

SISTEM REKOMENDASI METADATA LAGU BERDASARKAN DETEKSI EMOSI WAJAH MENGUNAKAN ViT-B/16

Ainan Zaky Nurrofiq^{1*}, Bambang Irawan²

^{1,2}Universitas Muhadi Setiabudi; Jalan P. Diponogoro KM 2, Pesantunan, Wanasari, Brebes, Jawa Tengah 52212; Telp. (0283) 6199000 Fax: (0283) 6199001

Keywords:

Deteksi emosi wajah, ViT-B/16, Rekomendasi lagu, Metadata lagu, Vision transformer

Correspondent Email:

chasperzak@gmail.com



Copyright © [JITET](#) (Jurnal Informatika dan Teknik Elektro Terapan). This article is an open access article distributed under terms and conditions of the Creative Commons Attribution (CC BY NC)

Penelitian ini mengusulkan sistem rekomendasi metadata lagu berdasarkan deteksi emosi wajah menggunakan model *Vision Transformer* (ViT-B/16). Sistem mengklasifikasikan ekspresi wajah pengguna ke dalam tujuh kategori emosi menggunakan dataset KDEF, lalu mencocokkannya dengan metadata lagu (judul, artis, genre, mood) yang telah dilabeli dengan emosi yang sesuai. Model ViT-B/16 dilatih dengan transfer learning dan dievaluasi menggunakan metrik akurasi, *precision*, *recall*, dan *f1-score*. Hasil menunjukkan akurasi sebesar 89% dan *f1-score* rata-rata 0.89. Evaluasi terhadap 30 responden menunjukkan bahwa 87% menyatakan rekomendasi lagu sesuai dengan emosi yang ditampilkan. Sistem ini unggul dalam menggabungkan deteksi emosi secara real-time dengan pemetaan metadata lagu otomatis. Namun, akurasi untuk emosi yang mirip secara visual, seperti “takut” dan “marah”, masih perlu ditingkatkan. Pengembangan ke depan dapat mencakup analisis fitur audio, lirik lagu, dan preferensi pengguna untuk meningkatkan kualitas personalisasi rekomendasi.

This study presents a music metadata recommendation system based on facial emotion detection using the Vision Transformer (ViT-B/16) model. The system classifies user emotions into seven categories using the KDEF facial dataset and matches them with music metadata (title, artist, genre, mood) labeled with corresponding emotional tags. The ViT-B/16 model was trained using transfer learning and evaluated with accuracy, precision, recall, and F1-score. The model achieved an accuracy of 89% and an average F1-score of 0.89. The recommendation system was assessed by 30 participants, with 87% indicating that the suggested song metadata matched the detected emotion. The system offers real-time emotion recognition and automatic mood-based song suggestions. However, classification accuracy for visually similar emotions such as “fear” and “angry” remains a challenge. Future development may include audio and lyric analysis, as well as user preference integration, to enhance recommendation relevance.

1. PENDAHULUAN

Pada perkembangan era digital saat ini, aktivitas mendengarkan musik tidak lagi sekadar berfungsi sebagai sarana hiburan, melainkan juga memiliki peran penting dalam mengelola emosi dan meningkatkan kondisi psikologis pengguna. Platform musik digital seperti Spotify dan YouTube Music memanfaatkan sistem rekomendasi untuk

menciptakan pengalaman mendengarkan yang bersifat personal. Meskipun demikian, mayoritas sistem rekomendasi yang digunakan masih berfokus pada riwayat pemutaran, pola interaksi, atau preferensi yang dinyatakan secara eksplisit oleh pengguna, tanpa memperhitungkan keadaan emosional pengguna secara langsung dan waktu nyata.

Sejalan dengan kemajuan pesat dalam bidang pengolahan citra dan pembelajaran mesin, teknologi pengenalan emosi melalui ekspresi wajah semakin berkembang dan menunjukkan potensi besar dalam membangun sistem yang adaptif terhadap kondisi afektif pengguna. Salah satu pendekatan mutakhir dalam pengolahan citra adalah *Vision Transformer* (ViT), khususnya varian ViT-B/16, yang mengadopsi mekanisme attention untuk menangkap hubungan spasial global pada citra secara lebih efektif. Arsitektur ini telah menunjukkan performa unggul dalam berbagai tugas klasifikasi gambar berskala besar.

