A Synergistic Approach to Real-Time Crash Risk Estimation at Signalized Intersections: Integrating Learning-Based Anomaly Detection with Extreme Value Theory

A. Yukai Wang[a], B. Tiantian Chen[a,*] and C. Kitae Jang[a]

[a] Cho Chun Shik Graduate School of Mobility, Korea Advanced Institute of Science and Technology, Daejeon, Korea
a.yukai@kaist.ac.kr, b.nicole.chen@kaist.ac.kr, c.kitae.jang@kaist.ac.kr
[*] Corresponding author

Extended abstract submitted for presentation at the Conference in Emerging Technologies in Transportation Systems (TRC-30)
September 02-03, 2024, Crete, Greece

April 28, 2024

Keywords: Crash risk estimation; Learning-based anomaly detection; Extreme value theory; Signalized intersections

1 INTRODUCTION

Real-time crash risk estimation at urban intersections is hampered by traditional data sources like loop detectors, which inadequately reflect traffic dynamics and lead to outcomes challenging to apply in real-time scenarios (Hossain *et al.*, 2019). Extreme Value Theory (EVT), particularly with its focus on traffic conflicts as precursors to accidents, offers a potent alternative (Arun *et al.*, 2021). This approach has been advanced by Zheng & Sayed (2020), who successfully integrated EVT with vehicle trajectory data to calculate safety indices, improving the prediction and management of crash risks.

Traditional EVT methods—Block Maxima (BM) and Peak Over Threshold (POT)—although foundational, have limitations in selecting relevant extremes (Ali *et al.*, 2023). Overcoming these, recent hybrid models blend Machine Learning (ML) and Deep Learning (DL) with EVT, significantly surpassing conventional models in precision, as demonstrated by Hussain *et al.* (2022). Such integration signifies a substantial leap in deploying ML-enhanced EVT for efficacious, real-time traffic safety analysis.

Building upon these advancements, this study develops a nuanced methodology for real-time crash risk estimation at signalized intersections, integrating cutting-edge video data processing and advanced analytical techniques. Illustrated in Figure 1, the approach starts with the sophisticated processing of video data to detect and track vehicle movements and calibrate camera feeds. The module 1 ensures the accurate capture of traffic dynamics, essential for effective application of Extreme Value Theory (EVT). Advancing to module 2, the approach employs both traditional and learning-based algorithms for anomaly detection and scoring, aimed at identifying extreme traffic conditions that signify precursors to potential accidents. To further enhance prediction accuracy, the module 3 incorporates Kernel Density Estimation for detailed probability analysis and validates the predictions against historical crash data, while also calculating safety indices to quantify risk levels. This comprehensive approach not only provides actionable insights but also facilitates proactive interventions to enhance real-time traffic safety at signalized intersections.

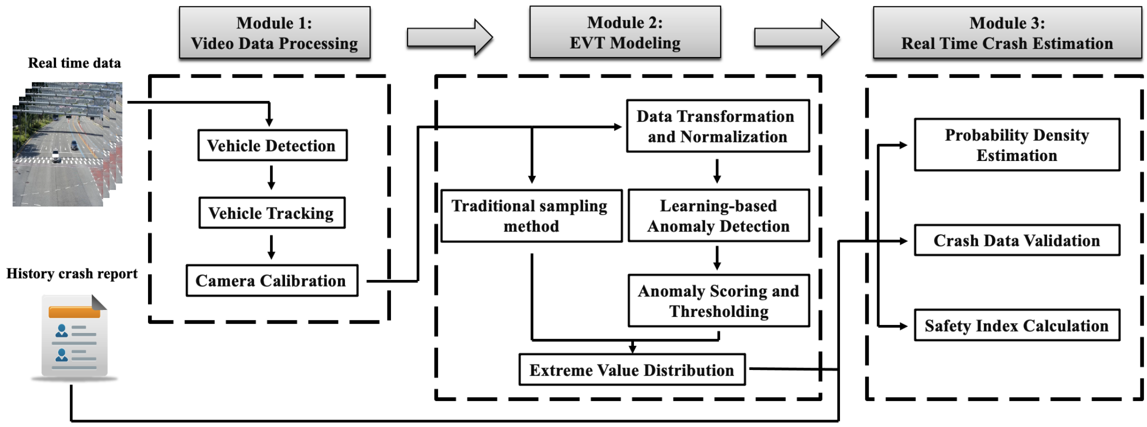


Figure 1 – *A schematic of the proposed framework*

2 METHODOLOGY

2.1 Traffic conflict indicators

In traffic safety analysis at signalized intersections, key metrics such as Time-to-Collision (TTC), Modified Time-to-Collision (MTTC), and Deceleration Rate to Avoid Collision (DRAC) are essential for evaluating collision risks. These indicators collectively enhance our understanding of the dynamic risks prevalent in signalized intersections. The mathematical formulations of these indicators are provided below:

$$\begin{cases} TTC & = d/v_{rel} \\ MTTC & = \Delta(s + \sqrt{\Delta s^2 + 2\Delta a \cdot d})/\Delta a \\ DRAC & = v_{rel}^2/2d \end{cases} \tag{1}$$

Here, d represents the distance between vehicles, v_{rel} is the relative velocity, Δs signifies speed difference, and Δa denotes acceleration difference. These formulas allow for precise risk assessments in complex traffic environments.

