SEMINAR NASIONAL 2026 Universitas Ma Chung

Prosiding Seminar Nasional Universitas Ma Chung 2026
(Sistem Informasi dan Informatika)
pp. 216-229

E-ISSN: 2808-7771

Perbandingan Akurasi Model VGG16, ResNet50, dan MobileNet dalam Pengenalan Ekspresi Wajah Sedih dan Senang pada Dataset FER2013

Alfond Manuel Joseph[1], Davin Valerian[2], dan Steven Surya Putra[3]

Teknik Informatika, Universitas Ma Chung, Villa Puncak Tidar N-01, Karangwidoro, Kecamatan Dau, Kabupaten Malang, Jawa Timur 65151.

Correspondence: Alfond Manuel Joseph (312310005@student.machung.ac.id)

Received: 27 05 2026 – Revised: 29 07 2026 - Accepted: 04 08 2026 - Published: 05 08 2026

Abstrak. Pengenalan ekspresi wajah adalah bidang krusial dalam interaksi manusia-komputer dan berbagai aplikasi visi lainnya, meskipun sering menghadapi tantangan signifikan seperti variasi ekspresi individu, kondisi pencahayaan yang beragam, dan oklusi. Penelitian ini bertujuan untuk mengevaluasi, serta membandingkan secara komprehensif performa akurasi model **VGG16, ResNet50, dan MobileNet** dalam mengenali dua ekspresi wajah fundamental: **sedih (sad)** dan **senang (happy)**. Evaluasi ini dilakukan menggunakan subset **dataset FER2013** yang telah difilter secara spesifik untuk kedua kelas emosi tersebut, memastikan fokus yang tajam pada tugas klasifikasi biner. Metode yang digunakan melibatkan adaptasi ketiga arsitektur **Convolutional Neural Network (CNN)** *pre-trained* tersebut. Kami memodifikasi lapisan *output* mereka agar sesuai dengan tugas klasifikasi biner yang spesifik, lalu melatih model-model ini secara iteratif menggunakan data citra "sad" dan "happy" dari FER2013. Hasil eksperimen menunjukkan bahwa **VGG16** berhasil mencapai akurasi pelatihan **85.94%** dan akurasi pengujian tertinggi sebesar **85.14%**, mengungguli model lainnya. **MobileNet** menunjukkan akurasi pelatihan **82.81%** dan akurasi pengujian **78.25%**, sementara **ResNet50** memperoleh akurasi pelatihan **65.62%** dan akurasi pengujian **64.63%**. Perbandingan ini secara jelas mengindikasikan bahwa VGG16, dengan arsitekturnya yang lebih dalam, memberikan kinerja superior dalam hal akurasi untuk pengenalan ekspresi "sad" dan "happy" pada dataset FER2013 yang difilter.

Kata kunci: Pengenalan Ekspresi Wajah, FER2013, VGG16, ResNet50, MobileNet.

PENDAHULUAN

Pengenalan ekspresi wajah merupakan salah satu bidang yang sangat penting dalam interaksi manusia-komputer. Dalam kehidupan sehari-hari, kemampuan untuk mengenali ekspresi wajah menjadi kunci dalam berbagai aplikasi, seperti pengawasan keamanan, analisis emosi dalam layanan pelanggan, serta aplikasi kesehatan mental. Namun, pengenalan ekspresi wajah menghadapi sejumlah tantangan, termasuk variasi ekspresi yang dimiliki tiap individu, perbedaan kondisi pencahayaan, dan potensi gangguan visual seperti oklusi wajah yang menghalangi fitur wajah. Untuk itu, penelitian ini dilakukan untuk mengevaluasi akurasi beberapa model deep learning dalam mengenali ekspresi wajah, khususnya ekspresi wajah sedih dan senang.

Penelitian ini menggunakan tiga model Convolutional Neural Network (CNN) yang telah banyak digunakan dalam bidang visi komputer, yaitu VGG16, ResNet50, dan MobileNet. Masing-masing model ini memiliki karakteristik dan keunggulan tersendiri yang memungkinkan perbandingan performa yang menarik dalam konteks pengenalan ekspresi wajah. Sebelumnya, berbagai penelitian telah mengembangkan model CNN untuk pengenalan ekspresi wajah, namun terdapat kekurangan dalam hal akurasi saat menghadapi variasi ekspresi atau kondisi pencahayaan yang berubah. Oleh karena itu, tujuan utama penelitian ini adalah untuk membandingkan ketiga model tersebut dalam pengenalan ekspresi wajah "sedih" dan "senang" pada dataset FER2013 yang telah difilter.

