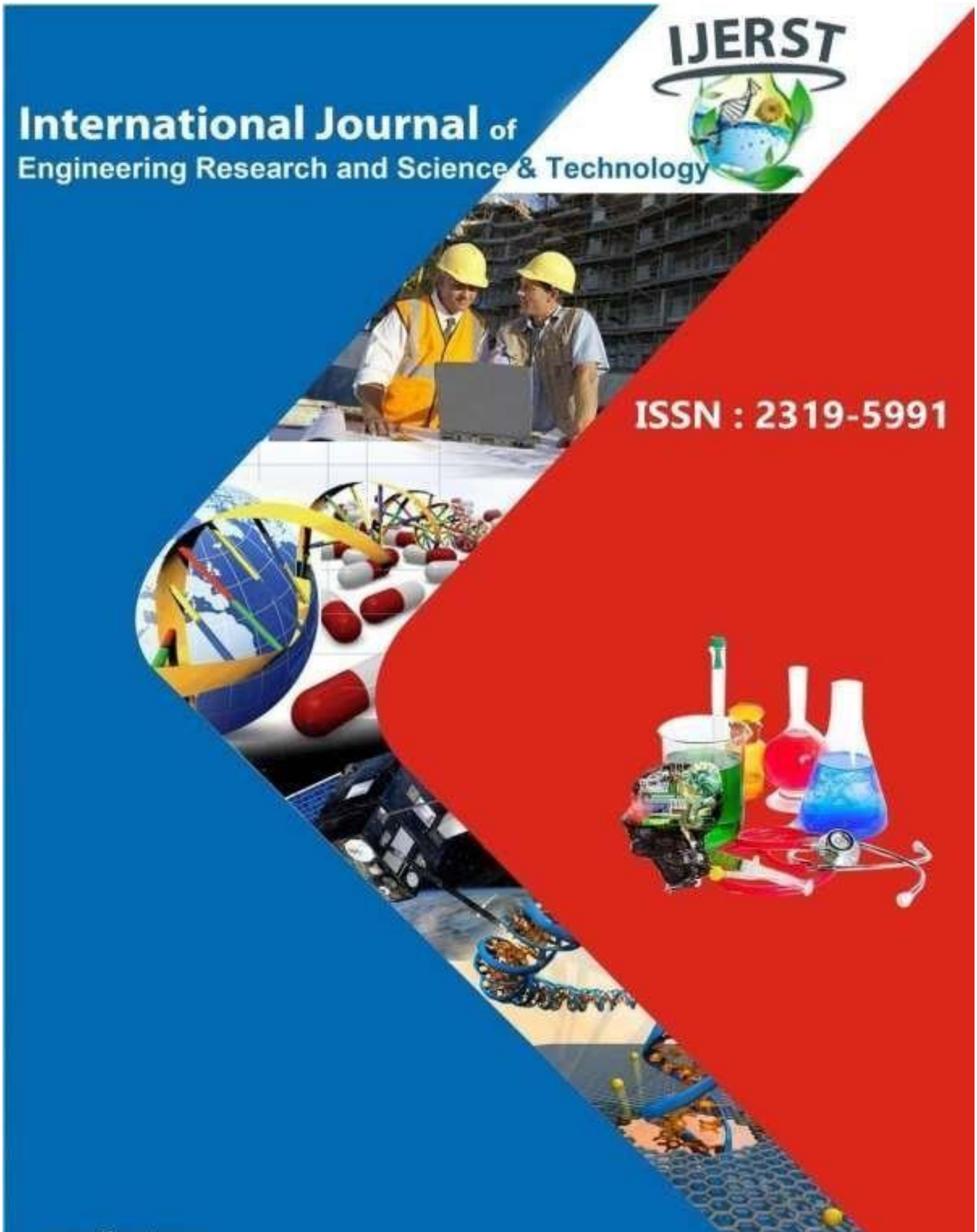


Email: editor@ijerst.com or editor.ijerst@gmail.com

ISSN 2319-5991 www.ijerst.com

Vol. 17, Issue.2, 2024

CAMPUS ABNORMAL BEHAVIOUR RECOGNITION WITH TEMPORAL SEGMENT TRANSFORMERS

[1]MR. N. CHANDIRAPRAKASH,[2]VENKATA RAMANA,[3]N.CHANDRAKANTH,[4]YARRAMSHETTI KAVYA,[5]VISHAL,[6]KOTHAKOTA DHANUSH

[1]Assistant Professor,department of information technology, malla reddy institute of engineering and technology(autonomous),dhulapally,secundrabad,chandrunit26@gmail.com