2.2 Learning-Based Anomaly Detection

In addressing the limitations of traditional EVT methods like BM and POT for traffic data, this study integrates advanced ML and DL for improved anomaly detection. ML algorithms, particularly Isolation Forest and Minimum Covariance Determinant, adeptly identify outliers signaling potential crash risks. One-Class SVM complements these by effectively distinguishing anomalies from typical traffic behavior.

Further enhancing EVT's robustness, DL methods like Autoencoders and Long Short-Term Memory (LSTM) networks unveil subtle patterns and temporal trends in traffic data. These techniques, by capturing deeper data complexities, offer refined insights into crash risk estimation, presenting a significant advancement over conventional EVT approaches.

2.3 Extreme Value Theory

Effective EVT modeling hinges on the precise identification of outliers within traffic conflict data, a task enhanced by learning-based anomaly detection methods that discern critical events from normal traffic fluctuations. This study's methodology tailors extreme values to fit two key distributions: the Generalized Extreme Value (GEV) for BM samples, and the Generalized Pareto Distribution (GPD) for POT data, ensuring an accurate representation of data extremities.

GEV distribution is defined as:

$$G(z; \mu, \sigma, \xi) = \exp\left[-\left(1 + \xi\left(\frac{z-\mu}{\sigma}\right)\right)^{-1/\xi}\right] \tag{2}$$

Here, z represents the variable of interest, μ denotes the location parameter, σ (must be positive) is the scale parameter, and ξ is the shape parameter. This formulation applies when $1 + \xi(z - \mu)/\sigma > 0$, ensuring the expression inside the exponential is positive and meaningful.

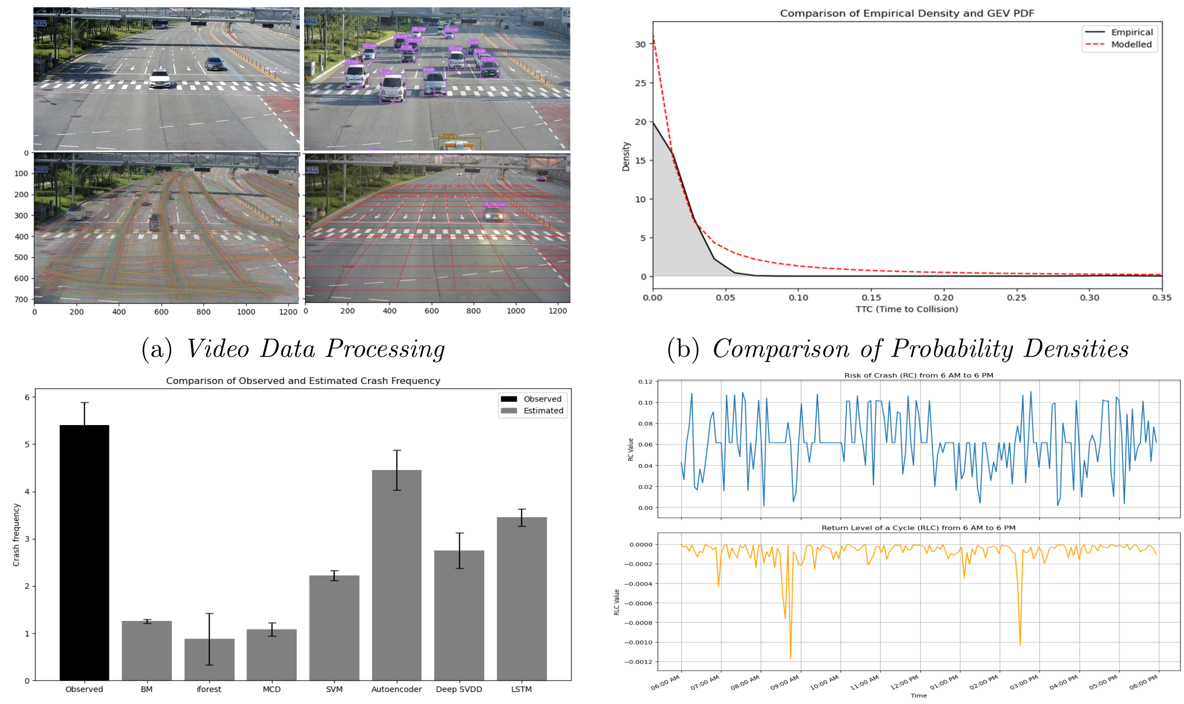
For POT method, which focuses on data exceeding a predefined threshold, GPD is employed:

$$H(z; \sigma, \xi) = 1 - \left(1 + \frac{\xi z}{\sigma}\right)^{-1/\xi} \tag{3}$$

The condition $1 + \xi z/\sigma > 0$ is necessary to ensure the terms within the parentheses remain positive for valid calculations.

