

Real-time Convolutional Neural Networks untuk Klasifikasi Emosi Wajah yang Terintegrasi dengan Rekomendasi Konten YouTube Berdasarkan Emosi Secara Rule-Based

Annisa Dinda Gantini¹, Adinda Faayza Malika², Berliana Elfada³ Priyanto Hidayatullah^{*4}, Muhammad Rizqi Sholahuddin⁵

^{1,2,3,4} Jurusan Teknik Komputer dan Informatika, Politeknik Negeri Bandung
annisa.dinda.tif421@polban.ac.id¹, adinda.faayza.tif421@polban.ac.id², berliana.elfada.tif421@polban.ac.id³,
priyanto@polban.ac.id^{*4}, muhammad.rizqi@polban.ac.id⁵

Abstract

The ability of a system to detect emotions in human facial images and provide YouTube content recommendations based on these detections can enhance human interaction with technology. This study proposes a novel integration of real-time facial emotion recognition and emotion-based multimedia content recommendation, implemented through a lightweight Convolutional Neural Network (CNN) framework. Two efficient CNN architectures—Sequential Fully-CNN and Mini-Xception—are compared for their performance in classifying seven emotions from the FER-2013 dataset. A key contribution of this work is the combination of real-time emotion classification and a rule-based mapping system that delivers personalized YouTube content recommendations based on the user's emotional state, offering a seamless, emotionally aware user experience. Experimental results show that the developed CNN models can achieve high accuracy in emotion recognition while maintaining computational efficiency suitable for real-time deployment. This approach is expected to improve the quality of human-computer interaction through responsive and emotionally relevant multimedia feedback.

Keywords: Emotion Detection, Convolutional Neural Networks, YouTube, Artificial Intelligence, Video Recommendation

Abstrak

Kemampuan sistem untuk mendeteksi emosi pada citra wajah manusia dan memberikan rekomendasi konten YouTube berdasarkan hasil deteksi tersebut dapat meningkatkan interaksi manusia dengan teknologi. Penelitian ini mengusulkan integrasi baru antara deteksi emosi wajah secara real-time dengan sistem rekomendasi multimedia berbasis emosi, menggunakan kerangka kerja ringan berbasis Convolutional Neural Networks (CNN). Dua arsitektur CNN yang efisien—Sequential Fully-CNN dan Mini-Xception—dibandingkan untuk mengklasifikasikan tujuh emosi pada dataset FER-2013. Kontribusi utama dari penelitian ini adalah penggabungan klasifikasi emosi secara real-time dengan sistem pemetaan berbasis aturan untuk merekomendasikan konten YouTube yang relevan secara emosional, menciptakan pengalaman pengguna yang lebih personal dan adaptif. Hasil pengujian menunjukkan bahwa model CNN yang dikembangkan mampu mendeteksi emosi dengan akurasi tinggi sekaligus menjaga efisiensi komputasi untuk aplikasi waktu nyata. Pendekatan ini diharapkan dapat meningkatkan kualitas interaksi manusia-komputer melalui umpan balik multimedia yang responsif dan relevan secara emosional.

Kata kunci: Deteksi Emosi, Convolutional Neural Networks, YouTube, Kecerdasan Buatan, Rekomendasi Video

I. PENDAHULUAN

Keberhasilan robotika layanan sangat bergantung pada interaksi yang lancar antara robot dan pengguna. Salah satu elemen penting dalam membangun interaksi yang efektif adalah kemampuan robot untuk mengenali dan merespons emosi pengguna. Emosi manusia umumnya tercermin melalui ekspresi wajah, yang menjadi sinyal nonverbal utama dalam komunikasi sosial. Namun, menafsirkan emosi melalui citra wajah bukanlah tugas yang sederhana. Variabilitas bentuk wajah, pencahayaan, sudut pandang, serta perbedaan ekspresi antar individu membuat proses klasifikasi emosi menjadi tantangan tersendiri. Akibatnya, model *Machine Learning*

(ML) yang digunakan sering kali memiliki kompleksitas tinggi dengan jutaan parameter, sehingga membutuhkan daya komputasi besar. Dalam konteks robotika layanan dan sistem *real-time*, efisiensi model menjadi aspek yang tidak dapat diabaikan [1].

Emosi sendiri merupakan konsep yang kompleks dan belum memiliki definisi tunggal dalam literatur. Emosi tidak selalu dapat dirasakan secara eksplisit seperti rasa sakit, melainkan menggambarkan kondisi internal yang bersifat esoterik dan kadang sulit diukur secara objektif [2]. Meski demikian, ekspresi wajah tetap menjadi indikator universal dan natural dalam menyampaikan emosi [3]. Studi menunjukkan bahwa akurasi manusia dalam mengklasifikasikan citra wajah ke dalam tujuh

emosi dasar {"marah", "jijik", "takut", "senang", "sedih", "terkejut", "netral"} pada *dataset* FER-2013 seperti pada Gambar 1 hanya mencapai sekitar $65\% \pm 5\%$, mencerminkan tantangan bahkan bagi manusia dalam mengidentifikasi emosi secara konsisten [4].

Seiring dengan berkembangnya teknologi pengenalan citra, CNN telah menjadi pendekatan utama dalam tugas-tugas visual seperti klasifikasi gambar dan deteksi objek. Keunggulan CNN dalam mengekstraksi fitur spasial menjadikannya sangat cocok untuk mendeteksi ekspresi wajah. Namun, arsitektur CNN konvensional sering kali berukuran besar dan tidak ideal untuk implementasi pada perangkat dengan sumber daya terbatas. Oleh karena itu, pengembangan model CNN yang ringan menjadi solusi menjanjikan karena menawarkan efisiensi komputasi yang lebih tinggi tanpa mengorbankan performa secara signifikan.

Di luar pengenalan berbasis wajah, pendekatan berbasis emosi juga telah diterapkan dalam berbagai kondisi emosional pengguna [5]. Sementara itu, CNN juga terbukti efektif dalam konteks pengenalan emosi berbasis sinyal otak (EEG), dengan model seperti M1 dan M2 yang mencapai akurasi mendekati sempurna, yaitu 99,89% dan 99,22% [6].

