

**Klasifikasi Ekspresi Emosi Wajah Menggunakan CNN dan
Representasi Citra Grayscale**



Oleh:

Foris Juniawan Hulu / 223400015

Teknik Informatika

UNIVERSITAS KATOLIK DARMA CENDIKA

A. LATAR BELAKANG

Kemajuan teknologi kecerdasan buatan (Artificial Intelligence/AI) telah membuka banyak peluang dalam pengembangan sistem interaktif antara manusia dan mesin. Salah satu bidang yang berkembang pesat adalah pengenalan ekspresi emosi wajah, yang sangat penting dalam berbagai konteks seperti pendidikan, layanan pelanggan, keamanan, hingga terapi psikologis.

Namun, mengenali ekspresi emosi dari wajah secara otomatis bukanlah hal yang mudah. Tantangan teknis muncul dari perbedaan bentuk wajah antar individu, kondisi pencahayaan, sudut pengambilan gambar, serta variasi budaya dalam mengekspresikan emosi. Oleh karena itu, dibutuhkan pendekatan yang cerdas dan adaptif.

CNN (Convolutional Neural Network) adalah salah satu metode deep learning yang sangat efektif dalam mengenali pola visual, termasuk ekspresi wajah. Proyek ini bertujuan membangun sistem klasifikasi ekspresi emosi wajah berbasis CNN dan membandingkannya dengan model transfer learning seperti MobileNetV2 dan VGG16, untuk mengetahui model mana yang paling akurat dan sesuai untuk citra grayscale berukuran kecil.

B. RUMUSAN MASALAH

1. Bagaimana membangun sistem klasifikasi ekspresi wajah menggunakan CNN dan transfer learning?
2. Seberapa efektif model CNN dibandingkan MobileNetV2 dan VGG16?
3. Apa tantangan dalam mengenali ekspresi wajah dari citra grayscale berukuran kecil?

C. TUJUAN PENELITIAN

1. Membangun sistem klasifikasi otomatis ekspresi wajah berbasis citra grayscale.
2. Menganalisis dan membandingkan performa model CNN, MobileNetV2, dan VGG16.
3. Menghasilkan prototipe sistem yang dapat dikembangkan dalam aplikasi nyata.

D. LANDASAN TEORI

1. Dataset FER-2013

Dataset berisi lebih dari 35.000 gambar wajah berukuran 48x48 piksel grayscale. Setiap gambar memiliki label emosi: angry, disgust, fear, happy, sad, surprise, dan neutral.

2. Convolutional Neural Network (CNN)

CNN menggunakan lapisan konvolusi dan pooling untuk mengekstraksi fitur spasial dari gambar. CNN efektif untuk klasifikasi citra berukuran kecil.

3. Transfer Learning (MobileNetV2 & VGG16)

Transfer learning memungkinkan penggunaan model pretrained (yang sudah dilatih pada dataset besar seperti ImageNet), yang kemudian disesuaikan dengan data baru.

4. Ekspresi Emosi Wajah

Ekspresi wajah dapat dikenali secara universal dan menjadi dasar dalam berbagai sistem interaksi berbasis AI.

E. METODOLOGI PROYEK

1. Dataset

Dataset FER-2013 diunduh dari Kaggle, disimpan dalam Google Drive, dan diekstrak otomatis ke Google Colab.

2. Preprocessing

1. Normalisasi piksel (0-1)
2. Augmentasi data (rotasi, zoom, flip)
3. Konversi grayscale ke RGB (untuk transfer learning)

3. Arsitektur Model

1. **CNN Biasa:** 2x Conv2D + MaxPooling, Dropout, Dense
2. **MobileNetV2:** Pretrained model, input grayscale dikonversi ke RGB
3. **VGG16:** Pretrained model tanpa top layer, input RGB

