

Implementasi *Convolutional Neural Network* untuk Klasifikasi Ekspresi Wajah Penonton *Stand Up Comedy*

¹Agus Kurnia Akbar, ²Pringgo Widyo Laksono, ³Wakhid Ahmad Jauhari

^{1,2,3} Universitas Sebelas Maret, Surakarta, Jawa Tengah, Indonesia

Email Korespondensi: aguskurnia123.ak@student.uns.ac.id

DOI: <https://doi.org/10.70115/ijsta.v2i2.215>

Article Info

Article history:

Received : Oct 29, 2024

Revised : Nov 27, 2024

Accepted : Dec 31, 2024

Keywords:

Red Neuronal

Convocional; Introduction to Facial Expressions,

Crowd of Spectators; Stand Up Comedy

ABSTRACT

Stand Up Comedy is one of the growing entertainment industries in Indonesia. The quality of the joke of the stand up comedian can be seen from the audience's response in the form of laughter and facial expressions. But it is very difficult to determine the audience's laughter level manually. Therefore, a tool is needed to measure the audience's laughter level more objectively. In this study, a machine learning model was designed and developed using a convolutional neural network to analyze the facial expressions of the stand-up comedian audience. The research stages include dataset collection, data preprocessing, model design, model training and model testing. The model was tested using artificial data from crowd emotion and live data from video crowd viewers. The results of this study show that the model can classify facial expressions in image and video data well. The model has an overall validation accuracy value of 72.42% with precision, recall, and F1-score values of 0.7341, 0.7309, and 0.7294, respectively, and a testing accuracy of 83.28% with precision, recall, and F1-score values of 0.8477, 0.8329, and 0.8297, respectively.



This work is licensed under a Creative Commons Attribution-ShareAlike 4.0 International License.

@2024 AHS Publisher

PENDAHULUAN

Industri hiburan menjadi salah satu sektor yang selalu berkembang bersamaan dengan perkembangan zaman. Setiap tahun, industri hiburan dunia mengalami peningkatan yang signifikan. Dalam laporan PwC yang berjudul “*Global Entertainment & Media Outlook 2019-2023*”, industri media dan hiburan dunia dari tahun 2014-2018 mengalami peningkatan pendapatan sebanyak 24% menjadi Rp 29.440 triliun dan dalam lima tahun kedepan diproyeksikan dapat menyentuh angka Rp 36.450 triliun. (PricewaterhouseCoopers, 2023). Sejalan dengan industri hiburan dunia, perkembangan industri hiburan di Indonesia mengalami

peningkatan. Berdasarkan publikasi dari Kemenparekraf (2021), pada tahun 2019 sub sektor ekonomi kreatif menyumbang Rp 1.153,4 Triliun PDB atau 7,3% terhadap PDB Nasional. Menurut Kominfo (2022), terdapat lebih dari 8,2 juta jumlah usaha kreatif di Indonesia yang terdiri dari 4 sub sektor ekonomi kreatif dengan pertumbuhan tercepat yaitu film, animasi, video dan seni pertunjukan. Pertumbuhan keempat sub sektor ekonomi kreatif tersebut harus sejalan dengan kualitas dari industri hiburan tersebut. Tontonan merupakan suatu media yang sangat besar pengaruhnya terhadap kualitas kehidupan masyarakat Indonesia karena dapat memberikan informasi sekaligus hiburan yang dapat mempengaruhi pola pikir mereka. Oleh karena itu, kualitas dari suatu hiburan menjadi sangat penting. Kualitas dari suatu hiburan dapat dilihat dari segi isi konten yang dibahas dan tingkat kepuasan dari penikmat hiburan tersebut. Kepuasan penonton menjadikan target utama untuk peningkatan kualitas tontonan (Sudarmawan, 2020).

Stand Up Comedy merupakan sebuah seni pertunjukkan komedi yang dilakukan oleh seorang penampil (*one man show*) atau biasa disebut komika. Komika harus memperhatikan unsur *personality* atau karakter dan *direct communication* atau komunikasi dua arah penonton dan penampil dalam sebuah penampilan. *Stand up comedy* merupakan penampilan yang membutuhkan umpan balik secara *real time* (MacRury, 2012). Hubungan antara komika dengan penontonnya merupakan hubungan timbal balik dimana penonton akan merespon *joke* komedian dengan ekspresi yang diungkapkan. Keberhasilan dari *stand up comedian* adalah membawakan materi yang sudah dipersiapkan dan dapat memuaskan seluruh penontonnya. Kepuasan dari penonton dapat di analisis dari suara tawa yang dihasilkan dan juga ekspresi wajah selama menonton penampilan *stand up comedy*. Ekspresi wajah dari penonton dapat digunakan sebagai acuan untuk menilai apakah materi yang disampaikan oleh komika memuaskan keseluruhan penontonnya.

Wajah merupakan rangsangan visual multidimensi dan menyediakan berbagai informasi seperti invidu sebagai identitas, jenis kelamin, usia, ras, suasana hati, dan niat (Sriyati, Setyanto, & Luthfi, 2020). Suasana hati dan niat dapat diidentifikasi melalui ekspresi wajah. Ekspresi wajah adalah perubahan wajah dalam menanggapi keadaan emosi, niat, atau komunikasi sosial seseorang (Tian, Kanade, & Cohn, 2011). Seseorang dapat mengontrol ekspresi wajah mereka tetapi pada umumnya ekspresi wajah timbul secara alami akibat dari emosi dan perasaan orang tersebut. Ekspresi wajah sangat menarik untuk diteliti karena merupakan bentuk komunikasi non-verbal yang digunakan untuk menyampaikan pesan sosial dan menggambarkan suasana hati seseorang. Menurut Paul Ekman (1992), terdapat 6 emosi dasar yang dapat dikenali melalui ekspresi wajah yaitu bahagia (*happy*), sedih (*sad*), jijik (*disgusting*), terkejut (*surprise*), marah (*angry*). Dalam pengaplikasian analisis ekspresi wajah ditambahkan satu kondisi non-emosional dengan ekspresi datar (*neutral*).

Pengenalan ekspresi wajah dilakukan untuk mengetahui emosi apa yang sedang dirasakan oleh seseorang yang kemudian dapat menentukan solusi yang tepat. Pada tahun 2019 di Jepang, pengenalan ekspresi wajah diterapkan oleh Toyota ke dalam mobil mereka yang dapat mendeteksi emosi penumpangnya melalui kamera dalam *dashboard*. Pendeteksian emosi sangat penting dilakukan untuk mengetahui perasaan apa yang sedang dirasakan pengendara. Pendeteksian emosi kemudian dianalisa oleh AI yang tertanam dalam sistem tersebut kemudian akan menghasilkan rekomendasi-rekomendasi yang menjadi solusi untuk pengendara. Selain itu, pendeteksi emosi juga sangat penting untuk mengetahui suasana dari suatu kerumunan orang. *Crowd Monitoring* atau pemantau kerumunan bisa berguna untuk mengidentifikasi situasi bahaya dan darurat. Menurut Holder (2017), *crowd monitoring* dapat melakukan analisis berdasarkan emosi yang dihasilkan dari kerumunan orang dan mampu memperkirakan kemungkinan yang akan terjadi dari kerumunan tersebut. Namun tidak semua emosi bisa

terlihat jelas melalui ekspresi wajahnya. Dengan banyaknya penonton *stand up comedy* maka tidak mungkin untuk melihat dan menganalisis ekspresi wajah satu per satu. Diperlukan klasifikasi citra ekspresi wajah untuk mengetahui kategori emosi apa yang dirasakan melalui ekspresi wajahnya. Salah satu metode yang cocok digunakan untuk klasifikasi citra ekspresi wajah adalah menggunakan metode *Deep Learning* yaitu *Convolutional Neural Network*.

