Predictive Modeling of Accident-Prone Road Zones and Action Recognition in Unstructured Traffic Scenarios using ADAS Systems at Population Scale

Thesis submitted in partial fulfillment
of the requirements for the degree of

Masters of Science
in
Computer Science and Engineering by Research

by

Ravi Shankar Mishra
2021701044
ravi.mishra@research.iiit.ac.in

International Institute of Information Technology, Hyderabad
(Deemed to be University)
Hyderabad - 500 032, INDIA
April, 2025

International Institute of Information Technology
Hyderabad, India

CERTIFICATE

It is certified that the work contained in this thesis, titled 'Predictive Modeling of Accident-Prone Road Zones and Action Recognition in Unstructured Traffic Scenarios using ADAS Systems at Population Scale' by Ravi Shankar Mishra, has been carried out under my supervision and is not submitted elsewhere for a degree.

_____ _____

Date Adviser: Prof. Ravi Kiran Sarvadevabhatla

_____ _____

Date Adviser: Prof. C V Jawahar

To, my family and friends

Acknowledgments

To begin with, I would like to extend my heartfelt gratitude to my advisors, Dr. Ravi Kiran S and Prof. C.V Jawahar, for his unwavering guidance and support throughout this incredible journey. His insightful advice has not only helped me become a better researcher but has also played a significant role in shaping my moral and ethical character. Dr. Ravi Kiran's meticulous attention to the finer details of my presentations has significantly enhanced my presentation skills, making me more confident and articulate in sharing my work.

Additionally, I would like to express my sincere appreciation to Prof. C.V Jawahar for their invaluable guidance and assistance throughout my research journey. The discussions we had about the project were intellectually stimulating and deeply enjoyable. Furthermore, our conversations about our future plans provided reassurance and a sense of control over our respective paths, fostering a collaborative and supportive environment.

I am also profoundly grateful to my friends Pranav Gupta, Zeeshan Khan, Madhav Agarwal, Varun Gupta and Seshadari Mazumdar. They have made my college life incredibly enjoyable, offering the much-needed balance between work and leisure. Our study sessions together were instrumental in helping me manage my coursework and research effectively. Their impact on my cultural tastes has been immense, and I feel extremely fortunate to have shared wonderful memories with such amazing friends over the years.

Finally, I would like to extend my deepest thanks to my family for their unwavering support and motivation. Their trust in me has allowed me to focus on my research without any worries, and their encouragement during times of distress and failure has been a crucial factor in helping me persevere and succeed.

This thesis would not have been possible without the support of many incredible people. I dedicate this work to all of them and sincerely hope to have their continued support in my future endeavors as well.

Abstract

This thesis addresses the critical challenge of improving road safety by introducing novel approaches to predictive modeling of accident-prone zones and action recognition in critical traffic scenarios. It makes two key contributions: the early identification of accident-prone zones using Advance Driving Assistance System (ADAS) data and the development of IDD-CRS, a comprehensive dataset for action recognition in unstructured road environments.

In the first study, geo-tagged collision alert data from a fleet of 200 ADAS-equipped city buses in Nagpur, India, is leveraged to proactively identify high-risk zones across urban road networks. Using Kernel Density Estimation (KDE), this study captures the spatiotemporal distribution of collision alerts, enabling the detection of emerging blackspots before accidents occur. A novel recall-based metric evaluates the alignment of these predicted zones with historical blackspots, while Earth Mover Distance (EMD)-based analysis identifies previously unreported accident-prone areas. This predictive framework provides civic authorities with actionable insights for targeted interventions, such as traffic-calming measures and infrastructure improvements, thereby enhancing public safety.

The second part of the thesis introduces the IDD-CRS dataset, a large-scale collection of traffic scenarios recorded using ADAS and dash cameras. IDD-CRS fills a critical gap in existing datasets by focusing on complex interactions between vehicles and pedestrians, with scenarios such as high-speed lane changes, unsafe vehicle approaches, and near-miss incidents. With precise temporal annotations powered by ADAS technology, the dataset ensures accurate event boundaries, providing a robust benchmark for action recognition and long-tail action recognition tasks. It includes 90 hours of footage spanning 5,400 one-minute videos and 135,000 frames, with hard negative examples to challenge existing models. Initial benchmarks highlight the limitations of current video backbones in recognizing rare events, emphasizing the need for further advancements.

Together, these contributions provide a holistic framework for improving road safety through proactive accident prevention and robust action recognition in traffic scenarios. By addressing both spatial accident prediction and temporal event recognition, this work offers foundational resources and actionable insights to advance research and practical solutions for safer road environments.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Motivation

Improving road safety has long been an intriguing and important challenge for researchers. Roads can be made safer through two primary approaches: reducing accidents and preventing them. Reducing accidents involves identifying potential causes and addressing them proactively. This requires pinpointing locations where accidents frequently occur, which are typically identified only after incidents have happened. On the other hand, preventing accidents can leverage advanced technology, such as systems that assist drivers by providing real-time warnings to avert potential collisions.

Research in road safety has traditionally focused on understanding accident-prone areas using statistical and machine learning models. Techniques such as crash frequency analysis, empirical Bayesian methods, and nonparametric density estimation like KDE have been employed to model historical accident data. However, these methods often focus on infrastructure factors (e.g., road geometry or traffic flow) rather than the behaviors leading to accidents. Pinpointing accident locations is typically achieved by analyzing accident statistics from various road locations using traditional statistical methods. However, this work leverages Advanced Driver Assistance System (ADAS) devices to identify high-risk locations proactively, before they evolve into accident-prone zones. By doing so, we aim to enhance road safety and prevent the loss of lives.

Developing technology to assist drivers effectively requires high-quality data. While numerous studies and datasets have been available to address road safety, most have prioritized pedestrian safety or ego-driver behavior. These datasets typically focus on capturing actions related to the ego vehicle like right/left lane change, U-turn, etc., or on the behavior of road agents concerning the ego-vehicle like yielding, cutting, overspeeding etc. This narrow focus limits the understanding of the broader interactions that occur on the road, particularly the risky behaviors of vehicles. Observing a vehicle changing lanes, a pedestrian appearing, or a car in front does not automatically indicate a safety issue. The real risk arises when these road agents are close to the ego-vehicle. Existing datasets fail to capture this crucial aspect. Human judgment naturally assesses safety by evaluating the distance between road agents and the ego-vehicle. we introduce a novel dataset called IDD-CRS, which addresses the gaps in existing

road safety-related datasets. This dataset was collected using ADAS devices, which provide precise temporal annotations for events. It also incorporates hard negative examples, a unique feature that enhances the dataset's utility in building robust models for video recognition tasks. We benchmarked the IDD-CRS dataset for two key video tasks: action recognition and long-tail action recognition. Popular existing methods were evaluated on this dataset, and their performance was reported, providing insights into their strengths and limitations in safety-critical scenarios.