Implementasi CNN juga mencapai kinerja yang lebih baik daripada metode tradisional pada kasus *real-time* [7]. Arsitektur 3D-CNN dengan *multi-scale convolutional kernels* juga telah menunjukkan peningkatan akurasi melalui ekstraksi fitur *spatiotemporal* pada data EEG [8]. Meskipun topik EEG tidak menjadi fokus utama, keberhasilan CNN dalam berbagai konteks ini menguatkan fleksibilitas arsitektur CNN untuk pengenalan emosi secara umum.



Gambar 1 : Sampel dari FER-2013 *emotion dataset*

Untuk meningkatkan kualitas interaksi manusia-robot yang lebih adaptif, sistem ini tidak hanya mengenali emosi wajah secara *real-time* menggunakan arsitektur CNN yang telah ada, tetapi juga mengintegrasikan fitur rekomendasi video YouTube berbasis emosi. Pendekatan seperti ini sejalan dengan penelitian [9] yang menggunakan sistem rekomendasi berdasarkan kemiripan metrik emosi pengguna, dimana vektor emosi historis disimpan dan dibandingkan untuk menghasilkan rekomendasi yang lebih relevan secara *real-time*.

Penambahan fitur ini bertujuan menghadirkan respons yang sesuai dengan suasana hati pengguna, sehingga interaksi menjadi lebih personal. Sebagai ilustrasi, ketika pengguna menunjukkan ekspresi sedih, sistem dapat merekomendasikan video yang bersifat menghibur untuk meningkatkan suasana hati. Sebaliknya, ketika pengguna tampak bahagia, sistem akan menyarankan video musik yang enerjik agar pengalaman tetap positif dan sesuai konteks emosional.

Kontribusi utama dari sistem ini terletak pada integrasi antara model klasifikasi emosi berbasis CNN menggunakan arsitektur yang diusulkan dengan sistem rekomendasi konten berbasis aturan (*rule-based*). Integrasi ini dirancang untuk meningkatkan kemampuan robotika layanan dalam merespons kondisi emosional pengguna. Dengan demikian, interaksi antara manusia dan mesin menjadi lebih empatik dan personal.

II. METODE PENELITIAN

Penelitian ini bertujuan untuk mengembangkan kerangka kerja berbasis CNN yang mampu mengenali emosi dari ekspresi wajah yang diintegrasikan dengan sistem rekomendasi konten YouTube berdasarkan emosi yang terdeteksi. Sistem ini menggunakan pendekatan *rule-based mapping* untuk menghubungkan kategori emosi dengan jenis konten video YouTube yang relevan. Dua model CNN yang dibandingkan dalam penelitian ini adalah Sequential Fully-CNN dan Mini-Xception.

2.1 Dataset

Dataset yang digunakan dalam penelitian ini adalah FER-2013 (*Facial Expression Recognition 2013*), yang berisi 35.887 citra wajah *grayscale* dengan resolusi 48x48 piksel yang digunakan untuk melatih model dalam mengklasifikasikan tujuh emosi berdasarkan emosi wajah, yaitu netral, senang, sedih, marah, takut, jijik, dan terkejut. *Dataset* ini dibagi menjadi 28.709 *training dataset* dan 7.178 *public test dataset*.

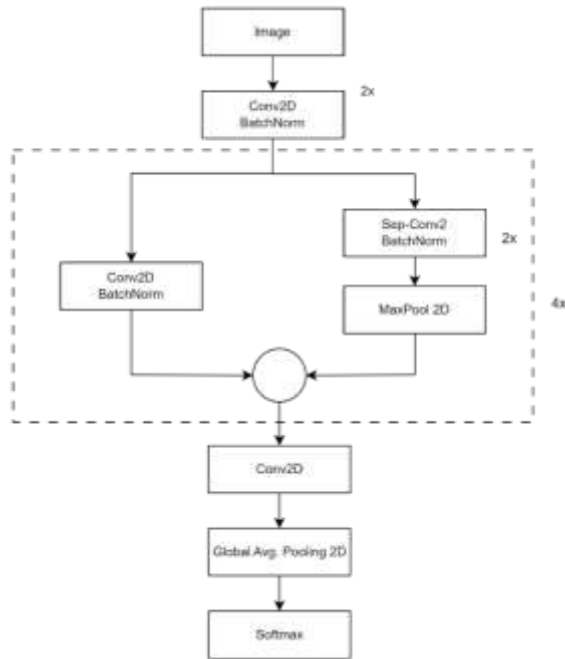
2.2 Model CNN

Dua arsitektur CNN dibandingkan dalam penelitian ini. Pemilihan model didasarkan pada prinsip Occam's Razor, yaitu menghasilkan model yang sederhana namun efektif untuk menangani klasifikasi emosi [10].

2.2.1 Model Pertama: Sequential Fully-CNN

Model pertama menggunakan *Global Average Pooling* (GAP) untuk menghilangkan *Fully Connected Layers*. *Fully Connected Layers* (*activation layers*) biasanya digunakan untuk mengklasifikasikan fitur yang diekstraksi oleh lapisan sebelumnya dalam jaringan dengan menghubungkan setiap neuron dari satu lapisan ke lapisan berikutnya [4]. Dengan menggantikan *Fully Connected Layers* menggunakan GAP, model dapat mengurangi dimensi *feature map* multidimensi menjadi vektor satu dimensi, yang kemudian dipetakan ke setiap kelas yang diharapkan.

Model ini terdiri dari 9 lapisan konvolusi, *ReLU Activation*, *Batch Normalization*, dan *Global Average Pooling* serta *Softmax* sebagai fungsi aktivasi output. Model ini disebut sebagai *Sequential Fully-CNN*.