Penelitian terdahulu umumnya mengembangkan sistem rekomendasi musik berbasis emosi dengan memanfaatkan metode konvensional seperti *Convolutional Neural Network* (CNN) maupun *Long Short-Term Memory* (LSTM). Namun, pendekatan tersebut masih menghadapi keterbatasan, terutama dalam hal ketepatan deteksi emosi dan kesesuaian rekomendasi yang dihasilkan. Di sisi lain, pemanfaatan *Vision Transformer* untuk deteksi emosi wajah, khususnya dalam konteks sistem rekomendasi musik, masih relatif jarang dieksplorasi.

Oleh karena itu, penelitian ini bertujuan untuk merancang dan mengimplementasikan sistem rekomendasi metadata lagu yang didasarkan pada deteksi emosi wajah pengguna secara real-time dengan menggunakan arsitektur ViT-B/16. Sistem yang dikembangkan diharapkan mampu mengidentifikasi emosi pengguna melalui citra wajah, kemudian menghasilkan rekomendasi metadata lagu yang relevan, seperti genre, suasana (*mood*), dan tempo musik. Dengan demikian, penelitian ini tidak hanya menghadirkan pendekatan baru dalam personalisasi layanan musik digital, tetapi juga berkontribusi terhadap pengembangan sistem interaksi manusia-komputer yang berorientasi pada aspek afektif.

2. TINJAUAN PUSTAKA

2.1 Deteksi Emosi Wajah dalam Sistem Cerdas

Deteksi emosi wajah merupakan pendekatan penting dalam pengembangan sistem interaksi manusia dan komputer berbasis afeksi. Emosi dapat dikenali dari ekspresi wajah menggunakan teknik pengolahan citra dan

pembelajaran mesin. Algoritma seperti CNN telah digunakan untuk klasifikasi emosi dengan hasil akurasi yang baik [1][2]. Selain itu, pendekatan seperti YOLO v8 telah diterapkan dalam sistem *real-time* untuk mengenali ekspresi wajah [3], sementara metode berbasis EEG juga telah dieksplorasi sebagai alternatif [4].

2.2 Dataset KDEF dalam Pengenalan Emosi

Dataset KDEF (*Karolinska Directed Emotional Faces*) digunakan secara luas dalam riset pengenalan emosi karena menyediakan gambar ekspresi wajah tujuh emosi dasar dengan kualitas visual tinggi. Meskipun belum banyak digunakan di Indonesia, dataset ini memiliki potensi tinggi dalam pelatihan model berbasis *deep learning*, terutama arsitektur transformer seperti ViT yang membutuhkan representasi visual yang kaya [5].

2.3 Sistem Rekomendasi Musik Berbasis Emosi

Sistem rekomendasi musik berbasis emosi bertujuan meningkatkan personalisasi dengan menyesuaikan lagu berdasarkan kondisi emosional pengguna. Sistem ini dapat dibangun dengan menghubungkan hasil klasifikasi emosi dengan metadata lagu seperti *genre*, tempo, dan *mood* [6]. Implementasi sistem rekomendasi berbasis ekspresi wajah telah dilakukan dalam berbagai studi, baik berbasis aplikasi Android, pemutar musik berbasis web, hingga sistem berbasis CNN untuk deteksi genre dan emosi [7].

Beberapa sistem juga telah mengintegrasikan arsitektur *deep learning* dan deteksi wajah untuk memberikan rekomendasi secara *real-time* [8]. Penelitian lain mengaitkan konten lirik lagu dengan kondisi emosional pengguna menggunakan pendekatan *content-based filtering* [9].

