

BAB I

PENDAHULUAN

1.1 Latar Belakang

Pengenalan emosi manusia memiliki peran penting dalam memahami perilaku dan respons seseorang, baik dalam bidang psikologi, pendidikan, maupun interaksi manusia-komputer. Metode tradisional seperti kuesioner atau wawancara sering kali menghadapi keterbatasan karena responden dapat memberikan jawaban yang tidak mencerminkan kondisi emosinya yang sebenarnya. Oleh karena itu, diperlukan pendekatan yang lebih objektif dalam mengenali kondisi emosional manusia. Salah satu pendekatan yang semakin dikembangkan adalah *Facial Expression Recognition* (FER), yaitu kemampuan sistem komputer untuk mengenali dan menginterpretasikan ekspresi wajah manusia yang terekam dalam citra ataupun video. Penelitian-penelitian nasional di Indonesia telah mengembangkan model berbasis *Convolutional Neural Network* (CNN) yang mampu mengklasifikasi ekspresi senang, sedih, marah, takut, terkejut, maupun netral dengan hasil yang menjanjikan. Penerapan FER tidak hanya terbatas pada klasifikasi ekspresi dasar, tetapi juga menjadi fondasi bagi sistem cerdas yang memahami kondisi emosional manusia secara lebih mendalam. Dengan adanya FER, peluang untuk menciptakan teknologi interaktif yang lebih alami, intuitif, dan responsif semakin terbuka, terutama dalam meningkatkan kualitas komunikasi antara manusia dan komputer[1].

Di era digital saat ini, aplikasi FER semakin luas dan tidak hanya terbatas pada penelitian akademik. Teknologi ini telah diadopsi dalam berbagai sektor seperti kesehatan, keamanan, pendidikan, dan hiburan. Misalnya, dalam bidang kesehatan, FER dapat digunakan untuk membantu mendiagnosis kondisi psikologis pasien berdasarkan perubahan ekspresi wajah yang muncul secara spontan. Dalam bidang keamanan, sistem ini dapat meningkatkan efektivitas *surveillance* dengan mendeteksi ekspresi mencurigakan sebagai indikator potensi ancaman. Sementara pada ranah hiburan digital seperti *gaming*, FER berfungsi untuk menyesuaikan mekanisme permainan secara dinamis berdasarkan emosi pemain yang terdeteksi, sehingga pengalaman bermain menjadi lebih personal, imersif, dan responsif[2].

Untuk mewujudkan hal tersebut, dibutuhkan metode komputasi yang andal dalam mengenali pola-pola kompleks dari ekspresi wajah. *Convolutional Neural Network* (CNN) telah menjadi pendekatan dominan dalam FER karena kemampuannya mengekstraksi fitur hierarkis dari data visual. Salah satu arsitektur CNN yang banyak digunakan adalah ResNet, yang diperkenalkan dengan mekanisme *residual block* sehingga mampu mengatasi masalah *vanishing gradient* pada jaringan dalam. ResNet terbukti memberikan akurasi tinggi dalam berbagai tugas pengenalan wajah dan ekspresi, menjadikannya kandidat ideal untuk aplikasi FER[3]. Di antara varian yang ada, ResNet50 banyak digunakan karena menawarkan keseimbangan antara kedalaman arsitektur, akurasi, dan efisiensi komputasi. Selain itu, penggunaan teknik *transfer learning* dengan memanfaatkan bobot *pre-trained* dari dataset besar seperti ImageNet dapat mempercepat proses pelatihan sekaligus menjaga performa meskipun dataset ekspresi wajah yang tersedia relatif terbatas[4].

Lebih lanjut, ResNet50V2 hadir sebagai pengembangan dari ResNet50 dengan penyempurnaan pada aliran gradien dan normalisasi batch, sehingga proses pelatihan menjadi lebih stabil dan efisien. Berbagai penelitian menunjukkan bahwa model berbasis ResNet menghasilkan performa lebih unggul dibandingkan metode tradisional berbasis fitur manual seperti Local Binary Patterns (LBP) atau Histogram of Oriented Gradients (HOG)[5]. Namun demikian, penerapan FER di dunia nyata tidak terlepas dari tantangan, terutama pada dataset dengan tingkat *noise* tinggi seperti FER2013 yang memiliki variasi besar pada pencahayaan, pose wajah, dan ekspresi tidak konsisten. Kondisi ini mengharuskan peneliti untuk mengombinasikan arsitektur *deep learning* dengan strategi tambahan seperti *fine-tuning*, *data augmentation*, serta penggunaan *classifier* yang lebih kuat[6].

Salah satu pendekatan yang menjanjikan adalah mengombinasikan *feature extractor* dari ResNet50V2 dengan *Support Vector Machine* (SVM) sebagai *classifier*. SVM dikenal efektif dalam menemukan margin pemisah optimal di ruang fitur berdimensi tinggi, sehingga dapat meningkatkan akurasi dalam tugas klasifikasi multi-ekspresi. Dibandingkan *softmax classifier* bawaan CNN, integrasi dengan SVM seringkali menghasilkan model yang lebih stabil, terutama ketika berhadapan dengan dataset berukuran kecil atau tidak seimbang. Beberapa

penelitian terdahulu bahkan membuktikan bahwa kombinasi *deep feature* dari CNN dengan SVM dapat memperbaiki hasil klasifikasi dan meningkatkan generalisasi model secara signifikan[7].