MASALAH

Masyarakat modern saat ini semakin bergantung pada teknologi untuk berbagai kebutuhan, termasuk interaksi sosial dan komunikasi non-verbal. Pengenalan ekspresi wajah menjadi isu penting, terutama dalam aplikasi yang berkaitan dengan pengawasan, layanan pelanggan, dan interaksi manusia dengan komputer. Namun, masih terdapat berbagai tantangan dalam meningkatkan akurasi pengenalan ekspresi wajah, terutama pada ekspresi yang lebih sulit seperti "sedih" dan "senang". Berbagai kondisi seperti pencahayaan yang tidak stabil, posisi wajah yang bervariasi, dan oklusi dapat memengaruhi hasil deteksi wajah. Oleh karena itu, ada kebutuhan mendesak untuk mengembangkan model yang lebih akurat dan dapat mengatasi berbagai tantangan tersebut dalam aplikasi dunia nyata.

Dalam konteks penelitian ini, masalah yang dihadapi adalah memilih model deep learning yang optimal untuk pengenalan ekspresi wajah pada dataset yang relatif kompleks.

VGG16, ResNet50, dan MobileNet masing-masing memiliki kelebihan dan kekurangan dalam hal kemampuan generalisasi dan akurasi, terutama dalam pengenalan ekspresi wajah dalam kondisi yang tidak ideal. Penelitian ini bertujuan untuk menilai dan membandingkan ketiga model ini secara objektif untuk menemukan model terbaik yang dapat diterapkan dalam pengenalan ekspresi wajah "sedih" dan "senang" secara lebih efektif.

METODE PENELETIAN

3.1. *Desain Penelitian*

Penelitian ini mengadopsi pendekatan kuantitatif dengan desain eksperimental komparatif. Tujuannya adalah mengevaluasi dan membandingkan kinerja akurasi tiga arsitektur Convolutional Neural Network (CNN) pra-terlatih: VGG16, MobileNetV2, dan ResNet50. Perbandingan ini dilakukan dalam konteks tugas klasifikasi biner untuk mengenali dua ekspresi wajah fundamental, yaitu "sedih" (sad) dan "senang" (happy).

3.2. *Dataset*

- **Nama Dataset:** FER2013 (Facial Expression Recognition 2013). Dataset ini merupakan *benchmark* standar yang banyak digunakan dalam penelitian pengenalan ekspresi wajah, dikenal karena variasi citra "dunia nyata" yang menantang [Goodfellow et al., 2013].
- **Pemilihan Data:** Dari dataset FER2013 asli, hanya citra yang memiliki label ekspresi **"sad"** dan **"happy"** yang digunakan. Pemilihan ini bertujuan untuk memfokuskan analisis pada kemampuan model dalam membedakan polaritas emosi yang jelas dan untuk menyederhanakan masalah klasifikasi menjadi biner.
- **Jumlah Citra:** Data yang difilter dibagi secara acak dengan stratifikasi (untuk menjaga proporsi kelas) menjadi set pelatihan (training), validasi (validation), dan pengujian (testing).
- **Karakteristik Citra:** Seluruh citra di-resize menjadi resolusi **48x48 piksel** dan diproses sebagai gambar **grayscale** (1 *channel*). Resolusi ini dipilih karena merupakan standar dalam dataset FER2013.

3.3. *Pra-pemrosesan Data*

1. **Ekstraksi dan Pemfilteran Data:** Dataset FER2013, yang awalnya dikemas dalam format ZIP, diekstrak ke direktori lokal Google Colab. Setelah ekstraksi, hanya gambar-gambar yang termasuk dalam sub-direktori "train/happy", "train/sad", "test/happy", dan "test/sad" yang dipilih untuk diproses lebih lanjut.
2. **Konversi Grayscale ke RGB:** Model-model pra-terlatih seperti VGG16, ResNet50, dan MobileNetV2 dilatih pada dataset ImageNet yang terdiri dari gambar RGB (3 *channel*). Oleh karena itu, citra *grayscale* (1 *channel*) dari FER2013 dikonversi menjadi 3 *channel* dengan menduplikasi *channel* yang ada.
3. ***Normalisasi Piksel:***

- **VGG16 & ResNet50:** Nilai piksel dinormalisasi dari rentang [0, 255] menjadi rentang [0, 1] dengan membagi setiap nilai piksel dengan 255.
- **MobileNetV2:** Normalisasi dilakukan menggunakan fungsi *pre-processing* khusus dari MobileNetV2 yang menormalisasi piksel ke rentang [-1, 1].

4. **Augmentasi Data:** Untuk meningkatkan *robustness* model dan mencegah *overfitting*, ImageDataGenerator diterapkan pada set pelatihan. Teknik augmentasi data secara artifisial memperluas dataset pelatihan dengan membuat versi gambar yang dimodifikasi, sehingga membantu model belajar fitur yang lebih general dan mengurangi ketergantungan pada variasi spesifik dalam data pelatihan asli [Shorten & Khoshgoftaar, 2019]. Parameter augmentasi yang digunakan meliputi: rotation_range=20, width_shift_range=0.15, height_shift_range=0.15, shear_range=0.15, zoom_range=0.15, horizontal_flip=True, brightness_range=[0.7, 1.3], channel_shift_range=10.0, dan fill_mode='nearest'.
5. **Pembagian Data:** Data dimuat dan dipisahkan menjadi set pelatihan (train_generator) dan set validasi/pengujian (validation_generator) menggunakan flow_from_directory dari ImageDataGenerator. Konfigurasi class_mode='binary' dan classes=['happy', 'sad'] digunakan untuk memastikan klasifikasi biner yang tepat.

