ResNet-50, dan ResMaskingNet?
- b. Apakah penerapan menyeimbangkan data yang digunakan dapat meningkatkan akurasi yang dihasilkan dari model VGG19, ResNet-50, dan ResMaskingNet?
- a. Model arsitektur apa yang paling cocok untuk meningkatkan akurasi deteksi wajah dengan dataset FEREC?

- b. Berapa akurasi yang dapat dicapai pada penerapan model arsitektur dengan menerapkan teknik balancing?

1.3. Batasan Masalah

Agar penelitian ini terarah dan pembahasan menjadi relevan, maka diberikan batasan-batasan dan ruang lingkup sebagai berikut:

- a. Dataset yang digunakan adalah FEREC.
- b. Model yang digunakan pada penelitian ini yaitu Residual Masking Network, Resnet50 dan VGG19.
- c. Data yang digunakan merupakan citra ekspresi wajah terdapat tujuh ekspresi yaitu sedih, senang, terkejut, marah, jijik, takut dan netral.
- d. Pengujian kinerja model dievaluasi menggunakan Confusion Matric
- e. Platform yang digunakan untuk melakukan pelatihan, pengujian, dan kebutuhan visualisasi adalah Google Colab Pro.

1.4. Tujuan Penelitian

Tujuan dari penelitian ini adalah untuk mendapatkan hasil akurasi dari metode yang diajukan sebagai upaya dalam mengembangkan dan melatih model Residual Masking Network yang efektif dalam mendeteksi dan mengklasifikasikan ekspresi wajah. Model ini akan mempelajari dan mengenali pola ekspresi wajah yang mengindikasikan respon emosi wajah.

1.5. Manfaat Penelitian

Berdasarkan tujuan yang ingin dicapai, penelitian ini diharapkan bermanfaat sebagai berikut:

- a. Hasil pemodelan ini dapat digunakan sebagai alat untuk mendapatkan pemahaman yang lebih mendalam dalam pengembangan dan evaluasi penelitian terkait klasifikasi dalam penelitian berikutnya.
- b. Model yang dihasilkan dapat menjadi alternatif dalam melakukan deteksi ekspresi pada wajah.



BAB II

TINJAUAN PUSTAKA

2.1. Tinjauan Pustaka

Dari peninjauan pustaka yang dilakukan, peneliti memperoleh beberapa kajian penelitian-penelitian terdahulu yang berkaitan dengan penelitian yang akan dilakukan.

Penelitian yang dilakukan oleh (Pham, 2020), melakukan pengujian terhadap kelayakan dari sistem yang dibuat untuk mengenali ekspresi pada wajah seseorang menggunakan metode *Convolutional Neural Network (CNN)*. Peneliti melakukan pengujian terhadap sistem deteksi ekspresi wajah menggunakan *Convolutional Neural Network (CNN)*. Mereka mengusulkan tiga inovasi utama: masking baru, Residual Masking Network, dan dataset baru bernama VEMO. Dengan menerapkan masking dalam empat blok, peneliti berhasil menghasilkan matriks nilai yang menunjukkan kinerja yang memuaskan pada sebagian besar ekspresi wajah, meskipun ketidakseimbangan dataset pelatihan antara FER2013 dan VEMO. Ekspresi-ekspresi seperti Happy dan Surprise menunjukkan skor yang tinggi, yakni 0,91 dan 0,81, serta 0,85 dan 0,71, sedangkan ekspresi Fear dan Sad hanya mencapai skor yang lebih rendah, yaitu 0,56 dan 0,44, serta 0,59 dan 0,63, menjadi skor terendah dalam evaluasi

Penelitian lainnya (Tawsin et al, 2019), membuat sistem identifikasi ekspresi wajah yang mampu mengkategorikan citra ke dalam tujuh kelas emosi yang berbeda. Dengan menggunakan augmentasi pada CNN, akurasi validasi

mencapai 96,24% setelah 120 epoch, sementara CNN tanpa augmentasi membutuhkan 260 epoch untuk mencapai akurasi validasi 92,95%. Augmentasi dilakukan dengan menciptakan gambar-gambar baru dari data gambar yang ada. Selain itu, penerapan Dropout dan EarlyStopping Callbacks dilakukan untuk mencegah overfitting. Evaluasi dilakukan dengan rasio 65% untuk pelatihan dan 35% untuk pengujian, menghasilkan akurasi tertinggi sebesar 95,87%. Meskipun sedikit lebih rendah dari akurasi sebelumnya dengan rasio pemisahan 80:20, hal ini dianggap wajar karena pengujian dilakukan dengan dataset yang lebih besar.

Peneliti (Jie Shao et al, 2019) mengusulkan tiga model CNN inovatif dengan struktur yang berbeda. Model pertama, yang dinamakan sebagai Light-CNN, adalah jaringan saraf konvolusional dengan kedalaman terbatas. Ini terdiri dari enam modul konvolusi residual yang dapat dipisahkan untuk mengatasi tantangan topologi yang kompleks dan overfitting. Model kedua adalah dual-branch CNN yang secara simultan mengekstrak fitur dari Local Binary Pattern (LBP) tradisional dan deep learning. Sedangkan model ketiga adalah CNN yang telah dipre-training dengan menggunakan teknik transfer learning, dimaksudkan untuk mengatasi kendala yang terkait dengan keterbatasan sampel pelatihan.

Penelitian berikutnya yang dilakukan oleh (Jason C et al, 2019), dalam studinya menyatakan bahwa meningkatkan kinerja jaringan FaceLiveNet dalam deteksi emosi, baik pada tingkat akurasi tinggi maupun rendah. Dalam konteks ini, sebuah framework baru yang disebut Dense_FaceLiveNet diusulkan pada dua model transfer learning. Penelitian ini menggunakan model inovatif yang disebut Dense_FaceLiveNet, berbasis arsitektur CNN, untuk deteksi emosi pada wajah

manusia. Dengan menggunakan dataset JAFFE dan KDEF, penelitian ini mengevaluasi kinerja model dengan mencapai akurasi tinggi, yaitu 90,97% dan 95,89% pada dataset masing-masing. Hasilnya menunjukkan bahwa model FER2013, yang dirancang sejalan dengan JAFFE dan KDEF, memiliki akurasi 70,02%, menunjukkan pengaruh signifikan dari ukuran dan kompleksitas data. Dari dataset FER2013, akurasi tertinggi yang dicapai adalah 84,59% menggunakan teknik transfer learning.

Penelitian lainnya yang pernah dilakukan oleh (Cao W et al, 2020) menjelaskan bahwa mereka memutuskan untuk menyertakan Convolutional Block Attention Module (CBAM) dalam beberapa lapisan jaringan VGG yang digunakan untuk meningkatkan akurasi dan stabilitas. Dengan menggunakan CBAM, penelitian ini menyederhanakan informasi yang perlu dianalisis oleh jaringan. Penelitian ini menguji CBAM pada dua basis data ekspresi wajah publik, termasuk CK+ dan FER-2013, dan membandingkan efek penyematan CBAM ke dalam jaringan asli VGG19. Hasilnya menunjukkan bahwa nilai loss konvergen menjadi hanya 0,067 dengan CBAM disematkan, sementara tanpa CBAM nilai loss konvergen mencapai 0,145, menggunakan SGD (Stochastic Gradient Descent) sebagai pengoptimal dalam kedua situasi pelatihan.

Penelitian terkait deteksi menggunakan KFSENet (Keyframe-based Feature Skeleton Estimation Network) yang dilakukan oleh (Le D, 2022). mengintegrasikan deteksi tindakan dengan deteksi wajah dan emosi, membentuk pendekatan terintegrasi dalam visi robot. Metode yang diusulkan diuji pada kedua standar benchmark publik dan dataset yang dikumpulkan sendiri, menggunakan framework

double-feature double-motion network (DDNet). Hasilnya menunjukkan bahwa menggunakan metode arcFace IR50 Backbone Network + Softmax Classifier menghasilkan akurasi sebesar 73,04%, sementara dengan metode Residual Masking Network, akurasi mencapai 74,14%.

Penelitian oleh (Mehendale, 2020) memanfaatkan dataset FEREC (Facial emotion recognition using convolutional neural networks) untuk mengenali emosi wajah menggunakan jaringan saraf konvolusional. Penelitian ini bertujuan mendeteksi emosi wajah dengan menggunakan dataset FEREC dan membandingkan struktur jaringan seperti Alexnet, VGG, GoogleNet, dan Resnet. Hasilnya menunjukkan bahwa menggunakan dataset FEREC menghasilkan akurasi tertinggi sebesar 78%. Metode ini memanfaatkan hanya 24 fitur EV (expressional vector) pada wajah, dan perlu dilakukan penyesuaian otomatis pada gambar ke depannya. Penelitian sebelumnya menggunakan dua skenario CNN, yaitu menghilangkan latar belakang gambar dan fokus pada fitur wajah EV.

Menurut (Rangaswamy S, 2020) model MobileNet dan CNN digunakan untuk mengklasifikasi ekspresi wajah orang yang menggunakan masker dan tidak menggunakan masker secara real-time melalui webcam. Dalam penelitian ini, gambar individu dikirimkan ke pengklasifikasi machine learning untuk menentukan apakah mereka mengenakan masker atau tidak. Dua model yang digunakan untuk membandingkan kinerja dalam hal akurasi, di mana Model CNN mencapai akurasi 99%, sementara Model MobileNet mencapai akurasi 98%.

2.2. Keaslian Penelitian

Tabel 2.1. Matriks Literature Review
PENERAPAN VGG19, RESNET50 DAN RESMASKNET PADA DETEKSI EKSPRESI WAJAH

No	Judul	Peneliti, Media Publikasi, dan Tahun	Tujuan Penelitian	Kesimpulan	Saran atau Kelemahan	Perbandingan
1	Facial Expression Recognition Using Residual Masking Network	Luan Pham, Huynh Vu, Tuan Anh Tran 2020, IEEE	Mengajukan ide masking baru untuk meningkatkan kinerja untuk deteksi ekspresi wajah menggunakan Residual Masking Network	Dari metode Residual Masking Network menghasilkan nilai akurasi 76,82 %	Dataset FER2013 dan VEMO tidak seimbang. Diharapkan kedepannya memakai dataset yang lebih seimbang.	Peneliti sebelumnya menggunakan dataset FER2013 dan VEMO guna mendeteksi ekspresi wajah. Penelitian Pham ini menggunakan 4 layer blok masking. Penelitian ini menggunakan dataset FERC guna mendeteksi ekspresi wajah. Dan menambahkan satu blok layer guna mempercepat waktu akurasi dalam deteksi ekspresi wajah.
2	Facial Expression Recognition using Convolutional Neural Network with Data Augmentation	Tawain Uddin Ahmed, Sazzad Hossain, Mohammad	Menciptakan sistem identifikasi ekspresi wajah yang mampu mengategorikan citra ke dalam tujuh	Dalam penelitian ini, akurasi tertinggi yang berhasil dicapai adalah sebesar 95,87%.	Pada penelitian yang akan datang, diharapkan akan terjadi perkembangan dalam hal klasifikasi emosi yang lebih kompleks, seperti	Mengklasifikasi emosi wajah tanpa memanfaatkan augmentasi data. Penelitian ini akan memanfaatkan augmentasi

Tabel 2.1. Matriks Literature Review
 PENERAPAN VGG19, RESNET50 DAN RESMASKNET PADA DETEKSI EKSPRESI WAJAH

No	Judul	Peneliti, Media Publikasi, dan Tahun	Tujuan Penelitian	Kesimpulan	Saran atau Kelemahan	Perbandingan
		Shahadat Hossain, Raihan Ul Islam, Karl Andersson 2019, IEEE Xplore	kelas emosi yang berbeda.		klasifikasi gabungan seperti bahagia dan terkejut, marah dan sedih, dan lain sebagainya.	data untuk membandingkan tingkat akurasi.
3	Three convolutional neural network models for facial expression recognition in the wild	Jie Shao, Yongsheng Qian 2019, Elsevier	Mengajukan tiga model CNN inovatif dengan struktur yang berbeda, yakni model Light-CNN, model Dual - branch CNN mengekstrak fitur dari Local Binary Pattern (LBP) tradisional dan deep learning, serta CNN yang telah di-pre-trained dengan menggunakan	Hasil penelitian ini menggunakan 3 model dan dengan 3 dataset yang berbeda, menghasilkan akurasi tertinggi dengan metode light-CNN 92,86%, dual-branch CNN 85,71% dan pre-trained CNN sebesar 95,20% pada dataset CK+.	Mengimplementasikan fitur yang lebih optimal untuk mengintegrasikan dual-branch CNN. Selain itu, memanfaatkan pelatihan parameter jaringan cross-database untuk mencapai tingkat generalisasi yang lebih baik.	Melakukan perbandingan antara struktur ResNet-50 dan VGG19 dengan tujuan menentukan hasil teroptimal dari setiap metode dalam proses klasifikasi emosi. Penelitian ini menggunakan perbandingan antara struktur VGG19, ResNet-50 dan ResMaskingNet untuk mengetahui ketepatan akurasi. Dan menggunakan OneSideSelection untuk menyeimbangkan data

Tabel 2.1. Matriks Literature Review
PENERAPAN VGG19, RESNET50 DAN RESMASKNET PADA DETEKSI EKSPRESI WAJAH

No	Judul	Peneliti, Media Publikasi, dan Tahun	Tujuan Penelitian	Kesimpulan	Saran atau Kelemahan	Perbandingan
			teknik transfer learning.			
4	Recognizing learning emotion based on convolutional neural networks and transfer learning	Jason C. Hung, Kuan-Cheng Lin, Nian-Xiang Lai 2019, ScientDirect	Penelitian ini bertujuan untuk meningkatkan kinerja jaringan FaceLiveNet dalam deteksi emosi, baik pada tingkat akurasi tinggi maupun rendah.	Penelitian ini menghasilkan tingkat akurasi dengan dataset JAFFE dan KDEF, memiliki 70,02%. Dan dengan menggunakan transfer learning dari dataset FER 2013, akurasi tertinggi yang dicapai adalah 84,59%.	Disarankan kedepannya untuk dapat menggunakan rasio pembagian data lain agar dapat mengetahui perbedaan akurasi dari dataset yang digunakan.	Menggunakan dataset FER2013 dan JAFFE dan KDEF untuk melakukan kasifikasi emosi pada gambar wajah manusia. Penelitian ini menggunakan dataset FER2013 untuk melakukan kasifikasi emosi pada gambar wajah manusia.
5	Facial Expression Recognition via a CBAM Embedded Network	Wenhao Caoa, Zhuoyu Fenga, Dongyao Zhanga, Yisiyuan Huang	Penelitian ini memilih untuk memasukkan Convolutional Block Attention Module (CBAM) ke dalam beberapa lapisan jaringan VGG yang di gunakan untuk meningkatkan	Hasil penelitian ini menggunakan 2 struktur yaitu CBAM dengan akurasi 71% dan struktur VGG dengan akurasi 69%.	Untuk perkembangan penelitian mendatang, apabila data tidak cukup dan mengalami overfit terlalu cepat, diharapkan menggunakan augmentasi data.	Proses deteksi ekspresi wajah ini menggunakan struktur CBAM dengan struktur VGG. Penelitian ini menggunakan perbandingan antara struktur VGG19, ResNet-50 dan ResMaskingNet untuk mengetahui ketepatan akurasi

Tabel 2.1. Matriks Literature Review
 PENERAPAN VGG19, RESNET50 DAN RESMASKNET PADA DETEKSI EKSPRESI WAJAH

No	Judul	Peneliti, Media Publikasi, dan Tahun	Tujuan Penelitian	Kesimpulan	Saran atau Kelemahan	Perbandingan
		2020, Elsevier	akurasi dan juga stabilitas.			
6	KFSENet: A Key Frame-Based Skeleton Feature Estimation and Action Recognition Network for Improved Robot Vision with Face and Emotion Recognition	Dinh-Son Le, Hai-Hong Phan, Ha Huy Hung, Van-An Tran The-Hung Nguyen and Dinh-Quan Nguyen 2022, MDPI	Membangun sistem deteksi tindakan dengan deteksi wajah dan emosi menggunakan KFSENet.	Metode yang diusulkan adalah metode arcFace IR50 Backbone Network + Softmax Classifier dan hasil akurasi sebesar 73.04 % serta deteksi menggunakan metode Residual Masking Network berhasil menunjukkan akurasi 74.14%.	Mengimplementasikan KFSENet dalam penggunaan keyframe dan metode yang diusulkan. Menghilangkan frame yang tidak perlu dan redundan, sehingga meningkatkan akurasi klasifikasi dan mengurangi biaya komputasi.	Menggunakan struktur KFSENet dan ResMaskingNet untuk pemrosesan gambar ekspresi wajah. Penelitian ini menggunakan VGG19, ResNet-50 dan ResMaskingNet architectures untuk deteksi ekspresi wajah
7	Facial emotion recognition using convolutional neural networks (FERC)	Mehendale, N. <i>S.V Applied Sciences</i> , 2(3).	Dalam penelitian ini, bertujuan mendeteksi emosi wajah menggunakan jaringan saraf konvolusional (FERC).	Dengan menggunakan dataset FERC pada perbandingan struktur Alexnet, VGG, GoogleNet dan Resnet, dataset ini mempunyai	Pada metode ini, hanya menggunakan 24 fitur EV (expressional vector) atau bayangan pada wajah, maka diperlukan kedepannya mengoreksi otomatis pada gambar	Peneliti sebelumnya menggunakan dua skenario CNN yaitu menghilangkan latar belakang dari gambar, dan berkonsentrasi pada fitur wajah ekstraksi vektor (EV).