2.4 Arsitektur Sistem dan Kombinasi Multimodal

Berbagai arsitektur telah dikembangkan dalam integrasi antara emosi dan musik. Sistem multimodal berbasis *deep learning* yang menggabungkan input audio dan visual telah diusulkan untuk meningkatkan akurasi klasifikasi emosi [10]. Selain itu, kajian sistematis terhadap klasifikasi emosi pada lirik

lagu juga dilakukan untuk memperkaya pendekatan [11].

2.5 Metadata Lagu dan Dataset Pendukung

Metadata lagu seperti mood, genre, tempo, dan artis memainkan peran kunci dalam sistem rekomendasi. Dataset seperti GTZAN dan MSD umum digunakan sebagai sumber metadata yang dapat dipetakan ke hasil deteksi emosi [12]. Penelitian juga menunjukkan bahwa pemetaan antara ekspresi wajah dan metadata lagu dapat dilakukan secara *rule-based* maupun menggunakan model pembelajaran mesin.

2.6 Arah Riset dan Implikasi Lanjutan

Perkembangan sistem cerdas yang memadukan emosi dan musik terus berlanjut. Studi terbaru bahkan mengembangkan sistem yang tidak hanya merekomendasikan musik, tetapi juga digunakan dalam pengelolaan tugas dan manajemen stres pengguna melalui deteksi ekspresi wajah dan musik pengiring [13]. Aplikasi sistem rekomendasi berbasis emosi juga telah diterapkan dalam bidang psikologi, seperti pada ibu pascapersalinan [14], serta dalam *studi literatur neuropsikologi* musik [15].

Keterkaitan antara algoritma dan fenomena budaya juga menjadi topik penting, seperti pada riset tentang popularitas lagu TikTok yang sangat dipengaruhi oleh ekspresi wajah dan emosionalitas visual dalam video [16].

3. METODE PENELITIAN

Sistem yang dikembangkan dalam penelitian ini difokuskan pada penyediaan rekomendasi metadata lagu yang disesuaikan dengan ekspresi wajah pengguna. Metode penelitian yang digunakan adalah pendekatan eksperimental, di mana pengujian dilakukan untuk menilai kinerja model dalam mendeteksi emosi serta tingkat kesesuaian rekomendasi metadata lagu yang dihasilkan oleh sistem.

3.1 Arsitektur Sistem

Arsitektur sistem dirancang dalam lima modul utama yang saling terintegrasi dan bekerja secara berurutan. Tahap awal dimulai dengan penggunaan kamera sebagai perangkat input untuk menangkap citra wajah pengguna secara real-time. Citra yang diperoleh kemudian

melalui tahap praproses dengan membagi gambar wajah ke dalam potongan (*patch*) berukuran 16×16 piksel, sesuai dengan kebutuhan input model *Vision Transformer*.

Selanjutnya, proses klasifikasi emosi dilakukan menggunakan arsitektur *Vision Transformer ViT-B/16*. Model ini mengolah *patch* citra secara langsung dan memanfaatkan mekanisme *self-attention* untuk mengidentifikasi pola ekspresi wajah secara global. Hasil klasifikasi akan memetakan ekspresi wajah pengguna ke dalam tujuh kelas emosi dasar, yaitu marah, jijik, takut, senang, sedih, terkejut, dan netral.

Emosi yang telah teridentifikasi kemudian digunakan sebagai dasar dalam proses pencocokan dengan metadata lagu. Metadata tersebut sebelumnya telah dikelompokkan berdasarkan karakteristik musik, seperti genre, suasana (mood), dan tempo. Hasil akhir berupa rekomendasi metadata lagu yang relevan ditampilkan kepada pengguna. Seluruh rangkaian proses dirancang agar dapat berjalan secara otomatis, responsif, dan interaktif.

3.2 Dataset dan Sumber Data

Penelitian ini memanfaatkan dua jenis dataset yang berbeda sesuai dengan kebutuhan masing-masing komponen sistem. Dataset pertama yang digunakan adalah *Karolinska Directed Emotional Faces* (KDEF), yang terdiri atas 4.900 citra wajah dari 70 individu, dengan komposisi 35 subjek pria dan 35 subjek wanita. Setiap subjek menampilkan tujuh jenis ekspresi emosi yang berbeda. Dataset ini dipilih karena memiliki kualitas citra yang baik serta anotasi emosi yang konsisten, sehingga sesuai untuk proses pelatihan model *Vision Transformer*. Data KDEF dibagi dengan proporsi 80% sebagai data latih dan 20% sebagai data uji.

