

PENGENALAN EKSPRESI WAJAH BERBASIS CONVOLUTIONAL NEURAL NETWORK DENGAN MODEL ARSITEKTUR VGG16

FACIAL EXPRESSION RECOGNITION BASED ON CONVOLUTIONAL NEURAL NETWORK WITH VGG16 ARCHITECTURE MODEL

Rizqy Joventus Gunawan¹, Budhi Irawan², Casi Setianingsih³

^{1,2,3} Universitas Telkom, Bandung

¹rizqyjoe@student.telkomuniversity.ac.id, ²budhiirawan@telkomuniversity.ac.id,

³setiacasie@telkomuniversity.ac.id

Abstrak

Teknologi sistem pengenalan ekspresi wajah dapat terus berkembang seiring waktu dengan penelitian-penelitian baru untuk menciptakan sistem yang lebih canggih dan akurat, salah satu perkembangannya dalam hal ini yaitu menggunakan metode Deep Learning. Deep Learning menghasilkan kinerja yang sangat baik pada sistem pengenalan ekspresi wajah yang menggunakan jumlah data yang banyak. Salah satu algoritma Deep Learning yaitu Convolutional Neural Network (CNN) merupakan algoritma yang cocok diimplementasikan pada sistem pengenalan ekspresi wajah dengan jumlah data yang banyak. Pada penelitian ini merancang sistem pengenalan ekspresi wajah menggunakan algoritma CNN, model yang digunakan merupakan arsitektur berbasis CNN yaitu Visual Geometry Group 16 (VGG16). Dataset yang digunakan pada sistem ini adalah FER2013, dimana dataset ini memiliki jumlah citra wajah yang banyak yaitu berjumlah 35.887 citra dengan 7 kategori emosi. Didapatkan hasil performansi terbaik pada model yang diajukan dengan yaitu model Modified VGG16 dengan parameternya yaitu menggunakan data augmentasi, epoch 100, dan learning rate 0.001 yang mencapai akurasi uji sebesar 70,63%, akurasi ini jauh lebih baik dibanding penelitian sebelumnya yang menggunakan basis model VGG16 dan dataset FER2013.

Abstract

Facial expression recognition system technology can continue to develop over time with new researches to create more sophisticated and accurate systems, one of the developments in this case is using the Deep Learning method. Deep Learning produces excellent performance on facial expression recognition systems that use large amounts of data. One of the Deep Learning algorithms, namely Convolutional Neural Network (CNN) is an algorithm that is suitable to be implemented in facial expression recognition systems with large amounts of data. In this research, to design a facial expression recognition system using the CNN algorithm, the model used is a CNN-based architecture, namely Visual Geometry Group-16 Weight Layer (VGG16). The dataset used in this system is FER2013, where this dataset has a large number of facial images, amounting to 35,887 images with 7 categories of emotions. The best performance results were obtained in the proposed model, namely the Modified VGG16 model with the parameters using augmentation data, epoch 100, and learning rate 0.001 which reached a test accuracy of 70.63%, this accuracy is much better than previous studies using the VGG16 model base and the FER2013 dataset.

Keywords: Facial Expression Recognition, Convolutional Neural Network, VGG16, FER2013.

1. Pendahuluan

Ekspresi wajah digunakan sebagai bentuk respon alami yang menggambarkan perasaan atau emosi seseorang dalam berinteraksi dengan suatu hal tertentu. Dalam interaksi antar sesama manusia, ekspresi digunakan sebagai bagian dari komunikasi. Pada hal lain, perkembangan teknologi dari masa ke masa sangatlah pesat, kunci perkembangan teknologi ada pada pengguna atau user. Oleh karena itu diciptakanlah sistem facial expression recognition (FER) dimana sistem ini dapat mengenali emosi manusia dari ekspresi wajah. Teknologi sistem facial expression recognition sudah banyak digunakan dengan macam-macam implementasinya dan banyak penelitian terkait sistem tersebut. Salah satu implementasinya yaitu merancang sistem pengenalan ekspresi wajah

menggunakan algoritma Convolutional Neural Network (CNN). Alasan penggunaan algoritma CNN yaitu karena CNN merupakan salah satu algoritma Deep Learning yang mampu melakukan klasifikasi data dari jumlah data yang banyak dan dapat memperoleh nilai akurasi yang tinggi dalam hal klasifikasi. Terkait hal tersebut terdapat salah satu penelitian yang dilakukan oleh Bodavarapu et al. [1] dimana penelitian tersebut merancang sistem pengenalan ekspresi wajah menggunakan algoritma CNN dan dataset FER2013. Hasil akurasi yang didapatkan dari penelitian tersebut mencapai 60% pada salah satu model arsitektur yang diuji yaitu Visual Geometry Group 16 (VGG16). Dari hasil akurasi yang diperoleh tersebut dapat diyakini bahwa akurasi dapat ditingkatkan lagi dengan melakukan modifikasi pada model tersebut.

Pada penelitian ini, merancang sistem pengenalan ekspresi wajah dengan menggunakan model arsitektur CNN yaitu VGG16 dan menggunakan dataset FER2013 yang bertujuan untuk menganalisis tingkat akurasi pada model yang diajukan. Model yang diajukan adalah model dengan basis VGG16 yang dimodifikasi, sehingga diharapkan penelitian ini ada peningkatan akurasi dari penelitian sebelumnya dalam facial expression recognition menggunakan algoritma CNN VGG16 dan dataset FER2013.

2. Dasar Teori

2.1. Facial Expression Recognition

Facial Expression Recognition (FER) atau Pengenalan Ekspresi Wajah merupakan teknik atau metode untuk mengenali emosi manusia yang dilihat dari mimik wajah manusia. FER terdiri dari tiga langkah utama, yaitu pre-processing, ekstraksi fitur, dan klasifikasi. Namun, FER pada pendekatan Deep Learning mengurangi ketergantungan pada pre-processing gambar dan ekstraksi fitur dengan menerapkan pembelajaran langsung end to end mulai dari data masukan (input) hingga hasil klasifikasi [2]. Pada FER, terdapat enam emosi dasar manusia sebagai klasifikasi utama pada sistem, diantaranya yaitu bahagia, terkejut, sedih, marah, jijik, dan takut [3]. Umumnya dataset terkait FER diberi label dengan enam emosi dasar ini.