[2,3,4,5,6]UG students, department of information technology,malla reddy institute of engineering and technology(autonomous),Dhulapally,Secundrabad

ABSTRACT

The intelligent campus surveillance system is beneficial to improve safety in school. Abnormal behavior recognition, a field of action recognition in computer vision, plays an essential role in intelligent surveillance systems. Computer vision has been actively applied to action recognition systems based on Convolutional Neural Networks (CNNs). However, capturing sufficient motion sequence features from videos remains a significant challenge in action recognition. This work explores the challenges of video-based abnormal behavior recognition on campus. In addition, a novel framework is established on long-range temporal video structure modeling and a global sparse uniform sampling strategy that divides a video into three segments of identical durations and uniformly samples each snippet. The proposed method incorporates a consensus of three temporal segment transformers (TST) that globally connects patches and computes self attention with joint spatiotemporal factorization. The proposed model is developed on the newly created campus abnormal behavior recognition (CABR50) dataset, which contains 50 human abnormal action classes with an average of over 700 clips per class. Experiments show that it is feasible to implement abnormal behavior recognition on campus and that the proposed method is competitive with other peer video recognition in terms of Top-1 and Top-5 recognition accuracy. The results suggest that TST-L+ can improve campus abnormal behavior recognition, corresponding to Top-1 and Top-5 accuracy results of 83.57% and 97.16%, respectively.

ISSN 2319-5991 www.ijerst.com

Vol. 17, Issue.2, 2024

I. INTRODUCTION

The "Campus Abnormal Behaviour Recognition with Temporal Segment Transformers" project represents a significant advancement in the domain of campus security and surveillance. With the increasing need for ensuring safety and security on educational campuses, there arises a demand for intelligent systems capable of detecting abnormal behaviors and potential security threats in real-time. Traditional surveillance methods often fall short in effectively monitoring large campus environments and identifying abnormal activities. In response to this challenge, this project proposes the implementation of Temporal Segment Transformers (TST), a state-of-the-art deep learning architecture, for recognizing abnormal behaviors from surveillance videos on campus premises. By leveraging the temporal information encoded in video segments, TST offers superior performance in detecting and classifying various types of abnormal behaviors, ranging from trespassing and vandalism to violence and emergencies. Through the development and deployment of TST-based surveillance systems, this project aims to enhance campus security, mitigate security risks, and ensure the safety and well-being of students, faculty, and staff.

II.EXISTING SYSTEM

Xie et al. [45] used spatiotemporal representations to learn the posture estimation of college students to identify abnormal behavior. They analyzed the behavior of sleeping and using mobile phones in the classroom. Other researchers [46], [47] have explicitly looked at abnormal behavior in the laboratory. Rashmi et al. [46] apply YOLOv3 to locate and recognize student actions in still images from surveillance video in school laboratories. Unlike Rashmi's image recognition-based analysis of students' abnormal behavior in the laboratory, Banerjee et al. [47] used video. They propose a deep convolutional network architecture to detect and classify the behavioral patterns of students and teachers in computer-enabled laboratories. The above works can significantly demonstrate recognition of abnormal behavior in specific scenarios on campus. However, do not aim to reveal the feasibility of numerous abnormal behaviors in multiple scenarios on campus. Although there is limited research on campus aberrant behavior

ISSN 2319-5991 www.ijerst.com

Vol. 17, Issue.2, 2024

recognition, it is a form of video understanding. The following is a review of video understanding research relevant to our work.

