# Real-Time Accident Risk Prediction via Stereo Vision and Driver Behavior Analysis

Isha Nitin Shetye, Anurag Rajesh Mahajan, and Alina Vereshchaka

Department of Computer Science and Engineering,
University at Buffalo, Buffalo, USA
{ishetye, anuragra, avereshc}@buffalo.edu

**Abstract.** Road accidents remain a major concern throughout the world, often caused by driver fatigue, distraction, and aggressive behavior. Existing vehicle safety systems focus on measures such as emergency braking and lane departure warnings but lack anticipate accidents before they occur. This research proposes a real-time accident risk prediction framework that integrates driver monitoring with external scene analysis. Our system includes driver behavior recognition, object detection, stereo vision-based depth estimation, and speed tracking using deep learning techniques to assess both internal and external risk factors. A final module computes a dynamic risk score and triggers early alerts when thresholds are exceeded. Additionally, the system generates a report to support emergency response and post-incident analysis. This research presents a novel integration of driver monitoring and road environment analysis, resulting in a more robust and proactive vehicle safety system. Check our website for more details: https://sites.google.com/view/real-time-accident-prediction/.

**Keywords:** Accident Prediction · Deep Learning · Driver Behavior Analysis · Stereo Vision · Risk Assessment · Intelligent Safety Systems.

## 1 Introduction

Road accidents remain one of the leading causes of injury and death worldwide, accounting for over 1.3 million fatalities each year, according to the World Health Organization [20]. Traditional Advanced Driver Assistance Systems (ADAS) such as lane keeping, adaptive cruise control, emergency braking, and antilock braking system have significantly improved road safety. However, these systems are limited by their design as they activate only after a potentially dangerous event. For example, automatic braking may activate only moments before impact, providing little time for accident prevention.

One of the limitation of many existing systems is the lack of integration with human factors. Studies have shown that driver distraction, drowsiness, and aggressive or inattentive behavior are major contributors to traffic accidents [6,9]. However, most ADAS frameworks rely mostly on external factors. These systems also struggle with dynamic road conditions, including poor weather, changing traffic density and unpredictable behaviour of nearby vehicles.

These challenges highlight the need for a proactive intelligent safety system that combines road situations with driver behavior analysis to assess and predict risk in real-time before an accident happen.

To address this gap, we present *Be Safe on the Road* a deep learning framework for real-time accident risk prediction. Our system integrates data from stereo vision, driver activity recognition, speed estimation, and depth inference to assess dynamic accident risk. The main contributions of our approach are:

- Robust depth and motion estimation. We use stereo cameras and SGBM disparity maps to estimate object depth, refined through bounding box cropping. Object motion is estimated by tracking bounding box centroids across frames and converting pixel displacement to real-world speed.
- We add a human factor into the risk assessment pipeline by integrating a CNN-based classifier that classifies driver behavior based on the in-cabin video stream (e.g., texting, drinking, talking).
- We introduce an LSTM-based module that is based on the spatial (depth), temporal (speed), and behavioral (driver state) modules to compute a risk score. If the predicted risk score exceeds the threshold, the system generates a report with the vehicle's location and details about the potential accident, which is sent to both the driver and external systems.

Our work contributes towards building an intelligent copilot system, that improves driver awareness, reducing human error, and smarter safer transportation systems.

## 2 Related Work

Accident risk prediction has received attention in recent years, with research exploring vision-based speed estimation, driver behavior modeling, and temporal risk forecasting. Traditional ADAS systems often rely on external sensors such as RADAR or LIDAR; however, recent work uses monocular and stereo vision for more accessible, cost-efficient alternatives.

**Vision-Based Speed and Risk Estimation.** Camera-based methods estimate vehicle speed using frame-to-frame displacement and calibrated optics [4, 12]. Optical flow and object tracking algorithms (e.g., KLT, SORT) have been applied in monocular settings [19], while stereo vision enables accurate depth perception for 3D reconstruction and motion estimation [13]. Temporal models such as LSTM, ConvLSTM, and attention-based networks improve accident prediction by learning from sequential traffic data [2, 10, 16].

