

BAB 1

PENDAHULUAN

1.1 Latar Belakang

Facial Expression Recognition (FER) merupakan bidang dalam *computer vision* yang mempelajari pengenalan ekspresi wajah manusia secara otomatis dari citra digital. Dalam interaksi manusia, keadaan emosi tidak selalu disampaikan secara langsung melalui kata-kata. Sebagian informasi tersebut ditunjukkan melalui tanda nonverbal, salah satunya perubahan ekspresi wajah. Manusia dapat menafsirkan tanda tersebut dalam komunikasi sehari-hari, sedangkan sistem komputasi yang hanya mengolah data tekstual atau numerik tidak secara otomatis mampu memahaminya. Akibatnya, respons sistem dapat tidak sesuai dengan keadaan afektif pengguna. FER diperlukan untuk mengubah pola visual pada wajah menjadi informasi yang dapat diproses oleh komputer sehingga interaksi manusia-komputer dapat berlangsung lebih alami dan responsif [2].

Informasi ekspresi yang diperoleh melalui FER dapat digunakan sebagai salah satu masukan bagi sistem yang berinteraksi dengan manusia. Sebagai contoh, robot sosial atau asisten virtual dapat memanfaatkan hasil pengenalan ekspresi untuk menyesuaikan gaya komunikasinya [1]. Contoh tersebut menunjukkan bahwa tujuan FER bukan untuk menentukan keseluruhan keadaan psikologis seseorang, melainkan mengenali pola ekspresi yang tampak pada wajah agar dapat diproses oleh sistem. Seiring berkembangnya *deep learning*, *Convolutional Neural Network* (CNN) banyak diterapkan pada FER karena mampu mempelajari ciri visual secara otomatis, mulai dari pola dasar seperti tepi dan tekstur hingga representasi ekspresi wajah yang lebih kompleks [3, 4]. Oleh karena itu, CNN relevan digunakan sebagai pendekatan dasar untuk membangun dan mengevaluasi model klasifikasi ekspresi wajah.

Meskipun kinerja FER terus berkembang, penerapannya pada kondisi nyata masih menghadapi sejumlah tantangan. Dataset yang diperoleh dari lingkungan *in-the-wild* umumnya memiliki variasi visual yang lebih tinggi dibandingkan dataset laboratorium. Variasi tersebut mencakup perbedaan pose kepala, pencahayaan, oklusi, identitas, usia, kualitas citra, hingga ekspresi yang ambigu [5, 6, 7]. Kondisi ini membuat model lebih sulit mengenali ekspresi secara konsisten karena wajah dengan label ekspresi yang sama tidak selalu memiliki karakteristik visual yang

serupa. Salah satu dataset yang dirancang untuk menguji FER pada kondisi dunia nyata adalah *Real-world Affective Faces Database* (RAF-DB). Pada subset ekspresi dasar, RAF-DB terdiri dari tujuh kelas, yaitu *angry*, *disgust*, *fear*, *happy*, *sad*, *surprise*, dan *neutral*, dengan 12.271 citra latih dan 3.068 citra uji [1, 8].

Selain variasi visual yang besar, RAF-DB juga memiliki distribusi jumlah data yang tidak merata antar kelas. Pada data latih resmi, kelas *happy* memiliki 4.772 sampel, sedangkan kelas *fear* hanya memiliki 281 sampel [1]. Dengan demikian, kelas mayoritas muncul hampir 17 kali lebih sering daripada kelas minoritas. Zhou dkk. menyatakan bahwa RAF-DB cenderung mengikuti pola distribusi *long-tailed*, sehingga model lebih mudah mengenali kelas mayoritas, tetapi kurang optimal pada kelas dengan jumlah sampel relatif sedikit seperti *fear*, *disgust*, dan *angry* [9]. Kondisi ini mendesak untuk ditangani karena akurasi keseluruhan dapat terlihat tinggi meskipun sebagian ekspresi minoritas sering tidak dikenali.

Penanganan ketidakseimbangan tidak cukup hanya dengan menerapkan transformasi citra pada distribusi yang tetap didominasi kelas mayoritas. Model perlu menerima paparan kelas yang terkendali, sementara salinan kelas minoritas perlu dibuat cukup beragam agar tidak sekadar mengulang citra yang sama. Oleh karena itu, penelitian ini menggunakan *random oversampling* pada data latih untuk membentuk jumlah yang sama pada tujuh kelas dan *balanced batch* untuk memastikan setiap kelas muncul dalam jumlah identik pada setiap *batch*. Augmentasi hanya diberikan pada baris tambahan hasil *oversampling*. Rancangan tersebut memisahkan fungsi penyeimbangan jumlah dari fungsi penambahan variasi visual.

