Facial Expression Recognition Dengan Pemodelan Berbasis CNN Pada Wajah Bermasker

Ahmad Maulana Indidharmanto
Fakultas Informatika
Universitas Telkom
Bandung, Indonesia
maulanaindi@students.telkomuniversity.ac.id

Aditya Firman Ihsan
Fakultas Informatika
Universitas Telkom
Bandung, Indonesia
adityaihsan@telkomuniversity.ac.id

Mahmud Dwi Sulistyo
Fakultas Informatika
Universitas Telkom
Bandung, Indonesia
mahmuddwis@telkomuniversity.ac.id

***Abstrak* — Pandemi Covid-19 yang telah menyebabkan penggunaan masker wajah menjadi hal yang umum di masyarakat sebagai upaya untuk mencegah penyebaran virus. Namun, hal ini menimbulkan tantangan baru dalam pengenalan ekspresi wajah atau *Facial Expression Recognition* (FER). *Facial Expression Recognition* digunakan untuk memahami bagaimana manusia berperilaku, sehingga membantu dalam strategi pencegahan penyebaran virus dan menghilangkan halangan manusia untuk saling bersosialisasi walau dalam keterbatasan. Dalam hal keamanan, dapat dimanfaatkan untuk membedakan orang yang memiliki intensi buruk dibalik ekspresi yang tertutup masker. Kekurangan yang ada pada *Facial Expression Recognition* saat ini yaitu terbatasnya pendeteksian berbagai macam jenis ekspresi dikarenakan hilangnya informasi penting dari area mulut dan hidung yang tertutup masker. Penelitian ini bertujuan untuk menemukan model arsitektur *Convolutional Neural Network* (CNN) dengan akurasi dalam beberapa ekspresi seperti bahagia, marah, sedih, netral, dan terkejut. Penelitian ini mengevaluasi performa tiga model, yaitu ResNet50, Emotion Ensemble Model, dan VGG19. ResNet50 menunjukkan performa dengan akurasi 89.51%, Emotion Ensemble Model dengan akurasi 82.49%. Sementara itu VGG19 mencapai akurasi 72.44%. Kontribusi utama penelitian ini adalah pengembangan *ensemble* yang cukup akurat pada dataset dengan variasi tinggi, serta analisis terhadap keunggulan dan kelemahan setiap model. Membantu mengenal pemilihan arsitektur model yang tepat untuk pengenalan ekspresi berbasis citra pada kondisi terbatas.**

***Kata kunci*— facial expression recognition, convolutional neural network, ResNet50, masker wajah, deep learning, Masked-Fer2013**

I. PENDAHULUAN

Isi Pandemi Covid-19 telah menyebabkan penggunaan masker wajah menjadi hal yang umum di masyarakat sebagai upaya untuk mencegah penyebaran virus, bahkan setelah pandemi mereda. Namun, penggunaan masker ini menimbulkan tantangan baru dalam pengenalan ekspresi wajah atau Facial Expression Recognition (FER), yang merupakan aspek penting dalam berbagai aplikasi, termasuk pemantauan perilaku sosial, sistem keamanan, dan pengembangan teknologi yang lebih responsif terhadap emosi manusia. Dalam bidang keamanan, FER dapat membantu membedakan individu yang menggunakan masker dari orang lain, terutama dalam mendeteksi potensi ancaman kriminal. Selain itu, FER juga dapat digunakan untuk memahami bagaimana individu bereaksi terhadap situasi tertentu ketika mengenakan masker, sehingga mendukung pengembangan sistem interaksi yang lebih adaptif.

Facial Expression Recognition telah berkembang pesat dengan berbagai model yang mencapai tingkat akurasi tinggi dalam mendeteksi ekspresi wajah tanpa masker. Namun, keberadaan masker yang menutupi area mulut dan hidung menyebabkan kehilangan informasi penting yang berdampak pada performa model FER. Bagian wajah yang tertutup tersebut berperan besar dalam membedakan ekspresi seperti bahagia, sedih, marah, atau terkejut. Selain itu, sebagian besar dataset yang digunakan untuk pelatihan model FER terdiri dari wajah tanpa masker, sehingga menurunkan akurasi sistem ketika diterapkan pada wajah bermasker. Beberapa dataset yang memang terdiri dari wajah bermasker juga memiliki keterbatasan, seperti kurangnya variasi orientasi wajah serta label emosi yang terbatas hanya pada kategori netral, positif, dan negatif.

Berbagai pendekatan telah dikembangkan untuk meningkatkan performa Facial Expression Recognition pada wajah bermasker, termasuk penggunaan model berbasis Convolutional Neural Network (CNN) dan Visual Geometry Group (VGG). Studi terdahulu juga mengeksplorasi metode ekstraksi fitur tambahan, seperti Local Binary Pattern (LBP), untuk meningkatkan akurasi pengenalan ekspresi. Salah satu pendekatan lain adalah dengan menambahkan masker secara otomatis pada dataset wajah berekspresi guna menciptakan dataset baru yang lebih relevan dengan kondisi penggunaan sebenarnya. Meskipun demikian, akurasi pengenalan ekspresi wajah bermasker masih lebih rendah dibandingkan dengan pengenalan wajah tanpa masker, sehingga diperlukan

upaya lebih lanjut dalam mengembangkan model yang lebih optimal.

Penelitian ini bertujuan untuk menentukan model CNN terbaik dalam mengenali ekspresi wajah yang tertutup masker. Model yang diusulkan adalah ResNet50, sebuah arsitektur Deep Learning dalam kategori CNN yang menggunakan koneksi residual untuk meningkatkan stabilitas dan efisiensi pelatihan. Pembandingan akan dilakukan dengan model lainnya, seperti CNN-LBP, MobileNet, dan VGG, guna mengevaluasi keunggulan dan kelemahan masing-masing model dalam mengenali ekspresi wajah bermasker. Selain itu, penelitian ini juga berupaya meningkatkan kemampuan klasifikasi ekspresi yang lebih rinci, mencakup emosi seperti marah, senang, netral, terkejut, dan sedih. Diharapkan bahwa penelitian ini dapat memberikan kontribusi dalam pengembangan sistem pengenalan ekspresi wajah yang lebih akurat dan andal dalam situasi di mana penggunaan masker tetap menjadi bagian dari kehidupan sehari-hari.