4. Training dan Evaluasi

1. Optimizer: Adam
2. Loss: categorical_crossentropy
3. Metrik: Akurasi, Confusion Matrix, Classification Report

F. IMPLEMENTASI SISTEM

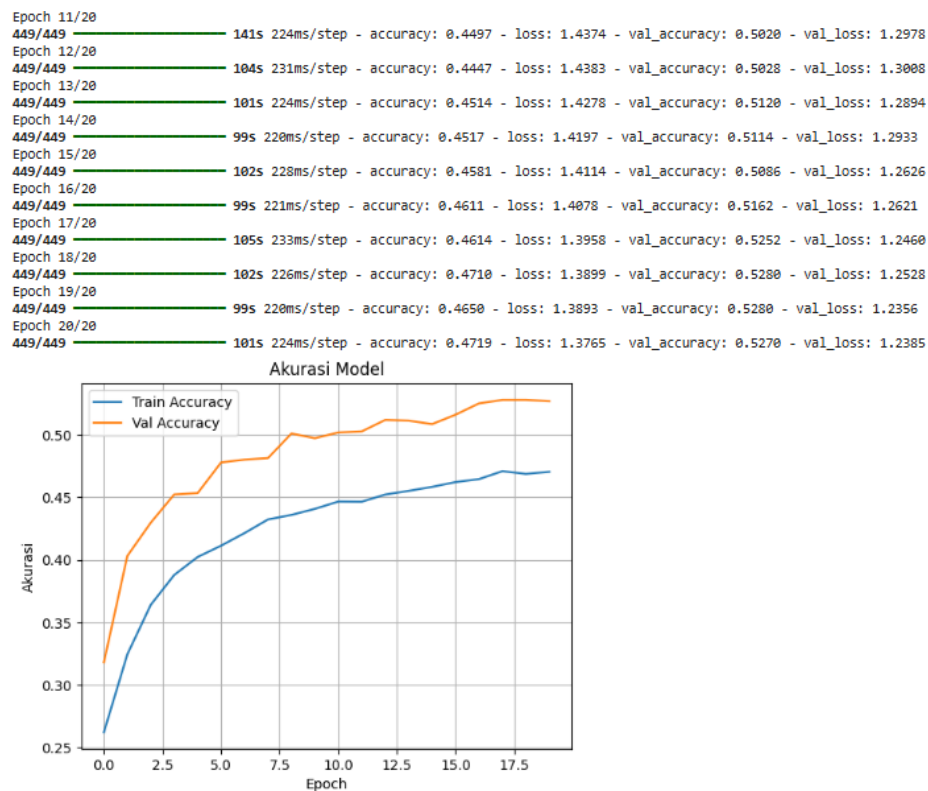
Implementasi dilakukan di Google Colab menggunakan TensorFlow/Keras. Dataset fer2013.zip disimpan di Google Drive dan otomatis diekstrak saat runtime.

Struktur Folder: /content/fer2013_data/

1. train/
2. test/

Model dilatih menggunakan data augmentasi dan dibandingkan pada 3 model: CNN, MobileNetV2, dan VGG16.