Tabel 2.1. Matriks Literature Review
 PENERAPAN VGG19, RESNET50 DAN RESMASKNET PADA DETEKSI EKSPRESI WAJAH

No	Judul	Peneliti, Media Publikasi, dan Tahun	Tujuan Penelitian	Kesimpulan	Saran atau Kelemahan	Perbandingan
		2020, Springer Nature		akurasi tertinggi sebesar 78%.		Penelitian ini menggunakan tahapan OneSideSelection untuk menyeimbangkan data menggunakan VGG19, ResNet-50 dan ResMaskingNet Architectures.
8	Real Time Face Mask Detection Using CNN	Rangaswamy, S. SSRN 2020, Elsevier	Penelitian ini bertujuan mengklasifikasi ekspresi wajah orang yang menggunakan masker dan tidak menggunakan masker menggunakan model MobileNet dan CNN.	Dalam penelitian ini, dua model digunakan untuk membandingkan kinerja dari segi akurasi. Model CNN memberikan akurasi sebesar 99%, sementara Model MobileNet memberikan akurasi sebesar 98%.	Diharapkan pengembangan kedepannya dapat melakukan deteksi emosi pada wajah manusia dengan menggunakan data berupa gambar, atau video.	Menggunakan penelitian bukan dengan deteksi dataset gambar, dan hanya bisa dengan real time. Penelitian ini menggunakan dataset sebagai ekstrasi deteksi ekspresi wajah.

2.3. Landasan Teori

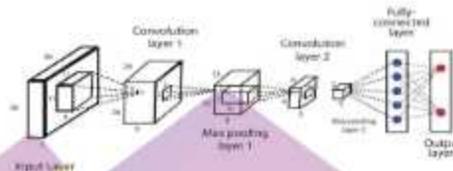
2.3.1. Arsitektur CNN

Convolutional Neural Network (CNN) adalah jenis arsitektur jaringan saraf yang dirancang khusus untuk menangani dan menganalisis data gambar. CNN sendiri merupakan sebuah konstruksi matematis yang secara umum arsitekturnya terdiri dari feature learning (*convolutional layer, activation layer, dan pooling layer*) dan structural condition classification (*flatten, fully connected layer, dan softmax*) (Wang et al., 2019).

Layer konvolusi memindai gambar dengan filter untuk mengekstrak pola-pola visual seperti tepi, sudut, dan tekstur. Layer pooling digunakan untuk mereduksi dimensi spasial dari peta fitur, mempertahankan informasi penting tanpa membebani komputasi. Fully connected layer kemudian menyatukan informasi ini untuk klasifikasi akhir. Fungsi aktivasi, seperti ReLU, diterapkan untuk memperkenalkan elemen non-linear dan meningkatkan kemampuan model untuk memahami hubungan yang kompleks dalam data.

CNN memiliki keunggulan dalam mengenali pola-pola spasial dan hierarki fitur dalam gambar, membuatnya sangat efektif dalam tugas-tugas seperti klasifikasi gambar dan deteksi objek. Model ini telah menunjukkan keberhasilan besar dalam berbagai bidang, termasuk visi komputer, deteksi wajah, dan kendaraan otonom. Keunggulan utama CNN adalah kemampuannya untuk secara otomatis mengekstrak fitur-fitur yang relevan dari data gambar tanpa memerlukan ekstraksi fitur manual, membuatnya sangat adaptif dan efisien untuk berbagai tugas

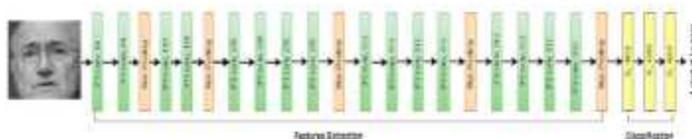
pengolahan gambar. Pada Gambar 2.1. berikut merupakan contoh arsitektur (Ali, Balamurali, & Varamini, 2022) yang nantinya akan digunakan pada penelitian ini



Gambar 2.1. Proses CNN

2.3.2. VGG19

VGG19 adalah jaringan saraf convolutional yang diperkenalkan oleh (Meena GMohbey KIndian A et al.) yang terdiri dari 19 lapisan, 16 lapisan konvolusi, dan tiga lapisan yang *fully connected*. Lapisan konvolusi tersebut memiliki kernel dengan ukuran 3×3 , dan setiap lapisannya dilapisi dengan fungsi aktivasi ReLU. Penggunaan kernel 3×3 tersebut memungkinkan model untuk mempelajari fitur-fitur yang lebih kecil dan local pada gambar, diikuti dengan lapisan max-pooling yang membantu dalam pengurangan dimensi serta mencegah terjadinya *overfitting*. Kedalaman tersebut membantu model untuk mengekstraksi fitur-fitur yang semakin kompleks dari gambar. Pada Gambar 2.2. merupakan arsitektur pada VGG-19 (Meena, Mohbey, Indian, & Kumar, 2022).



Gambar 2.2. Arsitektur VGG-19

Lapisan input awal berukuran 224×224 dengan kedalaman 3. Layer 1 dan 2 dari CNN Conv2D memiliki kedalaman 64. Kedalamannya mewakili jumlah filter yang digunakan untuk menghasilkan *feature map*. Setiap filter sesuai dengan pola yang berbeda dalam konvolusi *input* di sekitar gambar dan menghasilkan *feature map* (Vidhya & Uthra, 2021). Meskipun arsitektur VGG-19 tergolong tua, namun tetap menunjukkan kinerja yang baik dalam berbagai tugas (Kamal & Ez-zahraouy, 2023; Meena et al., 2022).

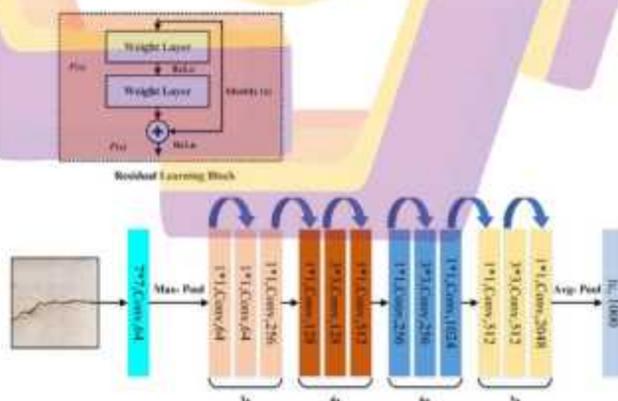
2.3.3. ResNet-50

ResNet-50 (Residual Network with 150 layers) adalah arsitektur jaringan saraf konvolusional (CNN) yang dikembangkan oleh Microsoft Research. ResNet merupakan sebuah model *Deep Neural Network* (DNN) yang dikemukakan oleh (He et al., 2016) pada IEEE Conference on Computer Vision and Pattern Recognition pada tahun 2016. Arsitektur ini terkenal karena memecahkan masalah degradasi yang dihadapi oleh jaringan yang lebih dalam, di mana penambahan layer tidak selalu meningkatkan kinerja model.

Salah satu fitur utama dari ResNet-50 adalah penggunaan blok residu atau residual block. Pada blok residu, *shortcut connections* (juga dikenal sebagai *skip connections*) dibuat untuk mengalirkan informasi langsung dari input ke lapisan keluaran, melibatkan operasi penambahan. Hal ini memungkinkan model untuk "meminimalkan" perubahan yang dilakukan oleh lapisan-lapisan di dalam blok residu. Dengan menggunakan *shortcut connections*, ResNet-50 dapat dengan efektif melatih jaringan dengan kedalaman yang lebih besar tanpa mengalami

masalah degradasi kinerja. Lapisan yang cukup dalam memungkinkan model untuk mengekstraksi fitur-fitur hierarkis yang sangat kompleks dari gambar yang rumit.

ResNet-50 terdiri dari beberapa blok residu yang masing-masing terdiri dari beberapa lapisan konvolusi dan batch normalization. Berkat adanya blok residu, ResNet50 cenderung lebih mudah untuk melatih dan memiliki kemampuan generalisasi yang lebih baik pada dataset yang beragam. Selain itu, dalam ResNet-50, terdapat juga penggunaan teknik pengurangan dimensi (downsampling) pada blok residu tertentu menggunakan konvolusi 1x1 untuk memastikan bahwa dimensi feature map dapat diatur dengan baik. Kombinasi dari konsep residual learning, kedalaman, kemampuan generalisasi, dan kinerjanya membuat ResNet50 menjadi arsitektur yang sangat baik untuk berbagai tugas klasifikasi gambar (Halim et al., 2023; Kamal & Ez-zahraouy, 2023). Pada Gambar 2.3. merupakan gambaran dari arsitektur ResNet-50 (Millenia, Naufal, & Siswanto, 2022).



Gambar 2.3. Arsitektur ResNet-50

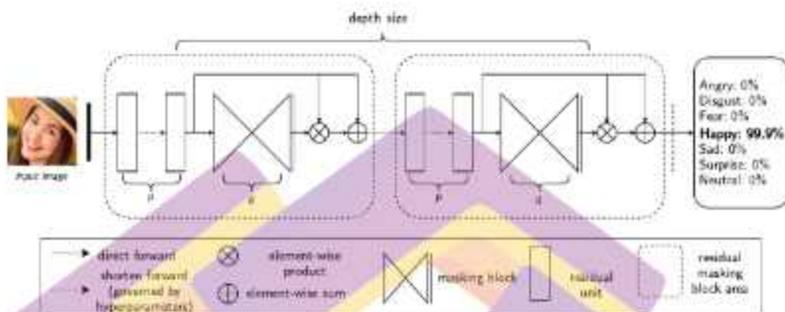
2.3.4. Residual Masking Network

The Residual Masking Network (ResMaskingNet) adalah model jaringan saraf konvolusional end-to-end yang menggabungkan deep residual networks dengan blok masking. Blok masking membantu memfokuskan perhatian model pada wilayah lokal yang penting untuk menyempurnakan peta fiturnya agar dapat membuat prediksi yang lebih rinci, dan struktur residual membantu menjaga kinerja dalam lapisan yang lebih dalam (Cheong J et al, 2021).

ResMaskingNet bertujuan adalah untuk menghasilkan foto berkualitas tinggi dengan menghilangkan gangguan atau objek yang tidak diinginkan dari gambar asli. Model ini menggunakan konsep residual learning yang mengarah pada peningkatan kinerja model dengan cara mempelajari selisih antara prediksi model dan gambar asli. Dalam kasus ini, selisih tersebut dapat berupa bagian gambar yang ingin dihilangkan. ResMaskingNet terdiri dari beberapa lapisan konvolusi yang digunakan untuk mengekstraksi fitur gambarnya. Selain itu, model ini juga menggunakan teknik "masking" untuk memisahkan objek yang tidak diinginkan dari latar belakang dan menghilangkannya. Dengan menggunakan Residual Masking Network, gambar hasilnya memiliki kualitas yang lebih baik dan lebih bersih, dengan gangguan atau objek yang tidak diinginkan dihilangkan secara efektif.

Metode yang diusulkan dari alur ResMaskingNet adalah jaringan yang terdiri dari empat blok utama Residual Masking (Resmasking Blocks). Setiap blok Residual Masking, yang mengoperasikan ukuran fitur yang berbeda, terdiri dari

Lapisan Residual dan Blok Masking. Pada Gambar 2.4. merupakan arsitektur dari ResMaskNet (Pham et al., n.d.).



Gambar 2.4. ResMaskNet

2.3.5. One Sided Selection

One Sided Selection merupakan metode pengambilan sampel ulang (re-sampling) yang ditargetkan dengan mencoba menyeimbangkan proporsi kelas. Teknik One Sided Selection mengambil sampel dibawah kelas mayoritas, dengan menghapus contoh-contoh yang dianggap noise, melewati batas, atau berlebihan. Diasumsikan bahwa contoh yang noise membanjiri kelas minoritas, sehingga dengan menghapusnya, dapat meningkatkan kinerja dari pembelajaran dengan data yang tidak seimbang. Singkatnya, metode One Sided Selection yaitu metode undersampling yang mencoba mengurangi sample dari kelas mayoritas untuk membuat dataset lebih seimbang (Ding et al., 2020).

2.3.6. Deteksi

Deteksi ekspresi wajah adalah proses mengenali dan menginterpretasi emosi atau ekspresi yang terlihat pada wajah seseorang melalui penggunaan teknologi komputer atau sistem kecerdasan buatan. Metode ini secara umum

mencakup teknik deteksi pola, pengolahan citra, dan pembelajaran mesin untuk mengidentifikasi ekspresi wajah seperti senang, sedih, marah, takut, terkejut, jijik, atau netral.

Ekspresi wajah, sebagai salah satu sarana terpenting bagi manusia untuk mengekspresikan emosi dan konotasi dalam proses komunikasi, memainkan peran penting dalam pengembangan sistem interaksi manusia-komputer. Dalam beberapa tahun terakhir, deteksi ekspresi wajah (Facial Expression Recognition/ FER) telah menjadi topik penelitian penting karena aplikasi-aplikasinya yang beragam, seperti dalam perawatan kesehatan dan animasi berbasis data. Tujuan dari sistem FER adalah untuk mengidentifikasi keadaan emosional manusia berdasarkan gambar wajah yang diberikan (Hui Ma and Turgay Celik, 2019)

Deteksi ekspresi wajah penting dalam banyak aplikasi, termasuk deteksi emosi, analisis video, manajemen interaksi manusia-komputer, keamanan, atau game interaktif. Teknologi deteksi ekspresi wajah ini dapat digunakan dalam berbagai jenis sistem, mulai dari kamera CCTV hingga kamera pada perangkat ponsel yang memungkinkan penggunaan pada berbagai platform.

Untuk memperoleh informasi ekspresi wajah, sistem deteksi ekspresi wajah menggunakan algoritma untuk menganalisis fitur-fitur wajah yang dikumpulkan dari video atau gambar, seperti mata, mulut, hidung, atau kerutan di wajah. Algoritma ini kemudian membandingkan fitur-fitur tersebut dengan pola yang telah dipelajari sebelumnya untuk mengidentifikasi ekspresi wajah yang tepat.

Dalam beberapa tahun terakhir, dengan perkembangan teknologi kecerdasan buatan yang lebih canggih, deteksi ekspresi wajah menjadi lebih akurat dan dapat mengenali ekspresi dengan tingkat ketepatan yang tinggi. Teknologi ini terus dikembangkan dan diimplementasikan untuk berbagai aplikasi yang berhubungan dengan interaksi manusia dan mesin.



BAB III

METODE PENELITIAN

3.1. Jenis, Sifat, dan Pendekatan Penelitian

Penelitian ini merupakan jenis penelitian eksperimental yang melibatkan serangkaian langkah, mulai dari pengumpulan data, pra-pemrosesan data, ekstraksi fitur, hingga pemodelan data menggunakan berbagai metode, dan kemudian melalui proses pelatihan dan pengujian. Sifat dari penelitian yang dilakukan dalam hal ini adalah mandiri, menggunakan metode deskriptif terhadap data yang telah terkumpul, dilanjutkan dengan pelatihan dan pengujian. Tujuan dari pengujian adalah untuk menemukan akurasi ekspresi wajah dari hasil yang diperoleh.

