Face\_Expression\_Recognition Menggunakan Convolutional Neural Network (CNN)

**Muhammad Hafizha Ramadani 1, Rahman Fauza,** **Elza Nuraripin1**, **M Yunus1**

1Sarjana Teknik Informatika, Fakultas Ilmu Komputer Dan Sistem Informasi,

Universitas Kebabangsaan Republik Indonesia, Bandung, 40263, Indonesia

|  |  |  |
| --- | --- | --- |
| **Article Info** |  | **ABSTRACT** |
| ***Article history:***  Received month dd, yyyy  Revised month dd, yyyy  Accepted month dd, yyyy |  | Klasifikasi ekspresi wajah merupakan salah satu tugas penting dalam bidang pengenalan pola dan visi komputer. Penelitian ini bertujuan untuk mengembangkan model AI berbasis arsitektur VGG16 untuk klasifikasi ekspresi wajah pada dataset AffectNet dan FER13. Model ini dirancang untuk mengenali ekspresi seperti bahagia, sedih, marah, takut, jijik, terkejut, dan netral. Arsitektur VGG16 dipilih karena kemampuannya yang unggul dalam ekstraksi fitur visual melalui penggunaan lapisan konvolusi yang dalam dan filter kecil berukuran 3×3.  Dataset diproses melalui teknik augmentasi data, termasuk rotasi, flipping, dan penyesuaian kontras, untuk meningkatkan keragaman data dan ketahanan model terhadap kondisi dunia nyata, seperti variasi pencahayaan, ekspresi intensitas rendah, dan sudut wajah yang berbeda. Transfer learning digunakan dengan memanfaatkan bobot awal yang telah dilatih pada dataset ImageNet untuk mempercepat konvergensi dan meningkatkan akurasi. |
| ***Keywords:***  ***Convolutional Neural Network***  ***Face Expression*** |
| *This is an open access article under the* [*CC BY-SA*](https://creativecommons.org/licenses/by-sa/4.0/) *license.*  A picture containing text, clipart  Description automatically generated |
| ***Corresponding Author:***  Muhammad Hafizha Ramadani, Rahman Fauza, Elza Nuraripin.  Sarjana Informatika, Fakultas Informatika dan Sistem Informasi, Universitas Kebangsaan Indonesia, Jalan Jl.Terusan Halimun No.37 Bandung 40263, Indonesia.  Email: support@iuk.ukri.org | | |

1. **INTRODUCTION**

Wajah merupakan stimulus sosial penting yang sering menjadi fokus penelitian ilmiah. Ekspresi wajah adalah salah satu karakteristik perilaku yang dapat merefleksikan emosi seseorang [1]. Berdasarkan perubahan pada otot-otot wajah, penelitian sebelumnya mengidentifikasi enam bentuk emosi dasar yang secara universal diterima sejak tahun 1992, yaitu marah, jijik, takut, bahagia, sedih, dan terkejut. Dalam konteks modern, dataset seperti AffectNet dan FER13 telah banyak digunakan untuk mengkaji pengenalan emosi berbasis ekspresi wajah. Dataset ini mencakup berbagai ekspresi wajah yang lebih beragam dan anotasi yang mendukung analisis emosi, memungkinkan pengembangan model pengenalan emosi yang lebih akurat dan komprehensif. Melalui ekspresi wajah yang didokumentasikan dalam dataset ini, emosi individu dapat dipahami dengan lebih mendalam.

Dengan pembagian data menjadi dua yaitu train dan validation dengan jumlah gabungan dari dataset AffectNet dan FER13 mendapatkan data train sebanyak 51.161 dan validation sebanyak 12.686. Penggabungan dataset AffectNet dan FER13 memberikan berbagai keuntungan yang signifikan dalam pengembangan model pengenalan emosi berbasis ekspresi wajah. Dengan menggabungkan kedua dataset, keragaman data meningkat, mencakup variasi ekspresi emosi, intensitas, serta cakupan demografi seperti usia, gender, dan etnis yang lebih luas. Hal ini mendukung pengembangan model yang lebih generalis dan mampu menangani data dari domain yang berbeda. Selain itu, dataset yang lebih besar membantu mencegah overfitting dan meningkatkan akurasi model, terutama dalam mengenali emosi dengan frekuensi rendah seperti jijik dan takut. Penggabungan ini juga memberikan manfaat bagi model pembelajaran mendalam, yang membutuhkan volume data besar untuk mencapai performa optimal. Lebih lanjut, cakupan kategori emosi yang lebih luas dari AffectNet melengkapi keterbatasan FER13, memungkinkan pengenalan emosi yang lebih detail dan kompleks. Dengan dataset gabungan ini, model dapat dilatih dan dievaluasi secara lebih komprehensif, mendukung aplikasi dunia nyata seperti interaksi manusia-komputer, analisis perilaku, dan diagnosis kesehatan mental.