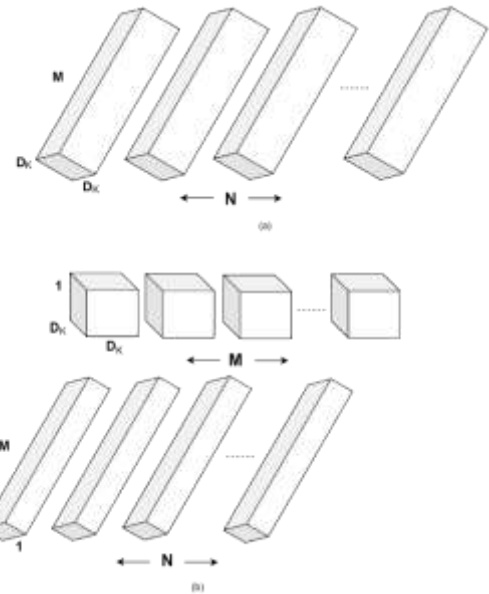
Gambar 2 : Mini-Xception

2.2.2 Model Kedua: Mini-Xception

Model kedua terinspirasi dari arsitektur Xception, yang dikenal karena menggunakan *depth-wise seperable convolutions* dengan *residual connections* untuk meningkatkan efisiensi CNN. Arsitektur Xception memanfaatkan *linear stack* dari *depth-wise seperable convolutions* yang memungkinkan modifikasi arsitektur hanya dengan beberapa baris kode menggunakan Keras atau TensorFlow-Slim [11].

Dalam model ini, *depth-wise seperable convolutions* digunakan untuk mengurangi jumlah parameter dengan membagi proses konvolusi menjadi dua tahap, yaitu *depth-wise convolutions* dan *pointwise convolutions* seperti yang diilustrasikan pada Gambar 3. *Depth-wise convolutions* menerapkan filter $D \times D$ pada setiap saluran masukan secara independen. *Pointwise convolutions* menggabungkan saluran masukan menjadi saluran dengan konvolusi $1 \times 1 \times M$. Pendekatan ini bertujuan untuk memisahkan korelasi spasial dari korelasi saluran, sehingga jumlah parameter dan perhitungan dapat dikurangi secara signifikan [12]. Secara matematis, *residual function* dalam model ini dinyatakan sebagai:

$$H(x) = F(x) + x$$



Gambar 3 : Perbedaan antara (a) *standard convolutions* dan (b) *depth-wise separable convolutions*

Penggunaan *depth-wise seperable convolutions* dalam model ini mengurangi kompleksitas komputasi dengan faktor $\frac{1}{N} + \frac{1}{D^2}$. Penelitian sebelumnya menunjukkan bahwa CNN dengan *multi-scale kernels* dapat menangkap pola lokal dan global dari sinyal EEG, yang menghasilkan prediksi emosi yang lebih akurat [13]. Selain itu, teknik ensembling dengan CNN 1D dan *attention mechanism* juga terbukti meningkatkan performa dalam pengenalan emosi berbasis suara secara *real-time* [14].

Arsitektur final dari model kedua, yang disebut Mini-Xception, terdiri dari 4 konvolusi *residual depth-wise seperable convolutions*, *batch normalization*, dan *ReLU activation function* pada setiap *convolutional layer*. Lapisan terakhir menerapkan GAP dan *softmax activation* untuk klasifikasi. Ilustrasi model ini dapat dilihat pada Gambar 2.

2.3 Alat dan Lingkungan Pengembangan

Pengembangan sistem dilakukan menggunakan perangkat keras berupa *laptop* dengan spesifikasi prosesor Intel Core i5-1135G7 generasi ke-11, memori RAM sebesar 16 GB, serta kartu grafis Intel(R) Iris(R) Xe Graphics yang memiliki VRAM sebesar 128 MB. Sistem operasi yang digunakan adalah Windows 11 64-bit. Bahasa pemrograman yang digunakan adalah Python versi 3.9 yang dijalankan pada *virtual environment* untuk menjaga konsistensi dependensi.

Model klasifikasi emosi dibangun menggunakan *framework* TensorFlow versi 2.18.0 dan Keras versi 3.8.0 yang sudah terintegrasi di dalamnya. Beberapa *library* pendukung juga digunakan, antara lain OpenCV versi 4.11.0.86 untuk pemrosesan citra dan pengambilan video secara *real-time* melalui *webcam*, NumPy versi 1.26.4 untuk pengolahan data numerik, Pandas versi 2.2.2 untuk manipulasi data, Matplotlib versi 3.9.2 untuk visualisasi hasil, serta Scikit-learn versi 1.5.1 untuk evaluasi

performa model menggunakan *confusion matrix*. Sistem juga diintegrasikan dengan YouTube Data API versi 15.0.0 untuk menampilkan rekomendasi konten video berdasarkan emosi pengguna secara *real-time*.

2.4 Pelatihan Model

Pelatihan model dilakukan menggunakan *Adam Optimizer*, yang dipilih karena efisiensinya dalam menyesuaikan *learning rate* secara adaptif untuk setiap parameter. *Hyperparameter* yang digunakan dalam penelitian ini meliputi jumlah *epoch* sebanyak 300, *learning rate* awal sebesar 0.001, *weight decay* sebesar 10^{-6} , dan *batch size* sebanyak 15. *Dataset* FER-2013 dibagi menjadi *training set* untuk pembaruan parameter model dan *validation set* untuk memantau performa selama pelatihan. Selama pelatihan, gambar dari *dataset* pelatihan dimasukkan dalam *batch* ke dalam model, kemudian hasil prediksi dibandingkan dengan label menggunakan *loss function CrossEntropyLoss*. Gradien yang dihasilkan dari perhitungan ini digunakan untuk memperbarui bobot model melalui algoritma *backpropagation*.

Regularisasi diterapkan menggunakan *weight decay* untuk mengurangi risiko *overfitting*, dan *learning rate scheduler* digunakan untuk menyesuaikan *learning rate* berdasarkan performa validasi. Jika tidak ada perbaikan dalam *validation loss* selama 40 *epoch* berturut-turut, *learning rate* akan dikurangi secara otomatis.