II. KAJIAN TEORI

A. Pandemi Covid-19

Pandemi COVID-19, yang dimulai pada akhir tahun 2019, telah menyebabkan perubahan signifikan dalam berbagai aspek kehidupan manusia, termasuk interaksi sosial dan penggunaan teknologi. Salah satu perubahan yang paling mencolok adalah penggunaan masker wajah secara luas untuk mencegah penularan virus. Perubahan ini telah mempengaruhi pengenalan ekspresi wajah (FER) karena sebagian besar wajah sekarang tertutup masker, yang menghadirkan tantangan dalam menafsirkan emosi dengan akurat [6][7][8]. Penelitian telah menunjukkan bahwa penggunaan masker wajah mempengaruhi pengenalan emosi baik pada populasi umum maupun pada individu dengan ciri-ciri autistik, menyoroti praktik global memakai masker sebagai bagian dari upaya menjaga jarak sosial selama pandemi [2]. Selain itu, studi telah menunjukkan bahwa keberadaan masker wajah bedah mengurangi intensitas emosi yang dirasakan, sehingga lebih sulit untuk membedakan emosi seperti jijik, terkejut, dan sedih [7].

B. Pengenalan Ekspresi Wajah

Pengenalan ekspresi wajah adalah teknologi yang digunakan untuk mengidentifikasi ekspresi ataupun emosi seseorang berdasarkan ekspresi wajahnya. Dengan adanya pandemi COVID-19, penggunaan masker wajah menjadi umum, yang menyebabkan tantangan baru dalam pengenalan ekspresi wajah. Beberapa penelitian telah dilakukan untuk mengatasi tantangan ini dengan mengembangkan dataset wajah bermasker dan metode untuk meningkatkan akurasi pengenalan ekspresi wajah pada wajah yang tertutup masker. Salah satu metode yang diusulkan adalah menggunakan dataset yang disimulasikan dengan masker dan melatih model FER pada dataset tersebut [6].

C. Model Klasifikasi CNN

Pengenalan ekspresi wajah adalah teknologi yang digunakan untuk mengidentifikasi ekspresi ataupun emosi seseorang berdasarkan ekspresi wajahnya. Dengan adanya pandemi COVID-19, penggunaan masker wajah menjadi umum, yang menyebabkan tantangan baru dalam pengenalan ekspresi wajah. Beberapa penelitian telah dilakukan untuk mengatasi tantangan ini dengan mengembangkan dataset wajah bermasker dan metode untuk meningkatkan akurasi pengenalan ekspresi wajah pada wajah yang tertutup masker. Salah satu metode yang diusulkan adalah menggunakan dataset yang disimulasikan dengan masker dan melatih model FER pada dataset tersebut [6]. Convolutional Neural Network (CNN) adalah salah satu jenis arsitektur deep learning yang paling populer dan efektif untuk tugas-tugas pengenalan gambar, termasuk pengenalan ekspresi wajah. CNN terdiri dari beberapa lapisan, termasuk lapisan konvolusi, lapisan pooling, dan lapisan fully connected. CNN telah terbukti sangat efektif dalam mengekstraksi fitur dari gambar dan mengklasifikasikan ekspresi wajah dengan akurasi tinggi.

[9] mengusulkan metode pengenalan ekspresi wajah berbasis jaringan saraf dan ekstraksi fitur menggunakan CNN. Bodavarapu dan Srinivas [10] mengembangkan metode pengenalan ekspresi wajah untuk gambar resolusi rendah menggunakan CNN dan teknik denoising. Dhivyaa et al memperkenalkan Multi-feature Integrated Concurrent Neural Network, sebuah arsitektur CNN untuk pengenalan ekspresi wajah manusia [11].

Salah satu model CNN yang dipertimbangkan yaitu penggunaan model ResNet50, ResNet50 adalah jenis khusus dari Residual Network (ResNet) yang terdiri dari 50 lapisan. ResNet dikembangkan untuk mengatasi masalah vanishing gradients yang sering ditemui dalam jaringan neural yang dalam dengan menggabungkan shortcut connections, yang memungkinkan pelatihan jaringan yang lebih dalam tanpa kehilangan informasi penting. ResNet50 menggunakan desain bottleneck untuk blok residu, yang mengurangi jumlah parameter dan mempercepat pelatihan setiap lapisan [9].

D. Hyperparameter Tuning

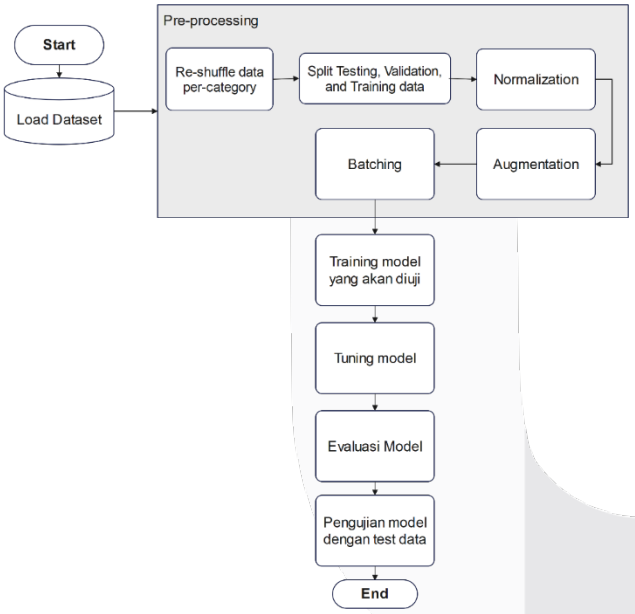
Proses tuning hyperparameter adalah langkah penting yang bertujuan untuk mengoptimalkan parameter yang tidak dipelajari selama pelatihan model, seperti learning rate, jumlah layer, dan ukuran batch. Berbagai metode, termasuk grid search, random search, dan algoritma optimasi seperti Particle Swarm Optimization (PSO), digunakan untuk tuning hyperparameter [12].