Dalam pengenalan emosi berbasis ekspresi wajah, ukuran dan keragaman dataset memainkan peran penting dalam meningkatkan performa model deep learning. Dengan menggabungkan dataset AffectNet dan FER13, yang keduanya memiliki karakteristik unik, diperoleh dataset yang lebih besar dan beragam. AffectNet menawarkan anotasi mendalam dengan cakupan emosi yang lebih luas, sementara FER13 menyediakan data yang lebih terstruktur dan sederhana. Penggabungan ini mendukung pelatihan model berbasis arsitektur VGG16, yang dikenal karena kedalamannya dan kemampuan ekstraksi fitur visual yang efektif melalui jaringan convolutional. Dataset gabungan membantu meningkatkan generalisasi model, meminimalkan overfitting, dan meningkatkan akurasi dalam pengenalan emosi, terutama pada kategori emosi dengan frekuensi rendah seperti jijik atau takut. Selain itu, arsitektur VGG16 memanfaatkan dataset besar ini untuk mendukung transfer learning, di mana bobot awal yang dilatih pada dataset besar dapat disesuaikan untuk aplikasi spesifik. Dengan demikian, kombinasi dataset AffectNet dan FER13 bersama arsitektur VGG16 menghasilkan model yang lebih adaptif dan akurat dalam mendeteksi emosi dari ekspresi wajah di berbagai domain.

Dataset yang digunakan dalam penelitian ini dibagi menjadi dua bagian, yaitu *training dataset* dan *validation dataset*. *Training dataset* digunakan untuk melatih model VGG16, di mana data diperkaya melalui augmentasi seperti rotasi, flipping, dan penyesuaian kontras untuk meningkatkan keragaman dan ketahanan model terhadap kondisi dunia nyata.

Sementara itu, *validation dataset* digunakan untuk mengevaluasi performa model selama pelatihan, memastikan model dapat menggeneralisasi data yang belum terlihat. Dataset ini tidak dikenai augmentasi tambahan agar mencerminkan kondisi data sesungguhnya. Pembagian dilakukan dengan proporsi 80:20, sehingga model dapat dilatih pada data yang cukup besar sambil tetap dievaluasi secara obyektif.

Hasil yang diperoleh dari penelitian ini masih jauh dari kata memuaskan. Proses pelatihan dilakukan selama 20 epoch, namun akurasi yang dicapai masih sangat rendah. Model menghasilkan *training accuracy* sebesar 30.66% dan *validation accuracy* sebesar 30.93%, yang menunjukkan performa model belum optimal dalam mengenali ekspresi wajah.

Salah satu faktor utama yang memengaruhi hasil ini adalah keterbatasan perangkat keras yang digunakan. Dalam penelitian ini, proses pelatihan dilakukan menggunakan laptop dengan spesifikasi kelas entry-level. Hal ini berdampak signifikan pada durasi pelatihan, di mana untuk menyelesaikan 20 epoch saja membutuhkan waktu rata-rata 3 jam. Dengan dataset yang besar, arsitektur VGG16 yang memiliki jutaan parameter membutuhkan waktu komputasi yang jauh lebih tinggi untuk melakukan iterasi yang cukup guna mencapai konvergensi.

Selain itu, jumlah epoch yang terbatas juga menjadi kendala. Sebagai model berbasis deep learning, VGG16 membutuhkan lebih banyak epoch agar dapat mempelajari pola-pola dari dataset yang kompleks. Dengan hanya 20 epoch, model belum sempat mempelajari fitur secara optimal, sehingga hasil akurasi rendah tidak terhindarkan.

Keterbatasan ini menunjukkan bahwa perangkat keras berperforma tinggi, seperti GPU yang dirancang khusus untuk pembelajaran mendalam, sangat diperlukan untuk menangani dataset besar dan model yang kompleks. Selain itu, strategi seperti *transfer learning*, *early stopping*, atau pengurangan dimensi dataset dapat dipertimbangkan untuk mengurangi kebutuhan sumber daya komputasi tanpa terlalu mengorbankan performa model.

1. **METHOD**

Penelitian ini dilakukan dalam beberapa tahap. Tahapan-tahapan tersebut dapat dilihat pada gambar1



**Gambar. 1** Tahapan metode

Penelitian ini dimulai dengan mengimpor *library* penting seperti NumPy, Pandas, dan TensorFlow, yang diperlukan untuk mendukung proses pelatihan dan pengujian model. Langkah berikutnya adalah mengimpor dataset, yaitu FER13 dan AffectNet, yang diunduh langsung untuk digunakan dalam penelitian ini.

Setelah dataset diimpor, dilakukan proses image preprocessing untuk mempersiapkan citra agar dapat diproses oleh model. Preprocessing ini mencakup penyesuaian ukuran gambar, normalisasi nilai piksel, dan augmentasi data. Tujuannya adalah memastikan data siap untuk proses ekstraksi fitur menggunakan Convolutional Neural Network (CNN).

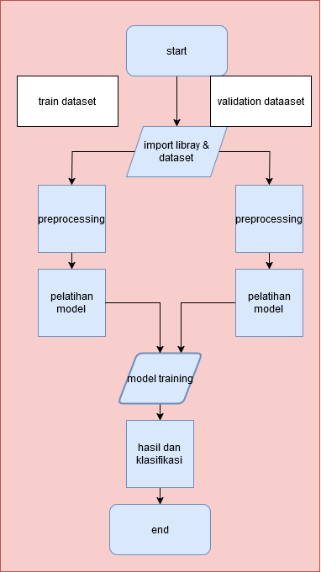
Arsitektur CNN yang digunakan dalam penelitian ini adalah VGG16, yang terdiri dari 16 lapisan. Model ini dirancang untuk mengenali pola visual pada citra dengan akurasi tinggi. Setelah arsitektur dirancang, model dilatih menggunakan dataset yang telah diproses. Proses ini menghasilkan model terlatih yang dapat digunakan untuk melakukan klasifikasi ekspresi wajah.