Penerapan metode yang digunakan dalam penelitian ini adalah pendekatan kuantitatif untuk mengukur tingkat akurasi dalam deteksi ekspresi wajah. Hasil penelitian ini akan mencerminkan tingkat keberhasilan dalam kecepatan akurasi dan waktu mengenali tujuh ekspresi wajah yang sedang diteliti yang akan ditampilkan dalam bentuk confusion matrix. Dalam proses deteksi ekspresi wajah, digunakan metode Convolutional Neural Networks (CNN) serta menerapkan Residual Masking Network pada data yang telah dikumpulkan, serta menggunakan arsitektur Resnet-50 untuk pembandingan guna mengetahui akurasi tertinggi dari deteksi ekspresi wajah.

3.2. Metode Pengumpulan Data

Langkah awal dalam penelitian adalah pengumpulan data sebelum tahap analisis data dimulai. Pengumpulan data merupakan langkah strategis dalam penelitian. Data yang diperlukan untuk penelitian ini adalah data deteksi ekspresi wajah dengan berbagai variasi ekspresi. Untuk tujuan penelitian ini, data ekspresi wajah diperoleh dari platform Kaggle dalam Facial Expression Recog Image Ver of (FERC) Dataset. Dataset terdiri dari 35.888 citra wajah dengan variasi ekspresi yang terbagi dalam tujuh kategori. FERC mencakup data deteksi ekspresi emosi pada gambar wajah yang memiliki variasi kompleks dalam kondisi pencahayaan, sudut pandang, dan latar belakang. Hal ini dapat memberikan tantangan tambahan dan membantu meningkatkan kinerja model ResMaskingNet di situasi yang lebih realistis. Ciri-ciri ekspresi dari ketujuh kategori yang digunakan dapat dilihat pada Tabel 3.1.

Tabel 3.1. Ciri ciri Ekspresi Wajah

No	Kategori	Ciri-ciri
1	Senang	<ul style="list-style-type: none"> ● Senyuman dengan bibir yang terangkat. ● Mata yang melebar dan mungkin muncul kerutan di sekitar mata. ● Positifnya ekspresi di seluruh wajah.
2	Sedih	<ul style="list-style-type: none"> ● Posisi dan bentuk mulut yang melipat ke bawah. ● Penurunan ujung alis dan mata yang terlihat sayu. ● Adanya kerutan di sekitar mata dan dahi.
3	Marah	<ul style="list-style-type: none"> ● Alis yang ditekuk dan mungkin mendekati satu sama lain. ● Tekanan di sekitar area mata dan garis wajah yang tegang. ● Bibir yang mungkin terlipat atau menonjol.
4	Terkejut	<ul style="list-style-type: none"> ● Mata yang melebar dan mungkin terbuka lebar. ● Mulut yang terbuka atau sedikit terbuka. ● Mungkin ada anggukan atau posisi tubuh yang menunjukkan kejutan.
5	Takut	<ul style="list-style-type: none"> ● Mata yang melebar dengan pupil yang mungkin membesar. ● Alis yang meninggi dan mungkin terpisah. ● Mungkin ada tegangan di area mulut.
6	Jijik	<ul style="list-style-type: none"> ● Lipatan di hidung atau gerakan hidung yang menunjukkan ketidaknyamanan. ● Kekokohan di sekitar area mulut atau bibir yang terangkat.
7	Netral	<ul style="list-style-type: none"> ● Wajah tanpa ekspresi khusus, dengan bibir dan mata dalam posisi biasa. ● Alis dan mulut dalam posisi relatif rileks. ● Tidak ada tanda-tanda khusus dari ekspresi emosi tertentu.

Dengan menggunakan dataset ini, peneliti dapat melakukan pengujian terhadap model Residual Masking Network. Tujuh ekspresi emosi wajah (Mahendar, 2023) pada tabel diatas bisa dilihat pada Gambar 3.1. berikut.



Gambar 3.1. Ekspresi Wajah Dalam Dataset FER2011

Dalam penelitian ini, dilakukan analisis data pada dataset yang telah disusun untuk penelitian, terdiri dari 35.888 data deteksi ekspresi wajah yang terbagi ke dalam tujuh kategori: senang, sedih, marah, terkejut, netral, jijik, takut. Data terdiri dari gabungan dua dataset train dengan total 32.299 dan dataset test dengan total 3.589 (Tabel 3).

Tabel 3.2. Distribusi Dataset

No	Kategori	Emosi	Jumlah
1	Dataset Train	Senang	8110
2		Sedih	5483
3		Marah	4462
4		Terkejut	3586
5		Takut	4593
6		Jijik	493
7		Netral	5572
8	Dataset Testing	Senang	879
9		Sedih	594
10		Marah	491
11		Terkejut	416
12		Takut	528
13		Jijik	55
14		Netral	626

Pada tabel diatas merupakan persebaran dataset. Setelah dataset yang diperlukan terkumpul, langkah selanjutnya adalah melakukan proses pra-pemrosesan data menggunakan Google Colab. Data yang berhasil pre-processing

kemudian akan dialokasikan untuk dibagi menjadi data *training* dan data *testing*. Proses alokasi akan berubah sesuai dengan nilai yang dihasilkan. Selanjutnya persiapan pemodelan dalam prediksi data. Hasil dari pemodelan akan di latih dan diuji yang nantinya didapatkan akurasi paling tinggi.

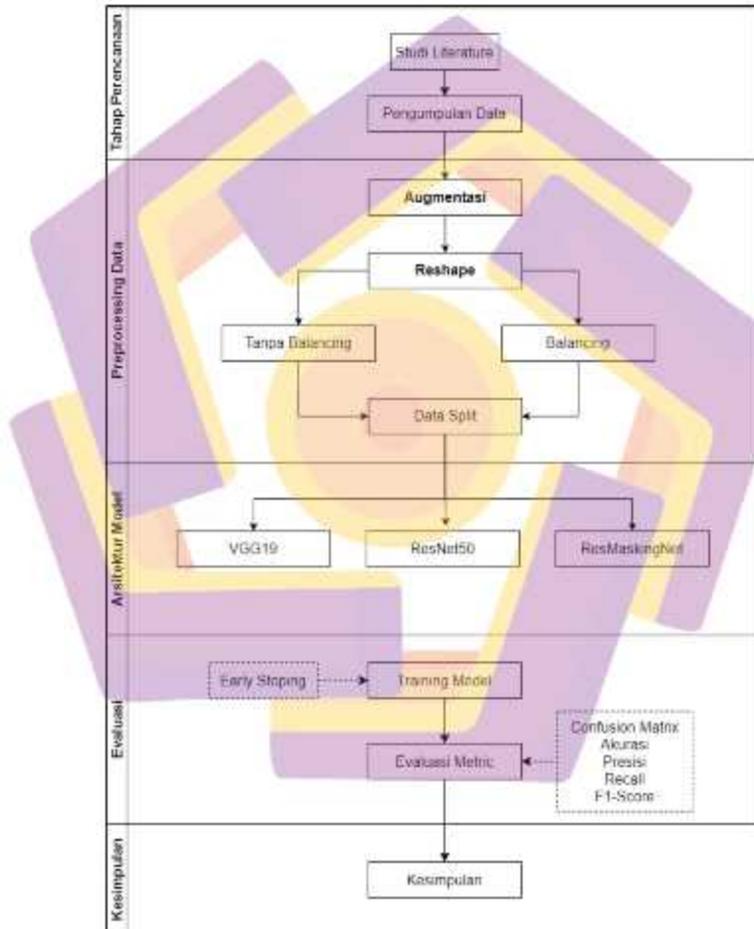
Pada penelitian ini menambahkan lapisan konvolusi atau blok residual tambahan untuk meningkatkan kapasitas representatif model. Ini dapat membantu model memahami pola-pola yang lebih kompleks pada gambar wajah.

Dalam proses pemodelan, perbandingan dengan model ResMaskingNet, ResNet50 dan VGG19 diharapkan mampu memberikan hasil yang lebih baik. Proses deteksi dievaluasi tidak hanya dari segi keakuratan tetapi juga kecepatan komputasi. Meskipun ResMaskingNet dapat memberikan hasil akurat, waktu komputasi yang diperlukan tetap dalam batas waktu yang dapat diterima. Penggunaan struktur ResNet50 dan VGG19 dalam pendekatan ini bertujuan untuk memaksimalkan perbandingan akurasi deteksi ekspresi wajah dengan mempertimbangkan kontribusi dari setiap elemen yang terlibat, baik data asli maupun data yang telah di augmentasi.

Beberapa faktor yang mempengaruhi akurasi deteksi ekspresi wajah pada metode Residual Masking Network adalah pengoptimalan ResMaskingNet dalam kompleksitas struktur jaringan yang mempengaruhi akurasi, kualitas data latih (*training*), dan parameter fine-tuning (*penyempurnaan*) yang digunakan dalam proses pengembangan Pada model yang sudah ada atau yang telah dilatih sebelumnya pada tugas terkait. Ini dapat membantu model menyesuaikan diri dengan karakteristik spesifik dari dataset deteksi ekspresi wajah. Dengan

memperhatikan aspek-aspek ini, teknologi deteksi ekspresi wajah berbasis ResMaskingNet dapat ditingkatkan untuk memberikan hasil yang lebih baik.

3.3. Alur Penelitian



Gambar 3.2. Alur Penelitian

Berdasarkan pada Gambar 3.2 alur penelitian diatas, pada penelitian ini akan melewati beberapa tahapan-tahapan sebagai berikut:

1. Pada tahap awal penelitian, langkah pertama yaitu melakukan studi literatur. Proses ini dilakukan dengan membaca buku dan jurnal yang berkaitan dengan masalah yang akan diteliti untuk mendapatkan pemahaman yang lebih baik tentang konteks masalah yang akan diteliti. Kemudian dilanjutkan dengan mengumpulkan data yang akan digunakan untuk penelitian. Pengumpulan data dilakukan dengan mengambil dataset dari Kaggle. Dataset yang digunakan pada penelitian ini adalah FERC yang memiliki 35.888 gambar.
2. Tahap kedua dalam penelitian ini adalah *Preprocessing* data. Tahapan ini merupakan tahap krusial dalam mempersiapkan gambar sebelum memasuki proses pelatihan. Terdapat beberapa tahapan dalam *preprocessing* data, yaitu augmentasi data, *reshape*, *Balancing*, dan *Split data*.

- a. Augmentasi Data

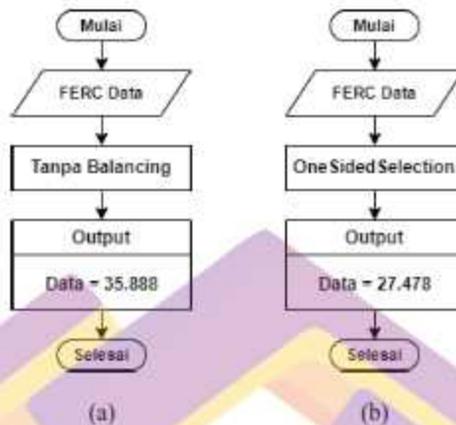
Tahap *preprocessing* yang pertama yaitu augmentasi data. Augmentasi data digunakan untuk memperluas jumlah data pelatihan dengan membuat variasi baru dari data yang sudah ada. Augmentasi terlebih dahulu dilakukan sebelum gambar memasuki proses training. Augmentasi yang dilakukan yaitu rotation, shear, width shift range, height shift range, zoom, dan horizontal flip.

b. *Reshape*

Reshape pada penelitian ini digunakan untuk mengubah bentuk array satu dimensi menjadi matriks dua dimensi yang merepresentasikan gambar wajah. Proses ini dilakukan dengan tujuan agar data gambar dapat diproses lebih lanjut oleh model yang digunakan.

c. *Balancing*

Tahapan *balancing* adalah salah satu teknik dalam pengolahan data yang digunakan untuk menyeimbangkan distribusi kelas yang tidak seimbang dalam dataset. Teknik *OneSideSelection* dipilih untuk mengurangi jumlah sample dalam kelas mayoritas, sambil mempertahankan sampel di kelas minoritas. Dengan teknik ini diharapkan model pembelajaran mesin dapat dilatih dengan lebih baik karena lebih memperhatikan kelas minoritas dan menghindari bias yang terlalu condong ke arah kelas mayoritas, sehingga dapat digunakan untuk meningkatkan kinerja model terutama dalam kasus distribusi kelas sangat tidak seimbang.



Gambar 3.3. Proses Balancing

Sebelum (a), dan Sesudah (b) diterapkan Balancing.

Pada Gambar 3.3 merupakan proses balancing data yang dilakukan pada dataset FERC. Pada perlakuan tanpa balancing, dataset tetap pada nilai yang sebenarnya yaitu sebanyak 35.888 data. Setelah diterapkan balancing dengan OneSidedSelection, dataset pada FERC menurun hingga menjadi 27.478 data. Setelah penerapan teknik balancing tersebut, menunjukkan perubahan jumlah total data yang ada. Secara umum, OSS bekerja dengan cara mengurangi jumlah sampel dari kelas minoritas untuk mencapai keseimbangan dalam dataset. OSS sendiri secara selektif menghapus Sebagian sampel dari kelas mayoritas untuk menciptakan keseimbangan antara kelas mayoritas dan kelas minoritas. Dengan begitu, model dapat lebih focus pada fitur-fitur yang relevan dari kelas minoritas, sehingga meningkatkan kinerja model klasifikasi.

d. Split Data

Setelah data melalui augmentasi data, dilakukan pembagian data sebesar 80% untuk data training, 10% validasi data, dan 10% data testing.

3. Tahapan selanjutnya melibatkan proses pelatihan data training dari dataset FERC yang telah selesai dilakukan preprocessing. Proses pelatihan ini dilakukan dengan 2 skenario yang berbeda dengan menggunakan arsitektur VGG19, ResNet-50, dan ResMaskingNet. Skenario pertama hanya menggunakan model arsitektur asli tanpa dilakukan balancing data. Skenario yang kedua menggunakan model arsitektur dengan balancing data.
4. Memasuki tahapan yang keempat yaitu evaluasi. Pada tahap pelatihan, parameter model seperti optimizer, batch size, dan learning rate disesuaikan. Untuk mencegah terjadinya overfitting serta mengoptimalkan performa model, early stopping diterapkan. Dengan diterapkannya early stopping, diharapkan model dapat mengoptimalkan performa dengan menghentikan pelatihan pada titik yang tepat, dan menghemat waktu pelatihan. Setelah training model selesai, dilakukan evaluasi model. Metode pengujian menggunakan Confusion Matrix untuk mendapatkan nilai accuracy, precision, recall, dan F1-Score.
5. Tahapan terakhir dalam penelitian ini adalah pengambilan kesimpulan. Setelah melalui tahapan studi literatur hingga evaluasi. Pada tahap ini, temuan hasil dari setiap skenario dipaparkan dengan membuat kesimpulan, dan hasil eksperimen yang telah dilakukan dan disajikan.

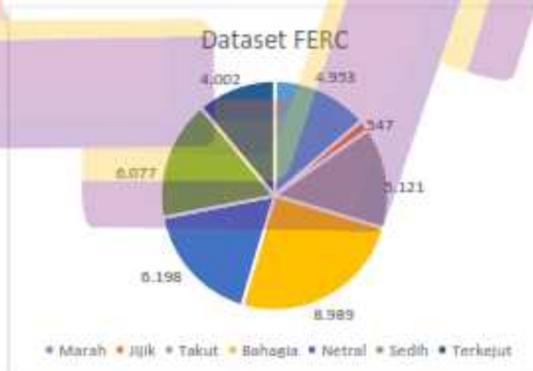
BAB IV

HASIL PENELITIAN DAN PEMBAHASAN

Pada bab IV ini berisi pemaparan hasil penelitian dan pembahasan yang telah dilakukan oleh peneliti secara terperinci mengenai langkah-langkah untuk menjawab rumusan masalah.

4.1. Pengumpulan Data

Dalam penelitian ini data yang digunakan adalah dataset FEREC yang diperoleh dari Kaggle dengan jumlah gambar 35.888, terbagi kedalam tujuh kelas yang berbeda. Dataset ini telah menjadi subjek beberapa penelitian sebelumnya. Distribusi data pada dataset FEREC tidak merata mencakup 4.953 data untuk kelas marah, 547 data untuk kelas jijik, 5.121 data untuk kelas takut, 8.989 data untuk kelas bahagia, 6.198 data untuk kelas netral, 6.077 data untuk kelas sedih, dan 4.002 data untuk kelas terkejut. Hal ini menunjukkan adanya ketidakseimbangan dalam data. Distribusi dari dataset dapat dilihat pada Gambar 4.1 berikut.