1.2 Thesis Contributions

The contributions of the thesis are as follows:

1. **Enhancing Road Safety: Predictive Modeling of Accident-Prone Zones with ADAS-Equipped Vehicle Fleet Data:**

   (a) **Early Identification of Accident-Prone Zones:** This work presents a novel methodology to identify potential accident-prone zones proactively using geo-tagged collision alert data collected from a fleet of 200 city buses equipped with Advanced Driver Assistance Systems (ADAS). To the best of our knowledge, this is the first research to utilize ADAS alerts for early detection of high-risk areas in a large-scale urban road network.

   (b) **Innovative Modeling and Evaluation Techniques:** The study employs Kernel Density Estimation (KDE) to model the spatiotemporal distribution of alert data across stratified time intervals. Additionally, a novel recall-based measure is introduced to evaluate the correspondence between KDE-identified zones and manually reported accident-prone areas (blackspots). This approach significantly outperforms existing methods using the recall-based metric.

   (c) **Prediction of Previously Unidentified Zones:** A new Earth Mover Distance-based linear assignment measure is proposed to predict accident-prone zones that were previously unidentified. The methodology demonstrates the potential of using ADAS alert data to support civic planners in recognizing emerging high-risk zones, enabling timely traffic-calming interventions and ultimately enhancing road safety.

2. **IDD-CRS: A Comprehensive Video Dataset for Critical Road Scenarios in Unstructured Environments:**

   (a) **Introduction of IDD-CRS Dataset:** This work presents IDD-CRS, a large-scale dataset focused on critical road scenarios, captured using Advanced Driver Assistance Systems (ADAS) and dash cameras. Unlike existing datasets that primarily emphasize pedestrian safety, IDD-CRS provides a comprehensive perspective by incorporating both vehicle and pedestrian behaviors, along with complex interactions between road agents.

(b) **Diverse Scenarios and Precise Annotations:** The dataset includes diverse scenarios, such as high-speed lane changes, unsafe vehicle approaches to pedestrians and cyclists, and close interactions between ego vehicles and other agents. By leveraging ADAS technology, IDD-CRS ensures precise temporal annotations for events, resulting in highly reliable data for safety-critical analysis.

(c) **Benchmarking and Advancing Model Development:** With 90 hours of video footage comprising 5400 one-minute-long videos and 135,000 frames, IDD-CRS introduces new vehicle-related and hard negative classes. Baselines for action recognition and long-tail action recognition tasks are established, highlighting the limitations of existing models and providing insights for future advancements in road safety technology.

1.3 Organization of Thesis

Chapter (2) provides a comprehensive overview of our work "Enhancing Road Safety: Predictive Modeling of Accident-Prone Zones with ADAS-Equipped Vehicle Fleet Data". It includes discussions on our data collection, experimental setup, results, and insights into improving model performance.

Chapter (3) provides a comprehensive overview of our work "IDD-CRS: A Comprehensive Video Dataset for Critical Road Scenarios in Unstructured Environments". It includes discussions on our data collection, experimental setup, results, and insights into improving model performance.

Chapter (4) concludes the thesis with final remarks. This section also highlights publications stemming from our research group.

Chapter (5) presents the concluding thoughts and future works.

Chapter 2

Enhancing Road Safety:Predictive Modeling of Accident-Prone Zones with ADAS-Equipped Vehicle Fleet Data

2.1 INTRODUCTION

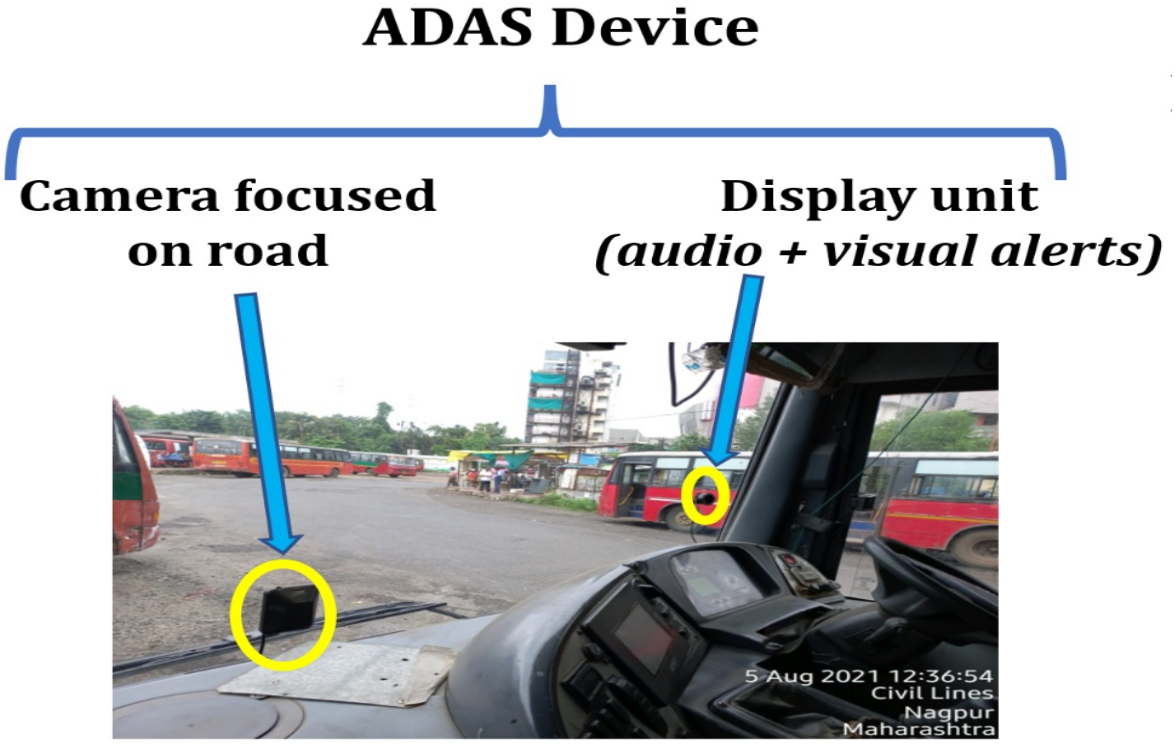

Figure 2.1 Inside view from a bus installed with an Advanced Driver Assistance System (ADAS). The ADAS system comprises (i) a camera installed inside on a windshield and monitoring the road ahead of the vehicle, and (ii) a small display with a buzzer to provide audio and visual alerts to the driver.

