



Pengembangan Aplikasi Mobile untuk Deteksi Ekspresi Wajah Menggunakan CNN dan Rekomendasi Musik

Larasati Puspita Candra Dewi¹, Rizky Arifiansyah^{2*} [Times New Roman 10] *for the corresponding author

¹Department, Faculty, Institution, City, Country [Times New Roman 9]

²Department, Faculty, Institution, City, Country [Times New Roman 9]

¹email, ²email, [Times New Roman 9]

Abstract

Gambar wajah manusia mengandung banyak informasi, salah satunya adalah ekspresi wajah yang bisa berkaitan dengan kondisi mental dan suasana hati seseorang. Emosi seperti bahagia, sedih, marah, tenang, takut, dan lain-lain secara alami tercermin melalui ekspresi wajah. Kemampuan untuk mendeteksi emosi manusia melalui ekspresi wajah memiliki fungsi yang penting untuk mengetahui suasana hati dan kondisi seseorang. Selain itu, musik sering digunakan sebagai bentuk ekspresi emosi atau pelampiasan perasaan, menjadikannya sarana yang kuat dalam mengelola suasana hati. Oleh karena itu penelitian ini bertujuan untuk merancang dan mengembangkan aplikasi mobile yang dapat mendeteksi ekspresi wajah menggunakan CNN dan juga memberikan rekomendasi musik yang sesuai dengan ekspresi tersebut. Pendekatan pengembangan dilakukan dengan menggunakan model Convolutional Neural Network (CNN) yang dilatih menggunakan dataset CK+ yang telah difilter, hanya mencakup empat jenis emosi, yaitu senang, marah, sedih, dan netral. Evaluasi model dilakukan menggunakan data uji yang mencakup gambar-gambar di luar dataset untuk mengukur kinerjanya secara lebih luas. Sistem memberikan rekomendasi musik yang sesuai berdasarkan ekspresi wajah yang telah dideteksi oleh model CNN. Model mencapai akurasi pelatihan sebesar 92% dan akurasi validasi sebesar 85%. Hasil ini mengindikasikan bahwa model mampu mempelajari fitur-fitur dataset secara efektif sambil tetap menjaga generalisasi. Namun, hasil menunjukkan adanya tantangan terkait dengan ketidakseimbangan kelas, di mana metrik seperti precision, recall, dan F1-score pada kelas tertentu seperti 'neutral' dan 'sadness' menunjukkan nilai yang lebih rendah. Penelitian ini menunjukkan bahwa model CNN dapat mendeteksi ekspresi wajah dengan akurasi yang bervariasi, memberikan hasil yang memadai pada data uji, termasuk gambar di luar dataset. Temuan ini membuka peluang untuk aplikasi lebih lanjut dalam bidang pengelolaan suasana hati melalui deteksi emosi wajah.

Keywords: Deteksi Emosi, Ekspresi Wajah, Rekomendasi Musik, Convolutional Neural Network (CNN)

How to Cite: [Caption completed by the editor]

DOI:

1. Pendahuluan

Semua orang setuju bahwa gambar wajah manusia mengandung informasi tentang fitur tertentu, seperti usia, jenis kelamin, emosi, etnis, dan ras [1]. Setiap orang memiliki berbagai jenis emosi yang bisa dilihat melalui gerakan tubuh dan ekspresi wajah. Emosi adalah perasaan dan pikiran yang unik, serta merupakan kondisi fisik dan mental yang mendorong seseorang untuk bereaksi [2]. Emosi seperti bahagia, sedih, marah, tenang, takut, dan lain-lain secara alami tercermin melalui ekspresi wajah. Ekspresi emosi dapat dikenali dari raut wajah yang berubah, kerutan pada kening, kedipan mata, maupun perubahan warna pada kulit wajah [3]. Seiring perkembangan teknologi, salah satu kemajuan penting adalah kemampuan untuk mendeteksi emosi manusia. Kemampuan menentukan keadaan emosional manusia adalah hal yang unik dan

istimewa yang dapat digunakan sebagai dasar untuk semua jenis pengenalan emosi [4].

