

BAB 5

SIMPULAN DAN SARAN

Bagian ini menyajikan kesimpulan dari keseluruhan proses penelitian dan implementasi model, serta memberikan saran untuk pengembangan di masa depan berdasarkan temuan dan tantangan yang dihadapi.

5.1 Simpulan

Penelitian ini diinisiasi untuk menjawab tantangan keamanan siber yang ditimbulkan oleh audio *deepfake* dan juga berhasil memenuhi poin-poin dari 1.4 Tujuan Penelitian, yaitu:

1. Berhasil konversi audio menjadi Mel-Spektrogram
2. Berhasil membangun model DCGANs untuk mendeteksi audio *deepfake*
3. Berhasil mengevaluasi kinerja model DCGANs dalam membedakan audio *real* dan *fake*

Metodologi yang digunakan mencakup beberapa tahap kunci, mulai dari pra-pemrosesan data audio mentah menjadi representasi Mel-spektrogram, hingga proses pelatihan model yang terkontrol dengan berbagai teknik regularisasi seperti Dropout dan L2 untuk menekan *overfitting*.

Berdasarkan hasil pengujian pada data uji, model yang dikembangkan berhasil mencapai akurasi akhir sebesar 87.8%, dengan nilai AUC 0.9386. Hasil ini menunjukkan bahwa model telah berhasil mempelajari pola-pola dasar untuk membedakan audio asli dan palsu, serta memiliki kemampuan deteksi yang jauh lebih baik daripada tebakan acak. Proses pelatihan tercatat berhenti pada epoch ke-19, dan model dari epoch ke-4 digunakan sebagai model final, yang menunjukkan bahwa mekanisme *EarlyStopping* berhasil mencegah *overfitting* yang lebih parah.

Analisis metrik yang lebih dalam justru menunjukkan performa yang kuat dan seimbang pada model akhir. Model ini tidak hanya menunjukkan Recall yang sangat tinggi untuk kelas *fake* (92%), tetapi juga mempertahankan Recall yang kuat untuk kelas *real* (82%). Hal ini mengindikasikan bahwa model mampu menjadi sensitif dalam mendeteksi ancaman tanpa menjadi terlalu "curiga" atau menghasilkan banyak kesalahan *false positive*.

Keseimbangan performa ini diperkuat oleh nilai Equal Error Rate (EER) yang rendah, yaitu 13.29%, yang mengonfirmasi bahwa model telah mencapai titik ekuilibrium yang sangat baik antara kesalahan pada audio asli dan palsu. Kesimpulannya, penelitian ini berhasil membuktikan bahwa arsitektur berbasis DCGAN merupakan pendekatan yang efektif dan andal untuk deteksi audio *deepfake*, dengan kemampuan untuk mencapai keseimbangan performa yang optimal.

5.2 Saran

Berdasarkan proses penelitian, implementasi, dan hasil analisis, terdapat beberapa saran yang dapat menjadi fokus untuk pengembangan penelitian ini di masa depan:

1. Fokus pada Pengurangan Bias dan *False Positive*: Mengingat model saat ini memiliki kecenderungan untuk salah mengklasifikasikan audio asli sebagai palsu (Recall rendah untuk kelas *real*), penelitian selanjutnya disarankan untuk fokus pada teknik yang dapat menyeimbangkan performa. Ini bisa mencakup:
 - (a) Penyesuaian Bobot Kelas (*Class Weighting*): Memberikan bobot kerugian (*loss weight*) yang lebih tinggi pada kelas *real* selama pelatihan untuk "mengajarkan" model agar tidak terlalu mudah memberikan prediksi *fake*.
 - (b) Tuning Ambang Batas Klasifikasi (*Threshold*): Melakukan analisis untuk menemukan ambang batas probabilitas yang optimal (tidak harus 0.5) yang dapat menyeimbangkan antara *precision* dan *recall*.
2. Eksperimen dengan Arsitektur dan Regularisasi: Disarankan untuk mengeksplorasi arsitektur CNN yang berbeda (misalnya, ResNet atau EfficientNet) yang mungkin lebih tahan terhadap *overfitting*. Selain itu, melakukan *tuning* yang lebih ekstensif pada hyperparameter regularisasi seperti laju Dropout dan nilai *weight decay* pada optimizer AdamW dapat membantu meningkatkan stabilitas.
3. Gunakan lebih dari satu dataset: Untuk menguji dan memvalidasi kemampuan generalisasi model secara lebih mendalam. Sebuah model yang dilatih hanya pada satu dataset berisiko mempelajari "jalan pintas" dengan

menghafal karakteristik unik dari dataset tersebut, seperti jenis artefak dari algoritma *deepfake* tertentu yang digunakan, kondisi rekaman (misalnya, jenis mikrofon atau tingkat kebisingan), atau aksen pembicara yang dominan. Dengan menguji model pada dataset lain yang memiliki karakteristik berbeda (proses yang dikenal sebagai cross-dataset *evaluation*), kita dapat memastikan bahwa model benar-benar belajar untuk mengidentifikasi pola-pola fundamental dari audio palsu, bukan hanya keunikan dari data latihnya. Penggunaan lebih dari satu dataset juga memungkinkan dilakukannya *benchmarking* atau perbandingan performa yang adil dengan penelitian-penelitian lain di bidang yang sama, sehingga meningkatkan kredibilitas dan validitas hasil penelitian.

4. Pengujian pada Dataset Eksternal: Untuk memvalidasi kemampuan generalisasi model secara lebih komprehensif, disarankan untuk menguji model yang telah dilatih pada dataset audio *deepfake* lain yang tidak digunakan selama pelatihan. Ini akan memberikan gambaran yang lebih baik tentang seberapa baik model beradaptasi pada data dari distribusi yang berbeda.
5. Gunakan Aspek Generator dari DCGANs untuk Data *Fake*: Saran ini bertujuan untuk meningkatkan jumlah dan variasi data audio palsu (*fake*) yang digunakan untuk melatih model diskriminator. Keterbatasan jumlah data pada dataset yang ada dapat diatasi dengan menggunakan *Generator* dari arsitektur DCGANs untuk menghasilkan sampel-sampel audio palsu baru yang realistis. Dengan melatih model detektor (*Diskriminator*) secara *adversarial* melawan *Generator* yang terus-menerus mencoba untuk "menipu"-nya, *Diskriminator* akan menjadi jauh lebih tangguh. Proses ini memaksa detektor untuk tidak hanya mengenali contoh-contoh yang ada, tetapi juga untuk mempelajari pola dasar atau artefak fundamental yang membedakan suara sintetis dari suara asli. Hasilnya adalah model deteksi yang memiliki kemampuan generalisasi yang lebih baik, karena ia telah dilatih untuk menghadapi berbagai macam variasi audio palsu yang mungkin belum pernah ada di dataset publik.