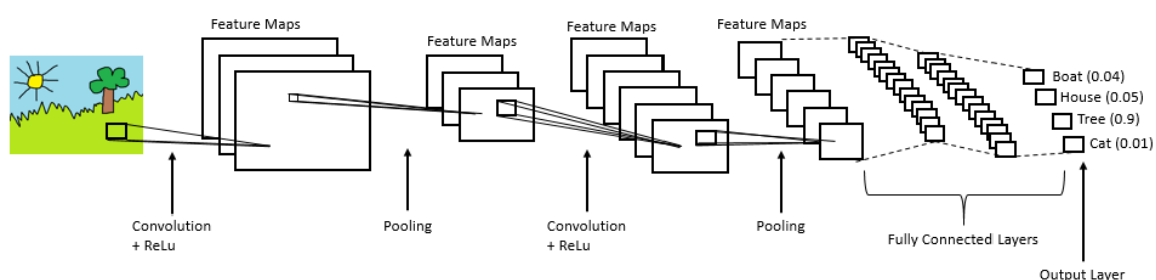
Gambar 1 Enam emosi dasar [4].

2.2. Deep Learning

Deep Learning merupakan bentuk dari Machine Learning yang bertujuan untuk memodelkan data abstrak tingkat tinggi dengan menggunakan arsitektur deep network yang terdiri dari beberapa transformasi linier atau non-linier [5]. Deep learning merupakan sistem cerdas yang meniru cara kerja otak manusia dalam merepresentasikan data kompleks dari skenario dunia nyata dan membantu dalam membuat keputusan cerdas. Deep learning dapat diterapkan dalam berbagai penelitian seperti pemodelan grafis, pengenalan pola, pemrosesan sinyal, computer vision, face recognition, dan lain-lain.

2.3. Convolutional Neural Network (CNN)

Convolutional Neural Network (CNN) merupakan salah satu algoritma pada deep learning. CNN adalah tipe deep neural network yang biasa digunakan pada data dua dimensi atau citra. CNN juga memiliki arsitektur dimana terdapat fitur ekstraksi dan klasifikasi yang digunakan untuk melatih data.

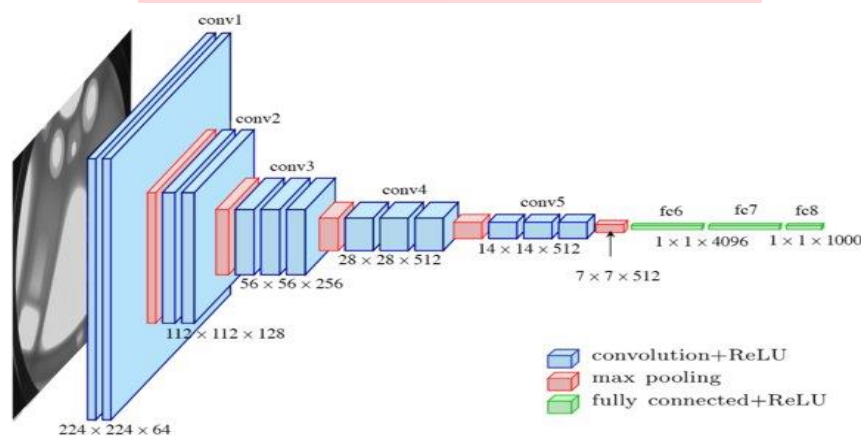


Gambar 2 Contoh arsitektur CNN [6].

CNN terdiri dari tiga jenis layer yaitu layer convolutional, layer pooling, dan layer fully connected. Layer convolutional memiliki sekumpulan filter yang dapat menggabungkan seluruh input gambar dan menghasilkan berbagai jenis feature map. Feature map adalah sebuah output yang berupa map yang dihasilkan oleh proses konvolusi. Layer pooling mengikuti layer convolutional dan digunakan untuk mengurangi ukuran spasial feature map dan beban pada komputasi jaringan. Average pooling dan max pooling adalah dua strategi pengambilan downsampling nonlinier yang paling umum digunakan untuk menerjemahkan invarian. Layer fully connected (FC) umumnya digunakan di akhir jaringan untuk memastikan bahwa semua neuron di layer tersebut sepenuhnya terhubung ke activation di layer sebelumnya dan untuk mengaktifkan feature map 2D untuk diubah menjadi feature map 1D untuk fitur representasi dan klasifikasi lanjutan [7].

2.4. Visual Geometry Group-16 Weight Layer (VGG16)

VGG16 merupakan model CNN dari Karen Simonyan dan Andrew Zisserman [8]. Model ini menunjukkan bahwa deep networks merupakan komponen penting untuk kinerja yang baik. Arsitektur VGG16 memiliki total 16 weight layer yang terdiri dari Convolutional layer dan FC. Keunggulan model ini menampilkan arsitektur yang sangat homogen yang hanya melakukan convolutional 3x3 dan pooling 2x2 dari awal hingga akhir. Kelemahan dari VGG16 adalah lebih berat untuk mengevaluasi dan menggunakan lebih banyak memori dan parameter mencapai 138 juta.



Gambar 3 Arsitektur VGG16 State-of-the-Art [9].

2.5. Konfusi Matriks (Confusion Matrix)

Perhitungan nilai akurasi pada pembelajaran mesin biasa disebut dengan konfusi matriks (*Confusion Matrix*). Pada dasarnya, konfusi matriks memberikan informasi perbandingan hasil klasifikasi yang dilakukan oleh sistem dalam bentuk model dengan hasil klasifikasi sebenarnya. Pada bagian ini ada beberapa matriks performansi yang umum dan sering digunakan, yaitu *accuracy*, *precision*, dan *recall*. Berikut merupakan persamaan untuk mencari nilai *accuracy*, *precision*, dan *recall* [10].

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (1)$$

$$Precision = \frac{TP}{TP+FP} \quad (2)$$

$$Recall = \frac{TP}{TP+FN} \quad (3)$$

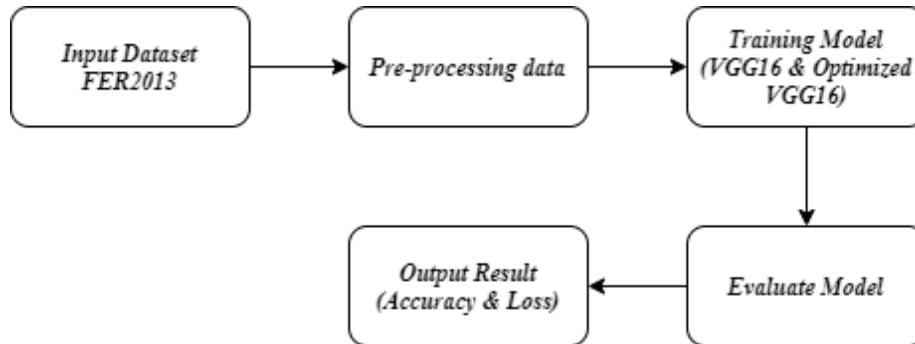
Penjelasan:

1. *True Positive* (TP), nilai pada hasil model benar dalam memprediksi kelas positif.
2. *True Negative* (TN), nilai pada hasil model benar memprediksi kelas negatif
3. *False Positive* (FP), hasil model salah dalam memprediksi kelas positif.
4. *False Negative* (FN), hasil model salah dalam memprediksi kelas negatif.