Tahap akhir adalah pengujian model untuk mengevaluasi performanya dalam mengenali ekspresi wajah. Hasil pelatihan dan pengujian disimpan dalam bentuk model terlatih, yang kemudian digunakan untuk melakukan klasifikasi pada citra baru.

* 1. **import library & dataset**

Tahap awal dalam penelitian ini adalah mengimpor *library* yang diperlukan untuk mendukung proses pelatihan dan pengujian model. *Library* utama yang digunakan meliputi NumPy untuk operasi numerik seperti normalisasi nilai piksel citra, Pandas untuk membaca dan mengelola dataset, serta Matplotlib atau Seaborn untuk memvisualisasikan data seperti distribusi kelas dan hasil pelatihan. Selain itu, TensorFlow dan Keras digunakan sebagai *framework* utama dalam membangun dan melatih model Convolutional Neural Network (CNN) dengan arsitektur VGG16.

Setelah *library* diimpor, langkah berikutnya adalah mengimpor dataset yang digunakan dalam penelitian ini, yaitu FER13 dan AffectNet. Dataset ini diunduh dari sumber resmi, di mana FER13 menyediakan gambar wajah terklasifikasi dalam beberapa kategori emosi, sementara AffectNet menawarkan dataset yang lebih besar dan kompleks dengan variasi kondisi dunia nyata.



**Gambar. 2** flow system

Dataset yang telah diimpor kemudian diproses untuk memastikan kompatibilitasnya dengan model. Proses ini meliputi membaca data, memisahkannya menjadi *training set* dan *validation set* dengan proporsi 80:20, serta melakukan *image preprocessing*. Preprocessing ini mencakup penyesuaian ukuran gambar menjadi 224x224 piksel sesuai dengan input arsitektur VGG16, normalisasi nilai piksel ke rentang [0, 1], dan augmentasi data untuk meningkatkan variasi dataset. Tahapan ini memastikan bahwa data siap digunakan untuk melatih dan menguji model sehingga dapat menghasilkan performa yang optimal.

Hal penting yang perlu diperhatikan adalah bahwa dataset AffectNet berisi citra ekspresi wajah dalam format RGB, sedangkan dataset FER13 menggunakan format grayscale. Untuk memastikan konsistensi data dan mempermudah proses *preprocessing* serta pelatihan, format kedua dataset harus diseragamkan. Dalam penelitian ini, citra pada dataset AffectNet diubah dari format RGB menjadi grayscale agar sesuai dengan format dataset FER13. Konversi ini dilakukan dengan cara menghitung nilai rata-rata dari tiga saluran warna (merah, hijau, dan biru) di setiap piksel untuk menghasilkan gambar grayscale. Dengan menyamakan format kedua dataset, proses *preprocessing* menjadi lebih sederhana, dan model dapat memproses data tanpa perbedaan format yang dapat memengaruhi performa. Langkah ini memastikan bahwa model dapat dilatih dengan dataset yang konsisten, sehingga hasil klasifikasi menjadi lebih akurat dan proses pelatihan berjalan lebih efisien.

* 1. **Image preprocessing**

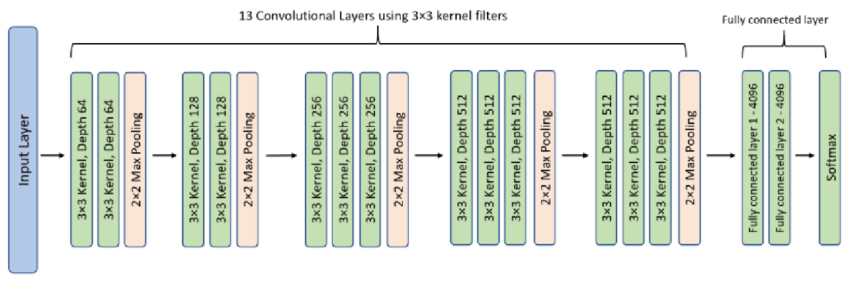
Tahap *image preprocessing* dilakukan untuk mempersiapkan citra sebelum digunakan dalam pelatihan dan pengujian model. Proses ini mencakup langkah-langkah berikut:

1. Convert Grayscale to RGB  
   Proses ini mengubah citra masukan dari format grayscale menjadi RGB dengan menduplikasi nilai piksel grayscale ke saluran merah, hijau, dan biru. Langkah ini dilakukan untuk menyelaraskan dataset FER13 (grayscale) dengan AffectNet (RGB).
2. Resize  
   Semua citra diubah ukurannya menjadi 224x224 piksel untuk menyesuaikan dengan dimensi input yang dibutuhkan oleh arsitektur VGG16.
3. Normalisasi  
   Nilai piksel citra dinormalisasi ke rentang 0 hingga 1 dengan membagi setiap nilai piksel dengan nilai maksimalnya (255).
4. AugmentasiCitra  
   Teknik augmentasi seperti rotasi, flipping horizontal, zoom, dan translasi diterapkan untuk meningkatkan variasi dataset tanpa menambah data baru.
5. DataFrame  
   Dataset diorganisasi ke dalam format DataFrame menggunakan *library* Pandas. DataFrame menyimpan informasi penting seperti jalur file citra, label emosi, dan kategori (*training* atau *validation set*). Format ini mempermudah pengelolaan data, seperti memisahkan data berdasarkan kelas, menghitung distribusi label, atau memantau hasil preprocessing.
6. Image data Generator

Untuk mempermudah augmentasi selama pelatihan, digunakan Image Data Generator dari Keras. Alat ini menghasilkan citra baru secara dinamis dengan menerapkan augmentasi secara real-time, seperti rotasi, flipping, atau zoom. Dengan menggunakan generator ini, proses augmentasi tidak memerlukan penyimpanan fisik citra hasil augmentasi, sehingga menghemat ruang dan mempercepat alur kerja.

