

BAB I

PENDAHULUAN

1.1 Latar Belakang

Ekspresi wajah sebagai elemen penting dalam interaksi manusia-komputer telah mengalami kemajuan besar sebagai hasil dari pengembangan teknologi deep learning dalam bidang visi komputer ((M, 2025); (Sultan et al., 2023)). Memahami emosi secara otomatis dapat digunakan untuk banyak hal, seperti sistem pemantauan keamanan, evaluasi respons iklan, dan platform e-learning yang fleksibel. Menciptakan model yang tidak hanya akurat dalam kondisi terkontrol tetapi juga kuat dalam skenario dunia nyata dengan variasi pencahayaan, pose, dan okulasi adalah tantangan utama yang masih dihadapi ((Singh et al., 2023); (Alkhaleeli, 2024)).

Tinjauan literatur terbaru menunjukkan bahwa sistem pengenalan emosi masih memiliki tingkat akurasi yang beragam dan seringkali belum optimal untuk data yang kompleks. Studi (Sarvakar et al., 2025) menemukan bahwa dataset FER2013 memiliki akurasi di bawah 77%. Di sisi lain, penelitian (Mohana et al., 2023) menemukan bahwa model menghadapi kesulitan membedakan emosi yang mirip, seperti sedih dan marah. (Manalu & Rifai, 2024) menemukan hasil serupa, tetapi akurasi model hybrid CNN-RNN tetap berada di kisaran 59–66% untuk dataset video. Karena kurangnya akurasi ini, keandalan sistem tidak dapat digunakan dalam situasi real-time yang membutuhkan presisi tinggi (Achmad et al., 2023).

Muncul peluang untuk mengeksplorasi arsitektur hybrid alternatif yang lebih efisien. Gated Recurrent Unit (GRU) menawarkan mekanisme yang lebih sederhana dengan hanya dua gate, yang secara teoritis dapat mengurangi beban komputasi tanpa mengorbankan kinerja. Beberapa penelitian seperti (Jayaraman & Mahendran, 2025) dan (Makhmudov et al., 2024) telah membuktikan efektivitas GRU dalam domain pengenalan emosi dari audio, namun penerapannya secara

hybrid dengan CNN untuk klasifikasi ekspresi wajah masih relatif terbatas. Eksplorasi ini penting mengingat kebutuhan akan model yang tidak hanya akurat tetapi juga efisien untuk aplikasi real-time.

Convolutional Neural Network (CNN) telah luas diterapkan dalam klasifikasi citra karena kemampuannya mengekstraksi fitur spasial secara bertingkat dari data visual. Dalam pengenalan ekspresi wajah, CNN dapat mendeteksi pola-pola krusial seperti bentuk mata, posisi alis, dan konfigurasi mulut yang menjadi penanda emosi. Meskipun demikian, CNN memproses setiap gambar secara terpisah sehingga tidak menangkap dinamika perubahan ekspresi yang berlangsung secara berurutan.

Di sisi lain, penelitian terdahulu menunjukkan bahwa Gated Recurrent Unit (GRU) efektif dalam memodelkan data berurutan dengan struktur yang lebih sederhana dibandingkan LSTM (Lim & Handhayani, 2025). GRU memiliki dua mekanisme gerbang, yaitu update gate dan reset gate, yang mampu mempertahankan informasi relevan serta mengatasi permasalahan vanishing gradient pada RNN konvensional (Saputra et al., 2023) dan (Mulyani, 2025). Beberapa studi juga menunjukkan bahwa GRU menghasilkan tingkat error yang rendah serta stabil dalam pemodelan data sekuensial (Ardi Lestari et al., 2025). Efisiensi jumlah parameter pada GRU menjadikannya lebih ringan secara komputasi, sehingga lebih sesuai untuk aplikasi yang memerlukan proses inferensi cepat (Danestiara et al., 2024).

Berdasarkan hal tersebut, penggabungan CNN dan GRU dalam arsitektur hibrida menjadi pendekatan yang rasional untuk klasifikasi emosi wajah berbasis rangkaian frame. CNN bertugas mengekstraksi fitur spasial dari tiap citra, sementara GRU mempelajari keterkaitan temporal antar fitur dalam urutan waktu (Nugraha et al., 2025). Kombinasi ini diharapkan mampu meningkatkan akurasi dalam membedakan emosi yang secara visual mirip, sekaligus tetap menjaga efisiensi komputasi untuk penerapan secara real-time.

Tujuan dari penelitian ini adalah untuk mengembangkan dan menguji kinerja algoritma hybrid CNN-GRU untuk klasifikasi emosi berdasarkan ekspresi wajah. Arsitektur model (CNN-GRU) adalah variabel bebas dan akurasi klasifikasi adalah variabel terikat (Achmad et al., 2023)(Mohammad Sufa Ammar Habibi et al., 2025) dan (Handhayani et al., 2023). Fokus penelitian adalah mengembangkan model algoritma hybrid CNN-GRU untuk pengenalan wajah dalam klasifikasi emosi yang mampu mengekstraksi fitur spasial dan temporal secara lebih efektif (Mas Diyasa et al., 2025).

Penelitian ini menggunakan pendekatan kuantitatif karena tujuan penelitian adalah untuk mengukur pengaruh penerapan arsitektur hybrid CNN-GRU terhadap akurasi klasifikasi yang lebih baik. Metode ini memungkinkan pengujian hipotesis secara statistik, perbandingan numerik model, dan generalisasi hasil dari eksperimen dengan data yang terkontrol dan terukur. Oleh karena itu, pendekatan kuantitatif akan memberikan bukti yang jelas tentang seberapa efektif algoritma yang diusulkan untuk menyelesaikan masalah efisiensi dan akurasi komputasi pada sistem pengenalan emosi.