Deep learning merupakan salah satu bagian dari *Machine Learning* menggunakan metode *Artificial Neural Network* (ANN) yang dapat mempercepat proses pembelajaran di *Neural Networks* dalam penggunaan *multiple layer* (Hania, 2017). Penelitian mengenai deteksi emosi sudah dilakukan beberapa peneliti, dan menunjukkan bahwa metode *Deep Learning* yang memiliki tingkat keberhasilan yang signifikan dalam pengenalan citra wajah adalah dengan *Convolutional Neural Network* (CNN) (Fukushima, 1980). Salah satu bagian dari pembelajaran mendalam adalah *Convolutional Neural Network* (Parkavi, 2020). CNN merupakan salah satu jenis *Deep Neural Network* yang didesain untuk mengolah data dua dimensi yang banyak diaplikasikan ke data citra. CNN berusaha untuk meniru *visual cortex* manusia dalam pengenalan citra sehingga memiliki kemampuan mengolah informasi citra (Eka Putra W. S., 2016). CNN memiliki kemampuan kecepatan dalam proses *training data* serta tersedia pada *library Keras* sehingga tidak memerlukan pembuatan fitur ekstraksi secara manual (Kusdiananggalih & Rachmawati, 2021). Kumpulan data yang digunakan untuk proses pelatihan dan validasi termasuk IMED (Liliana, Basaruddin, & Oriza, 2018), FER2013 dan JAFFE. Sementara itu, proses pengujian menggunakan *dataset Self Built* dan *Video Crowd* penonton. CNN merupakan teknik *deep learning* yang paling banyak digunakan untuk ekstraksi fitur yang lebih baik dilakukan pada kumpulan data besar (Sahu & Dash, 2020). Pada kondisi tertentu, model CNN dapat mengalami *overfitting* yaitu model terlalu banyak mempelajari data pelatihan sehingga performanya sangat tinggi sedangkan pada data baru (*validation* dan *testing*) memiliki performa yang menurun. Dua cara utama dilakukan untuk mengurangi *overfitting* yaitu *data augmentation* dan penambahan *dropout layer* (Krizhevsky, Sutskever, & Hinton, 2017).

Facial Expression Recognition merupakan penerapan dan pemanfaatan dari *computer vision* yang dapat menganalisis ekspresi wajah seseorang atau kelompok baik dari gambar ataupun video untuk mengungkapkan informasi mengenai emosi seseorang. Dalam penelitian ini, *Facial Expression Recognition* dilakukan dengan menerapkan Arsitektur *Convolutional Neural Network* untuk mendeteksi ekspresi wajah penonton *stand up comedy*.

METODE

Material

Model *Face Expression Recognition* untuk analisis ekspresi penonton *stand-up comedy* dibangun dengan arsitektur *Convolutional Neural Network* (Amaanullah, Pasfica, Nugraha, Zein, & Adhinata, 2022). Proses pelatihan dan validasi model menggunakan *dataset* IMED, FER2013 dan JAFFE. Proses pengujian menggunakan *dataset Selt Built* dan *video crowd* penonton *stand-up comedy* dari *Channel Youtube* “Wendy Cagur”.

Indonesian Mixed Emotion Dataset (IMED) adalah *dataset* yang menyimpan data ekspresi wajah dari berbagai suku di Indonesia. *Dataset* ini mencakup 19 kategori emosi, seperti marah, bahagia, sedih, dan lainnya, yang dikumpulkan dari 6 subjek laki-laki dan 9 subjek perempuan asal Indonesia (Liliana, Basaruddin, & Oriza, 2018). Subjek *dataset* berusia antara 17 hingga 32 tahun dan berasal dari berbagai etnis, seperti Minang, Batak, Sunda, Jawa, serta Manado. *Dataset* ini mencakup 6 ekspresi emosi, satu ekspresi wajah netral, serta 12 emosi campuran dalam format video dan gambar. Resolusi data yang digunakan adalah 720 x 480 piksel dengan warna RGB dalam format JPG. *Dataset* ini dikembangkan oleh Dewi Yanti

Liliana sebagai bagian dari Proyek Penelitian Doktor di Fakultas Ilmu Komputer, Universitas Indonesia (<https://imed.cs.ui.ac.id>).



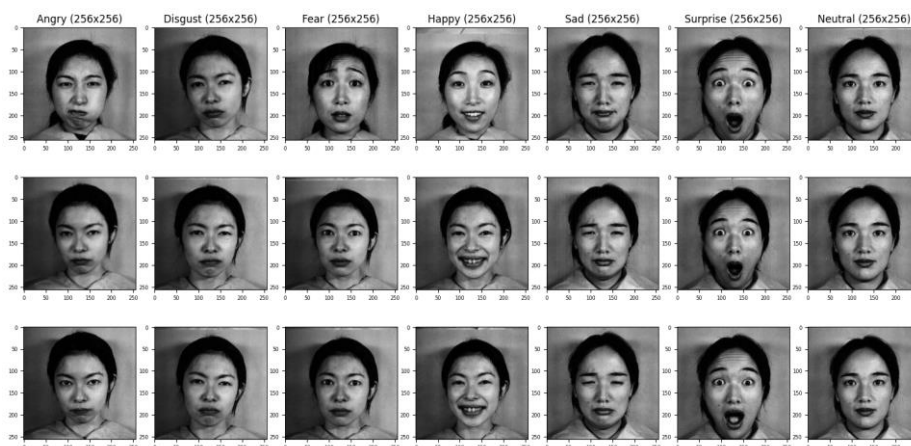
Gambar 1. Visualisasi *Dataset* IMED

Facial Expression Recognition (FER2013) adalah *dataset* global yang menyimpan data ekspresi wajah dari berbagai belahan dunia. *Dataset* ini terbagi menjadi data pelatihan (80%) dan validasi (20%) dengan total data yang mencakup 6 ekspresi emosi serta 1 ekspresi wajah netral. *Dataset* ini menggunakan resolusi 48 x 48 piksel dalam warna abu-abu dengan format JPG. Informasi mengenai jumlah jenis kelamin, usia, dan ras tidak tersedia. *Dataset* ini dapat diakses melalui Kaggle (<https://www.kaggle.com>).



Gambar 2. Visualisasi *Dataset* FER2013

Japanese Female Facial Expression (JAFFE) adalah *dataset* yang berisi ekspresi wajah perempuan Jepang. *Dataset* ini terdiri dari 213 gambar dari 10 subjek perempuan Jepang, dengan 7 kategori emosi, termasuk 6 ekspresi emosi dan 1 netral. *Dataset* ini menggunakan resolusi 256 x 256 piksel dalam warna abu-abu dengan format TIFF. Data ini tersedia di Kaggle (<https://www.kaggle.com>).



Gambar 3. Visualisasi *Dataset* JAFFE

Dataset Self-built digunakan dalam proses pengujian. *Dataset* ini diambil dari ekspresi wajah mahasiswa Teknik Industri UNS yang terdiri dari 15 subjek dengan 9 subjek laki-laki dan 6 subjek perempuan dengan usia rentan 20 hingga 25 tahun. Gambar diambil menggunakan Kamera Sony ILCE-6000. Gambar berukuran 4000 x 6000 piksel dalam warna sRGB dengan format JPG.



Gambar 4. Visualisasi *Dataset Self Built*

Berikut merupakan jumlah data setiap kelas ekspresi dari *dataset* yang sudah dikumpulkan pada penelitian ini yang dapat dilihat pada Tabel 1.

Tabel 1. Kuantitas *Dataset*

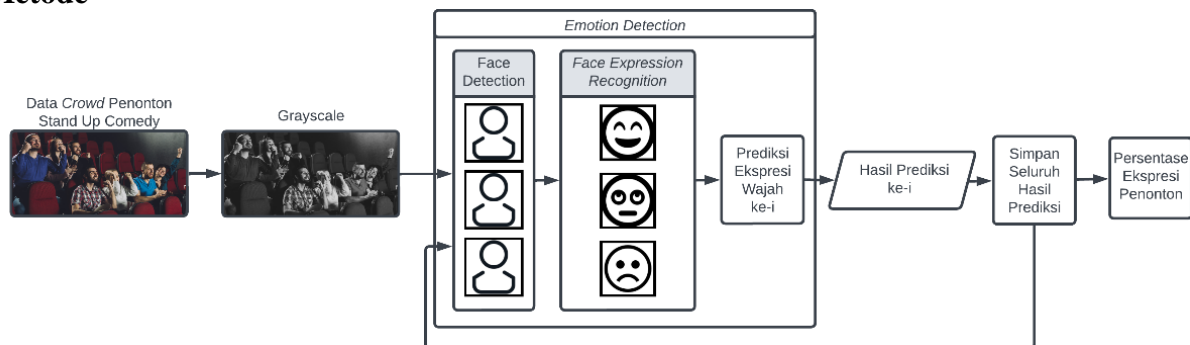
No	Emotion	Quantity				
		IMED	FER2013		JAFFE	Self Built
			Train	Validation		
1	Angry	1583	3993	960	30	45
2	Disgusting	1413	436	111	29	45
3	Fear	1466	4103	1018	32	45
4	Happy	1319	7164	1825	32	45
5	Neutral	480	4982	1216	30	45
6	Sad	1793	4938	1139	30	45
7	Surprise	950	3205	797	30	45

Data Video *Crowd* penonton *stand up comedy* dari Channel Youtube “Wendy Cagur” dalam acara “Lugas Season 2” digunakan untuk *testing* model. Cuplikan video penonton tertawa diambil dengan rentan waktu 6 detik dengan ukuran 1366 x 660 piksel dan *frame rate* 30 fps. Berikut merupakan *screenshot* dari video tersebut yang dapat dilihat pada Gambar 5.