Pelatihan model juga didukung dengan implementasi teknik visualisasi *guided backpropagation* secara *real-time* untuk memvalidasi fitur yang dipelajari oleh model. *Guided backpropagation* memungkinkan peneliti mengamati piksel mana dalam citra yang mengaktifkan elemen *higher-level feature map*. Dalam CNN yang menggunakan ReLU sebagai fungsi aktivasi untuk *intermediate layers*, teknik ini mengambil turunan dari setiap elemen (x,y) pada citra input I terhadap elemen (i,j) dari *feature map* f^L di lapisan L . Citra yang direkonstruksi (R) menyaring semua gradien negatif, sehingga hanya gradien positif yang meningkatkan nilai elemen terpilih dari *feature map* yang dipertahankan. Rekonstruksi pada lapisan L dalam CNN berbasis ReLU dirumuskan sebagai [15]:

$$R_{i,j}^L = (R_{i,j}^{L+1} > 0) \cdot R_{i,j}^{L+1}$$

Teknik ini memungkinkan pengamatan terhadap piksel-piksel pada citra wajah yang mengaktifkan *higher-level feature map*, seperti bagian alis, mulut, atau mata. Selama pelatihan, metrik seperti *train loss*, *validation loss*, *accuracy*, *precision*, dan *recall* dicatat menggunakan TensorBoard untuk memantau perkembangan model. Hasil pelatihan disimpan dalam bentuk *checkpoint* setiap *epoch* untuk memastikan keberlanjutan proses jika terjadi gangguan.

2.5 Evaluasi Model

Evaluasi dilakukan pada *test set* FER-2013 untuk mengukur performa model dalam klasifikasi emosi secara *real-time*. Model terbaik dari hasil pelatihan dievaluasi

menggunakan metrik seperti *accuracy*, *precision*, *recall*, dan *confusion matrix*. Visualisasi *guided backpropagation* digunakan selama evaluasi untuk memvalidasi apakah model benar-benar mempelajari fitur wajah yang relevan dalam pengambilan keputusan. Proses ini memastikan bahwa model mampu mengklasifikasikan emosi secara andal dan efisien.

2.6 Integrasi Sistem Rekomendasi Konten YouTube

Setiap emosi pengguna dipetakan ke kategori video tertentu menggunakan *rule-based mapping*. Misalnya, emosi "sedih" dapat dipetakan ke video motivasi atau hiburan, sedangkan "marah" dipetakan ke video yang menenangkan. Setelah emosi terdeteksi melalui model CNN, sistem secara otomatis memilih dan memutar video dari daftar tautan YouTube yang telah ditentukan. Sistem juga menampilkan *prompt* interaktif kepada pengguna untuk memberikan kontrol penuh terhadap video yang akan diputar. Integrasi ini dirancang agar sistem tidak hanya mendeteksi emosi secara *real-time*, tetapi juga memberikan respons yang sesuai melalui konten yang relevan.

III. HASIL PENELITIAN

3.1 Hasil Evaluasi Model

Hasil evaluasi menunjukkan bahwa model Mini-Xception mampu mengklasifikasikan emosi wajah dengan cukup baik. Tabel 1 menampilkan metrik performa model yang mencakup *validation loss*, *accuracy*, *precision*, dan *recall*.

Tabel 1: Metrik Performa Model

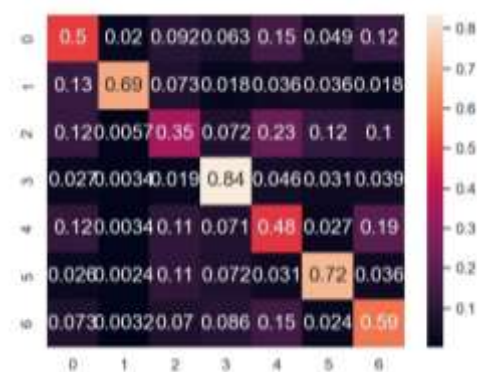
Metrik	Nilai
<i>Validation Loss</i>	1,134
<i>Accuracy</i>	0,602
<i>Precision</i>	0,587
<i>Recall</i>	0,596
Total Parameter	56.951

Nilai *validation loss* sebesar 1.134 mengindikasikan tingkat kesalahan prediksi yang moderat. *Accuracy* menunjukkan bahwa sekitar 60,2% dari seluruh sampel berhasil diklasifikasikan dengan benar. Selanjutnya, nilai *precision* menggambarkan bahwa 58,7% dari prediksi positif model Mini-Xception adalah benar, sementara *recall* mengindikasikan bahwa 59,6% dari semua kasus positif berhasil diidentifikasi oleh model Mini-Xception. Model Mini-Xception diimplementasikan dengan arsitektur yang efisien yang hanya membutuhkan total 56.951 parameter.

Berdasarkan hasil evaluasi, model Mini-Xception dipilih dalam penelitian ini karena memiliki jumlah parameter yang jauh lebih kecil, yaitu 56.951, dibandingkan dengan Sequential Fully-CNN. Sequential Fully-CNN memiliki sekitar 600.000 parameter dengan *accuracy* 66% pada dataset FER-2013 [4]. Meskipun *accuracy* Mini-Xception (60,2%) lebih rendah dibandingkan Sequential Fully-CNN, pengurangan

jumlah parameter hingga hampir 10 kali lipat memberikan keuntungan dalam efisiensi komputasi. Dengan kompleksitas yang lebih rendah, Mini-Xception lebih ringan dan cepat, sehingga lebih sesuai untuk aplikasi yang membutuhkan efisiensi tanpa mengorbankan terlalu banyak *accuracy*.

Untuk memahami lebih lanjut performa model Mini-Xception dalam mengenali setiap emosi, Gambar 1 menyajikan *Normalized Confusion Matrix* yang menampilkan persentase keberhasilan dan kesalahan klasifikasi untuk setiap kategori emosi. *Heatmap* ditampilkan dengan angka yang mewakili tujuh label emosi, yaitu 0 (marah), 1 (jijik), 2 (takut), 3 (senang), 4 (sedih), 5 (terkejut) dan 6 (netral).