* 1. **Perancangan Model CNN**

**2.3.1 Arsitektur VGG16**



**Gambar. 3** Arsitektur VGG16

Model yang digunakan dalam penelitian ini adalah arsitektur VGG16, yang merupakan salah satu model konvolusional yang terkenal dan banyak digunakan untuk tugas pengenalan gambar. Model ini terdiri dari beberapa lapisan yang disusun secara berurutan, masing-masing dengan fungsi tertentu dalam ekstraksi fitur dan klasifikasi gambar. Berikut adalah rincian lapisan-lapisan dalam model VGG16 yang digunakan.

1. Lapisan Input (Input Layer)

* Output Shape: (None, 48, 48, 3)
* Parameter: 0
* Lapisan input menerima gambar dengan ukuran 48x48 piksel dan 3 saluran warna (RGB). Dimensi (None, 48, 48, 3) menunjukkan bahwa batch ukuran gambar dapat bervariasi, tetapi setiap gambar memiliki dimensi 48x48 dengan 3 saluran warna.

2. Blok 1: Lapisan Konvolusi dan Pooling

* block1\_conv1 (Conv2D):
  + Output Shape: (None, 48, 48, 64)
  + Parameter: 1792
  + Lapisan konvolusi pertama dengan 64 filter berukuran kernel 3x3, bertujuan untuk mengekstraksi fitur dasar seperti tepi dan tekstur dari gambar input.
* block1\_conv2 (Conv2D):
  + Output Shape: (None, 48, 48, 64)
  + Parameter: 36,928
  + Lapisan konvolusi kedua dengan 64 filter tambahan, yang berfungsi untuk menangkap fitur tingkat lanjut dari gambar.
* block1\_pool (MaxPooling2D):
  + Output Shape: (None, 24, 24, 64)
  + Parameter: 0
  + Lapisan pooling dengan ukuran 2x2 dan stride 2 digunakan untuk mengurangi dimensi spasial gambar dari 48x48 menjadi 24x24 sambil menjaga fitur penting.

3. Blok 2: Lapisan Konvolusi dan Pooling

* block2\_conv1 (Conv2D):
  + Output Shape: (None, 24, 24, 128)
  + Parameter: 73,856
  + Lapisan konvolusi pertama di blok kedua, dengan 128 filter berukuran 3x3, bertujuan untuk mengekstraksi fitur yang lebih kompleks dari gambar.
* block2\_conv2 (Conv2D):
  + Output Shape: (None, 24, 24, 128)
  + Parameter: 147,584
  + Lapisan konvolusi kedua di blok kedua, juga dengan 128 filter, untuk menangkap fitur tingkat lanjut yang lebih dalam.
* block2\_pool (MaxPooling2D):
  + Output Shape: (None, 12, 12, 128)
  + Parameter: 0
  + Lapisan pooling yang mengurangi dimensi spasial gambar menjadi 12x12, yang memungkinkan kompresi informasi sambil mempertahankan fitur penting.

4. Blok 3: Lapisan Konvolusi

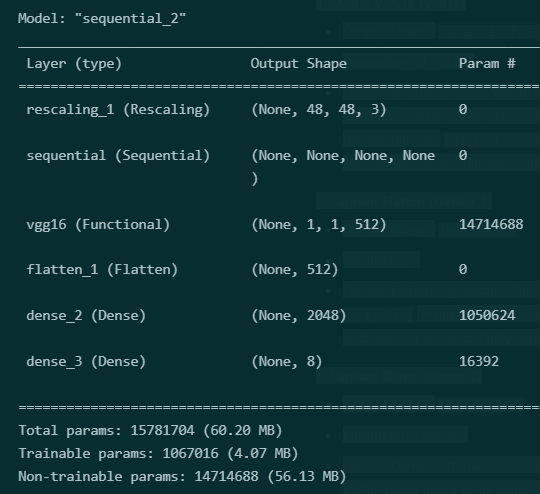
* block3\_conv1 (Conv2D):
  + Output Shape: (None, 12, 12, 256)
  + Parameter: 295,168
  + Lapisan konvolusi pertama di blok ketiga, dengan 256 filter, yang memperdalam ekstraksi fitur pada gambar yang telah diproses sebelumnya.
* block3\_conv2 (Conv2D):
  + Output Shape: (None, 12, 12, 256)
  + Parameter: 590,080
  + Lapisan konvolusi kedua di blok ketiga, meningkatkan jumlah fitur yang terdeteksi.
* block3\_conv3 (Conv2D):
  + Output Shape: (None, 12, 12, 256)
  + Parameter: 590,080
  + Lapisan konvolusi ketiga di blok ketiga, melanjutkan ekstraksi fitur yang semakin mendalam.

5. Total Parameter

* Total Parameters: 14,714,688 (56.13 MB)
* Trainable Parameters: 14,714,688 (56.13 MB)
* Non-trainable Parameters: 0 (0.00 Byte)

Secara keseluruhan, model ini memiliki 14,714,688 parameter yang dapat dilatih, yang mencakup lapisan-lapisan konvolusi dan pooling. Tidak ada parameter non-trainable dalam model ini, karena seluruh model dilatih sepenuhnya.