Gambar 5. Data *Testing* Video Kerumunan Penonton *Stand Up Comedy*

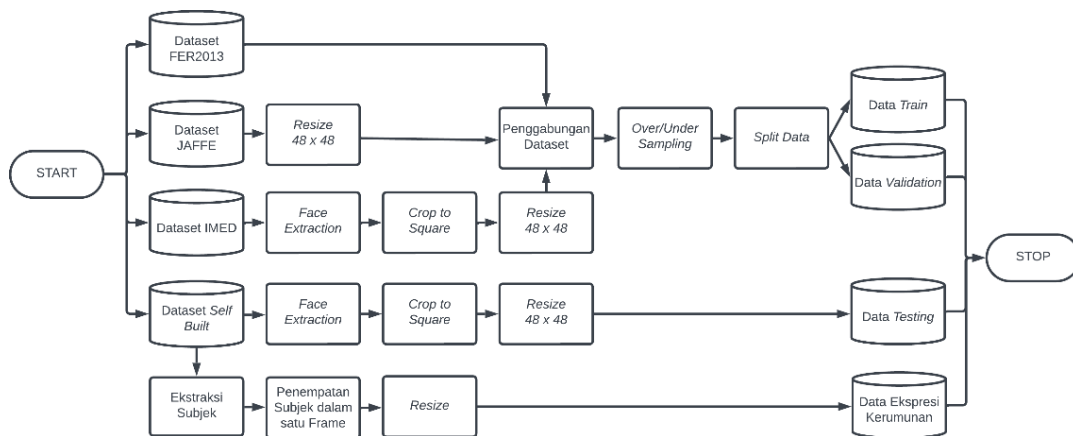
Metode



Gambar 6. Diagram Alur *Face Expression Recognition*

Model *Face Expression Recognition* yang dibangun dengan arsitektur *Convolutional Neural Network* ditujukan untuk menentukan ekspresi penonton *stand up comedy*. Model ini terdiri dari beberapa tahap utama yang dapat dilihat pada Gambar 6. Metode penelitian yang dilakukan untuk membangun model tersebut dijelaskan sebagai berikut.

Preprocessing Data

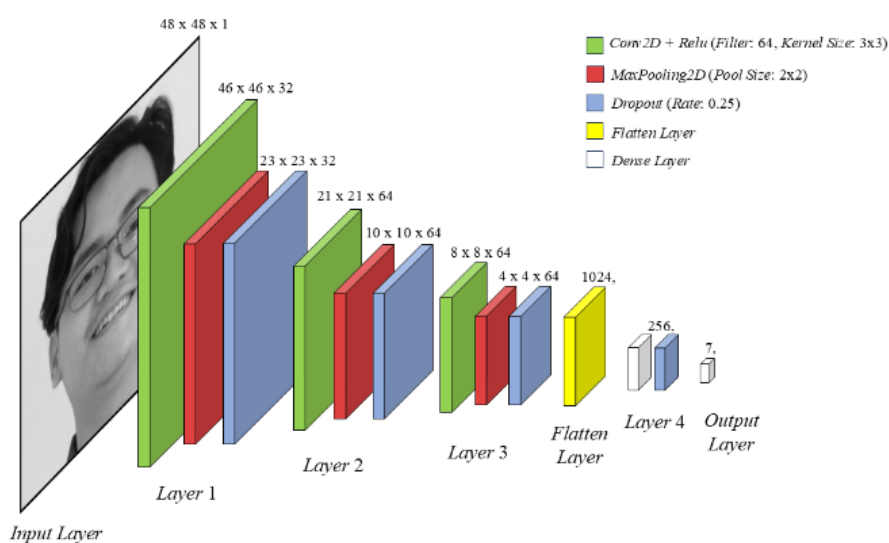


Gambar 7. Diagram Alur *Preprocessing* Data

Preprocessing data dilakukan terhadap *dataset* IMED, FER2013, JAFFE, dan *Self Built*. Proses *preprocessing* mencakup beberapa langkah antara lain ekstraksi wajah, penghapusan latar belakang, pemotongan menjadi bentuk persegi, pengubahan ukuran, augmentasi data (over/under sampling), serta pembagian data. *Preprocessing* dilakukan untuk meningkatkan kualitas data input sehingga menghasilkan kinerja model yang lebih optimal (Utama, Wibawa, Muladi, & Nafalski, 2022). Berbagai metode, seperti augmentasi data, normalisasi, dan pengurangan noise, telah digunakan untuk meningkatkan akurasi dan ketahanan CNN (Kumar & Singh, 2024). Hasil dari proses ini adalah data yang lebih terstandarisasi dan seragam, dengan kualitas yang konsisten pada setiap gambar dengan ukuran 48 x 48 piksel dengan format JPG dan PNG.

Keseluruhan *dataset* digabungkan untuk membentuk kumpulan data yang lebih besar. Penggunaan *dataset* yang besar bertujuan untuk mengurangi bias data dan meningkatkan kemampuan generalisasi model. Setelah penggabungan *dataset*, proses pembagian data dilakukan untuk menghasilkan *dataset* yang digunakan pada tahap pelatihan dan validasi. Berdasarkan penelitian sebelumnya, alokasi data sebesar 70-80% untuk pelatihan dan 20-30% untuk pengujian memberikan hasil yang optimal dalam pengembangan model machine learning (Gholamy, Kreinovich, & Kosheleva, 2018).

Arsitektur Model CNN



Gambar 8. Arsitektur Model CNN

Convolutional Neural Network (CNN) merupakan salah satu arsitektur dalam *deep neural network* karena memiliki jaringan pemrosesan yang mendalam dan banyak diaplikasikan untuk pemrosesan citra (Eka Putra W. , 2016). CNN merupakan salah satu jenis jaringan saraf tiruan yang telah menarik banyak perhatian dalam beberapa tahun terakhir berkat kemampuannya yang unggul dalam menyelesaikan berbagai tugas di bidang visi komputer dan pembelajaran mesin (Sakib, Ahmed, Kabir, & Ahmed, 2018). Arsitektur CNN sangat cocok digunakan untuk pengenalan citra wajah (Ravi, Yadhukrishna, & prithviraj, 2020). Pada penelitian ini, arsitektur CNN digunakan untuk pengenalan citra wajah penonton *stand up comedy*. Terdapat beberapa lapisan pada CNN yang berperan aktif dalam pemrosesan ekstraksi fitur dan klasifikasi antara lain *convolutional*, *pooling* dan *fully connected layers* (Lee, 2023).

a. *Convolution Layer*

Operasi konvolusi adalah proses menggeser kernel konvolusi (filter) di atas gambar input, mengalikan kernel konvolusi dan nilai piksel pada posisi yang sesuai dari gambar input, dan menjumlahkannya untuk mendapatkan peta fitur (Zhao, et al., 2024). Tujuan utama konvolusi pada data citra adalah untuk mengekstraksi fitur-fitur penting dari citra input. Proses ini menghasilkan transformasi *linear* dari data *input* berdasarkan informasi spasial yang terkandung dalam fitur data tersebut. Berikut merupakan rumus dari *convolutional layer* dengan ukuran filter 3 x 3.

$$O(i, j, k) = \sum_{m=1}^3 \sum_{n=1}^3 \sum_{c=1}^C I(i + m, j + n, c) \cdot K(m, n, c, k) \quad (1)$$

b. Fungsi Aktivasi

Fungsi aktivasi merupakan komponen matematis dalam *neural network* yang mengatur *neuron* menghasilkan *output* berdasarkan *input*-nya. Pada penerapannya ke dalam model, aktivasi berfungsi untuk membatasi *output* dari *neuron*, menyesuaikan skala data, mengaktifkan dan menonaktifkan *neuron* dan lainnya. Fungsi Aktivasi *ReLU* untuk mengubah nilai negatif menjadi nilai nol sambil menjaga nilai positif tidak berubah (Kuo, 2016). Aktivasi *ReLU* membantu jaringan mempelajari pola kompleks dengan menyaring nilai negatif menjadi nol.

$$ReLU(x) = \max(0, x) \quad (2)$$

Aktivasi *SoftMax* merupakan fungsi aktivasi yang digunakan untuk mengubah nilai skor mentah dari model menjadi distribusi probabilitas pada klasifikasi multi-kelas. Aktivasi *SoftMax* biasanya digunakan dalam lapisan keluaran jaringan CNN untuk klasifikasi multi-kelas dengan menormalkan keluaran ke distribusi probabilitas (Wasef & Rafla, 2021).

