

BAB 1

PENDAHULUAN

1.1 Latar Belakang Penelitian

Perkembangan teknologi kecerdasan buatan (*Artificial Intelligence*), khususnya pada bidang *Computer Vision*, telah memberikan kontribusi signifikan dalam analisis perilaku manusia melalui citra dan video [1]. Salah satu implementasi yang berkembang pesat adalah pengenalan ekspresi wajah (*facial expression recognition*) untuk mendeteksi kondisi emosional individu. Secara psikologis dan fisiologis, kondisi stres sering kali bermanifestasi melalui perubahan fisik, termasuk ketegangan otot wajah atau ekspresi negatif tertentu [2]. Hal ini menjadikan ekspresi wajah sebagai salah satu indikator visual yang dapat diukur untuk mengidentifikasi tingkat stres seseorang [3].

Tingkat stres merupakan indikator penting dalam kesehatan mental. Jika tidak terdeteksi secara dini, stres dapat berdampak pada penurunan produktivitas hingga gangguan kesehatan fisik dan psikologis. Deteksi tingkat stres secara konvensional umumnya dilakukan melalui kuesioner atau wawancara psikologis. Namun, pendekatan tersebut memiliki keterbatasan karena memakan waktu, bersifat subjektif, dan sangat bergantung pada kejujuran responden. Keterbatasan inilah yang mendorong perlunya sistem berbasis teknologi yang mampu melakukan identifikasi tingkat stres secara otomatis, objektif, dan non-invasif melalui analisis ekspresi wajah. Sistem semacam ini berimplikasi luas, mulai dari pemantauan kesehatan mental di lingkungan kerja, dunia pendidikan, hingga pengembangan kebijakan layanan kesehatan psikologis yang lebih responsif.

Pendekatan konvensional dalam klasifikasi emosi dan ekspresi wajah umumnya menggunakan metode *Convolutional Neural Network* (CNN) [4]. Meskipun CNN memiliki performa yang baik dalam mengekstraksi fitur lokal pada citra, model tersebut kurang mampu menangkap hubungan global antar bagian wajah secara menyeluruh [5]. Perkembangan terbaru dalam *deep learning* memperkenalkan arsitektur *Vision Transformer* (ViT) yang mengadaptasi mekanisme *self-attention* untuk memproses citra sebagai rangkaian *patch* [6]. Pendekatan ini memungkinkan model

untuk memahami keterkaitan antar fitur wajah secara lebih komprehensif dan efektif [7].

Beberapa penelitian terdahulu telah mengimplementasikan CNN maupun *hybrid* CNN-LSTM untuk pengenalan emosi wajah dengan tingkat akurasi yang bervariasi. Namun, pemanfaatan *Vision Transformer* khusus untuk klasifikasi tingkat stres berbasis ekspresi wajah masih relatif terbatas, terutama pada pemetaan ekspresi ke dalam beberapa level kategori stres [8]. Berdasarkan celah penelitian tersebut, penelitian ini mengusulkan penerapan metode *Vision Transformer* (ViT) sebagai pendekatan alternatif yang lebih komprehensif untuk mendeteksi kondisi psikologis seseorang secara otomatis dan non-invasif.

1.2 Rumusan Masalah Penelitian

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

1. Bagaimana merancang dan membangun sistem klasifikasi tingkat stres berbasis ekspresi wajah menggunakan metode *Vision Transformer* (ViT)?
2. Bagaimana performa model *Vision Transformer* dalam mengklasifikasikan tingkat stres berdasarkan parameter evaluasi seperti *accuracy*, *precision*, *recall*, dan *F1-score*?

1.3 Tujuan dan Manfaat Penelitian

1.3.1 Tujuan Penelitian

Tujuan dari penelitian ini adalah:

1. Merancang dan membangun sistem klasifikasi tingkat stres berbasis citra ekspresi wajah menggunakan arsitektur *Vision Transformer* (ViT).
2. Mengevaluasi dan menganalisis performa metode *Vision Transformer* dalam mendeteksi tingkat stres berdasarkan metrik pengujian (*accuracy*, *precision*, *recall*, dan *F1-score*).

1.3.2 Manfaat Penelitian

Manfaat dari penelitian ini adalah:

1. Memberikan kontribusi dalam pengembangan sistem deteksi stres berbasis *Computer Vision*.
2. Menjadi referensi akademik dalam penerapan *Vision Transformer* pada klasifikasi emosi dan tingkat stres.
3. Memberikan alternatif metode berbasis *deep learning* yang mampu menangkap hubungan global fitur wajah secara lebih efektif dibandingkan metode konvensional.
4. Mendukung pengembangan sistem pemantauan kondisi psikologis secara non-invasif dan otomatis.