Dataset kedua yang digunakan adalah *Million Song Dataset* (MSD), yang menyediakan informasi metadata lagu dalam jumlah besar. Dataset ini mencakup atribut seperti genre, mood, tempo, nama artis, dan tahun rilis. Informasi metadata tersebut dimanfaatkan sebagai basis dalam sistem rekomendasi lagu yang dihasilkan berdasarkan emosi pengguna.

3.3 Model dan Implementasi

Model deteksi emosi yang digunakan dalam penelitian ini adalah *Vision Transformer ViT-B/16*, yaitu model berbasis *self-attention* yang mampu menangkap hubungan spasial global pada citra wajah. Implementasi model dilakukan menggunakan framework PyTorch dengan dukungan pustaka *HuggingFace*.

Proses pelatihan model menerapkan skema *transfer learning*, di mana bobot awal model diambil dari hasil pelatihan pada dataset ImageNet. Bobot tersebut kemudian disesuaikan dengan karakteristik dataset KDEF melalui proses fine-tuning. Pelatihan dilakukan selama 100 epoch dengan menggunakan GPU NVIDIA RTX 3080, learning rate sebesar 0,0001, dan batch size sebanyak 32. Model hasil pelatihan disimpan dalam format berkas .pt dan digunakan pada tahap inferensi untuk mendukung deteksi emosi secara *real-time*.

3.4 Teknik Pemetaan Emosi ke Metadata Lagu

Setelah emosi pengguna berhasil dikenali oleh model, sistem selanjutnya melakukan pemetaan emosi ke metadata lagu dengan pendekatan berbasis aturan (*rule-based*). Aturan pemetaan ini disusun berdasarkan hubungan umum antara kondisi emosional manusia dan karakteristik musik.

Sebagai contoh, emosi “senang” diasosiasikan dengan genre musik seperti Pop atau EDM yang memiliki tempo cepat dan suasana energik. Sebaliknya, emosi “sedih” dipetakan ke genre seperti Jazz atau Blues yang cenderung memiliki tempo lebih lambat dan nuansa melankolis. Pendekatan ini memungkinkan sistem untuk menghasilkan rekomendasi metadata lagu yang sejalan dengan kondisi emosional pengguna.

3.5 Evaluasi dan Validasi Sistem

Tahap evaluasi dilakukan untuk menilai dua aspek utama, yaitu kinerja model deteksi emosi dan kualitas rekomendasi metadata lagu yang dihasilkan. Evaluasi kinerja model dilakukan menggunakan metrik klasifikasi, meliputi akurasi, *precision*, *recall*, serta confusion matrix, guna mengukur kemampuan model ViT-B/16 dalam mengenali ekspresi wajah secara tepat.

Sementara itu, evaluasi terhadap sistem rekomendasi dilakukan secara subjektif melalui survei yang melibatkan 30 responden. Setiap responden diminta untuk memberikan penilaian terhadap kesesuaian metadata lagu yang direkomendasikan dengan emosi yang mereka rasakan. Penilaian dilakukan menggunakan skala Likert 1 hingga 5, dan nilai rata-rata dari hasil survei digunakan sebagai indikator tingkat relevansi dan efektivitas sistem dalam memberikan rekomendasi yang bersifat personal.