Gambar 4.1. Distribusi Dataset FEREC

4.2. Skenario Percobaan

Selanjutnya, akan dilakukan skenario eksperimen terhadap dataset FERC berdasarkan analisis data sebelumnya, Skenario eksperimen ini akan diterapkan pada tiga arsitektur, yaitu VGG19, ResNet-50, dan ResMaskingNet. Tujuan dari penelitian ini adalah untuk mendapatkan pemahaman yang lebih baik mengenai bagaimana model dapat memengaruhi akurasi kinerja model dalam dataset FERC. Terdapat 2 skenario berbeda yang akan diuji. Setiap skenario akan melibatkan proses pelatihan model selama 100 epoch. Selama proses pelatihan, data dibagi menjadi 80% untuk training, dan 10% untuk validasi, dan sisanya untuk testing. Detail dari skenario yang dilakukan dapat dilihat pada Tabel 4.1.

Tabel 4.1. Skenario Percobaan

No	Skenario	Arsitektur		
		VGG19	ResNet50	ResMaskingNet
1	S1	VGG19	ResNet50	ResMaskingNet
2	S2	VGG19 + Balancing	ResNet50 + Balancing	ResMaskingNet + Balancing

Pada tabel tersebut dapat dilihat penggunaan skenario dengan tiga arsitektur berbeda. Pada skenario 1 menggunakan arsitektur model tanpa balancing data. Pada skenario 2 melibatkan penggunaan arsitektur model bersamaan dengan teknik balancing.

4.3. Preprocessing Data

Langkah selanjutnya yaitu preprocessing data. Preprocessing data merupakan langkah penting dalam proses ekstraksi, dan analisis data. Dataset FERC diubah menjadi format yang dapat dimengerti dan dianalisis oleh komputer.

Pada sintaks kode tersebut, memuat parameter yang digunakan untuk augmentasi data. Augmentasi ini bertujuan untuk meningkatkan variasi data pelatihan dengan membuat variasi baru dari data yang ada. Sehingga dapat membantu model dalam mempelajari pola yang lebih umum dan lebih tahan terhadap variasi dalam data masukan. Parameter yang digunakan yaitu `rotation_range`, parameter ini mengontrol rotasi secara acak dengan rentang 15 derajat. Parameter `shear_range` berguna untuk mengontrol seberapa banyak gambar yang dapat diubah bentuknya secara acak dengan menggeser gambar hingga 15% ukuran aslinya. Parameter `zoom_range` ini berarti gambar diperbesar atau diperkecil hingga 15% dari ukuran aslinya. Parameter `width_shift_range` dan `height_shift_range`, mengontrol rentang pergeseran horizontal dan vertical sebanyak 15%. Parameter terakhir yaitu `horizontal_flip`, parameter ini melakukan flip horizontal secara acak. Metode ini dapat membantu model untuk belajar pola-pola yang lebih umum dan dapat mengurangi terjadinya overfitting.

4.4.2. Reshape

Selanjutnya dilakukan proses reshape pada data. Proses reshape ini bertujuan untuk mengubah data gambar yang disimpan dalam bentuk string menjadi bentuk array 2D dengan ukuran 48x48 pixel. Berikut merupakan sintaks yang digunakan untuk reshape.

```
img_array = df.pixels.apply(lambda x: np.array
(x.split('')).reshape(48, 48).astype('float32'))
img_array = np.stack(img_array, axis = 0)

img_array.shape
```

Pertama, kode tersebut menggunakan metode `apply` dari `pandas DataFrame` (`df`) untuk menerapkan sebuah fungsi lambda pada setiap elemen dalam kolom `pixels`. Fungsi lambda tersebut mengambil setiap string yang dipisahkan oleh spasi dalam kolom `pixels`, mengubahnya menjadi array numpy dengan `np.array(x.split(' '))`, kemudian melakukan `reshape` array tersebut menjadi bentuk 2D dengan ukuran `48x48` pixel menggunakan `.reshape(48, 48)`. Akhirnya, semua nilai dalam array tersebut dikonversi menjadi tipe data `float32` dengan `.astype('float32')`. Setelah dilakukan `reshape` pada setiap elemen dalam kolom `pixels`, kode tersebut menggunakan `np.stack` untuk menggabungkan semua array menjadi satu array tunggal. `axis=0` digunakan untuk menentukan bahwa penggabungan dilakukan pada sumbu pertama, yaitu sumbu baris. Terakhir, kode tersebut mencetak bentuk (`shape`) dari array hasil penggabungan. `Shape` ini akan memberikan informasi tentang jumlah data gambar (baris) dan dimensi gambar (kolom), dalam hal ini berupa (`jumlah_data, 48, 48`), di mana `48x48` merupakan ukuran gambar yang telah direshape.

4.4.3. *Balancing*

Proses selanjutnya melakukan `balancing data` dengan menggunakan teknik `undersampling`. `Undersampling` adalah salah satu teknik dalam pemrosesan data yang digunakan untuk menangani masalah ketidakseimbangan kelas dalam dataset. Dalam konteks kelas yang tidak seimbang, kelas mayoritas memiliki jumlah sampel yang jauh lebih banyak daripada kelas minoritas. Ketidakseimbangan kelas dapat menjadi masalah serius dalam pembelajaran mesin, karena model yang dibangun cenderung memiliki bias terhadap kelas mayoritas, yang dapat

menghasilkan performa yang buruk dalam mengklasifikasikan kelas minoritas. Dengan menerapkan teknik ini, dapat digunakan untuk mengurangi jumlah sampel dari kelas mayoritas sehingga distribusi antara kelas mayoritas dan minoritas menjadi lebih seimbang. Berikut merupakan sintaks dalam perlakuan balancing data.

```
from imblearn.under_sampling import OneSidedSelection
oversampler = OneSidedSelection(random_state=42)
X_train_balanced, y_train_balanced =
oversampler.fit_resample(X_train.reshape(len(X_train), -1),
np.argmax(y_train, axis=1))
X_train_balanced = X_train_balanced.reshape(-1, img_width,
img_height, img_depth)
y_train_balanced = y_train_balanced.reshape(-1, 1)
y_train_balanced = to_categorical(y_train_balanced, num_classes)
Print(X_train_balanced.shape, y_train_balanced.shape)
```

Sintaks tersebut digunakan untuk mengatasi masalah ketidakseimbangan kelas dalam data pelatihan. Hal ini dikarenakan jumlah sampel dalam data FERC tidak seimbang, sehingga dapat menyebabkan model memiliki kecenderungan dalam mempelajari kelas mayoritas dengan baik sehingga mengabaikan kelas minoritas. Pada penelitian ini, untuk mengatasi hal tersebut dilakukan proses undersampling menggunakan metode 'OneSideSelection'. Teknik ini bertujuan untuk meningkatkan jumlah sampel pada kelas minoritas dengan menghasilkan sampel sintesis yang serupa dengan sampel sample pada kelas minoritas, sementara tetap menjaga jumlah sampel pada kelas mayoritas agar tidak berlebihan. Dengan teknik tersebut, dapat digunakan untuk mengatasi ketidakseimbangan kelas dalam dataset sehingga model yang dihasilkan dapat belajar dengan lebih baik dari kelas minoritas dan menghasilkan prediksi yang lebih akurat.

4.4.4. Split Data

Selanjutnya, data dibagi ke dalam set *training*, *validation*, dan *testing*. Metode ini biasanya digunakan untuk mengevaluasi kinerja model, dan mencegah *overfitting*. Rasio yang paling umum digunakan yaitu 80% untuk *training*, 10% untuk *validation*, dan 10% untuk *testing*. Set *training* merupakan bagian terbesar dari data dan digunakan untuk melatih model. Hal ini dikarenakan model menggunakan pola dan fitur dalam set *training* untuk membuat prediksi tentang data yang belum pernah dilihat sebelumnya. Set validasi digunakan untuk menyempurnakan *hyperparameter* model dan menilai kinerjanya selama proses pelatihan. Set validasi dapat membantu mencegah *overfitting* dengan memberikan evaluasi yang tidak biasa terhadap kinerja model. Terakhir set *testing* digunakan untuk menilai kinerja akhir dari model yang dilatih. Ini memberikan evaluasi yang tidak bias terhadap kemampuan model untuk menggeneralisasi data baru.

4.4. Arsitektur Model

Langkah selanjutnya setelah melewati *augmentasi* dan *balancing* yaitu menggunakan arsitektur CNN untuk mengekstraksi fitur yang dapat digunakan untuk tugas khusus seperti klasifikasi, segmentasi, deteksi, dan tugas lainnya. Penelitian ini menggunakan arsitektur model VGG19, ResNet50 dan ResMaskingNet. Proses pembelajaran (*training*) melewati beberapa tahapan, untuk ekstraksi fitur dilakukan pada *convolutional layer*, dan *pooling layer*. Untuk proses deteksi dilakukan pada *fully connected layer*.

Sebelumnya, arsitektur tersebut telah dilakukan *fine-tuning* dengan dataset ImageNet. Ini merujuk pada proses dalam pembelajaran mesin dimana model yang

telah dilatih sebelumnya (pre-trained model) disesuaikan kembali untuk melakukan tugas spesifik yang berbeda untuk meningkatkan kinerja model pada dataset baru. Dengan fine-tuning, model pre-trained dapat disesuaikan untuk mempelajari representasi-fitur yang lebih spesifik untuk tugas baru, sambil tetap memanfaatkan pengetahuan yang sudah ada dari pelatihan sebelumnya. Hal ini dapat menghasilkan model yang lebih baik dan lebih efisien dalam menyelesaikan tugas yang baru.

4.4.1. Arsitektur VGG19

Arsitektur VGG19 merupakan salah satu arsitektur Convolutional Neural Network (CNN) yang sangat terkenal dalam pengolahan citra. Arsitektur ini terdiri dari sejumlah besar lapisan konvolusi. Setiap lapisan konvolusi diikuti oleh lapisan aktivasi ReLU, yang membantu dalam mengekstraksi fitur-fitur yang bersifat non-linear. Lapisan konvolusi ini bertanggung jawab untuk mengekstraksi fitur-fitur dari gambar input dengan menerapkan filter konvolusi. Pada VGG19 terdapat lapisan input layer, yaitu lapisan yang menerima gambar sebagai input dengan ukuran yang telah ditetapkan. Arsitektur ini terdiri dari sejumlah besar lapisan konvolusi yang diikuti oleh lapisan activation ReLU. Lapisan ini untuk mengekstraksi fitur-fitur dari gambar input dengan menerapkan filter konvolusi. Selanjutnya pooling layer, lapisan ini untuk mengurangi dimensi fitur serta menyederhanakan representasi yang dihasilkan. Kemudian Fully Connected Layer, merupakan lapisan yang terdiri dengan beberapa neuron yang terhubung penuh, dan bertugas untuk menghubungkan fitur-fitur yang diekstraksi dengan label kelas yang diinginkan. Ini digunakan untuk tugas klasifikasi gambar. lapisan terakhir yaitu Output Layer, pada lapisan ini, lapisan fully connected dihubungkan ke lapisan

softmax yang menghasilkan probabilitas prediksi untuk setiap kelas. Berikut merupakan syntax yang dilakukan untuk mengubah dua lapisan terakhir pada arsitektur VGG19.

```
def build_model(bottom_model, classes):
    model = bottom_model.layers[-2].output
    model = GlobalAveragePooling2D()(model)
    model = Dense(classes, activation='softmax',
name='out_layer')(model)

    return model
```

Sintaks tersebut digunakan untuk membangun model klasifikasi menggunakan arsitektur dasar (*bottom_model*) dengan menambahkan lapisan global average pooling dan lapisan dense softmax. Dengan mengambil output dari lapisan kedua terakhir dari model dasar. Lapisan yang di ambil merupakan lapisan pooling sebelum lapisan fully connected pada VGG19. Selanjutnya menambahkan lapisan *GlobalAveragePooling2D* untuk meratakan fitur-fitur spasial menjadi satu vector fitur rata-rata. Yang terakhir yaitu menambahkan lapisan dense softmax untuk melakukan klasifikasi dengan jumlah kelas sebanyak '*classes*'.

Secara singkat, VGG19 mengekstraksi fitur-fitur dengan memanfaatkan lapisan konvolusi untuk mengenali pola-pola signifikan dalam gambar. Lalu, melalui lapisan pooling, dimensi fitur tersebut dikurangi, dan kemudian, melalui lapisan fully connected, fitur-fitur tersebut dihubungkan dengan label kelas. Akhirnya, model menghasilkan probabilitas prediksi untuk masing-masing kelas, yang dapat digunakan untuk mengklasifikasi gambar.

4.4.2. Arsitektur ResNet50

Arsitektur ResNet50 merupakan salah satu arsitektur Convolutional Neural Network (CNN) yang sangat terkenal dan berpengaruh dalam pengolahan citra. ResNet50 terdiri dari 50 lapisan, termasuk lapisan konvolusi, batch normalization, aktivasi ReLU, dan lapisan pooling, serta lapisan-lapisan shortcut. Kedalaman yang dimiliki memungkinkan arsitektur untuk mengekstraksi fitur yang semakin kompleks dan abstrak dari gambar, untuk tugas-tugas seperti klasifikasi dan deteksi objek. Arsitektur ResNet50 menggunakan blok residual yang terdiri dari beberapa lapisan konvolusi bertumpuk. Setiap blok residual memiliki dua cabang utama yang berisi lapisan konvolusi dan cabang shortcut yang melewati blok tanpa perubahan atau penyesuaian dimensi. Dengan adanya blok residual memungkinkan jaringan untuk beradaptasi dengan baik terhadap dataset yang beragam dan mengatasi masalah gradien yang menghilang (vanishing gradient) yang sering terjadi dalam jaringan yang dalam. Berikut merupakan sintak yang dilakukan untuk mengubah lapisan terakhir pada arsitektur ResNet50.

```
def build_model(bottom_model, classes):
    model = bottom_model.layers[-2].output
    model = GlobalAveragePooling2D()(model)
    model = Dense(classes, activation='softmax',
name='out_layer')(model)

    return model
```

Pada sintak tersebut merupakan sebuah fungsi untuk membangun model klasifikasi menggunakan arsitektur dasar ResNet50 dengan menambahkan dua lapisan di akhir model untuk menyesuaikan jumlah kelas output. Dapat dilihat pada kode `model = bottom_model.layers[-2].output` yang berarti

mengambil dua lapisan terakhir dalam model ResNet50, yang kemudian digantikan menjadi lapisan *Global Average Pooling 2D* dan Lapisan *Dense* dengan menggunakan aktivasi softmax dan diterapkan untuk tujuh kelas. Dengan demikian, fungsi ini memungkinkan untuk mengambil model paling dasar pada ResNet50 dan menyesuaikannya dengan jumlah kelas output.