3.4. *Arsitektur Model*

Penelitian ini memanfaatkan pendekatan **transfer learning** dengan menggunakan tiga arsitektur CNN pra-terlatih. Model-model ini dimuat dengan bobot yang telah dilatih pada dataset ImageNet. Lapisan klasifikasi asli dari model pra-terlatih dihapus dan diganti dengan lapisan kustom yang didesain untuk tugas klasifikasi biner. *Transfer learning* memungkinkan model untuk memanfaatkan fitur-fitur yang telah dipelajari dari dataset besar, sehingga mengurangi kebutuhan data pelatihan yang besar dan waktu pelatihan yang signifikan [Yosinski et al., 2014]. Setiap model dilatih dalam dua fase:

1. **Fase 1 (Pelatihan Lapisan Kustom):** Lapisan dasar model pra-terlatih dibekukan (*frozen*), sehingga bobotnya tidak diperbarui. Hanya lapisan *dense* kustom yang ditambahkan di atasnya yang dilatih.
2. **Fase 2 (Fine-tuning):** Beberapa lapisan terakhir dari model dasar dibuka (*unfrozen*) dan dilatih kembali bersama dengan lapisan kustom. Pada fase ini, *learning rate* yang sangat kecil digunakan untuk melakukan penyesuaian yang halus pada bobot model dasar, memungkinkan model untuk lebih beradaptasi dengan fitur spesifik dari dataset FER2013.

3.4.1. *Model VGG16*

- **Arsitektur:** VGG16 adalah arsitektur CNN yang terkenal karena kesederhanaan dan kedalamannya, terdiri dari 16 lapisan konvolusi dan *pooling*. Model ini pertama kali diperkenalkan oleh Simonyan dan Zisserman [2014].
- **Input Layer:** Model menerima input gambar dengan bentuk (48, 48, 1) (grayscale), yang kemudian dikonversi menjadi 3 *channel* (RGB) menggunakan lapisan

Lambda sebelum masuk ke model dasar VGG16.

- **Base Model:** VGG16(weights='imagenet', include_top=False, input_tensor=rgb_input) digunakan sebagai *feature extractor*.
- **Lapisan Kustom:** Output dari base_model diratakan (Flatten), diikuti oleh lapisan Dense(256, activation='relu', kernel_regularizer=l2(0.001)), BatchNormalization(), Dropout(0.6), dan diakhiri dengan Dense(1, activation='sigmoid') sebagai lapisan output biner.
- **Fine-tuning:** Pada Fase 2, sekitar **6 lapisan terakhir** dari *base model* VGG16 dibuka (*unfrozen*) dan dilatih kembali.

3.4.2. *Model MobileNetV2*

- **Arsitektur:** MobileNetV2 dirancang untuk efisiensi komputasi dan cocok untuk perangkat *mobile* atau *embedded*. Arsitektur ini menggunakan *inverted residual blocks* dan *depthwise separable convolutions* untuk mengurangi jumlah parameter dan komputasi secara signifikan tanpa mengorbankan akurasi secara drastis [Sandler et al., 2018].
- **Input Layer:** Input gambar (48, 48, 1) (grayscale) dikonversi ke 3 *channel* (RGB), lalu diproses menggunakan fungsi *preprocessing* bawaan MobileNetV2 yang menormalisasi piksel ke rentang [-1, 1].
- **Base Model:** MobileNetV2(weights='imagenet', include_top=False, input_tensor=processed_input) digunakan sebagai *feature extractor*.
- **Lapisan Kustom:** Output dari base_model dilewatkan melalui GlobalAveragePooling2D(), diikuti oleh Dense(128, activation='relu', kernel_regularizer=l2(0.001)), BatchNormalization(), Dropout(0.6), dan Dense(1, activation='sigmoid') sebagai lapisan output biner.
- **Fine-tuning:** Sekitar **60 lapisan terakhir** dari *base model* MobileNetV2 dibuka (*unfrozen*) dan dilatih ulang pada Fase 2.

3.4.3. *Model ResNet50*

- **Arsitektur:** ResNet50 adalah bagian dari keluarga ResNet yang memperkenalkan konsep *residual learning* melalui *skip connections*, memungkinkan pelatihan jaringan saraf yang sangat dalam tanpa degradasi kinerja, yang merupakan masalah umum pada jaringan yang dalam [He et al., 2016].
- **Input Layer:** Model menerima input gambar (48, 48, 1) (grayscale), yang kemudian dikonversi menjadi 3 *channel* (RGB) menggunakan lapisan Lambda sebelum masuk ke model dasar ResNet50.
- **Base Model:** ResNet50(weights='imagenet', include_top=False, input_tensor=rgb_input) digunakan sebagai *feature extractor*.
- **Lapisan Kustom:** Output dari base_model dilewatkan melalui GlobalAveragePooling2D(), diikuti oleh Dense(256, activation='relu', kernel_regularizer=l2(0.001)), BatchNormalization(), Dropout(0.6), dan Dense(1,

activation='sigmoid') sebagai lapisan output biner.