Gambar 4 : *Normalized Confusion Matrix*

Berdasarkan Gambar 1, model Mini-Xception menunjukkan performa terbaik dalam mengenali emosi "senang", "terkejut", dan "jijik" dengan tingkat akurasi masing-masing sekitar 84%, 72%, dan 69%. Emosi "netral" dan "marah" memiliki tingkat akurasi sedang, masing-masing sekitar 59% dan 50%. Performa terendah terlihat pada klasifikasi emosi "sedih" dan "takut" dengan akurasi masing-masing sekitar 48% dan 35%. Selain itu, beberapa pola kesalahan klasifikasi juga terungkap. Contohnya, sebanyak 23% emosi "sedih" salah diklasifikasikan sebagai "takut" dan 19% emosi "netral" dikategorikan sebagai "sedih".

3.2 Hasil Implementasi

Penelitian ini diperluas dengan menambahkan rekomendasi video YouTube berdasarkan deteksi emosi, dengan harapan dapat meningkatkan pengalaman pengguna melalui pemberian konten multimedia yang sesuai dengan emosi yang terdeteksi. Beberapa URL video YouTube dikumpulkan dan diklasifikasikan ke dalam lima kategori, yaitu *Energetic*, *Funny*, *Motivational*, *Adorable*, dan *Calming*. Selanjutnya, setiap kategori emosi dicocokkan dengan tipe-tipe video yang dianggap dapat meningkatkan suasana hati. Pembagian tersebut meliputi:

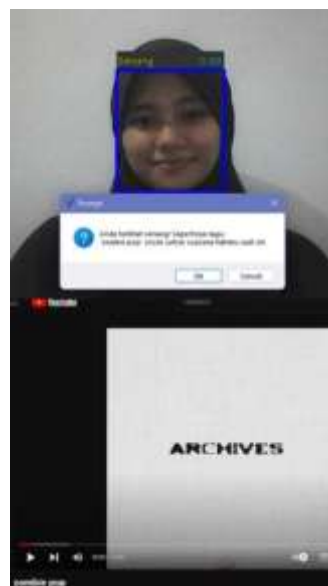
- Senang: *Energetic*
- Sedih: *Funny*, *Motivational*, *Adorable*
- Jijik: *Adorable*, *Funny*
- Marah: *Calming*, *Adorable*
- Takut: *Motivational*, *Calming*
- Terkejut: *Calming*

- Netral: *Energetic*, *Funny*, *Motivational*, *Adorable*, *Calming*

Setiap emosi wajah memiliki interpretasi berbeda yang menentukan rekomendasi konten. Emosi "senang" dikaitkan dengan perasaan gembira yang positif, sehingga sistem merekomendasikan konten energik untuk mempertahankan suasana hati positif dan meningkatkan semangat. Sebaliknya, "sedih" diinterpretasikan sebagai perasaan kecewa atau kehilangan, mendorong sistem merekomendasikan konten yang menghibur dan memberikan dukungan emosional. Pada emosi "takut" yang didefinisikan sebagai kekhawatiran akan sesuatu, sistem merekomendasikan video motivasi untuk memberikan dukungan positif. Saat mengidentifikasi emosi "marah" yang menunjukkan ketidakpuasan atau emosi intens, sistem menyajikan konten untuk membantu mengelola emosi dan menenangkan diri.

Untuk emosi "jijik" yang mencerminkan ketidaknyamanan terhadap sesuatu, sistem menyarankan video jenaka atau menggemaskan untuk mengalihkan perhatian dari stimulus yang memicu rasa jijik. Emosi "terkejut" diinterpretasikan sebagai respons spontan terhadap hal tak terduga, sehingga sistem merekomendasikan konten untuk meredakan keterkejutan. Sementara itu, pada emosi "netral" yang mencerminkan keadaan emosional yang stabil, sistem merekomendasikan semua kategori konten untuk memberikan variasi pilihan sesuai preferensi pengguna.

Hasil klasifikasi emosi dan rekomendasi video YouTube secara *real-time* ditunjukkan pada Gambar 5, Gambar 6, Gambar 7, Gambar 8, Gambar 9, Gambar 10, dan Gambar 11. Pada setiap gambar, bagian atas menampilkan hasil deteksi emosi pengguna beserta judul konten YouTube yang direkomendasikan berdasarkan emosi terdeteksi, sementara bagian bawah menampilkan halaman konten YouTube yang muncul setelah pengguna menekan tombol "Ok".



Gambar 5 : Deteksi emosi "senang"



Gambar 6 : Deteksi emosi "sedih"



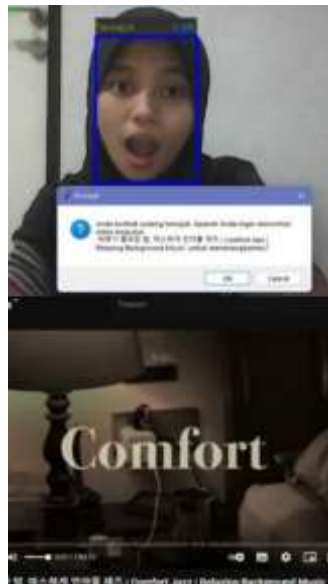
Gambar 7 : Deteksi emosi "jijik"



Gambar 8 : Deteksi emosi "marah"



Gambar 9 : Deteksi emosi "takut"



Gambar 10 : Deteksi emosi "terkejut"



Gambar 11 : Deteksi emosi "netral"

3.3 Pengujian Sistem

Pengujian sistem dilakukan untuk mengevaluasi akurasi klasifikasi emosi dan kesesuaian rekomendasi video YouTube yang dihasilkan berdasarkan emosi yang terdeteksi. Dalam pengujian ini, lima partisipan diminta untuk menampilkan ekspresi wajah yang sesuai dengan emosi yang telah ditentukan sebagai target emosi. Dari itu didapatkan emosi terdeteksi, nilai *confidence*, dan kategori video yang dipetakan berdasarkan emosi.