Data augmentation digunakan untuk meningkatkan keragaman visual data latih melalui transformasi yang tidak mengubah label ekspresi. Shorten dan Khoshgoftaar mengelompokkan augmentasi citra ke dalam transformasi geometrik, fotometrik atau ruang warna, teknik berbasis *erasing*, pencampuran citra, augmentasi ruang fitur, dan pendekatan berbasis GAN [10]. Penelitian ini memusatkan pengujian pada tiga kategori yang relevan terhadap variasi RAF-DB, yaitu geometrik, fotometrik, dan *erasing/oklusi*. Transformasi geometrik mencakup rotasi, pergeseran, *zoom*, *horizontal flip*, dan *RandomCrop*; transformasi fotometrik mencakup perubahan kecerahan dan *ColorJitter*; sedangkan *CoarseDropout* digunakan untuk mensimulasikan oklusi parsial yang secara konsep berkaitan dengan *Random Erasing* [11].

Beberapa penelitian FER sebelumnya telah menggunakan kombinasi

augmentasi yang beragam. LA-Net menerapkan *random cropping*, *random horizontal flipping*, *random erasing*, dan *random color jitter* [12]. FER-HA menggunakan *random horizontal flip*, rotasi, *CoarseDropout*, dan *ColorJitter* [13]. MixAugment dan MixCut juga menunjukkan bahwa rancangan augmentasi dapat memengaruhi performa FER [14, 15]. Namun, penggunaan arsitektur dan strategi pelatihan yang berbeda membuat kontribusi penyeimbangan kelas dan komposisi augmentasi sulit dipisahkan. Selain itu, augmentasi yang digunakan tanpa pembuktian distribusi kelas seimbang belum menjamin bahwa model tidak tetap didominasi kelas mayoritas.

Berdasarkan permasalahan tersebut, penelitian ini dilakukan pada RAF-DB untuk mengetahui bagaimana ketidakseimbangan kelas memengaruhi CNN dan bagaimana kombinasi penyeimbangan data serta augmentasi dapat memperbaiki performa antarkelas. Parameter model tidak ditentukan secara subjektif, tetapi dipilih melalui sepuluh percobaan *random search*; intensitas augmentasi dipilih melalui 12 kandidat berbasis data validasi. Eksperimen akhir dirancang dalam tujuh skenario, yaitu distribusi asli tanpa augmentasi, *oversampling* seimbang tanpa augmentasi, augmentasi umum, tiga penambahan teknik secara terpisah, dan augmentasi gabungan. Setiap skenario dievaluasi menggunakan beberapa *random seed*, *macro F1-score*, metrik per kelas, distribusi prediksi, dan *confusion matrix*. Dengan rancangan tersebut, penelitian dapat mengukur dampak ketidakseimbangan pada kelas minoritas RAF-DB, menentukan konfigurasi berdasarkan data validasi, dan menilai apakah penyeimbangan serta kombinasi augmentasi memperbaiki performa secara konsisten.

1.2 Rumusan Masalah

1. Bagaimana *class imbalance* RAF-DB memengaruhi performa CNN pada distribusi asli, khususnya dalam mengenali kelas *fear*, *disgust*, dan *angry*?
2. Bagaimana pengaruh *oversampling* seimbang dan *balanced batch* terhadap performa global, pemerataan *recall*, dan kecenderungan prediksi kelas mayoritas dibandingkan CNN pada distribusi asli?
3. Bagaimana pengaruh augmentasi umum serta penambahan *RandomCrop*, *ColorJitter*, dan *CoarseDropout* terhadap performa global dan performa setiap kelas ekspresi?

4. Apakah penggabungan transformasi geometrik, fotometrik, dan oklusi pada data latih seimbang memberikan solusi yang lebih efektif terhadap dampak *class imbalance* dibandingkan penggunaan masing-masing teknik secara terpisah?

1.3 Batasan Permasalahan

1. Dataset yang digunakan adalah RAF-DB subset ekspresi dasar dengan tujuh kelas, 12.271 citra latih, dan 3.068 citra uji; subset ekspresi majemuk tidak digunakan.
2. Data dibatasi pada citra wajah statis RGB berukuran 100×100 piksel, sesuai hasil audit dimensi citra sumber, tanpa melibatkan video, audio, teks, atau data multimodal. Nilai piksel diskalakan dari 0–255 menjadi 0–1 tanpa mengubah tiga kanal warna.
3. Data latih resmi dibagi secara terstratifikasi menjadi 85% data latih efektif dan 15% data validasi, sedangkan 3.068 citra uji resmi hanya digunakan pada evaluasi akhir.
4. Model yang digunakan adalah CNN dengan komponen utama *Conv2D*, *Batch Normalization*, *Max Pooling*, *Dropout*, *Global Average Pooling*, *Dense*, dan *Softmax*. Parameter model dipilih melalui sepuluh percobaan *random search* pada ruang parameter yang telah ditetapkan.
5. Penelitian tidak menggunakan arsitektur atau metode kompleks seperti ResNet, EfficientNet, Vision Transformer, Swin Transformer, *attention model*, GAN, *pseudo-labeling*, *focal loss*, dan *dynamic label weighting*.
6. Teknik yang diteliti dibatasi pada *data augmentation* berbasis transformasi citra, yaitu augmentasi umum (*horizontal flip*, rotasi, *zoom*, *width/height shift*, dan *brightness adjustment*) serta augmentasi tambahan (*RandomCrop*, *ColorJitter*, dan *CoarseDropout*).
7. Penanganan ketidakseimbangan dibatasi pada *random oversampling*, *balanced batch*, dan augmentasi pada baris tambahan hasil *oversampling*. Penelitian tidak menggunakan *class weighting*, *undersampling*, SMOTE, augmentasi berbasis GAN, atau penambahan label baru.