Teknologi pendeteksi emosi wajah telah berkembang pesat berkat penggunaan algoritma *machine learning*, khususnya *Convolutional Neural Network* (CNN). CNN adalah mesin pembelajaran paling populer yang model pembelajarannya ditujukan untuk klasifikasi langsung data dua dimensi, seperti gambar, video, teks, atau suara [5]. Metode pengenalan emosi manusia berdasarkan ekspresi wajah atau face detection dan peningkatan kemampuan perangkat keras telah mendorong banyak penelitian untuk melakukan pendeteksian wajah dan pengenalan ekspresi wajah [6]. Beberapa penelitian telah menunjukkan bahwa sistem yang menggunakan CNN mampu mendeteksi ekspresi wajah dengan akurasi yang sangat tinggi, memungkinkan teknologi ini diterapkan secara luas

dalam berbagai aplikasi. [7], [8], [9]. Namun, beberapa tantangan tetap ada, seperti kebutuhan akan data latih besar, kompleksitas pemrosesan data, variabilitas ekspresi antar individu, serta kondisi pencahayaan yang memengaruhi akurasi deteksi [8]

Selain itu, musik dikenal memiliki dampak besar pada ekspresi dan suasana hati seseorang [10]. Musik berperan sebagai media ekspresi emosional dan dapat meningkatkan kesejahteraan psikologis dan membantu mengelola suasana hati [11]. Penelitian sebelumnya menunjukkan bahwa pemilihan musik yang sesuai dengan keadaan emosional dapat membantu meningkatkan suasana hati, mengurangi stres, atau meningkatkan fokus [10].

Dengan demikian, tujuan dari penelitian ini adalah untuk merancang dan mengembangkan aplikasi mobile yang dapat mendeteksi ekspresi wajah menggunakan CNN dan juga memberikan rekomendasi musik yang sesuai dengan ekspresi tersebut. Hasil dari penelitian ini diharapkan dapat meningkatkan akurasi dan efisiensi dalam mendeteksi ekspresi wajah secara real-time, sehingga menciptakan pengalaman yang lebih personal dan interaktif bagi pengguna.

2. Metode Penelitian

2.1 Pengumpulan Dataset

Data penelitian ini dikumpulkan menggunakan dataset CK+, yang dapat diakses secara terbuka melalui platform Kaggle dan merupakan salah satu dataset yang paling banyak digunakan untuk penelitian yang berfokus pada pengenalan emosi berbasis gambar wajah. dataset CK+ terdiri dari serangkaian gambar wajah yang dikumpulkan dari berbagai individu yang memiliki ekspresi emosi yang berbeda. Terdapat 8 ekspresi dari dataset CK+ yaitu marah, jijik, takut, senang, sedih, terkejut, netral, mengejek. Namun, pada penelitian ini kami hanya menggunakan 4 ekspresi yaitu marah, senang, sedih, dan netral.

Tabel 1. Jumlah Dataset

Ekspresi	Jumlah Gambar
Marah	135
Senang	207
Netral	54
Sedih	84

2.2 Preprocessing

Faktor yang paling signifikan dan berpengaruh dalam kinerja generalisasi algoritma pembelajaran mesin adalah *preprocessing* data [12]. Pada penelitian ini, *preprocessing* data dilakukan untuk memastikan bahwa model dapat mempelajari pola dari dataset dengan cara yang efektif dan menghindari potensi masalah seperti data yang tidak seimbang yang ditunjukkan pada Tabel 1. Langkah-langkah *preprocessing* yang dilakukan adalah sebagai berikut:

1. Normalisasi Gambar: Setiap gambar pada dataset dinormalisasi dengan membagi nilai pikselnya dengan 255 sehingga berada dalam rentang [0, 1]. Hal ini dilakukan untuk mempercepat proses pelatihan model karena skala piksel yang lebih kecil membantu algoritme konvergen lebih cepat.

2. Resize Gambar: Semua gambar diubah ukurannya menjadi 128×128 piksel untuk memastikan konsistensi dimensi masukan model. Resolusi ini dipilih untuk menjaga keseimbangan antara kecepatan komputasi dan representasi visual yang cukup detail.

