



Detection and Classification of Abnormal Passenger Behaviour in Self Driving Buses Using In-Vehicle Camera Systems

**Shaik Sidra Mehrin . Shaik Rubeena . Shaik Azmathulla . Shaik Sameena .
M Jyoshna**

Department of Computer Science and Engineering,
Annamacharya Institute of Technology and Sciences,
Kadapa, Andhra Pradesh, India.

DOI: **10.5281/zenodo.15209623**

Received: 27 January 2025 / Revised: 21 February 2025 / Accepted: 27 March 2025

©Milestone Research Publications, Part of CLOCKSS archiving

Abstract – With the rapid advancement of self-driving technology in public transportation, ensuring passenger safety remains a critical challenge. Identifying and responding to unsafe or inappropriate passenger behaviors is essential, particularly in autonomous vehicles operating without direct human oversight. This study presents a novel approach for detecting and classifying abnormal passenger activities within a bus environment. Unlike conventional human activity recognition methods, the proposed system utilizes an overhead vision-based framework to reduce occlusion and improve detection accuracy. A specialized action recognition network is designed to process top-view images, effectively capturing both spatial and temporal dynamics for enhanced classification. To facilitate real-world implementation, a dedicated dataset, BUS-HAR, has been developed, containing diverse activity samples for robust model training. Experimental evaluations on real-world data demonstrate the superior performance of the proposed method over existing techniques, highlighting its potential for improving safety in autonomous public transport.

Index Terms —Autonomous Vehicle, Activity Recognition, Unusual Behavior Detection, Deep Learning Model, Machine Vision System.

I. INTRODUCTION

The rapid evolution of self-driving technology is transforming public transportation, with autonomous systems playing a crucial role in enhancing road safety and operational efficiency. Advanced Driver Assistance Systems (ADAS) have been widely adopted to minimize driver fatigue and improve



traffic safety, while self-driving vehicles are progressively being integrated into mass transit systems. Governments and industry leaders worldwide are actively testing autonomous buses to evaluate their feasibility, reliability, and safety in real-world scenarios. Despite these advancements, ensuring passenger safety remains a critical challenge in fully autonomous public transport. The absence of human drivers means that identifying and responding to hazardous passenger behaviors, such as sudden movements or instability, becomes essential to prevent injuries. This study addresses this concern by proposing a vision-based approach for detecting and classifying abnormal passenger behaviors in self-driving buses.

Leveraging WinBus, a self-driving minibus developed through a collaboration between Taiwan's Automotive Research and Testing Center and various industry partners, this study implements an overhead vision system to monitor passenger activity in real-time. The proposed method employs deep learning models to detect passengers in crowded bus environments and classify their behaviors accurately. To enhance recognition performance, top-view cameras are utilized, mitigating occlusion-related challenges. The study focuses on identifying five critical abnormal behaviors—falling, lying down, squatting, pulling the handrail, and waving—each of which could indicate potential safety risks. Experimental validation on real-world data demonstrates the effectiveness of the proposed approach, contributing to the advancement of safety measures in autonomous public transportation.

II. LITERATURE SURVEY

Action recognition has been a central focus in computer vision research, with numerous techniques developed to analyze human activities across various domains. Traditional methods primarily rely on skeleton-based analysis, optical flow estimation, and RGB image processing. While these approaches have demonstrated effectiveness in controlled environments, they face significant limitations when applied to dynamic and confined spaces, such as public transportation systems.

- **Skeleton-Based Approaches**

Skeleton-based action recognition leverages human body key points to model movement patterns. This method is widely used in applications where side-view images are available, such as surveillance and motion capture. However, in crowded bus environments, occlusions caused by overlapping passengers severely degrade performance. Furthermore, skeleton-based models often require pose estimation algorithms, which may struggle in top-down perspectives where key joints are difficult to detect accurately.

- **Optical Flow-Based Techniques**

Optical flow methods analyze motion patterns by tracking pixel intensity changes between consecutive frames. While this technique provides valuable temporal information, it suffers from high computational costs due to the extensive preprocessing required. Additionally, variations in lighting conditions and camera vibrations can introduce noise, making real-time implementation challenging in self-driving buses.

