

# Implementasi Algoritma *Convolutional Neural Network* Arsitektur *MobileNetV2* Untuk Klasifikasi Ekspresi Wajah Pada Dataset FER

Devfris Dhimas Permana Putra<sup>1</sup>, Galang Kurnia Anaga<sup>2</sup>, Wahyu Tia Fitriyana<sup>3</sup>

<sup>1,2,3</sup>Teknik Informatika, Fakultas Teknik dan Ilmu Komputer, Universitas Nusantara PGRI Kediri

E-mail: <sup>1</sup>[\\*<sup>1</sup>devfrisdhimasp@gmail.com](mailto:devfrisdhimasp@gmail.com), <sup>2</sup>[galangkurniaa1@gmail.com](mailto:galangkurniaa1@gmail.com), <sup>3</sup>[wtiafitriyana@gmail.com](mailto:wtiatifitriyana@gmail.com)

**Abstrak** – Wajah merupakan bagian tubuh yang menjadi fokus ketika melakukan interaksi sosial. Manusia menunjukkan emosinya melalui ekspresi pada wajahnya. Face processing merupakan salah satu teknologi dalam bidang computer vision yang sedang berkembang saat ini. Salah satunya adalah pengenalan ekspresi wajah (facial expression recognition). Dalam penelitian ini, kami mengusulkan penggunaan algoritma *Convolutional Neural Network* (CNN) dengan arsitektur *MobileNetV2* untuk klasifikasi ekspresi wajah. Model CNN dilatih menggunakan dataset FER yang berisi 4200 gambar wajah dengan tujuh kelas ekspresi, yaitu senang, marah, sedih, takut, jijik, terkejut, dan netral. Hasil penelitian menunjukkan bahwa model CNN dengan arsitektur *MobileNetV2* dapat mencapai akurasi sebesar 42,68%. Untuk meningkatkan akurasi model, perlu dilakukan eksplorasi lanjutan dengan teknik-teknik seperti augmentasi data dan dropout. Selain itu, disarankan untuk menggunakan data latih yang lebih besar.

**Kata Kunci** — klasifikasi ekspresi wajah, CNN, *MobileNetV2*, FER-2013

## 1. PENDAHULUAN

Wajah merupakan bagian tubuh yang menjadi fokus Ketika melakukan interaksi sosial. Manusia menunjukkan emosinya melalui ekspresi pada wajahnya. Wajah manusia menyediakan banyak informasi. Banyak hal menarik yang bisa diperhatikan dan dipelajari. Face processing merupakan salah satu teknologi dalam bidang computer vision yang sedang berkembang saat ini. Salah satunya adalah pengenalan ekspresi wajah (facial expression recognition). Ekspresi pada wajah merupakan bentuk dari emosi seseorang. Pada bidang marketing, kepuasan pelanggan terhadap pelayanan penting untuk diperhatikan oleh perusahaan. Salah satu indikator yang dapat menunjukkan kepuasan pelanggan terhadap pelayanan adalah ekspresi wajah atau emosi pelanggan. Dengan adanya pengenalan ekspresi wajah, perusahaan dapat mengevaluasi setiap pelayanan yang telah diberikan melalui sistem pengenalan ekspresi wajah. Oleh, karena itu peneliti tertarik untuk mengimplementasikan salah satu algoritma *Deep Learning* terhadap kasus ini.

