BAB 5
SIMPULAN DAN SARAN

5.1 Simpulan

Penelitian ini bertujuan untuk mengimplementasikan dan mengevaluasi enam model deteksi video *deepfake* berbasis arsitektur InceptionV3 dan LSTM yang dilatih menggunakan dataset FaceForensics++ dengan lima metode manipulasi *deepfake* berbeda: Deepfakes, Face2Face, FaceSwap, NeuralTextures, dan FaceShifter. Evaluasi dilakukan menggunakan skenario *cross-subset testing* serta lima metrik performa, yaitu akurasi, presisi, *recall*, F1-score, dan AUC-ROC.

Seluruh model menggunakan arsitektur yang sama, di mana InceptionV3 (*frozen*) tanpa *top classification layer* dan GlobalAveragePooling2D mengekstrak fitur spasial dari setiap *frame*, yang kemudian dinormalisasi dan diproses secara temporal oleh LSTM 128 unit, lalu diklasifikasikan oleh lapisan *Dense* 64-neuron dan *Dense* 1-neuron. Berdasarkan hasil penelitian, kesimpulan yang dapat diambil adalah sebagai berikut.

1. Model deteksi video *deepfake* berhasil diimplementasikan dengan memanfaatkan kombinasi ekstraksi fitur spasial dari InceptionV3 dan pemodelan hubungan temporal antar *frame* menggunakan LSTM. Proses pelatihan dilakukan pada video yang telah diproses menjadi urutan gambar wajah.

2. Model yang dilatih secara spesifik pada satu metode manipulasi video *deepfake*, terutama Deepfakes dan FaceSwap, menunjukkan kinerja terbaik pada data uji dari metode yang sama. Model Deepfakes, misalnya, mencapai AUC-ROC sebesar 0.9007 serta F1-score sebesar 82.8%.

3. Kemampuan generalisasi model terhadap metode manipulasi video *deepfake* yang berbeda tergolong rendah. Hal ini terlihat dari penurunan nilai AUC-ROC yang signifikan saat model diuji pada *subset* yang berbeda dari data pelatihan, dengan nilai rata-rata di luar diagonal hanya sekitar 0.5–0.65.

4. Model gabungan yang dilatih menggunakan seluruh subset metode manipulasi video *deepfake* menunjukkan kinerja yang lebih seimbang di seluruh metode, namun tidak unggul pada metode manapun secara spesifik.

AUC-ROC tertinggi model gabungan adalah 0.735 pada subset Deepfakes, dan rata-rata AUC keseluruhan sebesar 0.6620.

5. Fenomena *overfitting* ditemukan pada hampir semua model, terutama setelah epoch ke-8, meskipun telah diterapkan strategi mitigasi seperti dropout, *batch normalization*, regularisasi L2, dan augmentasi. Hal ini kemungkinan disebabkan oleh jumlah data yang terbatas untuk tiap metode, arsitektur model yang terlalu kompleks, serta jumlah frame (15) yang tidak cukup untuk menangkap pola temporal yang lebih kompleks.

Dengan demikian, penelitian ini menunjukkan bahwa pendekatan InceptionV3 dan LSTM dapat digunakan untuk membangun sistem deteksi video *deepfake*, namun performa dan kemampuan generalisasinya masih terbatas, terutama jika jumlah frame rendah dan variasi data tiap metode tidak mencukupi. Model yang dilatih dengan gabungan metode manipulasi video *deepfake* memiliki potensi dalam meningkatkan generalisasi, namun diperlukan strategi pelatihan dan pemilihan data yang lebih efektif untuk mengimbangi beragamnya karakteristik metode manipulasi pada video *deepfake*.

5.2 Saran

Berdasarkan hasil dan keterbatasan penelitian ini, maka saran yang dapat diberikan untuk penelitian selanjutnya adalah sebagai berikut.

1. Meningkatkan jumlah dan variasi data pelatihan, khususnya dengan memperbanyak metode manipulasi video *deepfake*, identitas, ekspresi wajah, dan kondisi pencahayaan untuk meningkatkan kemampuan generalisasi model terhadap data yang tidak dikenal.

2. Menggunakan jumlah *frame* yang lebih banyak sebagai *input*, agar model LSTM dapat menangkap pola temporal dan inkonsistensi gerakan wajah secara lebih efektif.

3. Mengeksplorasi arsitektur *pre-trained* lain seperti Xception, EfficientNet, atau kombinasi CNN dengan Transformer.

4. Mengeksplorasi kembali penggunaan InceptionV3 dan LSTM dengan penyesuaian kompleksitas arsitektur untuk mengurangi *overfitting* dan meningkatkan performa.

5. Melakukan evaluasi pada skenario dunia nyata, seperti video dari media sosial atau *dataset* dengan kompresi tinggi lainnya, untuk menguji ketahanan model terhadap *noise* dan variasi alami.