3. Perancangan Sistem

3.1. Desain Sistem

Secara umum sistem yang akan dibangun adalah sebuah sistem pengenalan ekspresi wajah yang menggunakan dataset FER2013 dengan model dasarnya yaitu model arsitektur VGG16 berbasis Convolutional Neural Network (CNN). Pada penelitian ini pengenalan ekspresi wajah akan berfokus pada akurasi yang diperoleh dari model yang dibuat. Gambaran sistem yang dibuat untuk penelitian ini dapat dilihat pada gambar 3.1.



Gambar 4 Gambaran umum sistem.

Sistem yang dirancang menggunakan input data untuk memasukkan dataset yang digunakan. Setelah data diinput terlebih dahulu dilakukan proses pre-processing data untuk mengolah data sedemikian rupa sehingga hasil data yang telah diproses pada pre-processing dilanjutkan ke dalam proses training. Hasil data yang diproses pada pre-processing menjadi tiga. Pada proses training, model dilatih untuk melakukan pengenalan ekspresi wajah pada data latih dan data validasi. Model yang telah dilatih kemudian dievaluasi untuk menampilkan akurasi yang didapat pada masing-masing data latih, data validasi, dan data uji menggunakan model yang telah dilatih.

3.2. Dataset yang digunakan

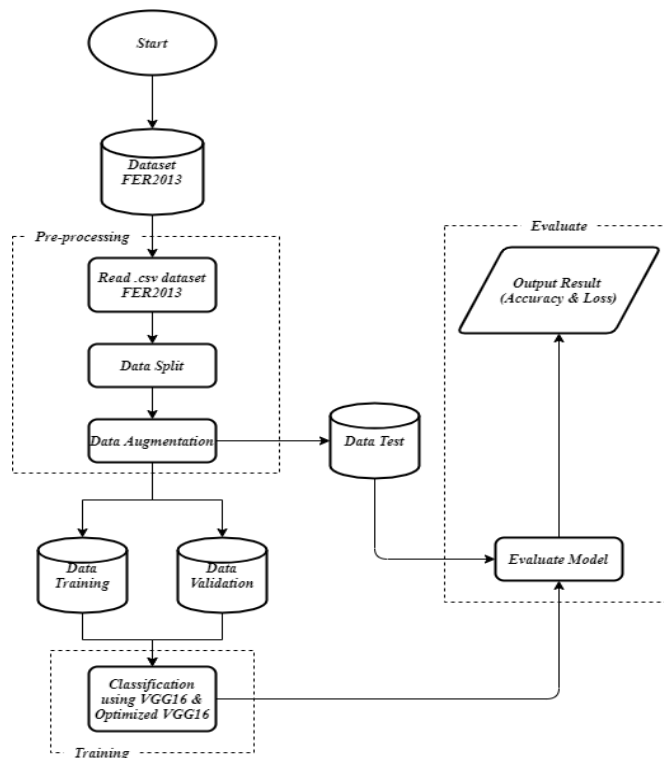
Dataset yang digunakan pada sistem ini yaitu FER2013. Dataset ini memiliki 35,887 citra yang berukuran 48x48 piksel dengan warna grayscale. Dataset FER2013 tersedia dalam file dengan format .csv dimana file tersebut berisi kolom emotion, pixels, dan usage. Kolom emotion, berisi nilai kelas yang memiliki range nilai dari 0 sampai 6 masing-masing mewakili emosi wajah (0 = marah, 1 = jijik, 2 = takut, 3 = senang, 4 = sedih, 5 = terkejut, 6 = netral). Pada kolom pixels, berisi matriks citra berukuran 48x48 dengan nilai warna dalam format 0-255 (Grayscale). Pada kolom Usage, berisi penamaan yang ditujukan untuk beberapa citra berdasarkan pemakaiannya (Training, PublicTest, PrivateTest). Pada penelitian ini, berdasarkan kategori Usage pada dataset FER2013 data Training digunakan untuk data latih, data PublicTest digunakan untuk data validasi, dan data PrivateTest digunakan untuk data uji.



Gambar 5 Tujuh kelas emosi pada dataset FER2013.

3.3. Rancangan Sistem

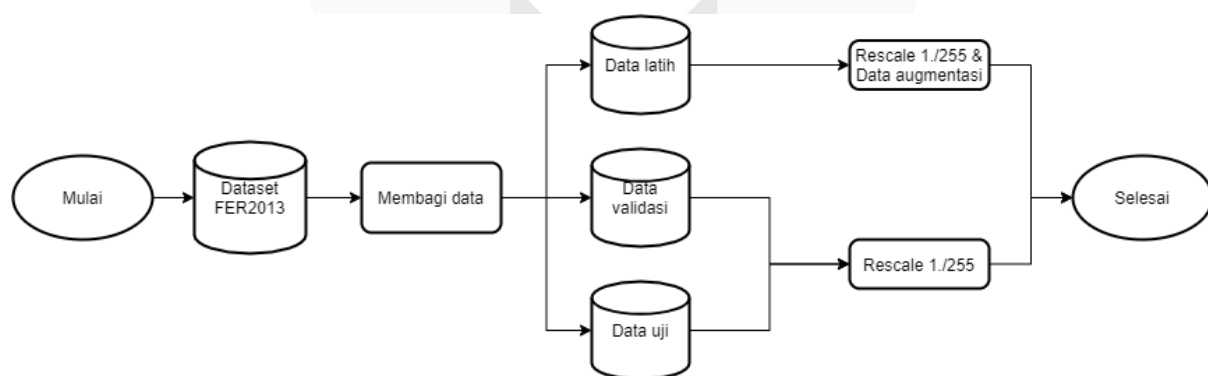
Dalam penelitian ini, perancangan sistem diperlukan sebelum sistem ini dibangun. Dari rancangan sistem ini, dapat diidentifikasi bahwa terdapat beberapa proses yang harus dilalui untuk mencapai tujuan. Berikut merupakan rancangan sistem yang akan dibangun pada tugas akhir ini dalam bentuk Flowchart.



Gambar 6 Flowchart rancangan sistem.

Dari rancangan sistem pada gambar 6, sistem yang akan dibangun terdiri dari beberapa bagian yaitu Pre-processing, Training, dan Evaluate. Dari tiga bagian tersebut masing-masing terdapat beberapa proses.

Pada bagian pre-processing, pertama dilakukan pembacaan dataset FER2013 yang masih dalam format .csv dimana terdapat piksel-piksel yang kemudian diolah kedalam bentuk citra dengan masing-masing kategori kelas emosi dan usage. Pada pembagian atau split data, dataset dikategorikan menjadi tiga bagian berdasarkan usage yaitu data Training dikategorikan ke data latih, PublicTest dikategorikan menjadi data validasi, dan PrivateTest dikategorikan menjadi data uji. Rasio pembagian dataset ini adalah 80% data latih, 10% data validasi, dan 10% data uji. Pada penelitian ini, data latih dan data validasi digunakan untuk training model, sedangkan data uji hanya digunakan untuk evaluasi model.



