Project Summary: Vision Transformer-Based Classification of Authentic vs. AI-Generated Human Faces

This project addresses the critical issue of detecting AI-generated human faces (deepfakes) to protect digital media authenticity. The team fine-tuned a Vision Transformer (ViT-Base/16) model on a large-scale Kaggle dataset of 200,001 images, with a balanced split between authentic and AI-generated faces.

The methodology involved comprehensive data preprocessing (resizing, normalization) and light augmentation to improve the model's generalization. The model was trained for 20 epochs using the AdamW optimizer and cross-entropy loss, achieving a final validation accuracy of 92.67%.

On a held-out test set of 30,001 images, the model demonstrated robust performance, with an overall accuracy of 93%, an AUC of 0.99, and a per-class F1-score of 0.93. The project also included an analysis of misclassified images, highlighting challenges with low-resolution or non-standard poses. The final repository structure includes a summary, the main Python code, a requirements file, and a README to guide potential users and collaborators.