**Driver Behavior and Risk Modeling.** Research has shown that integrating driver behavior into predictive models significantly improves accident risk forecasting. CNN and LSTM models have been used to classify in-cabin driver behaviors [5], while probabilistic and attention-based approaches integrate environmental and behavioral features to predict potential risks earlier [7].

**Depth Estimation and Multi-View Perception.** Stereo disparity methods (e.g., SGBM) and monocular depth estimation with CNNs have shown reliable performance in real-time setup [8,11]. Efficient stereo pipelines for embedded deployment are being actively explored in different research papers [18].

**System-Level Integrations.** Recent frameworks combine YOLO-based detection with IoT and blockchain for secure, real-time traffic monitoring [3, 17]. Other pipelines use license plate tracking or in-vehicle detection for accurate speed and event estimation [1, 3]. Frameworks like SAARTHI automate emergency response using visual inspection and geolocation [15], while stereo-based ADAS still face limitations in long-range fusion performance [12].

## 3 Proposed Framework

We present *Be Safe on the Road*, an end-to-end deep learning framework for real-time accident risk prediction. Our system combines object detection, stereo depth estimation, speed estimation, driver behavior classification, and temporal accident risk prediction to assess accident risk in complex traffic environments. As shown in Fig. 1, the pipeline includes four core modules. Each module is described in details below.

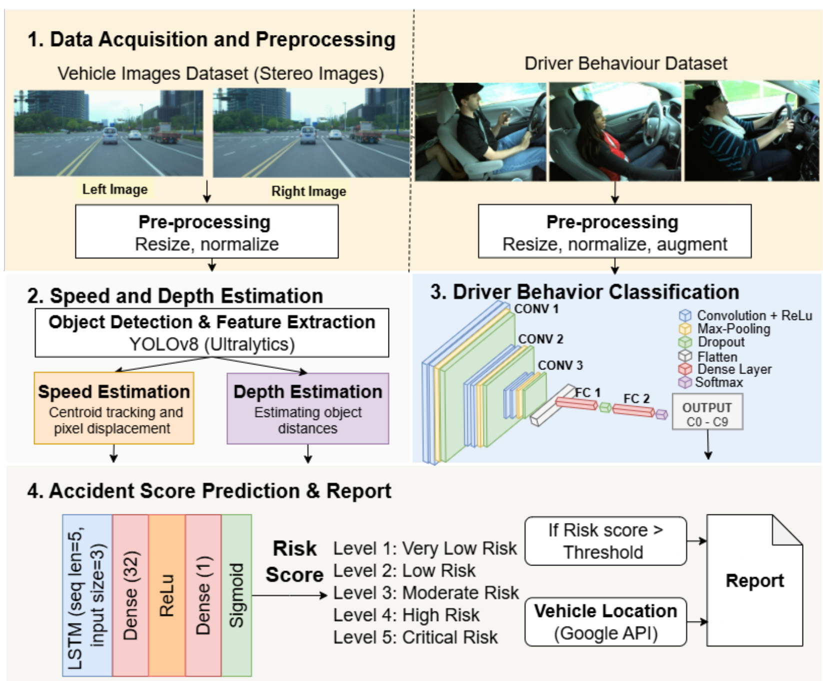


Fig. 1: Architecture of the proposed accident risk prediction framework. The pipeline includes: (1) data acquisition and preprocessing, (2) stereo vision for depth and speed estimation using centroid tracking, (3) CNN-based driver behavior classification, (4) LSTM-based accident risk scoring and report generator.

### 3.1 Data Acquisition and Preprocessing

We use two publicly available datasets: DrivingStereo for road scene analysis, and the State Farm Distracted Driver Detection dataset for behavior classification.

- DrivingStereo: Contains 170,000 stereo image pairs from synchronized left/right cameras. We use a subset of 2,000 pairs for depth/speed estimation.
- State Farm Dataset: Includes 22,424 labeled in-cabin images with 10 classes (e.g., safe driving, texting, drinking).