1.2 Identifikasi Masalah

Berdasarkan latar belakang tersebut, terdapat beberapa permasalahan yang dapat diidentifikasi:

1. Model CNN-GRU berukuran 48×48 masih sulit menangkap detail spesial halus pada area wajah seperti “*takut* dan *jijik*”.
2. Belum tersedia implementasi real-time berbasis webcam yang secara langsung memanfaatkan model 48×48 dengan pendekatan sequence frame untuk menangkap perubahan emosi antar-frame pada dataset FER-2013.

3. Belum ada pengujian terhadap performa hybrid CNN-GRU pada kondisi real-time seperti variasi pencahayaan buruk, pose wajah miring, dan ekspresi spontan dalam aplikasi webcam real-time.

1.3 Rumusan Masalah

Berdasarkan latar belakang tersebut, terdapat beberapa permasalahan yang dapat diidentifikasi:

1. Bagaimana merancang arsitektur hybrid CNN-GRU yang mampu menangkap fitur spasial dan temporal ekspresi wajah secara lebih optimal dibandingkan model CNN konvensional?
2. Bagaimana efektivitas model hybrid CNN-GRU dalam meningkatkan akurasi klasifikasi emosi, khususnya untuk emosi yang memiliki kemiripan seperti sedih-marah atau netral-spontan?
3. Bagaimana performa model hybrid CNN-GRU ketika diterapkan pada kondisi dunia nyata yang memiliki variasi pencahayaan, pose wajah, dan ekspresi spontan dalam proses inferensi real-time?

1.4 Tujuan Penelitian

Penelitian ini bertujuan untuk:

1. Mengembangkan arsitektur hybrid CNN-GRU yang mampu mengekstraksi fitur spasial dan temporal.
2. Menguji akurasi dan keandalan model hybrid CNN-GRU dalam mengklasifikasikan tujuh emosi wajah.
3. Mengevaluasi performa model hybrid CNN-GRU dalam skenario real-time (pencahayaan, pose, ekspresi spontan).

1.5 Manfaat Penelitian

Penelitian ini diharapkan memberikan manfaat sebagai berikut:

1. Menambah wawasan dan referensi terkait penerapan algoritma hybrid CNN-GRU dalam klasifikasi emosi.
2. Model CNN-GRU dapat diterapkan pada berbagai aplikasi seperti e-learning, monitoring, kesehatan digital.
3. Menjadi dasar bagi pengembangan sistem kecerdasan buatan yang mampu melakukan analisis ekspresi wajah secara real-time dengan kecepatan dan akurasi yang memadai.
4. Menghasilkan model dan dataset terstruktur berbasis sequence tiga frame yang dapat digunakan sebagai dasar pengembangan dan penelitian lanjutan pada sistem analisis ekspresi wajah berbasis temporal

1.6 Batasan Masalah

Agar penelitian lebih terarah, ruang lingkup penelitian ini dibatasi pada hal-hal berikut:

1. Input data yang digunakan berupa citra ekspresi wajah grayscale berukuran 48×48 piksel, sesuai proses preprocessing pada tahap pelatihan dan pengujian model.
2. Model hybrid CNN-GRU dilatih dan diuji menggunakan urutan tiga frame (sequence length = 3) untuk merepresentasikan dinamika temporal ekspresi wajah.
3. Proses real-time pada webcam hanya mendeteksi satu wajah setiap kali prediksi, mengikuti mekanisme deteksi wajah Haar Cascade pada sistem.
4. Klasifikasi emosi dibatasi pada tujuh kelas: angry, disgust, fear, happy, sad, surprise, dan neutral, sesuai struktur dataset dan konfigurasi model.
5. Evaluasi model difokuskan pada akurasi dan kestabilan prediksi, tanpa menilai aspek komputasi seperti kecepatan inferensi, FPS, atau penggunaan memori.
6. Pengujian kondisi real-time terbatas pada variasi pencahayaan, pose wajah, dan ekspresi spontan melalui webcam tanpa skenario ekstrem tambahan.

1.7 Sistematika Penulisan

Skripsi ini disusun dengan sistematika sebagai berikut:

- Bab I Pendahuluan:

Memberikan penjelasan tentang latar belakang, rumusan masalah, batasan dan tujuan penelitian, serta keuntungan dan kerugian dari penelitian.

- Bab II Tinjauan Pustaka:

Membahas penelitian sebelumnya, serta teori dan konsep penelitian ini.

- Bab III Metodologi Penelitian:

Memberikan penjelasan tentang metode, arsitektur CNN-GRU.

- Bab IV Hasil dan Pembahasan:

Menampilkan hasil penerapan pada data.

- Bab V Kesimpulan dan Saran:

Menampilkan kesimpulan penelitian.

1.8 Mata Kuliah

Matakuliah yang mendasari skripsi sebagai berikut:

- Pengolahan Citra Digital – Mempelajari teknik pra-pemrosesan citra, segmentasi, dan ekstraksi fitur visual untuk mendukung sistem klasifikasi berbasis wajah.
- Algoritma – Membahas perancangan algoritma efisien dan analisis kompleksitas untuk mendukung pengembangan arsitektur CNN-GRU.
- Deep Learning – Fokus pada arsitektur CNN dan GRU, pelatihan model, serta evaluasi performa dalam klasifikasi emosi berbasis ekspresi wajah