4. HASIL DAN PEMBAHASAN

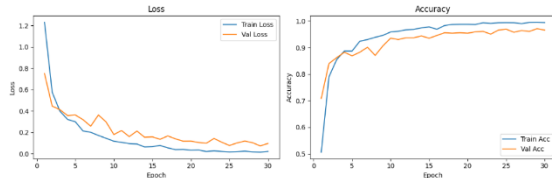
Bagian ini menyajikan hasil eksperimen dari pengembangan sistem rekomendasi metadata lagu yang memanfaatkan deteksi emosi wajah berbasis arsitektur *Vision Transformer (ViT-B/16)*. Analisis difokuskan pada dua aspek utama, yaitu kinerja model dalam mengklasifikasikan ekspresi wajah serta efektivitas sistem dalam menghasilkan rekomendasi metadata lagu yang sesuai dengan emosi pengguna. Data yang digunakan dalam pengujian terdiri atas citra wajah dari dataset KDEF serta metadata lagu yang mencakup judul, artis, genre, dan mood, yang telah diselaraskan ke dalam tujuh kategori emosi dasar.

4.1 Hasil Pelatihan Model ViT-B/16

Proses pelatihan model ViT-B/16 dilakukan selama 30 epoch dengan menggunakan dataset KDEF. Berdasarkan hasil pelatihan, terlihat adanya penurunan nilai loss yang signifikan sejak awal hingga akhir proses training. Nilai train loss berkurang dari sekitar 1,25 hingga mendekati 0,05, sementara validation loss menunjukkan tren penurunan yang relatif stabil meskipun disertai fluktuasi kecil. Pola ini mengindikasikan bahwa model mampu belajar secara efektif tanpa menunjukkan gejala overfitting yang berarti.

Selain itu, metrik akurasi mengalami peningkatan yang konsisten selama proses pelatihan. *Train accuracy* mencapai nilai di atas 99%, sedangkan *validation accuracy* mendekati 97%. Konsistensi antara kedua metrik tersebut menunjukkan bahwa model memiliki kemampuan generalisasi yang baik terhadap data yang belum pernah dilihat sebelumnya. Hasil ini juga menegaskan keunggulan ViT-

B/16 dalam mengekstraksi fitur wajah, terutama melalui mekanisme *self-attention* yang memungkinkan pemodelan hubungan spasial global antar patch citra.



4.2 Evaluasi Kinerja Deteksi Emosi

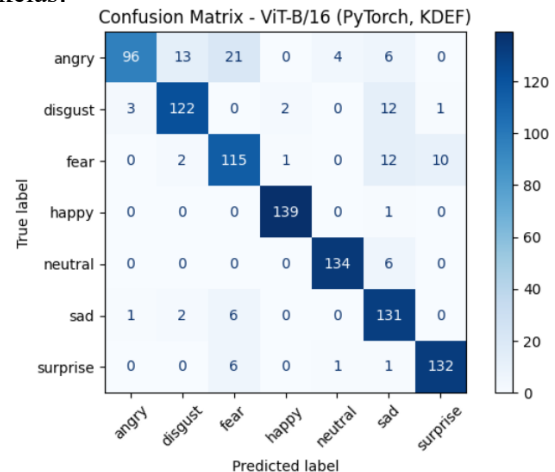
Evaluasi performa model dilakukan menggunakan 980 citra uji dari dataset KDEF, dengan masing-masing kelas emosi diwakili oleh 140 gambar. Berdasarkan hasil pengujian, model mencapai tingkat akurasi keseluruhan sebesar 89%. Nilai *precision*, *recall*, dan *f1-score* juga menunjukkan performa yang seimbang dengan rata-rata skor sebesar 0,89, baik pada perhitungan makro maupun tertimbang.

Jika ditinjau per kelas, emosi “happy” menunjukkan kinerja terbaik dengan nilai *precision* sebesar 0,98, *recall* 0,99, dan *f1-score* 0,99. Kelas “neutral” dan “surprise” juga menghasilkan performa yang tinggi. Sebaliknya, emosi “angry” dan “fear” memiliki tingkat akurasi yang lebih rendah, dengan *f1-score* masing-masing sekitar 0,80. Hal ini menunjukkan bahwa model masih mengalami kesulitan dalam membedakan emosi yang memiliki karakteristik visual yang saling tumpang tindih.