3. Augmentasi Data: Augmentasi data diterapkan untuk memperluas variasi dataset yang tersedia dan meningkatkan kemampuan generalisasi model. Teknik augmentasi yang digunakan dalam penelitian ini meliputi rotasi gambar untuk mencerminkan perbedaan orientasi wajah, perubahan skala (zoom) untuk menangkap variasi ukuran wajah, serta flip horizontal untuk menambah variasi posisi wajah secara lateral. Dengan penerapan teknik-teknik ini, model diharapkan dapat mengenali pola emosi dengan lebih baik, bahkan pada kondisi wajah yang memiliki perbedaan sudut dan ukuran yang signifikan.

4. Pembagian Dataset: Dataset dibagi menjadi dua subset menggunakan parameter `validation_split = 0.2` pada `ImageDataGenerator`, yaitu data pelatihan dan data validasi. Data pelatihan, yang mencakup 80% dari keseluruhan dataset, digunakan untuk melatih model agar dapat mengenali pola dan fitur pada gambar. Sementara itu, data validasi, yang terdiri dari 20% dari dataset, digunakan untuk mengevaluasi performa model selama proses pelatihan. Pembagian ini memastikan bahwa model dapat diuji terhadap data yang tidak dilihat selama pelatihan, sehingga menghindari masalah *overfitting* dan meningkatkan kemampuan model dalam menggeneralisasi hasil pada data yang belum pernah ditemui sebelumnya.

5. Label: Label pada dataset diubah menjadi format numerik menggunakan `class_indices` dari `ImageDataGenerator`, agar bisa diproses oleh model dalam format one-hot encoding. Label ini kemudian disimpan dalam file `label_map.pkl`, yang berguna untuk evaluasi dan prediksi di tahap berikutnya.

Dengan menerapkan langkah-langkah preprocessing ini, dataset yang digunakan menjadi lebih siap untuk dimasukkan ke dalam model CNN. Proses ini memastikan bahwa model dapat mempelajari pola emosi secara efisien dan efektif dari data gambar wajah.

2.3 Pelatihan Model

Setelah dataset melalui proses preprocessing, langkah berikutnya adalah pelatihan model menggunakan algoritma pembelajaran mendalam *deep learning*. Pada penelitian ini, model yang digunakan adalah Convolutional Neural Network (CNN), yang telah

terbukti efektif dalam mengenali pola visual pada data gambar [13], [14], [15].

1. Arsitektur Model

Arsitektur Convolutional Neural Network (CNN) yang digunakan dalam penelitian ini dirancang untuk mengekstraksi fitur-fitur penting dari gambar ekspresi wajah. Model ini terdiri dari tiga layer konvolusi utama. Layer pertama menggunakan 32 filter dengan ukuran kernel 3x3 dan aktivasi ReLU, diikuti oleh Max Pooling 2x2; layer kedua meningkatkan jumlah filter menjadi 64 dengan ukuran kernel tetap 3x3 dan aktivasi ReLU, diikuti oleh Max Pooling 2x2; dan layer ketiga menambah kompleksitas dengan 128 filter, kernel 3x3 dengan aktivasi ReLU, dan Max Pooling 2x2.

Setelah layer konvolusi, model menggunakan Flatten Layer untuk mengubah output konvolusi menjadi vektor satu dimensi. Dense Layer pertama memiliki 128 neuron dengan aktivasi ReLU, diikuti oleh Dropout Layer (0.5) untuk mencegah overfitting. Dense Layer output memiliki jumlah neuron sesuai dengan jumlah kelas (4 ekspresi) dan menggunakan aktivasi Softmax untuk mendapatkan probabilitas kelas.

2. Kompilasi Model

Tipe Lapisan	Bentuk Keluaran	Parameter
conv2d_3 (Conv2D)	(None, 126, 126, 32)	896
max_pooling_2d_3 (MaxPooling 2D)	(None, 63, 63, 32)	0
conv2d_4 (Conv2D)	(None, 61, 61, 64)	18,496
max_pooling_2d_4 (MaxPooling 2D)	(None, 30, 30, 64)	0
conv2d_5 (Conv2D)	(None, 28, 28, 128)	73,856
max_pooling_2d_5 (MaxPooling 2D)	(None, 14, 14, 128)	0
flatten_1 (Flatten)	(None, 25088)	0
dense_2 (Dense)	(None, 128)	3,211,392
dropout_1 (Dropout)	(None, 128)	0
dense_3 (Dense)	(None, 4)	516

Proses kompilasi model dilakukan dengan konfigurasi seperti Tabel 2. Optimizer yang digunakan adalah Adam, dipilih karena kemampuannya dalam mengadaptasi learning rate secara otomatis dan efektif untuk masalah optimasi dengan dataset gambar. *Loss Function* yang digunakan adalah *Categorical Crossentropy*, yang sesuai untuk masalah klasifikasi multi-kelas dan mengukur perbedaan antara distribusi probabilitas prediksi dengan label sebenarnya. Metrik yang digunakan adalah akurasi, yang digunakan untuk mengukur persentase prediksi yang benar selama pelatihan.

