V. KESIMPULAN DAN SARAN

5.1. Kesimpulan

Penelitian ini menunjukkan bahwa pendekatan *sequential fine-tuning* dua fase—dengan tahap pertama pada empat dataset bahasa Inggris (CREMA-D, RAVDESS, SAVEE, dan TESS), diikuti oleh tahap kedua menggunakan data kurasi dari podcast Indonesia—dapat secara efektif dimanfaatkan untuk mengenali emosi dalam audio berbahasa Indonesia menggunakan model Wav2Vec 2.0. Hasil penelitian menunjukkan bahwa pendekatan fine-tuning dua fase secara signifikan mengungguli model dasar (*baseline*) yang hanya dilatih menggunakan data berbahasa Indonesia, dengan capaian kinerja mendekati 91% dari performa manusia (*Human-Level Performance*). Model ini terbukti cukup efektif dalam mengenali emosi dasar, terutama pada audio berdurasi sedang.

Meskipun demikian, sejumlah keterbatasan teridentifikasi. Performa model menurun pada input dengan durasi sangat pendek maupun sangat panjang, serta menunjukkan akurasi yang rendah dalam mendeteksi emosi minoritas seperti *fear* (takut) dan *disgust* (jijik). Eksperimen dengan penambahan dataset TESS juga menegaskan bahwa kompatibilitas dan distribusi data memiliki pengaruh signifikan terhadap kemampuan generalisasi model. Kode yang digunakan untuk melatih dan melakukan evaluasi pada model ini tersedia di https://github.com/alianurrahman/SER

5.2. Rekomendasi

Untuk pengembangan lebih lanjut, pendekatan ensemble perlu dipertimbangkan guna meningkatkan akurasi dengan menggabungkan berbagai arsitektur model. Alternatif lainnya adalah metode *multi-modal* yang mengombinasikan fitur suara, teks, dan ekspresi wajah untuk mendeteksi emosi secara lebih menyeluruh. Diperlukan pula benchmark dataset berbahasa Indonesia yang lebih komprehensif, termasuk pelabelan valence dan arousal, guna menangkap dinamika emosi secara lebih akurat. Selain itu, penelitian mendatang dapat mengarahkan pengembangan model agar lebih inklusif terhadap individu dengan kebutuhan khusus *(neurodiversity)*. Rekomendasi ini diharapkan dapat memperkaya penelitian pengenalan emosi berbasis suara dan meningkatkan kontribusinya dalam aplikasi nyata yang relevan secara sosial dan budaya.