Tabel 2: Hasil Pengujian pada Emosi “Senang”

Partisipan	Emosi Target	Emosi Terdeteksi	Conf	Kategori Video
P1	Senang	Senang	1.0	<i>Energetic</i>
P2	Senang	Senang	0.73	<i>Energetic</i>
P3	Senang	Senang	0.99	<i>Energetic</i>
P4	Senang	Senang	1.0	<i>Energetic</i>
P5	Senang	Senang	0.99	<i>Energetic</i>

Berdasarkan hasil pada Tabel 2, sistem berhasil mendeteksi emosi “senang” dengan tingkat *confidence* yang tinggi pada seluruh partisipan. Seluruh hasil deteksi sesuai dengan target emosi dan rekomendasi video yang diberikan sesuai dengan kategori yang telah ditentukan untuk emosi “senang”, yaitu *Energetic*. Hasil ini menunjukkan bahwa sistem mampu mengidentifikasi emosi “senang” dengan akurat dan memberikan rekomendasi yang relevan.

Tabel 3: Hasil Pengujian pada Emosi “Sedih”

Partisipan	Emosi Target	Emosi Terdeteksi	Conf	Kategori Video
P1	Sedih	Sedih	0.78	<i>Funny</i>
P2	Sedih	Sedih	0.43	<i>Funny</i>
P3	Sedih	Sedih	0.48	<i>Motivatio nal</i>
P4	Sedih	Sedih	0.86	<i>Motivatio nal</i>
P5	Sedih	Sedih	0.74	<i>Funny</i>

Berdasarkan hasil pada Tabel 3, sistem berhasil mendeteksi emosi “sedih” dengan tingkat *confidence* yang bervariasi. Nilai *confidence* terendah tercatat pada partisipan P2 (0.43) dan P3 (0.48), sedangkan nilai tertinggi tercatat pada partisipan P4 (0.86). Di samping itu, sistem merekomendasikan video dalam kategori *Funny* dan *Motivational*. Kategori video tersebut sesuai dengan pemetaan yang telah ditentukan sebelumnya. Hasil ini menunjukkan bahwa sistem mampu mengidentifikasi ekspresi sedih dengan cukup baik serta tetap dapat memberikan rekomendasi yang relevan.

Tabel 4: Hasil Pengujian pada Emosi “Jijik”

Partisipan	Emosi Target	Emosi Terdeteksi	Conf	Kategori Video
P1	Jijik	Senang	0.77	<i>Energetic</i>
P2	Jijik	Netral	0.59	<i>Calming</i>

P3	Jijik	Netral	0.46	<i>Energetic</i>
P4	Jijik	Jijik	0.75	<i>Funny</i>
P5	Jijik	Jijik	0.5	<i>Funny</i>

Berdasarkan hasil pada Tabel 4, sistem hanya berhasil mendeteksi emosi yang sesuai dengan target pada dua partisipan, yaitu P4 dan P5, dengan nilai *confidence* masing-masing sebesar 0.75 dan 0.5. Sementara itu, hasil deteksi pada partisipan P1, P2, dan P3 menunjukkan ketidaksesuaian antara emosi target dan emosi yang terdeteksi, yaitu “senang” atau “netral”. Dari segi rekomendasi video, sistem berhasil merekomendasikan video yang sesuai untuk emosi “jijik”, yaitu video pada kategori *Funny*. Meskipun deteksi emosi pada beberapa partisipan tidak sesuai, sistem mampu memberikan rekomendasi yang relevan dengan emosi yang terdeteksi, yaitu kategori *Energetic* untuk emosi “senang” dan “netral” serta kategori *Calming* untuk emosi “netral”. Hasil ini menunjukkan bahwa sistem tepat dalam menyesuaikan rekomendasi video berdasarkan emosi yang terdeteksi, namun masih perlu peningkatan akurasi dalam mengenali emosi “jijik” secara tepat.

Tabel 5: Hasil Pengujian pada Emosi “Marah”

Partisipan	Emosi Target	Emosi Terdeteksi	Conf	Kategori Video
P1	Marah	Marah	0.6	<i>Adorable</i>
P2	Marah	Marah	0.77	<i>Calming</i>
P3	Marah	Marah	0.81	<i>Calming</i>
P4	Marah	Marah	0.74	<i>Adorable</i>
P5	Marah	Marah	0.55	<i>Adorable</i>

Berdasarkan hasil pada Tabel 5, sistem berhasil mendeteksi emosi “marah” dengan nilai *confidence* yang bervariasi dengan seluruh nilai berada di atas angka 0.5. Nilai terendah tercatat pada partisipan P5 (0.55) dan nilai tertinggi tercatat pada partisipan P3 (0.81). Seluruh hasil deteksi sesuai dengan target emosi dan rekomendasi video sesuai dengan kategori yang telah ditentukan untuk emosi “marah”, yaitu *Adorable* dan *Calming*. Hasil ini menunjukkan bahwa sistem mampu mengidentifikasi emosi “marah” dengan tingkat akurasi yang cukup baik serta memberikan rekomendasi konten yang relevan.

Tabel 6: Hasil Pengujian pada Emosi “Takut”

Partisipan	Emosi Target	Emosi Terdeteksi	Conf	Kategori Video
P1	Takut	Takut	0.69	<i>Motivation al</i>
P2	Takut	Takut	0.56	<i>Calming</i>
P3	Takut	Takut	0.84	<i>Motivation al</i>
P4	Takut	Senang	0.81	<i>Energetic</i>
P5	Takut	Netral	0.65	<i>Energetic</i>

Berdasarkan hasil pada Tabel 6, sistem berhasil mendeteksi emosi “takut” dengan tepat pada tiga partisipan, yaitu P1, P2, dan P3 dengan nilai *confidence*

bervariasi di atas angka 0.55. Sementara itu, pada partisipan P4 dan P5, sistem salah mendeteksi emosi menjadi “senang” dan “netral” dengan nilai *confidence* yang relatif tinggi, yaitu 0.81 dan 0.65. Dari segi rekomendasi video, sistem memberikan rekomendasi video yang tepat untuk emosi yang tepat dan keliru, yaitu video *Motivational* dan *Calming* untuk emosi “takut” serta video *Energetic* untuk emosi “senang” dan “netral”. Hasil ini menunjukkan bahwa sistem sudah tepat dalam menyesuaikan rekomendasi video berdasarkan emosi yang terdeteksi, namun masih perlu peningkatan akurasi dalam mengenali emosi “takut” secara tepat.