1.4 Batasan Masalah Penelitian

Agar penelitian lebih terarah dan fokus, maka batasan masalah dalam penelitian ini adalah:

1. Data yang digunakan berupa citra statis ekspresi wajah
2. Klasifikasi tingkat stres dibatasi pada 3 kategori, yaitu stres rendah, stres sedang, dan stres tinggi.
3. Proses *preprocessing* citra dibatasi pada tahapan *resizing* ke ukuran standar arsitektur ViT (misalnya 224×224 piksel) dan normalisasi sebelum diproses oleh model.
4. Model yang digunakan murni berbasis arsitektur dasar *Vision Transformer* (ViT), tanpa menggunakan kombinasi dengan metode lain seperti CNN atau LSTM.
5. Evaluasi performa model dilakukan menggunakan *confusion matrix* dengan metrik *accuracy*, *precision*, *recall*, dan *F1-score*.
6. Penelitian ini murni merupakan pendekatan komputasional (*Computer Vision*) dan tidak ditujukan sebagai alat diagnosis medis klinis.

1.5 Keterbaruan Penelitian

Untuk menegaskan kebaruan (*novelty*) dan posisi penelitian ini dibandingkan dengan penelitian-penelitian sebelumnya, dilakukan tinjauan pustaka terhadap tiga artikel jurnal terdahulu yang memiliki relevansi tinggi. Penelitian pertama dilakukan oleh Tian dkk. (2024) [9] dengan judul "*Facial Expression Recognition Based on Vision Transformer with Hybrid Local Attention*". Penelitian tersebut memanfaatkan arsitektur *Vision Transformer* (ViT) dengan mekanisme perhatian hibrida (*hybrid attention*) untuk meningkatkan akurasi pengenalan tujuh ekspresi emosi wajah secara umum. Persamaan penelitian tersebut dengan penelitian ini adalah sama-sama memanfaatkan algoritma ViT. Namun, keterbaruan utama dari penelitian ini terletak pada luaran (*output*) klasifikasi sistemnya. Jika penelitian terdahulu murni berfokus pada pengenalan emosi dasar seperti marah, jijik, atau terkejut, penelitian ini melangkah lebih jauh dengan memetakan akumulasi dari kategori emosi dasar tersebut (menggunakan dataset FER2013) ke dalam kondisi psikologis spesifik, yaitu pengklasifikasian tingkat stres menjadi kategori rendah, sedang, dan tinggi.

Selanjutnya, penelitian oleh Harshinni dkk. (2023) [10] yang berjudul "*A Deep Learning Approach for Human Stress Detection using Haar-Cascade Algorithm*", mengusulkan sistem pendeteksian stres psikologis manusia berdasarkan analisis ekspresi wajah menggunakan dataset FER2013 dengan menggabungkan algoritma deteksi *Haar-Cascade* dan metode regresi logaritmik. Persamaan penelitian tersebut dengan penelitian ini terletak pada tujuan akhir sistem, yaitu mengevaluasi indikasi stres manusia menggunakan dataset yang sama. Adapun keterbaruan utama pada penelitian ini terletak pada pergantian metode komputasi inti yang digunakan. Penelitian terdahulu menggunakan algoritma *machine learning* konvensional (*Haar-Cascade*) yang berfokus pada deteksi fitur wajah secara kaku, sedangkan penelitian ini mengimplementasikan algoritma mutakhir *deep learning*, yaitu *Vision Transformer* (ViT). Mekanisme *global self-attention* pada ViT memungkinkan model untuk mengevaluasi korelasi seluruh bagian wajah secara bersamaan, sehingga menawarkan pendekatan ekstraksi fitur yang jauh lebih komprehensif dibandingkan metode sebelumnya.

Terakhir, penelitian oleh Fadhlurrahman dkk. (2025) [11] dengan judul "*Pengenalan Ekspresi Wajah Peserta Didik di Ruang Kelas Menggunakan Vision*

Transformer (ViT)", menerapkan sistem pengenalan ekspresi wajah secara *real-time* menggunakan ViT untuk memahami kondisi emosional peserta didik saat proses belajar mengajar di dalam kelas. Persamaan dengan penelitian ini adalah penerapan metode ViT untuk menganalisis wajah manusia guna menginterpretasikan kondisi emosional di Indonesia. Keterbaruan utama dari penelitian ini dibandingkan penelitian tersebut adalah ruang lingkup pemrosesan dan tujuan spesifik diagnosis. Penelitian terdahulu merancang sistem tangkapan *real-time* yang difokuskan pada emosi umum untuk evaluasi metode pengajaran guru, sementara penelitian ini difokuskan pada pemrosesan komputasional menggunakan citra statis terstandarisasi. Tujuannya adalah membangun model evaluasi deteksi medis non-invasif yang secara khusus mengukur tingkat stres, yang tentunya memiliki landasan parameter berbeda dari sekadar respons emosional umum di ruang kelas.