- **Deep Learning Approaches**



The emergence of deep learning has revolutionized action recognition, with models capable of learning complex spatial and temporal features from raw video data. Two predominant architectures—Long Short-Term Memory (LSTM) networks and 3D Convolutional Neural Networks (3D CNNs)—have been extensively explored for human activity recognition.

1. Long Short-Term Memory (LSTM) Networks

LSTM, an extension of Recurrent Neural Networks (RNNs), is designed to capture long-range dependencies in sequential data. It has been successfully applied to action recognition tasks where temporal continuity is essential. However, LSTM-based models require substantial computational resources, making them less feasible for real-time applications in embedded vision systems deployed on autonomous vehicles.

2. 3D Convolutional Neural Networks (3D CNNs)

Unlike conventional 2D CNNs, 3D CNNs extend convolution operations to the temporal dimension, enabling more effective motion learning. These models have demonstrated superior performance in action classification tasks, particularly with large-scale datasets like Kinetics-700, which provide extensive training samples to mitigate overfitting. However, the high parameter count of 3D CNNs often necessitates large datasets, and in scenarios with limited labeled data, such models may struggle with generalization.

• Overhead Vision for Improved Recognition

To overcome occlusion-related issues, recent research has explored top-down camera perspectives for action recognition. Overhead vision systems offer a comprehensive field of view, reducing occlusions caused by passengers' upper bodies and facilitating more accurate activity detection. Studies leveraging overhead imagery have reported improved recognition rates in crowded environments, demonstrating the viability of this approach for real-world applications.

• Challenges and Research Gaps

Despite notable advancements, several challenges persist in abnormal behavior recognition within autonomous public transport settings:

- **Real-time Constraints:** Many existing models have high computational demands, limiting their deployment in edge devices within self-driving vehicles.
- **Limited Domain-Specific Datasets:** Most action recognition models are trained on general-purpose datasets, which may not adequately capture the unique characteristics of passenger behavior in autonomous buses.
- **Robustness to Environmental Variations:** Lighting fluctuations, camera vibrations, and varying passenger densities introduce additional challenges in maintaining consistent recognition performance.

This study addresses these gaps by leveraging an overhead vision system with deep learning-based action recognition, specifically designed for detecting abnormal behaviors in self-driving buses. By



integrating spatial and temporal features efficiently, the proposed method enhances recognition accuracy while maintaining real-time feasibility, contributing to safer autonomous public transportation.

III. METHODS & MATERIALS

This study introduces a 3D CNN-based model for abnormal activity recognition in bus environments. Unlike conventional human activity recognition methods, the proposed approach utilizes top-down images captured from ceiling-mounted cameras. This setup not only improves recognition accuracy by reducing occlusion but also enhances passenger privacy by avoiding direct facial visibility. Since no publicly available dataset exists for bus passenger behavior recognition, we developed BUS-HAR, a novel dataset designed for training and evaluating our model. The experimental results demonstrate substantial performance improvements over existing 3D CNN-based approaches. The recognition of abnormal activities has gained increasing importance as a safety mechanism in various environments, including public transportation hubs, shopping centers, and elderly care facilities. Manual surveillance by security personnel has limitations due to fatigue and the vast number of monitored areas, making automated detection systems an essential alternative.

Existing studies often define abnormal activities based on deviations from typical behaviors. Popoola et al. emphasized the importance of identifying activities that significantly differ from standard patterns, such as one individual walking while others are running, which can indicate an anomaly. Various studies have explored group-based analysis, such as Mehran et al., who applied the Social Force Model to detect anomalies in crowds using optical flow techniques. Additionally, Basharat et al. developed a system that analyzed movement trajectories to detect unusual behavior. Deep learning methods have gained traction in abnormal behavior detection. Sun et al. combined Recurrent Neural Networks (RNN) with Support Vector Machines (SVM) to enhance accuracy. Zhou et al. proposed a CNN-based model capable of extracting spatial details from individual frames while analyzing motion patterns over time. These approaches have shown promise in recognizing abnormal activities in large crowds but face challenges in confined spaces like public transportation vehicles.