8. Pencarian augmentasi dibatasi pada tiga tingkat intensitas augmentasi umum dan tiga tingkat intensitas untuk masing-masing teknik tambahan. Pemilihan konfigurasi menggunakan *validation macro F1-score*, bukan hasil data uji.
9. Eksperimen akhir dirancang menggunakan *random seed* 42, 123, dan 2026. Evaluasi mencakup *accuracy*, *balanced accuracy*, *macro precision*, *macro recall*, *macro F1-score*, *weighted F1-score*, *recall gap*, distribusi prediksi, laporan klasifikasi per kelas, dan *confusion matrix*.
10. Eksperimen dijalankan menggunakan Python dan TensorFlow/Keras pada Kaggle Notebook dengan akselerasi GPU, sehingga hasil penelitian terikat pada lingkungan komputasi tersebut.
11. Pembahasan dibatasi pada *benchmarking* algoritma klasifikasi ekspresi wajah dan tidak mencakup diagnosis, kesehatan mental, *mental score*, atau interpretasi psikologis dari hasil prediksi.

1.4 Tujuan Penelitian

1. Menganalisis pengaruh ketidakseimbangan distribusi RAF-DB terhadap performa CNN pada distribusi asli, terutama pada kelas *fear*, *disgust*, dan *angry*.
2. Mengevaluasi pengaruh *oversampling* seimbang dan *balanced batch* terhadap performa global, pemerataan *recall*, dan distribusi kelas prediksi.
3. Menganalisis pengaruh augmentasi umum, *RandomCrop*, *ColorJitter*, dan *CoarseDropout* terhadap performa global dan performa setiap kelas ekspresi.
4. Menentukan efektivitas gabungan transformasi geometrik, fotometrik, dan oklusi pada data latih seimbang berdasarkan *macro F1-score*, metrik per kelas, kestabilan antarseed, distribusi prediksi, dan *confusion matrix*.

1.5 Manfaat Penelitian

1. Memberikan bukti empiris mengenai pengaruh penyeimbangan kelas dan variasi *data augmentation* terhadap klasifikasi ekspresi wajah menggunakan CNN pada RAF-DB.

2. Menyajikan perbandingan terkontrol antara distribusi asli, *oversampling* seimbang, augmentasi umum, augmentasi tambahan, dan augmentasi gabungan.
3. Menunjukkan pentingnya metrik selain akurasi, seperti *balanced accuracy*, *macro F1-score*, *recall gap*, distribusi prediksi, dan analisis per kelas pada dataset *long-tailed*.
4. Menjadi referensi bagi penelitian berikutnya yang ingin mengevaluasi komposisi augmentasi geometrik, fotometrik, dan *erasing*/oklusi pada *Facial Expression Recognition*.
5. Menyediakan rancangan eksperimen yang dapat diaudit melalui pemeriksaan data, pencarian parameter berbasis validasi, *balanced batch*, dan pelaporan beberapa *random seed*.

1.6 Sistematika Penulisan

Sistematika penulisan skripsi ini disusun ke dalam lima bab yang saling berkaitan. Uraian singkat isi masing-masing bab adalah sebagai berikut.

1. **Bab I PENDAHULUAN** memuat latar belakang mengenai pengaruh *class imbalance* pada pengenalan ekspresi wajah menggunakan *dataset RAF-DB*. Bab ini juga menjelaskan rumusan masalah, batasan masalah, tujuan penelitian, manfaat penelitian, serta sistematika penulisan.
2. **Bab II LANDASAN TEORI** membahas konsep dan teori yang mendukung penelitian, meliputi *Facial Expression Recognition*, *dataset RAF-DB*, *Convolutional Neural Network (CNN)*, *class imbalance*, serta *data augmentation*. Bab ini juga memaparkan penelitian terdahulu yang berkaitan dengan augmentasi data dan klasifikasi ekspresi wajah.
3. **Bab III METODOLOGI PENELITIAN** menjelaskan audit dan pembagian *RAF-DB*, pembentukan data serta *batch* seimbang, pra-pemrosesan citra, arsitektur *CNN*, *random search* parameter model, pencarian intensitas augmentasi, tujuh skenario eksperimen, pelatihan beberapa *random seed*, dan metrik evaluasi.
4. **Bab IV HASIL DAN PEMBAHASAN** menyajikan bukti audit dan keseimbangan data, hasil pencarian parameter, perbandingan metrik global

dan per kelas, kestabilan antarseed, distribusi prediksi, serta *confusion matrix*. Pembahasan difokuskan pada pengaruh penyeimbangan dan variasi augmentasi terhadap kelas minoritas.

5. **Bab V PENUTUP** memuat kesimpulan yang menjawab rumusan masalah berdasarkan pengaruh penyeimbangan kelas dan variasi augmentasi terhadap performa CNN pada RAF-DB, serta saran untuk penelitian selanjutnya.

