

ABSTRAK

Perkembangan teknologi *Artificial Intelligence* (AI) khususnya pada pembuatan gambar sintesis telah mempermudah terciptanya gambar wajah palsu yang sulit dibedakan dari gambar asli. Penyalahgunaan teknologi tersebut dapat menimbulkan berbagai permasalahan seperti pelanggaran privasi, penipuan identitas, hingga penyebaran disinformasi. Oleh karena itu, diperlukan sistem klasifikasi yang mampu membedakan gambar wajah asli dan palsu hasil generatif AI secara akurat. Penelitian ini bertujuan untuk menerapkan model *Data-Efficient Image Transformer* (DeiT) dalam klasifikasi gambar wajah asli atau palsu hasil generatif AI serta mengetahui kinerja model DeiT pada dataset berukuran kecil. Metodologi penelitian yang digunakan adalah CRISP-DM yang meliputi tahapan *business understanding*, *data understanding*, *data preparation*, *modelling*, *evaluation*, dan *deployment*. Dataset yang digunakan terdiri dari 589 gambar wajah asli dan 700 gambar wajah palsu hasil AI. Penelitian menerapkan lima skenario persiapan data, yaitu data asli tanpa augmentasi, data asli dengan augmentasi, oversampling, undersampling, dan augmentasi sebelum *split dataset*, serta tiga rasio pembagian data yaitu 60:20:20, 70:15:15, dan 80:10:10. Model yang digunakan adalah DeiT-Tiny dengan evaluasi menggunakan *accuracy*, *precision*, *recall*, dan *F1-score*. Hasil penelitian menunjukkan bahwa model DeiT mampu bekerja dengan baik pada dataset terbatas. Performa terbaik diperoleh pada rasio pembagian data 80:10:10 dengan skenario data asli menggunakan augmentasi setelah pembagian dataset yang menghasilkan akurasi sebesar 95%, sedangkan skenario augmentasi sebelum *split dataset* memperoleh akurasi tertinggi sebesar 96% namun berpotensi mengalami *data leakage*. Hasil penelitian menunjukkan bahwa model *Data-Efficient Image Transformer* (DeiT) efektif digunakan untuk klasifikasi gambar wajah asli dan palsu hasil generatif AI pada dataset kecil.

Kata Kunci: *Deepfake*, Generatif AI, Klasifikasi Citra, *Data-Efficient Image Transformer* (DeiT), *Deep Learning*, *Computer Vision*.

ABSTRACT

The development of Artificial Intelligence (AI) technology, particularly in synthetic image generation, has facilitated the creation of fake facial images that are difficult to distinguish from real ones. The misuse of this technology may lead to various issues such as privacy violations, identity fraud, and the spread of misinformation. Therefore, a classification system capable of accurately distinguishing between real and AI-generated fake facial images is needed. This study aims to implement the Data-Efficient Image Transformer (DeiT) model for classifying real and fake facial images generated by AI and to evaluate the performance of the DeiT model on a small dataset. The research methodology used in this study is CRISP-DM, which consists of business understanding, data understanding, data preparation, modelling, evaluation, and deployment stages. The dataset used consists of 589 real facial images and 700 AI-generated fake facial images. This study applied five data preparation scenarios, namely original data without augmentation, original data with augmentation, oversampling, undersampling, and augmentation before dataset splitting, as well as three dataset split ratios of 60:20:20, 70:15:15, and 80:10:10. The model used in this study is DeiT-Tiny, evaluated using accuracy, precision, recall, and F1-score metrics. The results show that the DeiT model performs well on a limited dataset. The best performance was achieved using the 80:10:10 dataset split ratio with the original data augmentation after dataset splitting scenario, resulting in an accuracy of 95%, while the augmentation before split scenario achieved the highest accuracy of 96% but had the potential for data leakage. The findings indicate that the Data-Efficient Image Transformer (DeiT) model is effective for classifying real and fake facial images generated by AI on small datasets.

Keywords: *Deepfake, Generative AI, Image Classification, Data-Efficient Image Transformer (DeiT), Deep Learning, Computer Vision.*