

BAB V

KESIMPULAN DAN SARAN

Bab ini menyajikan kesimpulan dari keseluruhan proses penilitian yang telah dilakukan, mulai dari perancangan hingga pengujian model untuk mengklasifikasi emosi terhadap lagu. Kesimpulan diambil mencerminkan pencapaian tujuan penilitian serta efektivitas metode yang digunakan dalam menghasilkan emosi yang dapat merepresentasikan keseluruhan suatu lagu. Selain itu, bab ini juga memuat saran-saran yang ditujukan untuk pengembangan penelitian di masa mendatang agar menjadi lebih optimal dan aplikatif dalam kehidupan nyata.

5.1 Kesimpulan

Berdasarkan penelitian yang telah dilakukan mengenai implementasi CNN terhadap ekstraksi fitur audio didapatkan kesimpulan sebagai berikut,

1. Proses ekstraksi MFCC, *mel-spectrogram* serta beberapa fitur audio dan juga spektral tambahan diimplementasikan melalui beberapa tahap seperti *pre-emphasis* yang dimana membagi sinyal audio utuh menjadi frame kecil untuk mengimbangi kehilangan energi, windowed-framing untuk memperkuat akurasi dengan membagi-bagi menjadi beberapa sekmenn, *Fourier Transform* yang bertugas untuk mengubah sinyal audio menjadi sinyal frekuensi, *mel-filter bank* untuk mengubah skala frekuensi menjadi skala mel dalam pendekatan dengan cara manusia mendengar, *logarithmic compression* untuk mengubah menjadi skala logaritmik untuk mengurangi efek perbedaan besar amplitudo antar *frame*, serta *Discrete Cosine Transform* yang menghasilkan koefisien yang mewakili energi dari musik tersebut.
2. CNN diterapkan dengan memperlakukan fitur MFCC dan *mel-spectrogram* sebagai representasi 2D yang kemudian diproses melalui arsitektur berlapis untuk ekstraksi fitur hierarkis. Arsitektur terdiri dari empat blok konvolusi dengan iterasi penambahan *channel*, masing-masing dilengkapi *Batch Normalization* untuk stabilitas, ReLU activation untuk *non-linearity*, *Max Pooling* untuk reduksi dimensi, dan *Dropout* untuk regularisasi mencegah *overfitting*. Setelah ekstraksi fitur konvolusi, *Global Average Pooling* digunakan untuk agregasi spasial, diikuti *fully connected layers* untuk prediksi emosi. Model menggunakan pendekatan *regression-based classification* yang memprediksi nilai *valence* dan *arousal* yang kontinu, kemudian dikonversi ke kategori

emosi berdasarkan *threshold*. *Training* menggunakan *optimizer AdamW*, *MSE loss function*, *learning rate scheduler* adaptif, dan *early stopping* untuk mencegah *overfitting*. Pendekatan ini terbukti efektif dengan akurasi 63-66% untuk klasifikasi 4 kuadran dan meningkat signifikan menjadi 77% untuk klasifikasi 2 kuadran, menunjukkan bahwa CNN mampu menangkap pola emosi musik dengan baik ketika kompleksitas ruang emosi disederhanakan.

3. Penelitian mencapai akurasi bervariasi tergantung konfigurasi ekstraksi fitur dan kompleksitas ruang emosi. Untuk klasifikasi 4 kuadran, model dengan MFCC 13 koefisien mencapai akurasi 63%, meningkat menjadi 65% dengan MFCC 24 koefisien dan 66% dengan MFCC 30 koefisien. Model *mel-spectrogram* menunjukkan performa yang sebanding dengan akurasi 64 sampai 65% untuk berbagai resolusi bins (128, 256, 512), sementara model dengan augmentasi mencapai akurasi 65% dengan *F1-score* tertinggi 0.60. Namun, terdapat perbedaan signifikan antar kelas dengan *recall* 80 sampai 86% pada kelas mayoritas tetapi hanya 5 sampai 15% pada kelas minoritas, menghasilkan akurasi seimbang 47 sampai 55% dan *macro F1-score* 42-52%. Penyederhanaan dari klasifikasi 4 kuadran menjadi 2 kuadran (berdasarkan dimensi *valence*) menghasilkan peningkatan dramatis dengan akurasi 77%, *precision* 0.78, *recall* 0.77, dan *F1-score* 0.77, menunjukkan performa yang jauh lebih seimbang dengan ketidakseimbangan kelas yang lebih ringan (ratio 1.6:1 dibandingkan 4.4:1). Hasil ini mengkonfirmasi bahwa keterbatasan performa pada klasifikasi 4 kuadran disebabkan oleh kombinasi ketidakseimbangan kelas ekstrim dan kompleksitas ruang emosi yang terlalu detail, serta limitasi dasar MFCC dalam menangkap kompleksitas emosi musik seperti harmoni dan melodi, dengan performa pendekatan MFCC untuk 4 kuadran berada pada rentang 63 sampai 66%, namun meningkat signifikan menjadi 77% ketika disederhanakan menjadi 2 kuadran.

5.2 Saran

Adapun beberapa saran yang dapat dikembangkan untuk penelitian selanjutnya yaitu sebagai berikut :

1. Penanganan Ketidakseimbangan Kelas

Ketidakseimbangan kelas yang ekstrim (ratio 4.4:1) menjadi salah satu faktor utama rendahnya performa pada kelas minoritas, penelitian selanjutnya disarankan un-

tuk menerapkan teknik penanganan ketidakseimbangan kelas seperti SMOTE (Synthetic Minority Over-sampling Technique), *class weighting* dalam *loss function*, atau *focal loss* yang memberikan bobot lebih tinggi pada kelas minoritas. Selain itu, dapat dipertimbangkan penggunaan *ensemble methods* atau *cost-sensitive learning* untuk meningkatkan *recall* pada kelas minoritas yang saat ini hanya mencapai 5 sampai 15%.

2. Ekspansi Dataset dan Validasi Cross-Domain

Untuk meningkatkan generalisasi model, disarankan melakukan augmentasi dataset dengan teknik yang lebih rumit seperti *SpecAugment*, *mixup*, atau menggunakan dataset emosi musik tambahan dari sumber lain.