Although self-attention is added as a submodule to CNNs to improve this temporal modeling video understanding, the ability of remote modeling can be made more robust by applying a pure transformer. These algorithms [27], [29], [30] are related to decomposing spatiotemporal self-attention using different factorized methods. They achieved better results than previous pure CNN and methods for adding self-attention units to videos. However, depending on the global attention modeling, these methods lead to a geometric increase in complexity.

Subsequently, MVTs [28] and the video Swin transformer [31] presented the idea of multi-scale hierarchical modeling by calculating the self-attention of multi-scale windows, which is much lower than the computational complexity of global self-attention. Moreover, it exceeds the previous decomposition space-time modeling methods in terms of accuracy and efficiency. The latest research [32] proposed a multi-view transformer consisting of multiple independent encoders to represent different dimensional input views, fusing information across views through horizontal connections. Although they achieve state-of-the-art performance, their method relies on many unpublic datasets and trained views. It may limit their generalization performance when applied to new videos or tasks beyond their original training scope. Therefore, a fairer comparison is required. In the case of employing ImageNet-21K as pre-training data to initialize network weights, the transformer method established on multi-scale hierarchical modeling [31] has more competitive advantages in video understanding.

Disadvantages

- In the existing work, the system did not find Sensors and wireless data transmission for measuring food safety.
- This system is less performance due to lack of Real-time Prediction Campus Abnormal Behaviour.

III.PROPOSED SYSTEM

• We propose a consensus of three temporal segment transformers (TST) based on the video Swin transformer for the new campus abnormal behavior

International Journal of Engineering Research and Science & Technology
ISSN 2319-5991 www.ijerst.com
Vol. 17, Issue.2, 2024

recognition (CABR50) dataset. It enhances the ability to capture motion sequences and model long-range abnormal behavior on campus.

• We perform extensive comparative experiments with state-of-the-art methods for recognizing abnormal campus behaviors. The results show that it is feasible and can improve the accuracy of abnormal campus behavior recognition. In addition, we demonstrate the performance comparison of the TST and previous methods on the UCF-101 dataset. It indicates that our proposed TST model has acceptable generalization performance.

• Our research provides essential technical support for the identification and early warning of abnormal behavior on campus, which plays an essential role in intelligent campus surveillance systems.

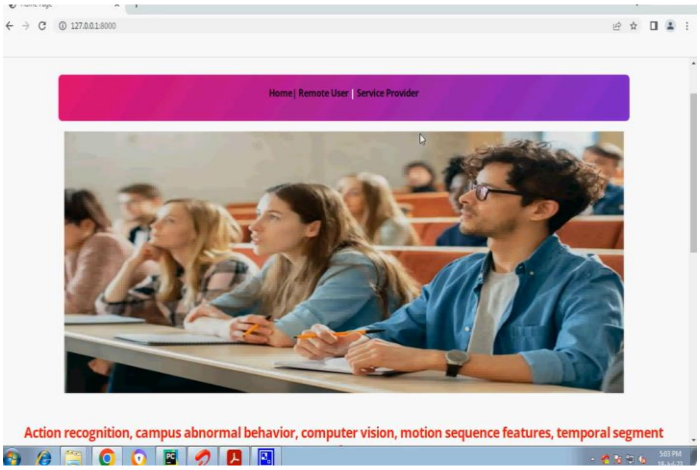
Advantages

- ❖ The system is more effective since it involves Convolutional Neural Networks (CNNs) method.
- ❖ The system finds more ADVANTAGES OF THE system which designed a campus abnormal behavior recognition framework called temporal segment transformer (TST) to exploit temporal action features and achieve video-level global modeling.