**2.3.2 Arsitektur sequential**

****

**Gambar .4** Arsitektur sequential

Model "sequential\_2" yang digunakan dalam penelitian ini merupakan model yang terdiri dari beberapa lapisan yang diatur secara berurutan untuk memproses input dan menghasilkan output klasifikasi. Model ini mengintegrasikan penggunaan VGG16 sebagai fitur ekstraktor dengan beberapa lapisan tambahan untuk klasifikasi ekspresi wajah. Berikut adalah rincian lapisan-lapisan dalam model tersebut.

1. Lapisan Rescaling (rescaling\_1)

* Output Shape: (None, 48, 48, 3)
* Parameter: 0
* Lapisan ini digunakan untuk melakukan penskalaan ulang pada gambar input, mengubah rentang nilai piksel gambar menjadi antara 0 dan 1. Ini penting untuk memastikan konsistensi dalam proses pelatihan, karena model lebih mudah belajar dengan data yang terstandarisasi.

2. Lapisan Sequential (sequential)

* Output Shape: (None, None, None, None)
* Parameter: 0
* Lapisan Sequential ini menyusun lapisan-lapisan lainnya secara berurutan. Meskipun tidak ada parameter yang dapat dilatih di sini, lapisan Sequential berfungsi untuk mengkoordinasikan urutan lapisan-lapisan yang ada, termasuk integrasi dengan VGG16.

3. Model VGG16 (vgg16)

* Output Shape: (None, 1, 1, 512)
* Parameter: 14,714,688
* Lapisan ini adalah model VGG16 yang telah dilatih sebelumnya sebagai fitur ekstraktor. Model VGG16 digunakan untuk mengekstraksi fitur dari gambar input dan menghasilkan representasi dengan dimensi (None, 1, 1, 512). Hasil ekstraksi fitur ini adalah representasi 512 dimensi dari gambar yang digunakan untuk klasifikasi lebih lanjut.

4. Lapisan Flatten (flatten\_1)

* Output Shape: (None, 512)
* Parameter: 0
* Lapisan Flatten digunakan untuk meratakan output dari VGG16, yang berupa tensor 4D (None, 1, 1, 512) menjadi vektor 1D dengan panjang 512. Proses ini penting untuk meneruskan hasil ekstraksi fitur ke lapisan fully connected berikutnya.

5. Lapisan Dense (dense\_2)

* Output Shape: (None, 2048)
* Parameter: 1,050,624
* Lapisan Dense pertama adalah lapisan fully connected yang terdiri dari 2048 unit. Lapisan ini bertanggung jawab untuk belajar representasi yang lebih kompleks dari fitur yang diekstraksi oleh VGG16. Fungsi aktivasi yang digunakan biasanya adalah ReLU, yang memperkenalkan non-linearitas pada model.

6. Lapisan Dense (dense\_3)

* Output Shape: (None, 8)
* Parameter: 16,392
* Lapisan Dense kedua adalah lapisan output yang terdiri dari 8 unit, yang sesuai dengan jumlah kelas yang akan diprediksi (misalnya, ekspresi wajah: bahagia, sedih, marah, takut, jijik, terkejut, dan netral). Fungsi aktivasi yang digunakan di sini adalah softmax, yang menghasilkan probabilitas untuk masing-masing kelas.

7. Total Parameter

* Total Parameters: 15,781,704 (60.20 MB)
* Trainable Parameters: 1,067,016 (4.07 MB)
* Non-trainable Parameters: 14,714,688 (56.13 MB)

Total parameter dalam model ini adalah 15,781,704, dengan 1,067,016 parameter yang dapat dilatih yang digunakan untuk melatih lapisan Dense setelah VGG16. Parameter yang tidak dapat dilatih sebanyak 14,714,688, berasal dari model VGG16 yang digunakan sebagai fitur ekstraktor dan telah dilatih sebelumnya.

* 1. **Pelatihan Model**

Pelatihan model dilakukan menggunakan VGG16 dengan dataset pelatihan dan validasi yang disediakan melalui train\_generator dan validation\_generator. Proses pelatihan ini bertujuan untuk mengoptimalkan parameter model dengan mengurangi nilai loss dan meningkatkan accuracy pada dataset pelatihan dan validasi. Model dilatih selama 20 epoch dengan konfigurasi tertentu, sebagai berikut:

1. Pengaturan Pelatihan

Model dilatih menggunakan metode fit yang diterapkan pada objek model VGG16\_model. Proses pelatihan dilakukan dengan parameter sebagai berikut:

* train\_generator: Generator yang menyediakan batch data pelatihan.
* validation\_generator: Generator yang menyediakan batch data validasi untuk evaluasi kinerja model setiap epoch.
* batch\_size: Ukuran batch yang digunakan untuk pelatihan, menentukan berapa banyak gambar yang diproses dalam satu langkah pelatihan.
* epochs: Jumlah iterasi yang dilakukan model terhadap seluruh dataset pelatihan. Dalam eksperimen ini, model dilatih selama 20 epoch.
* callbacks: Penggunaan callback tqdm\_callback memungkinkan pemantauan progres pelatihan melalui tampilan progress bar yang lebih informatif.