- **Fine-tuning:** Sekitar **20 lapisan terakhir** dari *base model* ResNet50 dibuka (*unfrozen*) dan dilatih ulang pada Fase 2.

3.5. *Konfigurasi Pelatihan*

- **Optimasi:** Adam optimizer digunakan untuk pelatihan di kedua fase. Adam dikenal karena efisiensinya dalam hal komputasi dan persyaratan memori yang rendah, serta kemampuannya untuk berkinerja baik dengan sedikit *tuning* [Kingma & Ba, 2014].
    - **Fase 1:** learning_rate=0.0001.
    - **Fase 2:** learning_rate=0.000005 (nilai yang jauh lebih kecil untuk *fine-tuning* yang presisi).
- **Fungsi *Loss*:** binary_crossentropy digunakan sebagai fungsi *loss* karena ini adalah masalah klasifikasi biner.
- **Metrik Evaluasi:** Akurasi (accuracy) digunakan sebagai metrik utama untuk memantau performa model selama pelatihan dan pada set pengujian.
- **Ukuran *Batch*:** 64.
- **Jumlah *Epochs*:**
    - **Fase 1:** 30 *epoch*.
    - **Fase 2:** 50 *epoch* tambahan, sehingga total *epoch* mencapai 80.
- **Callbacks:** Untuk mengoptimalkan proses pelatihan dan mencegah *overfitting*, beberapa *callback* digunakan:
    - ReduceLROnPlateau: Mengurangi *learning rate* jika val_accuracy tidak membaik selama 5 *epoch* (patience=5).
    - EarlyStopping: Menghentikan pelatihan lebih awal jika val_accuracy tidak menunjukkan peningkatan selama 10 *epoch* (patience=10), dan secara otomatis memulihkan bobot model terbaik dari *epoch* sebelumnya.
    - ModelCheckpoint: Menyimpan bobot model terbaik (berdasarkan akurasi validasi tertinggi) ke Google Drive.
- **Class Weight (Opsional):** class_weight dapat digunakan untuk menangani ketidakseimbangan kelas dalam dataset pelatihan, memberikan bobot lebih pada kelas minoritas jika proporsi sampel tidak seimbang.

3.6. *Lingkungan Pengembangan*

- **Bahasa Pemrograman:** Python.
- **Pustaka/Framework:** TensorFlow 2.x dan Keras, bersama dengan pustaka pendukung seperti NumPy, Matplotlib, Seaborn, dan Scikit-learn.
- **Perangkat Keras:** Pelatihan model dilakukan di Google Colab Pro, memanfaatkan akselerasi GPU (misalnya T4 GPU) untuk mempercepat

komputasi yang intensif.

HASIL DAN PEMBAHASAN

4. *Hasil dan Pembahasan*

Bagian ini menyajikan hasil eksperimen dari pelatihan dan pengujian ketiga model CNN (VGG16, MobileNetV2, ResNet50) dalam tugas klasifikasi biner ekspresi "sad" dan "happy" menggunakan dataset FER2013 yang telah difilter. Pembahasan komparatif mengenai kinerja masing-masing model juga disajikan.

4.1. *Hasil Pelatihan dan Pengujian Model*

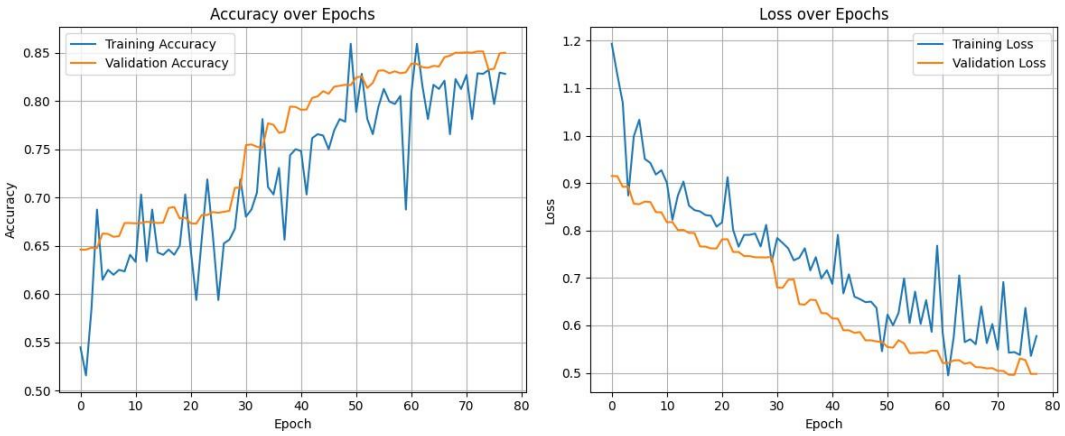
Tabel 4.1. meringkas akurasi pelatihan dan pengujian terbaik yang dicapai oleh masing-masing arsitektur CNN yang diuji. Akurasi pengujian mengacu pada akurasi terbaik yang dicatat pada set validasi/pengujian setelah pelatihan dua fase selesai dan bobot terbaik dimuat.