Road accidents pose a significant global challenge, not only leading to the loss of lives within families but also causing economic strain on dependent households and impacting the overall welfare of the country. This issue is particularly pronounced in developing nations, it is the tenth leading cause of death, and India, in particular, confronts a formidable task in enhancing road safety. The country ranked first globally in terms of road fatalities annually.

In this paper, we present our work in the city of Nagpur - a large city in India, with a population of three million. The city faces a severe road safety issue, partly attributed to its location at the intersection of two major national highways (NH) – NH-44 and NH-53. Typically, most of the steps taken by city and traffic planning authorities address issues in the road network only after an accident has occurred. Our work is precisely designed to address this problem of early detection of possible accident-prone zones which need attention from civic authorities. Specifically, we present a novel approach to identify possible accident-prone zones in a large city-scale road network using collision alert data from a vehicle fleet. Our geo-tagged alert data has been collected over a year from 200 city buses installed with an Advanced Driving Assistance System (ADAS). To model the distribution of alert data across stratified time intervals, we employ a nonparametric technique called Kernel Density Estimation (KDE). We introduce a novel recall-based method to assess the degree of support provided by our density-based approach for existing, manually determined accident-prone zones ('blackspots') provided by civic authorities. A quantitative comparison of our KDE approach with other approaches used for modeling accident data reveals that our approach significantly outperforms previous approaches in terms of the proposed recall-based method. We also introduce a novel linear assignment Earth Mover Distance-based measure to predict previously unidentified accident-prone zones.

Overall, our results and findings support the feasibility of utilizing ADAS's alert data from vehicle fleets to aid civic planners in assessing accident-zone trends and deploying traffic-calming measures, thereby improving overall road safety and saving lives. Although presented in a specific setting, the general nature of our approach and analysis techniques is an attractive option for replication at scale and useful for other countries and geographies.

2.2 RELATED WORKS

The accident is a daunting challenge in the current scenario. Work has been done in different geographical areas to identify the accident locations based on the past accident history. These works analyze accident data from various countries including Sweden [14], Turkey [8], and United States [27]. The time frame of the data ranges from 2 to 43 years. The datasets contain information on the location, time, and details of the accident, as well as information on the involved vehicles and road conditions at the time of the accident.

In recent years, there has been a growing interest in identifying accident-prone zones to improve road safety and reduce the number of accidents. Statistical methods such as crash frequency [1], empirical bayseian [23], linear regression [52] and negative binomial regression [7], have been used to

model the correlation between past accident data and non-behavioral factors such as highway geometry, environmental conditions, and traffic characteristics. Some studies have used machine learning methods such as K-Nearest Neighbor Classifier (KNN) [24], Support Vector Machine (SVM) [40] and Random Forest (RF) [32, 42] to classify small road segments as potential accident-prone zones. Nonparametric density estimation like Network KDE [50] and clustering methods like DBSCAN [12, 45, 15, 51, 39] and Monte Carlo Simulation [2] have also been used to cluster accident points in crucial sections of the road such as intersections and junctions. To incorporate both temporal and spatial dependencies, deep learning models such as Long Short-term Memory (LSTM) [37] and Stack denoising convolution autoencoder (SDCAE) [5] have been used. However, these works focus on non-behavioral factors and do not examine data arising from driver behaviors.

Reducing accidents involves more than just pinpointing their locations. Efforts are made to minimize accidents in real-time using advanced technologies, such as the automatic braking system [16], which slows down the vehicle when it is about to collide with other road entities. Some studies consider braking data as accident indicators, using them to identify accident locations [36, 48], but this data lacks the exact reason for the event. Events can occur due to poor driving behavior, making the location not inherently accident-prone. There is a chance of incorrect information in this data, particularly as it is collected from a single vehicle, introducing a potential bias.

The methods mentioned above aren't directly comparable to the newly introduced approach due to the unique characteristics of our data. Nevertheless, this work has incorporated several of these methods as baselines for comparison. The meticulous analysis of this work highlights KDE's superiority over other methods in distinct settings, as evidenced by strong alignment with existing blackspots data. Notably, Network KDE [50], a density-based approach, and DBSCAN [39], a clustering-based technique, demonstrate marginally improved performance compared to frequency-centric approaches like Crash Frequency [1] and Empirical Bayesian [23]. Despite these subtle variations, the backing from pre-existing blackspots data remains considerably limited for these methods (see Figure **??**), particularly when juxtaposed with the results achieved by KDE. Additionally, these methods encounter challenges when handling substantial data volumes.

2.3 Data and its collection setup

A large dataset is necessary to conduct a comprehensive analysis. Therefore, the ADAS alerts generated from Nagpur public transit buses on different routes for a year have been utilized. Figure 2.2 shows the route chosen for this study. This covers a total distance of 1600 kilometers. The mentioned distance represents approximately 85% of the entire road network in Nagpur. Furthermore, blackspot data given by the civic authorities, have been utilized to supplement the primary data analysis. This approach allows us to gain a comprehensive understanding of the alert patterns and accident-prone zones in Nagpur whereby the efficacy of ADAS devices in detecting these patterns is also assessed.