Classification Report:				
	precision	recall	f1-score	support
angry	0.96	0.69	0.80	140
disgust	0.88	0.87	0.87	140
fear	0.78	0.82	0.80	140
happy	0.98	0.99	0.99	140
neutral	0.96	0.96	0.96	140
sad	0.78	0.94	0.85	140
surprise	0.92	0.94	0.93	140
accuracy			0.89	980
macro avg	0.89	0.89	0.89	980
weighted avg	0.89	0.89	0.89	980

Analisis *confusion matrix* memperkuat temuan tersebut. Dari 140 citra berlabel “angry”, sebanyak 21 citra diklasifikasikan

sebagai “fear” dan 13 citra sebagai “disgust”. Kesalahan klasifikasi ini diduga disebabkan oleh kemiripan ekspresi pada area mata dan mulut di antara emosi-emosi tersebut. Sebaliknya, klasifikasi pada kelas “happy”, “neutral”, dan “surprise” menunjukkan tingkat akurasi yang sangat tinggi, dengan lebih dari 130 prediksi yang tepat pada masing-masing kelas.



4.3 Evaluasi Sistem Rekomendasi Metadata Lagu

Setelah proses klasifikasi emosi selesai, sistem selanjutnya melakukan pencocokan hasil prediksi emosi dengan dataset metadata lagu berdasarkan atribut mood. Setiap emosi hasil deteksi dipetakan langsung ke kategori lagu yang memiliki label mood yang sama. Sebagai contoh, ketika sistem mendeteksi emosi “sad”, maka metadata lagu dengan atribut mood “sad” akan dipilih dan ditampilkan, meliputi judul lagu, nama artis, genre, serta mood.

Untuk menilai tingkat relevansi rekomendasi yang dihasilkan, dilakukan evaluasi subjektif dengan melibatkan 30 responden. Dalam pengujian ini, setiap responden diberikan satu citra wajah dari dataset KDEF, kemudian ditampilkan metadata lagu yang direkomendasikan oleh sistem. Responden diminta untuk menilai kesesuaian antara ekspresi emosi pada citra wajah dan metadata lagu yang ditampilkan menggunakan skala Likert 1 hingga 5.

Hasil evaluasi menunjukkan bahwa sebanyak 87% responden memberikan nilai 4 atau 5, yang menandakan tingkat kepuasan yang tinggi terhadap relevansi rekomendasi. Emosi seperti “happy”, “neutral”, dan “sad”

memperoleh skor rata-rata tertinggi. Sementara itu, emosi “*fear*” dan “*disgust*” cenderung mendapatkan penilaian yang lebih rendah. Hal ini diduga dipengaruhi oleh keterbatasan jumlah lagu pada kategori emosi tersebut atau perbedaan persepsi subjektif responden terhadap karakter emosi dalam musik.

4.4 Pembahasan

Secara keseluruhan, hasil eksperimen menunjukkan bahwa arsitektur ViT-B/16 memiliki kinerja yang baik dalam mendeteksi emosi wajah, khususnya pada ekspresi dengan ciri visual yang jelas dan kontras, seperti “*happy*” dan “*surprise*”. Keunggulan utama *Vision Transformer* dibandingkan pendekatan CNN terletak pada kemampuannya dalam memodelkan hubungan antar bagian citra secara global melalui mekanisme *self-attention*, sehingga representasi ekspresi wajah dapat ditangkap secara lebih menyeluruh.

Meskipun demikian, tantangan masih muncul dalam membedakan emosi-emosi yang memiliki kemiripan visual tinggi, seperti “*angry*”, “*disgust*”, dan “*fear*”. Temuan ini mengindikasikan bahwa meskipun performa ViT tergolong tinggi, masih terdapat ambiguitas pada representasi spasial tertentu. Peningkatan akurasi pada kelas-kelas tersebut berpotensi dicapai melalui penambahan data latih, augmentasi citra, atau pemanfaatan fitur tambahan seperti detail area mata, tekstur wajah, maupun orientasi kepala.