All images are resized to $64 \times 64$ for efficiency. Stereo images are normalized and converted to grayscale for disparity calculation. We apply data augmentation (rotation, flipping, shifting) for driver images to improve generalization.

### 3.2 Depth and Speed Estimation

**Depth Estimation** We use stereo vision and YOLOv8 for detecting objects and estimating depth via the Semi-Global Block Matching (SGBM) algorithm. Detected bounding boxes from the left stereo frame are used to extract regions for disparity computation. The object-level depth is computed as $Z(u, v) = \frac{B \cdot f}{d(u,v)}$, where $f$ is the focal length (in pixels), $B$ is the baseline between cameras, and $d(u, v)$ is the disparity at pixel coordinates $(u, v)$.

The complete depth estimation methodology is illustrated in Fig. 2.

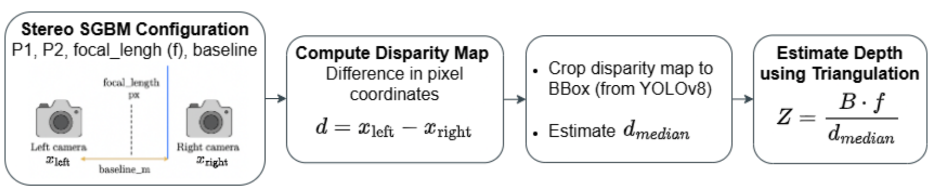


Fig. 2: Stereo images are used to calculate disparity, with detected objects (using YOLOv8) and their distance is estimated using Stereo SGBM - Triangulation method.

**Speed Estimation** For the motion-based risk indicators (e.g., abrupt halts, fast approach), we estimate the surrounding vehicles' speed using frame-to-frame centroid displacement. This part is based on the object bounding boxes produced by YOLOv8. For each detected vehicle in a frame, we compute the centroid of its bounding box to track its position across time. The bounding box coordinates are represented as $(x_1, y_1)$ for the top-left corner and $(x_2, y_2)$ for the bottom-right corner. The centroid $(c_x, c_y)$ of the bounding box is computed as $c_x = \frac{x_1+x_2}{2}$ and $c_y = \frac{y_1+y_2}{2}$. To estimate the object's movement between two frames, we calculate the Euclidean distance between centroids as $\Delta s_{\text{pixels}} = \sqrt{(c_x(t) - c_x(t-1))^2 + (c_y(t) - c_y(t-1))^2}$. Here, $(c_x(t), c_y(t))$ is the centroid in the current frame $t$, and $(c_x(t-1), c_y(t-1))$ is the centroid in the previous frame $t-1$.

To convert this pixel displacement into a real-world speed estimate, we use:

$$\text{Speed}_{\text{mph}} = 2.24 \cdot \left( \frac{\Delta s_{\text{pixels}}}{\text{pixels\_per\_meter}} \cdot \text{fps} \right) \tag{1}$$

, where pixels_per_meter = 30 (scale factor), fps = 5 (frame rate), and 2.24 converts speed from m/s to mph.

If a vehicle slows down from $> 35$ mph to $< 3mph$ within 1-2 frames, a possible accident is detected and passed to the risk assessment module. This speed tracking helps detect abrupt changes that could indicate a collision or hazard. The complete methodology for speed estimation is shown in Fig. 3.

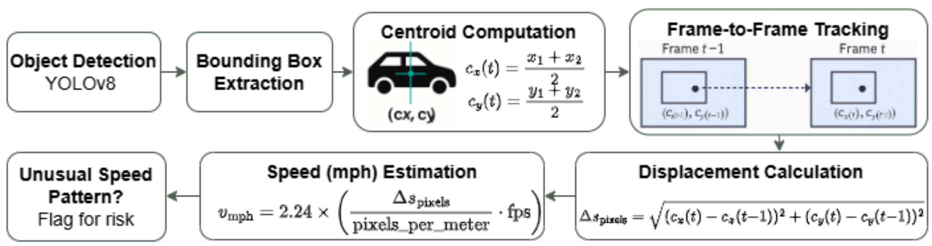


Fig. 3: Objects are detected and tracked frame-by-frame using YOLOv8. The speed is calculated from the movement of bounding box centroids, and unusual speed patterns are flagged as potential risks.