Gambar 7 Flowchart Pre-processing

Setelah melalui proses split data, dilakukan augmentasi data untuk meningkatkan performa yang lebih optimal. Dalam tugas akhir ini, penulis menggunakan pustaka Tensorflow dengan API Keras dimana modul ImageDataGenerator digunakan untuk menentukan augmentasi apa saja yang akan diterapkan pada dataset. Berikut merupakan augmentasi yang digunakan:

```

datagen = ImageDataGenerator(rescale=1./255,
                             rotation_range=10,
                             width_shift_range=0.2,
                             height_shift_range=0.2,
                             zoom_range=0.2,
                             horizontal_flip=True)

testgen = ImageDataGenerator(rescale=1./255)

```

Gambar 8 Augmentasi data.

Augmentasi pada variabel datagen hanya digunakan pada data latih. Sedangkan augmentasi pada variabel testgen digunakan pada data validasi dan data uji, hal ini dikarenakan semua data perlu dinormalisasi sehingga menghindari error ketika melakukan klasifikasi.

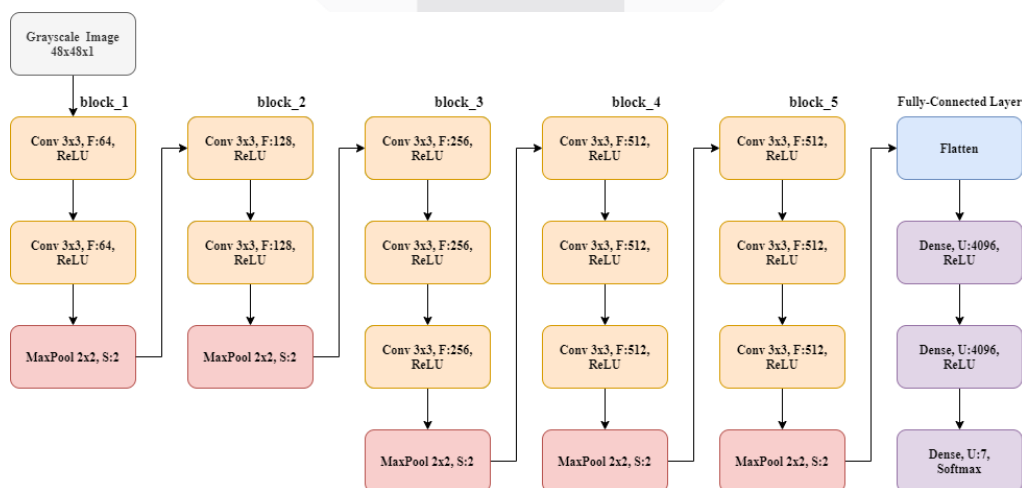
Pada tahap training, model dibangun terlebih dahulu untuk ekstraksi fitur dan klasifikasi pada proses training. Dalam penelitian ini, menggunakan algoritma CNN dengan model arsitektur VGG16 dan menggunakan model yang diajukan dimana model tersebut berbasis VGG16 yang dimodifikasi (Modified VGG16). Dua model tersebut tanpa menggunakan Transfer Learning maupun Pretrained. Model yang dibangun menggunakan pustaka Tensorflow dengan API Keras untuk menyusun layer neural network yang digunakan pada arsitektur VGG16. Setelah model dibangun, dilakukan compile model untuk dilanjutkan ke proses training. Pada compile model, terdapat optimizer dan adanya parameter learning rate pada optimizer tersebut. Dalam penelitian ini, optimizer yang digunakan yaitu Adam dan RMSprop.

Setelah model dicompile, dilakukan proses training model yang telah dibangun dan data yang digunakan yaitu data latih dan data validasi. Terdapat parameter yang digunakan untuk melakukan proses training, yaitu epoch dan learning rate. Epoch merupakan salah satu parameter training untuk mewakili jumlah iterasi yang harus dilakukan pada set data, dan learning rate merupakan salah satu parameter training untuk menghitung nilai koreksi bobot pada waktu proses training.

Setelah training model, model tersebut akan dievaluasi untuk menguji akurasi yang didapat dari model yang sudah dilatih pada proses training sebelumnya. Pengujian dilakukan dengan menggunakan data uji yang sebelumnya telah melalui proses preprocessing. Hasil yang ditampilkan berupa nilai akurasi dan loss dari model yang dievaluasi.

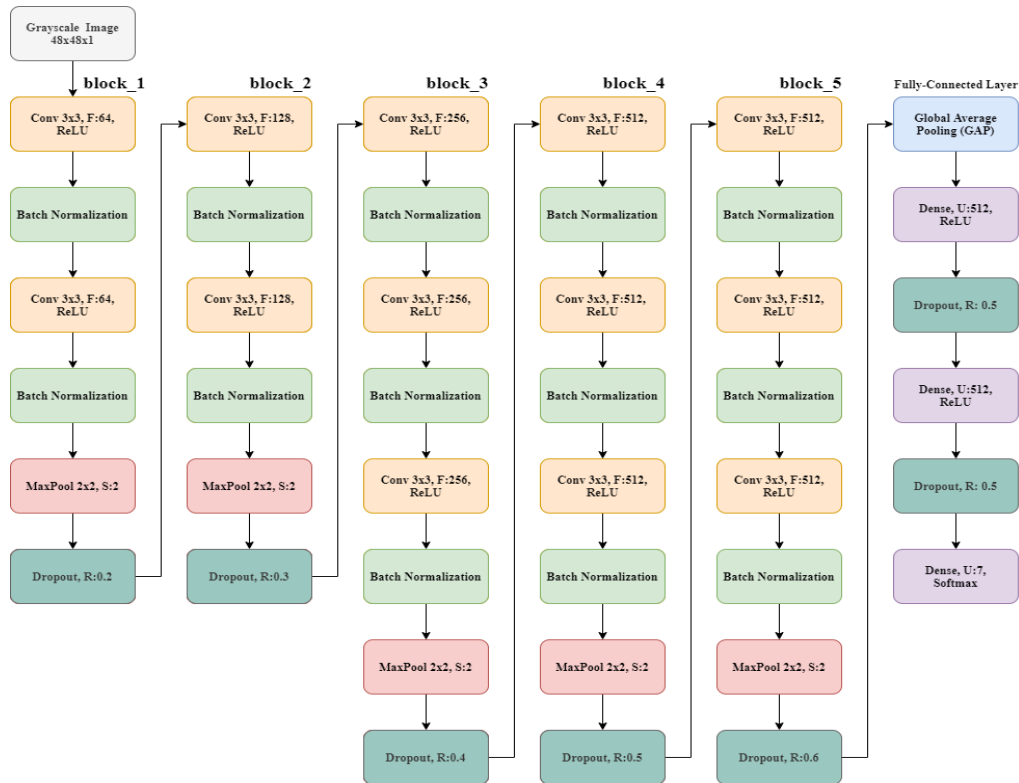
3.4. Model yang diajukan

Pada penelitian ini, model VGG16 dan VGG16 yang dimodifikasi digunakan sebagai ekstraksi fitur dan klasifikasi pada sistem pengenalan ekspresi wajah.