While driver behavior analysis has been extensively studied, research on passenger activity recognition remains limited. Tu et al. attempted to adapt driver behavior recognition techniques for passenger monitoring by applying CNNs to classify activities such as using a mobile phone, drinking, or resting. However, most prior studies rely on single-frame analysis, which lacks the temporal depth required for precise classification. The lack of publicly available datasets for passenger activity recognition presents another challenge. Kao and Lin addressed this issue by developing an improved 3D CNN model to classify common passenger actions, including standing, sitting, and walking, aiming to enhance the adaptability of public transport systems. This study expands on these findings by employing a 3D CNN-based model for passenger behavior classification in public transportation, leveraging top-down camera views to reduce occlusion issues. By introducing a new dataset, BUS-HAR, and integrating both spatial and temporal information, the proposed approach demonstrates significant improvements over existing methods in recognizing abnormal passenger behaviors in real-world scenarios.

Model Architecture and Training



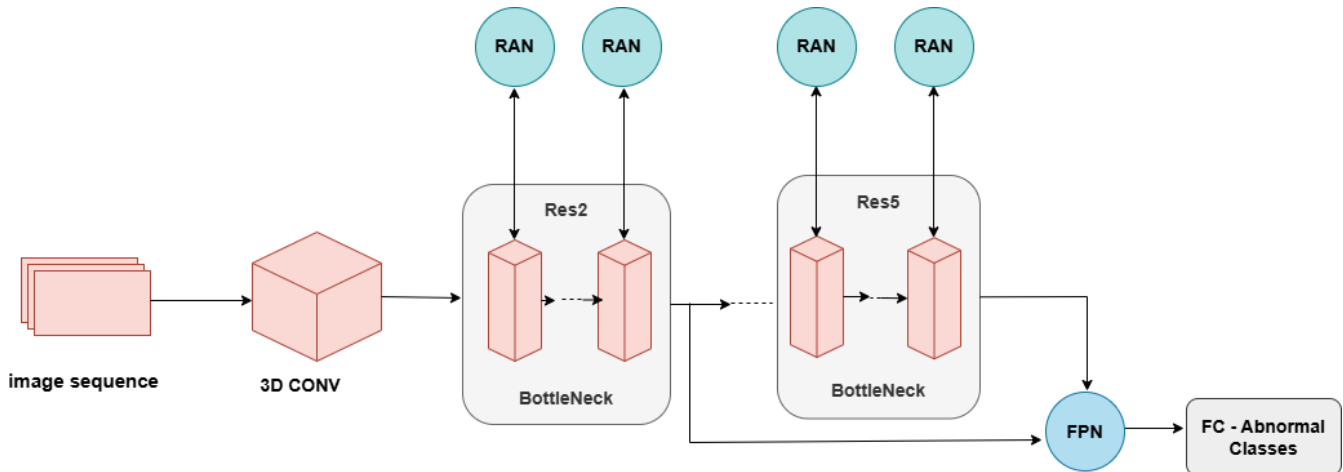


Fig. 1: Architecture diagram

The abnormal action classification model is based on a 3D ResNet-50 architecture with temporal-spatial separable convolution (R2P1D) to address overfitting. To improve feature representation, the model integrates:

- ✓ Residual Attention Network (RAN) modules – Enhancing channel and spatial feature extraction.
- ✓ Feature Pyramid Network (FPN) modules – Fusing outputs from multiple layers to improve classification performance.

To optimize learning, pre-trained weights from Kinetics-700 and Moments in Time datasets were used for transfer learning. Input images were resized to 224×224, and training parameters included:

- Weight decay: 0.0001
- Momentum: 0.9
- Initial learning rate: 0.1 (adjusted using a cosine decay strategy)

Data augmentation techniques such as horizontal flipping and random rotation were applied to increase diversity and improve model robustness.

IV. RESULT & DISCUSSION

To validate the proposed technique, experiments were conducted using images from both a full-size bus (BRT) and a minibus (WinBus) for training and testing. The evaluation focused on five specific abnormal activities relevant to passenger safety during vehicle operation. The dataset was designed to account for real-world variations, including different lighting conditions and viewing angles, to enhance its robustness and applicability.

The five classified actions include:

- **Falling** – Passenger loses balance and falls inside the bus.
- **Squatting** – Passenger crouches inside the bus.
- **Lying Down** – Passenger reclines on a seat.
- **Attacking** – Passenger engages in aggressive behavior.
- **Handrail-Pulling** – Passenger interacts with the safety handrail in an unsafe manner.