Tabel 4.1. Ringkasan Akurasi Model VGG16, ResNet50, dan MobileNetV2

Model	Akurasi Training Terbaik	Akurasi Pengujian Terbaik
VGG16	85.94%	85.14%
ResNet50	65.62%	64.63%
MobileNetV2	82.81%	78.25%

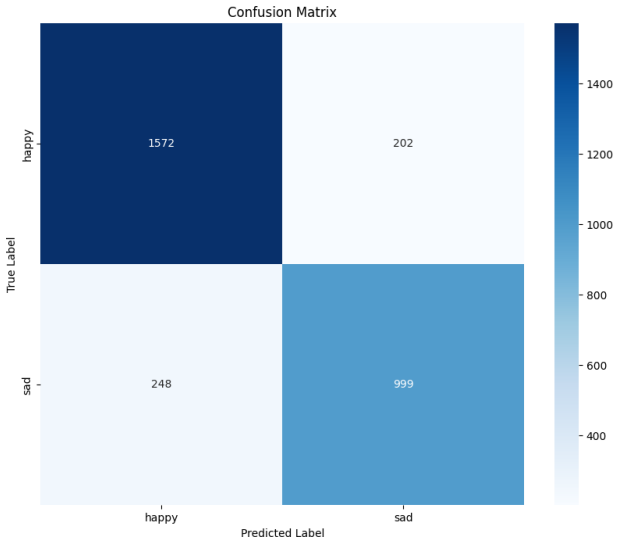
4.1.1. *Hasil Model VGG16*

Model VGG16 menunjukkan kinerja yang sangat baik dalam tugas klasifikasi biner ekspresi "sad" dan "happy" pada dataset FER2013 yang difilter. **Akurasi pelatihan tertinggi mencapai 85.94%**, dan yang lebih penting, **akurasi pengujian mencapai 85.14%**. Tingginya akurasi pengujian mengindikasikan bahwa model VGG16 mampu menggeneralisasi dengan baik pada data yang belum pernah dilihat sebelumnya, dengan *gap* yang relatif kecil antara akurasi pelatihan dan pengujian. Hal ini menunjukkan efektivitas arsitektur VGG16 dalam mengekstraksi fitur-fitur diskriminatif dari citra wajah 48x48 piksel untuk membedakan kedua ekspresi ini [Simonyan & Zisserman, 2014].



Gambar 4.1. Kurva Akurasi dan Loss Model VGG16 selama Pelatihan

Gambar 4.1. menampilkan kurva akurasi dan loss untuk data pelatihan (Training Accuracy dan Training Loss) serta data validasi (Validation Accuracy dan Validation Loss) seiring berjalannya epoch. Terlihat bahwa akurasi validasi menunjukkan tren peningkatan yang konsisten dan relatif mulus, mencapai puncaknya di sekitar 85%. Demikian pula, loss validasi menunjukkan penurunan yang stabil. Konsistensi dan gap yang minim antara kurva pelatihan dan validasi mengindikasikan bahwa model VGG16 tidak mengalami overfitting yang signifikan dan proses pembelajaran berlangsung secara efektif.



Gambar 4.2. Confusion Matrix Model VGG16

Gambar 4.2. menyajikan confusion matrix dari hasil pengujian model VGG16 pada set validasi. Dari matriks ini, dapat dianalisis bahwa:

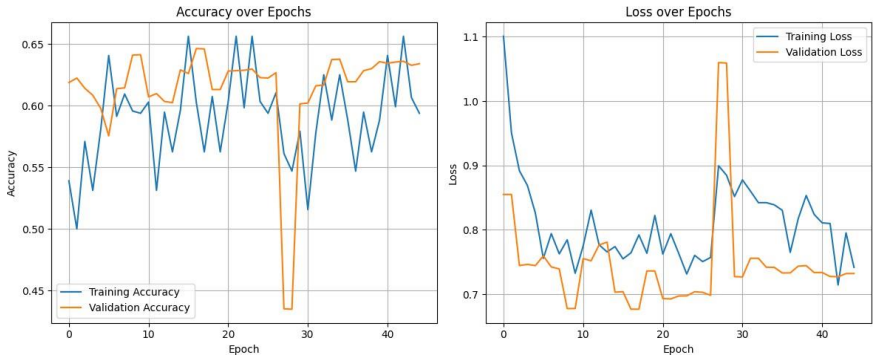
- Sebanyak **1572** citra "happy" diklasifikasikan dengan benar sebagai "happy" (True Positive).
- Sebanyak **999** citra "sad" diklasifikasikan dengan benar sebagai "sad" (True Negative).