$$Softmax(x) = \frac{e^{x_i}}{\sum_{j=1}^K e^{x_j}} \quad (3)$$

c. *Pooling Layer*

Pooling Layer merupakan komponen penting dalam CNN karena berfungsi untuk mengurangi dimensi spasial peta fitur dan mengontrol *overfitting* (Gholamalinezhad & Khosravi, 2020). *Pooling Layer* membagi keluaran dari *Convolutional Layer* menjadi beberapa *grid* kecil, lalu memilih nilai tertentu dari setiap *grid* untuk membentuk matriks citra yang ukurannya telah diperkecil. Terdapat dua jenis *Pooling Layer*, yaitu *max pooling* dan *average pooling*. *Max pooling* berfungsi untuk mengambil nilai maksimum dari area gambar yang dicakup oleh *kernel*, sementara *average pooling* digunakan untuk menghitung nilai rata-rata dari area tersebut. Berikut merupakan rumus dari *max pooling* pada *pooling layer*.

$$O(i, j, k) = \max_{m, n} (I(i + m, j + n, k)) \quad (4)$$

d. *Fully Connected Layer (Dense)*

Fully Connected Layer adalah lapisan yang berfungsi untuk mentransformasi data berdimensi tinggi menjadi format yang memungkinkan klasifikasi secara *linear*. Sebelum data dari *Convolutional Layer* dapat diproses oleh *Fully Connected Layer*, setiap *neuron* perlu diubah terlebih dahulu menjadi bentuk satu dimensi. Oleh karena itu, lapisan ini biasanya ditempatkan di bagian akhir jaringan. Berikut merupakan rumus dari *fully connected layer*.

$$y = f(W_x + b) \quad (5)$$

e. *Dropout*

Dropout adalah teknik atau regulasi yang ditambahkan pada *neural network* untuk mengurangi terjadinya *overfitting* dan meningkatkan kemampuan generalisasi model. Teknik ini bekerja dengan cara menonaktifkan sejumlah *neuron* secara acak selama fase pelatihan. Proses ini dapat membantu melatih model lebih *robust* dan mengurangi ketergantungan pada *neuron* tertentu.

f. *Flatten*

Lapisan *Flatten* merupakan lapisan yang penting dalam arsitektur jaringan konvolusional. Lapisan ini digunakan untuk mengubah data *input* dari yang berbentuk multidimensi seperti matriks menjadi satu bentuk dimensi yaitu vektor. Karena lapisan *fully connected* hanya bisa menerima input dalam bentuk vektor satu dimensi, maka perlu digunakan *flatten* untuk mengubah bentuk data input.

Evaluasi Model

Pada penelitian ini dilakukan evaluasi hasil pelatihan model CNN dengan *confusion matrix*. *Confusion matrix* merupakan alat yang mendasar yang digunakan pada *machine learning* untuk mengevaluasi hasil kinerja klasifikasi model (Amin, 2022). *Confusion matrix* diperoleh dari suatu model yang dilatih pada sebuah *dataset* dan menunjukkan korelasi antara kelas kumpulan data yang berbeda mengenai *neural network*. Dalam matriks ini terdapat elemen utama berupa *True Positive* (TP), *True Negative* (TN), *False Positive* (FP), dan *False Negative* (FN). Sebagai indikator kualitas model, *confusion matrix* membantu analisis data untuk memahami sejauh mana model dapat mengenali dan mengklasifikasikan dengan benar setiap kategori dalam data, memberikan wawasan yang berharga untuk meningkatkan kinerja sistem klasifikasi. *Confusion Matrix* digunakan dalam berbagai evaluasi matrix performa, antara lain:

a. Akurasi (*Accuracy*)

Fungsinya yaitu mengukur seberapa sering model menentukan prediksi yang benar. Rumusnya adalah sebagai berikut:

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (6)$$

b. Presisi (*Precision*)

Fungsi dari *precision* adalah untuk mengukur banyaknya prediksi positif yang benar-benar positif. Rumusnya adalah sebagai berikut:

$$Precision = \frac{TP}{TP+FP} \quad (7)$$

c. *Recall*

Fungsi dari *recall* adalah untuk mengukur seberapa baik model dalam mendeteksi kasus positif. Rumusnya adalah sebagai berikut:

$$Recall = \frac{TP}{TP+FN} \quad (8)$$

d. *F1-Score*

F1-Score adalah hasil rata-rata harmonis dari presisi dan *recall* dengan memberikan keseimbangan antara keduanya. Rumusnya adalah sebagai berikut:

$$F1 - Score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \tag{9}$$

Testing Model

Model *Face Expression Recognition* yang telah dilatih dilakukan *testing* menggunakan data *crowd self built* dan data video *crowd* penonton *stand up comedy*. *Testing* dilakukan untuk mengetahui apakah model dapat melakukan prediksi *crowd emotion* dalam bentuk foto dan video *footage* asli. Kerumunan sebagai atribut untuk pemahaman emosi kerumunan dapat menjembatani kesenjangan antara fitur visual tingkat rendah dan perilaku kerumunan tingkat tinggi (Rabiee, Haddadnia, & Mousavi, 2016).

HASIL DAN PEMBAHASAN

Preprocessing Data

Preprocessing Data dilakukan terhadap *Dataset* IMED, FER2013, JAFFE dan *Self Built* dengan Bahasa Pemrograman *Python* pada *Software Visual Studio Code*. Proses dilakukan secara bertahap terhadap setiap *dataset* dengan kondisi yang berbeda sehingga proses yang dilakukan berbeda-beda. Proses yang dilakukan antara lain *face extraction*, *crop to square*, *resize*, *over/undersampling*, *split data* dan *remove background*.

Setiap *dataset* memiliki ukuran dan format gambar yang berbeda. *Dataset* IMED memiliki ukuran awal 720 x 480 *pixels* dengan warna RGB dan format gambar JPG. *Dataset* FER2013 memiliki ukuran awal 48 x 48 *pixels* dengan warna *grayscale* dan format gambar JPG. *Dataset* JAFFE memiliki ukuran 256 x 256 *pixels* dengan warna *grayscale* dan format gambar TIFF. *Dataset Self Built* memiliki ukuran awal 4000 x 6000 *pixels* dengan warna sRGB dan format gambar JPG. Dari keseluruhan *dataset* tersebut, ukuran yang terkecil adalah 48 x 48 *pixels* dan format yang paling umum adalah JPG. Proses *preprocessing data* ditujukan untuk menyeragamkan dan menyetandarkan ukuran menjadi 48 x 48 *pixels* dan format gambar JPG. Dengan kualitas data input yang seragam akan menghasilkan kinerja model yang lebih optimal (Utama, Wibawa, Muladi, & Nafalski, 2022).

Setiap *dataset* memiliki jumlah data yang berbeda setiap kelasnya sehingga perlu dilakukan penyeimbangan data setiap kelas untuk mencegah terjadinya *oversampling* pada kelas tertentu. Ketidakseimbangan kelas dalam kumpulan data secara signifikan memengaruhi kinerja CNN dalam tugas klasifikasi gambar (Banerjee, dkk., 2021). Proses yang dilakukan yaitu *oversampling* pada kelas dengan jumlah kecil, *undersampling* pada kelas dengan jumlah data besar dan *split data* untuk membagi data menjadi data *train* dan *validation*. *Dataset* IMED, FER2013 dan JAFFE digunakan untuk proses *training* dan *validation*. *Dataset Self Built* digunakan untuk *testing* model. Berikut merupakan hasil dari proses *preprocessing data* yang dapat dilihat pada Tabel 2.

Tabel 2. Hasil *Preprocessing* Dataset

No	Emotion	Initial Dataset	Final Dataset		
		(IMED, FER2013, JAFFE)	Train	Validation	Testing (Self Built)
1	Angry	6566	4723	1114	45
2	Disgusting	1989	3837	945	45
3	Fear	6619	4891	1154	45
4	Happy	10340	4678	1099	45
5	Neutral	6708	5323	1257	45
6	Sad	7900	5139	1216	45
7	Surprise	4982	4352	1108	45

