

## ABSTRAK

Pembelajaran *online* telah berkembang pesat dalam beberapa tahun terakhir seiring kebutuhan di masa kini serta adanya disrupsi teknologi terhadap metode pengajaran. Dalam pelaksanaannya, engagement atau keterlibatan peserta menjadi salah satu faktor penting untuk menciptakan proses pembelajaran *online* yang efektif. Teknologi pengenalan ekspresi wajah (FER) telah digunakan untuk menangani masalah ini, namun pendekatan berbasis model *Convolutional Neural Network* (CNN) memiliki keterbatasan karena tidak dapat memanfaatkan informasi temporal dalam video. Oleh karena itu, penelitian ini mengusulkan penggunaan model 3D CNN pada data spasial-temporal yang lebih kompleks dan meningkatkan akurasi deteksi keterlibatan peserta berdasarkan ekspresi wajah. Penelitian ini bertujuan untuk mengevaluasi kemampuan model 3D CNN dibandingkan dengan model CNN dan mengoptimalkan parameter seperti epoch, batch size, dan learning rate untuk meningkatkan efisiensi pelatihan. Model dilatih menggunakan dataset video ekspresi wajah dengan enam kategori (*anger, sad, fear, neutral, happy, dan surprised*). Hasil penelitian menunjukkan bahwa bahwa model 3D CNN dapat menangkap informasi spasial-temporal secara lebih efektif dibandingkan model CNN. Model terbaik menggunakan arsitektur 3D Inception-ResNet + LSTM dengan konfigurasi parameter optimal mencapai akurasi 99.07% dengan waktu pelatihan yang lebih efisien.

**Kata Kunci:** facial expression recognition, spasial-temporal, 3D CNN, Inception-ResNet, 3D Inception-ResNet, LSTM.