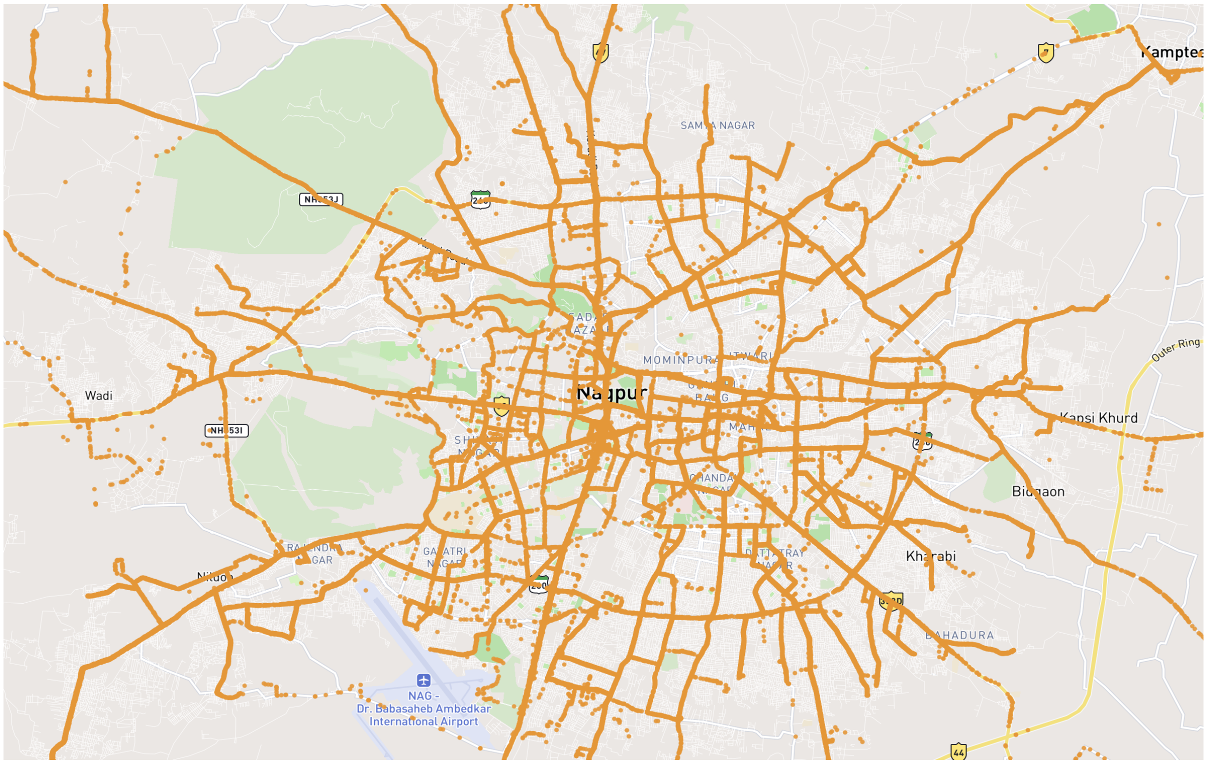

Figure 2.2 Comprehensive Vehicle Route in Nagpur: Mapped in orange, this route spans 1600 kilometers, representing approximately 85% of Nagpur's entire road network, as illustrated on the city map.

Table 2.1 Description of data types, events, and a brief description of data collected and analyzed in our work.

Data	**Event**	**Description**
Spatial	Location (GPS coordinates)	Latitude and longitude of the alert
Temporal	Date and time	Date and time of the alert
Alerts from Advanced Driver Assistance System (ADAS)	Pedestrian collision warning (PCW)	Alert for a potential collision with a pedestrian, in front of the driven vehicle
	Front collision warning (FCW)	Alert for a potential collision with another vehicle in the lane, in front of the driven vehicle
	Lane Departure Warning (LDW)	Alert when the driven vehicle moves out of a lane, without using a lane-change indicator

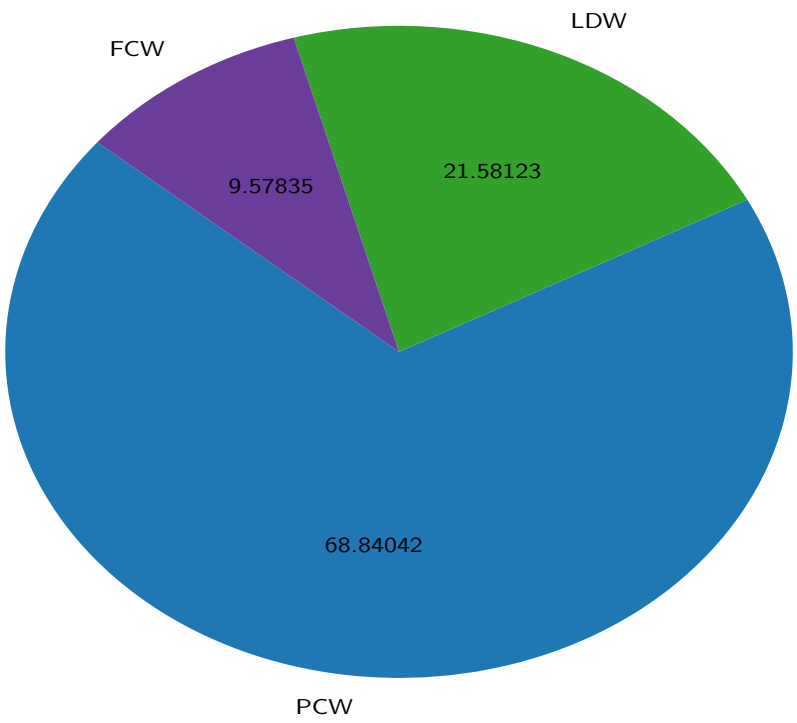

Figure 2.3 Distribution of alerts - FCW, LDW, and PCW.

2.3.1 Advanced Driver Assistance System (ADAS)

In this work, the camera-based Advanced Driver Assistance System (ADAS)[1] was utilized. The system is capable of detecting the presence of objects (stationary as well as moving) with type as well as their distance, including GPS coordinates around the vehicle, and accordingly sends visual and audio alarms. These alerts are given to the driver in the output unit if the vehicle is detected to be on an unsafe path (like lane departure), unsafely close to another vehicle/pedestrian/bicycle, or any infrastructure element (potential crash), etc. Based on these visual or audio alerts, the driver can potentially take corrective actions in driving to prevent or avoid an impending collision or any undue hit to infrastructure elements. Figure 3.1 shows the device installed in one of the buses from our study. The device has one AI-enabled camera (input) fitted on the dashboard of the bus and is focused toward the road at an optimum angle to detect various features such as pedestrians, cyclists, lane departure, chances of a collision, road features, and has a display unit (output) which gives visual as well as audio alerts to the driver while driving. To store the huge geo-tagged data coming from the ADAS-equipped bus fleets, a centralized server is used. Table 2.1 lists the types of events that are identified by the device and associated alerts.