*Convolutional Neural Network* (CNN) merupakan metode *deep learning* yang sering digunakan dalam pengenalan pola citra. CNN ideal digunakan untuk tugas-tugas seperti klasifikasi wajah, pengenalan objek, dan segmentasi citra. CNN juga dapat mengklasifikasikan jenis jenis ekspresi wajah dengan menggunakan serangkaian proses konvolusi untuk mengekstrak fitur-fitur pada citra wajah. Proses konvolusi adalah operasi matematika yang mengalikan bagian-bagian dari citra dengan filter. Filter inilah yang digunakan untuk mengekstrak fitur tertentu dari citra wajah. Oleh karena itu, CNN merupakan metode yang efektif untuk pengenalan wajah. *MobileNet* merupakan salah satu contoh arsitektur CNN. *MobileNet* dirancang untuk komputasi yang efisien dengan menggunakan teknik kompresi yang disebut *inverted residual blocks* untuk mengurangi jumlah parameter tanpa mengorbankan kinerjanya. *Inverted residual blocks* terdiri dari dua bagian utama, yaitu *Bottleneck* dan *Expander*. Pada penelitian ini, peneliti menggunakan arsitektur dari CNN yaitu *MobileNetV2*[2].

Dalam beberapa tahun terakhir, terdapat beberapa penelitian yang membahas tentang CNN dan arsitektur CNN *MobileNetV2* yang digunakan untuk mengklasifikasikan objek citra. Salah satu contohnya adalah penelitian yang dilakukan oleh Derry Alamsyah, dan Dicky Pratama pada tahun 2020. Penelitian tersebut membahas tentang implementasi CNN untuk klasifikasi ekspresi citra wajah pada dataset FER-2013 yang diambil dari situs Kaggle[9]. Dataset dibagi menjadi tujuh kelas, yaitu *angry*, *sad*, *disgusting*, *neutral*, *happy*, dan *fear*. Hasil akurasi tertinggi yang diperoleh dari penelitian ini adalah sebesar 66% dengan menggunakan *Adamax optimizer*.

Pada tahun 2023 penelitian lain yang dilakukan oleh Anti Nada Nafisa, Erika Nia Devina BR Purba, dkk CNN dengan arsitektur *MobileNetV2* digunakan untuk klasifikasi penyakit tumor otak *Glioma*, *Pituitary*, dan *Meningioma*[10]. Dataset yang digunakan dibagi menjadi tiga kelas sesuai jenis tumor yang sudah diuraikan

sebelumnya. Hasil akurasi yang diperoleh dengan data testing didapatkan nilai evaluasi sebesar 78% sedangkan hasil akurasi dengan data validasi didapatkan nilai evaluasi sebesar 83%.

Berdasarkan latar belakang yang telah diuraikan, maka peneliti memutuskan menggunakan algoritma *Convolutional Neural Network* (CNN) dengan arsitektur model *MobileNetV2* untuk klasifikasi ekspresi wajah pada data FER-2013.

### 1.1 Ekspresi Wajah

Manusia mempunyai kemampuan untuk menampilkan emosi dalam bentuk ekspresi wajah. Terdapat 6 ekspresi wajah yang biasa digunakan oleh manusia yaitu : sedih, senang, marah, jijik, takut dan terkejut. Setiap ekspresi wajah tersebut memiliki perbedaan pada tarikan otot-otot wajah[1]. Salah satu gerak tubuh yang sering digunakan dalam proses komunikasi adalah ekspresi. Ekspresi wajah mencakup kombinasi isyarat, di mana setiap isyarat memiliki makna dan dapat memengaruhi pesan verbal yang akan diungkapkan. Komunikasi juga dapat dipancarkan melalui bahasa tubuh dan Gerakan (*sign language*)[1].

### 1.2 *Convolutional Neural Network*

*Convolutional Neural Network* (CNN) tergolong dalam kategori deep neural network, mirip dengan *neural network* pada umumnya, yang terdiri dari sejumlah besar *neuron* memiliki bobot dan bias yang dapat diatur melalui proses pelatihan[2]. Setiap *neuron* dalam CNN menerima berbagai input, menjalankan operasi komputasi, menghasilkan output, dan terhubung dengan neuron lainnya. Penggunaan istilah “konvolusi” mengacu operasi aljabar linier yang melibatkan perkalian matriks filter pada citra yang sedang diolah. Proses ini, yang dikenal sebagai lapisan konvolusi, merupakan salah satu jenis lapisan yang dapat ada dalam struktur jaringan CNN, meskipun sering dianggap sehingga elemen paling vital. Lapisan lain yang umumnya digunakan melibatkan *Pooling Layer*, *Normalization Layer*, dan *Fully Convolutional (FC) Layer*. Keistimewaan CNN terletak pada asumsi eksplisit bahwa inputnya selalu berupa citra digital, memepermudah penyesuaian parameter arsitektur jaringan dan meningkatkan efisiensi secara keseluruhan.