### 3.3 Driver Behavior Analysis

We use a CNN trained on the State Farm dataset to classify driver behavior into 10 categories. The architecture includes three convolutional layers followed by fully connected layers. It captures spatial features and performs well in recognizing behaviors such as "safe driving," "drinking," and "texting" (Fig. 4).

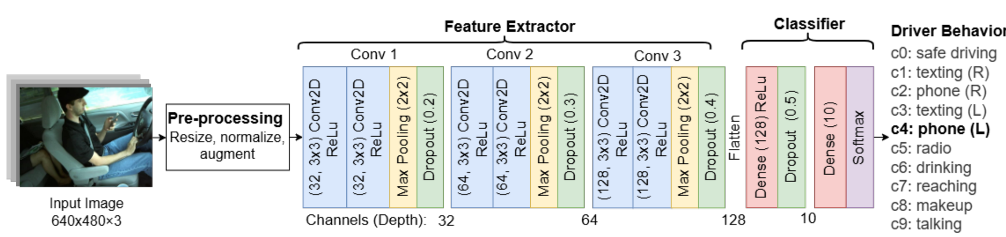


Fig. 4: CNN architecture for driver behavior classification. The pipeline includes preprocessing, convolutional layers with ReLU, pooling, dropout, and fully connected layers with softmax for classifying 10 behavior types.

### 3.4 Accident Risk Prediction & Report Generation

We use an LSTM model to capture temporal patterns in driver behavior, speed, and proximity. Input sequences consist of five-frame windows grouped by vehicle ID, enabling risk prediction based on recent motion and behavior history. The model outputs a normalized risk score in [0, 1], which is mapped to five risk levels. If the score exceeds a threshold, a report is generated with vehicle location and event summary. See Fig. 5 for details.

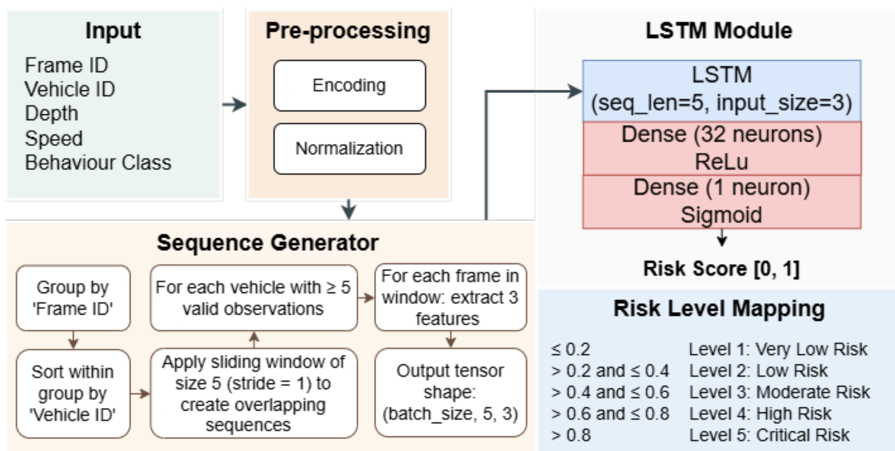


Fig. 5: LSTM-based accident risk prediction pipeline. Input features are grouped by Frame ID and Vehicle ID, then processed into 5-step sequences using a sliding window. The LSTM module outputs a risk score in [0, 1], which is mapped to five discrete risk levels from Very Low to Critical

## 4 Experimental Results

We evaluate our real-time accident risk prediction framework across five key components: object detection, depth estimation, speed estimation, driver behavior classification and temporal risk prediction. Experiments were conducted using the subsets of the DrivingStereo and State Farm datasets, descrived in Section 3.1

### 4.1 Object Detection

We compared YOLOv8 and Faster R-CNN [14] for vehicle detection on the stereo image dataset used for depth and speed estimation. YOLOv8 achieved an average inference time of 0.015 seconds per image-over 18 times faster than Faster R-CNN making it more suitable for real-time use (Fig. 6) Although Faster R-CNN showed slightly better accuracy in some cases, YOLOv8 provided clearer and more consistent detections with a lower latency and a smaller model size.

