ABSTRAK

Proliferasi audio deepfake, yang dihasilkan dengan menggunakan teknologi canggih seperti WaveNet dan Generative Adversarial Networks (GAN), menimbulkan ancaman yang signifikan terhadap keamanan digital, termasuk pencurian identitas, informasi yang salah, dan penipuan. Untuk mengatasi tantangan ini, penelitian ini mengusulkan kerangka kerja end-to-end untuk deteksi deepfake audio yang memanfaatkan Mel Spectrograms sebagai fitur input dan model Xception sebagai arsitektur tulang punggung. Metodologi ini mencakup teknik prapemrosesan yang dioptimalkan, seperti normalisasi dan pengubahan ukuran, serta strategi augmentasi data yang kuat untuk meningkatkan kualitas fitur dan generalisasi model. Kerangka kerja ini dievaluasi menggunakan dataset Automatic Speaker Verification (ASV) spoof 2021, yang mencapai akurasi pengujian yang tinggi yaitu 95,86% dengan presisi, recall, dan F1-skor yang seimbang untuk klasifikasi 'asli' dan 'palsu'. Analisis komparatif menunjukkan bahwa model Xception mengungguli ResNet50 dan MobileNetV2 dalam hal akurasi dan generalisasi. Meskipun hasilnya menyoroti ketangguhan dan efisiensi kerangka kerja yang diusulkan, penelitian di masa depan dapat mengeksplorasi lebih jauh jalur preprocessing yang canggih, arsitektur hibrida, dan set data yang beragam untuk lebih meningkatkan kinerja pendeteksian. Penelitian ini memberikan solusi yang andal dan efisien untuk melindungi dari ancaman yang terus meningkat yang ditimbulkan oleh pemalsuan audio.

Kata Kunci: Audio Deepfakes, Mel Spectrograms, Xception Model, Pembelajaran Mendalam, Keamanan Digital