

BAB 1

PENDAHULUAN

1.1 Latar Belakang Masalah

Perkembangan teknologi kecerdasan buatan (*Artificial Intelligence*) dan *Computer Vision* telah membawa transformasi signifikan dalam berbagai aspek kehidupan manusia [3]. Kemampuan mesin untuk memahami informasi visual, khususnya dalam konteks interaksi manusia, menjadi area riset yang berkembang pesat seiring meningkatnya kebutuhan sistem otomatis yang mampu memahami perilaku dan kondisi emosional manusia secara real-time [4]. Teknologi *deep learning* telah memungkinkan mesin melakukan tugas kompleks seperti pengenalan pola visual dan interpretasi makna dari data visual, yang sebelumnya hanya dapat dilakukan oleh manusia [5].

Facial Expression Recognition (FER) merupakan salah satu cabang penting dalam bidang *Computer Vision* yang bertujuan mengidentifikasi ekspresi wajah manusia secara otomatis [3]. Ekspresi wajah adalah komponen fundamental komunikasi non-verbal, di mana Paul Ekman mendefinisikan tujuh emosi dasar universal yang diakui lintas budaya: *happy, sad, angry, fear, surprise, disgust*, dan *neutral* [6]. Teknologi FER memiliki aplikasi yang semakin luas dalam kehidupan modern, di antaranya pemantauan keterlibatan dan kondisi emosional siswa dalam pembelajaran daring [7, 8, 9], deteksi dini kondisi psikologis seperti depresi dan kecemasan dalam sistem kesehatan mental [10], serta pengembangan sistem interaksi manusia-komputer yang lebih adaptif dan responsif [11]. Luasnya cakupan aplikasi ini mendorong kebutuhan akan sistem FER yang akurat, efisien secara komputasi, dan dapat diinterpretasikan.

Dataset FER2013 merupakan *benchmark* standar yang digunakan secara luas dalam penelitian FER [12]. Namun demikian, dataset asli ini memiliki kualitas anotasi yang rendah akibat proses *crowdsourcing* melalui Amazon Mechanical Turk oleh satu annotator per gambar, sehingga akurasi pengenalan manusia pada dataset ini hanya mencapai ~65% [8]. Untuk mengatasi permasalahan tersebut, penelitian ini menggunakan FERPlus, yaitu versi re-anotasi dari FER2013 yang dikembangkan oleh Microsoft melalui *majority voting* dari 10 annotator per gambar, sehingga kredibilitas label menjadi lebih baik [8]. Setelah proses pembersihan (pembuangan kelas *contempt, unknown*, dan *NF*, serta gambar dengan hasil voting

seri), dataset FERPlus yang digunakan pada penelitian ini berisi 33.500 gambar resolusi rendah (48×48 piksel) yang terbagi menjadi *training set* (26.778 gambar), *validation set* (3.372 gambar), dan *test set* (3.350 gambar). Meskipun kualitas label meningkat, dataset ini tetap memiliki tantangan fundamental berupa distribusi kelas yang sangat tidak seimbang (*class imbalance*): kelas *neutral* mendominasi dengan 9.365 sampel pada *training set*, sementara kelas *disgust* hanya memiliki 175 sampel, menghasilkan rasio ketidakseimbangan sebesar 53,5:1 yang bahkan lebih ekstrem dibandingkan FER2013 (16,5:1). Kondisi ini menyebabkan model cenderung bias terhadap kelas mayoritas dan mengabaikan kelas minoritas. Karakteristik inilah yang menjadikan FERPlus sebagai dataset yang menantang sekaligus relevan untuk mengukur ketangguhan suatu metode FER.

Berbagai penelitian terdahulu telah berupaya mengatasi tantangan tersebut, namun masing-masing masih memiliki keterbatasan. Huo et al. [13] mengajukan arsitektur *lightweight* berbasis kombinasi Xception dan MobileNetV2 dan memperoleh akurasi 70,76% pada FER2013, namun model ini menunjukkan kelemahan signifikan pada kondisi oklusi parsial seperti wajah tertutup sebagian. Fang et al. [14] mengusulkan pendekatan *deep joint learning* dengan teknik augmentasi yang menghasilkan akurasi 75,42%, tetapi augmentasi yang digunakan bersifat seragam untuk semua kelas sehingga tidak efektif dalam menangani kelas minoritas seperti *disgust* dan *fear*. Liao et al. [1] mengintegrasikan ResNet dengan mekanisme *attention* berbasis CBAM dan LBP serta mencapai akurasi 73,71%, namun penelitian ini tidak menyertakan strategi eksplisit untuk menangani *class imbalance* pada data pelatihan. Selain itu, hampir seluruh penelitian tersebut menggunakan model yang bersifat *black box* tanpa menyertakan mekanisme *explainability* yang memungkinkan pemahaman terhadap keputusan prediksi model [5]. Berdasarkan tinjauan tersebut, terdapat kesenjangan penelitian berupa belum adanya pendekatan yang secara simultan mengintegrasikan mekanisme *attention* berbasis CBAM, penanganan *class imbalance* melalui *Focal Loss* tanpa modifikasi komposisi data, dan *explainability* melalui Grad-CAM dalam satu arsitektur *lightweight* untuk FER pada dataset FERPlus. Keterbatasan ini menjadi dasar motivasi penelitian yang diusulkan dalam skripsi ini.