Optimasi hyperparameter memainkan peran signifikan dalam meningkatkan kinerja model machine learning. Teknik seperti multi-objective hyperparameter tuning (MOHPT) telah diusulkan untuk memilih hyperparameter secara efektif menggunakan kerangka kerja optimasi metaheuristik multi-objektif yang dilakukan oleh Naseri et al [13]. Selain itu, pendekatan optimasi Bayesian, termasuk Sequential Model-Based Optimization (SMBO), telah menunjukkan efektivitas dalam tuning hyperparameter, mengungguli metode tradisional seperti grid search dan random search [14][15].

III. METODE

Sistem yang dibangun pada penelitian ini adalah pemodelan untuk pendeteksian ekspresi wajah pada wajah yang bermasker dengan menggunakan model klasifikasi CNN seperti ResNet50, VGG19 dan ensemble model. Sistem pemodelan akan menerima input dari dataset yang telah melalui *preprocessing* terlebih dahulu. Diagram alur ini dapat dilihat pada Gambar 3.1. Dataset yang telah tersimpan akan di *Re-shuffe* terlebih dahulu saat *preprocessing* dilakukan sebelum nantinya dataset akan dipisah menjadi tiga jenis penyimpanan yaitu *testing data, training data* dan *validation data.* Proses selanjutnya dataset akan dilakukan normalisasi kepada agar membantu stabilitas dan konvergensi pelatihan model nantinya, augmentasi di lakukan selanjutnya untuk memberikan variasi dalam dataset agar menambah keberagaman data *training* untuk memperluas rekognisi yang dilakukan model nantinya, diakhiri dengan *Batching* yang mengelompokkan data ke dalam ukuran yang tetap dikarenakan dataset berjumlah besar.

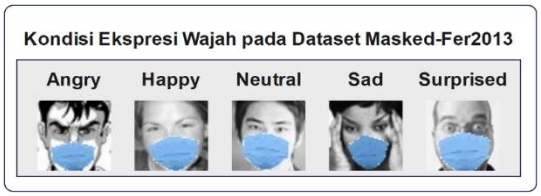
Setelah *preprocessing* dilewati, model-model CNN seperti ResNet50, VGG19 dan ensemble model akan di bangun terlebih dahulu dan dilakukan training pada dataset yang dipersiapkan sebelumnya.



Alur Sistem Pemodelan

A. Dataset

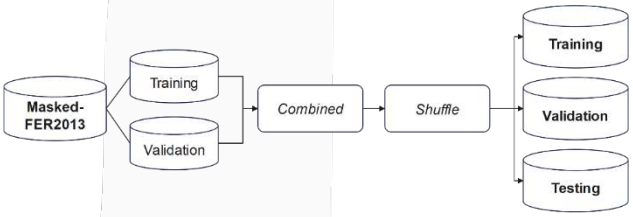
Langkah pertama dalam sistem ini adalah pengumpulan data wajah yang akan digunakan. Dataset yang digunakan dalam penelitian ini adalah "Masked-FER2013", yaitu dataset FER-2013 yang sudah ada sebelumnya digunakan untuk training ekspresi wajah, akan tetapi telah ditambahkan input gambar masker pada setiap gambar yang terkumpulnya dan telah terbagi menjadi data *training* dan data validasi. Data ini berupa berbagai ekspresi wajah dalam *grayscale* dan memiliki 5 jenis emosi wajah seperti marah, takut, netral, senang, sedih, dan terkejut yang dapat dilihat pada gambar 3.2. sebagai contoh-contoh sample data yang telah terkumpul dan belum melalui tahap preprocessing lain.



GAMBAR 3.2

B. *Preprocessing*

Langkah pertama dalam sistem ini adalah pengumpulan data wajah yang akan digunakan. Dataset yang digunakan dalam penelitian ini adalah "Masked-FER2013", yaitu dataset FER-2013 yang sudah ada sebelumnya digunakan untuk training ekspresi wajah, akan tetapi telah ditambahkan input gambar masker pada setiap gambar yang terkumpulnya dan telah terbagi menjadi data *training* dan data validasi. Data ini berupa berbagai ekspresi wajah dalam *grayscale* dan memiliki 5 jenis emosi wajah seperti marah, takut, netral, senang, sedih, dan terkejut yang dapat dilihat pada gambar 3.2. sebagai contoh-contoh sample data yang telah terkumpul dan belum melalui tahap preprocessing lain. Pada *Preprocessing* ini melibatkan proses seperti normalisasi, pengubahan ukuran gambar, jumlah *batch*, dan augmentasi data untuk meningkatkan variasi dalam dataset. *Preprocessing* bertujuan untuk mempersiapkan data agar sesuai dengan input yang diperlukan oleh model CNN yang dipakai. *Preprocessing* mengikuti tahapan sebagai berikut.



GAMBAR 3.3
Alur Sistem Pemodelan

1. Menggabungkan Data

Dataset Masked-Fer2013 yang terdiri training dan validation akan di gabungkan terlebih dahulu untuk mengubah persentasi pembagian data yang akan digunakan untuk training model dan yang akan digunakan untuk pengujian nantinya agar mengurangi beban pelatihan nantinya.

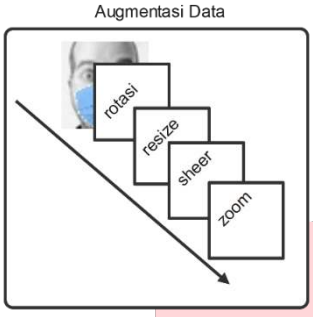
2. *Shuffle Data*

Dataset tersebut akan diacak terlebih dahulu sebagai salah satu step tambahan untuk meningkatkan efektifitas klasifikasi nantinya dengan harapan dapat mencegah bias dan meningkatkan generalisasi.

3. *Split Data*

Data yang diacak tadi akan dibagi dengan bobotnya masing-masing kedalam tiga buah kategori, yaitu Training, Validation, dan Testing.

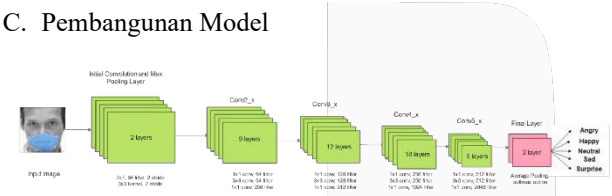
4. Augmentasi Data



GAMBAR 3.4
Augmentasi Data

Augmentasi data dilakukan untuk menyesuaikan dataset terlebih dahulu agar sesuai dengan model yang dibangun, teknik yang dilakukan berupa rotasi gambar, perubahan ukuran, *shear image,* dan zoom, proses ini dapat dilihat pada Gambar 3.4.