### 1.3 *MobileNetV2*

Keberhasilan metode *Deep learning* pada berbagai tugas visi komputer telah menyebabkan penerapannya yang luas [5] dan gambaran umum penerapannya di *The Face Expression Recognition* (FER) diberikan dalam penelitian [7]. Pada bagian ini, perubahan arsitektur yang diperkenalkan di *MobileNetV1* [6] dan *MobileNetV2* yang secara signifikan mengurangi biaya komputasi dibandingkan dengan metode *Deep learning* lainnya yang ditujukan untuk pengenalan objek dengan hanya sedikit penurunan kinerja[6].

*MobileNet*, merupakan salah satu arsitektur CNN yang dapat digunakan untuk mengatasi kebutuhan akan *computing resource* berlebih. Perbedaan mendasar antara arsitektur *MobileNet* dan arsitektur CNN pada umumnya adalah penggunaan lapisan atau layer konvolusi dengan ketebalan filter yang sesuai dengan ketebalan dari input *image*[6].

Perbedaan mendasar antara arsitektur *MobileNet* dan arsitektur CNN pada umumnya adalah penggunaan lapisan atau layer konvolusi dengan ketebalan filter yang sesuai dengan ketebalan dari input *image*. *MobileNet* membagi konvolusi menjadi *depthwise convolution* dan *pointwise convolution*. Lapisan pertama disebut *depthwise convolution*, karena melakukan penyaringan ringan dengan menerapkan filter konvolusional tunggal per saluran masukan. Lapisan kedua adalah konvolusi  $1 \times 1$ , yang disebut *pointwise convolution*, yang bertanggung jawab untuk membangun fitur baru melalui komputasi kombinasi linier dari saluran input.

*MobileNetV2* meningkatkan kinerja model seluler pada berbagai tugas dan tolok ukur serta di seluruh spektrum ukuran model yang berbeda. Sama seperti *MobileNetV1*, *MobileNetV2* masih menggunakan *depthwise* dan *pointwise convolution*. *MobileNetV2* menambahkan dua fitur baru yaitu: 1) linear bottleneck, dan 2) shortcut connections antar bottlenecks. Pada bagian bottleneck terdapat input dan output antara model sedangkan lapisan atau layer bagian dalam meng-enkapsulasi kemampuan model untuk mengubah input dari konsep tingkat yang lebih rendah seperti *pixels* ke deskriptor tingkat yang lebih tinggi seperti kategori gambar). Pada akhirnya, seperti halnya koneksi residual pada CNN tradisional, *shortcut* antar *bottlenecks* memungkinkan training atau pelatihan yang lebih cepat dan akurasi yang lebih baik [8]. Pada beberapa penelitian menunjukkan bahwa model *MobiExpressNet*, mencapai akurasi tinggi 67,96% pada set data FER2013 dengan ukuran model sekitar 75000 parameter dan komputasi  $1 \times 10^6$  FLOP[5].

## 2. METODE PENELITIAN

### 2.1 Pengumpulan Dataset

Gambar yang digunakan pada penelitian ini adalah data yang bersifat publik. Data citra yang digunakan adalah dataset FER yang didapatkan dari situs Kaggle. Dataset yang digunakan sebanyak 4200 citra. Dataset terbagi menjadi 2 kelas dimana setiap kelas terbagi lagi menjadi 7 sub-kelas. Sub-kelas tersebut meliputi kelas happy, disgust, fear, angry, neutral, surprise, dan sad. Data terdiri dari gambar wajah dalam *grayscale* yang berukuran 48x48 piksel. Gambar 1 adalah contoh dataset FER yang digunakan dalam penelitian ini.