4.4.3. ResMaskingNet

ResmaskingNet adalah sebuah arsitektur jaringan saraf konvolusi (CNN) yang menggabungkan konsep dari dua pendekatan yang berbeda, yaitu residual learning (ResNet) dan masking blocks. Dalam ResmaskingNet, konsep residual learning digunakan untuk mengatasi masalah gradien yang menghilang atau meledak selama pelatihan jaringan yang dalam jaringan yang sangat dalam. Sementara itu, konsep masking blocks digunakan untuk meningkatkan kemampuan jaringan dalam menangkap fitur-fitur yang kompleks dari gambar. Code di atas mendefinisikan ResmaskingNet menggunakan TensorFlow dan Keras. Pertama, sebuah blok masking dan blok Resmasking diimplementasikan menggunakan fungsi `masking_block` dan `resmasking_block`. Selanjutnya, model CNN dibangun menggunakan blok-blok Resmasking tersebut, di mana lapisan-lapisan konvolusi dan pooling diikuti oleh beberapa blok Resmasking yang bertujuan untuk mengekstraksi fitur-fitur dari gambar input. Setelah itu, model dikompilasi dengan menggunakan optimizer Adam, fungsi loss categorical cross-entropy, dan akurasi sebagai metrik evaluasi. Hasilnya adalah sebuah model CNN yang siap untuk dilatih dan dievaluasi menggunakan dataset yang sesuai. Berikut merupakan sintak yang digunakan untuk membuat model ResMaskNet.

```

import tensorflow as tf
from tensorflow.keras import layers, Model

# Define masking block
def masking_block(input_tensor, filters, kernel_size=(3, 3),
activation='relu'):
    x = layers.Conv2D(filters, kernel_size,
padding='same')(input_tensor)
    x = layers.BatchNormalization()(x)
    x = layers.Activation(activation)(x)
    return x

def resmasking_block(input_tensor, filters, num_blocks):
    x = input_tensor
    for _ in range(num_blocks):
        # Residual layer
        residual = x

        # Masking blocks
        for _ in range(num_blocks):
            x = masking_block(x, filters)

        # Match the shape of the residual tensor
        if residual.shape[-1] != x.shape[-1]:
            residual = layers.Conv2D(filters, (1, 1),
padding='same')(residual)

        # Residual connection
        x = layers.Add()([x, residual])
        x = layers.Activation('relu')(x)

    return x

# Define the convolutional neural network
def create_model(input_shape, num_classes):
    inputs = tf.keras.Input(shape=input_shape)

    # Conv1
    x = layers.Conv2D(64, (7, 7), strides=(2, 2),
padding='same', activation='relu')(inputs)
    x = layers.MaxPooling2D((3, 3), strides=(2, 2),
padding='same')(x)

    # Resmasking Block 1
    x = resmasking_block(x, filters=32, num_blocks=1)

    # Resmasking Block 2
    x = resmasking_block(x, filters=64, num_blocks=2)

    # Resmasking Block 3
    x = resmasking_block(x, filters=128, num_blocks=3)

```

Lanjutan script

```

# Resmasking Block 4
x = resmasking_block(x, filters=64, num_blocks=2)

# Resmasking Block 5
x = resmasking_block(x, filters=32, num_blocks=1)

# Average pooling
x = layers.GlobalAveragePooling2D()(x)
x = Dropout(0.5)(x)

# Fully connected layer with softmax activation
outputs = layers.Dense(num_classes,
activation='softmax')(x)

model = Model(inputs, outputs)
return model

# Define input shape and number of classes
input_shape = (48, 48, 3) # Example input shape
num_classes = 7

# Create the model
model = create_model(input_shape, num_classes)

# Compile the model
model.compile(optimizer='adam',
loss='categorical_crossentropy', metrics=['accuracy'])

# Print model summary
model.summary()

```

Kode tersebut merupakan implementasi Convolutional Neural Network dengan menggunakan blok-blok konvolusi yang disebut masking block dan resmasking block. Pada fungsi Fungsi masking_block digunakan untuk membuat blok konvolusi sederhana dengan batch normalization dan aktivasi ReLU. Blok ini mengambil input tensor dan mengaplikasikan operasi konvolusi, batch normalization, dan aktivasi ReLU pada tensor tersebut. Fungsi resmasking_block digunakan untuk membuat blok residu yang terdiri dari beberapa masking blocks. Blok ini mengambil input tensor, jumlah filter, dan jumlah blok, dan kemudian

mengulangi pembangunan beberapa blok masking dalam blok residu. Blok residu juga mengimplementasikan koneksi residu, yaitu mengambil input asli sebelum operasi konvolusi dan menambahkannya kembali ke output setelah operasi konvolusi. Kemudian, model dibuat dan digunakan untuk membuat model CNN dengan menggunakan blok-blok yang telah didefinisikan sebelumnya. Model ini terdiri dari beberapa lapisan konvolusi awal, diikuti oleh beberapa blok residu yang dibangun dengan menggunakan masking blocks. Setelah model dibuat, kode tersebut mengkompilasi model dengan menggunakan optimizer Adam dan fungsi loss categorical crossentropy, serta menentukan metrik evaluasi akurasi untuk pelatihan. Dengan demikian, kode tersebut merupakan implementasi CNN yang menggunakan blok-blok konvolusi dengan koneksi residu untuk membangun arsitektur yang lebih dalam dan efektif dalam deteksi pola pada gambar.

4.5. Hasil Pelatihan dan Evaluasi Model

Setelah menentukan arsitektur model, dilakukan kompilasi model dengan menggunakan optimizer Adam, fungsi loss categorical cross-entropy, dan akurasi sebagai metrik evaluasi. Ini menentukan cara model akan dilatih dan dievaluasi. Keseluruhan, langkah-langkah ini menggambarkan proses ekstraksi fitur, pelatihan, evaluasi, dan analisis menggunakan model ResNet50, memberikan pemahaman tentang kinerja model dan kemampuannya dalam mengklasifikasikan gambar dengan akurasi.

4.5.1. Hasil Pelatihan Model

Setelah melalui tahap preprocessing dan proses pembelajaran pada setiap skenario, berbagai hasil klasifikasi berhasil diperoleh sesuai dengan jumlah

skenario yang diterapkan. Model yang terbentuk mewakili pengetahuan terhadap pembelajaran ekspresi emosi pada wajah manusia. Dengan model ini, kemampuan untuk mengidentifikasi emosi manusia berdasarkan karakteristik unik dari setiap kelas dapat dilakukan. Proses identifikasi citra dilakukan pada lapisan konvolusi, tempat ekstraksi fitur dari citra berlangsung, sebelum selanjutnya dilakukan penyederhanaan dimensi pada lapisan pooling untuk memperjelas ciri-ciri pada setiap citra. Setelah fitur-fitur citra diperoleh, langkah berikutnya adalah proses deteksi dan identifikasi pada lapisan fully connected yang menentukan kelas dari citra wajah. Dilanjutkan dengan tahap pengujian untuk mengevaluasi akurasi dan kinerja masing-masing model.

Tabel 4.2. Hasil Pelatihan Model

Skenario	Acc Train	Loss Train	Acc Val	Loss Val	Time Consumption
VGG19					
S1	0.9818	0.0495	0.6920	2.0286	2878.418 s
S2	0.9313	0.1949	0.6884	1.3282	1691.823 s
ResNet50					
S1	0.9779	0.0653	0.6652	1.9003	2304.839 s
S2	0.9710	0.0824	0.6622	1.7048	2323.586 s
ResMaskingNet					
S1	0.7664	0.6798	0.6532	1.0990	2672.472 s
S2	0.7860	0.6188	0.6519	1.1840	1605.807 s

Pada Tabel 4.2. diperlihatkan hasil dari pengujian yang telah dilakukan terhadap arsitektur VGG19, ResNet50 dan ResMaskingNet. Tabel hasil eksperimen

ini menggambarkan performa berbagai skenario model yang diuji. Terdapat beberapa poin penting yang harus diperhatikan dalam analisis ini.

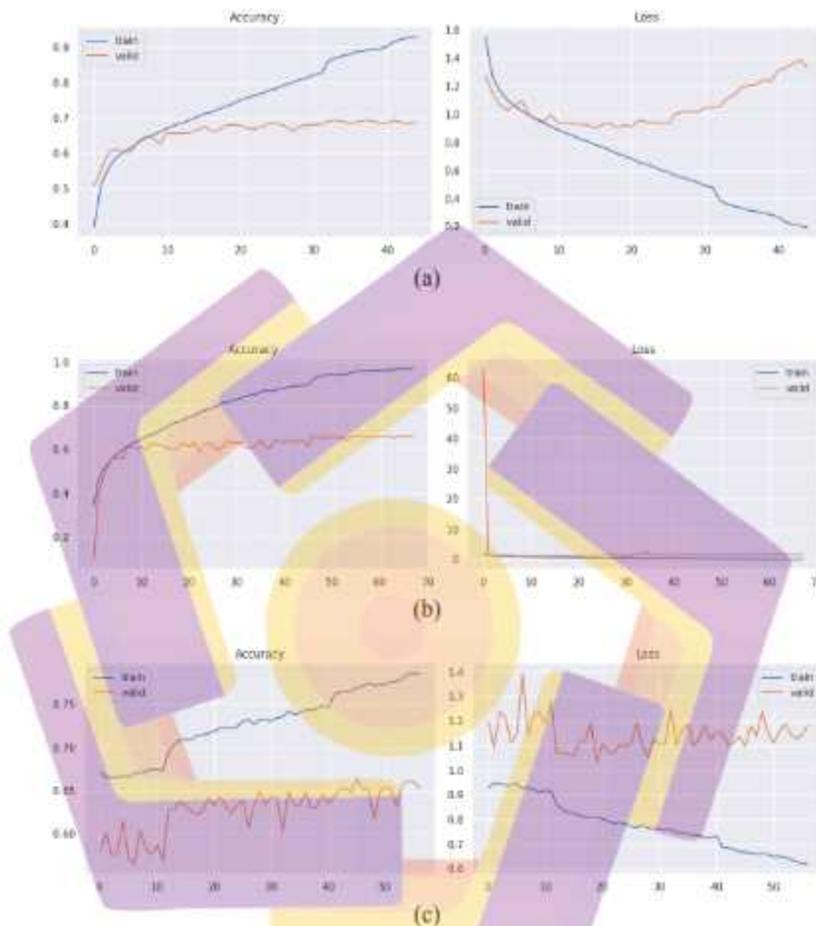
Pertama, waktu yang diperlukan untuk melatih (training) masing-masing skenario berbeda secara signifikan. Skenario S2 pada ResMaskingNet memiliki waktu pelatihan yang paling singkat, sekitar 1605.807 detik, sementara Skenario S1 pada VGG19 memiliki waktu pelatihan terlama, sekitar 2878.418 detik.

Kedua, dalam hal kualitas pelatihan, Skenario S1 pada VGG19 menunjukkan training loss yang paling rendah, yaitu 0.0495. Ini mencerminkan kemampuan model ini untuk menyesuaikan diri secara efektif dengan data pelatihan. Skenario S1 pada Resnet50 juga memiliki training loss yang rendah, yaitu 0.0653.

Ketiga, dalam konteks kemampuan akurasi, Skenario S1 pada ResMaskingNet mencapai akurasi yang relatif rendah yaitu 0.7664, dan akurasi training yang baik terdapat pada S1 pada VGG19 yaitu 0.9818.

Pada penelitian ini early stopping digunakan untuk mengoptimalkan performa dan mencegah overfitting memungkinkan model untuk berhenti pelatihan pada titik yang optimal, menghemat waktu dan sumber daya yang diperlukan. Callback ReduceLROnPlateau juga diterapkan untuk memantau akurasi validasi dan menyesuaikan laju pembelajaran jika tidak ada peningkatan dalam jumlah epoch yang telah ditentukan, mempercepat konvergensi model dan meningkatkan kinerja. Spesifikasi hyperparameter model telah ditentukan untuk pengujian, termasuk learning rate 0.0001, optimizer Adam, batch size 32, dan 100 epoch.

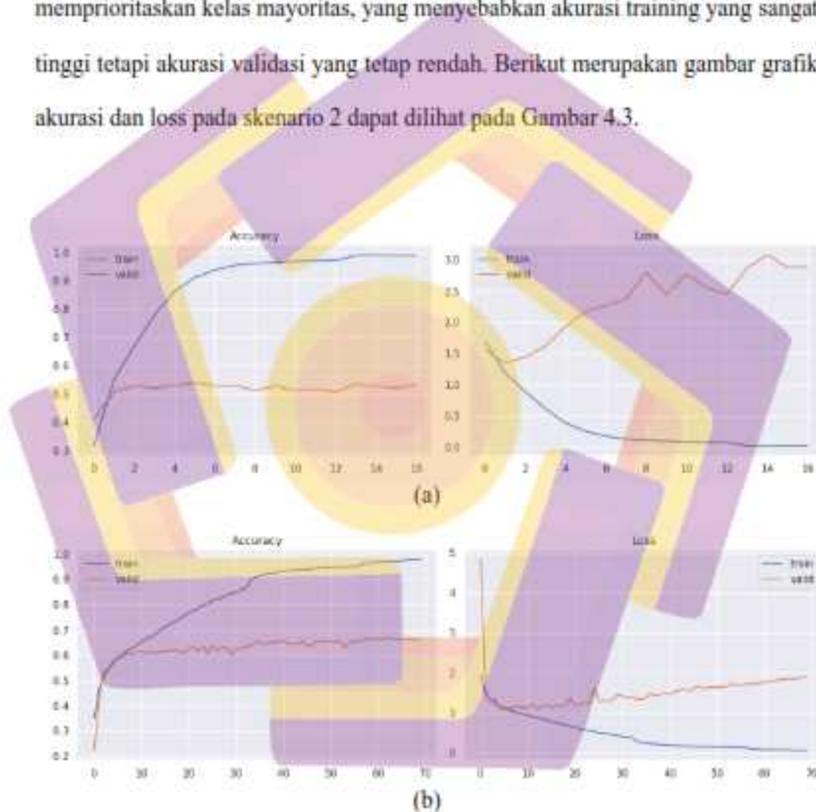
Selanjutnya akan dipaparkan hasil akurasi dari pelatihan dan validasi. Accuracy training dan accuracy validation adalah metrik penting dalam mengevaluasi kinerja model machine learning. Akurasi training mengindikasikan seberapa baik model dapat memprediksi data yang digunakan selama pelatihan, membantu dalam memahami dan menyesuaikan pola yang ada dalam data tersebut. Sementara itu, akurasi validasi digunakan untuk menilai kemampuan model dalam memprediksi data yang belum pernah dilihat sebelumnya selama pelatihan (data validasi). Penilaian akurasi validasi penting untuk menentukan apakah model dapat melakukan prediksi yang efektif pada data baru dan untuk mencegah overfitting, di mana model terlalu beradaptasi dengan data pelatihan sehingga tidak dapat menggeneralisasi pada data baru. Secara keseluruhan, akurasi training mengukur kinerja model pada data pelatihan, sementara akurasi validasi mengevaluasi kemampuan model untuk menerapkan pola pada data yang belum pernah dilihat selama pelatihan. Grafik akurasi training dan validasi untuk berbagai skenario dapat dilihat pada Gambar 4.2.

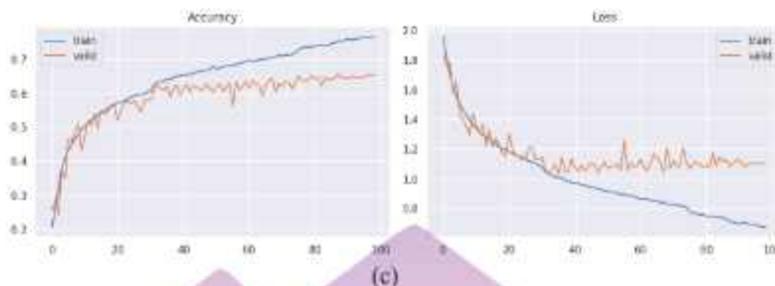


Gambar 4.2. Grafik Akurasi dan Loss data Unbalance, pada VGG19 (a), ResNet50 (b), dan ResMaskNet (c).

Pada Gambar 4.2. menunjukkan plot akurasi training dan validasi pada skenario 1 yang menunjukkan terjadinya overfitting. Kenaikan akurasi training secara terus-menerus, sementara akurasi validasi cenderung stagnan bahkan menurun, menunjukkan perbedaan kurva antara kedua metrik tersebut. Fenomena ini mengindikasikan bahwa model cenderung terlalu terfokus pada data pelatihan

yang spesifik dan kehilangan kemampuannya untuk menggeneralisasi pola pada data baru. Pada skenario 1, menggunakan data asli yang tidak seimbang menyebabkan kesenjangan besar antara akurasi training dan akurasi validasi. Kesenjangan tersebut mungkin disebabkan oleh model yang terlalu memprioritaskan kelas mayoritas, yang menyebabkan akurasi training yang sangat tinggi tetapi akurasi validasi yang tetap rendah. Berikut merupakan gambar grafik akurasi dan loss pada skenario 2 dapat dilihat pada Gambar 4.3.

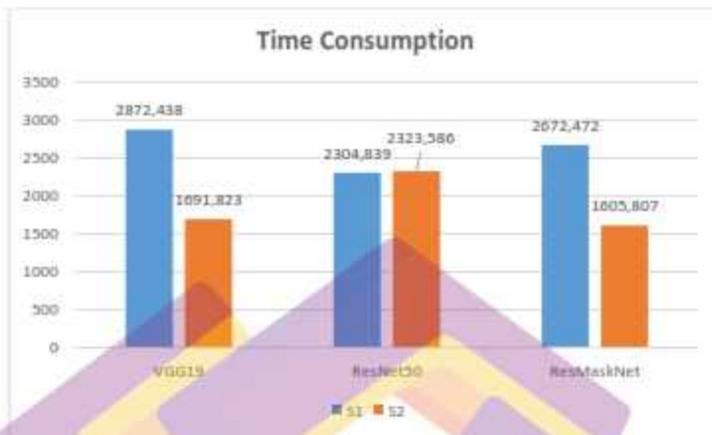




Gambar 4.3. Grafik Akurasi dan Loss data Balanced, pada VGG19 (a), ResNet50 (b), dan ResMaskNet (c).