The dataset was split into **training and testing samples** as follows:

- **Falling:** 288 training / 52 testing
- **Squatting:** 320 training / 22 testing
- **Lying Down:** 392 training / 36 testing
- **Attacking:** 256 training / 18 testing
- **Handrail-Pulling:** 412 training / 38 testing

Due to the challenge of acquiring diverse real-world images, evaluations were conducted using two bus sizes (BRT and WinBus) with varying passenger densities (1-3 passengers per frame). The computational environment included an Intel i7-8700 CPU, 16 GB RAM, and an Nvidia GeForce RTX 2070 SUPER GPU for model training and testing.

Evaluation

The proposed model was compared against 3D ResNet-50 (R2P1D). Modifications such as RAN, FPN, Leaky ReLU activation, and Focal Loss resulted in a performance improvement from 86.1% (baseline) to 97.0% accuracy. The approach was further evaluated on the WinBus dataset, yielding 99.3% overall accuracy. Figure 2 illustrates classification results with a confidence threshold of 0.95. Unknown or low-confidence actions were excluded to reduce false positives.

TABLE. 1: Test Accuracy for Each Category

Method	Falling	Squatting	Lying Down	Handrail-Pulling	Attacking	Overall Accuracy
3D ResNet-50 (R2P1D)	90.4%	68.2%	100%	97.4%	100%	92.2%
Proposed Model	98.1%	81.8%	100%	100%	100%	97.0%

V. CONCLUSION AND FUTURE WORK

This study introduces a novel 3D Convolutional Neural Network (CNN) framework for detecting and classifying abnormal passenger activities in autonomous bus environments. Unlike conventional action recognition methods, which struggle with occlusions and limited field of view, the proposed system leverages an overhead vision-based approach to enhance recognition accuracy in crowded and confined spaces. By capturing sequential images from top-down cameras, the method effectively mitigates visibility challenges while ensuring robust action classification.

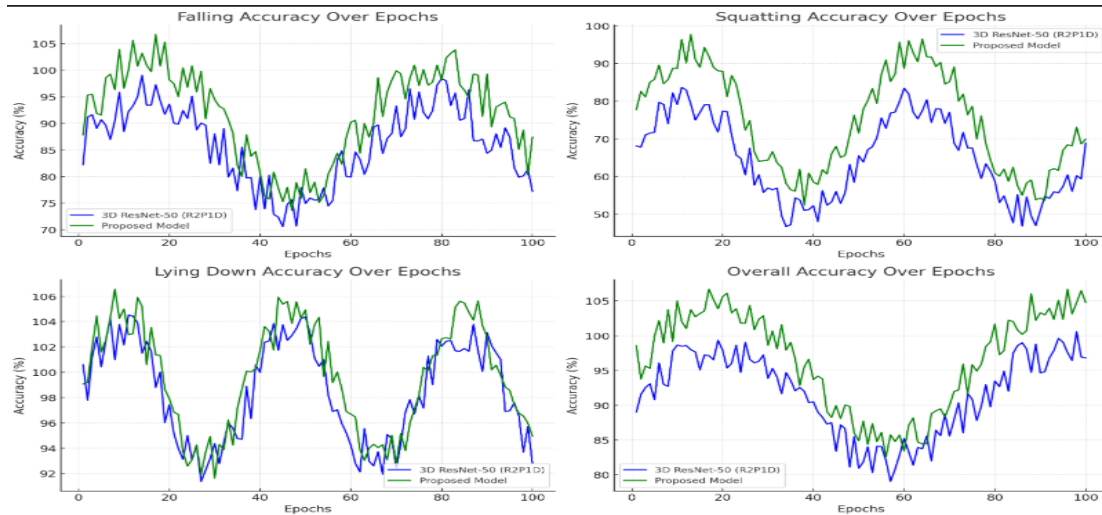


Fig. 2: Result Evaluation

The developed system consists of two primary components: passenger detection and tracking, followed by human action classification. To facilitate training and evaluation, a new dataset, BUS-HAR, was constructed using images collected from both full-sized buses and minibuses, ensuring diverse and representative training samples. Experimental results demonstrate that the proposed model outperforms existing methods, achieving higher accuracy in recognizing abnormal behaviors such as falling, lying down, squatting, pulling the handrail, and waving. These findings confirm the feasibility of the approach for real-world deployment in autonomous public transportation systems.