Gambar 9 Arsitektur VGG16.

Pada gambar 9 arsitektur ini dasarnya sama dengan model VGG16 pada state-of-the-art, namun hanya ada sedikit perubahan diantaranya ukuran input diubah menjadi 48x48x1 (grayscale) untuk menyesuaikan dataset FER2013, dan unit pada klasifikasi berjumlah 7 sesuai dengan jumlah kelas emosi pada dataset FER2013.



Gambar 10 Arsitektur Modified VGG16.

Pada gambar 10 merupakan arsitektur dari model yang diajukan, arsitektur ini menggunakan model dasar VGG16 namun dengan perubahan-perubahan diantaranya terdapat tambahan layer Dropout dan layer Batch Normalization pada, Global Average Pooling (GAP) menggantikan layer Flatten, dan perubahan parameter unit pada layer fully connected. Perubahan ini bertujuan untuk meningkatkan performansi saat melakukan training.

4. Hasil Pengujian Dan Analisis

4.1. Skenario Pengujian

Pengujian dilakukan dengan menggunakan satu nilai dari tiap parameter untuk sekali operasi pada sistem, parameter tersebut digunakan pada sistem yang keluarannya berupa akurasi dari model yang digunakan dengan parameter tertentu. Berikut merupakan tabel parameter yang digunakan pada pengujian ini.

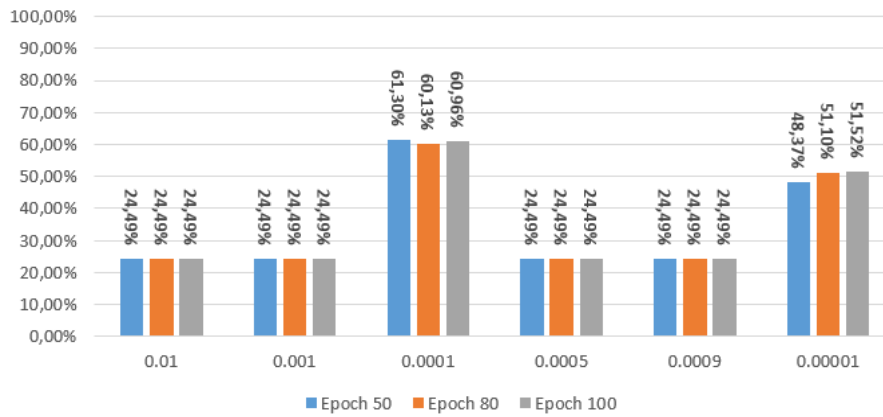
Tabel 1 Parameter dan Nilai

Parameter	Nilai
Epoch	50, 80, 100
Optimizer	Adam, RMSprop
Learning Rate	0.01, 0.001, 0.0001, 0.0005, 0.0009
Augmentasi	Dengan augmentasi, tanpa augmentasi

4.2. Hasil Pengujian

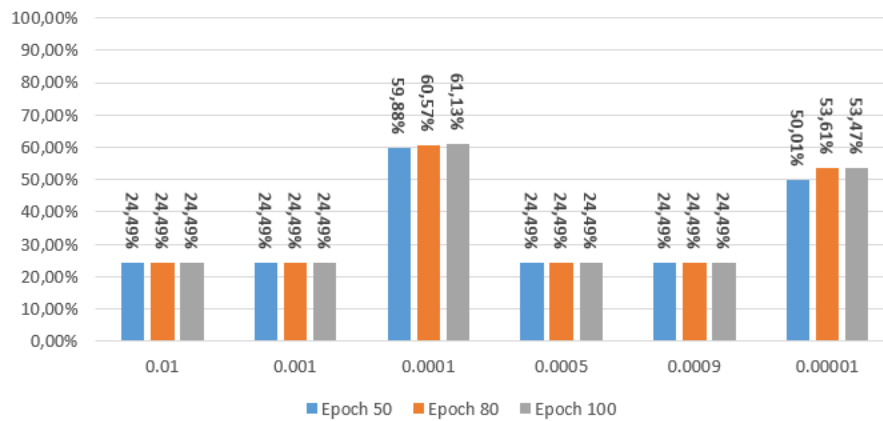
Dari semua pengujian yang dilakukan pada sistem yang telah dibangun, hasil yang diperoleh merupakan akurasi dari model yang digunakan. Berikut merupakan seluruh hasil akurasi pada data uji dari masing-masing model dengan parameter-parameter tertentu.

Test Accuracy in VGG16 using Adam Optimizer & Without Augmentation



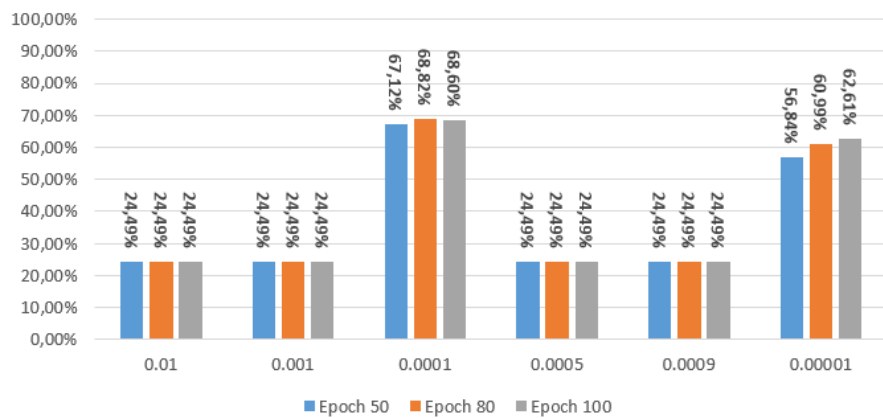
Gambar 11 Non-augmentation VGG16 Adam optimizer