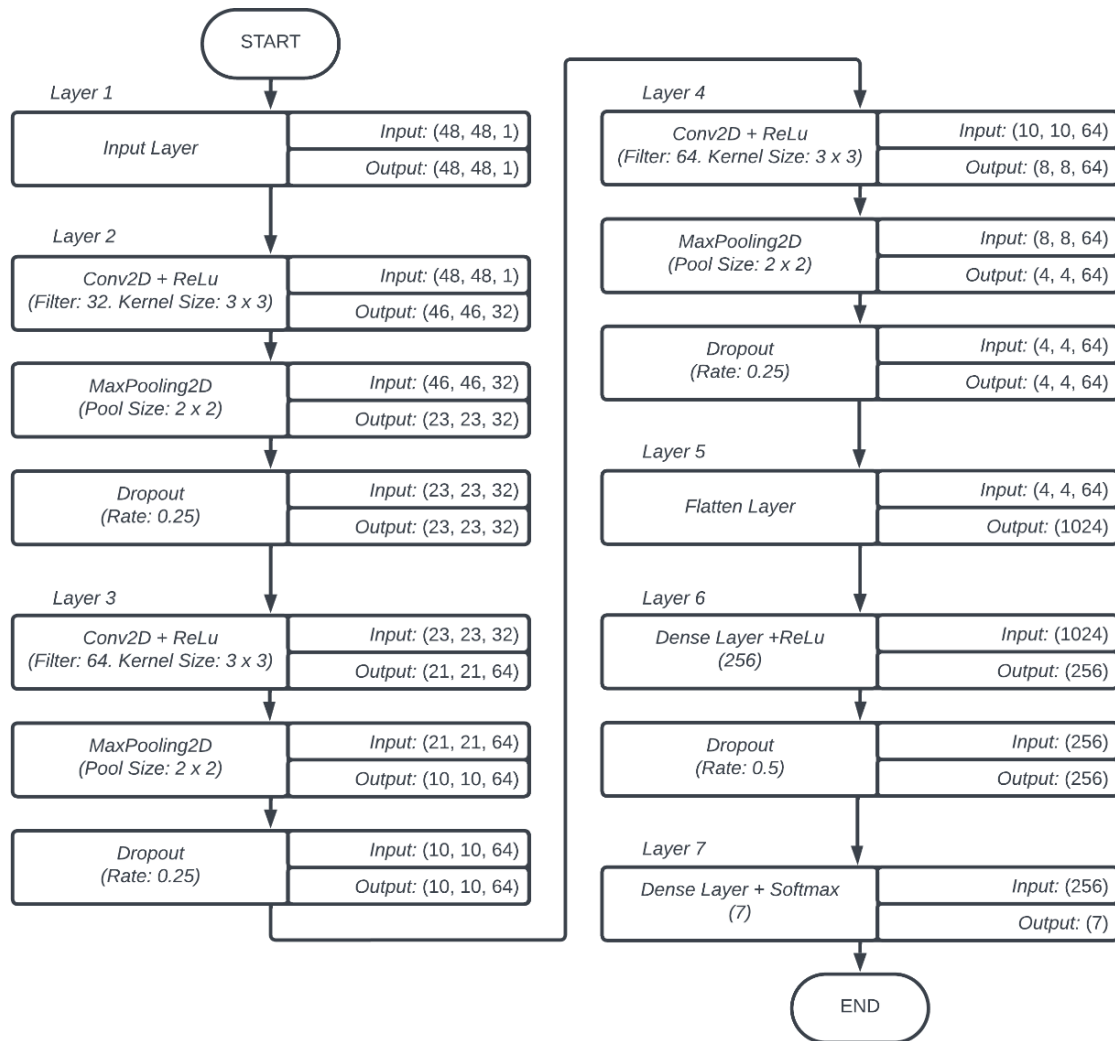
Dataset Self-built juga digunakan untuk simulasi penonton *stand up comedy* dengan menggabungkan beberapa ekspresi ke dalam satu gambar. Data berupa gambar dengan jumlah 15 orang dengan 7 ekspresi yang dibangun secara acak. Data ini digunakan untuk menguji model apakah dapat menganalisis ekspresi dalam bentuk gambar. Berikut merupakan *screenshot* dari video tersebut yang dapat dilihat pada Gambar 9.



Gambar 9. Data *Crowd Self Built*

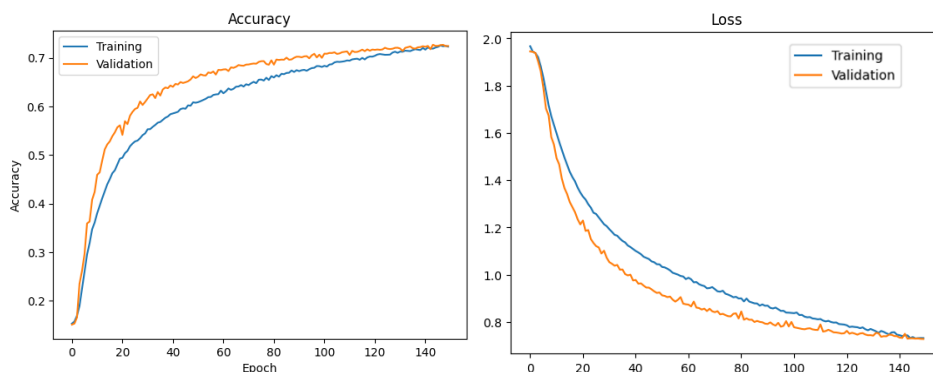
Convolutional Neural Network (CNN)

Pada penelitian ini digunakan arsitektur CNN sederhana yang dimodifikasi dari arsitektur VGG. *Visual Geometry Group* (VGG) merupakan arsitektur model CNN yang populer untuk pemrosesan klasifikasi gambar (Zakaria & Hassim, 2023). Arsitektur model CNN pada penelitian ini terdiri dari *input layer*, *convolutional layer*, *maxpooling layer*, *dropout layer*, *flatten layer* dan *dense layer*. *Input layer* merupakan lapisan penerima data citra dengan dimensi yang ditetapkan sesuai *dataset* yang digunakan yaitu $48 \times 48 \text{ pixel}$. Lapisan 2, 3 dan 4 terdiri dari 3 lapisan yang sama. Lapisan pertama adalah *convolutional layer* yang berfungsi untuk mengekstrak fitur dari gambar dengan menggeser filter dengan ukuran 3×3 dengan tambahan aktivasi *ReLU* untuk mengubah data menjadi *non-linear*. Lapisan kedua adalah *maxpooling layer* yang berfungsi untuk mengurangi dimensi dari input dengan memilih data yang paling besar dalam satu *filter maxpooling* 2×2 . Lapisan ketiga adalah *dropout layer* yang berfungsi untuk menonaktifkan beberapa unit neuron secara acak dengan rasio 0.25. *Flatten layer* berfungsi untuk mengubah data *array* dua dimensi menjadi satu dimensi. *Dense layer* berfungsi untuk melakukan pemetaan dari setiap fitur yang telah terekstraksi oleh lapisan konvolusi dan menjadi keputusan klasifikasi. Pada lapisan terakhir terdapat aktivasi *softmax* yang berfungsi untuk mengubah *output* dari *dense layer* menjadi matriks probabilitas yaitu 7 klasifikasi ekspresi. Berikut merupakan gambaran arsitektur CNN yang digunakan pada penelitian ini yang dapat dilihat pada Gambar 10.



Gambar 10. Diagram Alur Arsitektur Model CNN

Model dilatih dengan *final dataset* yang dapat dilihat pada Tabel 2 menggunakan Adam *Optimizer* dengan *learning rate* sebesar 0.00005. Ditambahkan *loss function* yaitu *categorical crossentropy* yang sesuai dengan kasus klasifikasi multi-kelas. Pelatihan model dilakukan dengan *batch size* 32 dan *epoch* 150. Berikut merupakan grafik *accuracy* dan *loss* pada proses *training* dan *validation*.



Gambar 11. Grafik Accuracy dan Loss pada Proses Training dan Validation

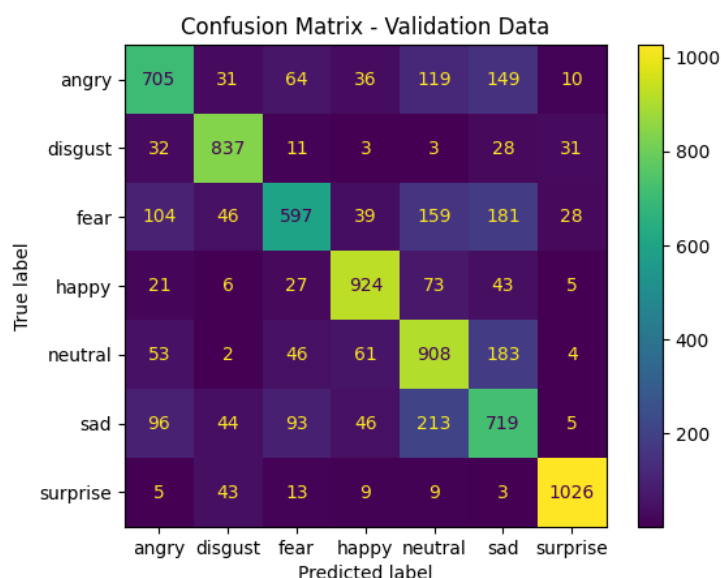
Hasil dari proses pelatihan model CNN yang telah dilakukan digambarkan pada grafik *accuracy* dan *loss*. Proses pelatihan model dilakukan dengan *epoch* 150 dan *batch size* 32 yang

menunjukkan bahwa grafik *accuracy* dan *loss* memiliki tren yang stabil dan konsisten. Grafik *accuracy* mengindikasikan peningkatan yang signifikan pada tahap awal pelatihan (*epoch* 0–20) untuk data *training* dan *validation*. Hal tersebut mencerminkan kemampuan model untuk dengan mempelajari pola dasar dari *dataset* dengan cepat. Setelah mencapai sekitar *epoch* ke-100, peningkatan akurasi mulai melambat dan berakhir stabil. Selain itu, *accuracy* pada proses *validation* memiliki nilai yang lebih tinggi dibandingkan *accuracy* pada proses *training* menunjukkan kemampuan generalisasi model yang baik serta mengindikasikan tidak adanya *overfitting*. Pada grafik *loss*, terlihat penurunan yang konsisten pada nilai *loss* untuk proses *training* dan *validation* sepanjang proses pelatihan. Penurunan yang paling signifikan terjadi pada fase awal (*epoch* 0–20), yang mengindikasikan bahwa model sedang melakukan adaptasi terhadap data. Setelah mendekati *epoch* ke-100, penurunan nilai *loss* melambat, menandakan bahwa model hampir mencapai konvergensi. Di akhir pelatihan, nilai *training loss* dan *validation loss* mendekati kesamaan, menunjukkan bahwa model telah belajar secara optimal tanpa mengalami *overfitting*.