[1] from Mobileye

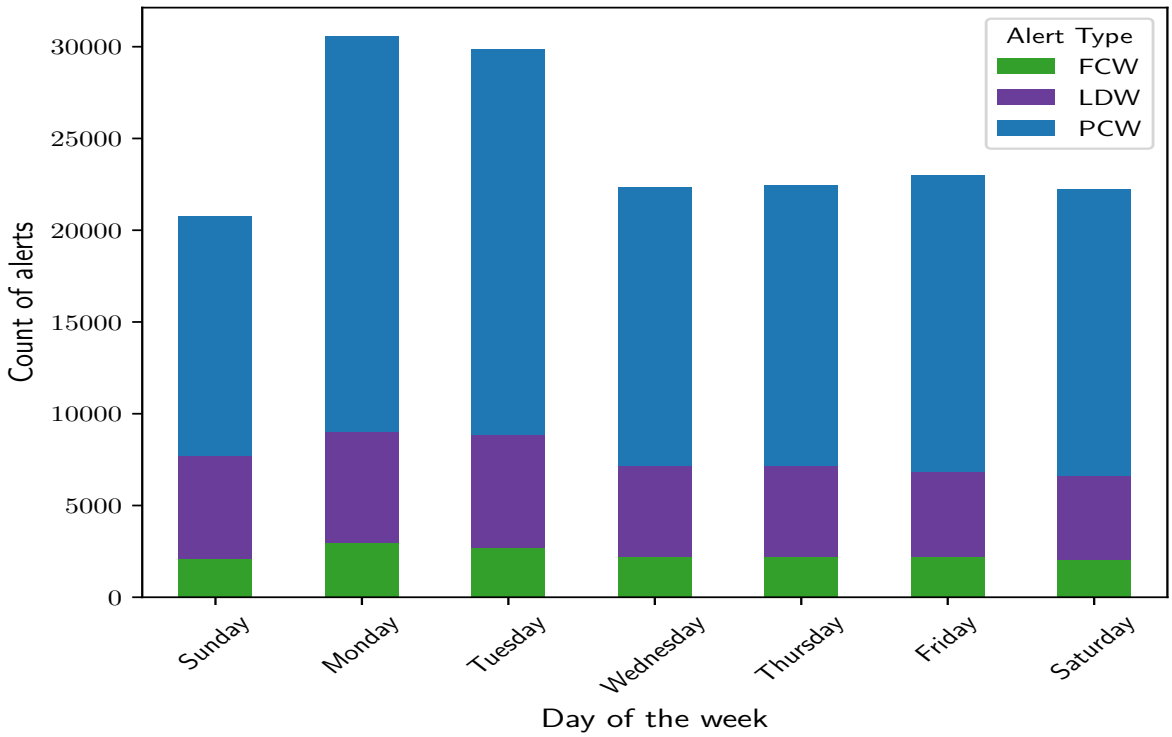

Figure 2.4 Weekly trends in alerts.

2.4 Methodology

The goal is to identify road zones at high risk for accidents (colloquially referred to as ”Blackspots”) by using alert data from ADAS devices. Since traffic patterns vary throughout the day, analyzing the entire data together would not give an accurate prediction. An alternative is to stratify the day into different intervals based on the traffic flow pattern of that zone. Later sections show, that this enables accurate localization of accident-prone zones on the road. From the ADAS alert attributes (Table 2.1), PCW and FCW alert data are used. A nonparametric Kernel Density Estimation (KDE) is employed to model the alert probability density. Prominent (high-probability) locations are subsequently identified (Sec. 2.4.2) and utilized for blackspot verification (Sec. 2.4.3) and prediction (Sec. 2.4.4).

2.4.1 Kernel Density Estimation

Using a sample of observations, the probability density function (PDF) of a random variable can be calculated statistically using Kernel Density Estimation (KDE) [44]. It operates by utilizing a kernel function, a probability distribution function used to weight the data points to smooth a histogram of the sample data. The outcome is an estimation of the population’s underlying PDF from which the sample was drawn. KDE is a valuable tool for datasets that might not follow conventional parametric

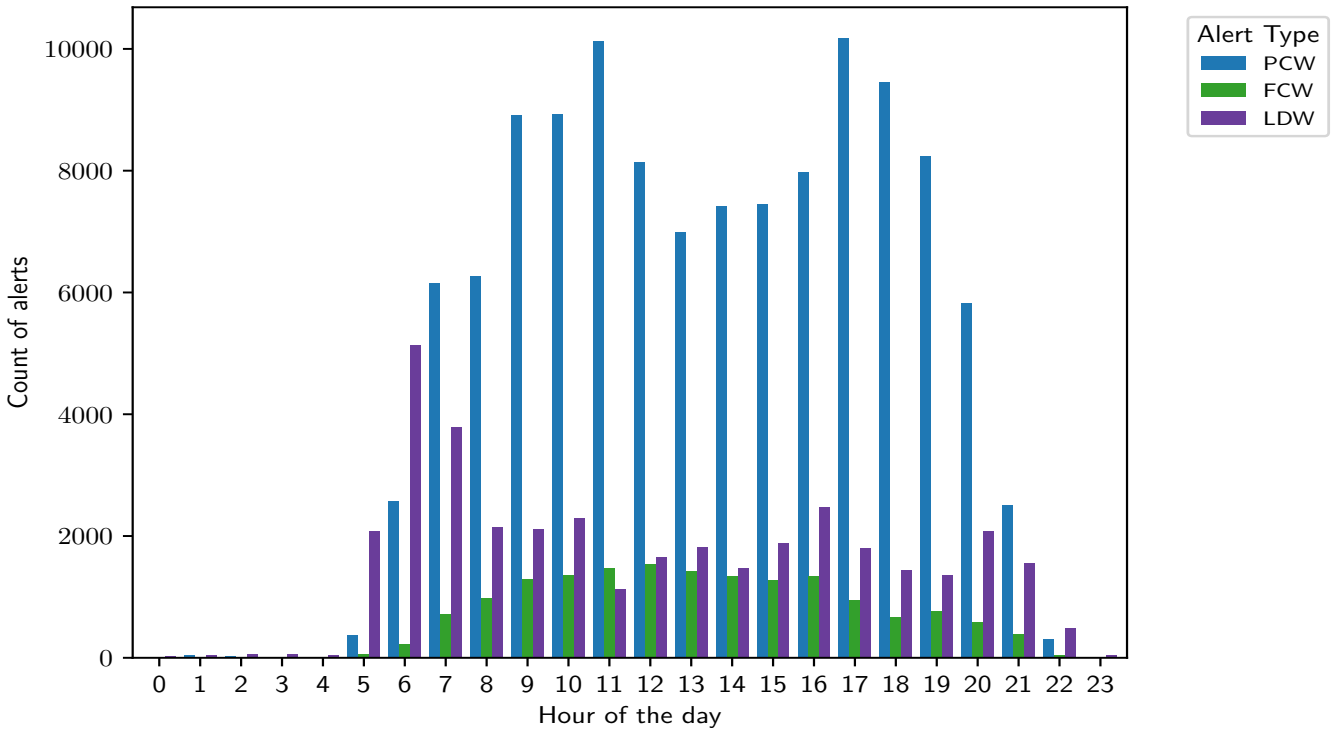

Figure 2.5 Hourly distribution of alerts showing the patterns of peak and non-peak hours in traffic.

distributions since it makes no assumptions about the underlying distribution of the data, which is one of its key advantages.

