ABSTRACT

Online learning has rapidly developed in recent years due to current demands and the disruption of traditional teaching methods by technology. In its implementation, engagement or participant involvement is one of the critical factors in creating an effective online learning process. Facial Expression Recognition (FER) technology has been utilized to address this issue, but approaches based on Convolutional Neural Network (CNN) models face limitations as they cannot leverage temporal information in videos. Therefore, this study proposes the use of a 3D CNN model on more complex spatial-temporal data and improve the accuracy of participant engagement detection based on facial expressions. This research aims to evaluate the performance of the 3D CNN model compared to the CNN model and optimize parameters such as epoch, batch size, and learning rate to enhance training efficiency. The model was trained using a video dataset of facial expressions with six categories (anger, sad, fear, neutral, happy, surprised). The results indicate that the 3D CNN model effectively captures spatial-temporal information compared to the CNN model. The best model utilized the 3D Inception-ResNet + LSTM architecture with optimal parameter configurations, achieving 99.07% accuracy with more efficient training time.

Keywords: facial expression recognition, spatial-temporal, 3D CNN, Inception-ResNet, 3D Inception-ResNet, LSTM.