Pada Gambar 4.3. menunjukkan plot akurasi dan loss pada data training dan validasi Skenario 2, penggunaan balancing OneSideSelection untuk menyeimbangkan data. Hal ini menandakan bahwa model yang menggunakan balanced data mengalami overfitting pada data validasi. Kurva pada data train cenderung konvergen dan tetap stabil seiring dengan berjalannya iterasi pelatihan, menunjukkan bahwa model telah mencapai titik dimana peningkatan kinerja yang signifikan.

Selanjutnya akan dibahas mengenai *Time Consumption*. *Time Consumption* merupakan waktu yang diperlukan sebuah model atau system untuk menyelesaikan tugas tertentu atau melakukan prediksi. Waktu konsumsi ini dapat digunakan untuk mengukur sejauh mana model dapat memberikan hasil dalam waktu yang wajar. Model *machine learning* yang efisien akan memiliki waktu konsumsi yang lebih rendah. Pada Gambar 4.3. menunjukkan rata-rata *time consumption* pada masing-masing scenario pada kedua model.



Gambar 4.4. Time Consumption Pada VGG19, ResNet50, ResMaskNet

Gambar 4.4. diatas menunjukkan perbandingan waktu yang digunakan untuk setiap scenario yang diuji. Waktu yang dikonsumsi (time consumption) pada S1 dan S2 dapat memberikan pemahaman mengenai efisiensi dan kompleksitas pelatihan model pada masing-masing skenario. Pada arsitektur VGG19, waktu yang dikonsumsi adalah 2878.418 s, sedangkan pada S2 terjadi peningkatan efisiensi waktu pelatihan dengan waktu yang dikonsumsi hanya 1691.823 detik. Adanya penurunan waktu consumption ini bisa disebabkan oleh beberapa faktor, seperti penggunaan teknik augmentasi data yang mungkin mempercepat proses pelatihan dengan memperbesar diversitas dataset, dan juga dipengaruhi oleh teknik balancing sehingga lebih efisien.

Selanjutnya pada arsitektur model ResNet50, pada S1 waktu yang dibutuhkan adalah 2304.839 detik. Pada S2, waktu consumption yang dibutuhkan sedikit meningkat menjadi 2323.586 detik. Perbedaan waktu yang tidak signifikan antara S1 dan S2 mungkin menunjukkan bahwa penggunaan balancing data pada

skenario kedua tidak memberikan dampak besar pada efisiensi waktu pelatihan model ResNet50.

Selanjutnya pada model ResMaskingNet, pada S1, waktu yang diperlukan adalah 2673.472 detik. Sementara pada S2, terjadi peningkatan efisiensi waktu pelatihan menjadi 1605.807 detik yang dibutuhkan. Penurunan yang signifikan dalam waktu consumption pada S2 menunjukkan bahwa penggunaan augmentasi data memiliki dampak yang positif dalam meningkatkan efisiensi pelatihan model ResMaskingNet.

Teknik balancing digunakan untuk menyeimbangkan distribusi kelas dalam dataset, khususnya dalam kasus di mana kelas-kelas memiliki jumlah sampel yang tidak seimbang. Dengan menyeimbangkan distribusi kelas, model dapat belajar dengan lebih baik dari setiap kelas, yang pada gilirannya dapat meningkatkan kemampuannya untuk melakukan prediksi yang akurat pada kelas yang kurang mewakili. Teknik balancing dapat mengurangi bias yang mungkin muncul dari ketidakseimbangan kelas dan membantu model dalam mempelajari pola yang lebih umum dalam dataset.

Dengan demikian, penggunaan kedua teknik ini dapat memberikan kontribusi terhadap penurunan waktu consumption dalam pelatihan model. Keduanya tidak hanya membantu dalam meningkatkan kinerja model, tetapi juga dapat mengoptimalkan penggunaan sumber daya komputasi yang tersedia.

4.5.2. Hasil Pengujian Model

Setelah model selesai dilatih, langkah berikutnya adalah mengujinya menggunakan berbagai metrik evaluasi. Salah satu metrik yang umum digunakan

adalah confusion matrix. Confusion matrix membantu dalam mengevaluasi kinerja model dengan membandingkan prediksi model dengan kebenaran sebenarnya. Dalam confusion matrix, ada empat istilah utama: TP (True Positives), TN (True Negatives), FP (False Positives), dan FN (False Negatives). TP mengindikasikan jumlah sampel positif yang berhasil diprediksi dengan benar, TN mencerminkan jumlah sampel negatif yang diprediksi dengan benar, FP menggambarkan jumlah sampel negatif yang salah diprediksi sebagai positif, dan FN menunjukkan jumlah sampel positif yang salah diprediksi sebagai negatif.

Selain confusion matrix, ada beberapa metrik evaluasi lainnya yang penting. Akurasi mengukur seberapa banyak sampel yang diklasifikasikan dengan benar dibandingkan dengan jumlah total sampel dalam set pengujian. Presisi mengukur seberapa banyak sampel yang diklasifikasikan sebagai positif yang sebenarnya positif, dibandingkan dengan total jumlah sampel yang diprediksi sebagai positif. Recall, juga dikenal sebagai sensitivitas, mengukur seberapa banyak sampel positif yang diprediksi dengan benar dibandingkan dengan total jumlah sampel positif dalam data sebenarnya. F1-Score adalah metrik yang menggabungkan presisi dan recall dengan menggunakan rata-rata tertimbang, memberikan gambaran keseluruhan tentang kinerja model.

Evaluasi metrik ini sangat penting karena membantu dalam menilai keandalan dan kegunaan model dalam pengambilan keputusan. Dengan memahami kelemahan dan kelebihan model, langkah-langkah perbaikan dan peningkatan kinerja dapat diambil secara tepat.

4.5.2.1. Pengujian Skenario 1 dan 2

Untuk mengukur kinerja model, confusion matrix digunakan. Confusion matrix menunjukkan seberapa akurat model dapat memprediksi kelas pada set data.



Gambar 4.5. Confusion Matrix Skenario 1 pada (a) VGG19, (b) ResNet50, (c) ResMaskNet, dan Skenario 2 pada (d) VGG19, (e) ResNet50, (f) ResMaskNet

Pada skenario 1, akan ditunjukkan hasil dari confusion matrix dari model arsitektur ResNet50, VGG19, dan ResMaskNet. Pada Gambar 4.4 merupakan confusion matrix pada setiap model dan skenario penelitian yang dilakukan. Pada Gambar 4.5. tersebut, keenamnya memiliki kesamaan yaitu pada label kelas 3 memiliki Tingkat True Positive yang tinggi. Label kelas 3 sendiri merupakan kelas Happiness, dimana dalam dataset memiliki jumlah data paling banyak. True Positive merupakan jumlah sampel secara benar diprediksi sebagai kelas yang sebenarnya. True Positive menunjukkan jumlah prediksi yang benar pada seluruh kelas. Semakin gelap warnanya, maka semakin besar prediksi benar. Hal ini bisa terjadi karena adanya ketidakseimbangan kelas dalam data sehingga menyebabkan nilai peningkatan nilai True Positive (TP) pada kelas dengan jumlah data yang lebih banyak. Ini terjadi karena model cenderung memprediksi kelas mayoritas lebih akurat daripada kelas minoritas.

Pada gambar 4.5. (a) model menunjukkan kemampuan yang cukup baik dalam mendeteksi kelas 3 dengan True Positive sebanyak 1538 data, 21.43% dari total data. Namun, masih terdapat kesalahan klasifikasi yang cukup tinggi, dengan 5186 data yaitu 72.24% dari total data terdeteksi salah. Gambar 4.5. (b) hasil menunjukkan 20.89% dari total data merupakan sampel yang benar diklasifikasikan sebagai kelas 3 yaitu sebanyak 1506 data, sedangkan presentase kesalahan klasifikasi sebanyak 70.81% dari total data salah diklasifikasikan sebagai kelas 3 sebanyak 5083 data. Selanjutnya, pada gambar 4.5 (c) 1534 data termasuk dalam kategori True Positive dengan presentase 21.37%, dan 5110 data True Negative dengan presentase sebesar 71.18%.

Setelah dilakukan balancing data, gambar 4.5. (d) menunjukkan presentase 21.77% dari keseluruhan data terdeteksi benar sebanyak 1563 data, sedangkan 5170 data sepadan 72.04% terdeteksi salah. Pada gambar 4.5 (e), 1531 data atau 21.32% dari total data terdeteksi benar, dan 5105 data atau 71.12% dari total data terdeteksi salah. Terakhir, pada Gambar 4.5 (f) menunjukkan sebanyak 1571 data dengan presentase 21.88% terdeteksi benar dan 5034 data dengan presentase 70.13% terdeteksi salah. Berdasarkan hasil yang telah dipaparkan, ini menunjukkan bahwa tidak terdapat perbedaan hasil yang signifikan antara skenario 1 dan skenario 2. Secara keseluruhan hasil skenario, model sering kali salah dalam mengklasifikasikan kelas 3 dengan kelas lainnya. Tingginya presentase kesalahan klasifikasi pada setiap skenario menunjukkan bahwa model masih memerlukan perbaikan lebih lanjut dalam model untuk meningkatkan akurasi.

Sedangkan pada label 1, yaitu kelas disgust nilai True Positive pada kelas tersebut memiliki nilai yang paling sedikit diantara kelas lainnya. Hal ini dikarenakan kurangnya representasi data untuk kelas tersebut. Kelas minoritas, yang merupakan kelas disgust, memiliki jumlah sampel yang jauh lebih sedikit dibandingkan dengan kelas mayoritas. Oleh karena itu, model cenderung memiliki akses terbatas terhadap data yang relevan untuk memahami pola yang terkait dengan kelas minoritas. Akibatnya, kemampuan model dalam mengenali dengan tepat instance yang termasuk dalam kelas minoritas menjadi terbatas. Dengan demikian, model kurang efektif dalam mempelajari pola yang berkaitan dengan kelas minoritas karena kurangnya representasi dalam data pelatihan.

4.6. Evaluasi Model

Pada model RMN, Resnet50 dan VGG19 sebagai arsitektur jaringan yang berfungsi sebagai deteksi ekspresi wajah menggunakan dataset FERF. ResNet50 dan VGG19 dapat diintegrasikan sebagai basis arsitektur lainnya, dan dengan model ini akan dievaluasi menggunakan subset pengujian matrik evaluasi yaitu akurasi, presisi, dan recall dan F1-Score. Ini dapat digunakan untuk mengukur kinerja model dalam deteksi ekspresi wajah. Selanjutnya hasil dari confusion matrix dari skenario 1 hingga 2 pada ketiga model arsitektur dapat dilihat pada Gambar 4.5 berikut.



Gambar 4.5. Evaluasi Metric pada Skenario 1

Pada Gambar 4.5. menunjukkan evaluasi metric pada Skenario 1, VGG19 memiliki presisi sebesar 69%, ResNet50 memiliki presisi sebesar 66%, dan ResMaskingNet memiliki presisi sebesar 65%. Ini menunjukkan bahwa VGG19 memiliki kemampuan yang sedikit lebih baik dalam menghindari membuat prediksi positif yang salah dibandingkan dengan ResNet50 dan ResMaskNet. Pada Recall,

VGG19 memiliki recall sebesar 67%, ResNet50 memiliki recall sebesar 66%, dan ResMaskingNet juga memiliki recall sebesar 65%. Ini menunjukkan bahwa VGG19 memiliki kemampuan yang sedikit lebih baik dalam mengidentifikasi sampel positif dibandingkan dengan ResNet50 dan ResMaskingNet. Kemudian, VGG19 memiliki F1-Score sebesar 68%, ResNet50 memiliki F1-Score sebesar 66%, dan ResMaskingNet juga memiliki F1-Score sebesar 65%. Ini menunjukkan bahwa VGG19 memiliki keseimbangan yang sedikit lebih baik antara presisi dan recall dibandingkan dengan ResNet50 dan ResMaskingNet. Selanjutnya, VGG19 memiliki akurasi sebesar 69%, ResNet50 memiliki akurasi sebesar 67%, dan ResMaskingNet juga memiliki akurasi sebesar 66%. Secara keseluruhan, VGG19 adalah model yang paling unggul dari ketiga model tersebut dalam hal kinerja evaluasi yang diberikan. Selanjutnya, berikut adalah hasil pengujian pada skenario 2 dapat dilihat pada Gambar 4.6 berikut.



Gambar 4.6. Evaluasi Metric pada Skenario 2

Pada gambar 4.6. tersebut, terdapat empat metrik evaluasi kinerja model yaitu (Presisi, Recall, F1-Score, dan Akurasi) untuk tiga model yang berbeda. Presisi mengukur seberapa banyak prediksi positif yang benar dibandingkan dengan total prediksi positif yang dilakukan model. VGG19 memiliki presisi sebesar 70%, ResNet50 dan ResMaskNet memiliki presisi yang sama yaitu 66%. Ini berarti VGG19 memiliki kemampuan yang sedikit lebih baik dalam menghindari membuat prediksi positif yang salah dibandingkan dengan ResNet50 dan ResMaskNet. Recall yang juga dikenal sebagai sensitivitas mengukur seberapa banyak sampel positif yang berhasil diprediksi dengan benar oleh model dibandingkan dengan total jumlah sampel positif dalam data sebenarnya. VGG19 memiliki recall sebesar 68%, sedangkan ResNet50 memiliki recall sebesar 66%, dan ResMaskNet memiliki recall sebesar 63%. Ini menunjukkan bahwa VGG19 dan ResNet50 memiliki kemampuan yang lebih baik dalam mengidentifikasi sampel positif dibandingkan dengan ResMaskNet. Pada F1-Score, VGG19 sebesar 69%, sementara ResNet50 dan ResMaskNet memiliki F1-Score yang sama, yaitu 66%. Ini menunjukkan bahwa VGG19 memiliki keseimbangan yang sedikit lebih baik. Akurasi pada VGG19 lebih baik dari ResNet50 dan ResMaskNet yaitu 70%. Meskipun VGG19 memiliki kinerja yang lebih baik dalam hal presisi, recall, dan F1-Score, akurasi keduanya hampir sama dengan ResNet50 dan hanya sedikit lebih tinggi dari ResMaskingNet.

Dilihat dari keseluruhan hasil, VGG19 lebih unggul daripada ResNet50, dan ResMaskingNet. Berdasarkan kedalaman yang dimiliki, VGG19 memiliki kedalaman sebanyak 19 lapisan, ResNet50 sebanyak 50 lapisan, dan

ResMaskingNet sebanyak 10 lapisan. Dataset FEREC mungkin tidak terlalu besar atau kompleks secara visual. Dalam konteks ini, VGG19 dengan kedalamannya sudah cukup untuk mengekstraksi fitur wajah yang relevan. Sedangkan ResMaskingNet kurang kompleks karena hanya 10 lapisan, sedangkan ResNet50 terlalu kompleks dengan 50 lapisan ketika digunakan pada dataset FEREC. Selain itu, Residual Learning mungkin tidak memberikan keuntungan yang signifikan dalam dataset FEREC. Sehingga, VGG19 yang lebih sederhana, lebih cocok jika diterapkan pada dataset tersebut. Dalam penelitian ini, VGG19 telah bekerja lebih baik dalam tugas klasifikasi gambar pada dataset FEREC. Dalam kasus dimana tidak ada fitur khusus dari dataset yang membutuhkan presentasi yang lebih kompleks, VGG19 sudah cukup untuk melakukan tugas face recognition pada dataset tersebut.

Fakta tersebut relevan dengan konsep umum dalam machine learning, terutama dalam konteks ketidakseimbangan kelas atau fitur dalam data pelatihan. Masalah ketidakseimbangan kelas dapat menyebabkan hasil evaluasi yang bias terhadap kelas mayoritas dan tidak menguntungkan bagi kelas minoritas. Oleh karena itu, diperlukan pendekatan khusus untuk menangani ketidakseimbangan ini, seperti penggunaan metrik evaluasi yang sesuai dan penerapan teknik-teknik seperti oversampling, undersampling, atau pengaturan ulang bobot kelas. Metode seperti oversampling (menambahkan sampel dari kelas minoritas) atau undersampling (mengurangi sampel dari kelas mayoritas) adalah strategi yang digunakan untuk menangani ketidakseimbangan kelas dalam data. Dengan menerapkan teknik-teknik ini, distribusi kelas dalam dataset dapat diubah sehingga setiap kelas memiliki jumlah representasi yang lebih seimbang. Hal ini bertujuan untuk

membantu model dalam mempelajari pola-pola dari semua kelas dengan lebih baik, sehingga meningkatkan kemampuan model untuk mengidentifikasi pola-pola yang mungkin terjadi pada kelas atau fitur yang kurang umum.