Test Accuracy in VGG16 using RMSprop Optimizer & Without Augmentation



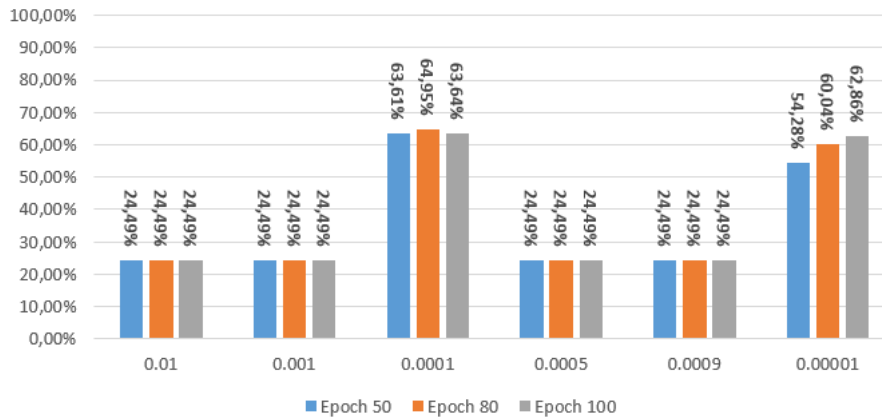
Gambar 12 Non-augmentation VGG16 RMSprop optimizer

Test Accuracy in VGG16 using Adam Optimizer & With Augmentation



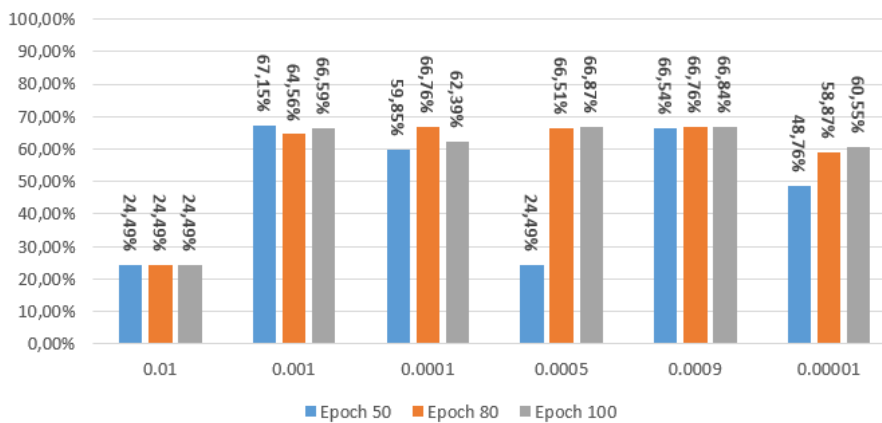
Gambar 13 Augmentation VGG16 Adam optimizer

Test Accuracy in VGG16 using RMSprop Optimizer & With Augmentation



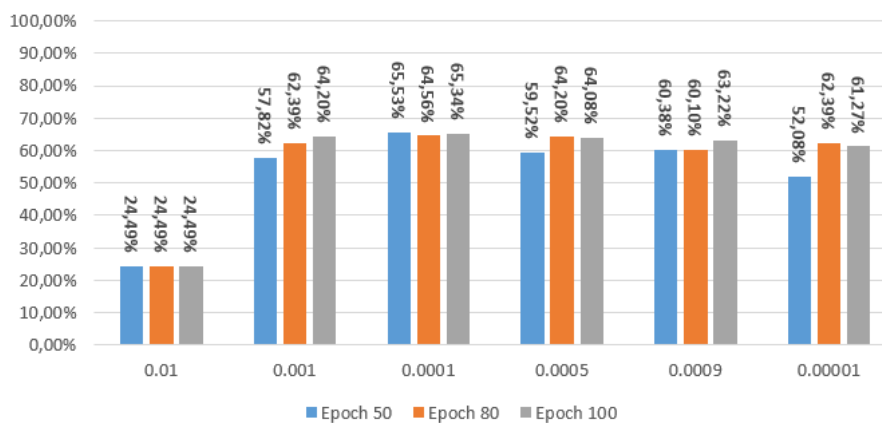
Gambar 14 Augmentation VGG16 RMSprop optimizer

Test Accuracy in Modified VGG16 using Adam Optimizer & Without Augmentation

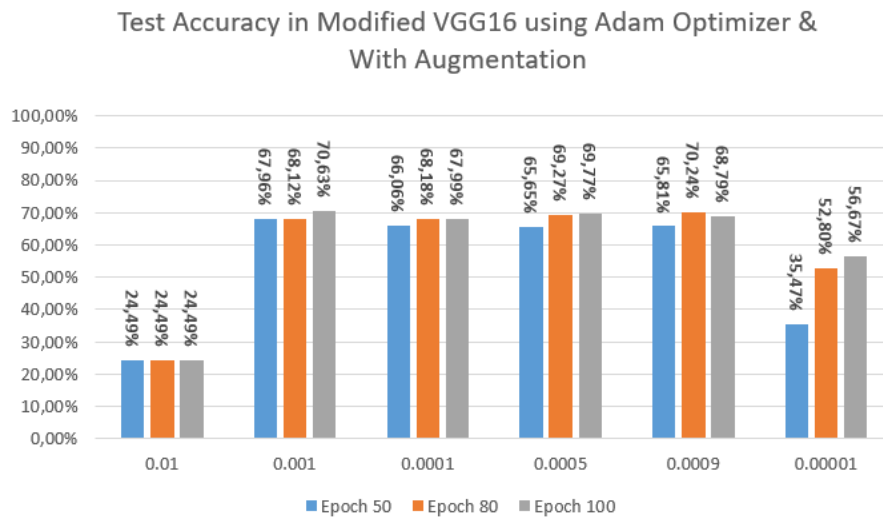


Gambar 15 Non-augmentation modified VGG16 Adam optimizer

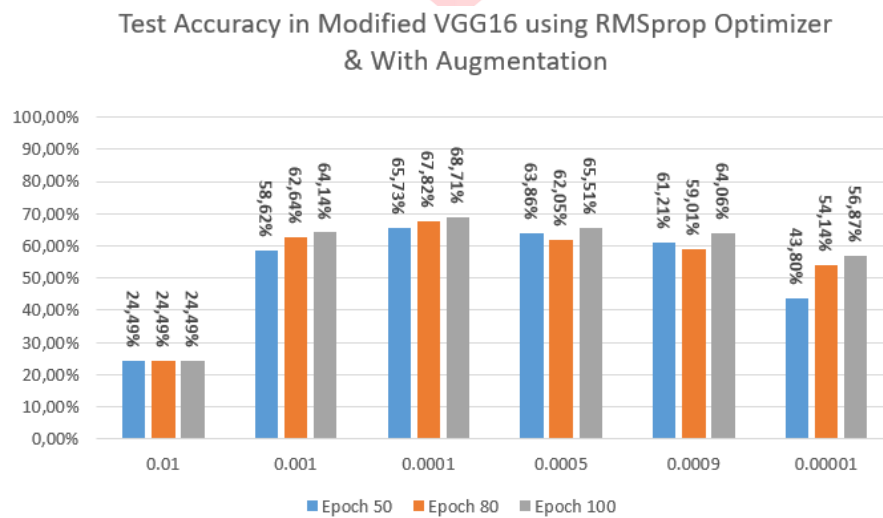
Test Accuracy in Modified VGG16 using RMSprop Optimizer & Without Augmentation