Gambar 1. Data FER

### 2.2 Cleaning Data

Persebaran data yang tidak merata dapat menyebabkan terjadinya kesalahan ketika melakukan klasifikasi objek. Oleh karena itu diperlukan proses cleaning data atau pembersihan data. Hal tersebut dapat diatasi dengan mengurangi data yang berlebih hingga setiap kelas memiliki jumlah data yang sama. Cleaning data citra pada penelitian ini dilakukan dengan mengurangi dataset yang semula berjumlah 35000 dikurangi menjadi 4200 dimana setiap kelasnya memiliki data yang sama. Data *grayscale* yang sudah ada dilakukan proses *resize* semua data citra yang semula memiliki ukuran 48x48 piksel *dirresize* ke ukuran 96x96 piksel untuk menjalankan arsitektur *MobileNetV2*[4].

### 2.3 Pembagian Dataset

Dalam membuat model *deep learning* diperlukan data untuk melatih model yang dibuat. Pertimbangan utama dalam membangun dataset adalah untuk menentukan ukuran setiap set. Pembuatan model *deep learning Convolutional Neural Network* pada penelitian ini, dataset dibagi ke dalam 2 kelas, yaitu data *train*, dan data *validation*. Pada kelas data *train*, terdapat 7 sub-kelas yang masing-masing sub-kelas memiliki 500 data. Sementara pada kelas data *validation*, terdapat 7 sub-kelas yang masing-masing sub-kelas memiliki 100 data. Tabel 1 adalah tabel pembagian data yang dilakukan pada penelitian ini.

Tabel 1. Pembagian Dataset

	Data Train	Data Validasi
Happy	500	100
Disgust	500	100

	Data Train	Data Validasi
<b>Fear</b>	500	100
<b>Angry</b>	500	100
<b>Neutral</b>	500	100
<b>Surprise</b>	500	100
<b>Sad</b>	500	100

#### 2.4 Perancangan Arsitektur *MobileNetV2*

Model klasifikasi objek akan dibangun dengan teknik *transfer learning* menggunakan pretrained model *MobileNetV2*. Arsitektur *MobileNetV2* dipilih karena model ini memiliki performa yang tinggi dengan komputasi yang lebih efisien. *MobileNetV2* dan arsitektur CNN memiliki perbedaan pada *convolutional layer*. *Convolutional layer* pada *MobileNetV2* menggunakan ketebalan filter yang sesuai dengan ketebalan dari gambar input. Tabel 2 merupakan arsitektur *MobileNetV2* pada penelitian ini.

Tabel 2. Arsitektur *MobileNetV2*

Bagian	Layer	Filter	Stride	Output
Input	-	-	-	96 x 96 x 3 48 x 48 x
Stem	Conv2D	3 x 3	2	32
Depthwise Separable Conv2D	3 x 3	1	48 x 48 x 128	
Depthwise Separable Conv2D	3 x 3	2	24 x 24 x 128	
Inverted Residual Block	-	-	-	24 x 24 x 128
Inverted Residual Block	-	-	-	24 x 24 x 128
...	...	...	...	...
Inverted Residual Block	-	-	-	12 x 12 x 128
Depthwise Separable Conv2D	3 x 3	1	12 x 12 x 128	
Depthwise Separable Conv2D	3 x 3	2	6 x 6 x 128	
Classifier	Global Average Pooling	-	-	128
Fully Connected	-	-	-	1000

```
Model: "sequential_4"
Layer (type)                Output Shape                Param #
-----
mobilenetv2_1.00_96 (Func    (None, 3, 3, 1280)         2257984
tional)
flatten_4 (Flatten)         (None, 11520)              0
dense_12 (Dense)            (None, 128)                1474688
dense_13 (Dense)            (None, 64)                 8256
dense_14 (Dense)            (None, 7)                  455
-----
Total params: 3741383 (14.27 MB)
Trainable params: 3707271 (14.14 MB)
Non-trainable params: 34112 (133.25 KB)
```

