ABSTRACT

Deepfake is an artificial intelligence-based technology with deep learning capabilities that can create or manipulate a person's face realistically. In the latest VIDA survey, "Where's The Fraud – Protecting Indonesia Business from AI Generated Fraud," a 1,540% increase in deepfake fraud cases was found in the APAC region from 2022 to 2023, while in Indonesia, there was a 1,550% increase in fraud cases. This significant increase in deepfake cases highlights new challenges in detecting visual manipulation using artificial intelligence (AI) technology. While the use of deep learning technology offers benefits, it can also pose a serious threat in cases of fraud and extortion. Therefore, AI-based detection methods are needed to effectively and efficiently detect deepfakes. This research focuses on the development and application of a deepfake detection method based on Convolutional Neural Network (CNN) using the Residual Network 50 (ResNet50) architecture modified with the Convolution Block Attention Module (CBAM) to enhance accuracy in detecting artifact patterns that appear in deepfake images, as a contribution to efforts to mitigate the increasing misuse of deepfake technology. This research uses a Kaggle dataset consisting of 8,000 datasets. The dataset will be used to compare data authenticity through normalization, model training, and performance evaluation with accuracy metrics consisting of original images and images that have been deepfaked using AI-based technology. In a previous study, the use of ResNet50 in classification showed a result of 78.87%, while the integration of the attention mechanism in ResNet50 with the Long-Distance Attention Module achieved an accuracy of 94.30% and an AUC of 98.70%. The results of this study show that the use of CBAM on 8,000 datasets produced accuracy metrics of 68.44%, precision of 72.26%, recall of 53.73%, and an F1-score of 61.63%. and from observations on a 6,000-dataset, the results showed an accuracy of 75.27%, precision of 67.91%, recall of 95.80%, and an F1-score of 79.43%. With this approach, it is hoped that it can become an effective solution to enhance the security of biometric-based systems and prevent the misuse of deepfake technology across various sectors.

Keywords: *Deepfake*, *Convolutional Neural Network*, *ResNet50, Convolutional Block Attention Module, Gradient-weighted Class Activation Mapping, Rectified Linear Unit, Sigmoid.*