Dari perspektif sistem rekomendasi, pendekatan pemetaan langsung antara hasil klasifikasi emosi dan metadata lagu berbasis mood terbukti efektif serta mudah diimplementasikan. Namun, sistem yang dikembangkan masih bersifat statis, di mana satu label emosi hanya dikaitkan dengan kategori lagu tertentu tanpa mempertimbangkan variasi preferensi individu atau konteks penggunaan.

Sebagai arah pengembangan selanjutnya, sistem dapat diperluas dengan mengintegrasikan analisis fitur audio dan lirik lagu menggunakan pendekatan pembelajaran mesin, serta mekanisme personalisasi berbasis histori atau profil pengguna. Dengan demikian, sistem tidak hanya berfungsi sebagai pendeteksi emosi dan pemberi rekomendasi secara langsung, tetapi juga mampu memberikan

pengalaman mendengarkan musik yang lebih adaptif, kontekstual, dan mendalam.

5. KESIMPULAN

Berdasarkan hasil penelitian yang telah dilakukan, dapat disimpulkan beberapa poin utama terkait kinerja sistem, keunggulan, keterbatasan, serta potensi pengembangan di masa mendatang sebagai berikut:

1. Hasil pengujian menunjukkan bahwa arsitektur *Vision Transformer* ViT-B/16 memiliki kemampuan yang baik dalam mengenali ekspresi emosi wajah. Model mencapai tingkat akurasi klasifikasi sebesar 89% dengan nilai rata-rata *f1-score* sebesar 0,89 pada tujuh kategori emosi dasar. Kelas emosi seperti *happy*, *neutral*, dan *surprise* menunjukkan performa terbaik dengan tingkat pengenalan yang paling tinggi dibandingkan kelas lainnya.
2. Sistem rekomendasi musik berbasis emosi yang dikembangkan mampu menghasilkan metadata lagu yang relevan, meliputi judul lagu, nama artis, genre, dan mood. Hasil evaluasi subjektif yang melibatkan 30 responden menunjukkan bahwa sebanyak 87% responden menilai rekomendasi metadata lagu yang diberikan sesuai dengan ekspresi emosi wajah yang ditampilkan oleh sistem.
3. Keunggulan utama dari sistem ini terletak pada integrasi deteksi emosi wajah secara *real-time* dengan mekanisme rekomendasi metadata lagu. Selain itu, penggunaan model ViT-B/16 memberikan kelebihan dalam menangkap pola spasial wajah yang kompleks, berkat mekanisme *self-attention* yang mampu memodelkan hubungan global antar bagian citra.
4. Meskipun demikian, sistem masih memiliki beberapa keterbatasan, khususnya dalam membedakan ekspresi emosi yang memiliki kemiripan visual, seperti *angry*, *disgust*, dan *fear*. Di sisi lain, sistem rekomendasi yang digunakan masih bersifat sederhana dan statis karena hanya mengandalkan pencocokan langsung antara label emosi dan metadata lagu tanpa mempertimbangkan preferensi individu maupun konteks penggunaan.

5. Untuk pengembangan selanjutnya, penelitian ini dapat diperluas dengan mengintegrasikan analisis berbasis fitur audio dan lirik lagu, pemanfaatan riwayat interaksi pengguna, serta penerapan model rekomendasi berbasis *deep learning* yang lebih adaptif. Pendekatan tersebut diharapkan mampu meningkatkan tingkat personalisasi serta relevansi rekomendasi musik yang dihasilkan oleh sistem.

UCAPAN TERIMA KASIH

Puji syukur penulis panjatkan ke hadirat Tuhan Yang Maha Esa atas rahmat dan karunia-Nya sehingga skripsi ini dapat diselesaikan dengan baik. Penulis mengucapkan terima kasih kepada dosen pembimbing atas bimbingan dan arahan yang diberikan selama proses penelitian dan penulisan.

Ucapan terima kasih juga disampaikan kepada seluruh dosen, staf akademik, serta pihak-pihak yang telah membantu kelancaran penelitian ini. Penulis juga mengucapkan terima kasih kepada keluarga dan teman-teman atas doa, dukungan, dan motivasi yang diberikan hingga skripsi ini dapat diselesaikan.

