

ABSTRACT

Liveness detection is a crucial component for securing face recognition systems against presentation attacks. Although models based on Convolutional Neural Networks (CNNs) and Transformers show great potential, they still face challenges regarding generalization and balance in detecting various attack types. This research proposes and evaluates an ensemble stacking model that combines a Swin Transformer and an EfficientNetV2 to improve the performance of multi-class liveness detection. The method was evaluated quantitatively using accuracy, BPCER, and APCER metrics, and implemented in a web application for practical validation. The experimental results show that the proposed ensemble model achieved a superior accuracy of 98.17%, with a BPCER of 0.0049 and an ACER of 0.0150, significantly outperforming its individual base models. Although the implementation in a real-time web application confirmed the model's functionality, testing also revealed several failure cases in extreme attack scenarios involving scale and distance, highlighting the current model's limitations.

Keywords: *Liveness Detection, Swin Transformer, EfficientNetV2, Ensemble Stacking.*