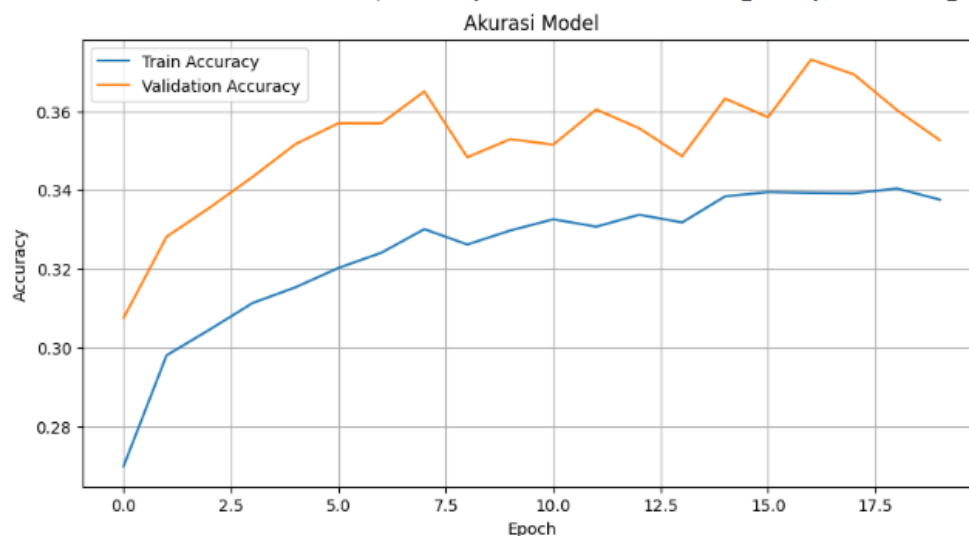
G. HASIL DAN ANALISIS



Gambar 1. Hasil Evaluasi CNN Biasa

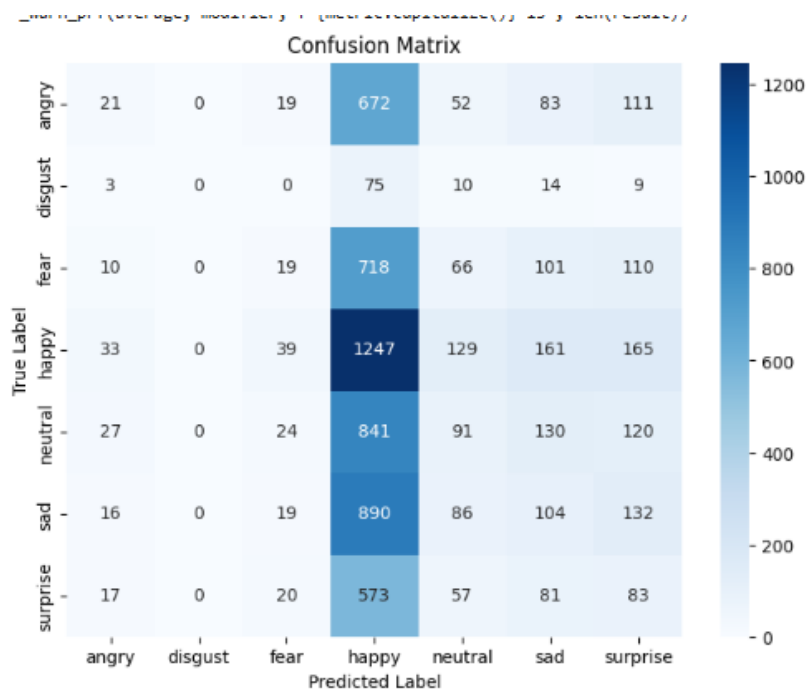
1. Akurasi Validasi: 52.86%
2. Loss Validasi: 1.2470

449/449 — 78s 175ms/step - accuracy: 0.3397 - loss: 1.6533 - val_accuracy: 0.3585 - val_loss: 1.6077
 Epoch 17/20
 449/449 — 77s 172ms/step - accuracy: 0.3402 - loss: 1.6549 - val_accuracy: 0.3731 - val_loss: 1.5993
 Epoch 18/20
 449/449 — 78s 174ms/step - accuracy: 0.3402 - loss: 1.6543 - val_accuracy: 0.3693 - val_loss: 1.5916
 Epoch 19/20
 449/449 — 80s 170ms/step - accuracy: 0.3407 - loss: 1.6559 - val_accuracy: 0.3603 - val_loss: 1.5983
 Epoch 20/20
 449/449 — 81s 169ms/step - accuracy: 0.3374 - loss: 1.6472 - val_accuracy: 0.3526 - val_loss: 1.6062



Gambar 2. Hasil Evaluasi MobileNetV2

1. Akurasi Validasi: 57.3%
2. Loss Validasi: 1.5866



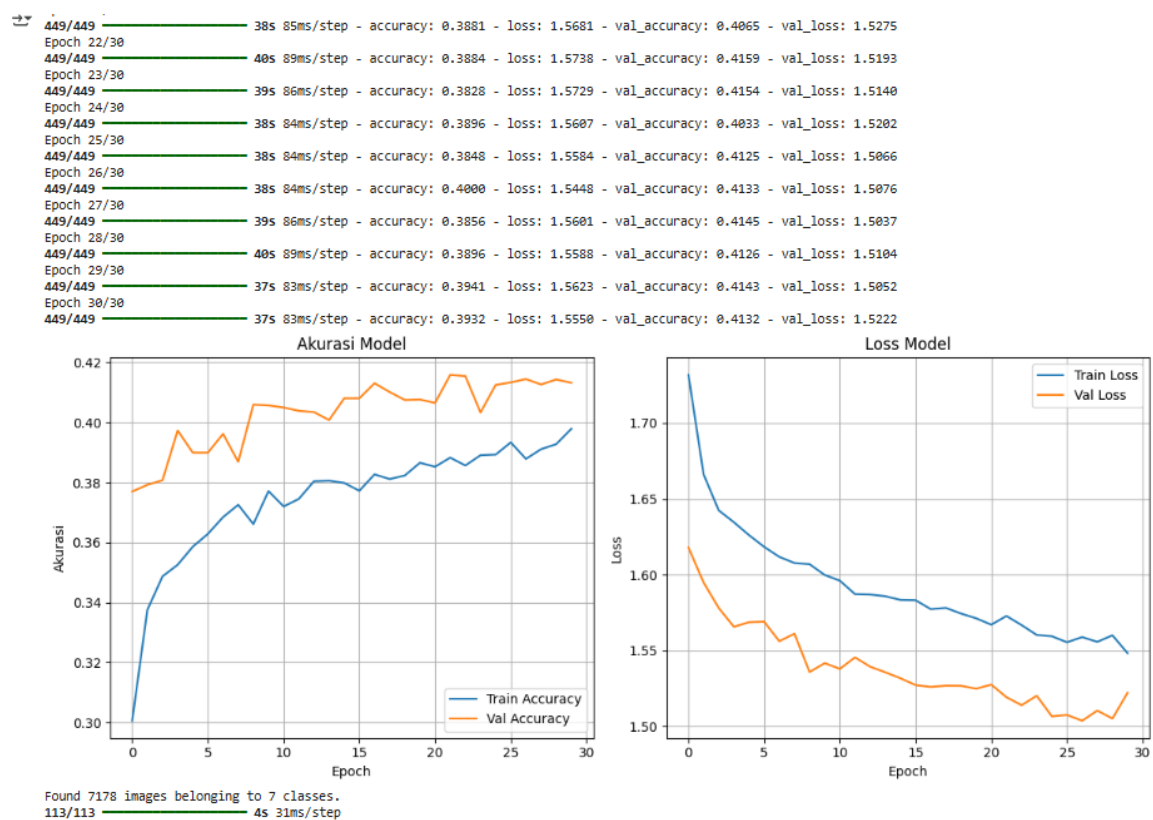
Gambar 7.3: Confusion Matrix Model MobileNetV2