Gambar 16 Non-augmentation modified VGG16 RMSprop optimizer



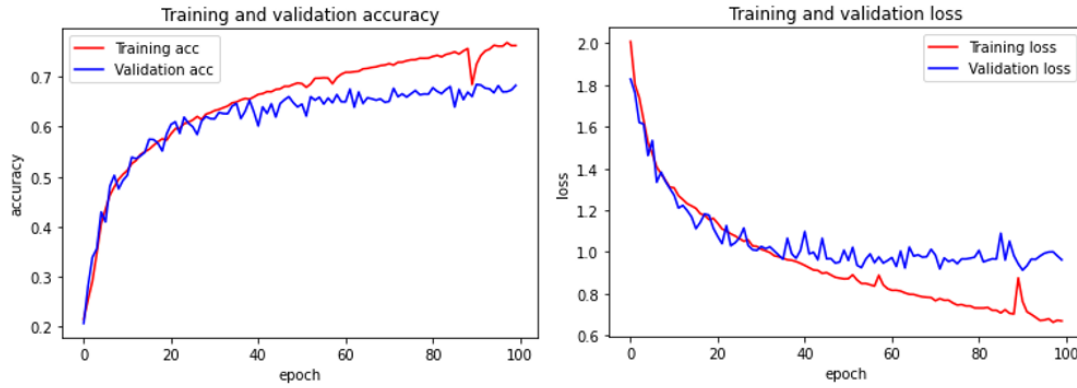
Gambar 17 Augmentation modified VGG16 Adam optimizer



Gambar 18 Augmentation modified VGG16 RMSprop optimizer

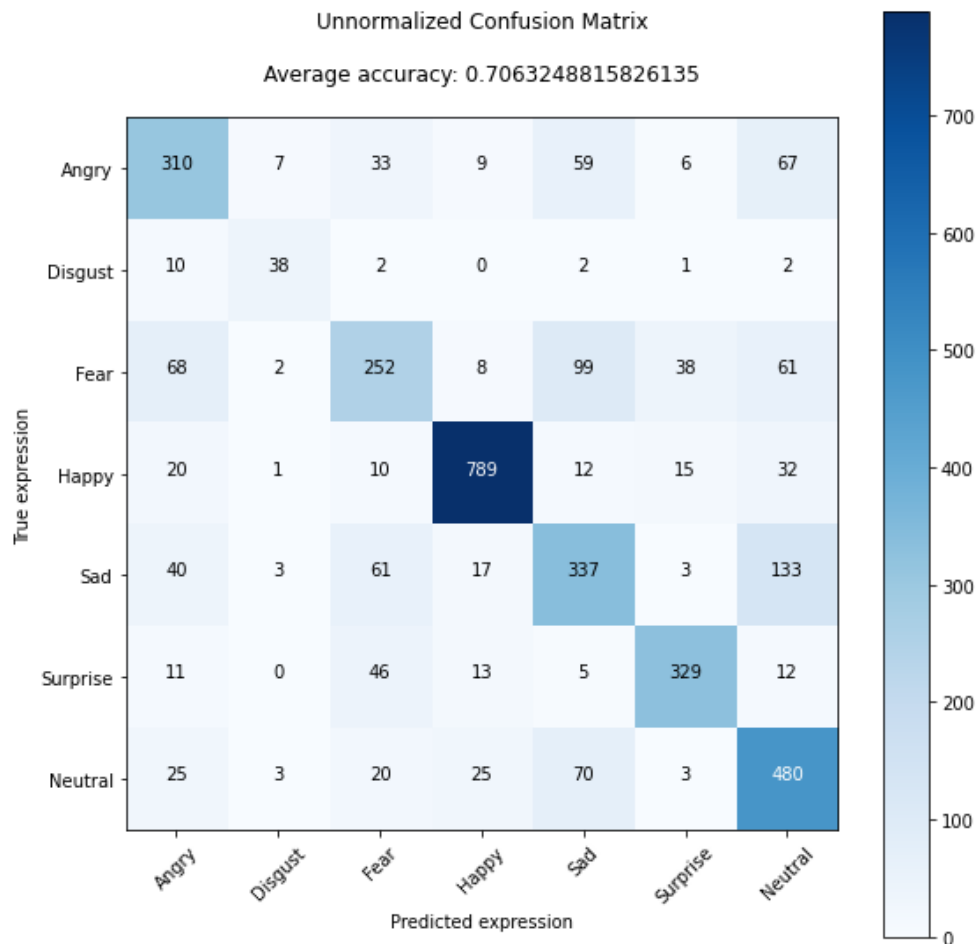
4.3. Analisis hasil pengujian

Pada hasil pengujian, model Modified VGG16 memperoleh akurasi lebih baik dibandingkan model VGG16. Pada hasil pengujian yang terbaik dari yang telah diuji coba baik dari pengujian parameter dan augmentasi hasil yang optimum/terbaik adalah model Modified VGG16 dengan augmentasi dengan parameter epoch 100, learning rate 0.001, dan menggunakan optimizer Adam.



Gambar 19 Grafik akurasi latih dan validasi pada parameter terbaik.

Pada gambar 19 merupakan grafik dari hasil proses training dimana jarak antara akurasi latih dan akurasi validasi tidak jauh berbeda yaitu akurasi latih yang diperoleh 76,71% dan akurasi validasi yang diperoleh 68,28%, untuk itu hasil ini dianggap tidak mengalami overfitting maupun underfitting. Pada hasil evaluasi didapatkan masing-masing hasil akurasi yaitu akurasi latih sebesar 80,88%, validasi sebesar 68,26%, dan uji sebesar 70,63%.



Gambar 20 Matriks konfusi pada parameter terbaik

Pada gambar 20 menunjukkan bahwa label kelas terbaik yaitu kelas emosi senang dimana memiliki nilai yang cukup tinggi. Sedangkan nilai kelas terendah yaitu pada kelas emosi jijik, hal ini disebabkan karena jumlah dataset untuk kelas emosi jijik sangat sedikit.

Tabel 2 Precision, Recall, dan Akurasi.