IV.MODULES

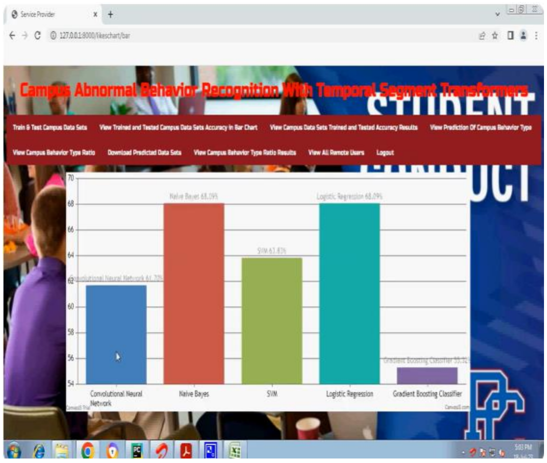
➢ Service Provider

In this module, the Service Provider has to login by using valid user name and password.



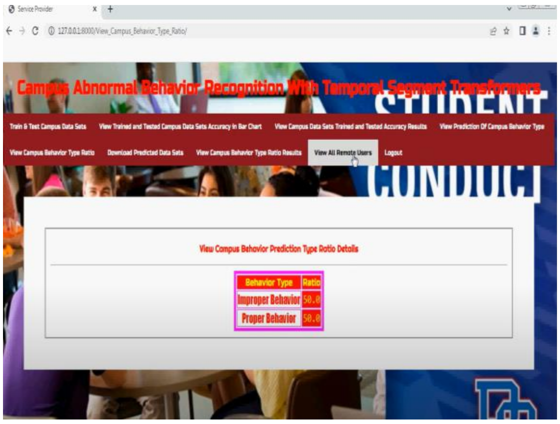
After login successful he can do some operations such as Browse Datasets and Train & Test Data Sets, View Trained and Tested Accuracy in Bar Chart,

International Journal of Engineering Research and Science & Technology
ISSN 2319-5991 www.ijerst.com
Vol. 17, Issue.2, 2024



View Trained and Tested Accuracy Results,

View Predicted Type, View Type Ratio, Download Predicted Data Sets, View Type Ratio Results,



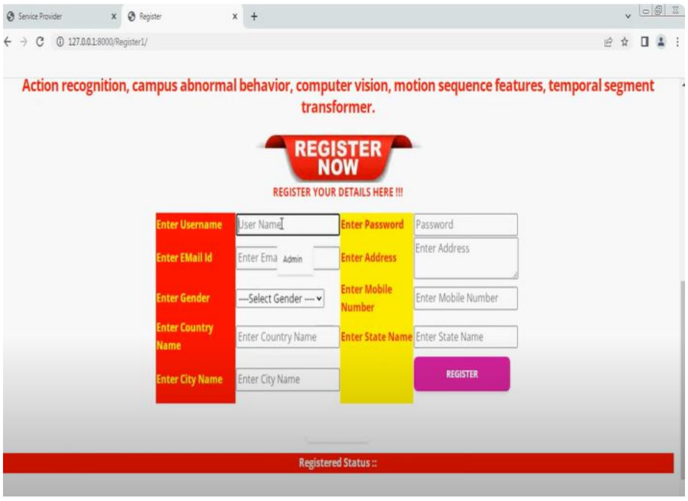
View All Remote Users.

➢ View and Authorize Users

In this module, the admin can view the list of users who all registered. In this, the admin can view the user's details such as, user name, email, address and admin authorizes the users.

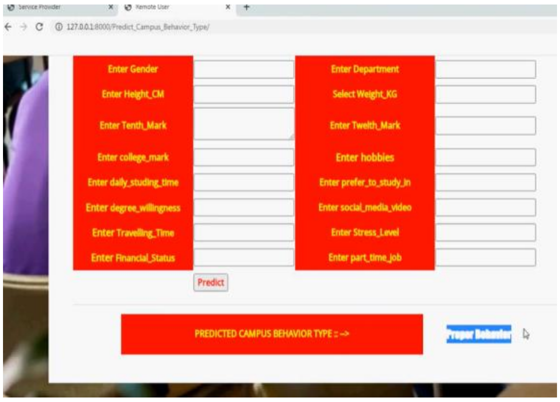
➢ Remote User

In this module, . User should register before doing any operations. Once user registers, their details will be stored to the database.