Hasil ini menunjukkan bahwa model CNN dapat mempelajari *dataset* dengan efektif dan memiliki kemampuan generalisasi yang baik. Selanjutnya dilakukan evaluasi dengan *dataset testing* untuk memastikan bahwa model mampu mempertahankan kinerja yang baik pada data yang belum pernah dilihat sebelumnya.

Validation Model

Pada penelitian ini dilakukan proses *validation* yang bertujuan untuk mengevaluasi performa model selama tahap *training* dengan menggunakan data *validation* yang terpisah dari data *training*. Proses *validation* berfungsi untuk melihat kemampuan model dalam menggeneralisasi data baru serta mencegah terjadinya *overfitting*. Evaluasi dilakukan dengan mengukur *confusion matrix* seperti *precision*, *recall*, *F1-score*, dan *accuracy*. *Confusion matrix* dapat memberikan gambaran tentang kemampuan model dalam mengklasifikasikan ekspresi wajah pada data *validation*. Selain itu, *confusion matrix* digunakan untuk menganalisis pola prediksi model dan mengidentifikasi kesalahan prediksi yang dominan pada setiap kelas kategori. Hasil validasi ini menjadi acuan penting dalam menentukan parameter optimal model sebelum pengujian akhir dilakukan. Berikut merupakan *confusion matrix* dan *classification report* pada proses *validation* yang dapat dilihat pada Gambar 12 dan Tabel 3.



Gambar 12. *Confusion Matrix* pada Proses *Validation* Model

Tabel 3. *Classification Report Validation Model*

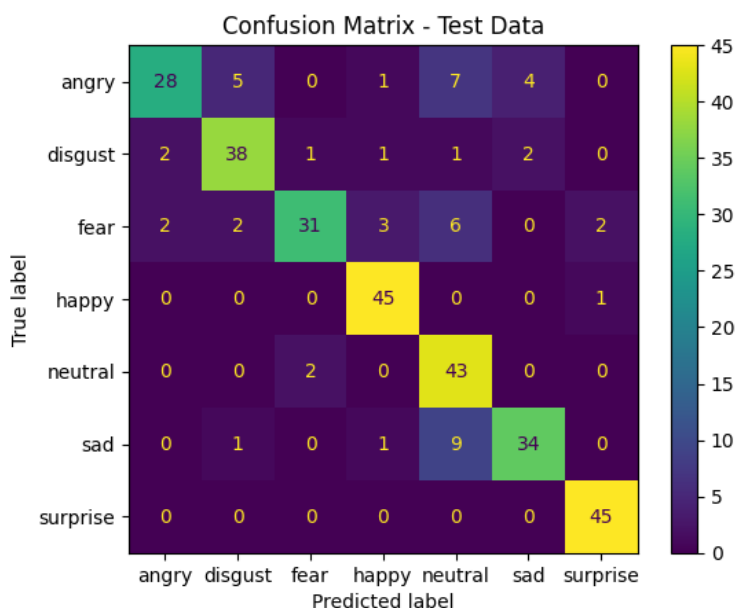
<i>Classification Report for Validation Data (32 x 150)</i>				
<i>Category</i>	<i>Precision</i>	<i>Recall</i>	<i>F1-Score</i>	<i>Support</i>
<i>angry</i>	0.6939	0.6329	0.662	1114
<i>disgust</i>	0.8295	0.8857	0.8567	945
<i>fear</i>	0.7015	0.5173	0.5955	1154
<i>happy</i>	0.8265	0.8408	0.8336	1099
<i>neutral</i>	0.6119	0.7224	0.6625	1257
<i>sad</i>	0.5505	0.5913	0.5702	1216
<i>surprise</i>	0.9252	0.926	0.9256	1108
<i>accuracy</i>			0.7242	7893
<i>macro avg</i>	0.7341	0.7309	0.7294	7893
<i>weighted avg</i>	0.727	0.7242	0.7224	7893

Berdasarkan *confusion matrix* dan *classification report* hasil dari proses *validation* menunjukkan performa model yang beragam untuk masing-masing kategori emosi. Pada *confusion matrix*, kategori *surprise* memiliki jumlah prediksi yang benar tertinggi dengan nilai diagonal mencapai 1026, diikuti oleh kategori *happy* dan *disgust*. Hal ini konsisten dengan hasil pada *classification report*, di mana *surprise* memiliki nilai tertinggi yaitu pada *precision* sebesar 0.925, *recall* 0.926, dan *F1-score* 0.926. Ini menunjukkan bahwa model mampu mengenali ekspresi *surprise* dengan baik. Sebaliknya, kategori *sad* memiliki performa yang paling rendah, dengan *precision* sebesar 0.5505, *recall* 0.5913, dan *F1-score* 0.5702. Hal ini menunjukkan bahwa model mengalami kesulitan dalam mengklasifikasikan emosi *sad*, yang kemungkinan disebabkan oleh tumpang tindih fitur dengan kategori lain, seperti *fear* atau *neutral*. Selain itu, kategori *neutral* juga memiliki nilai *precision* dan *recall* yang cukup rendah, masing-masing 0.6119 dan 0.7224, meskipun jumlah sampel kategori ini cukup besar.

Secara keseluruhan, akurasi model pada data *validation* adalah 72.42%, dengan nilai *macro average* pada *precision*, *recall* dan *F1-score* masing-masing sebesar 0.7341, 0.7309 dan 0.7294. Hal ini menunjukkan bahwa model memiliki performa generalisasi yang cukup baik, meskipun terdapat ketidakseimbangan dalam akurasi pengenalan antar kategori. Untuk meningkatkan performa pada kategori dengan *F1-score* rendah, pendekatan seperti data *augmentation* yang lebih spesifik, *rebalancing dataset*, atau modifikasi arsitektur model dapat dipertimbangkan. Analisis ini mengindikasikan bahwa model cukup efektif untuk mendeteksi emosi tertentu, terutama *surprise* dan *happy*, tetapi masih memerlukan perbaikan untuk kategori dengan performa rendah seperti *sad* dan *fear*.

Testing Model

Pada penelitian ini dilakukan poses *testing* yang bertujuan untuk mengevaluasi performa klasifikasi ekspresi wajah pada data *testing* yang belum pernah dilihat oleh model sebelumnya. Evaluasi dilakukan untuk mengukur sejauh mana model mampu mengenali emosi dengan akurasi tinggi, berdasarkan *confusion matrix* seperti *precision*, *recall*, *F1-score*, dan *accuracy*. *Confusion matrix* digunakan untuk memberikan gambaran mendalam mengenai distribusi prediksi model, termasuk kekeliruan yang terjadi antar kategori emosi. Analisis ini dapat menentukan kekuatan dan kelemahan model dalam menangani berbagai kategori emosi.