Selain itu hal lain yang menyebabkan performa model tidak bisa optimal yaitu dataset yang tidak bagus. Data training yang berkualitas rendah adalah penyebab utama kegagalan sistem machine learning di industri. Alasan mengapa data tidak optimal adalah karena dalam kumpulan data FEREC, terdapat objek yang sebenarnya tidak perlu tapi terdapat dalam frame data. Sebagai contoh, dalam dataset ekspresi wajah, elemen yang paling penting adalah fitur-fitur seperti kontur wajah, ekspresi di area alis, mata, hidung, dan bibir. Namun, dataset tersebut masih mencakup banyak elemen yang tidak relevan seperti bahu, tangan, rambut, dan make up. Hal ini mengakibatkan penurunan kinerja algoritma dalam mengenali fitur-fitur yang sebenarnya penting, karena keberadaan terlalu banyak elemen yang sebenarnya tidak diperlukan dalam data tersebut.

Berdasarkan penelitian sebelumnya menunjukkan bahwa penerapan teknik balancing pada dataset dapat meningkatkan tingkat akurasi, hasil ini konsisten dengan temuan-temuan sebelumnya dalam bidang machine learning dan data mining. Terutama, hal ini terbukti efektif ketika dataset mengalami ketidakseimbangan, di mana beberapa kelas memiliki jumlah sampel yang jauh lebih sedikit daripada yang lain. Dalam penelitian yang dilakukan oleh (Masana et al., 2020), penekanannya pada masalah ketidakseimbangan dataset yang dapat menyebabkan performa klasifikasi yang buruk pada kelas minoritas menjadi relevan. Ketidakseimbangan ini dapat menyebabkan model lebih cenderung

memprioritaskan kelas mayoritas, sehingga meningkatkan kesalahan prediksi pada kelas minoritas. Penggunaan teknik balancing data seperti oversampling atau undersampling membantu mengatasi masalah ini dengan memberikan representasi yang lebih seimbang untuk setiap kelas dalam data pelatihan.

4.7. Perbandingan dengan Penelitian Sebelumnya

Untuk memudahkan perbandingan dengan penelitian serupa yang pernah dilakukan sebelumnya, dapat dilihat pada gambar yang merupakan grafik perbandingan dengan penelitian terdahulu terkait penelitian yang dilakukan dalam deteksi emosi wajah menggunakan dataset FERF.



Gambar 4.6. Perbandingan Akurasi

Berdasarkan hasil dari 4 penelitian yang ditampilkan pada Gambar 4.7, penggunaan teknik balancing masih belum mencukupi untuk memberikan peningkatan akurasi pada dataset FERF. Pada penelitian pertama yang dilakukan oleh (Ansari, Kulkarni, Rajesh, & Gurudas, 2023) melakukan ekstraksi fitur pada model, dengan akurasi hasil sebesar 70.14% pada dataset FERF. Namun, hasil

penelitian tersebut menunjukkan terjadinya overfitting. Pada hasil confusion matrix, penelitian tersebut mengalami banyak miss-classification pada keseluruhan kelasnya. Hal tersebut dapat menjadi masalah karena jika model dilatih pada dataset yang tidak seimbang dapat menjadi bias terhadap kelas mayoritas dan kurang mampu memprediksi kelas minoritas.

Penelitian oleh (Kavitha & Kannan, 2023), didasarkan pada transfer learning dan pengoptimalan hyperparameter untuk pengenalan wajah dengan mengusulkan kombinasi dari CNN dan LSTM menghasilkan akurasi sebesar 66%. Selain itu proses preprocessing menggunakan Gaussian Filter untuk mengurangi noise pada dataset FEREC. Walaupun begitu, nampaknya metode yang dilakukan masih belum cukup untuk meningkatkan akurasi. Hal ini dikarenakan dataset FEREC dipengaruhi oleh perbedaan antar-subjek yang signifikan berdasarkan karakteristik manusia seperti usia, jenis kelamin, etnis, dan ekspresi. Kebingungan juga bertambah dikarenakan adanya variasi postur, dan juga pencahayaan.

Penelitian yang dilakukan oleh (Singh et al., 2024) membuat metode baru untuk mengidentifikasi emosi manusia berdasarkan Convolutional Neural Network. Melakukan pendekatan konvensional dengan memanfaatkan potensi metode pembelajaran mesin modern AlexNet dengan memanfaatkan dataset FEREC dan RAFD. Hasil menunjukkan akurasi pada dataset FEREC lebih baik sebesar 89.76%. Penyetelan hyperparameter dengan pendekatan ensemble dan mencoba fungsi aktivasi dapat meningkatkan kinerja model. Namun teknologi pengenalan emosi akan menimbulkan masalah etika termasuk privasi dan bias. Peneliti tersebut menyarankan untuk dapat meningkatkan model terhadap bias, serta berkolaborasi

dengan melibatkan professional psikologi, sehingga dapat meningkatkan pengenalan emosi. Serta meningkatkan pemahaman tentang emosi dan petunjuk visual yang terkait.

Penelitian mengenai face recognition seringkali menantang. Masalah utama yang sering muncul dalam penelitian ini, termasuk penelitian sebelumnya yaitu keragaman emosi dalam dataset yang secara signifikan dapat memengaruhi kinerja model. Faktor yang mempengaruhi keragaman dalam dataset antara lain yaitu seperti ras yang berbeda, usia, kondisi pencahayaan, oklusi (halangan pada wajah), wajah dari samping, postur tubuh, dan gambar skala abu-abu dengan ukuran 48 x 48 pixel. Selain itu, masalah lainnya yaitu terdapat label yang hilang dan kesalahan label, sehingga membuat penelitian pengenalan ekspresi wajah menggunakan dataset FERECN menantang (Yaermaimaiti, Kari, & Zhuang, 2022)(Alsharekh, 2022). Adanya keragaman emosi dalam dataset tersebut menimbulkan tantangan lainnya seperti ketidakseimbangan data. Dataset yang tidak seimbang ini dapat menghasilkan model yang bias berkinerja baik pada kelas mayoritas namun buruk pada kelas minoritas. Besarnya ukuran dataset juga menyebabkan terjadinya overfitting. Overfitting terjadi ketika sebuah model terlalu kompleks dan menangkap pola yang spesifik pada data pelatihan daripada pola yang dapat digeneralisasi. Selain itu, kualitas gambar dalam dataset sangat bervariasi, sehingga menyulitkan model untuk menggeneralisasi dengan baik. Gambar dalam dataset FERECN memiliki resolusi, kondisi pencahayaan, ekspresi wajah yang berbeda sehingga mempengaruhi kinerja model.

Meskipun teknik oversampling telah diterapkan dalam penelitian ini, hasilnya masih belum memuaskan. Kemungkinan terjadinya overfitting pada kelas minoritas menjadi penyebabnya. Sesuai dengan prinsip kerja dari oversampling one sided selection, yang bertujuan menambah jumlah sampel pada kelas minoritas dengan menduplikasi data yang ada. Hal tersebut bisa membuat model terlalu fit pada data yang dihasilkan secara sintetis, akibatnya model tidak mampu menggeneralisasi dengan baik pada data sebenarnya. Selain itu, jika data yang telah dioversampling memiliki kualitas rendah atau label salah, model akan belajar dari data yang salah atau tidak relevan sehingga dapat mengurangi akurasi dan kinerja model. Selain itu, faktor lain seperti variasi dalam dataset membuat tugas pengenalan ekspresi wajah menjadi sangat kompleks. Penambahan jumlah sampel saja tidak selalu cukup untuk menangani kerumitan ini. Sehingga, hanya menggunakan teknik oversampling tidak selalu cukup untuk menangani tantangan tersebut.

Hasil temuan tersebut menggaris bawahi pentingnya pemahaman mendalam tentang karakteristik data, terutama pada data dengan kelas yang tidak seimbang. Walaupun penelitian ini belum sukses mencapai tujuannya, namun telah memberikan kontribusi berharga dalam pemahaman tentang karakteristik data FEREC. Dataset FEREC sendiri merupakan dataset dengan kualitas yang buruk dan memiliki ekspresi dalam data yang kurang terkontrol (Singh et al., 2024). Sehingga dataset FEREC sulit digunakan karena kealamian intrinsik gambar-gambar dalam dataset yang mencerminkan situasi kehidupan nyata dan kurangnya kendali dalam pengambilan gambar. Namun, keadaan yang kurang terkontrol tersebut juga

membuat dataset tersebut menjadi alat pengujian yang berharga untuk identifikasi dan pelatihan gambar, karena dapat meniru kondisi dunia nyata yang kompleks. Dengan demikian, dataset FEREC memiliki nilai sebagai sumber data yang menantang namun realistis untuk pengembangan dan evaluasi system pengenalan ekspresi wajah (Liu, Wang, Li, & Pan, 2023).

Dari keseluruhan perbandingan, penelitian ini masih belum bisa melebihi penelitian sebelumnya. Penelitian yang dilakukan oleh (Singh et al., 2024), memanfaatkan dataset wajah yang luas dan mengaplikasikan arsitektur AlexNet untuk mengekstraksi fitur yang kompleks dari wajah. Dengan memperpanjang waktu pelatihan dan memperbesar dataset secara signifikan sehingga dapat meningkatkan kinerja model. Sedangkan pada penelitian, hanya menggunakan waktu pelatihan sebanyak 100 epoch saja. Hal ini berpengaruh terhadap performa model yang dihasilkan. Waktu pelatihan yang terbatas mungkin kurang optimal dibandingkan dengan penelitian sebelumnya yang menggunakan waktu pelatihan lebih besar. Akibatnya, Tingkat akurasi dan kemampuan generalisasi model mungkin tidak sebaik yang dihasilkan oleh penelitian sebelumnya, yang memanfaatkan pelatihan yang lebih intensif. Selain itu, terbatasnya sumber daya komputasi yang digunakan juga mempengaruhi hasil akhir. Perangkat yang digunakan dalam penelitian ini tidak cukup kuat sehingga mempengaruhi proses pelatihan model menjadi lebih lambat dan kurang efisien. Ini merupakan faktor yang membatasi kemampuan peneliti untuk menjalankan lebih banyak epoch atau menggunakan dataset yang lebih besar, yang pada gilirannya mempengaruhi kinerja model secara keseluruhan. Selain itu, dataset dengan banyak kelas seperti pada

FERC ini dapat mempengaruhi model yang berdampak pada kinerja dan efektivitas model. Semakin banyak kelas yang diidentifikasi model, maka semakin kompleks tugas klasifikasi tersebut. Model yang terus belajar membedakan antara banyak kelas yang berbeda, dapat meningkatkan kesulitan dalam pelatihan dan optimasi model. Masalah keseimbangan data pada setiap kelas juga menyebabkan model bias terhadap kelas yang lebih banyak data dan mengabaikan kelas yang kurang terwakili. Semakin banyak kelas yang harus dikenali, semakin sulit bagi model untuk mencapai akurasi tinggi. Model mungkin mengalami kesulitan dalam membedakan kelas yang serupa, sehingga dapat menurunkan performa secara keseluruhan.

Penelitian yang dilakukan hanya mencapai akurasi 70% dibandingkan penelitian (Singh et al., 2024) yang memiliki akurasi 89.76%, ini disebabkan oleh faktor termasuk waktu pelatihan yang terbatas, dataset yang kurang representative, serta keterbatasan sumber daya komputasi. Meskipun demikian, penelitian ini tetap memberikan kontribusi penting yang berharga. Penelitian ini memperkenalkan metode baru dengan menggunakan CNN bersama dengan oversampling *One Sided Selection* untuk pengenalan emosi menggunakan deep learning. Metode ini mungkin memiliki potensi yang belum sepenuhnya terekplorasi dan bisa dioptimalkan lebih lanjut. Selain itu, penelitian ini juga memberikan dasar bagi peneliti lain untuk mengembangkan dan menyempurnakan metode yang digunakan. Kelemahan dan kekurangan yang ditemukan dapat menjadi Pelajaran berharga untuk perbaikan dan inovasi lebih lanjut.

BAB V

PENUTUP

5.1. Kesimpulan

Berdasarkan hasil pada percobaan yang telah dilakukan terhadap arsitektur VGG19, ResNet50 dan ResMakingNet dengan dua scenario percobaan dapat disimpulkan bahwa.

Penelitian ini menunjukkan bahwa kombinasi dataset FEREC dengan arsitektur model VGG19 dengan teknik oversampling dalam proses balancing data telah menghasilkan hasil akurasi tertinggi. VGG19 menonjol sebagai arsitektur yang efektif dalam tugas ini. Walaupun demikian, penelitian ini masih belum melebihi penelitian sebelumnya dikarenakan waktu pelatihan yang terbatas, dataset yang kurang representative, dan keterbatasan sumber daya komputasi. Waktu pelatihan yang terbatas menyebabkan model tidak optimal. Sehingga Tingkat akurasi dan kemampuan untuk generalisasi model tidak sebaik yang dihasilkan oleh penelitian sebelumnya yang memanfaatkan waktu pelatihan lebih intensif. Dataset FEREC yang kurang representative, ini berarti sampel data yang digunakan tidak mencerminkan secara akurat gambar yang ingin diidentifikasi. Penting untuk memastikan bahwa sampel yang diambil mencakup keragaman yang ada atau sesuai dengan metodologi yang tepat untuk pengenalan emosi wajah. Selain itu, terbatasnya sumber daya komputasi mempengaruhi kemampuan untuk menjalankan aplikasi yang membutuhkan banyak daya komputasi, pemrosesan data besar, dan kecerdasan buatan yang membutuhkan

pengolahan mendalam. Keterbatasan sumber daya komputasi ini dapat mempengaruhi pada waktu pemrosesan yang lebih lama.

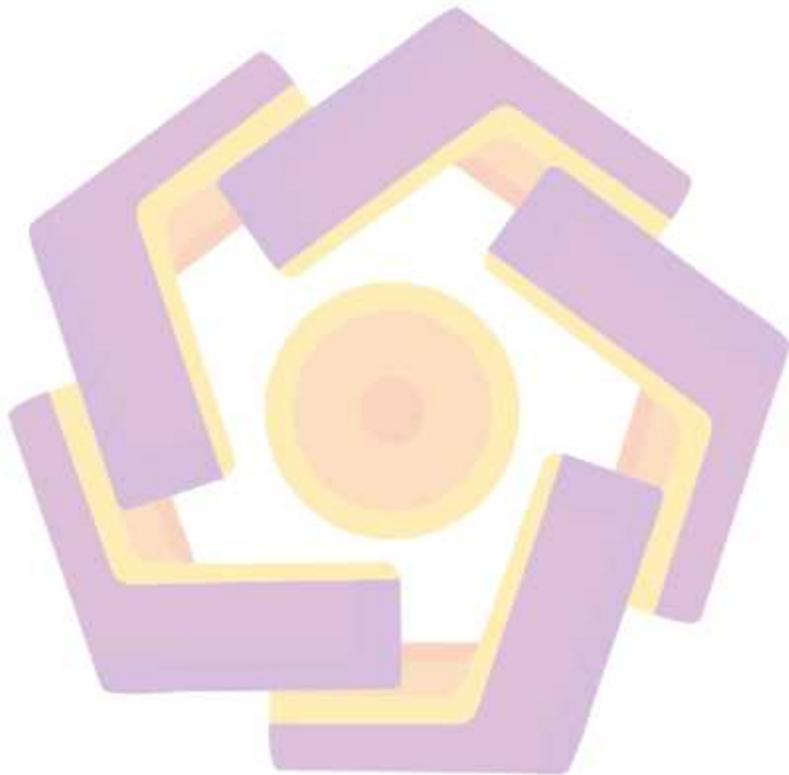
Meskipun teknik oversampling *One Sided Selection* telah diterapkan pada penelitian ini, hasilnya masih belum memuaskan. Terjadinya overfitting pada kelas minoritas menjadi penyebabnya. Sesuai dengan prinsip kerjanya, tujuan dari oversampling yaitu untuk menambah jumlah sampel pada kelas minoritas dengan menduplikasi data yang ada sehingga menyebabkan model terlalu terbiasa dengan data yang dihasilkan secara sintesis. Sehingga model tidak mampu menggeneralisasi dengan baik pada data sebenarnya. Data yang telah dioversampling mungkin akan memiliki kualitas rendah, model dapat mempelajari pola yang tidak relevan atau bahkan salah, yang akan mengurangi akurasi dan kinerja model. Kompleksitas dari dataset, seperti variasi dalam ekspresi wajah, membuat tugas pengenalan tersebut menjadi sangat rumit. Dalam konteks ini, penambahan jumlah sampel saja tidak cukup untuk menangani kompleksitas tersebut. Penggunaan oversampling tidaklah cukup untuk mengatasi tantangan yang dihadapi dalam penelitian tersebut. Diperlukan pendekatan yang lebih komprehensif atau kombinasi dari berbagai teknik untuk meningkatkan kinerja model.