2. Hasil Pelatihan per Epoch

Setiap epoch terdiri dari dua bagian utama: pelatihan pada dataset pelatihan dan evaluasi pada dataset validasi. Berikut adalah contoh hasil yang diperoleh setelah menjalani beberapa epoch pelatihan:

Epoch 1/20:

* Loss: 1.9261
* Accuracy: 0.2582
* Val\_loss (Loss Validasi): 1.8750
* Val\_accuracy (Akurasi Validasi): 0.2857

Pada epoch pertama, model menunjukkan loss sebesar 1.9261 pada dataset pelatihan dan accuracy sebesar 25.82%. Pada dataset validasi, model memiliki loss 1.8750 dan accuracy 28.57%. Nilai loss yang lebih tinggi menunjukkan bahwa model masih dalam tahap pembelajaran dan perlu diperbaiki untuk meningkatkan akurasi.

Epoch 2/20:

* Loss: 1.8908
* Accuracy: 0.2714

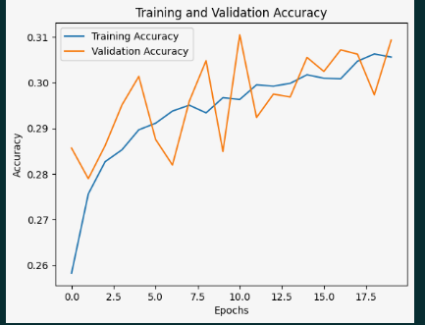
Pada epoch kedua, loss mengalami penurunan menjadi 1.8908, dan accuracy meningkat sedikit menjadi 27.14%. Ini menunjukkan bahwa model mulai belajar dari data pelatihan dengan lebih baik, meskipun akurasi pada dataset pelatihan dan validasi masih perlu ditingkatkan.

Proses Pelatihan Selama 20 Epoch

* Dalam setiap epoch berikutnya, model terus melatih dirinya dengan memperbarui bobot berdasarkan gradien yang dihitung selama backpropagation.
* Loss diharapkan berkurang seiring berjalannya pelatihan, yang menandakan bahwa model semakin baik dalam mengklasifikasikan data.
* Akurasi pada dataset pelatihan dan validasi akan dievaluasi setiap epoch untuk memantau kemajuan pelatihan dan menghindari overfitting.

1. **RESULTS AND DISCUSSION**

**3.1.** **Hasil Klasifikasi**

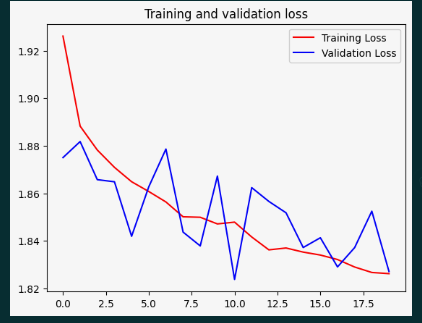
****

**Gambar .5** Akurasi training dan validation

Hasil dari penelitian ini menunjukkan bahwa performa model masih belum memuaskan. Setelah pelatihan selama 20 epoch, akurasi yang diperoleh sangat rendah, dengan **training accuracy** sebesar 30.66% dan **validation accuracy** 30.93%. Ini menunjukkan bahwa model belum berhasil mengenali ekspresi wajah dengan baik.

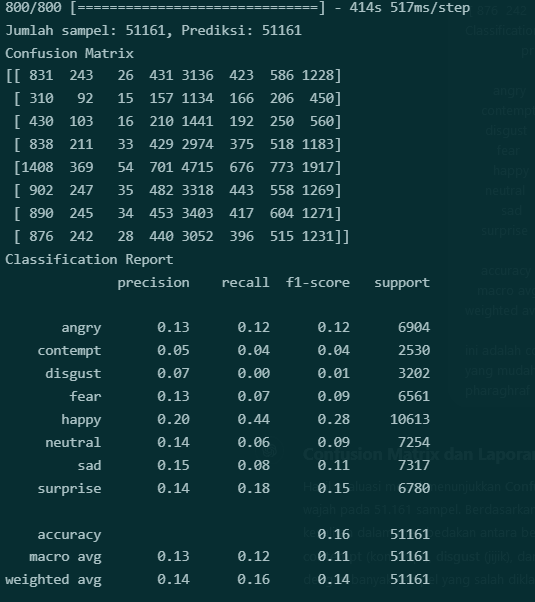
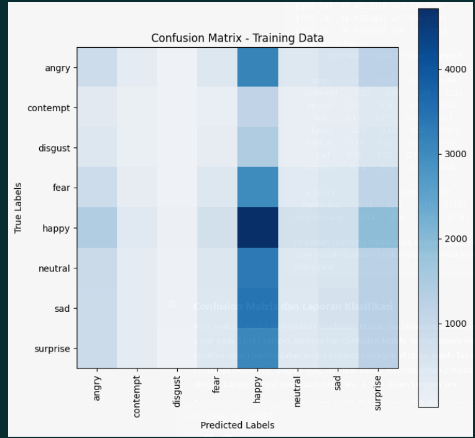
Salah satu penyebab utama rendahnya akurasi adalah keterbatasan perangkat keras yang digunakan dalam penelitian ini. Pelatihan dilakukan menggunakan laptop dengan spesifikasi entry-level, yang mengakibatkan waktu pelatihan yang cukup lama, sekitar 3 jam untuk 20 epoch. Model VGG16, yang memiliki jutaan parameter, membutuhkan waktu komputasi lebih banyak untuk mencapai konvergensi, terutama dengan dataset yang besar. Selain itu, jumlah epoch yang terbatas juga menjadi faktor penghambat. VGG16 memerlukan lebih banyak epoch untuk mempelajari fitur-fitur kompleks dari dataset, dan dengan hanya 20 epoch, model belum cukup waktu untuk belajar secara optimal.