Gambar 2. Arsitektur Pre-trained MobileNetV2

### 2.5 Pelatihan Model

Pelatihan model pada penelitian ini menggunakan 2 jenis *optimizer*, Adam dan SGD. Optimizer Adam yaitu algoritma yang merupakan perkembangan dari algoritma Stochastic Gradient Descent (SGD) klasik dimana bobot network telah diperbarui. Sedangkan *optimizer* SGD penurunan gradien stokastik secara acak membagi kumpulan observasi menjadi minibatch. Untuk setiap minibatch, gradien dihitung dan vektor dipindahkan. Sementara itu data dibagi kedalam data *train* dan data *validation* dengan 45 *epoch*.

### 2.6 Evaluasi

Evaluasi pada penelitian ini menggunakan *confusion matrix* dengan menggunakan nilai akurasi, akurasi train, dan akurasi val.

## 3. HASIL DAN PEMBAHASAN

### 3.1 Model

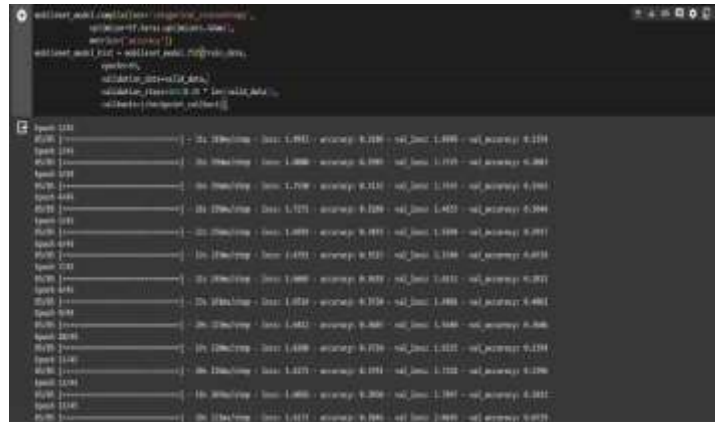
Pada gambar 3 Model ini memiliki 2283604 parameter, yang dapat dibagi menjadi dua kategori. Parameter yang dapat dilatih dan parameter yang tidak dapat dilatih. Parameter yang dapat dilatih 25620 parameter yang dapat diubah selama pelatihan model. Parameter yang tidak dapat dilatih 2257984 parameter yang tidak dapat diubah selama pelatihan model. Dengan model yang memiliki enam lapisan. Input layer ini menerima input dari data gambar. Data *augmentation* layer ini menerapkan transformasi acak ke gambar input untuk meningkatkan keragaman dataset. *MobileNetV2* layer ini merupakan arsitektur dasar model yang bertanggung jawab sebagai pengklasifikasi gambar. *Global average pooling layer* ini mengurangi dimensi output dari *MobileNetV2* layer. Output layer ini menghasilkan prediksi kelas untuk gambar input.

```
[ ] mobilenet_model.summary()
Model: "model_3"
Layer (type)                Output Shape                Param #
-----
input_layer (InputLayer)    [(None, 224, 224, 3)]      0
data_augmentation (Sequenti (None, None, None, 3)      0
al)
mobilenetv2_1.00_224 (Func    (None, None, None, 1280)   2257984
tional)
global_average_pooling (Gl    (None, 1280)              0
obalAveragePooling2D)
output_layer (Dense)        (None, 7)                 8967
-----
Total params: 2260951 (8.65 MB)
Trainable params: 8967 (35.03 KB)
Non-trainable params: 2257984 (8.61 MB)
```