C. Pembangunan Model



GAMBAR 3.5
Arsitektur Model

Model yang digunakan dalam penelitian ini adalah ResNet50, VGG19 dan Ensemble model. Pada model ResNet50 telah di-pretrained dengan bobot dari ImageNet yang terdapat pada Gambar 3.5 yang menerima input gambar dalam ukuran (48,48,3) untuk menyesuaikan dimensi pada dataset, untuk model seperti VGG19 akan dilakukan penyesuaian tersendiri dimana model tersebut hanya dapat menerima gambar dengan ukuran diatas 224x224 piksel. Pada proses pengujian model-model CNN itu, setiap model akan dilakukan *tuning* terlebih dahulu pada lapisan atas dari model ini untuk menyesuaikan dengan tugas pengenalan ekspresi wajah dengan melakukan *freezing layers* terkecuali pada empat lapisan teratas. Beberapa layer tambahan seperti *Global Average Pooling* yang akan mengurangi fitur pada proses yang dilakukan untuk mengantisipasi *overfitting*, dilakukan juga normalisasi pada *batch* yang digunakan untuk menstabilkan dan mempercepat pelatihan model, *dropout layer* ditambahkan yang akan melakukan mengatur fraksi dari input unit menjadi nol ketika training dilakukan, berikutnya ditambahkan juga dense layer sebagai dasar aktivasi fungsi ReLU dan terdapat regularisasi yang menambahkan penalti pada fungsi *loss*.

$$J(\theta) = J_{original}(\theta) + \alpha \sum_{i=1}^{m} |w_i| + \beta \sum_{i=1}^{m} w_i^2 \quad (3.1)$$

Dimana $J(\theta)$ berupa regularisai loss function, $J_{orignal}(\theta)$ berupa loss function originalnya, $\alpha \; dan \; \beta$ berupa hyperparameter yang mengkontrol L1 dan L2 masing-masing regularisasi. Dan pada *output layer* berupa *dense layer* dengan 5 unit kelas dan sebuah aktivasi softmax untuk menghasilkan keluaran berupa klasifikasi akhir dari pelatihan model.

Untuk meningkatkan performansi yang di dapat juga, model di terapkan optimasi *hyperparamater* yang diharapkan untuk mendapatkan performansi yang terbaik. Optimasi *hyperparameter* yang dilakukan meliputi perubahan *dropout rate*, jumlah unit pada lapisannya, dan kecepatan *training* yang dilakukan, untuk memastikan tidak terjadi-nya *overfitting* akan ditangani oleh sistem *callback* yang akan menghentikan jika tidak ada peningkatan pada validasi *loss* setelah beberapa epoch terlewati.

D. Evaluasi Kinerja

Untuk Tahap terakhir yaitu dilakukannya evaluasi yang bertujuan untuk mengukur performansi model. Model akan dievaluasi menggunakan Metriks Evaluasi yang berupa *accuracy, precision, Recall, F1-Score* dan *Loss Function*. Berikut penjelasan dari performansi metriks yang didapat saat dilakukannya evaluasi.

1. *Accuracy*

Accuracy berupa rasio prediksi yang benar dari total keseluruhan data yang di gunakan.

$$Accuracy = \frac{\text{Jumlah prediksi yang benar}}{\text{Total jumlah prediksi}} \quad (3.2)$$

2. *Precision*

Precision berupa rasio antara True Positive dan total hasil positif yang diprediksi. Precision menunjukkan seberapa banyak kasus positif yang sebenarnya diidentifikasi oleh model.

$$\text{Precision} = \frac{True\ Positives}{True\ Positives\ +\ False\ Positives} \quad (3.3)$$

3. *Recall*

Recall berupa rasio antara True Positive dan total kasus positif yang sebenarnya. *Recall* menunjukkan seberapa banyak kasus positif yang sebenarnya diidentifikasi oleh model.

$$Recall = \frac{True\ Positives}{True\ Positives\ +\ False\ Negatives} \quad (3.4)$$

4. *F1-Score*

F1-Score Berupa rasio antara precision dan recall. *F1 Score* menunjukkan seberapa baik model menisahkan kasus positif dan negatif, serta seberapa akurat model dalam memprediksi hasil yang benar.

$$F1\ Score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (3.5)$$

5. *Loss*

Loss menghitung performa klasifikasi model yang mengeluarkan nilai diantara 0 sampai 1.

$$Loss = \sum_{i=1}^{N} y_i \log(y_i) \quad (3.6)$$

Dimana $^{y}{}_{i}$ berupa *true label* dan y berupa kemungkinan prediksi.

IV. HASIL DAN PEMBAHASAN

A. Skenario Pengujian

Gambar dinomori secara berurutan. Letak penulisannya di bawah gambar disertai dengan penjelasan. Contoh: Gambar 1(A) Pada percobaan yang dilakukan, data dari dataset Masked-FER2013 yang terdiri dari ekspresi wajah tertutup masker ini yang telah dibagi menjadi tiga bagian: data latih, data validasi dan data testing. Setiap kategori ekspresi (marah, bahagia, netral, sedih, dan terkejut) memiliki sekitar 2000 gambar. Percobaan ini melibatkan pengelolaan data seperti:

Resizing gambar: Semua citra diubah ukurannya menjadi 48x48 piksel untuk memastikan keseragaman dimensi input model untuk mengatasi jika ada kekeliruan ukuran pada salah satu gambar.

Normalisasi: Nilai piksel dinormalisasi ke rentang [0, 1], yang bertujuan untuk mempercepat konvergensi model selama pelatihan.

Augmentasi data: Teknik augmentasi seperti rotasi, flipping horizontal, dan penambahan noise untuk meningkatkan keragaman data pelatihan, sehinga model lebih *robust* terhadap variasi data nyata.

Kemudian mengimplementasikan pendekatan untuk memisahkan bagian atas wajah dan wajah penuh untuk meningkatkan akurasi deteksi ekspresi pada wajah. Dimana setiap model-model CNN yang digunakan akan dilatih pada dua buah bagian wajah yang telah di siapkan dalam 50 epochs dalam 32 batch yang dibantu dengan *early stopping*.