We selected YOLOv8 for its speed and practicality. Detection was applied to both stereo frames, enabling integration with downstream depth estimation. Fig. 7 illustrates sample vehicle detections.

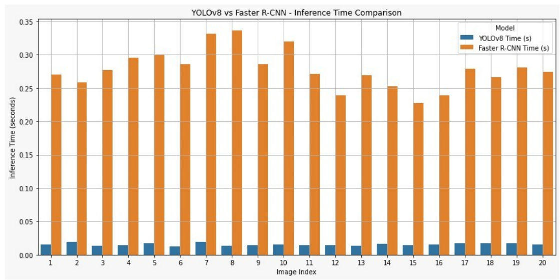


Fig. 6: The bar chart shows that YOLOv8 achieves lower inference times (∼0.015s) across all images compared to Faster R-CNN (∼0.25-0.34s).

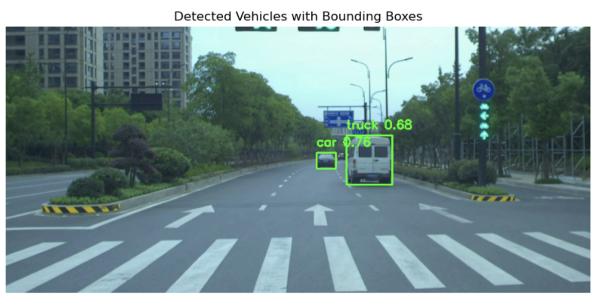


Fig. 7: Detected vehicles using YOLOv8 with confidence scores of 0.76 and 0.68 for car and truck classes, respectively.

### 4.2 Depth & Speed Estimation

**Depth Estimation.** We benchmarked four methods for depth estimation: Stereo SGBM, MiDaS (Monocular Transformer-based), SIFT + Triangulation, and Stereo Block Matching. Table 1 summarizes their performance. MiDaS offered deployment ease but lacked geometric reliability; SIFT + Triangulation was computationally heavy. Stereo SGBM provided the best balance of accuracy, runtime, and geometric consistency, making it suitable for real-time applications. Fig. 8 shows result of depth calculation for a frame generated from the dataset.

**Speed Estimation** We experimented with four speed estimation methods: Centroid Tracking, Vision-Based Estimation, Optical Flow and Background Subtraction. Centroid Tracking provided consistent results in clear visibility but was sensitive to occlusion. Vision-Based Estimation used object scale and frame rate but required calibration and consistent viewpoints. Optical Flow captured motion dynamics well but was affected by camera and background movement.

Table 1: Comparison of Depth Estimation Methods

| Method | Abs Rel ↓ | RMSE ↓ | $\delta < 1.25$ ↑ |
|---|---|---|---|
| Stereo SGBM | **0.11** | **3.4** | **85.2%** |
| MiDaS Monocular | 0.16 | 5.1 | 71.3% |
| SIFT + Triangulation | 0.19 | 6.7 | 68.5% |
| StereoBM | 0.22 | 7.5 | 62.1% |

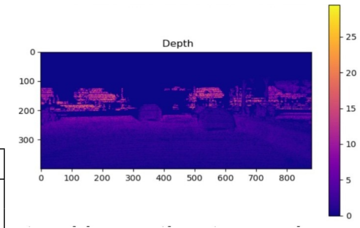


Fig. 8: Stereo SGBM disparity map sample: two vehicles detected at 96.35 ft and 128.01 ft.