Gambar 3. Model

### 3.2 Train dan valid data

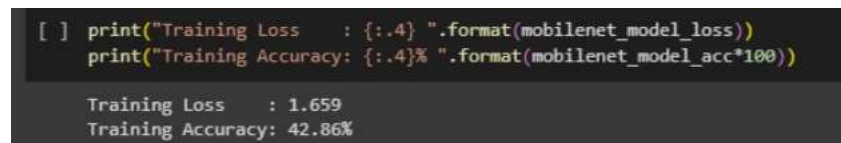
Pada gambar 4, tersebut menunjukkan hasil pelatihan model. Hasil pelatihan ini dapat diukur dengan menggunakan metrik akurasi dan presisi. Berdasarkan gambar tersebut, model ini memiliki akurasi 42,86%.



Gambar 4. Train dan valid data

### 3.3 Evaluasi Data

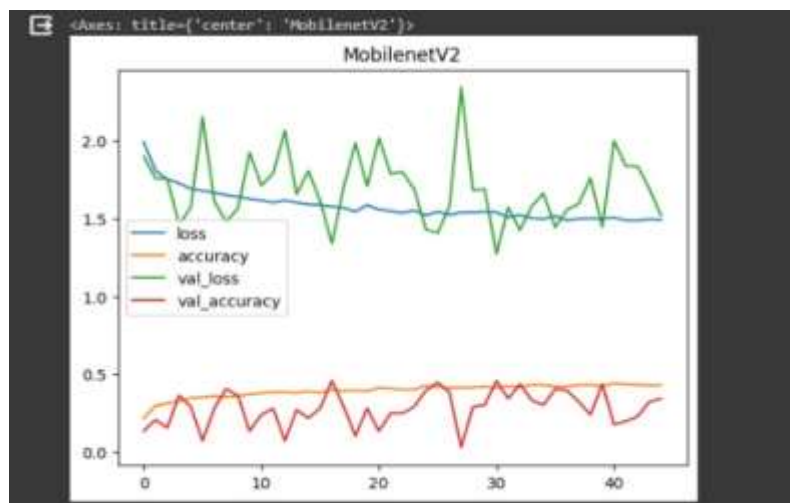
Hasil dari evaluasi model klasifikasi gambar tulisan aksara Jawa menggunakan arsitektur MobileNetV2. Berdasarkan gambar 5 tersebut, model ini memiliki akurasi 42,86% dan presisi 41,95%.



Gambar 5. Evaluasi data

### 3.4 Analisa Hasil

Pada gambar 6 merupakan hasil klasifikasi ekspresi wajah manusia dilakukan menggunakan metode CNN arsitektur MobileNetV2. Dataset yang digunakan adalah dataset tulisan aksara Jawa yang tersedia di Kaggle. Dataset tersebut berisi sekitar 4200 gambar ekspresi wajah manusia yang dibagi menjadi 7 ekspresi, yaitu kelas senang, marah, sedih, takut, jijik, terkejut, dan netral.



Gambar 6. Analisa hasil

#### 4. SIMPULAN

Berdasarkan hasil penelitian, dapat disimpulkan bahwa model CNN dengan arsitektur *MobileNetV2* tidak cukup baik digunakan untuk klasifikasi ekspresi wajah dengan akurasi yang cukup buruk. Akurasi yang didapatkan hanya sebesar 42,68% ketika dilakukan simulasi model dengan citra diluar dataset. Model ini perlu eksplorasi lanjutan untuk meningkatkan akurasinya sebelum dikembangkan untuk penelitian-penelitian selanjutnya.