Sejumlah penelitian sebelumnya juga menyoroti bahwa integrasi deep learning dengan metode klasifikasi tradisional, khususnya CNN sebagai feature extractor dan SVM sebagai classifier, dapat menghasilkan performa yang lebih unggul dibandingkan penggunaan CNN murni. Hal ini terutama terlihat pada kondisi dataset yang terbatas, tidak seimbang, atau mengandung noise, di mana CNN sering kali mengalami *overfitting*. Dengan adanya SVM, proses klasifikasi dapat lebih robust karena algoritma ini memaksimalkan margin antar kelas pada ruang fitur yang kompleks. Beberapa studi bahkan melaporkan peningkatan akurasi signifikan ketika SVM digunakan untuk menggantikan lapisan klasifikasi fully connected pada CNN, sehingga kombinasi ini dianggap lebih stabil dan adaptif dalam menghadapi variasi data[8].

Berdasarkan penelitian-penelitian di atas, dapat disimpulkan bahwa topik *Facial Expression Recognition* (FER) terus mengalami perkembangan pesat seiring dengan kemajuan teknologi *deep learning*. Penelitian terdahulu menunjukkan bahwa FER berperan penting dalam pengembangan sistem interaksi manusia dan komputer yang mampu memahami kondisi emosional penggunanya secara lebih alami. Metode *Convolutional Neural Network* (CNN) menjadi pendekatan dominan karena kemampuannya dalam mengekstraksi fitur visual yang kompleks dari citra wajah. Salah satu arsitektur CNN yang banyak digunakan adalah ResNet, khususnya varian ResNet50 dan ResNet50V2. ResNet50V2 dikembangkan untuk menyempurnakan stabilitas pelatihan melalui mekanisme *residual block* dan *batch normalization* yang lebih efisien, sehingga mampu menghasilkan representasi fitur yang lebih kuat. Namun, tantangan masih muncul pada dataset ekspresi wajah yang terbatas, tidak seimbang, dan mengandung *noise*, yang dapat menyebabkan model mengalami *overfitting* dan penurunan akurasi. Untuk mengatasi hal tersebut, penelitian ini mengusulkan kombinasi antara ResNet50V2 sebagai *feature extractor* dan *Support Vector Machine* (SVM) sebagai *multiclass classifier*. Pendekatan ini diharapkan mampu meningkatkan akurasi klasifikasi setidaknya sekitar 2% dibandingkan penelitian sebelumnya, dengan hasil yang lebih stabil dan

generalisasi yang lebih baik. Selain itu, metode ini juga memiliki potensi untuk diterapkan pada analisis perilaku, maupun *human-computer interaction* secara real-time. Penelitian ini diharapkan dapat memberikan kontribusi nyata dalam pengembangan sistem FER yang lebih akurat, efisien, dan adaptif, serta menjadi dasar bagi penelitian selanjutnya untuk menyempurnakan model melalui strategi *fine-tuning* dan *data augmentation* yang lebih optimal.

1.2 Perumusan Masalah

1.2.1 Permasalahan

Deteksi ekspresi emosi pada citra wajah telah mengalami perkembangan signifikan seiring dengan kemajuan *deep learning*, namun penerapannya masih menghadapi berbagai kendala. Variasi pencahayaan, sudut pandang, occlusion, serta ekspresi yang tidak lengkap dapat menurunkan akurasi sistem. Selain itu, meskipun banyak arsitektur CNN seperti ResNet telah terbukti memberikan hasil baik, sebagian model memiliki kompleksitas berlebihan sehingga membutuhkan sumber daya komputasi yang tinggi. Kondisi ini menimbulkan tantangan ketika model diterapkan pada lingkungan dengan keterbatasan perangkat keras maupun kebutuhan inferensi cepat. Di sisi lain, metode klasifikasi bawaan CNN (softmax) kadang kurang optimal pada dataset yang terbatas dan tidak seimbang, sehingga dibutuhkan pendekatan alternatif yang dapat meningkatkan akurasi sekaligus menjaga efisiensi.

1.2.2 Metode Usulan

Penelitian ini mengusulkan pemanfaatan arsitektur ResNet50V2 dengan pendekatan *transfer learning* untuk mengekstraksi fitur citra wajah, yang kemudian dikombinasikan dengan *Multiclass Support Vector Machine* (SVM) sebagai *classifier*. Kombinasi ini diharapkan mampu meningkatkan akurasi klasifikasi ekspresi wajah dibandingkan penggunaan CNN murni, serta memberikan generalisasi yang lebih baik pada dataset yang terbatas dan mengandung *noise*.

