# CONCLUSION

Abnormal behavior recognition plays a significant role in intelligent campus surveillance systems. In this study, a CABR50 dataset was created, and a framework named temporal segment transformers was proposed to address the problem of identifying abnormal behavior on campus. Specifically, TST divides the input video into three segments of equal duration from the original video and obtains snippets uniformly sampled from its segment. These snippets are used as the inputs for the backbone network. Each snippet produces its initial prediction of the class, followed by a consensus function between the snippets exported as the final prediction, enabling dynamic global video modeling. Extensive experiments are carried out on the three split CABR50 datasets to verify the classification accuracy: TSN, Slow fast, Swin -B, and TST methods. In addition, the superiority of the proposed method in classifying abnormal behaviors on campus is verified in terms of the analysis result. To explore the model’s generalization performance, we experimented with it on the UCF-101 dataset and achieved promising results. In summary, this work demonstrates the feasibility of using abnormal campus behavior recognition. In addition, our proposed TST can effectively model long-range behavior and achieve competitive results on CABR50. However, the model should perform better on the abnormal campus behavior categories of coughing, debating, and yelling that belong to hard data. In the future, we will try to combine multimodal approaches , such as extracting audio features from videos, to assist in classifying hard data for abnormal campus behavior. In addition, we can take an efficient approach to reduce the complexity of the model and overcome the problem of imbalance in the corresponding category data, such as GAN, which generates new training data for unbalanced categories