

BAB I

PENDAHULUAN

1.1 Latar Belakang

Emosi merupakan respons tersinkronisasi pada subsistem organisme terhadap evaluasi suatu stimulus yang relevan bagi kepentingan individu (Scherer, 2022). Dalam komputasi afektif, representasi emosi telah berkembang dari label diskret menuju model dimensional *Valence-Arousal* (VA) yang memetakan kondisi afektif pada skala kontinu kesenangan dan aktivasi (Boateng dkk., 2022; Pinto dkk., 2019). Kemampuan VA dalam membedakan gradasi emosi secara halus menjadikannya sangat relevan untuk menganalisis ekspresi anak yang memiliki kompleksitas afektif tinggi dengan keterbatasan dalam komunikasi verbal.

Menggambar merupakan sarana penting bagi anak untuk mengekspresikan kondisi internal dan representasi mentalnya, dipandang sebagai jendela menuju dunia batin anak (Ferro dkk., 2025), di mana coretannya tidak sekadar salinan realistis dunia luar melainkan gambaran dari apa yang mereka rasakan, pikirkan, dan ketahui tentang realitas di sekitarnya (Quaglia dkk., 2015). Anak dapat menyampaikan emosi dasar melalui isyarat grafis dan kemampuan ini berkembang seiring usia (Cannoni dkk., 2023). Namun analisis gambar secara tradisional bersifat subjektif dan memakan waktu sehingga penerapan kecerdasan buatan untuk interpretasi otomatis menjadi alternatif yang menjanjikan meskipun menghadapi tantangan dalam mengekstraksi sinyal afektif dari elemen visual yang bersifat abstrak (Khlaif dkk., 2025).

Çiftçi dkk. (2025) memperkenalkan dataset KIDO yang mencakup 10.860 gambar coretan dari 5.430 siswa di Turki, di mana DenseNet121 mencapai F1 79,03% pada modalitas visual dan ALBERT mencapai akurasi 97,47% pada

modalitas teks. Namun pendekatan tersebut bersifat unimodal dimana gambar dan teks refleksi diproses secara independen sehingga interaksi lintas modalitas dan konteks emosional yang lebih utuh belum tereksplorasi.

Zhu dkk. (2022) mengembangkan *Image-Text Interaction Network* (ITIN) yang mengintegrasikan wilayah afektif citra dengan representasi kata melalui *cross-modal alignment* dan *adaptive gating*, mencapai akurasi 75,19% pada MVSA-Single. Namun ITIN masih terbatas pada sentimen diskret dan menggunakan mekanisme *cross-modal attention* satu tahap (*single-stage*) sehingga fitur yang masuk ke modul interaksi belum melalui penguatan representasi intramodal yang optimal. Pendekatan *two-stage attention* pada TAG-fusion (Zhang dkk., 2025) mengatasi hal ini dengan memisahkan tahap penguatan intramodal dan *cross-modal interaction* secara bertahap.

Pada sisi visual, ResNet efektif mengekstraksi fitur tekstur berkat *skip connections* (He dkk., 2016), namun terbatas dalam menangkap hubungan spasial global akibat bias induktif lokal (Dosovitskiy dkk., 2021), keterbatasan ini dapat diatasi melalui deteksi wilayah berbasis *bounding box* yang memfokuskan model pada elemen visual relevan (Anderson dkk., 2018). Pada sisi tekstual, BERT unggul dalam menangkap konteks dua arah (Devlin dkk., 2019) namun memiliki kompleksitas komputasi kuadratik terhadap panjang urutan (Zhao dkk., 2022). BiGRU mengatasi keterbatasan tersebut dengan merangkum dependensi kontekstual dua arah secara lebih efisien (Han dkk., 2022). Kombinasi keduanya dalam ITIN memungkinkan *cross-modal attention* dan *adaptive gating* mereduksi pasangan region–kata yang tidak selaras secara emosional (Bin Saeed & Cha, 2025;

Zhu dkk., 2022), meskipun efektivitasnya tetap bergantung pada kualitas deteksi wilayah yang dihasilkan (Anderson dkk., 2018).

Meskipun ITIN mampu mengintegrasikan modalitas secara adaptif, model ini masih terbatas pada label emosi diskret dan belum melakukan eksplorasi dimensi intensitas afektif kontinu. Keterbatasan ini dapat diatasi dengan mengintegrasikan fitur VA yang diekstraksi melalui model regresi berbasis RoBERTa (Mendes & Martins, 2023), sehingga sistem dapat membedakan emosi secara lebih presisi melalui representasi nilai numerik kontinu dalam ruang *Valence-Arousal* dibandingkan kategori diskret (Lee dkk., 2022).