Background Subtraction worked efficiently in static scenarios but failed in dynamic environments. After comparing performance across accuracy and computational cost, we selected Centroid Tracking as the preferred method. Fig. 9 shows result of speed estimation on one of the frames.


Fig. 9: Speed estimated as 40.1 mph and 30.3 mph for the two detected cars with confidence score as 0.76 and 0.72 respectively

### 4.3 Driver Behavior Classification

Multiple architectures were evaluated for driver behavior classification. The CNN-based model demonstrated the best performance for real-time deployment. Table 2 summarizes the comparative results across accuracy, precision, recall, and F1-score.

Table 2: Comparison of Driver Behavior Analysis Methods

| Method | Accuracy | Precision | Recall | F1-Score |
|---|---|---|---|---|
| CNN-based Classification | **92.3%** | 91.7% | **92.0%** | **91.8%** |
| LSTM-based Temporal Modeling | 86.5% | 85.0% | 86.2% | 85.4% |
| Transformer + Pose Estimation | 90.8% | **92.1%** | 88.9% | 90.4% |
| Transformer over Temporal Features | 91.5% | 90.3% | 91.6% | 90.9% |

### 4.4 Risk Prediction and System Integration

Our LSTM-based module accurately predicted high-risk scenarios involving sudden braking, close following, and distracted driving. To demonstrate full system integration, we developed an interactive Streamlit dashboard that simulates the entire pipeline. More details are available on our project page:
https://sites.google.com/view/real-time-accident-prediction/.

## 5 Conclusion

Our proposed *Be Safe on the Road* framework improves traditional vehicle safety systems by introducing a deep learning system that predicts accident risk in advance. By combining stereo-based depth estimation and speed estimation, along with driver behavior analysis, the system provides a better risk assessment pipeline. This enables early warnings about possible risks and provides timely alerts to prevent accidents before they occur.

While the current system is vision-based, future work will explore incorporating additional sensor modalities such as LiDAR, RaDaR, and biometric inputs to improve accuracy and adaptability across diverse driving conditions.

## References

1. Ahmad, F., Ansari, M.Z., Hamid, S., Saad, M.: A computer vision based vehicle counting and speed detection system. In: 2023 International Conference on Recent Advances in Electrical, Electronics & Digital Healthcare Technologies (REEDCON). pp. 487–492 (2023). https://doi.org/10.1109/REEDCON57544.2023.10151423
2. Baek, J.W., Chung, K.H.: Accident risk prediction model based on attention-mechanism lstm using modality convergence in multimodal. Personal and Ubiquitous Computing **27**, 1179–1189 (2023). https://doi.org/10.1007/s00779-021-01552-1
3. Beric, D., Havzi, S., Lolic, T., Simeunovic, N., Stefanovic, D.: Development of the mes software and integration with an existing erp software in industrial enterprise. In: 2020 19th International Symposium INFOTEH-JAHORINA (INFOTEH). pp. 1–6 (2020). https://doi.org/10.1109/INFOTEH48170.2020.9066345
4. Bhatlawande, S., Shilaskar, S., Mahajan, A., Shete, M.: Vision-based vehicle speed estimation. In: 2024 International Conference on Distributed Computing and Optimization Techniques (ICDCOT). pp. 1–5 (2024). https://doi.org/10.1109/ICDCOT61034.2024.10516237
5. Brühwiler, L., Fu, C., Huang, H., Longhi, L., Weibel, R.: Predicting individuals' car accident risk by trajectory, driving events, and geographical context. Computers, Environment and Urban Systems **93**, 101760 (2022). https://doi.org/10.1016/j.compenvurbsys.2022.101760
6. Guo, F., Klauer, S.G., Fang, Y., Hankey, J.M., Antin, J.F., Perez, M.A., Lee, S.E., Dingus, T.A.: The effects of age on crash risk associated with driver distraction. International journal of epidemiology **46**(1), 258–265 (2017). https://doi.org/10.1093/ije/dyw234
7. Hu, W., Xiao, X., Xie, D., Tan, T., Maybank, S.: Traffic accident prediction using 3-d model-based vehicle tracking. IEEE Transactions on Vehicular Technology **53**(3), 677–694 (2004). https://doi.org/10.1109/TVT.2004.825772

