BAB V
KESIMPULAN DAN SARAN

5.1 Kesimpulan

Berdasarkan hasil dari penelitian yang dilakukan mengenai klasifikasi aksi peserta didik di dalam ruang kelas menggunakan pendekatan model Vision Transformer berbasis video, maka didapatkan kesimpulan sebagai berikut:

1. Pembentukan dataset video pengenalan dan klasifikasi aksi peserta didik berhasil dilakukan. Data diambil pada ruang kelas nyata, terdiri dari total 403 klip video dengan lima kelas aksi, yaitu mengangguk, mengangkat tangan, menggunakan ponsel, menopang kepala di atas meja, dan menunduk yang dilakukan oleh 53 peserta didik/aktor.
2. Implementasi VideoMAE berhasil dilakukan untuk tugas pengenalan dan klasifikasi aksi peserta didik di dalam ruang kelas dengan penerapan fine-tuning model pre-trained VideoMAE. Dilakukan eksperimen dengan kombinasi parameter epoch, batch size, dan learning rate. Penambahan parameter *weight decay* dan learning rate scheduler terbukti dapat mencegah terjadinya *overfitting*. Penambahan jumlah epoch juga dapat menaikan akurasi karena memberikan waktu lebih kepada model untuk mengeksplorasi data lebih luas.
3. Performa VideoMAE untuk tugas pengenalan dan klasifikasi aksi peserta didik di dalam ruang mencapai nilai akurasi sebesar 87.23%, F1-score 87.41%, recall 87.28%, dan precision 88.33% dengan konfigurasi parameter 4 epoch, 4 batch size, 5e-5 learning rate, 0.01 *weight decay*, dan linear learning rate scheduler. Hasil ini menunjukkan bahwa VideoMAE mampu mengenali aksi peserta didik dengan cukup baik, terutama dalam konteks pembelajaran yang dilakukan di ruang kelas.
4. Performa VideoMAE unggul dalam seluruh matriks evaluasi dengan selisih 6.37%, recall sebesar 6.37%, precision sebesar 6.81% dan F1-score sebesar 6.42% dibandingkan dengan TimeSformer. Performa TimeSformer untuk tugas pengenalan dan klasifikasi aksi peserta didik di dalam ruang mencapai nilai akurasi sebesar 79.80%, F1-score 79.80%, recall 79.80%, dan precision

81.03% dengan konfigurasi parameter 8 epoch, 4 batch size, 5e-5 learning rate, 0.01 *weight decay*, dan linear learning rate scheduler.

5.2 Saran

Berdasarkan penelitian yang telah dilakukan terdapat keterbatasan dan kekurangan, berikut beberapa saran yang dapat digunakan sebagai pertimbangan pada penelitian selanjutnya:

1. Dataset yang digunakan pada penelitian ini berukuran kecil dengan total 403 video. Untuk mendapatkan performa model yang lebih baik, dapat dilakukan perluasan baik dari jumlah maupun variasi video dataset.
2. Untuk implementasi VideoMAE pada dataset kecil disarankan untuk melakukan eksplorasi variasi hyperparameter. Dalam penelitian ini percobaan variasi parameter meliputi epoch, learning rate, dan batch size beserta penambahan .
3. Menerapkan strategi fine-tune lebih luas, seperti menggunakan scheduler agar model lebih stabil, menerapkan layer freezing atau unfreeze (fine-tune penuh) dan regularisasi untuk mengurangi overfitting selama proses pelatihan.