1.2.3 Pertanyaan Penelitian

1. Bagaimana rancangan dan implementasi sistem deteksi ekspresi emosi berbasis *transfer learning* ResNet50V2 dengan Multiclass SVM pada citra wajah?
2. Bagaimana kinerja sistem yang diusulkan dalam mendeteksi ekspresi emosi, ditinjau dari metrik evaluasi seperti akurasi, presisi, *recall*, dan *F1-score*, serta dibandingkan dengan pendekatan sebelumnya?

1.3 Tujuan dan Manfaat

1.3.1 Tujuan Penelitian

Penelitian ini bertujuan untuk:

1. Merancang dan mendapatkan sistem deteksi ekspresi emosi pada citra wajah menggunakan *transfer learning* ResNet50V2 yang dikombinasikan dengan Multiclass SVM.
2. Mengevaluasi dan mendapatkan hasil kinerja sistem yang dibangun berdasarkan metrik akurasi, presisi, *recall*, dan *F1-score*, serta membandingkannya dengan pendekatan sebelumnya untuk melihat peningkatan performa.

1.3.2 Manfaat Penelitian

Manfaat yang diharapkan dari penelitian ini meliputi:

1. Memberikan kontribusi dalam pengembangan metode deteksi ekspresi wajah berbasis *deep learning* dengan akurasi yang lebih baik.
2. Menyediakan alternatif pendekatan bagi peneliti atau praktisi dalam mengintegrasikan CNN dengan *classifier* tradisional untuk meningkatkan performa klasifikasi.
3. Mendukung penelitian lebih lanjut mengenai penerapan *transfer learning* dan SVM pada dataset ekspresi wajah yang terbatas maupun tidak seimbang.
4. Menjadi acuan awal bagi pengembangan sistem berbasis emosi yang membutuhkan deteksi wajah secara akurat dan efisien.

1.4 Batasan Masalah

1. Penelitian ini hanya berfokus pada deteksi ekspresi emosi dasar (seperti senang, sedih, marah, takut, terkejut, jijik, dan netral).
2. Dataset yang digunakan terbatas pada dataset citra wajah publik yang umum digunakan di bidang FER.
3. Model dikembangkan untuk konteks citra wajah 2D dan tidak mengoptimalkan performa khusus untuk perangkat berbasis VR atau sistem multimodal.
4. Evaluasi kinerja dilakukan melalui pengujian model pada dataset, tanpa melibatkan uji coba langsung dengan pengguna dalam skala besar.
5. Penelitian ini hanya memanfaatkan informasi visual dari citra wajah dan tidak mempertimbangkan faktor lain seperti suara, gerakan tubuh, atau konteks lingkungan sebagai indikator tambahan emosi.

1.5 Sistematika Penulisan Skripsi

Sistematika penulisan laporan penelitian ini adalah sebagai berikut:

BAB I PENDAHULUAN : Bab ini berisi latar belakang penelitian, perumusan masalah, tujuan, serta manfaat yang ingin dicapai. Selain itu, dijelaskan pula batasan penelitian agar ruang lingkup lebih terarah. Bagian akhir memuat sistematika penulisan laporan.

BAB II TINJAUAN PUSTAKA : Bab ini membahas teori-teori dasar yang mendukung penelitian, seperti konsep *computer vision*, pengenalan ekspresi emosi, dan metode deteksi berbasis *deep learning*. Uraian dilanjutkan dengan *transfer learning*, arsitektur ResNet50V2, serta algoritma *Support Vector Machine (SVM)*. Selain itu, ditinjau juga penelitian terdahulu sebagai pembandingan dan acuan pengembangan.

BAB III METODOLOGI PENELITIAN : Bab ini memaparkan desain penelitian yang digunakan serta penjelasan mengenai dataset yang digunakan . Tahapan yang dibahas meliputi *preprocessing* data, ekstraksi fitur dengan ResNet50V2, dan klasifikasi dengan SVM. Selain itu, dijelaskan juga metode pelatihan serta validasi model untuk mengevaluasi kinerja sistem deteksi emosi.

BAB IV HASIL DAN PEMBAHASAN :

Bab ini membahas hasil implementasi sistem deteksi ekspresi emosi menggunakan *transfer learning* ResNet50V2 dan *Multiclass* SVM. Uraian meliputi proses pelatihan model, hasil pengujian pada dataset CK+ dan FER2013, serta analisis performa model menggunakan metrik akurasi, *precision*, *recall*, *F1-score*, *confusion matrix*, *learning curve*, dan *ROC curve*. Selain itu, dijelaskan pula implementasi sistem berbasis Streamlit untuk prediksi ekspresi emosi secara interaktif.

BAB V PENUTUP :

Bab ini berisi kesimpulan dari hasil penelitian yang telah dilakukan serta saran untuk pengembangan penelitian selanjutnya. Kesimpulan mencakup pencapaian tujuan penelitian dan performa model yang dihasilkan, sedangkan saran memuat rekomendasi perbaikan metode, penggunaan dataset yang lebih beragam, maupun pengembangan sistem deteksi emosi di masa mendatang.