8. Kaneko, A.M., Yamamoto, K.: Two-view monocular depth estimation by optic-flow-weighted fusion. IEEE Robotics and Automation Letters **4**(2), 830–837 (2019). https://doi.org/10.1109/LRA.2019.2893426
9. Klauer, S.G., Guo, F., Simons-Morton, B.G., Ouimet, M.C., Lee, S.E., Dingus, T.A.: Distracted driving and risk of road crashes among novice and experienced drivers. New England journal of medicine **370**(1), 54–59 (2014)
10. Li, P., Abdel-Aty, M., Yuan, J.: Real-time crash risk prediction on arterials based on lstm-cnn. Accident Analysis & Prevention **135**, 105371 (2020). https://doi.org/10.1016/j.aap.2019.105371
11. Liu, F., Shen, C., Lin, G., Reid, I.: Learning depth from single monocular images using deep convolutional neural fields. IEEE Transactions on Pattern Analysis and Machine Intelligence **38**(10), 2024–2039 (2016). https://doi.org/10.1109/TPAMI.2015.2505283
12. Luvizon, D.C., Nassu, B.T., Minetto, R.: A video-based system for vehicle speed measurement in urban roadways. IEEE Transactions on Intelligent Transportation Systems **18**(6), 1393–1404 (2017). https://doi.org/10.1109/TITS.2016.2606369
13. Miljković, M., Vranješ, M., Mijić, D., Cakić, M.: Vehicle distance estimation based on stereo camera system with implementation on a real adas board. In: 2022 International Conference on Software, Telecommunications and Computer Networks (SoftCOM). pp. 1–6 (2022). https://doi.org/10.23919/SoftCOM55329.2022.9911360
14. Ren, S., He, K., Girshick, R., Sun, J.: Faster r-cnn: Towards real-time object detection with region proposal networks. Advances in neural information processing systems **28** (2015)
15. Sanghvi, C., Vereshchaka, A.: Saarthi: Smart auto assessment and roadside technical help interface. In: Social, Cultural, and Behavioral Modeling. SBP-BRiMS 2024, Lecture Notes in Computer Science, vol. 14972. Springer, Cham (2024). https://doi.org/10.1007/978-3-031-72241-7_4
16. Siswanto, J., Hendry, Rahardja, U., Sembiring, I., Hartomo, K.D., Purnomo, H.D., Iriani, A.: Number of road accidents predicting using deep learning-based lstm development models. In: 2023 11th International Conference on Cyber and IT Service Management (CITSM). pp. 1–6 (2023). https://doi.org/10.1109/CITSM60085.2023.10455169
17. Soma, K., Shibu, L., N, M.: A real-time vehicle detection and speed estimation using yolo v8. In: 2024 International Conference on Advances in Data Engineering and Intelligent Computing Systems (ADICS). pp. 1–6 (2024). https://doi.org/10.1109/ADICS58448.2024.10533551
18. Su, H., Zhu, J., Yin, Z., Dong, Y., Zhang, B.: Efficient and robust semi-supervised learning over sparse-regularized graph. In: Leibe, B., Matas, J., Sebe, N., Welling, M. (eds.) Computer Vision – ECCV 2016, Lecture Notes in Computer Science, vol. 9912, pp. 339–354. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-46484-8_35
19. Thakurdesai, H.M., Aghav, J.V.: Computer vision based position and speed estimation for accident avoidance in driverless cars. In: Tuba, M., Akashe, S., Joshi, A. (eds.) ICT Systems and Sustainability, Advances in Intelligent Systems and Computing, vol. 1077, pp. 539–547. Springer, Singapore (2020). https://doi.org/10.1007/978-981-15-0936-0_47
20. World Health Organization: Global status report on road safety 2023 (2023), https://www.who.int/publications/i/item/9789240086517, accessed: 2025-06-29