Let $x_1, x_2, ..., x_n$ be independently and identically distributed samples from a distribution with an unknown density f at any given point x. The kernel density estimator is:

$$f_h(x) = \frac{1}{nh} \sum_{i}^{n} K(\frac{x - x_i}{h}); i = 1, 2, 3, ..., n \tag{2.1}$$

Here, K is the kernel — a non-negative function — and $h > 0$ is a smoothing parameter called the bandwidth. It is crucial to pick the correct kernel function and bandwidth when utilizing kernel density estimation. The weights of the data points are determined by the kernel function, which can be Gaussian, Epanechnikov, triangular, etc. depending on the type of data being used. To provide accurate density estimations, the bandwidth parameter h, which regulates the degree of smoothing applied to the data, must be set to the proper value.

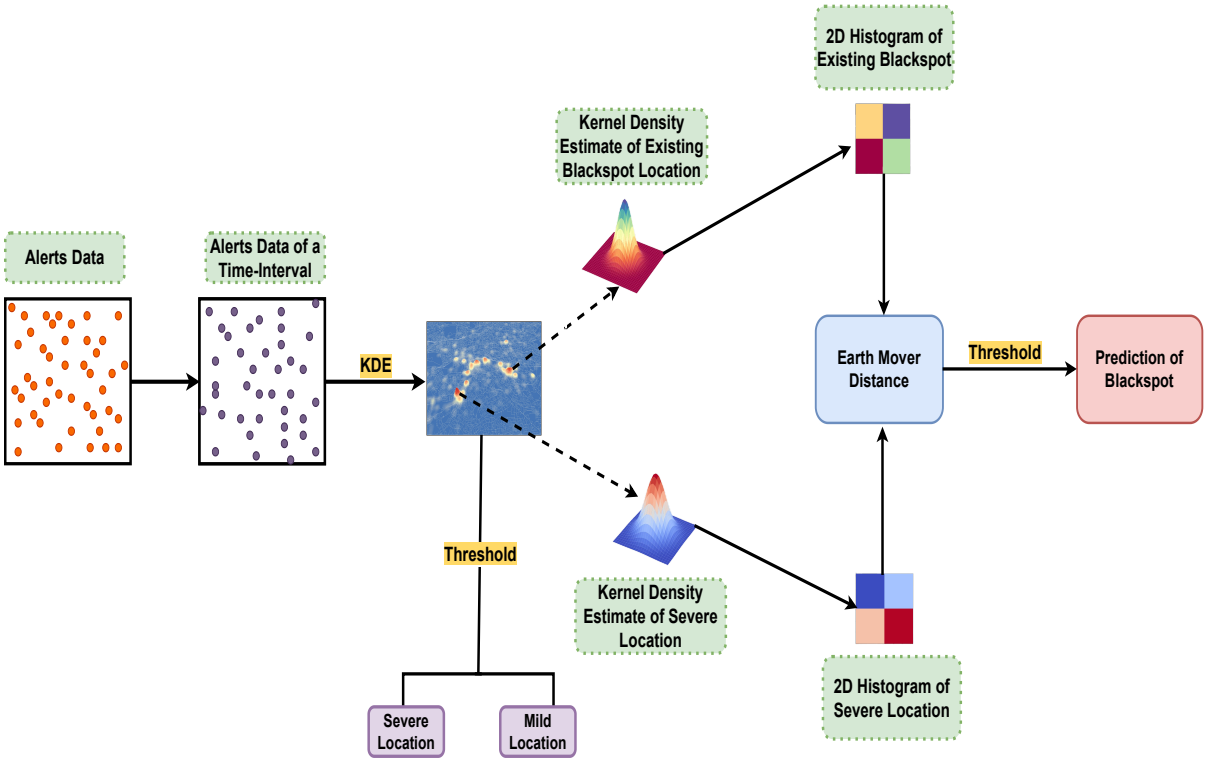

Figure 2.6 The figure outlines our blackspot prediction methodology. The initial step involves segmenting the alert data into distinct time intervals, followed by the application of Kernel Density Estimation (KDE) to model the spatial distribution of alert occurrences. Utilizing a statistically determined threshold, the method identifies locations with varying degrees of severity, classifying them as severe or mild. To forecast new potential blackspots, a rigorous analysis is performed. This analysis encompasses the evaluation of data distributions at existing blackspots and locations of severe alerts, facilitated by the utilization of a 2D Histogram. The comparison employs the Earth Mover Distance (EMD) as a metric. Subsequently, a threshold mechanism is employed to discern and designate emerging blackspot candidates. The amalgamation of these steps forms the basis of our predictive approach.