3. Proses Pelatihan dan Evaluasi

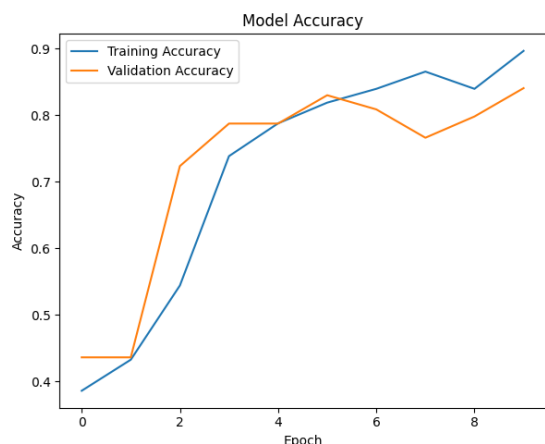
Proses pelatihan model dilakukan dengan menggunakan data generator yang telah dipersiapkan sebelumnya. Data generator terdiri dari training generator yang menggunakan 80% dataset dan validation generator yang menggunakan 20% dataset, dengan batch size sebesar 32 dan ukuran gambar 128x128 piksel. Konfigurasi pelatihan meliputi jumlah epoch sebanyak 10, dan model dilatih menggunakan metode `model.fit()` dengan validasi dilakukan pada setiap epoch. Selama proses pelatihan, model akan mempelajari fitur dari gambar ekspresi wajah, menyesuaikan bobot untuk meminimalkan kesalahan prediksi, dan melakukan validasi pada setiap epoch untuk memantau performa.

Setelah proses pelatihan selesai, langkah berikutnya adalah mengevaluasi performa model dengan menggunakan metrik-metrik yang relevan untuk tugas klasifikasi multi-kelas. Evaluasi dilakukan untuk memastikan bahwa model dapat mengenali ekspresi emosi pada gambar wajah dengan akurasi tinggi dan mampu melakukan generalisasi pada data baru yang tidak terlihat selama pelatihan.

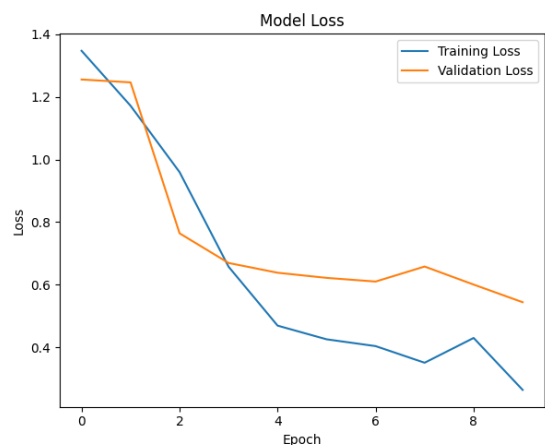
3. Hasil dan Pembahasan

3.1 Hasil

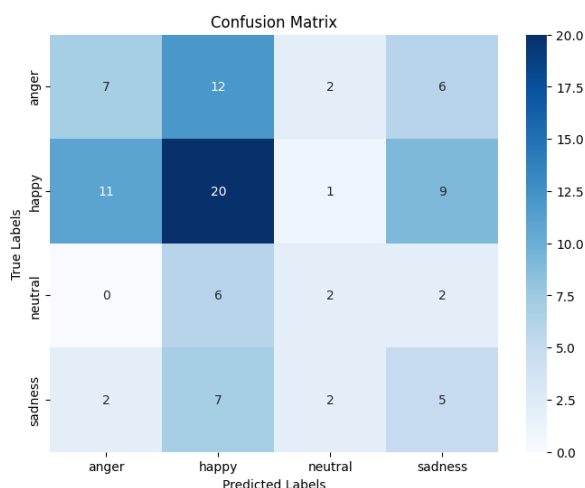
Hasil penelitian menunjukkan efektivitas model CNN dalam mengklasifikasikan ekspresi wajah ke dalam empat kategori: marah, senang, sedih, dan netral. Model mencapai akurasi pelatihan sebesar 92% dan akurasi validasi sebesar 85%. Hasil ini mengindikasikan bahwa model mampu mempelajari fitur-fitur dataset secara efektif sambil tetap menjaga generalisasi. Meskipun akurasi yang dicapai cukup tinggi, analisis lebih lanjut terhadap *classification report* menunjukkan beberapa metrik yang kurang memuaskan pada Tabel 3.