Berdasarkan kesenjangan tersebut serta keterbatasan dataset KIDO yang hanya menyediakan dua label emosi diskret sehingga belum mampu merepresentasikan variasi intensitas perasaan anak secara lebih kaya, penelitian ini mengusulkan model multimodal untuk klasifikasi emosi coretan anak dengan mengintegrasikan fitur afektif kontinu VA ke dalam arsitektur ITIN yang dimodifikasi. Kebaruan penelitian ini terletak pada penggabungan tiga sumber informasi yaitu fitur visual gambar, fitur tekstual refleksi anak, dan nilai VA sebagai pengayaan afektif, yang diharapkan menghasilkan sistem deteksi emosi yang lebih akurat dan mampu menangkap intensitas perasaan anak secara lebih objektif untuk membantu praktisi pendidikan maupun kesehatan mental.

1.2 Rumusan Masalah

Berdasarkan latar belakang tersebut, rumusan masalah dalam penelitian ini adalah sebagai berikut:

1. Bagaimana mengembangkan arsitektur klasifikasi emosi coretan anak berbasis *Image-Text Interaction Network* dengan integrasi *Valence-Arousal*

dalam mengatasi subjektivitas interpretasi manual dan keterbatasan model emosi diskret?

2. Bagaimana pengaruh integrasi fitur afektif kontinu *Valence-Arousal* terhadap peningkatan akurasi dan pemahaman nuansa emosional dibandingkan dengan model multimodal standar?
3. Sejauh mana model yang diusulkan mampu mengintegrasikan interaksi antara elemen visual pada gambar coretan dan teks refleksi anak untuk menentukan label emosi yang tepat?

1.3 Tujuan Penelitian

Berdasarkan rumusan masalah yang telah diuraikan, tujuan dari penelitian ini adalah sebagai berikut:

1. Membangun model klasifikasi emosi multimodal yang mengintegrasikan fitur *Valence-Arousal* pada arsitektur ITIN untuk mengolah data gambar dan teks coretan anak.
2. Mengevaluasi performa model dalam mengklasifikasikan emosi menjadi kelas *Happiness* dan *Sadness* dengan tingkat akurasi yang lebih optimal melalui pengayaan fitur afektif kontinu.
3. Membuktikan efektivitas mekanisme *cross-modal interaction* dalam mengintegrasikan pesan emosional antara gambar dan tulisan refleksi anak pada dataset KIDO.

1.4 Manfaat Penelitian

Penelitian ini memberikan sejumlah manfaat dalam ranah keilmuan dan berikut merupakan manfaat secara teoritis yaitu:

1. Memberikan kontribusi teoritis pada ranah *affective computing* mengenai pentingnya pemodelan emosi dimensional dalam memperkuat sistem inferensi multimodal.
2. Menghasilkan sistem pendeteksi dini (*screening*) masalah kesehatan mental anak yang lebih objektif melalui analisis coretan visual.
3. Memberikan referensi teknis bagi pengembang kecerdasan buatan dalam merancang sistem yang mampu memahami intensitas afeksi pada lintas domain aplikasi.

Adapun manfaat secara praktis dari penelitian ini dijabarkan dalam beberapa poin sebagai berikut:

1. Menyediakan sebuah alat bantu skrining awal pendeteksi kondisi emosional yang objektif dan non-intrusif dalam aktivitas menggambar untuk anak-anak usia sekolah dasar (SD) hingga menengah pertama (SMP).
2. Memfasilitasi praktisi pendidikan di lingkungan keseharian anak seperti Guru Bimbingan Konseling (BK), Wali Kelas, dan Orang Tua dengan informasi afektif yang kaya sebagai jembatan untuk membuka komunikasi empatik dengan anak yang kesulitan berekspresi secara verbal.

1.5 Batasan Masalah

Untuk menjaga fokus dan kelayakan penelitian, ditetapkan batasan-batasan sebagai berikut:

1. Penelitian ini menggunakan dataset publik KIDO (Çiftçi dkk., 2025) dengan fokus pada dua kategori emosi utama saja yaitu *happiness* dan *sadness*.
2. Modalitas masukan terbatas pada gambar pindaian coretan tangan anak dan transkripsi teks refleksi dalam bahasa Inggris.

3. Fitur *Valence-Arousal* diekstrak menggunakan model eksternal yang telah dilatih pada korpus emosi standar yaitu EmoBank (Buechel & Hahn, 2017).
4. Evaluasi performa model dilakukan menggunakan metrik standar berupa akurasi, presisi, *recall*, dan skor F1 dengan pembagian rasio dataset 85:15 untuk *training* dan *testing* (Çiftçi dkk., 2025).
5. Populasi target dan sampel data yang mendasari hasil validasi hanya mempresentasikan individu pembuat coretan yang berada direntang populasi uji anak usia 6-14 tahun sesuai rekam jejak bawaan dataset (Çiftçi dkk., 2025).
6. Penelitian ini tidak mempertimbangkan spektrum warna sebagai fitur analisis emosi karena ekstraksi fitur visual dilakukan melalui pendekatan berbasis region (Zhu dkk., 2022).