1. ResNet50

Arsitektur model ResNet50 dilatih pada dua bagian dataset yang telah dipisahkan yaitu bagian atas wajah dan seluruh wajah, dimana model telah di pretrained pada ImageNet dan dapat mengadaptasi input citra grayscale dengan melakukan modifikasi pada lapisan konvolusi pertama agar hanya menerima satu channel warna saja dimana model dicoba dilatih pada satu kondisi saja yaitu grayscale.

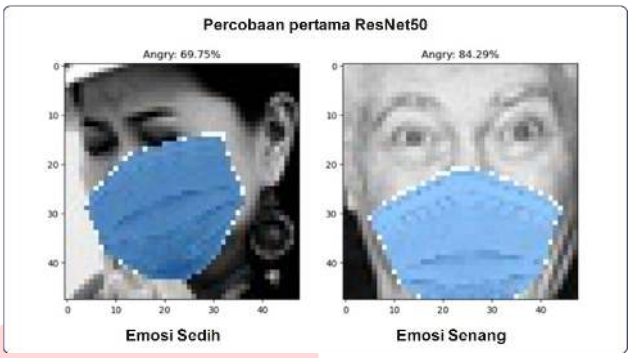
Dalam percobaan pertama Model di latih dengan pengaturan berikut:

- Learning rate: 0.001
- Optimizer: Adam
- Epochs: 50
- Batch size: 32

ResNet50 dengan pengaturan parameter tadi, early stop pada model CNN ini bekerja pada epoch 9 menyebabkan pelatihan berhenti lebih awal dengan hasil:

Training Loss: 0.823 | Training Acc: 64.74% Val Loss: 2.907 | Val Acc: 51.44%

Model tersebut lalu dilakukan testing pada data testing yang telah dipisahkan sebelumnya, ketika dilakukan pengujian pada salah satu citra ekspresi ‘sedih’ citra tersebut terdeteksi sebagai ekspresi ‘marah’, begitupun dengan salah satu citra ‘senang’ terdeteksi sebagai ekspresi ‘marah’. Setelah melakukan pengujian untuk beberapa citra lainnya, pengujian ini menghasilkan konklusi bahwa disini terjadi overfitting pada salah satu ekspresi wajah yang diprediksi yaitu ekspresi ‘marah’ seperti pada Gambar 4.1.



Emosi Sedih dan Senang Terdeteksi Sebagai Emosi Marah

Pada percobaan selanjutnya ditambahkan pengaturan berikut:

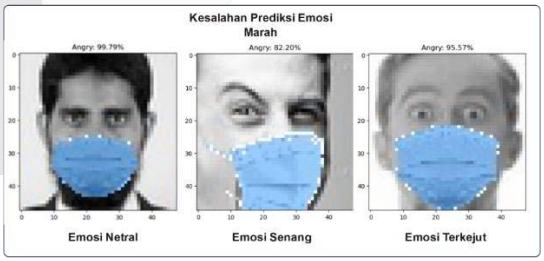
- *Learning Rate* diturunkan menjadi 0.0005.

Hal ini berdampak pada konvergensinya menjadi lebih lambat dan waktu pelatihan menjadi lebih lama, tetapi *validation loss* lebih stabil dibanding sebelumnya yang menandakan penurunan *overfitting*. Pada percobaan ini masih terjadi *overfitting* untuk salah satu kategori ekspresi yang terdeteksi.

Di percobaan terakhir model ditambahkan pengaturan berikut:

- *Weight decay*: 0.01 pada optimizer Adam.
- *Dropout rate*: 0.5 pada *fully connected layer.*

Setelahnya, model mulai dapat melakukan prediksi dengan cukup stabil pada ekspresi wajah ‘sedih’, ‘senang’, dan ‘terkejut’ seperti yang ditampilkan pada Gambar 4.1 Pada ekspresi ‘marah’ masih cukup terdapat *overfit* yang terjadi, pada ekspresi ‘netral’ terjadi kesulitan prediksi dimana perubahan ekspresi yang terlihat tidak sebesar ekspresi lainnya menjadikannya citra yang sering terjadi salah prediksi seperti pada Gambar 4.2.



Kesalahan Prediksi Emosi Marah

2. *Ensemble Model*

Arsitektur model CNN pada percobaan ini yaitu *Ensemble Model* dirancang dengan dua komponen utama, yaitu:

a. Fitur Ekstraktor

Lapisan konvolusi pertama menerima input citra dengan 3 channel dan 64 fitur beserta kernel size ‘3’ dan *padding* ‘1’

untuk membuat keluaran tetap sama. RelU *Activation* digunakan pada setiap lapisan konvolusi untuk menambahkan non-linearitas beserta tiga tingkat konvolusi dengan meningkatkan jumlah filter (64 → 128 → 256) agar dapat mengatasi pola yang lebih kompleks.

b. Klasifikasi

Lapisan diubah dari dimensi spasial menjadi vektor data dilanjut dengan *Fully Connected Layer* yang menerapkan Linear pertama (256 * 6 * 6 → 512) agar fitur yang didapat lebih ringkas dilanjut dengan *dropout* untuk mengurangi *overfitting.*

Percobaan pertama model dilatih dengan pengaturan berikut:

- *Optimizer*: AdamW
- *Learning Rate:* 0.001
- ReduceLRonPlateau digunakan untuk mencegah terjadinya *stagnant* tanpa penurunan validasi *loss.*

Pada percobaan pertama, *Emotion Ensemble Model* ini sudah memberikan akurasi yang cukup bagus akan tetapi *validation loss* yang cukup tinggi yang mengacu pada model yang bisa saja *overfit* ataupun *underfit*, ketika dilakukan pengujian dengan *data testing* ditemukan bahwa pada ekspresi 'marah' cukup terjadi *overfit* dan ekspresi yang cenderung mirip dengan kategori lain pada beberapa citra terjadi *underfit*. Dari kedua hal ini didapatkan konklusi bahwa *behavior* dari kedua model CNN yang digunakan untuk membangun *ensemble model* ini masih terbawa, termasuk kekurangan dari kedua model CNN tersebut yaitu ResNet50 dan VGG19.