Gambar 1. Grafik Akurasi Model



Gambar 2. Grafik Kehilangan Model



Gambar 3. Confusion Matrix

Tabel 3. Laporan Klasifikasi

Class	Precision	Recall	F1-Score
Anger	0.35	0.26	0.30
Happy	0.44	0.49	0.47
Neutral	0.29	0.20	0.24
Sadness	0.23	0.31	0.26
Accuracy	0.36		
Macro avg	0.33	0.31	0.32
Weighted avg	0.36	0.36	0.36

Secara umum, meskipun akurasi pada data pelatihan dan validasi tergolong tinggi, nilai precision, recall, dan F1-score untuk beberapa kelas, terutama anger dan sadness, menunjukkan performa yang rendah. Model memiliki precision dan recall yang cukup rendah pada kelas-kelas tersebut, yang berarti model kurang efektif dalam mendeteksi ekspresi tertentu.

Namun, hasil prediksi untuk gambar di luar dataset menunjukkan bahwa model mampu memberikan prediksi yang sesuai, meskipun memiliki performa yang kurang baik pada data uji. Ini mengindikasikan bahwa model mungkin masih dapat mengenali pola yang lebih umum meskipun ketepatan pada data uji tidak maksimal.

3.2 Pembahasan

Hasil yang diperoleh menunjukkan adanya beberapa isu yang perlu dicermati, terutama berkaitan dengan imbalansi data dan performa per kelas. Meskipun akurasi pelatihan tinggi, hal ini seringkali tidak cukup menggambarkan kinerja model secara menyeluruh. Pada kasus ini, rendahnya metrik precision, recall, dan F1-score pada kelas-kelas tertentu mengindikasikan bahwa model mengalami kesulitan dalam mengenali ekspresi tertentu dengan baik, terutama untuk kelas dengan jumlah gambar yang lebih sedikit, seperti neutral dan sadness.

Salah satu penyebab utama rendahnya metrik-metrik ini adalah adanya ketidakseimbangan kelas dalam dataset. Seperti yang terlihat dalam tabel distribusi dataset, kelas neutral dan sadness memiliki jumlah gambar yang jauh lebih sedikit dibandingkan dengan happy dan anger. Ketidakseimbangan ini dapat menyebabkan model lebih cenderung untuk memprediksi kelas yang

lebih dominan, sehingga precision dan recall untuk kelas yang lebih sedikit menjadi rendah.

Meskipun model menunjukkan hasil yang kurang optimal pada data uji, hasil yang diperoleh pada gambar di luar dataset lebih baik. Ini mengindikasikan bahwa model mungkin lebih robust terhadap variasi data yang lebih besar atau gambar baru yang memiliki kemiripan dengan gambar pelatihan, meskipun tidak sepenuhnya berhasil mengenali semua variasi ekspresi pada dataset yang lebih kecil.

4. Kesimpulan

Penelitian ini menunjukkan bahwa model Convolutional Neural Network (CNN) mampu mendeteksi ekspresi wajah dengan tingkat akurasi yang bervariasi pada data uji, dengan hasil yang cukup memadai pada pengujian gambar di luar dataset. Model memiliki kelemahan dalam menghadapi ketidakseimbangan kelas, yang menyebabkan performa rendah pada ekspresi tertentu. Namun, hasil kualitatif menunjukkan potensi penggunaan model ini dalam aplikasi praktis untuk mendeteksi ekspresi emosi dasar manusia. Untuk meningkatkan performa, penelitian di masa depan dapat menggunakan dataset yang lebih besar dan seimbang, serta teknik peningkatan data lainnya.

