

ABSTRAK

Perkembangan pesat teknologi *artificial intelligence* tidak hanya membuka banyak peluang di berbagai bidang, tetapi juga membawa tantangan baru, utamanya yang menyangkut manipulasi citra digital atau yang dikenal dengan istilah *deepfake*. Teknologi ini mampu menghasilkan sebuah citra atau video palsu dengan presisi tinggi menggunakan bantuan GANs (*Generative Adversarial Networks*) yang dapat menimbulkan kekhawatiran masif di masyarakat, utamanya yang berhubungan dengan keamanan dan privasi digital serta penyebaran berita atau konten *hoax* yang mudah disalahgunakan oleh pihak tertentu. Metode klasifikasi citra wajah asli dan *deepfake* yang akurat menjadi urgensi tersendiri untuk menangani masalah ini. Oleh karena itu, penelitian ini bertujuan untuk mengembangkan model klasifikasi citra wajah asli dan *deepfake* menggunakan *transfer learning* dengan arsitektur Xception. Arsitektur Xception menjadi pilihan dikarenakan kemampuannya yang sangat baik dalam melakukan diferensiasi antara citra wajah asli dan *deepfake* yang memiliki fitur cukup kompleks dengan performa yang efisien. Penerapan *transfer learning* dengan arsitektur Xception dilakukan menggunakan dataset *Deepfake Faces* yang diperoleh dari Kaggle. Dalam penelitian ini digunakan 20.000 citra yang terbagi secara seimbang ke dalam dua kelas, yaitu asli dan *deepfake*. Hasil penelitian menunjukkan bahwa arsitektur Xception mampu menghasilkan performa klasifikasi yang tinggi, dengan akurasi tertinggi mencapai 90,78%. Penelitian ini diharapkan dapat memberikan kontribusi positif dalam keamanan digital, forensik media, dan upaya pemberantasan konten palsu atau *hoax* di masyarakat yang cukup rumit untuk diatasi.

Kata kunci : *Deepfake*, Klasifikasi Citra Wajah, *Transfer Learning*, Arsitektur Xception