2.4.2 KDE in our problem

For data, the Forward Collision Warning (FCW) and Pedestrian Collision Warning (PCW) are employed, as these collision types characterize accident-prone zones. The objective is to utilize KDE to estimate the density of alerts at various locations, employing a Gaussian kernel:

$$h(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right) \tag{2.2}$$

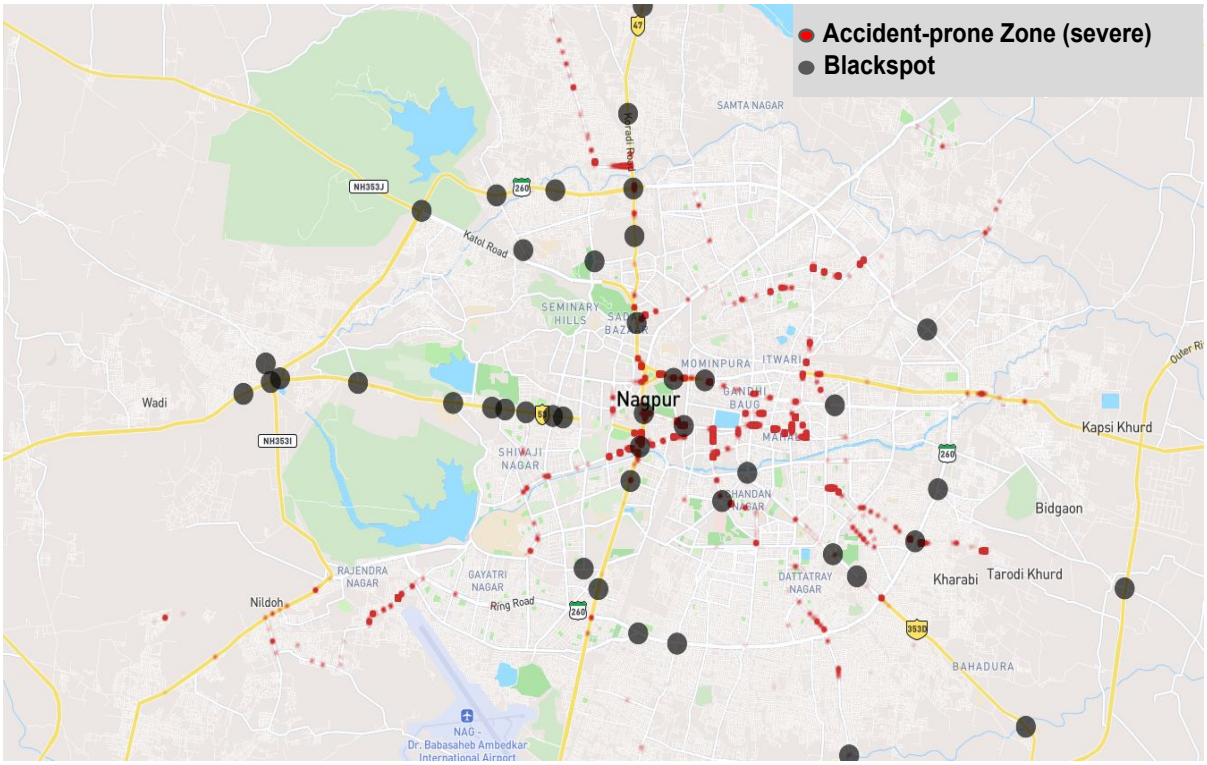

Figure 2.7 Accident-prone zones (severe) determined by KDE after thresholding using ADAS device-based alerts (in red) and manually identified 'blackspots' (black circles) overlaid on the map of Nagpur city during the time interval 8:00-11:59 hours, corresponding to the morning peak commute time.

where μ is the mean and σ is the standard deviation of the distribution.

Typically, the traffic patterns vary in quantity and semantic type (e.g. vehicles). Ignoring these attributes and considering all alerts the same can lead to inaccurate identification of accident-prone zones. To tackle this issue, alerts are distributed in four-time intervals based on the traffic flow pattern of Indian roads (8:00-11:59), (12:00-15:59), (16:00-19:59), and (20:00-23:59), and density is estimated for these time intervals separately. Time intervals early in the morning (00:00-7:59) are not considered as the number of alerts during these times is relatively low. Note that the time intervals can potentially be modified as per modeling needs for a different city, country, etc. To categorize the severity of the identified accident-prone zone, statistically determined thresholds are applied to alert density values. The third quartile (Q3) value of the kernel density estimate is considered as a threshold in this work paper. An example of the resulting density map can be seen in Figure **??** - alert density values greater than a threshold are considered severe (in red) and rest as mild.

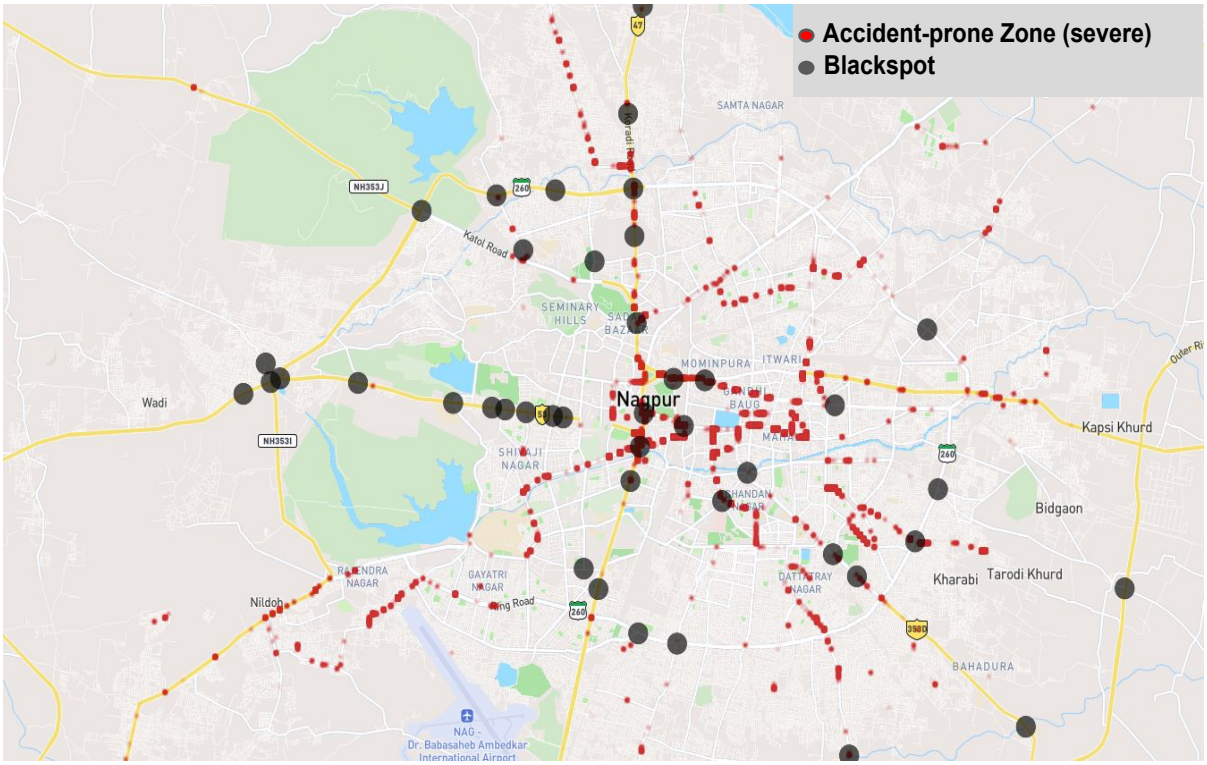

Figure 2.8 Accident-prone zones (severe) during the time interval 12:00-15:59 hours, a period of lower traffic in the afternoon.