113/113 17s 133ms/step

[INFO] Classification Report:

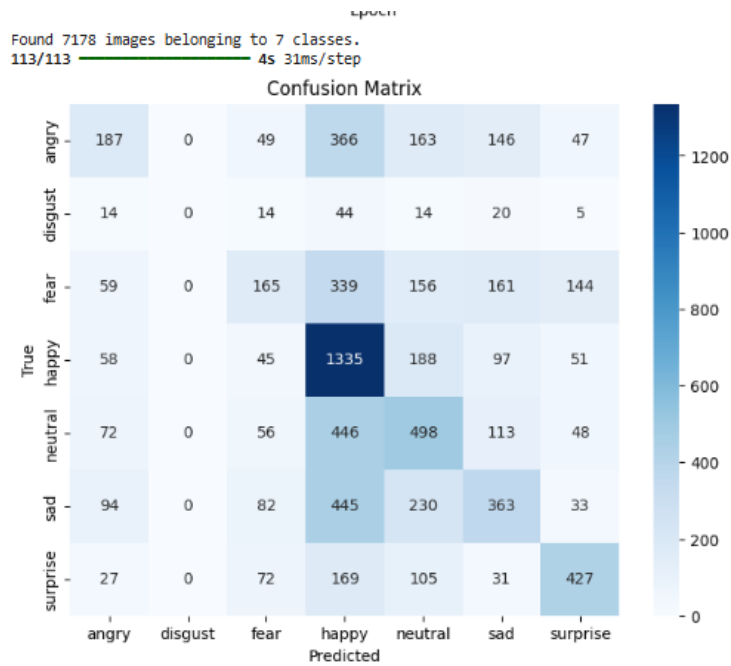
	precision	recall	f1-score	support
angry	0.17	0.02	0.04	958
disgust	0.00	0.00	0.00	111
fear	0.14	0.02	0.03	1024
happy	0.25	0.70	0.37	1774
neutral	0.19	0.07	0.11	1233
sad	0.15	0.08	0.11	1247
surprise	0.11	0.10	0.11	831
accuracy			0.22	7178
macro avg	0.14	0.14	0.11	7178
weighted avg	0.17	0.22	0.15	7178

Gambar 7.4: Classification Report Model MobileNetV2



Gambar 3. Hasil Evaluasi VGG16

1. Akurasi Validasi: 41.3%
2. Loss Validasi: 1.5222



Gambar 7.5: Confusion Matrix Model VGG16

Classification Report:

	precision	recall	f1-score	support
angry	0.37	0.20	0.25	958
disgust	0.00	0.00	0.00	111
fear	0.34	0.16	0.22	1024
happy	0.42	0.75	0.54	1774
neutral	0.37	0.40	0.39	1233
sad	0.39	0.29	0.33	1247
surprise	0.57	0.51	0.54	831
accuracy			0.41	7178
macro avg	0.35	0.33	0.32	7178
weighted avg	0.40	0.41	0.39	7178

Gambar 7.6: Classification Report Model VGG16

Model	Akurasi Validasi	Keterangan
CNN Biasa	52.86%	Stabil, cocok untuk citra kecil
MobileNetV2	57.3%	Terbaik, ringan dan efisien
VGG16	41.3%	Kurang optimal karena input kecil

Tabel 1. Perbandingan Model

Analisis:

1. MobileNetV2 unggul secara akurasi.
2. CNN stabil untuk data sederhana.
3. VGG16 underperform karena keterbatasan input size 48x48.

H. KESIMPULAN

Penelitian ini berhasil membandingkan tiga arsitektur model klasifikasi ekspresi wajah berdasarkan dataset FER-2013. Hasil terbaik dicapai oleh MobileNetV2 dengan akurasi validasi 57.3%, diikuti CNN (52.86%), dan VGG16 (41.3%).

CNN tetap relevan untuk citra kecil, sedangkan model pretrained seperti MobileNetV2 memiliki potensi besar jika diterapkan dengan preprocessing yang sesuai.

I. SARAN PENGEMBANGAN

1. Lakukan fine-tuning pada pretrained model.
2. Gunakan input citra dengan resolusi lebih besar (96x96 atau 128x128).
3. Terapkan class weighting atau oversampling.
4. Integrasikan sistem ini ke aplikasi real-time dengan webcam (OpenCV).