- Terdapat **202** citra "sad" yang salah diklasifikasikan sebagai "happy" (False Positive).
- Terdapat 248 citra "happy" yang salah diklasifikasikan sebagai "sad" (False Negative).

Berdasarkan confusion matrix, akurasi keseluruhan model dihitung sebagai (1572+999)/(1572+202+248+999)=2571/3021≈0.8510, atau 85.10%. Angka ini sangat konsisten dengan akurasi pengujian terbaik yang tercatat, memvalidasi kinerja model secara keseluruhan. Meskipun ada beberapa kesalahan klasifikasi, jumlahnya relatif kecil dibandingkan dengan prediksi yang benar, menegaskan performa kuat VGG16. Model sedikit lebih sering salah mengklasifikasikan ekspresi "happy" sebagai "sad" dibandingkan sebaliknya.

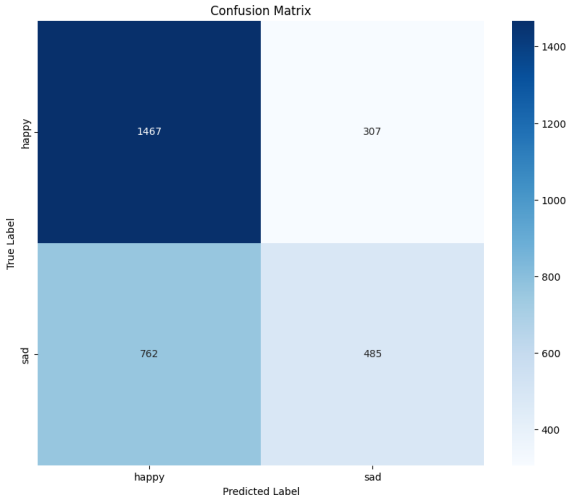
4.1.2. *Hasil Model ResNet50*

Model ResNet50 menunjukkan akurasi pelatihan terbaik keseluruhan sebesar **65.62%** dan akurasi pengujian terbaik keseluruhan sebesar **64.63%**. Kinerja ini menempatkan ResNet50 di posisi ketiga di antara model yang diuji dalam studi ini.



Gambar 4.3. Kurva Akurasi dan Loss Model ResNet50 selama Pelatihan

Gambar 4.3. menampilkan kurva akurasi dan loss selama pelatihan model ResNet50. Kurva akurasi pelatihan dan validasi menunjukkan fluktuasi yang cukup signifikan, yang mungkin mengindikasikan kurangnya stabilitas dalam proses pelatihan atau kesulitan model dalam mempelajari fitur yang konsisten dari dataset. Meskipun ada peningkatan akurasi, puncaknya tidak setinggi model lain, dan terdapat beberapa penurunan tajam pada kurva akurasi validasi, terutama di sekitar epoch 28-30, di mana loss validasi melonjak tinggi. Hal ini menunjukkan adanya ketidakstabilan atau mungkin overfitting sementara pada beberapa epoch, yang mungkin disebabkan oleh kompleksitas arsitektur ResNet50 dan karakteristik dataset 48x48 piksel.



Gambar 4.4. Confusion Matrix Model ResNet50

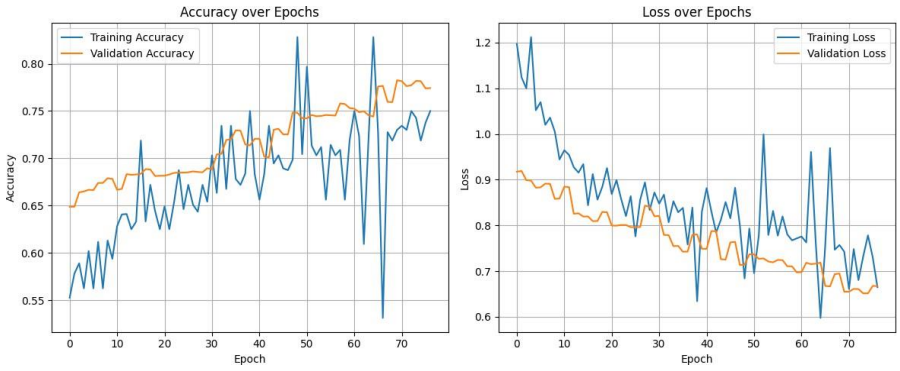
Gambar 4.4. menyajikan confusion matrix dari hasil pengujian model ResNet50. Dari matriks ini, dapat dianalisis bahwa:

- Sebanyak **1467** citra "happy" diklasifikasikan dengan benar sebagai "happy" (True Positive).
- Sebanyak **485** citra "sad" diklasifikasikan dengan benar sebagai "sad" (True Negative).
- Terdapat **307** citra "sad" yang salah diklasifikasikan sebagai "happy" (False Positive).
- Terdapat 762 citra "happy" yang salah diklasifikasikan sebagai "sad" (False Negative).

