ABSTRACT

Traffic accidents peaked in 2023 with 148,575 cases, 20% of which were linked to drowsiness tripling crash risk due to impaired alertness and delayed reaction. This research proposes a real-time method to extract Eye Aspect Ratio (EAR) and Mouth Aspect Ratio (MAR) from sequential facial images using MediaPipe. EAR and MAR are computed from eye and mouth landmarks, then structured sequentially to capture temporal changes in subject condition. This representation effectively models drowsiness transitions and serves as input for deep learning-based detection.The study covers five main components: classification method testing, input structuring, image enhancement, augmentation, and system pipeline. Using the NTHU-DDD dataset, data are windowed in 60-frame segments and processed with MediaPipe for EAR and MAR extraction. Experimental results demonstrate that the CNN-LSTM model is capable of effectively processing sequential EAR and MAR features for drowsiness detection. A full representation with an implicit input shape of (120, 1) yielded superior performance compared to limited or separately processed features. The use of Synthetic Minority Oversampling Technique (SMOTE) helped improve performance by balancing class distribution. However, several augmentation methods that performed well under hold-out validation exhibited unstable results when evaluated using cross-validation. Overall, the CNN-LSTM-120FT model without image enhancement or augmentation proved to be the most stable and reliable across various testing scenarios, achieving an accuracy of 85.59% and a highest precision of 92.31%.

Keywords: *Traffic Accidents, Drowsiness, Eye Aspect Ratio (EAR), Mouth Aspect Ratio (MAR), Deep Learning, Sequential Images*