Training Loss: 0.179 | Training Acc: 93.4% Val Loss: 2.9353 | Val Acc: 40.42%

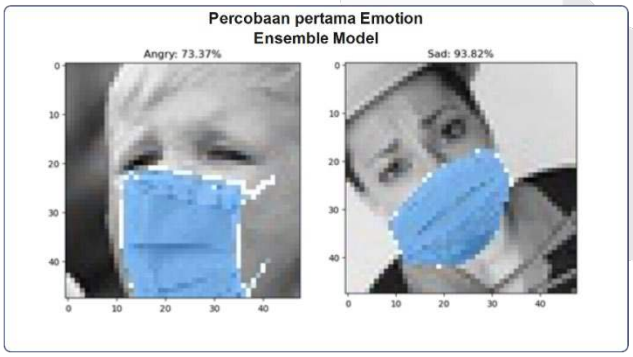
Pada percobaan selanjutnya dilakukan pengaturan sebagai berikut:

- Augmentasi ulang data

Untuk memberikan generalisasi model pada data yang divalidasi.

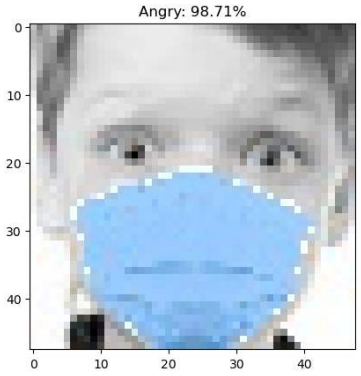
- *Dropout Rate*: 0.5 pada *fully connected layer.*

ResNet50 dan VGG19 *dituning* pada *layer* terakhir dengan *learning rate* 0.0001



GAMBAR 4.3
Prediksi Emosi Yang Tepat

Model CNN *Ensemble Model* ini dapat memprediksi dengan cukup stabil untuk beberapa jenis ekspresi dengan wajah yang terlihat jelas seperti pada Gambar 4.3 akan tetapi waktu pelatihan menjadi lebih lama dari sebelumnya. Hal lainnya yang terlihat yaitu ekspresi 'terkejut' yang cukup banyak terjadi salah prediksi dikarenakan memiliki kemiripan ekspresi wajah dengan emosi 'marah' dalam hal mata yang terbuka lebar, dan juga ekspresi 'sedih' dalam hal alis yang turun dan melebar menjadikan prediksi untuk beberapa ekspresi 'terkejut' yang tidak terlihat jelas menjadi salah prediksi seperti yang ditampilkan pada Gambar 4.4.



GAMBAR 4.4
Prediksi Emosi Yang Tepat

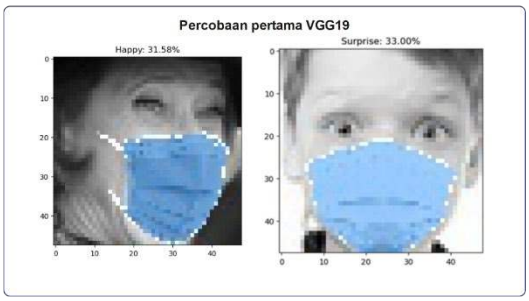
1. VGG19

Arsitektur model ResNet50 dilatih pada dua bagian dataset yang telah dipisahkan yaitu bagian atas wajah dan seluruh wajah, dimana model telah di *pretrained* pada ImageNet dan dapat mengadaptasi input citra *grayscale* dengan melakukan modifikasi pada lapisan konvolusi pertama agar hanya menerima satu channel warna saja dimana model dicoba dilatih pada satu kondisi saja yaitu *grayscale*. Arsitektur model VGG19 membutuhkan perlakuan khusus dimana ukuran citra dataset di ubah menjadi 224x224 piksel. VGG19 ini akan di bagi dua dalam memproses citra pada keseluruhan wajah dan memfokuskan pada bagian atas wajah. Kemudian kedua output tersebut akan di gabungkan melalui lapisan yang terhubung ke klasifikasi akhir dalam menentukan jenis emosi yang ada.

Pada percobaan pertama model dilatih dengan pengaturan berikut:

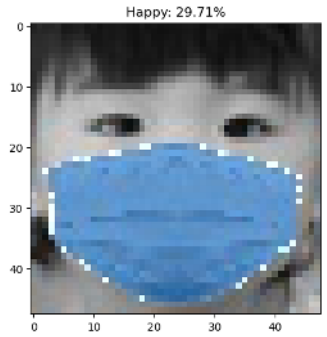
- Adam *optimizer* dengan *learning rate* di 0.001,
- *Cross-Entropy Loss* untuk mengukur kesalahan klasifikasi.

Pada semua percobaan yang dilakukan, pada model VGG19 mendapat akurasi yang lebih rendah dibandingkan kedua model CNN yang telah dibahas sebelumnya akan tetapi pada beberapa citra ketika mendeteksi ekspresi wajah walaupun memiliki tingkat *confidence* yang cukup rendah, model ini dapat memprediksi ekspresi wajah yang benar, seperti yang bisa dilihat pada gambar 4.5 menampilkan ekspresi 'terkejut' walau dengan *confidence level* di 33%.



Emosi Prediksi Yang Benar

Untuk beberapa citra wajah yang memiliki kecondongan pada dua jenis ekspresi yang mirip tanpa menampilkan perubahan raut wajah yang berlebihan untuk menggambarkan emosi yang dirasakan, masih terjadi kesalahan prediksi seperti pada Gambar 4.6 yang menampilkan ekspresi 'netral' seorang gadis perempuan akan tetapi dikarenakan kecondongan bentuk mata nya yang mewakili salah satu ekspresi 'senang' model CNN VGG19 ini pun akhirnya memprediksi bahwa citra tersebut adalah citra dengan ekspresi 'senang'.