#### 5. SARAN

Meskipun model CNN dengan arsitektur *MobileNetV2* menunjukkan hasil yang kurang memuaskan dalam penelitian klasifikasi ekspresi wajah dengan akurasi 42,68%, potensinya untuk ditingkatkan masih terbuka lebar. Berikut beberapa saran untuk mengeksplorasi peningkatan akurasi model:

- a. Gunakan teknik augmentasi data untuk meningkatkan keragaman data latih. Hal ini dapat membantu model untuk belajar pola-pola yang lebih umum dan mencegah *overfitting*.
- b. Gunakan teknik dropout untuk mencegah *overfitting*. Hal ini dapat membantu model untuk belajar lebih tergantung pada informasi yang lebih umum dan mengurangi ketergantungan pada informasi yang spesifik.
- c. Gunakan teknik regularization untuk mengurangi kompleksitas model dan meningkatkan generalisasi model. Hal ini dapat membantu model untuk belajar pola-pola yang lebih umum dan dapat diterapkan pada data baru.

Dengan melakukan eksplorasi lanjutan dengan teknik-teknik ini, diharapkan dapat meningkatkan akurasi model klasifikasi ekspresi wajah berbasis *MobileNetV2*.

#### DAFTAR PUSTAKA

- [1] LIESTANTYO H, N. B., Widodo, A. P., & Suryono, S. (2021). RANCANG BANGUN SISTEM DAN EVALUASI MODEL MACHINE LEARNING MOBILENETSV2 DAN NASNET UNTUK MENDETEKSI EMOSI PADA WAJAH (Doctoral dissertation, School of Postgraduate Studies).
- [2] Sihombing, R. S. I., Siregar, R. N. T., Sitorus, V., & Sitompul, T. S. (2023). Pengenalan Ekspresi Wajah Menggunakan Convolutional Neural Network (CNN). *Journal of Creative Student Research*, 1(6), 89-97.
- [3] Alamsyah, Derry, and Dicky Pratama. "Implementasi Convolutional Neural Networks (CNN) untuk Klasifikasi Ekspresi Citra Wajah pada FER-2013 Dataset." (*JurTI*) *Jurnal Teknologi Informasi* 4.2 (2020): 350-355.
- [4] Guntoro, AL Sigit, Edy Julianto, and Djoko Budiyo. "Pengenalan Ekspresi Wajah Menggunakan Convolutional Neural Network." *Jurnal Informatika Atma Jogja* 3.2 (2022): 155-160.
- [5] Cotter, S. F., 2020, *MobiExpressNet: A Deep Learning Network for Face Expression Recognition on Smart Phones*, Dalam 2020 IEEE International Conference on Consumer Electronics (ICCE) (hlm. 1–4). Las Vegas, NV, USA: IEEE
- [6] Howard, A. G., Zhu, M., Chen, B., Kalenichenko, D., Wang, W., Weyand, T., Andreetto, M., dan Adam, H., 2017, *MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications*, ArXiv:1704.04861 [Cs].
- [7] Li, S., dan Deng, W., 2020, *Deep Facial Expression Recognition: A Survey*, *IEEE Transactions on Affective Computing*, 1–1.
- [8] Sandler, M., Howard, A., Zhu, M., Zhmoginov, A., dan Chen, L.-C., 2018, *MobileNetV2: Inverted Residuals and Linear Bottlenecks*, Dalam 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition (hlm. 4510–4520). Salt Lake City, UT: IEEE.
- [9] Alamsyah, D., & Pratama, D. (2020). Implementasi Convolutional Neural Networks (CNN) untuk Klasifikasi Ekspresi Citra Wajah pada FER-2013 Dataset. (*JurTI*) *Jurnal Teknologi Informasi*, 4(2), 350-355.
- [10] Harahap, F. A. A., Nafisa, A. N., Purba, E. N. D. B., & Putri, N. A. (2023). Implementasi Algoritma Convolutional Neural Network Arsitektur Model MobileNetV2 dalam Klasifikasi Penyakit Tumor Otak Glioma, Pituitary dan Meningioma. *Jurnal Teknologi Informasi, Komputer, dan Aplikasinya (JTika)*, 5(1), 53-61.