Pemilihan MobileNetV2 sebagai *backbone* dalam penelitian ini didasarkan pada beberapa pertimbangan teknis yang relevan. MobileNetV2 menggunakan arsitektur *inverted residual blocks* dengan *depthwise separable convolution* yang secara signifikan mengurangi jumlah parameter dan kompleksitas komputasi dibandingkan model CNN dalam seperti VGG-16 atau ResNet-50, tanpa

mengorbankan kemampuan ekstraksi fitur secara substansial [15]. Dengan total parameter sekitar 2,4 juta, MobileNetV2 cocok untuk skenario dengan sumber daya komputasi terbatas. Lebih lanjut, penggunaan bobot *pretrained* dari dataset ImageNet memungkinkan *transfer learning* yang efektif, di mana fitur-fitur umum seperti tepi, tekstur, dan pola visual yang telah dipelajari dapat dimanfaatkan kembali untuk tugas FER [11]. Penelitian Zhu et al. [15] juga telah membuktikan bahwa MobileNetV2 yang dimodifikasi mampu mencapai performa kompetitif pada FER2013 dengan efisiensi komputasi yang lebih tinggi dibandingkan arsitektur yang lebih besar.

Untuk mengatasi keterbatasan yang ditemukan pada penelitian terdahulu, penelitian ini mengintegrasikan tiga komponen tambahan ke dalam arsitektur MobileNetV2. Pertama, *Convolutional Block Attention Module* (CBAM) ditambahkan setelah lapisan fitur MobileNetV2 untuk memungkinkan model memfokuskan perhatian secara selektif pada region wajah yang paling diskriminatif, seperti area mata, alis, dan mulut, melalui mekanisme *channel attention* dan *spatial attention* secara sekuensial [1]. Kedua, untuk mengatasi *class imbalance* tanpa melakukan modifikasi pada komposisi data, digunakan *Focal Loss* yang dikombinasikan dengan *class weights* berbasis akar kuadrat frekuensi invers. *Focal Loss* bekerja dengan menurunkan bobot *loss* pada sampel yang sudah diprediksi dengan benar (*easy samples*), sehingga proses pelatihan lebih terfokus pada sampel yang sulit diprediksi, yang umumnya berasal dari kelas minoritas [14]. Pendekatan ini dipilih karena teknik *data balancing* seperti *oversampling* berisiko memperbanyak duplikasi pada kelas minoritas *disgust* yang jumlahnya sangat terbatas pada dataset FERPlus. Ketiga, *Grad-CAM* (*Gradient weighted Class Activation Mapping*) diimplementasikan sebagai mekanisme *explainability* untuk menghasilkan visualisasi *heatmap* yang menunjukkan region wajah yang paling berpengaruh dalam setiap keputusan klasifikasi [2].

Berdasarkan analisis terhadap penelitian terdahulu dan kebutuhan yang teridentifikasi, penelitian ini mengusulkan sistem FER yang mengintegrasikan MobileNetV2 dengan CBAM, *Focal Loss* dengan *class weights*, dan Grad-CAM dalam satu *framework* kohesif. Integrasi ini bertujuan untuk secara simultan mengatasi tiga tantangan utama dalam FER: (1) efisiensi komputasi melalui arsitektur *lightweight* MobileNetV2, (2) ketidakseimbangan kelas melalui *Focal Loss* yang adaptif, dan (3) kebutuhan interpretabilitas model melalui Grad-CAM. Dengan pendekatan ini, diharapkan dapat dicapai sistem FER yang tidak hanya akurat, tetapi juga dapat dipahami secara visual oleh pengguna akhir, khususnya

dalam konteks aplikasi pemantauan emosional pada pendidikan daring dan layanan kesehatan mental [16].

1.2 Rumusan Masalah

Berdasarkan latar belakang yang telah diuraikan, rumusan masalah dalam penelitian ini adalah sebagai berikut:

1. Bagaimana mengimplementasikan MobileNetV2 dengan integrasi CBAM dan *Focal Loss* untuk mendeteksi ekspresi wajah pada data ekspresi wajah?
2. Berapa tingkat akurasi yang dihasilkan oleh model MobileNetV2+CBAM+*Focal Loss* pada dataset FERPlus?

1.3 Batasan Permasalahan

Agar penelitian ini lebih terfokus dan terarah, maka ditetapkan batasan masalah sebagai berikut:

1. Dataset yang digunakan adalah FERPlus, yaitu versi re-anotasi dari FER2013 yang dikembangkan oleh Microsoft, tanpa melibatkan dataset lain di luar itu.
2. Klasifikasi yang dilakukan mencakup tujuh ekspresi wajah, yaitu *happy*, *sad*, *angry*, *fear*, *surprise*, *disgust*, dan *neutral*.