Tabel 7: Hasil Pengujian pada Emosi “Terkejut”

Partisipan	Emosi Target	Emosi Terdeteksi	Conf	Kategori Video
P1	Terkejut	Terkejut	0.98	<i>Calming</i>
P2	Terkejut	Terkejut	0.93	<i>Calming</i>
P3	Terkejut	Terkejut	0.89	<i>Calming</i>
P4	Terkejut	Terkejut	0.76	<i>Calming</i>
P5	Terkejut	Terkejut	0.73	<i>Calming</i>

Berdasarkan hasil pada Tabel 7, sistem berhasil mendeteksi emosi “terkejut” dengan seluruh nilai *confidence* di atas angka 0.7. Seluruh hasil deteksi sesuai dengan target emosi dan rekomendasi video sesuai dengan kategori yang telah ditentukan untuk emosi “terkejut”, yaitu *Calming*. Hasil ini menunjukkan bahwa sistem mampu mengidentifikasi emosi “terkejut” dengan tingkat akurasi tinggi serta memberikan rekomendasi konten yang relevan.

Tabel 8: Hasil Pengujian pada Emosi “Netral”

Partisipan	Emosi Target	Emosi Terdeteksi	Conf	Kategori Video
P1	Netral	Netral	0.86	<i>Adorable</i>
P2	Netral	Netral	0.9	<i>Calming</i>
P3	Netral	Netral	0.93	<i>Energetic</i>
P4	Netral	Netral	0.9	<i>Funny</i>
P5	Netral	Netral	0.9	<i>Energetic</i>

Berdasarkan hasil pada Tabel 8, sistem berhasil mendeteksi emosi “netral” dengan nilai *confidence* sangat tinggi, mayoritas di atas nilai 0.85. Seluruh hasil deteksi sesuai dengan target emosi dan rekomendasi video sesuai dengan kategori yang telah ditentukan untuk emosi “netral”. Hasil ini menunjukkan bahwa sistem mampu mengidentifikasi emosi “netral” dengan tingkat akurasi yang tinggi serta memberikan rekomendasi konten yang relevan.

Berdasarkan hasil pengujian terhadap lima responden, sistem berhasil mendeteksi berbagai emosi dengan tingkat akurasi yang bervariasi. Deteksi pada emosi “senang”, “terkejut”, dan “netral” menunjukkan nilai *confidence* yang sangat tinggi dengan seluruh deteksi sesuai target. Di samping itu, emosi “sedih” dan “marah” juga terdeteksi dengan nilai *confidence* yang cukup tinggi

dengan seluruh hasil deteksinya sesuai dengan emosi yang ditargetkan. Namun, sistem masih memerlukan peningkatan dalam mendeteksi emosi “jijik” dan “takut” karena rata-rata nilai *confidence*-nya belum cukup tinggi serta masih terdapat beberapa kesalahan deteksi yang mengarah pada emosi “senang” dan “netral”. Dari segi rekomendasi video, sistem telah berhasil memberikan rekomendasi yang sesuai dengan emosi yang terdeteksi.

IV. PEMBAHASAN

4.1 Analisis Hasil Evaluasi Model

Berdasarkan hasil evaluasi yang telah dilakukan, model Mini-Xception menunjukkan kemampuan yang cukup baik dalam mengklasifikasikan emosi wajah dari dataset FER-2013. Dengan nilai *accuracy* sebesar 60,2%, model ini mampu mengidentifikasi emosi dengan tingkat keberhasilan yang moderat. Dari segi jumlah parameter, model Mini-Xception hanya memiliki 56.951 parameter, yang jauh lebih kecil dibandingkan model Sequential Fully-CNN yang memiliki sekitar 600.000 parameter. Hal ini menunjukkan bahwa model Mini-Xception memiliki efisiensi komputasi yang lebih baik dengan tetap mempertahankan performa yang kompetitif. Pengurangan jumlah parameter ini dicapai melalui penggunaan *depth-wise separable convolutions* yang mampu mengurangi kompleksitas komputasi tanpa mengorbankan kemampuan ekstraksi fitur yang signifikan.

4.2 Implikasi dari Hasil Penelitian

Berdasarkan analisis di atas, hasil ini menunjukkan bahwa Mini-Xception sangat potensial untuk diterapkan dalam skenario nyata, terutama di lingkungan dengan keterbatasan perangkat keras seperti perangkat *mobile*. Efisiensi parameter memungkinkan proses deteksi emosi dilakukan secara real-time, tanpa membutuhkan sumber daya komputasi tinggi.

Kontribusi ilmiah dari sistem ini adalah pada integrasi sistem rekomendasi konten YouTube berbasis *rule-based mapping*. Dengan memetakan hasil klasifikasi emosi ke jenis konten tertentu, sistem mampu meningkatkan personalisasi pengalaman pengguna secara emosional. Hal ini membuka peluang pengembangan aplikasi serupa di berbagai domain seperti terapi digital hingga hiburan berbasis afeksi.

Namun demikian, pendekatan *rule-based* memiliki keterbatasan konseptual karena tidak dapat menangkap nuansa emosi yang lebih kompleks atau perubahan emosional yang bersifat dinamis. Ini menunjukkan bahwa kontribusi sistem ini lebih bersifat *foundational*, yang memberikan dasar awal bagi pengembangan sistem cerdas yang lebih adaptif. Integrasi dengan metode *machine learning* berbasis *reinforcement learning* atau *deep recommendation systems* akan sangat menjanjikan untuk menciptakan sistem yang benar-benar responsif terhadap dinamika emosi pengguna.

4.3 Keterbatasan dan Potensi Pengembangan

Terdapat beberapa keterbatasan utama yang perlu dikaji untuk pengembangan lebih lanjut. Pertama, sistem *rule-based recommendation* yang digunakan tidak mampu menyesuaikan diri terhadap perubahan emosi yang bersifat temporal dan kontekstual. Oleh karena itu, pendekatan ini masih terbatas dalam memberikan rekomendasi yang benar-benar *personalized*. Integrasi metode *reinforcement learning*, yang mampu melakukan pemetaan adaptif antara emosi dan jenis konten berdasarkan umpan balik pengguna, merupakan arah pengembangan yang bernilai secara ilmiah dan praktis.

