

BAB 1 PENDAHULUAN

1.1. Latar Belakang

Teknologi pendidikan dan pembelajaran *online* mengalami perkembangan pesat, hal ini ditandai dengan kemunculan berbagai pembelajaran *online* dan peralihan banyak kelas dari universitas dan lembaga pendidikan ke ranah *online* [1]. Salah satu masalah dalam pembelajaran *online* adalah engagement atau keterlibatan peserta dalam pembelajaran *online* [2]. Keterlibatan setiap peserta dalam pembelajaran *online* dapat dinilai dari status attentive dan inattentive-nya. Ketika peserta tidak terlibat selama pembelajaran *online*, hal tersebut dapat mengurangi hasil dari upaya pengajar dan bantuan instruksional mereka, sehingga menghasilkan pengalaman belajar yang kurang efektif [3]. Dalam hal ini, teknologi AI dan visi komputer dapat dikembangkan melalui pendekatan pengenalan ekspresi wajah atau *Facial Expression Recognition* (FER) berdasarkan ekspresi wajah yang tertangkap oleh kamera untuk dapat mengenali tingkat keterlibatan peserta dalam pembelajaran *online*.

Penggunaan teknologi FER dalam pembelajaran *online* dapat membantu untuk memantau keterlibatan peserta sehingga dapat menghasilkan pengalaman belajar yang efektif. Meskipun terdapat beberapa penelitian tentang penilaian formatif berbasis seluler terhadap keterlibatan peserta, pada pembelajaran *online* dengan jumlah peserta yang relatif tinggi, video wajah dari setiap peserta tidak dapat membantu pengajar dalam menilai keterlibatan peserta, sehingga deteksi otomatis keterlibatan peserta merupakan solusi yang tepat [1].

Dalam penerapan FER pada pembelajaran *online* terdapat tantangan pada data yang bersifat spasial dan temporal, misalnya data video yang mencakup informasi spasial dari fitur wajah dan informasi temporal dari

perubahan ekspresi sepanjang waktu. Meskipun model *Convolutional Neural Network* (CNN) telah menunjukkan kinerja tinggi dalam klasifikasi video dengan mengekstrak fitur spasial dari setiap frame RGB secara independen dan memberikan akurasi pengenalan yang baik, pendekatan ini tidak memanfaatkan informasi temporal dalam representasi data [4].

Beberapa penelitian sebelumnya telah menunjukkan hasil yang signifikan dalam pengenalan ekspresi wajah menggunakan berbagai model deep learning. Sebagai contoh, penelitian menggunakan ResNet-50 dan CBAM pada dataset RAF-DB dan FER-2013 berhasil mencapai akurasi sebesar 87.62% dan 88.13% [5]. Pendekatan lain seperti Ensemble Learning yang menggabungkan VGG19 dan ResNet-50 mampu meningkatkan akurasi hingga 93.11% pada dataset CK+ [3]. Namun, model-model ini terutama memanfaatkan fitur spasial dan belum secara optimal memanfaatkan informasi temporal yang terdapat dalam video. Selain itu, penelitian yang menggunakan model CNN-RF hanya mencapai akurasi 71% pada dataset FER-2013 [6], sementara model temporal seperti Softmax Stokastik Temporal dan TFEN masing-masing menunjukkan akurasi sebesar 80% [4] dan 81.73% [7]. Meskipun ada peningkatan performansi pada model-model tertentu, keterbatasan dalam menangkap informasi spasial-temporal secara efektif masih menjadi tantangan utama.

Model 3D CNN diusulkan untuk menangani tantangan pada data yang bersifat spasial dan temporal. Model ini dapat memproses data spasial dan temporal secara bersamaan dengan lebih efektif dalam menangkap perubahan ekspresi wajah dari waktu ke waktu, sehingga lebih cocok untuk mengenali ekspresi wajah dalam video yang mengandung informasi spasial dan temporal [7].

Dengan mempertimbangkan kebutuhan akan analisis temporal yang efektif dalam mengenali ekspresi wajah peserta dalam video pembelajaran *online*, tujuan penelitian ini adalah mengevaluasi efektivitas model 3D CNN untuk mengatasi masalah ini.