Keterbatasan ini menunjukkan bahwa perangkat keras dengan performa lebih tinggi, seperti GPU, sangat dibutuhkan untuk menangani dataset besar dan model yang kompleks. Selain itu, pendekatan seperti transfer learning, early stopping, atau pengurangan dimensi dataset bisa dipertimbangkan untuk mengurangi kebutuhan komputasi tanpa mengorbankan performa model secara signifikan.

****

**Gambar .6** training dan validation loss

Hasil dari pelatihan menunjukkan bahwa training loss sebesar 1.8262 dan validation loss sebesar 1.8272. Nilai loss yang hampir serupa antara dataset pelatihan dan validasi menunjukkan bahwa model tidak mengalami overfitting, yang berarti model tidak hanya menghafal data pelatihan, tetapi dapat menggeneralisasi dengan cukup baik pada data yang tidak terlihat (validasi). Meskipun demikian, nilai loss yang relatif tinggi pada kedua dataset menunjukkan bahwa model masih belum optimal dalam memahami pola-pola pada ekspresi wajah. Hal ini mengindikasikan bahwa model memerlukan lebih banyak pelatihan atau optimasi tambahan, seperti penyesuaian arsitektur atau penggunaan teknik regulasi, untuk mencapai performa yang lebih baik.

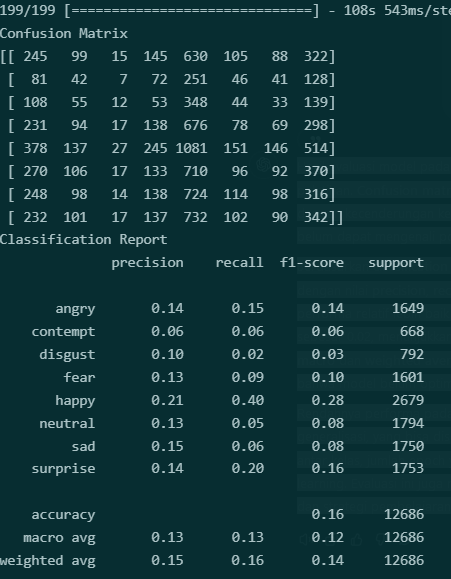
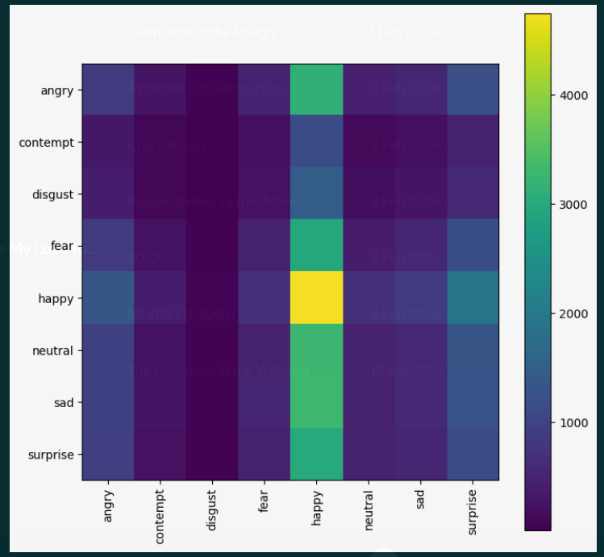
****

**Gambar .7** Classification Report danConfusion Matrix

Hasil evaluasi model pada dataset pelatihan menunjukkan confusion matrix yang mengindikasikan distribusi prediksi model terhadap label yang sebenarnya. Dari matriks tersebut, terlihat bahwa model cenderung mengalami kesulitan dalam membedakan kategori emosi tertentu, terutama pada kelas dengan jumlah sampel besar seperti "happy" dan "angry." Hal ini terlihat dari banyaknya prediksi yang tersebar ke berbagai kelas lain, menunjukkan bahwa model belum mampu mempelajari pola dengan baik.

Berdasarkan classification report, akurasi keseluruhan model hanya mencapai 16%, dengan nilai precision, recall, dan f1-score yang relatif rendah di semua kelas. Kategori "happy" memiliki performa terbaik dengan recall 0.44, namun kelas lain seperti "disgust" dan "contempt" memiliki recall yang sangat rendah, masing-masing 0.00 dan 0.04, mengindikasikan model hampir tidak mengenali kelas tersebut.

Rendahnya performa ini dapat disebabkan oleh beberapa faktor, seperti ketidakseimbangan jumlah sampel antar kelas, kompleksitas dataset, serta keterbatasan pelatihan model yang hanya dilakukan selama 20 epoch. Selain itu, arsitektur VGG16 dengan parameter yang besar mungkin belum cukup teroptimasi untuk menangani data ekspresi wajah yang kompleks tanpa fine-tuning atau penggunaan teknik transfer learning.

