ABSTRACT

Digital manipulation tools like deepfakes have ad vanced in sophistication because to the quick development of deep learning and artificial intelligence. Face swapping, in which one person's face is swapped out for another, is one of the most alarming types of deepfakes. This technique produces incredibly lifelike movies that may deceive viewers. Detecting these manipulated videos is crucial to mitigating their negative impact on privacy and security. This paper proposes an ensemble approach to detecting face swap deepfakes by combining the Swin Transformer and Bidirectional Long Short-Term Memory (BiLSTM) with an attention mechanism. The Swin Transformer is employed for spatial feature extraction, while the BiLSTM captures temporal patterns between frames, and the attention mechanism focuses on the most relevant timesteps. The model is evaluated on the FaceForensics++ dataset, achieving a validation accuracy of 93.81% with a validation loss of 0.19, outperforming the Long Short-Term Memory (LSTM), Fully Convolutional Network (FCN), and Convolutional Neural Network- Bidi rectional Long Short-Term Memory (CNN-BiLSTM) models. Experimental results demonstrate the superior ability of the Swin-BiLSTM With Attention model to accurately detect face swap manipulations, even under varying facial poses, lighting conditions, and motion variations. The proposed method shows promise in addressing the challenges of deepfake detection, offering potential applications in privacy protection, misinfor mation prevention, and security. Future work may explore the integration of additional data modalities or advanced techniques to further enhance detection accuracy and robustness.

Keywords: *Deepfake Detection, Face Swap, Swin Transformer, BiLSTM, Attention Mechanism.*