GAMBAR 4.6
Emosi Netral Terdeteksi Senang

B. Hasil Pengujian

1. Hasil ResNet50

Hasil Percobaan model CNN ResNet50 menunjukkan hasil performa berikut:

TABEL 4.1
PERFORMA RESNET50

Model	Akurasi (%)	Presisi (%)	Recall (%)	F1-Score (%)
ResNet50	89.5	87.3	86.7	87.0

Berdasarkan Tabel 4.1, Kinerja lebih detail dari ResNet50 dalam mendeteksi ekspresi wajah bermasker dianalisis untuk setiap kategori ekspresi berikut:

TABEL 4.2
AKURASI EKSPRESI RESNET50

Ekspresi	Akurasi (%)
Bahagia	87.0
Marah	87.5
Sedih	88.2
Netral	87.1
Terkejut	91.4

Performa ResNet50 pada Tabel 4.2 menunjukkan kemampuan yang sangat baik dalam mengenali pola visual pada citra wajah bermasker.

2. Hasil *Ensemble Model*

Hasil Percobaan model CNN *Ensemble Model* menunjukkan hasil performa berikut:

TABEL 4.3
PERFORMA ENSEMBLE MODEL

Model	Akurasi (%)	Presisi (%)	Recall (%)	F1-Score (%)
Emotion Ensemble	82.4	84.7	82.3	83.5

Berdasarkan Tabel 4.3, Kinerja lebih detail dari Kinerja *Emotion Ensemble* dalam mendeteksi ekspresi wajah bermasker dianalisis untuk setiap kategori ekspresi:

TABEL 4.4
AKURASI EKSPRESI ENSEMBLE MODEL

Ekspresi	Akurasi (%)
Bahagia	82.5
Marah	88.9
Sedih	76.8
Netral	75.0
Terkejut	89.0

Pada Tabel 4.4 menampilkan pendekatan *ensemble model* terbukti efektif dalam meningkatkan akurasi dan stabilitas model.

3. Hasil VGG19

Hasil Percobaan model CNN VGG19 menunjukkan hasil performa berikut:

TABEL 4.5
PERFORMA VGG19

Model	Akurasi (%)	Presisi (%)	Recall (%)	F1-Score (%)
VGG19	72.4	73.0	72.1	72.0

Pada Tabel 4.5 Kinerja VGG19 dalam mendeteksi ekspresi wajah bermasker dianalisis untuk setiap kategori ekspresi:

TABEL 4.6
AKURASI EKSPRESI VGG19

Ekspresi	Akurasi (%)
Bahagia	70.0
Marah	77.5
Sedih	60.0
Netral	62.1
Terkejut	83.0

Arsitektur VGG19, meskipun lebih sederhana dibandingkan ResNet50, menghasilkan performa yang cukup baik.

C. Analisis Hasil Pengujian

Analisis ini dilakukan untuk mengevaluasi performa dan efisiensi dari masing-masing model berdasarkan hasil pelatihan dan evaluasi menggunakan metrik utama seperti akurasi, presisi, recall, dan F1-Score. Selain itu, analisis yang didapat menjadi pertimbangan keunggulan dan kelemahan dari setiap arsitektur model CNN yang dipakai dalam menangani Masked-FER2013 dengan kompleksitas tinggi.

1. Analisis ResNet50

Arsitektur model CNN ResNet50 menunjukkan performa yang cukup tinggi di 89.51%, walau dengan sedikit *overfit* pada emosi marah, model ini dapat mengenali pola kompleks pada citra wajah bermasker. Namun, Kebutuhan komputasi model ini lebih tinggi dibandingkan model lainnya. Model ResNet50 ini memiliki stabilitas yang baik dalam kategori ekspresi bahagia dan terkejut.

Beberapa faktor yang dapat terlihat yaitu penggunaan *learning rate* di 0.001 memberikan konvergensi yang cukup stabil dikarenakan ketika *learning rate* di atur pada nilai tinggi membuat model melewati batas minimum lokal. *Batch size* menjadi pengaruh besar pada lama nya pelatihan dikarenakan jika terlalu besar akan mengurangi kemampuan model untuk mendeteksi variasi dalam data dan penggunaan 50 *epoch* memberikan cukup waktu untuk melatih model untuk mengenali pola kompleks secara sederhana dan cepat.

Resnet50 ini dirancang untuk menangkap *global feature* dari citra yang di latih, efektif untuk emosi dengan ekspresi yang jelas. Namun, pendekatan ini kurang optimal untuk mengenali ekspresi dengan akurasi tinggi. Walaupun ResNet50 dapat menangani *vanishing gradient*, model ini masih mengalami kesulitan dalam mendeteksi ekspresi yang lebih halus dan tidak memungkinkan untuk menangkap semua jenis ekspresi dengan tepat.

2. Analisis *Ensemble Model*

Arsitektur model CNN *Ensembel Model* mencapai akurasi 82.49% dengan metode pendekatan *ensemble* yang memanfaatkan kelebihan dari model ResNet50 dan VGG19 yang telah dibangun secara kolektif dapat meningkatkan performa dan stabilitas model, terutama untuk dataset dengan variasi tinggi. Model ini dapat mengatasi bias dan *overfitting* yang sering terjadi karna mampu menangani kompleksitas dataset dengan kombinasi fitur tadi. Namun, kompleksitas tambahan dari penggabungan model ini membutuhkan daya komputasi tinggi dan memerlukan *tuning* yang cermat untuk menghindari *overfit* tersebut.

3. Analisis VGG19

Arsitektur model CNN VGG19 terlihat lebih dangkal dibanding kedua model yang ditampilkan, menghasilkan akurasi lebih rendah di 72.44%. Efisiensi VGG19 dalam menangkap pola kompleks pada wajah kurang optimal, terbatas dalam kapasitas untuk menangkap berbagai variasi emosi yang terlihat dan setelah beberapa kali pelatihan model pada dataset tetap menunjukkan performa yang lebih rendah dibandingkan dengan kedua model CNN lainnya. Dataset yang kompleks menjadi halangan, terutama untuk emosi yang tidak terlalu jelas, penggunaan arsitektur model *dual-path* untuk membagi atensi pada bagian wajah atas dan keseluruhan wajah cukup meningkatkan deteksi ekspresi yang bergantung pada ekspresi di bagian atas wajah seperti ekspresi marah.

V. KESIMPULAN

Berdasarkan eksperimen yang dilakukan pada tiga model, yaitu ResNet50, *Ensemble Model,* dan VGG19 diperoleh beberapa temuan utama dimana masing-masing arsitektur memiliki keunggulan dan kelemahan dalam menangani dataset Masked-FER2013 yang kompleks.