Berdasarkan confusion matrix, akurasi keseluruhan model dihitung sebagai (1467+485)/(1467+307+762+485)=1952/3021≈0.6461, atau 64.61%. Angka ini sangat konsisten dengan akurasi pengujian terbaik yang tercatat. Analisis kesalahan menunjukkan bahwa model ResNet50 mengalami kesulitan yang lebih besar dalam membedakan kedua kelas, terutama dalam mengklasifikasikan "happy" sebagai "sad" (762 kasus), yang jauh lebih tinggi dibandingkan kesalahan sebaliknya (307 kasus). Hal ini mengindikasikan bias atau kelemahan model dalam menangkap fitur spesifik dari ekspresi "happy" yang membedakannya dari "sad" secara efektif dalam konteks dataset ini, meskipun ResNet50 adalah arsitektur yang sangat dalam dan canggih [He et al., 2016].

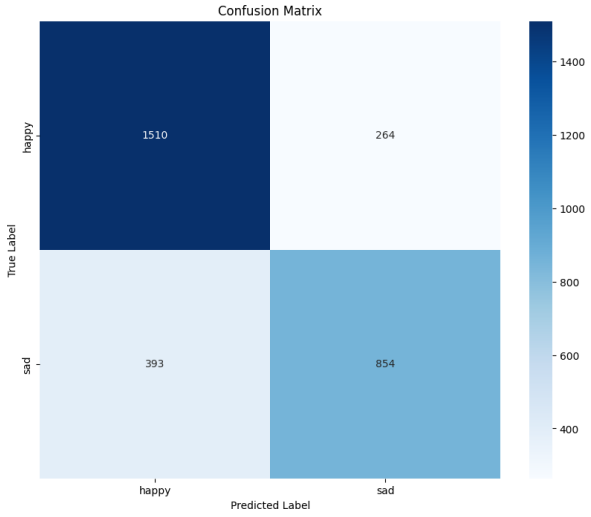
4.1.3. *Hasil Model MobileNetV2*

MobileNetV2 mencapai akurasi pelatihan terbaik keseluruhan sebesar **82.81%** dan akurasi pengujian terbaik keseluruhan sebesar **78.25%**. Kinerja ini menempatkan MobileNetV2 di posisi kedua di antara ketiga model yang diuji, menunjukkan keseimbangan yang baik antara akurasi dan efisiensi.



Gambar 4.5. Kurva Akurasi dan Loss Model MobileNetV2 selama Pelatihan

Gambar 4.5. menampilkan kurva akurasi dan loss selama pelatihan model MobileNetV2. Kurva akurasi validasi menunjukkan peningkatan yang stabil dan mencapai puncaknya di sekitar 78-79%, meskipun terdapat beberapa fluktuasi pada akurasi pelatihan. Kurva loss untuk pelatihan dan validasi menunjukkan penurunan yang konsisten, mengindikasikan bahwa model terus belajar dan mengoptimalkan bobotnya. Performa yang relatif stabil pada validasi menunjukkan kemampuan generalisasi yang baik dari MobileNetV2.



Gambar 4.6. Confusion Matrix Model MobileNetV2

Gambar 4.6. menyajikan confusion matrix dari hasil pengujian model MobileNetV2. Dari matriks ini, dapat dianalisis bahwa:

- Sebanyak **1510** citra "happy" diklasifikasikan dengan benar sebagai "happy" (True Positive).
- Sebanyak **854** citra "sad" diklasifikasikan dengan benar sebagai "sad" (True Negative).
- Terdapat **264** citra "sad" yang salah diklasifikasikan sebagai "happy" (False Positive).
- Terdapat 393 citra "happy" yang salah diklasifikasikan sebagai "sad" (False Negative).

Berdasarkan confusion matrix, akurasi keseluruhan model dihitung sebagai (1510+854)/(1510+264+393+854)=2364/3021≈0.7825, atau 78.25%. Angka ini sangat konsisten dengan akurasi pengujian terbaik yang tercatat. Model ini menunjukkan kinerja yang solid untuk kedua kelas. Kesalahan klasifikasi antara "happy" dan "sad" cukup seimbang, meskipun sedikit lebih banyak "happy" yang salah diklasifikasikan sebagai "sad".

4.2. *Analisis Perbandingan Kinerja Model*

Dari hasil yang diperoleh, terlihat bahwa **VGG16 adalah model dengan kinerja terbaik** dalam mengklasifikasikan ekspresi "sad" dan "happy" pada dataset FER2013 yang difilter, mencapai akurasi pengujian **85.14%**. Superioritas VGG16 dalam kasus ini dapat diatribusikan pada kemampuannya untuk mempelajari fitur hierarkis yang kaya dan representasi visual yang kuat, meskipun dengan jumlah parameter yang lebih besar [Simonyan & Zisserman, 2014].