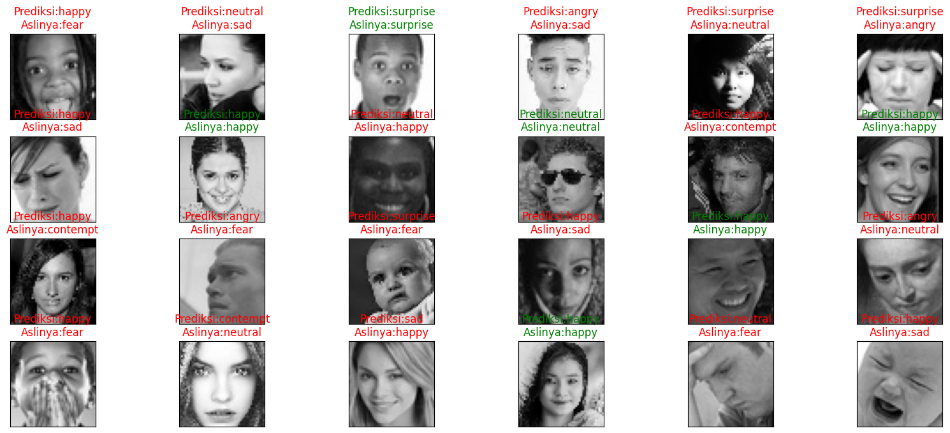
****

**Gambar .7** Classification Report danConfusion Matrix validaton

Hasil evaluasi model pada dataset validasi menunjukkan performa yang serupa dengan dataset pelatihan. Confusion matrix menggambarkan bahwa prediksi model tersebar luas di berbagai kelas, dengan kecenderungan kesalahan prediksi yang signifikan. Hal ini mengindikasikan bahwa model belum dapat mengenali pola ekspresi wajah secara konsisten pada data validasi.

Berdasarkan classification report, akurasi keseluruhan model pada dataset validasi adalah 16%, dengan nilai precision, recall, dan f1-score yang rendah di semua kelas. Kelas "happy" memiliki performa relatif lebih baik dengan recall 0.40, namun kelas lain seperti "disgust" hanya memiliki recall sebesar 0.02, menunjukkan bahwa model hampir tidak mengenali kategori tersebut. Rata-rata nilai macro dan weighted average menunjukkan skor yang rendah (sekitar 0.12 hingga 0.16), menandakan bahwa model belum optimal.

Rendahnya performa pada dataset validasi ini menegaskan bahwa model mengalami kesulitan dalam generalisasi, yang bisa disebabkan oleh beberapa faktor, seperti ketidakseimbangan jumlah sampel antar kelas, jumlah epoch yang terlalu sedikit, atau kurangnya teknik optimasi seperti transfer learning. Evaluasi ini juga menggarisbawahi pentingnya penggunaan perangkat keras yang lebih baik dan strategi pembelajaran tambahan untuk meningkatkan hasil klasifikasi.

****

**Gambar .8** Outpu akhir model

1. **CONCLUSION**

Penelitian ini bertujuan untuk mengklasifikasikan ekspresi wajah menggunakan dataset FER13 dan AffectNet dengan arsitektur VGG16. Hasil menunjukkan bahwa performa model masih rendah, dengan akurasi pelatihan dan validasi hanya mencapai sekitar 16%. Nilai precision, recall, dan f1-score juga rendah di semua kelas emosi, menunjukkan bahwa model belum mampu mengenali pola ekspresi wajah dengan baik. Rendahnya performa ini disebabkan oleh keterbatasan perangkat keras tanpa GPU, kompleksitas dataset yang tinggi, dan kebutuhan optimasi tambahan pada model. Untuk meningkatkan hasil, diperlukan strategi seperti transfer learning, augmentasi data, penggunaan perangkat keras yang lebih baik, dan penyesuaian hyperparameter. Dengan langkah-langkah ini, model diharapkan dapat menangani kompleksitas dataset dan mencapai performa yang lebih baik.

**ACKNOWLEDGEMENTS**

Pekerjaan ini didukung oleh Program Sarjana Ilmu Komputer dan Sistem Informasi Universitas Kebangsaaan Republik Indonesia

**REFERENCES**

[1] Augyeris Lioga Seandrio1, Awang Hendrianto Pratomo2, Mangaras Yanu Florestiyanto3 Implementasi Convolutional Neural Network (CNN) Pada Pengenalan Ekspresi

Wajah Vol. 18, No. 2, Juni 2021, pp. 211-221

[2] Josephine Roosandriantini1, Ryan Putranda2, Yulia Wahyuningsih3\*, Yohana ChristelaOktaviani4, Elisabeth Yolanda Christin5, Face Expression Recognizer Dengan Convolutional Neural Network Untuk Membantu Penderita Autisme Mengenali Ekspresi Wajah Seseorang Vol. 11 No. 3, pISSN: 2303-0577 eISSN: 2830-7062

[3] Reza Rizki Reynaldo1, Irfan Maliki2\* Pengenalan Ekspresi Wajah Dengan Metode Viola Jones dan Convolutional Neural Network Komputika: Jurnal Sistem KomputerVolume 9, Nomor 1, April 2020, hlm. 1 - 9 Terakreditasi Peringkat 3, SK No. 28/E/KPT/2019 DOI: 10.34010/komputika.v10i1.4119

[4] Yusril Ihza \*1, Danang Lelono2 Face Expression Classification in Children Using CNN IJCCS (Indonesian Journal of Computing and Cybernetics Systems)Vol.16, No.2, April 2022, pp. 159~168 ISSN (print): 1978-1520, ISSN (online): 2460-7258 DOI: 10.22146/ijccs.72493

[5] Arief Saputro1, Syahri Mu’min2, Moch. Lutfi3, Helmanita Putri4 Deep Transfer Learning Dengan Model Arsitektur Vgg16 Untuk Klasifikasi Jenis Varietas Tanaman Lengkeng Berdasarkan Citra Daun Vol. 6 No. 2, September 2022