Gambar 13. Confusion Matrix pada Proses Testing Model

Tabel 3. Classification Report Testing Model

Classification Report for Test Data (32 x 150)				
Category	Precision	Recall	F1-Score	Support
<i>angry</i>	0.875	0.6222	0.7273	45
<i>disgust</i>	0.8261	0.8444	0.8352	45
<i>fear</i>	0.9118	0.6739	0.775	46
<i>happy</i>	0.8824	0.9783	0.9278	46
<i>neutral</i>	0.6515	0.9556	0.7748	45
<i>sad</i>	0.85	0.7556	0.8	45
<i>surprise</i>	0.9375	1	0.9677	45
<i>accuracy</i>			0.8328	317
<i>macro avg</i>	0.8477	0.8329	0.8297	317
<i>weighted avg</i>	0.8481	0.8328	0.8298	317

Pada tahap *testing*, model diuji menggunakan 317 sampel data *testing* untuk mengevaluasi performanya dalam mengklasifikasikan tujuh kategori emosi yaitu *angry*, *disgust*, *fear*, *happy*, *neutral*, *sad* dan *surprise*. Berdasarkan *confusion matrix*, performa model menunjukkan distribusi prediksi yang beragam. Model berhasil mengklasifikasikan kategori *surprise* dengan 45 sampel benar dengan nilai *recall* 1.0 dan *F1-score* 0.9677. Hal tersebut menunjukkan kemampuan model sangat baik dalam mengenali emosi tersebut. Kategori *happy* juga memiliki performa tinggi, dengan nilai *recall* 0.9783 dan *F1-score* 0.9278. Kategori *neutral* memiliki nilai *precision* terendah yaitu 0.6515, meskipun *recall*-nya cukup tinggi yaitu sebesar 0.9556. Hal tersebut mengindikasikan banyak prediksi positif palsu pada kategori ini. Sementara itu, kategori *angry* memiliki nilai *recall* sebesar 0.6222, menunjukkan bahwa model kesulitan mengenali emosi ini meskipun nilai *precision*-nya cukup baik yaitu sebesar 0.875. Kategori lain seperti *fear*, *disgust*, dan *sad* memiliki performa sedang, dengan nilai *F1-score* masing-masing sebesar 0.775, 0.8352, dan 0.8.

Secara keseluruhan, akurasi *testing* mencapai 83.28%, dengan nilai *macro average* untuk *precision*, *recall*, dan *F1-score* masing-masing sebesar 0.8477, 0.8329, dan 0.8297. Hasil ini menunjukkan bahwa model memiliki kemampuan klasifikasi yang baik secara umum, meskipun terdapat kekeliruan dalam membedakan beberapa kategori emosi. Analisis ini menunjukkan bahwa model lebih efektif untuk kategori dengan pola emosi yang lebih jelas seperti *surprise* dan *happy*, sedangkan pada ekspresi yang tidak khas diperlukan perbaikan lebih lanjut.

Penerapan Data Crowd

```

START
IMPORT libraries: cv2, numpy, networkx, distance, Counter
IMPORT library: tensorflow.keras.models as models
DEFINE load_cnn_model:
    MODEL = models.load_model("path_to_your_cnn_model.h5")
    RETURN MODEL
DEFINE preprocess_faces:
    FOR each detected face:
        CALCULATE face center
        RESIZE face to (48, 48)
        NORMALIZE face
        ADD face and center to lists
    RETURN processed faces and centers
DEFINE process_image:
    LOAD image and convert to grayscale
    DETECT faces in image
    IF faces detected:
        PROCESSED_FACES = preprocess_faces(detected_faces)
        PRINT "Predicted emotions:", MODEL.predict(PROCESSED_FACES)
    ELSE:
        PRINT "No faces detected"
DEFINE process_video:
    OPEN video
    WHILE video has frames:
        LOAD frame and convert to grayscale
        DETECT faces in frame
        IF faces detected:
            PROCESSED_FACES = preprocess_faces(detected_faces)
            PRINT "Predicted emotions:", MODEL.predict(PROCESSED_FACES)
        IF 'q' pressed, STOP
    CLOSE video
MODEL = load_cnn_model()
process_image(MODEL)
process_video(MODEL)
END

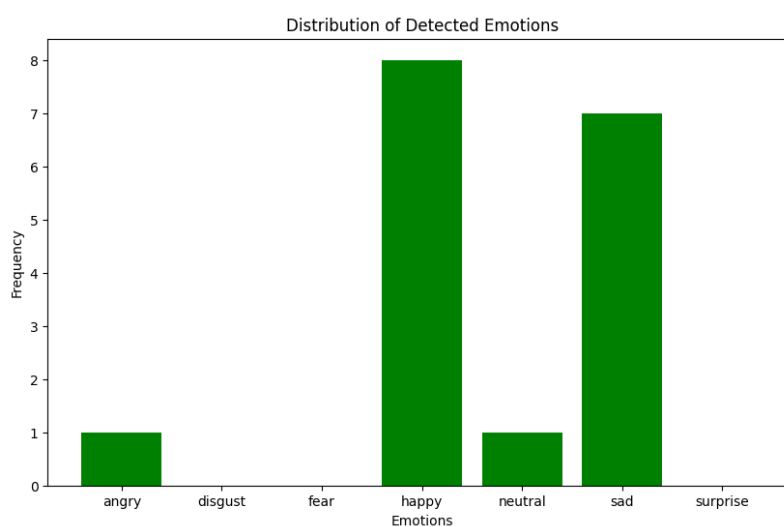
```

Gambar 14. Pseudocode Face Expression Recognition

Pada penelitian ini dilakukan pengujian model *Convolutional Neural Network* (CNN) yang telah dilatih dengan data foto *crowd self built* dan video *crowd* penonton *stand up comedy*. Proses dilakukan dengan tiga tahapan utama antara lain pemuatan model, *preprocessing* data wajah dan prediksi emosi pada gambar maupun video. Berikut merupakan *pseudocode* yang digunakan untuk melakukan prediksi emosi pada gambar dan video yang dapat dilihat pada Gambar 14.



Gambar 15. Hasil Pengujian Model dengan Data *Crowd Self Built*



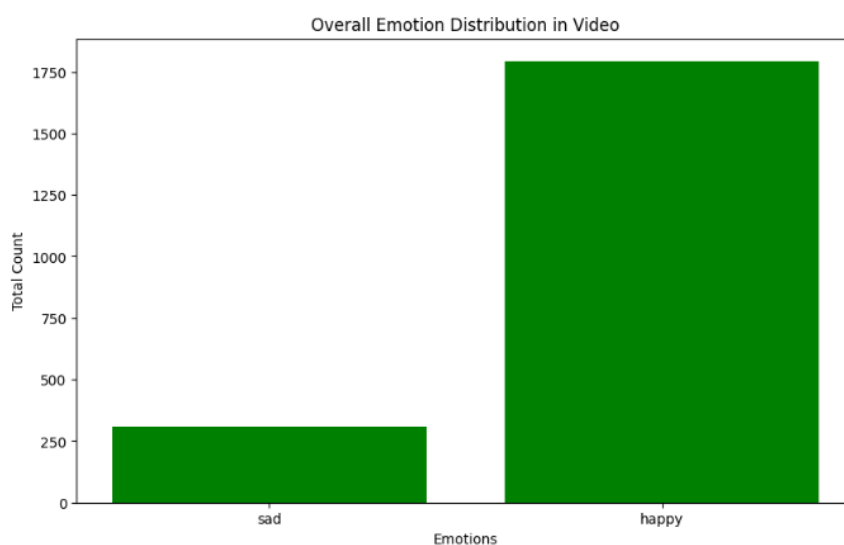
Gambar 16. Distribusi Hasil Deteksi Emosi pada Data *Crowd Self Built*

Hasil pengujian model deteksi ekspresi wajah pada *crowd* dengan data *self built* menunjukkan hasil klasifikasi yang cukup akurat dengan mendeteksi berbagai ekspresi pada data gambar atau *frame*. Pada Gambar x, model berhasil mendeteksi tujuh kategori ekspresi wajah (*angry, disgust, fear, happy, neutral, sad, dan surprise*) sesuai dengan jumlah data latih yang digunakan sebelumnya. Berdasarkan hasil deteksi, ekspresi *happy* dan *sad* mendominasi distribusi ekspresi yang terdeteksi pada kerumunan, sebagaimana divisualisasikan pada grafik distribusi di Gambar 16.

Distribusi Ekspresi pada grafik menunjukkan bahwa terdapat ekspresi yang diprediksi dalam *frame*. Ekspresi *happy* merupakan ekspresi yang paling sering muncul, dengan total frekuensi sebanyak 8 individu. Ekspresi *sad* berada di urutan kedua, dengan 6 individu yang terdeteksi memiliki ekspresi ini. Ekspresi *neutral* hanya terdeteksi pada 1 individu, sedangkan *angry* terdeteksi pada 1 individu lainnya. Tidak ada deteksi untuk ekspresi *disgust, fear, maupun surprise* dalam kerumunan ini. Kondisi ini mencerminkan kemampuan model untuk memberikan prediksi yang sesuai dengan suasana atau situasi di dalam gambar, seperti dominasi ekspresi *happy* yang mengindikasikan suasana positif dalam acara tersebut.