1.2. Rumusan Masalah

Tantangan utama dalam pembelajaran *online* adalah menilai keterlibatan peserta dengan efektif. Meskipun FER telah digunakan untuk menangani masalah ini, masih terdapat tantangan pada data yang bersifat spasial dan temporal. Pertanyaan yang muncul dalam penelitian ini adalah:

1. Bagaimana efektivitas 3D CNN pada data spasial dan temporal dibandingkan dengan CNN?
2. Bagaimana mengoptimalkan efektivitas 3D CNN pada data spasial dan temporal?

1.3. Tujuan dan Manfaat

Tujuan utama dari penelitian ini adalah untuk mengevaluasi efektivitas 3D CNN pada data spasial data temporal yang dihasilkan dari video pembelajaran *online* dibandingkan dengan model CNN. Penelitian ini juga bertujuan untuk mengoptimalkan efektivitas 3D CNN agar mampu menangkap informasi spasial dan temporal dengan lebih efektif. Dengan menggunakan metrik evaluasi seperti *accuracy*, *precision*, *recall*, dan *F1-score*, penelitian ini akan mengukur seberapa baik model dapat mendeteksi keterlibatan peserta berdasarkan ekspresi wajah.

Penelitian ini diharapkan dapat memberikan dampak yang positif bagi pengajar dan peserta didik. Hasil dari penelitian ini dapat memberikan solusi bagi pengajar dan lembaga pendidikan untuk memantau keterlibatan peserta secara lebih efektif untuk meningkatkan kualitas pembelajaran *online*. Peserta didik juga diharapkan dapat meningkatkan hasil pembelajaran *online* dengan pengalaman belajar yang lebih terarah dan personal. Penelitian ini diharapkan mampu menjadi langkah awal yang signifikan dalam penerapan teknologi AI dan visi komputer untuk mendukung pendidikan di era digital.

1.4. Batasan Masalah

Penelitian ini memiliki beberapa batasan masalah yang perlu diperhatikan agar penelitian dapat terfokus dan terarah. Berikut adalah batasan-batasan masalah yang diidentifikasi dalam penelitian ini:

1. Ekspresi Wajah

Penelitian ini terbatas pada pengenalan ekspresi wajah dasar yang terdiri dari *anger, sad, fear, neutral, happy, dan surprised*. Ekspresi wajah yang lebih kompleks atau campuran dari beberapa ekspresi wajah tidak akan menjadi fokus pada penelitian ini.

2. Pembelajaran *Online*

Penelitian ini dibatasi pada konteks pembelajaran online. Penelitian tidak akan mencakup analisis keterlibatan dalam konteks tatap muka atau *blended learning*. Selain itu, penelitian ini berfokus pada analisis ekspresi wajah terhadap keterlibatan peserta. Faktor-faktor lain yang dapat mempengaruhi keterlibatan peserta seperti interaksi peserta dan pengajar, nilai peserta, dan keaktifan peserta tidak dipertimbangkan dalam penelitian ini.

3. Klasifikasi Keterlibatan

Keterlibatan dinilai hanya dari status *attentive* dan *inattentive* berdasarkan ekspresi wajah peserta pembelajaran *online*. Ekspresi wajah *anger, sad, dan fear* dikategorikan sebagai *inattentive*, sedangkan ekspresi wajah *neutral, happy, dan surprised* dikategorikan sebagai *attentive* [3].

4. *Dataset*

Penelitian ini menggunakan *dataset* yang sebagian besar memiliki pencahayaan baik. Sehingga model tidak dapat memberikan hasil maksimal saat digunakan pada data dengan pencahayaan buruk.

1.5. Metode Penelitian

Penelitian ini dilakukan melalui beberapa tahapan utama, dimulai dengan studi literatur untuk memahami konsep dasar dan teknologi terkait, dilanjutkan dengan pengumpulan data berupa *dataset* video ekspresi wajah

yang relevan. Tahap berikutnya adalah pemrosesan data, seperti ekstraksi frame dari video dan normalisasi data. Setelah itu, dilakukan implementasi dan pengoptimalan model menggunakan arsitektur 3D CNN untuk menangkap informasi spasial-temporal. Model yang dikembangkan kemudian melalui proses pengujian untuk mengevaluasi performanya menggunakan metrik seperti akurasi, precision, recall, dan F1-score. Hasil dari pengujian ini dianalisis pada tahap evaluasi untuk menilai efektivitas model dan mencapai tujuan penelitian.