Kedua, dari sisi akurasi klasifikasi emosi, model masih rentan terhadap gangguan lingkungan, seperti pencahayaan yang buruk, ekspresi wajah yang tidak jelas, sudut wajah yang tidak frontal, serta kualitas citra rendah. Faktor-faktor ini menurunkan reliabilitas sistem dalam skenario penggunaan nyata. Oleh karena itu, pengembangan lebih lanjut dengan *dataset* yang lebih representatif, mencakup variasi kondisi pencahayaan, ras, usia, dan ekspresi yang lebih beragam, akan memperkuat kemampuan *generalization* model secara signifikan.

Pengembangan lebih lanjut dapat diarahkan pada pendekatan seperti yang digunakan pada penelitian [15], yaitu sistem rekomendasi berdasarkan kemiripan metrik emosi pengguna. Dengan menyimpan dan membandingkan vektor emosi historis, sistem dapat memberikan rekomendasi yang lebih adaptif dan relevan secara *real-time* sesuai dengan kondisi emosional pengguna.

V. KESIMPULAN

Berdasarkan hasil implementasi dan evaluasi yang telah dilakukan, aplikasi yang dibangun mampu untuk mendeteksi emosi wajah senang, sedih, marah, jijik, terkejut, takut, dan netral dengan menggunakan model Mini-Xception yang telah diuji pada *dataset* FER-2013. Model deteksi emosi wajah ini tidak hanya mampu mengenali emosi pengguna, tetapi juga telah diintegrasikan dengan fitur rekomendasi video YouTube berbasis *rule-based mapping* yang menyesuaikan konten dengan emosi pengguna sehingga memberikan pengalaman yang lebih interaktif dan personal.

Namun, pendekatan *rule-based* masih memiliki keterbatasan karena kurang responsif terhadap perubahan emosi wajah manusia yang bersifat dinamis dan kompleks. Selain itu, faktor lingkungan seperti pencahayaan, sudut pengambilan gambar, kualitas kamera, dan variasi ekspresi wajah individu juga mempengaruhi akurasi deteksi emosi. Sebagai saran untuk penelitian selanjutnya, pengembangan dapat dilakukan dengan meningkatkan arsitektur model atau menggunakan arsitektur yang lebih kuat, memperluas variasi *dataset* yang digunakan, menerapkan pendekatan berbasis *reinforcement learning*, dan melakukan implementasi serta pengujian di perangkat *mobile* atau sistem berbasis *edge computing*.

VI. REFERENSI

- [1] A. M. Abdulazeez ja Z. S. Ageed, "Face Emotion Recognition Based on Machine Learning: A Review," *International Journal of Informatics Information System and Computer Engineer*, p. 35, 2024.
- [2] S. A. Fatima, A. Kumar ja S. S. Raoof, "Real Time Emotion Detection of Humans Using Mini-Xception Algorithm," *IOP Conference Series: Materials Science and Engineering*, p. 9, 2021.
- [3] S. Zhang, Y. Ding, Z. Wei ja C. Guan, "Continuous Emotion Recognition with Audio-visual Leader-follower Attentive Fusion," *osa/vuosik*. 2, p. 8, 2021.
- [4] O. Arriaga, M. Valdenegro-Toro ja P. Plöger, "Real-time Convolutional Neural Networks for Emotion and Gender Classification," 2017.
- [5] R. Goel, S. Susan, S. Vashisht ja A. Dhanda, "Emotion-Aware Transformer Encoder for Empathetic Dialogue Generation," *International Conference on Affective Computing and Intelligent Interaction Workshops and Demos (ACIIW)*, p. 6, 2019.
- [6] S. Akter, R. A. Prodhan, T. S. Pias ja D. Eisenberg, "M1M2: Deep-Learning-Based Real-Time Emotion Recognition from Neural Activity," *Sensors*, p. 27, 2022.
- [7] S. Minaee, M. Minaee ja A. Abdolrashidi, "Deep-Emotion: Facial Expression Recognition Using Attentional Convolutional Network," *Computer Vision and Pattern Recognition*, *osa/vuosik*. 1, p. 8, 2019.
- [8] Y. Su, Z. Zhang, X. Li, B. Zhang ja H. Ma, "The multiscale 3D convolutional network for emotion recognition based on electroencephalogram," *Frontiers in Neuroscience*, p. 14, 2022.
- [9] D. N. Amali, A. R. Barakbah ja A. R. A. Besari, "Semantic Video Recommendation System Based on Video Viewers Impression from Emotion Detection," *2018 International Electronics Symposium on Knowledge Creation and Intelligent Computing (IES-KCIC)*, 2018.
- [10] D. P. Kingma ja J. L. Ba, "Adam: A Method for Stochastic Optimization," *ICLR*, *osa/vuosik*. 9, p. 15, 2015.
- [11] F. Chollet, "Xception: Deep Learning with Depthwise Separable Convolutions," *osa/vuosik*. 3, p. 8, 2017.
- [12] L. S, "Facial Expression Recognition based on Mini_Xception," *Highlights in Science, Engineering and Technology*, *osa/vuosik*. 39, p. 10, 2023.
- [13] T.-D.-T. Phan, S.-H. Kim, H.-J. Yang ja G.-S. Lee, "EEG-Based Emotion Recognition by Convolutional Neural Network with Multi-Scale Kernels," *Sensors*, p. 13, 2021.

-
- [14] W. Alsabhan, "Human-Computer Interaction with a Real-Time Speech Emotion Recognition with Ensembling Techniques 1D Convolution Neural Network and Attention," *Sensors*, osa/vuosik. 2, p. 21, 2023.
- [15] J. T. Springenberg, A. Dosovitskiy, T. Brox ja M. Riedmiller, "Striving for Simplicity: The All Convolutional Net," 2014.