MobileNetV2 menempati posisi kedua, menunjukkan *trade-off* yang baik antara akurasi dan efisiensi. Akurasinya yang mencapai **78.25%** relatif dekat dengan VGG16, namun dengan ukuran model yang jauh lebih kecil dan kecepatan inferensi yang lebih cepat berkat desain *depthwise separable convolutions*nya [Sandler et al., 2018]. Hal ini menjadikan MobileNetV2 kandidat kuat untuk aplikasi *real-time* atau pada perangkat dengan sumber daya komputasi terbatas.

Di sisi lain, **ResNet50 menunjukkan akurasi terendah** di antara ketiga model, yaitu **64.63%**. Ini adalah hasil yang menarik mengingat kedalaman dan kemampuan ResNet50 untuk mengatasi masalah *vanishing gradient* dengan *residual connections* [He et al., 2016].

Perbedaan kinerja antar model juga dapat dipengaruhi oleh karakteristik dataset FER2013 yang merupakan citra *grayscale* 48x48 piksel. Resolusi yang relatif rendah mungkin tidak menyediakan detail spasial yang cukup untuk sepenuhnya memanfaatkan kedalaman dan kompleksitas arsitektur seperti ResNet50 tanpa *fine-tuning* yang sangat spesifik dan cermat [Shorten & Khoshgoftaar, 2019; Yosinski et al., 2014].

KESIMPULAN

Penelitian ini berhasil mencapai target utamanya, yaitu membandingkan tingkat akurasi tiga model CNN—VGG16, ResNet50, dan MobileNetV2—dalam pengenalan ekspresi wajah sedih dan senang pada dataset FER2013, dengan hasil menunjukkan bahwa VGG16 memiliki performa terbaik dengan akurasi pengujian sebesar 85.14%. Kegiatan ini memberikan dampak positif terhadap pemahaman dan penerapan arsitektur deep learning dalam klasifikasi ekspresi wajah, serta memperlihatkan potensi model VGG16 untuk diterapkan dalam aplikasi nyata yang membutuhkan pengenalan emosi secara akurat. Untuk kegiatan selanjutnya, disarankan melakukan eksplorasi terhadap model lightweight lainnya serta pengujian pada dataset yang lebih kompleks dan realistis, termasuk penambahan kelas emosi lainnya guna meningkatkan generalisasi dan relevansi model dalam skenario dunia nyata.

Prosiding Seminar Nasional Universitas Ma Chung 2026
(Sistem Informasi dan Informatika)
pp. 216-229
E-ISSN: 2808-7771

UCAPAN TERIMA KASIH

Kami ingin menyampaikan rasa terima kasih yang sebesar besar nya kepada Bapak Windra Swastika, Ph.D dan anggota kelompok yang berisikan Alfond Manuel Joseph, Davin Valerian dan Steven Surya Putra , atas kerjasama yang mereka berikan. Kelancaran kegiatan Perbandingan Akurasi Model VGG16, ResNet50, dan MobileNet dalam Pengenalan Ekspresi Wajah Sedih dan Senang pada Dataset FER2013 ini tidak akan terwujud tanpa bantuan yang luar biasa dari Bapak Windra Swastika, Ph.D dan anggota kelompok ini.

DAFTAR PUSTAKA

Goodfellow, I. J., Erhan, D., Luc, P., Mirza, M., Hamner, B., Cifka, W., Cubuk, E., Dahl, F., Dangel, P., Fiske, C., Glass, J., Guadarrama, S., Krizhevsky, A., Kumar, S., Lavoie, D., Lee, Y., Levenberg, J., Ma, X., Masci, J., ... & Bengio, Y. (2013). **Challenges in representation learning: A report on three machine learning contests**. *International Conference on Neural Information Processing Systems (NeurIPS) Workshops*.

He, K., Zhang, X., Ren, S., & Sun, J. (2016). **Deep residual learning for image recognition**. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 770-778).

Kingma, D. P., & Ba, J. (2014). **Adam: A method for stochastic optimization**. *arXiv preprint arXiv:1412.6980*.

Sandler, M., Howard, A., Zhu, M., Zhmoginov, A., & Chen, L. C. (2018). **MobileNetV2: Inverted residuals and linear bottlenecks**. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 4510-4520).

Shorten, C., & Khoshgoftaar, T. M. (2019). **A survey on image data augmentation for deep learning**. *Journal of Big Data*, 6(1), 1–48.

Simonyan, K., & Zisserman, A. (2014). **Very deep convolutional networks for large-scale image recognition**. *arXiv preprint arXiv:1409.1556*.

Yosinski, J., Clune, J., Erhan, D., & Lipson, H. (2014). **How transferable are features in deep neural networks?**. *Advances in Neural Information Processing Systems*, 27.