3 RESULTS

The dataset for this study comprises video recordings from traffic cameras at three signalized intersections in Daejeon, South Korea: Jinteo, Wongol, and Yuseong.The videos were captured from 6 AM to 6 PM on weekdays, encompassing both peak and non-peak traffic periods, thereby providing a comprehensive view of typical traffic flows and behaviors. Additionally, historical accident data spanning from 2017 to 2019 was utilized, with each incident precisely geolocated to match the corresponding intersections. Figure 2 depicts a suite of visualizations representing partial results from the results analysis.



(a) *Video Data Processing*

(b) *Comparison of Probability Densities*

(c) *Comparison of estimated and observed crashes*

(d) *Temporal Variations in Safety Indices*

Figure 2 – *Visualizations of Partial Results*

The overall process of video data processing is illustrated in Figure 2a. To ensure real-time estimation capabilities, lightweight object detection algorithms such as YOLOv8 were employed for vehicle recognition, paired with DEEPSORT for vehicle tracking. This combination yielded an average processing time of 2.3 milliseconds per frame, enabling the system to handle real-time analysis of video data at 30 frames per second. Such processing speed is critical for the timely evaluation and response to dynamic traffic conditions at signalized intersections.

Leveraging the swift data processing framework, the subsequent analysis involved computing the parameters for the GEV and GPD models from the extreme values detected in the traffic flow. The alignment of these statistical models with the empirical data underscores their efficacy in capturing the extremities of traffic conflict indicators, including TTC as depicted in Figure 2b.

Figure 2c conveys the performance of diverse anomaly detection techniques in estimating observed crash frequencies, within GEV distribution fitting. The graphical representation indicates that deep learning methodologies, specifically Autoencoder and LSTM networks, correlate more closely with actual crash frequencies. This concordance suggests that deep learning techniques might offer superior predictive performance by effectively deciphering the intricate and temporal patterns inherent in traffic data.

Figure 2d provides a comprehensive analysis, displaying the Risk of Crash (RC) and Return Level of a Cycle (RLC) from 6 AM to 6 PM. The upper plot shows RC values, with a consistent pattern of fluctuations throughout the day. This shows varying crash risks, potentially aligned with typical traffic patterns. In the lower plot, the RLC values remain stable with a few noticeable spikes. These significant peaks occur around 9 AM and 3 PM, suggesting periods of heightened crash risk, potentially due to shifts in traffic flow or other contributing factors. These insights offer a detailed time-based analysis, allowing us to understand when and how crash risks vary, contributing to more effective safety planning and traffic management.

4 DISCUSSION

This study demonstrates the application of advanced analytic techniques to improve traffic safety, emphasizing real-time processing of video data with learning-enhanced anomaly detection to refine EVT model accuracy. The calculation of RC and RLC indices offers quantitative insight into the temporal dynamics of crash risks at urban intersections. These findings underscore the potential for intelligent traffic systems to proactively manage and mitigate safety hazards, setting a precedent for future traffic safety interventions that are both responsive and data-driven.Future studies could focus on integrating broader datasets, encompassing varied traffic conditions and demographic factors, to further generalize the findings. Additionally, exploring the incorporation of vehicle-to-infrastructure communication could enhance predictive capabilities, paving the way for even more proactive and adaptive traffic safety systems.

References

Ali, Yasir, Haque, Md Mazharul, & Mannering, Fred. 2023. Assessing traffic conflict/crash relationships with extreme value theory: Recent developments and future directions for connected and autonomous vehicle and highway safety research. *Analytic methods in accident research*, 100276.

Arun, Ashutosh, Haque, Md Mazharul, Bhaskar, Ashish, Washington, Simon, & Sayed, Tarek. 2021. A bivariate extreme value model for estimating crash frequency by severity using traffic conflicts. *Analytic methods in accident research*, **32**, 100180.

Hossain, Moinul, Abdel-Aty, Mohamed, Quddus, Mohammed A, Muromachi, Yasunori, & Sadeek, Soumik Nafis. 2019. Real-time crash prediction models: State-of-the-art, design pathways and ubiquitous requirements. *Accident Analysis & Prevention*, **124**, 66–84.

Hussain, Fizza, Li, Yuefeng, Arun, Ashutosh, & Haque, Md Mazharul. 2022. A hybrid modelling framework of machine learning and extreme value theory for crash risk estimation using traffic conflicts. *Analytic methods in accident research*, **36**, 100248.

Zheng, Lai, & Sayed, Tarek. 2020. A novel approach for real time crash prediction at signalized intersections. *Transportation research part C: emerging technologies*, **117**, 102683.