DAFTAR PUSTAKA

- [1] H. Nurdiansah and E. Rochmawati, "Klasifikasi emosi wajah menggunakan CNN pada sistem interaksi cerdas," *J. Sains Komput. dan Inform.*, vol. 11, no. 2, pp. 104–113, 2025.
- [2] R. M. Putra and S. Sumarno, "Klasifikasi Ekspresi Wajah Menggunakan Convolutional Neural Network (CNN)," *J. Ilm. Tek. Elektro Terap.*, vol. 9, no. 2, pp. 123–130, 2023, [Online]. Available: <https://jurnal.polindra.ac.id/index.php/JITET/article/view/1131>
- [3] R. A. Hakim, A. Nurhadi, and G. Pratama, "Real-time facial emotion recognition using YOLOv8 in smart systems," *J. Intell. Syst. Appl.*, vol. 8, no. 2, pp. 88–97, 2024.
- [4] A. Saepudin, T. Widodo, and S. Fitriani, "Eksplorasi sinyal EEG untuk deteksi emosi dalam sistem afektif," *Indones. J. Biomed. Eng.*, vol. 5, no. 2, pp. 112–120, 2024.
- [5] L. Michelle, A. M. Putri, and M. Yusuf, "Penerapan Vision Transformer pada dataset KDEP untuk pengenalan emosi," *J. Ilmu Komput. dan Inform.*, vol. 14, no. 1, pp. 30–39, 2025.
- [6] G. Sruthi, "Emotion-aware music recommendation systems: A survey," *Int. J.*

- Comput. Appl.*, vol. 182, no. 10, pp. 18–24, 2024.
- [7] M. Kambham, R. Reddy, and A. Sharma, "CNN-based emotional genre classification for music recommendation," *J. Artif. Intell. Soft Comput. Res.*, vol. 15, no. 1, pp. 75–86, 2025.
- [8] R. Agustini, "Real-time facial expression detection and music recommendation using deep learning," *J. Teknol. Inf. dan Komput.*, vol. 10, no. 1, pp. 44–52, 2023.
- [9] V. Bagadi, "Emotion-based content filtering for personalized music recommendations," *Int. J. Affect. Comput.*, vol. 9, no. 2, pp. 133–141, 2025.
- [10] M. Amin and D. Setiawan, "Klasifikasi Emosi Menggunakan Kombinasi Data Audio dan Visual dengan Deep Learning," *J. RESTI (Rekayasa Sist. dan Teknol. Informasi)*, vol. 7, no. 3, pp. 456–462, 2023, [Online]. Available: <https://ejournal.unsri.ac.id/index.php/resti/article/view/14567>
- [11] Z. Mamieva, E. Gubanov, and D. Vasiliev, "Systematic review of emotion classification in song lyrics using NLP," *Procedia Comput. Sci.*, vol. 219, pp. 920–928, 2023, [Online]. Available: <https://doi.org/10.1016/j.procs.2023.01.215>
- [12] N. Sidora and M. Harani, "Pemanfaatan metadata musik untuk sistem rekomendasi berbasis emosi," *J. Teknol. dan Data*, vol. 8, no. 1, pp. 25–36, 2024.
- [13] S. Gupta, "Affective computing for task management: Integrating facial expression and background music for stress mitigation," *IEEE Trans. Human-Machine Syst.*, vol. 54, no. 3, pp. 213–225, 2024.
- [14] N. Khadijah, "Emotion-aware music recommendation for postpartum mothers: A psychological approach," *Psikol. dan Kesehat. Ment.*, vol. 12, no. 2, pp. 99–109, 2023.
- [15] D. R. Nugraha, T. M. Sari, and Y. Hartono, "Neuropsikologi musik dan deteksi emosi: Kajian literatur integratif," *J. Psikol. Musik*, vol. 2, no. 1, pp. 60–72, 2024.
- [16] A. P. Kusuma, "Cultural emotion analysis in TikTok music trends: The role of facial expression in popularity metrics," *J. Digit. Soc. Stud.*, vol. 6, no. 1, pp. 55–67, 2025.