Gambar 17. Hasil Pengujian Model dengan Data Video *Crowd Penonton Stand Up Comedy*



Gambar 18. Distribusi Hasil Deteksi Emosi pada Data Video *Crowd Penonton Stand Up Comedy*

Pengujian model selanjutnya dilakukan dengan data *video crowd* penonton *stand up comedy*. Hasil pengujian menunjukkan bahwa model dapat mendeteksi ekspresi pada data video dan disajikan dalam grafik distribusi keseluruhan emosi yang dapat dilihat pada Gambar 18. Hasil distribusi menunjukkan bahwa emosi keseluruhan dalam video terdiri dari emosi *happy* dengan lebih dari 1750 deteksi dan sebagian kecil emosi *sad* dengan jumlah lebih dari 250 deteksi. Hasil ini menunjukkan bahwa model CNN yang digunakan memiliki performa yang baik dalam mendeteksi ekspresi emosi dominan dalam konteks acara hiburan. Dominasi emosi "happy" sesuai dengan ekspektasi karena *stand-up comedy* bertujuan untuk menciptakan suasana humor dan kebahagiaan.

KESIMPULAN

Penelitian ini dilakukan untuk menerapkan model pengenalan ekspresi wajah dengan *Convolutional Neural Network* (CNN) untuk mendeteksi emosi penonton *stand up comedy*. Proses dilakukan dengan tahapan *preprocessing* data, pelatihan model, validasi, pengujian dan penerapan pada data kerumunan. *Dataset* dari berbagai sumber yaitu IMED, FER2013, JAFFE dan *Self Built* diproses untuk diseragamkan format dan ukurannya. *Dataset* juga dilakukan penyeimbangan kelas data untuk meningkatkan kinerja model. Model CNN yang dirancang

berdasarkan modifikasi arsitektur VGG menunjukkan hasil akurasi validasi sebesar 72% dan akurasi pengujian sebesar 83.28%. Pada data kerumunan, model dapat mendeteksi berbagai emosi dalam bentuk foto dan video. Perbaikan dapat dilakukan untuk meningkatkan performansi model yang dapat dilakukan dengan data *augmentation*, *rebalancing dataset* ataupun modifikasi arsitektur CNN. Penelitian selanjutnya juga bisa dilakukan dengan berfokus pada pengkalibrasian deteksi wajah ataupun pembuatan aplikasi secara *real time*. Penerapan model memiliki potensi yang tinggi untuk digunakan sebagai analisis ekspresi secara *real time* pada berbagai konteks sosial dan hiburan terutama *stand up comedy*.

UCAPAN TERIMA KASIH

Terimakasih kepada Universitas Sebelas Maret yang telah mendukung suasana akademik dalam penelitian ini. Terima kasih juga kepada Asisten Laboratorium Sistem Produksi Teknik Industri UNS yang bersedia menjadi responden dalam penelitian ini dan semua pihak yang telah menyumbangkan data dan memberikan informasi untuk mendukung penelitian ini.

DAFTAR PUSTAKA

- Amaanullah, R. F., Pasfica, G. R., Nugraha, S. A., Zein, M. R., & Adhinata, F. D. (2022). Implementasi Convolutional Neural Network Untuk Deteksi Emosi Melalui Wajah. *Jurnal Teknologi Informasi dan Multimedia*, 236 - 244.
- Amin, M. F. (2022). Confusion Matrix in Binary Classification Problems: A Step-by-Step Tutorial. *Journal of Engineering Research*.
- Banerjee, A., Ghosh, K., Sarkar, A., Bhattacharjee, M., & Chatterjee, S. (2021). Effects of Class Imbalance Problem in Convolutional Neural Network Based Image Classification. *Computer Science*.
- Eka Putra, W. S. (2016). Klasifikasi Citra Menggunakan Convolutional Neural Network (CNN) pada Caltech 101. *Jurnal Teknik ITS*.
- Ekman, P. (1992). Facial Expression of Emotion: New Findings, New Question. *Psychological Science*, 3(1), 34-38.
- Fukushima, K. (1980). *Neocognitron: A Self-Organizing Neural Network Modul for a Mechanism of Pattern Recognition Unaffected by Shift in Position*. Biological Cybernetics.
- Gholamalizadeh, H., & Khosravi, H. (2020). Pooling Methods in Deep Neural Networks, a Review. *arXiv.org*.
- Gholamy, A., Kreinovich, V., & Kosheleva, O. (2018). Why 70/30 or 80/20 Relation Between Training and Testing Sets: A Pedagogical Explanation. *Computer Science, Mathematics, Education*.
- Hania, A. A. (2017). Mengenal Artificial Intelligence, Machine Learning, Neural Network, dan Deep Learning. *Jurnal Teknologi Indonesia*.
- Hills, G. (2006). *A Guide to UK Performing Arts*. London.
- Holder, R. P. (2017). *Using Facial Expression Recognition for Crowd Monitoring*. Durban, South Africa: College of Agriculture, Engineering and Science, University of Kwazulu-Natal.
- Kominfo. (2022, January 14). *Pemerintah Dorong Optimalisasi Pertumbuhan Industri Kreatif Indonesia*. Diambil kembali dari Kominfo: <https://www.kominfo.go.id>

- Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2017). ImageNet Classification with Deep Convolutional Neural Networks Revisited. *Communication of the ACM Volume 60*, 84 - 90.
- Kumar, P., & Singh, D. (2024). Pre-Processing Techniques for Enhancing CNN Performance. *International Conference on Reliability, Infocom Technologies and Optimization (ICRITO)*.
- Kuo, C. -C. (2016). Understanding Convolutional Neural Networks with a Mathematical Model. *Journal of Visual Communication and Image Representation*.
- Kusdiananggalih, P., & Rachmawati, E. (2021). Pengenalan Ekspresi Wajah dari Cross Dataset Menggunakan Convolutional Neural Network (CNN). 3429-3445.
- Lee, Y. (2023). The CNN: The Architecture Behind Artificial Intelligence Development. *Journal of student-scientists' research*.
- Liliana, D., Basaruddin, T., & Oriza, I. (2018). The Indonesian Mixed Emotion Dataset (IMED): A Facial Expression Dataset for Mixed Emotion Recognition. *International Conference on Artificial Intelligence and Virtual Reality*, 56-60.
- MacRury, I. (2012). Humour as 'Social Dreaming': Stand-up Comedy as Therapeutic Performance. *Psychoanal Cult Soc 17*, 185-203.
- Parkavi, J. (2020). A Detailed Study of Deep Learning using Convolutional Neural Network Approach. *International Journal for Research in Applied Science and Engineering Technology*.
- PricewaterhouseCoopers. (2023). *Entertainment and media outlook 2019-2023*. Diambil kembali dari PricewaterhouseCoopers: <https://www.pwc.com/>
- Rabiee, H., Haddadnia, J., & Mousavi, H. (2016). Crowd behavior representation: an attribute-based approach. *SpringerPlus*.
- Ravi, R., Yadhukrishna, S., & prithviraj, R. (2020). A Face Expression Recognition Using CNN & LBP. *International Conference Computing Methodologies and Communication*.
- Sahu, M., & Dash, R. (2020). A Survey on Deep Learning: Convolution Neural Network (CNN). *Smart Innovation, Systems and Technologies*.
- Sakib, S., Ahmed, N., Kabir, A., & Ahmed, H. (2018). An Overview of Convolutional Neural Network: Its Architecture and Applications. *Computer Science*.
- Simske, S., & Vans, M. (2017). Learning before Learning: Reversing Validation and Training. *ACM Symposium on Document Engineering*.
- Sriyati, Setyanto, A., & Luthfi, E. T. (2020). Literature Review: Pengenalan Wajah Menggunakan Algoritma Convolutional Neural Network. *TIKomSIN*, 63-72.
- Sudarmawan, W. (2020). *Kualitas Acara Televisi: Peran Kepuasan Menonton Sebagai Mediator Pengaruh Terhadap Loyalitas Penonton*. Yogyakarta: Program Pasca Sarjana Ilmu Ekonomi Universitas Islam Indonesia.
- Tian, Y., Kanade, T., & Cohn, J. (2011). Facial Expression Recognition. *Handbook of Face Recognition*, 487-519.
- Utama, A. B., Wibawa, A. P., Muladi, & Nafalski, A. (2022). PSO based Hyperparameter tuning of CNN Multivariate Time-Series Analysis. *Jurnal Online Informatika (JOIN)*, 193-202; DOI: 10.15575/join.v7i2.858.
- Wasef, M. R., & Rafla, N. (2021). Hardware Implementation of Multi-Rate Input SoftMac Activation Function. *Midwest Symposium on Circuits and Systems*.

Zakaria, N., & Hassim, Y. M. (2023). Improved Image Classification Task Using Enhanced Visual Geometry Group of Convolution Neural Networks. *JOIV: International Journal on Informatics Visualization*.

Zhao, X., Wang, L., Zhang, Y., Han, X., Deveci, M., & Parmar, M. (2024). A Review of Convolutional Neural Networks in Computer Vision. *Artificial Intelligence Review*.