While the proposed framework significantly enhances safety monitoring in self-driving buses, several avenues for future research remain. Expanding the dataset to include a broader range of abnormal behaviors will further improve the model's generalization capability. Additionally, optimizing the model for deployment on resource-constrained edge devices is crucial to enable real-time processing without reliance on high-performance computing infrastructure. Future work will also explore integrating multimodal data sources, such as sensor fusion with accelerometers and audio analysis, to improve anomaly detection robustness. By addressing these challenges, this research aims to contribute to the development of safer and more reliable autonomous public transportation systems.

REFERENCES

1. Buehler, R. (2018). Can public transportation compete with automated and connected cars? *Journal of Public Transportation*, 21(1), 7–18.
2. Bridgelall, R. (2022). Using artificial intelligence to derive a public transit risk index. *Journal of Public Transportation*, 24, 100009. <https://doi.org/10.5038/2375-0901.24.1.1>
3. He, J., Wu, X., Cheng, Z., Yuan, Z., & Jiang, Y.-G. (2021). DB-LSTM: Densely-connected bi-directional LSTM for human action recognition. *Neurocomputing*, 444, 319–331. <https://doi.org/10.1016/j.neucom.2021.03.024>
4. Lee, H., Kim, Y.-S., Kim, M., & Lee, Y. (2021). Low-cost network scheduling of 3D-CNN processing for embedded action recognition. *IEEE Access*, 9, 83901–83912. <https://doi.org/10.1109/ACCESS.2021.3087776>