Model CNN ResNet50 menunjukkan performa tertinggi dengan akurasi 89.51% mampu menangkap pola kompleks pada citra wajah bermasker, namun membutuhkan daya komputasi tinggi dan rentan terhadap *overfitting* pada kategori ekspresi tertentu. Sementara itu, *Ensemble Model* berhasil menggabungkan kelebihan ResNet50 dan VGG19, mencapai akurasi 82.49% namun dengan tingkat kompleksitas yang lebih tinggi dan memerlukan *tuning* parameter yang teliti. Model CNN VGG19, meskipun menunjukkan efisiensi komputasi yang ringan, hanya mencapai akurasi 72.44% karena keterbatasannya menangkap pola kompleks pada dataset dengan variasi tinggi.

Secara keseluruhan, penelitian ini menunjukkan bahwa pendekatan ensemble dapat meningkatkan stabilitas prediksi dibandingkan dengan model individual, namun dengan konsekuensi peningkatan daya komputasi yang dibutuhkan dan model ResNet memiliki nilai akurasi yang tinggi dalam mendeteksi beberapa kategori ekspresi tertentu dibandingkan model lainnya. Secara tidak langsung menyadarkan bahwa pemilihan arsitektur model yang tepat dengan kompleksitas dataset dan tujuan sistem yang akan dibangun adalah hal yang penting dilakukan, ditambah dengan kompleksitas ekspresi yang dilatih seperti pada penelitian sebelumnya yang mendapatkan hasil akurasi yang lebih tinggi dan akurat dan dilatih pada kualitas dataset yang lebih baik akan tetapi hanya memiliki tiga jenis kategori ekspresi saja.

REFERENSI

[1] Yang, B., Wu, J., & Hattori, G. (2021, September 19). Face Mask Aware Robust Facial Expression Recognition During The Covid-19 Pandemic. https://doi.org/10.1109/icip42928.2021.9506047

[2] ELsayed, Y., Elsayed, A E M., & Abdou, M A. (2023, April 1). An automatic improved facial expression recognition for masked faces. Springer Science+Business Media, 35(20), 14963-14972. https://doi.org/https://doi.org/10.1007/s00521-023-08498-w

[3] Küntzler, T., Höfling, T T A., & Alpers, G W. (2021, May 5). Automatic Facial Expression Recognition in Standardized and Non-standardized Emotional Expressions. Frontiers Media, 12. https://doi.org/https://doi.org/10.3389/fpsyg.2021.627561

[4] Huang, B., Wang, Z., Wang, G., Jiang, K., He, Z., Zou, H., & Zou, Q. (2021, October 1). Masked Face Recognition Datasets and Validation. https://doi.org/https://doi.org/10.1109/iccvw54120.2021.00172

[5] Niu, B., Gao, Z., & Guo, B. (2021, January 12). Facial Expression Recognition with LBP and ORB Features. Hindawi Publishing Corporation, 2021, 1-10. https://doi.org/https://doi.org/10.1155/2021/8828245

[6] Pazhoohi, F., Forby, L., & Kingstone, A. (2021). Facial masks affect emotion recognition in the general population and individuals with autistic traits. Plos One, 16(9), e0257740. https://doi.org/10.1371/journal.pone.0257740
[7] Thomas, L., Castell, C., & Hecht, H. (2022). How facial masks alter the interaction of gaze direction, head orientation, and emotion recognition. Frontiers in Neuroscience, 16. https://doi.org/10.3389/fnins.2022.937939
[8] McCrackin, S., Capozzi, F., Mayrand, F., & Ristic, J. (2021). The influence of face masks on emotion recognition and the role of individual differences.. https://doi.org/10.31234/osf.io/dx94v
[9] Tegani, S., & Abdelmoutia, T. (2021). Using COVID-19 Masks Dataset to Implement Deep Convolutional Neural Networks For Facial Emotion Recognition. *2021 4th International Symposium on Advanced Electrical and Communication Technologies (ISAECT)*, 1-5.
[10] Bodavarapu, P. and Srinivas, P. (2021). Facial expression recognition for low resolution images using convolutional neural networks and denoising techniques. Indian Journal of Science and Technology, 14(12), 971-983. https://doi.org/10.17485/ijst/v14i12.14
[11] Dhivyaa, C., Dhivyaa, K., Nithya, K., & Karthika, S. (2022). Multi-feature integrated concurrent neural network for human facial expression recognition. 網際網路技術學刊, 23(6), 1263-1274. https://doi.org/10.53106/160792642022112306009
[12] Raji, I., Bello-Salau, H., Umoh, I., Onumanyi, A., Adegboye, M., & Salawudeen, A. (2022). Simple deterministic selection-based genetic algorithm for hyperparameter tuning of machine learning models. Applied Sciences, 12(3), 1186. https://doi.org/10.3390/app12031186
[13] Naseri, H., Waygood, E., Wang, B., & Patterson, Z. (2022). Application of machine learning to child mode choice with a novel technique to optimize hyperparameters. International Journal of Environmental Research and Public Health, 19(24), 16844. https://doi.org/10.3390/ijerph192416844
[14] Yogatama, D., Kong, L., & Smith, N. (2015). Bayesian optimization of text representations.. https://doi.org/10.18653/v1/d15-1251
[15] Joy, T., Rana, S., Gupta, S., & Venkatesh, S. (2019). Fast hyperparameter tuning using bayesian optimization with directional derivatives.. https://doi.org/10.48550/arxiv.1902.02416
[16] Yang, B., Wu, J., & Hattori, G. (2020). *Facial Expression Recognition with the advent of face masks. 19th International Conference on Mobile and Ubiquitous Multimedia.* doi:10.1145/3428361.3432075
[17] M. Saleem Abdullah, S., & Abdulazeez, A. M. . (2021). Facial Expression Recognition Based on Deep Learning Convolution Neural Network: A Review. *Journal of Soft Computing and Data Mining*, *2*(1), 53-65. https://publisher.uthm.edu.my/ojs/index.php/jscdm/article/view/7906
[18] P. Barros and A. Sciutti, "I Only Have Eyes for You: The Impact of Masks On Convolutional-Based Facial Expression Recognition," *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, Nashville, TN, USA, 2021, pp. 1226-1231, doi: 10.1109/CVPRW53098.2021.00134.