After registration successful, he has to login by using authorized user name and password. Once Login is successful user will do some operations like REGISTER AND LOGIN, after login we have to Predict Type,

International Journal of Engineering Research and Science & Technology
ISSN 2319-5991 www.ijerst.com
Vol. 17, Issue.2, 2024



VIEW YOUR PROFILE.

V.CONCLUSION

In conclusion, the "Campus Abnormal Behaviour Recognition with Temporal Segment Transformers" project presents a novel approach to enhancing campus security through advanced deep learning techniques. By leveraging Temporal Segment Transformers (TST), the project offers a robust solution for real-time detection and classification of abnormal behaviors in surveillance videos on educational campuses. The integration of TST-based surveillance systems enables proactive monitoring, early detection, and rapid response to security threats, thereby enhancing the safety and security of campus environments. Moving forward, the deployment of TST-based surveillance systems has the potential to revolutionize campus security practices, providing stakeholders with actionable insights and enabling timely intervention to mitigate security risks and ensure the well-being of campus communities.

VI.REFERENCES

- Wang, X., Girshick, R., Gupta, A., & He, K. (2018). Non-local neural networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (pp. 7794-7803).
- Lin, T. Y., Dollár, P., Girshick, R., He, K., Hariharan, B., & Belongie, S. (2017). Feature pyramid networks for object detection. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (pp. 2117-2125).
- Feichtenhofer, C., Fan, H., Malik, J., & He, K. (2019). SlowFast networks for video recognition. In Proceedings of the IEEE International Conference on Computer Vision (ICCV) (pp. 6202-6211).
- Zolfaghari, M., Singh, K., Brox, T., & Schiele, B. (2018). ECOfeatures: efficient convolutional features for high-level visual tracking. In Proceedings of the IEEE Conference on Computer Vision

ISSN 2319-5991 www.ijerst.com
Vol. 17, Issue.2, 2024

and Pattern Recognition (CVPR) (pp. 6704-6713).

- Chen, L. C., Zhu, Y., Papandreou, G., Schroff, F., & Adam, H. (2018). Encoder-decoder with atrous separable convolution for semantic image segmentation. In Proceedings of the European Conference on Computer Vision (ECCV) (pp. 801-818).
- Wang, L., Xiong, Y., Wang, Z., Qiao, Y., Lin, D., Tang, X., & Van Gool, L. (2016). Temporal segment networks: Towards good practices for deep action recognition. In European Conference on Computer Vision (ECCV) (pp. 20-36).
- Zhang, Z., Luo, P., Loy, C. C., & Tang, X. (2017). Facial landmark detection by deep multi-task learning. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (pp. 1285-1294).
- Redmon, J., Divvala, S., Girshick, R., & Farhadi, A. (2016). You only look once: Unified, real-time object detection. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (pp. 779-788).
- Simonyan, K., & Zisserman, A. (2014). Two-stream convolutional networks for action recognition in videos. In Advances in neural information processing systems (pp. 568-576).
- He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (pp. 770-778).
- Szegedy, C., Ioffe, S., Vanhoucke, V., & Alemi, A. A. (2017). Inception-v4, Inception-ResNet and the impact of residual connections on learning. In Proceedings of the AAAI Conference on Artificial Intelligence (AAAI) (pp. 4278-4284).
- Sandler, M., Howard, A., Zhu, M., Zhmoginov, A., & Chen, L. C. (2018). MobileNetV2: Inverted residuals and linear bottlenecks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (pp. 4510-4520).
- Dai, J., Qi, H., Xiong, Y., Li, Y., Zhang, G., Hu, H., ... & Wei, Y. (2017). Deformable convolutional networks. In Proceedings of the

ISSN 2319-5991 www.ijerst.com

Vol. 17, Issue.2, 2024

IEEE International Conference on Computer Vision (ICCV) (pp. 764-773).

- Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., ... & Polosukhin, I. (2017). Attention is all you need. In Advances in neural information processing systems (pp. 5998-6008).
- He, K., Gkioxari, G., Dollár, P., & Girshick, R. (2017). Mask R-CNN. In Proceedings of the IEEE International Conference on Computer Vision (ICCV) (pp. 2961-2969).