1.4 Tujuan Penelitian

Tujuan dari penelitian ini adalah:

1. Mengimplementasikan MobileNetV2 yang diintegrasikan dengan CBAM dan *Focal Loss* untuk mendeteksi ekspresi wajah pada data ekspresi wajah.
2. Mengetahui tingkat akurasi yang dihasilkan oleh model MobileNetV2+CBAM+*Focal Loss* pada dataset FERPlus.

1.5 Manfaat Penelitian

1. Memberikan kontribusi berupa *framework* yang mengintegrasikan *attention mechanism* (CBAM), *loss function* adaptif terhadap *class imbalance* (*Focal*

Loss), dan *explainability* melalui Grad-CAM dalam satu arsitektur kohesif untuk *Facial Expression Recognition*.

2. Memperkaya literatur ilmiah di bidang *Computer Vision*, khususnya dalam mengatasi tantangan fundamental dataset FER yang memiliki karakteristik *class imbalance* ekstrem dan kebutuhan interpretabilitas model.
3. Menghasilkan sistem FER yang dapat diaplikasikan dalam pemantauan keterlibatan dan kondisi emosional siswa pada pembelajaran daring untuk meningkatkan kualitas pendidikan jarak jauh.
4. Memberikan kontribusi pada deteksi dini kondisi psikologis dalam sistem kesehatan mental untuk mendukung diagnosis dan intervensi yang tepat waktu.

1.6 Sistematika Penulisan

Sistematika penulisan laporan penelitian ini adalah sebagai berikut:

- **Bab 1 PENDAHULUAN**

Bab ini berisi latar belakang masalah, rumusan masalah, batasan permasalahan, tujuan penelitian, manfaat penelitian, dan sistematika penulisan.

- **Bab 2 LANDASAN TEORI**

Bab ini membahas teori-teori yang mendasari penelitian, meliputi *Facial Expression Recognition*, MobileNetV2, CBAM, *Focal Loss*, Grad-CAM, dataset FERPlus, dan perbandingan penelitian terdahulu.

- **Bab 3 METODOLOGI PENELITIAN**

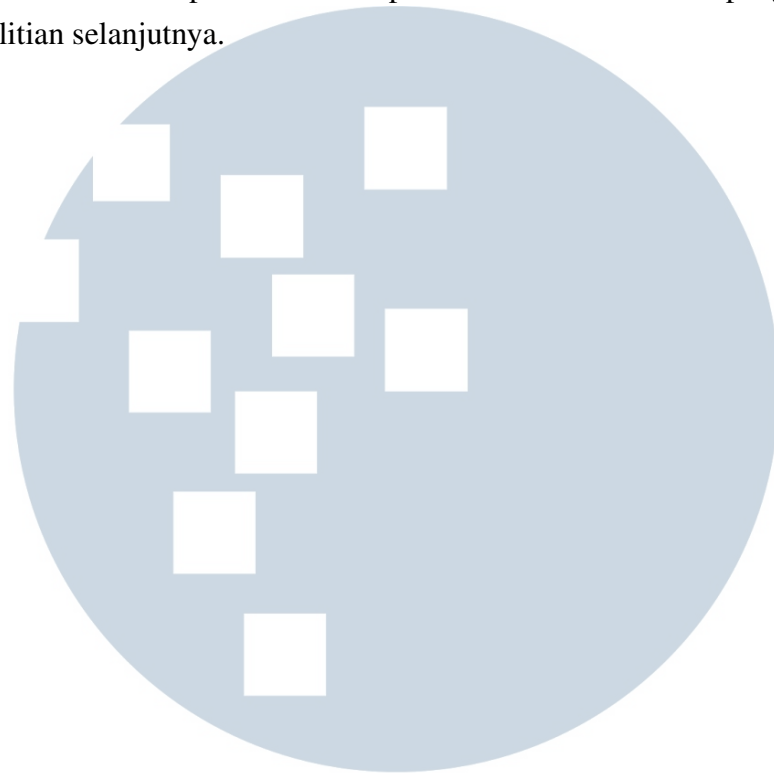
Bab ini menjelaskan metodologi penelitian yang digunakan, mencakup alur penelitian, pengumpulan data, analisis eksplorasi data, *preprocessing* dan augmentasi, arsitektur model, *Focal Loss*, konfigurasi pelatihan, metrik evaluasi, implementasi Grad-CAM, dan *hyperparameter tuning*.

- **Bab 4 HASIL DAN DISKUSI**

Bab ini menyajikan hasil eksperimen dan pembahasan, meliputi hasil pelatihan per skenario, evaluasi performa model, analisis *confusion matrix*, visualisasi Grad-CAM, dan perbandingan dengan penelitian terdahulu.

- **Bab 5 SIMPULAN DAN SARAN**

Bab ini berisi kesimpulan dari hasil penelitian dan saran untuk pengembangan penelitian selanjutnya.



UMN
UNIVERSITAS
MULTIMEDIA
NUSANTARA