Daftar Rujukan

- [1] S. S. Panna and Betrisandi, 'Klasifikasi Kelompok Usia Melalui Citra Wajah Berbasis Image Texture Analysis pada Sistem Automatic Video Filtering', *RESTI*, vol. 3, no. 3, pp. 429–434, Dec. 2019, doi: 10.29207/resti.v3i3.1280.
- [2] A. M. Khairi, 'Dosen Fakultas Ushuluddin dan Dakwah IAIN Surakarta Jln. Pandawa Pucangan kartasura Sukoharjo, Jawa Tengah', vol. 19, 2018.
- [3] I. Ekawati, F. N. R. Putra, M. Sumadyo, and R. Nugroho, 'Deteksi Emosi Menggunakan Convolutional Neural Network Berdasarkan Ekspresi Wajah', 2024.
- [4] Rendi Nurcahyo and Mohammad Iqbal, 'Pengenalan Emosi Pembicara Menggunakan Convolutional Neural Networks', *J. RESTI (Rekayasa Sist. Teknol. Inf.)*, vol. 6, no. 1, pp. 115–122, Feb. 2022, doi: 10.29207/resti.v6i1.3726.
- [5] Al Sigit Guntoro, Edy Julianto, and Djoko Budiyo, 'Pengenalan Ekspresi Wajah Menggunakan Convolutional Neural Network', *Jurnal Informatika Atma Joga*, vol. 3, no. 2, pp. 155–160, Nov. 2022, doi: 10.24002/jiaj.v3i2.6790.
- [6] M. Ihsan, R. K. Niswatin, and D. Swanjaya, 'DETEKSI EKSPRESI WAJAH MENGGUNAKAN TENSORFLOW', *JTI*, vol. 6, no. 1, p. 428, Mar. 2021, doi: 10.30736/jti.v6i1.554.
- [7] F. Chan, 'Perancangan Sistem Pendeteksi Emosional Siswa Menggunakan Algoritma CNN untuk Mengukur Tingkat Pengelolaan Kelas', *Intellect*, 2024.
- [8] M. A. Fadiilla, H. Setiawan, and M. Ramadhan, 'IMPLEMENTASI METODE CONVOLUTIONAL NEURAL NETWORK (CNN) PADA SISTEM DETEKSI EMOSI DARI EKSPRESI WAJAH MANUSIA DENGAN APLIKASI ANDROID SEBAGAI ANTARMUKA PENGGUNA', 2023.
- [9] Lia Farokhah, 'Perbandingan Metode Deteksi Wajah Menggunakan OpenCV Haar Cascade, OpenCV Single Shot Multibox Detector (SSD) dan DLib CNN', *RESTI*, vol. 5, no. 3, pp. 609–614, Jun. 2021, doi: 10.29207/resti.v5i3.3125.
- [10] C. Amelia and Y. Aryaneta, 'PENGARUH MUSIK TERHADAP EMOSI', 2022.
- [11] A. Lutfi, R. Aulia, F. Fatihatunnisa, and N. Khairunnisa, 'Harmoni Getaran dan Gelombang: Eksplorasi Emosi Manusia melalui Karya Seni Musik', 2023.
- [12] K. Maharana, S. Mondal, and B. Nemade, 'A review: Data pre-processing and data augmentation techniques', *Global Transitions Proceedings*, vol. 3, no. 1, pp. 91–99, Jun. 2022, doi: 10.1016/j.gltp.2022.04.020.
- [13] M. R. Edy, 'Deteksi Emosi dari Ekspresi Wajah dengan Deep Learning', vol. 3, no. 2, 2024.
- [14] A. F. Ibrahim and S. Raniprma, 'Implementasi Machine Learning Pada Alat Deteksi Emosi Untuk Sistem Kontrol Suhu Dan Pencahayaan Ruangan', 2022.
- [15] F. A. G. Zhafirah, R. Rokhana, R. Sigit, and B. S. B. Dewantara, 'Fusi Algoritma K-Means dan CNN untuk Klasifikasi Emosi pada Anak', *tc*, vol. 22, no. 3, pp. 622–634, Aug. 2023, doi: 10.33633/tc.v22i3.8667.