5. Carreira, J., & Zisserman, A. (2017). Quo vadis, action recognition? A new model and the Kinetics dataset. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 6299–6308). <https://doi.org/10.1109/CVPR.2017.667>
6. Tseng, C.-H., & Lin, H.-Y. (2022). A vision-based system for abnormal behavior detection and recognition of bus passengers. In *2022 IEEE 25th International Conference on Intelligent Transportation Systems (ITSC)* (pp. 2134–2139). IEEE. <https://doi.org/10.1109/ITSC55140.2022.9922289>
7. Ramanujam, E., Perumal, T., & Padmavathi, S. (2021). Human activity recognition with smartphone and wearable sensors using deep learning techniques: A review. *IEEE Sensors Journal*, 21(12), 13029–13040. <https://doi.org/10.1109/JSEN.2020.3045991>
8. Ehatisham-Ul-Haq, M., et al. (2019). Robust human activity recognition using multimodal feature-level fusion. *IEEE Access*, 7, 60736–60751. <https://doi.org/10.1109/ACCESS.2019.2915552>
9. Popoola, O. P., & Wang, K. (2012). Video-based abnormal human behavior recognition—a review. *IEEE Transactions on Systems, Man, and Cybernetics, Part C*, 42(6), 865–878. <https://doi.org/10.1109/TSMCC.2011.2178594>
10. Helbing, D., & Molnar, P. (1995). Social force model for pedestrian dynamics. *Physical Review E*, 51(5), 4282–4286. <https://doi.org/10.1103/PhysRevE.51.4282>
11. Mehran, R., Oyama, A., & Shah, M. (2009). Abnormal crowd behavior detection using social force model. In *2009 IEEE Conference on Computer Vision and Pattern Recognition* (pp. 935–942). IEEE. <https://doi.org/10.1109/CVPR.2009.5206737>
12. Basharat, A., Gritai, A., & Shah, M. (2008). Learning object motion patterns for anomaly detection and improved object detection. In *2008 IEEE Conference on Computer Vision and Pattern Recognition* (pp. 1–8). IEEE. <https://doi.org/10.1109/CVPR.2008.4587478>
13. Delgado, B., Tahboub, K., & Delp, E. J. (2014). Automatic detection of abnormal human events on train platforms. In *NAECON 2014 - IEEE National Aerospace and Electronics Conference* (pp. 169–173). IEEE. <https://doi.org/10.1109/NAECON.2014.7000085>
14. Sun, X., Zhu, S., Wu, S., & Jing, X. (2018). Weak supervised learning-based abnormal behavior detection. In *2018 24th International Conference on Pattern Recognition (ICPR)* (pp. 1580–1585). IEEE. <https://doi.org/10.1109/ICPR.2018.8546099>
15. Nayak, R., Pati, U. C., & Das, S. K. (2021). A comprehensive review on deep learning-based methods for video anomaly detection. *Image and Vision Computing*, 106, 104078. <https://doi.org/10.1016/j.imavis.2020.104078>
16. Zhou, S., et al. (2016). Spatial-temporal convolutional neural networks for anomaly detection and localization in crowded scenes. *Signal Processing: Image Communication*, 47, 358–368. <https://doi.org/10.1016/j.image.2016.06.007>
17. Yan, C., et al. (2015). Recognizing driver inattention by convolutional neural networks. In *2015 8th International Congress on Image and Signal Processing (CISP)* (pp. 680–685). IEEE. <https://doi.org/10.1109/CISP.2015.7408012>
18. Tu, I., et al. (2017). Deep passenger state monitoring using viewpoint warping. In *Image Analysis and Processing – ICIAP 2017* (pp. 137–148). Springer. https://doi.org/10.1007/978-3-319-68560-1_13
19. Tu, I., et al. (2018). Dual viewpoint passenger state classification using 3D CNNs. In *2018 IEEE Intelligent Vehicles Symposium (IV)* (pp. 2163–2169). IEEE. <https://doi.org/10.1109/IVS.2018.8500517>
20. Velastin, S. A., & Gómez-Lira, D. A. (2017). People detection and pose classification inside a moving train using computer vision. In *Advances in Visual Informatics* (pp. 319–330). Springer. https://doi.org/10.1007/978-3-319-70010-6_29
21. Kao, S.-F., & Lin, H.-Y. (2021). Passenger detection, counting, and action recognition for self-driving public transport vehicles. In *2021 IEEE Intelligent Vehicles Symposium (IV)* (pp. 572–577). IEEE. <https://doi.org/10.1109/IV48863.2021.9575887>
22. Madapuri, R. K., & Senthil Mahesh, P. C. (2017). HBS-CRA: Scaling impact of change request towards fault proneness: Defining a heuristic and biases scale (HBS) of change request artifacts (CRA). *Cluster Computing*, 22(S5), 11591–11599. <https://doi.org/10.1007/s10586-017-1424-0>





23. Dwaram, J. R., & Madapuri, R. K. (2022). Crop yield forecasting by long short-term memory network with Adam optimizer and Huber loss function in Andhra Pradesh, India. *Concurrency and Computation: Practice and Experience*, 34(27), e7310. <https://doi.org/10.1002/cpe.7310>
24. Ahmed, S. T., Sivakami, R., Banik, D., Khan, S. B., Dhanaraj, R. K., Mahesh, T. R., & Almusharraf, A. (2024). Federated learning framework for consumer IoMT-edge resource recommendation under telemedicine services. *IEEE Transactions on Consumer Electronics*.
25. Fathima, A. S., Basha, S. M., Ahmed, S. T., Khan, S. B., Asiri, F., Basheer, S., & Shukla, M. (2025). Empowering consumer healthcare through sensor-rich devices using federated learning for secure resource recommendation. *IEEE Transactions on Consumer Electronics*.
26. Ahmed, S. T., Patil, K. K., Shanraj, R. K., Khan, S. B., Alzahrani, S., & Rani, S. (2024). 6GTelMED: Resources recommendation framework on 6G enabled distributed telemedicine using Edge-AI. *IEEE Transactions on Consumer Electronics*.
27. Ramaiah, N. S., & Ahmed, S. T. (2022). An IoT-based treatment optimization and priority assignment using machine learning. *ECS Transactions*, 107(1), 1487.
28. Pasha, A., Ahmed, S. T., Painam, R. K., Mathivanan, S. K., Karthikeyan, P., Mallik, S., & Qin, H. (2024). Leveraging ANFIS with Adam and PSO optimizers for Parkinson's disease. *Heliyon*, 10(9).