Emosi	Precision	Recall
Marah	$\frac{310}{(310 + 484)} = 64\%$	$\frac{310}{(310 + 67)} = 63\%$
Jijik	$\frac{38}{(38 + 54)} = 70\%$	$\frac{38}{(38 + 55)} = 69\%$
Takut	$\frac{252}{(252 + 424)} = 59\%$	$\frac{252}{(252 + 528)} = 48\%$
Bahagia	$\frac{789}{(789 + 861)} = 92\%$	$\frac{789}{(789 + 879)} = 90\%$
Sedih	$\frac{337}{(337 + 584)} = 58\%$	$\frac{337}{(337 + 594)} = 57\%$
Terkejut	$\frac{329}{(329 + 395)} = 83\%$	$\frac{329}{(329 + 745)} = 44\%$
Netral	$\frac{480}{(480 + 626)} = 77\%$	$\frac{310}{(310 + 787)} = 61\%$
Akurasi	$\frac{2535}{3589} = 70,63\%$	

Pada tabel diatas, perhitungan precision, recall, dan akurasi dari tiap emosi dihitung secara manual untuk mencari hasil akurasi keseluruhan. Terbukti dari perhitungan tersebut bahwa hasil akurasi dari perhitungan manual dengan hasil akurasi dari sistem yang dirancang memperoleh kesamaan.

Tabel 3 Perbandingan akurasi dengan penelitian sebelumnya.

Penelitian	Model	Dataset	Test Accuracy
Pavan <i>et al.</i> [1]	VGG16	FER2013	60%
Yin <i>et al.</i> [11]	VGG16 (Pre-trained Imagenet)		68,24%
Amil <i>et al.</i> [12]	Fine Tuning VGG16 (Pre-trained VGG-Face)		70,2%
Model yang diajukan	VGG16		68,82%
	Modified VGG16	70,63%	

Pada tabel perbandingan diatas, hasil akurasi uji yang diperoleh tiga penelitian sebelumnya dijadikan pembandingan dengan hasil akurasi uji menggunakan model yang diajukan. Sama seperti pada tugas akhir ini, tiga penelitian tersebut menggunakan model basis VGG16 dan dataset FER2013.

Pada penelitian milik Pavan et al, akurasi uji yang diperoleh mencapai 60% menggunakan model VGG16. Pada penelitian milik Yin et al, akurasi uji yang diperoleh mencapai 68,24% menggunakan model VGG16 dengan Pre-trained Imagenet. Pada penelitian milik Amil et al, akurasi uji yang diperoleh mencapai 70,2% menggunakan model Fine Tuning VGG16 dengan Pre-trained VGG-Face. Pada tugas akhir ini menggunakan dua model dengan akurasi uji terbaik masing-masing mencapai 67,12% pada model VGG16 dan 70,63% pada model Modified VGG16.

Dari perbandingan akurasi uji tersebut, hasil yang terbaik merupakan model yang penulis ajukan yaitu Modified VGG16 dimana akurasi uji yang diperoleh mencapai 70,63% dengan parameter sebagai berikut; Menggunakan data augmentasi, epoch 100, optimizer Adam, dan learning rate 0,001.

5. Kesimpulan

1. Sistem yang telah dibangun dapat melakukan pengenalan ekspresi wajah dimana terdapat tiga bagian yaitu preprocessing, training, dan evaluate. Preprocessing digunakan untuk menyiapkan data yang akan dilatih pada proses training, data masukannya adalah dataset FER2013. Proses training dapat membuat dan melatih model berbasis CNN dengan model arsitektur VGG16. Pada proses Evaluate, sistem dapat melakukan proses pengujian pada model yang telah dilatih untuk menghasilkan keluaran berupa akurasi, loss, dan konfusi matriks.
2. Hasil akurasi dari keluaran sistem yaitu pada model VGG16 mencapai akurasi 83,95% pada data latih, 68,63% pada data validasi, dan 68,82% pada data uji. Pada model Modified VGG16 menghasilkan akurasi lebih baik dari model VGG16 yaitu mencapai 80,88% pada data latih, 68,26% pada data uji, dan 70,63% pada data uji
3. Akurasi pada data uji digunakan untuk membandingkan akurasi pada penelitian sebelumnya. Pada tugas akhir ini, akurasi terbaik yaitu menggunakan model modified VGG16 yang mencapai 70,63% dimana akurasi ini lebih unggul 0,43% dari milik Amil et al dengan akurasi 70,2%, lebih unggul 2,39% dari milik Yin et al dengan akurasi 68,24%, dan lebih unggul 10,63% dari milik Pavan et al.

Referensi

- [1] P. Bodavarapu and P. Srinivas, "Facial expression recognition for low resolution images using convolutional neural networks and denoising techniques," *Indian J. Sci. Technol.*, vol. 14, pp. 971–983, 2021, doi: 10.17485/IJST/v14i12.14.
- [2] Y. Huang, F. Chen, S. Lv, and X. Wang, "Facial Expression Recognition: A Survey," *Symmetry (Basel)*, vol. 11, p. 1189, 2019, doi: 10.3390/sym11101189.
- [3] P. Ekman, "An argument for basic emotions," *Cogn. Emot.*, vol. 6, pp. 169–200, 1992.
- [4] A. Kołakowska, A. Landowska, M. Szwoch, W. Szwoch, and M. Wróbel, "Modeling emotions for affect-aware applications," in *Information Systems Development and Applications*, Faculty of Management University of Gdańsk, 2015, pp. 55–67.
- [5] S. Almabdy and L. Elrefaei, "Deep Convolutional Neural Network-Based Approaches for Face Recognition," *Appl. Sci.*, vol. 9, p. 4397, 2019, doi: 10.3390/app9204397.
- [6] A. F. Yaseen, "A Survey on the Layers of Convolutional Neural Networks," *Int. J. Comput. Sci. Mob.*, vol. 7, no. 12, pp. 191–196, 2018.
- [7] S. Li and W. Deng, "Deep Facial Expression Recognition: A Survey," *IEEE Trans. Affect. Comput.*, vol. PP, 2018, doi: 10.1109/TAFFC.2020.2981446.
- [8] K. Simonyan and A. Zisserman, "Very Deep Convolutional Networks for Large-Scale Image Recognition," *ICLR 2015*, 2015.
- [9] M. Ferguson, R. Ak, Y.-T. Lee, and K. Law, "Automatic localization of casting defects with convolutional neural networks," in *Conference: IEEE International Conference on Big Data (Big Data)*, 2017, pp. 1726–1735, doi: 10.1109/BigData.2017.8258115.
- [10] S. Ghoneim, "Accuracy, Recall, Precision, F-Score & Specificity, which to optimize on?," *Towards Data Science*, 2019. <https://towardsdatascience.com/accuracy-recall-precision-f-score-specificity-which-to-optimize-on-867d3f11124> (accessed Jul. 25, 2021).
- [11] Y. Fan, X. Lu, D. Li, and Y. Liu, "Video-Based Emotion Recognition Using CNN-RNN and C3D Hybrid Networks," in *Proceedings of the 18th ACM International Conference on Multimodal Interaction*, 2016, pp. 445–450, doi: 10.1145/2993148.2997632.
- [12] A. Khanzada, C. Bai, and F. T. Celepcikay, "Facial Expression Recognition with Deep Learning," *CoRR*, vol. abs/2004.1, 2020, [Online]. Available: <https://arxiv.org/abs/2004.11823>.