5.2. Saran

Adapun beberapa saran yang dapat direkomendasikan untuk penelitian selanjutnya pada topik penelitian yang sama adalah sebagai berikut.

1. Memberikan model lebih banyak waktu untuk pelatihan ketika model diberikan lebih banyak waktu untuk pelatihan, ini memberikan kesempatan untuk algoritma pembelajaran mesin untuk menyesuaikan parameter-parameter dengan lebih baik terhadap data yang ada. Dengan demikian, model dapat menangkap dan memahami pola-pola yang lebih rumit dan mendalam dalam data tersebut. Dalam konteks arsitektur model seperti VGG19 yang kompleks, waktu pelatihan yang lebih lama memungkinkan lapisan-lapisan yang dalam untuk secara efektif mengekstraksi fitur-fitur yang lebih beragam dan rumit dari dataset. Dengan lebih banyak waktu pelatihan, model memiliki potensi yang lebih besar untuk mencapai tingkat akurasi dan kinerja yang lebih tinggi.
2. Menanggulangi keterbatasan sumber daya komputasi dengan menggunakan teknik regularisasi seperti dropout atau regularisasi L2 untuk membantu mengurangi overfitting serta memungkinkan pelatihan model dengan dataset dalam waktu yang lebih singkat. Selain itu perlu untuk mengoptimalkan hyperparameter seperti ukuran batch, learning rate, dan jumlah epoch untuk membantu mempercepat konvergensi model dan mengurangi waktu pelatihan.
3. Mengembangkan dataset yang lebih representative dengan menggunakan Generative Adversarial Networks (GANs). GANs sendiri digunakan untuk menghasilkan data sintesis yang menyerupai data asli, yang dapat membantu dalam meningkatkan representasi dataset. Namun, sebelum menggunakan

GANs penting untuk mengevaluasi kualitas data dari dataset yang akan digunakan terlebih dahulu. Perlu untuk melakukan validasi yang teliti terhadap data sintesis, untuk memastikan bahwa data tersebut dapat digunakan dalam pelatihan model.



DAFTAR PUSTAKA

PUSTAKA BUKU

- Bambang Y, Yusri I S, Mangaras Y F & Panji D A. (2019) *Face & Facial Expressions Recognition*. <http://upnyk.ac.id/26165>
- Garry Agustinus Safiro. (2020) *Deteksi Ekspresi Wajah Menggunakan Deep Learning*. <http://dinamika.ac.id/eprint>
- Mohamed Elgendy. (2020). *Deep Learning for Vision Systems*. Manning Publications.

PUSTAKA MAJALAH, JURNAL ILMIAH ATAU PROSIDING

- Ali, R., Balamurali, M., & Varamini, P. (2022). Deep Learning-Based Artificial Intelligence to Investigate Targeted Nanoparticles' Uptake in TNBC Cells. *International Journal of Molecular Sciences*, 23(24). <https://doi.org/10.3390/ijms232416070>
- Ansari, S., Kulkarni, P., Rajesh, T., & Gurudas, V. R. (2023). Facial Emotion Detection Using Deep Learning: A Survey. *Proceedings of IEEE InC4 2023 - 2023 IEEE International Conference on Contemporary Computing and Communications*, 1–5. <https://doi.org/10.1109/InC457730.2023.10262904>
- Dewan, A., Likhith Ashwin, D., Likith, M., Allabhaneni, G. V., & Janardhana, D. R. (2022). Face Identification and Real Time Emotion Tracker. *MysuruCon 2022 - 2022 IEEE 2nd Mysore Sub Section International Conference*. <https://doi.org/10.1109/MysuruCon55714.2022.9972729>
- Ding, H., Wei, B., Gu, Z., Yu, Z., Zheng, H., Zheng, B., & Li, J. (2020). KA-Ensemble: towards imbalanced image classification ensembling under-sampling and over-sampling. *Multimedia Tools and Applications*, 79(21–22), 14871–14888. <https://doi.org/10.1007/s11042-019-07856-y>
- Halim, A., El-Manfy, A., Badr, A. E. R., El-Khatib, A., El-Basir, M. A., El-Tabee, S., ... El-Khouly, A. (2023). Facial Expressions Analysis to Evaluate the Level

of Students' Understanding. *1st International Conference of Intelligent Methods, Systems and Applications, IMSA 2023*, 424–429. <https://doi.org/10.1109/IMSA58542.2023.10217489>

Kamal, & Ez-zahraouy, H. (2023). A comparison between the VGG16 , VGG19 and ResNet50 architecture frameworks for classification of normal and CLAHE processed medical images. *Research Square*, 0–16.

Kavitha, M. N., & Kannan, A. R. (2023). Hybrid Convolutional Neural Network and Long Short-Term Memory Approach for Facial Expression Recognition. *Intelligent Automation and Soft Computing*, 35(1), 689–704. <https://doi.org/10.32604/iasc.2023.025437>

Khairuddin, Y., & Chen, Z. (2021). *Facial Emotion Recognition: State of the Art Performance on FER2013*. Retrieved from <http://arxiv.org/abs/2105.03588>

Liu, F., Wang, Y., Li, Z., & Pan, Z. (2023). GEIKD: Self-knowledge distillation based on gated ensemble networks and influences-based label noise removal. *Computer Vision and Image Understanding*, 235(June), 103771. <https://doi.org/10.1016/j.cviu.2023.103771>

Meena, G., Mohbey, K. K., Indian, A., & Kumar, S. (2022). Sentiment Analysis from Images using VGG19 based Transfer Learning Approach. *Procedia Computer Science*, 204(2021), 411–418. <https://doi.org/10.1016/j.procs.2022.08.050>

Millenia, J., Naufal, M. F., & Siswanto, J. (2022). Melanoma Detection using Convolutional Neural Network with Transfer Learning on Dermoscopic and Macroscopic Images. *Journal of Information Systems Engineering and Business Intelligence*, 8(2), 149–161. <https://doi.org/10.20473/jjisebi.8.2.149-161>

Pham, L., Huynh Vu, T., Anh Tran, T., Chi Minh City, H., Trung Ward, L., & Duc District, T. (n.d.). *Facial Expression Recognition Using Residual Masking Network*. Retrieved from <http://github.com/phamquilian/ResidualMaskingNetwork>

Rangaswamy, S. (n.d.). *REAL TIME FACE MASK DETECTION USING CNN*. Retrieved from <https://ssrn.com/abstract=4139377>

- Singh, K. U., Kumar, A., Kumar, G., Singh, T., Kumar, S., & Yadav, S. P. (2024). An Autonomous Emotion Recognition Strategy Employing Deep Learning for Self-Learning. *2023 3rd International Conference on Technological Advancements in Computational Sciences (ICTACS)*, 883–888. <https://doi.org/10.1109/ictacs59847.2023.10389867>
- Tong, Y., Cao, W., Sun, Q., & Chen, D. (2021). The Use of Deep Learning and VR Technology in Film and Television Production From the Perspective of Audience Psychology. *Frontiers in Psychology*, 12. <https://doi.org/10.3389/fpsyg.2021.634993>
- Ali, R., Balamurali, M., & Varamini, P. (2022). Deep Learning-Based Artificial Intelligence to Investigate Targeted Nanoparticles' Uptake in TNBC Cells. *International Journal of Molecular Sciences*, 23(24). <https://doi.org/10.3390/ijms232416070>
- Ansari, S., Kulkarni, P., Rajesh, T., & Gurudas, V. R. (2023). Facial Emotion Detection Using Deep Learning: A Survey. *Proceedings of IEEE InC4 2023 - 2023 IEEE International Conference on Contemporary Computing and Communications*, 1–5. <https://doi.org/10.1109/InC457730.2023.10262904>
- Dewan, A., Likhith Ashwin, D., Likith, M., Allabhaneni, G. V., & Janardhana, D. R. (2022). Face Identification and Real Time Emotion Tracker. *MysuruCon 2022 - 2022 IEEE 2nd Mysore Sub Section International Conference*. <https://doi.org/10.1109/MysuruCon55714.2022.9972729>
- Ding, H., Wei, B., Gu, Z., Yu, Z., Zheng, H., Zheng, B., & Li, J. (2020). KA-Ensemble: towards imbalanced image classification ensembling under-sampling and over-sampling. *Multimedia Tools and Applications*, 79(21–22), 14871–14888. <https://doi.org/10.1007/s11042-019-07856-y>
- Halim, A., El-Manfy, A., Badr, A. E. R., El-Khatib, A., El-Basir, M. A., El-Tabee, S., ... El-Khouly, A. (2023). Facial Expressions Analysis to Evaluate the Level of Students' Understanding. *1st International Conference of Intelligent Methods, Systems and Applications, IMSA 2023*, 424–429. <https://doi.org/10.1109/IMSA58542.2023.10217489>
- Kamal, & Ez-zahraouy, H. (2023). A comparison between the VGG16 , VGG19 and ResNet50 architecture frameworks for class i cation of normal and CLAHE processed medical images. *Research Square*, 0–16.

- Kavitha, M. N., & Kannan, A. R. (2023). Hybrid Convolutional Neural Network and Long Short-Term Memory Approach for Facial Expression Recognition. *Intelligent Automation and Soft Computing*, 35(1), 689–704. <https://doi.org/10.32604/iasc.2023.025437>
- Khairuddin, Y., & Chen, Z. (2021). *Facial Emotion Recognition: State of the Art Performance on FER2013*. Retrieved from <http://arxiv.org/abs/2105.03588>
- Liu, F., Wang, Y., Li, Z., & Pan, Z. (2023). GEIKD: Self-knowledge distillation based on gated ensemble networks and influences-based label noise removal. *Computer Vision and Image Understanding*, 235(June), 103771. <https://doi.org/10.1016/j.cviu.2023.103771>
- Meena, G., Mohbey, K. K., Indian, A., & Kumar, S. (2022). Sentiment Analysis from Images using VGG19 based Transfer Learning Approach. *Procedia Computer Science*, 204(2021), 411–418. <https://doi.org/10.1016/j.procs.2022.08.050>
- Millenia, J., Naufal, M. F., & Siswanto, J. (2022). Melanoma Detection using Convolutional Neural Network with Transfer Learning on Dermoscopic and Macroscopic Images. *Journal of Information Systems Engineering and Business Intelligence*, 8(2), 149–161. <https://doi.org/10.20473/jisebi.8.2.149-161>
- Pham, L., Huynh Vu, T., Anh Tran, T., Chi Minh City, H., Trung Ward, L., & Duc District, T. (n.d.). *Facial Expression Recognition Using Residual Masking Network*. Retrieved from <http://github.com/phamquilian/ResidualMaskingNetwork>
- Rangaswamy, S. (n.d.). *REAL TIME FACE MASK DETECTION USING CNN*. Retrieved from <https://ssrn.com/abstract=4139377>
- Singh, K. U., Kumar, A., Kumar, G., Singh, T., Kumar, S., & Yadav, S. P. (2024). An Autonomous Emotion Recognition Strategy Employing Deep Learning for Self-Learning. *2023 3rd International Conference on Technological Advancements in Computational Sciences (ICTACS)*, 883–888. <https://doi.org/10.1109/ictacs59847.2023.10389867>
- Tong, Y., Cao, W., Sun, Q., & Chen, D. (2021). The Use of Deep Learning and VR Technology in Film and Television Production From the Perspective of

- Pham, L., Huynh Vu, T., Anh Tran, T., Chi Minh City, H., Trung Ward, L., & Duc District, T. (n.d.). *Facial Expression Recognition Using Residual Masking Network*. <http://github.com/phamquilian/ResidualMaskingNetwork>
- Cao, W., Feng, Z., Zhang, D., & Huang, Y. (2020). Facial Expression Recognition via a CBAM Embedded Network. *Procedia Computer Science*, 174, 463–477. <https://doi.org/10.1016/j.procs.2020.06.115>
- Revina, I. M., & Emmanuel, W. R. S. (2021). A Survey on Human Face Expression Recognition Techniques. In *Journal of King Saud University - Computer and Information Sciences* (Vol. 33, Issue 6, pp. 619–628). King Saud bin Abdulaziz University. <https://doi.org/10.1016/j.jksuci.2018.09.002>
- Li, S., & Deng, W. (2018). *Deep Facial Expression Recognition: A Survey*. <https://doi.org/10.1109/TAFPC.2020.2981446>
- Cheong, J. H., Jolly, E., Xie, T., Byrne, S., Kenney, M., & Chang, L. J. (n.d.). *PyFeat: Python Facial Expression Analysis Toolbox*.
- Li, Y., Guo, K., Lu, Y., & Liu, L. (2021). Cropping and attention based approach for masked face recognition. *Applied Intelligence*, 51(5), 3012–3025. <https://doi.org/10.1007/s10489-020-02100-9>
- Le, D. S., Phan, H. H., Hung, H. H., Tran, V. A., Nguyen, T. H., & Nguyen, D. Q. (2022). KFSENet: A Key Frame-Based Skeleton Feature Estimation and Action Recognition Network for Improved Robot Vision with Face and Emotion Recognition. *Applied Sciences (Switzerland)*, 12(11). <https://doi.org/10.3390/app12115455>
- Shao, J., & Qian, Y. (2019). Three convolutional neural network models for facial expression recognition in the wild. *Neurocomputing*, 355, 82–92. <https://doi.org/10.1016/j.neucom.2019.05.005>

- Cao, W., Feng, Z., Zhang, D., & Huang, Y. (2020). Facial Expression Recognition via a CBAM Embedded Network. *Procedia Computer Science*, 174, 463–477. <https://doi.org/10.1016/j.procs.2020.06.115>
- Institute of Electrical and Electronics Engineers, IEEE Computer Society, IEEE Computer Society Technical Committee on Machine Intelligence and Pattern Analysis, & International Conference on Activity and Behavior Computing 2019 Spokane, Wash. (n.d.). *2019 Joint 8th International Conference on Informatics, Electronics & Vision (ICIEV)*
- Hung, J. C., Lin, K. C., & Lai, N. X. (2019). Recognizing learning emotion based on convolutional neural networks and transfer learning. *Applied Soft Computing Journal*, 84. <https://doi.org/10.1016/j.asoc.2019.105724>
- Bialek, C., Matiolański, A., & Grega, M. (2023). An Efficient Approach to Face Emotion Recognition with Convolutional Neural Networks. *Electronics (Switzerland)*, 12(12). <https://doi.org/10.3390/electronics12122707>
- Ali, L., Alnajjar, F., Jassmi, H. al, Gochoo, M., Khan, W., & Serhani, M. A. (2021). Performance evaluation of deep CNN-based crack detection and localization techniques for concrete structures. *Sensors*, 21(5), 1–22. <https://doi.org/10.3390/s21051688>
- Li, Y., Guo, K., Lu, Y., & Liu, L. (2021). Cropping and attention based approach for masked face recognition. *Applied Intelligence*, 51(5), 3012–3025. <https://doi.org/10.1007/s10489-020-02100-9>
- Mehendale, N. (2020). Facial emotion recognition using convolutional neural networks (FERC). *SN Applied Sciences*, 2(3). <https://doi.org/10.1007/s42452-020-2234-1>
- Rangaswamy, S. (n.d.). *REAL TIME FACE MASK DETECTION USING CNN*. <https://ssrn.com/abstract=4139377>

- Meena, G., Mohbey, K. K., Indian, A., & Kumar, S. (2022). Sentiment Analysis from Images using VGG19 based Transfer Learning Approach. *Procedia Computer Science*, 204, 411–418. <https://doi.org/10.1016/j.procs.2022.08.050>
- Vidhya, J. V., & Uthra, R. A. (2021). *Violence detection in videos using Conv2D VGG-19 architecture and LSTM network.*