2.4.3 Evaluating Support for Blackspots from Alert Data

To determine the extent of support for a given blackspot in terms of alerts in its vicinity, the following recall-based evaluation metric is proposed. For a given time interval, let $B_1, B_2, \ldots B_m$ represent the accident-prone zone ("blackspot") geographical locations provided by the civic authority, considered as ground-truth. Let $A_1, A_2, \ldots A_r$ represent the 'severe' alert locations mentioned in the previous section. Let d be a threshold distance from a blackspot B. The total absence of alerts within distance d from the blackspot is considered a false negative, and let FN be the number of such false negatives. The recall for distance threshold d (Recall-d) is defined as $R_d = 1 - \frac{FN}{m}$, where m is the total number of blackspot locations. Intuitively, a higher recall for various distances d indicates stronger support for blackspots arising from the alert data.

2.4.4 Predicting Blackspots from Alert Data

Apart from determining the extent of support for blackspots in terms of alerts, a useful application is to predict potential' blackspots solely from alert data. To accomplish this, c blackspots with the highest alert density values are selected. For each of these blackspots, a $P \times Q$ spatial distance grid centered

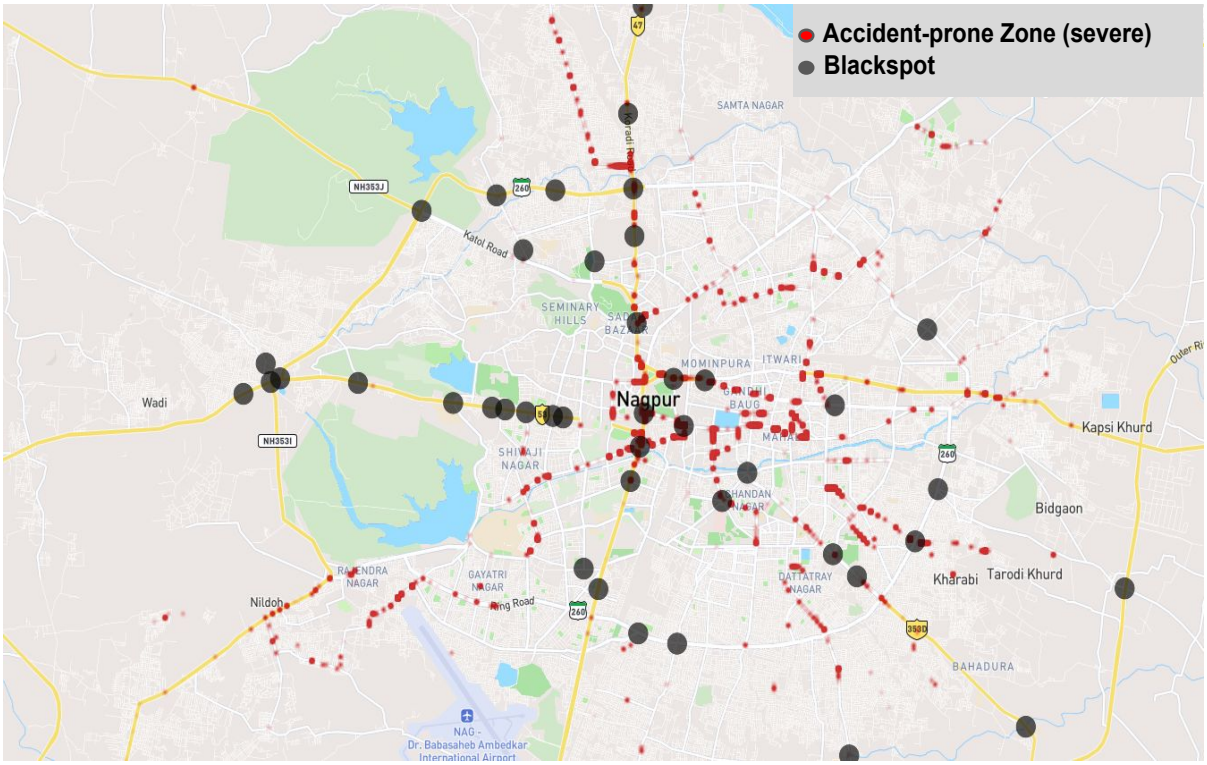

Figure 2.9 Accident-prone zones (severe) during the time interval 16:00-19:59 hours, corresponding to the evening peak commute time.

on the blackspot coordinates is established, and the $2D$ histogram of alerts within the grid is determined (refer to Figure 2.6). Let $H_{b_i}, i = 1, 2, \ldots c$ be the histogram corresponding to the i-th blackspot. For each alert location $A_1, A_2, \ldots A_m$, alert histograms $H_{a_j}, j = 1, 2, \ldots m$ are determined using the same-sized grid as that used for blackspots. The histogram distance' (EMD) with respect to each of the blackspot histograms is then computed. If at least one blackspot exists such that $EMD(H_{a_j}, H_{b_i})$ is smaller than a threshold, the j-th alert location is labeled as a potential blackspot.

For the histogram distance, a standard Earth Movers Distance [38] with linear sum assignment is used as the metric. Formally, let $H = (h_1, n_{h_1}), \ldots, (h_c, n_{h_c})$ be the histogram for a reference blackspot, where (h_x, n_{h_x}) represents the spatial index and normalized histogram count, and $c = P \times Q$ is the total number of histogram bins. Similarly, let $A = (a_1, n_{a_1}), \ldots, (a_c, n_{a_c})$ represent the histogram for an alert location. Let $\mathbf{D} = [d_{ij}]$ be the ground distance matrix where d_{ij} is the ground distance between histogram counts at locations h_i and a_j. The 'distance' between the histograms is formulated as finding the minimum cost flow $\mathbf{F} = [f_{ij}]$ between the histograms.

$$\mathbf{F} = min \sum_{i=1}^{c} \sum_{j=1}^{c} f_{ij} d_{ij} \tag{2.3}$$

subject to the following constraints:

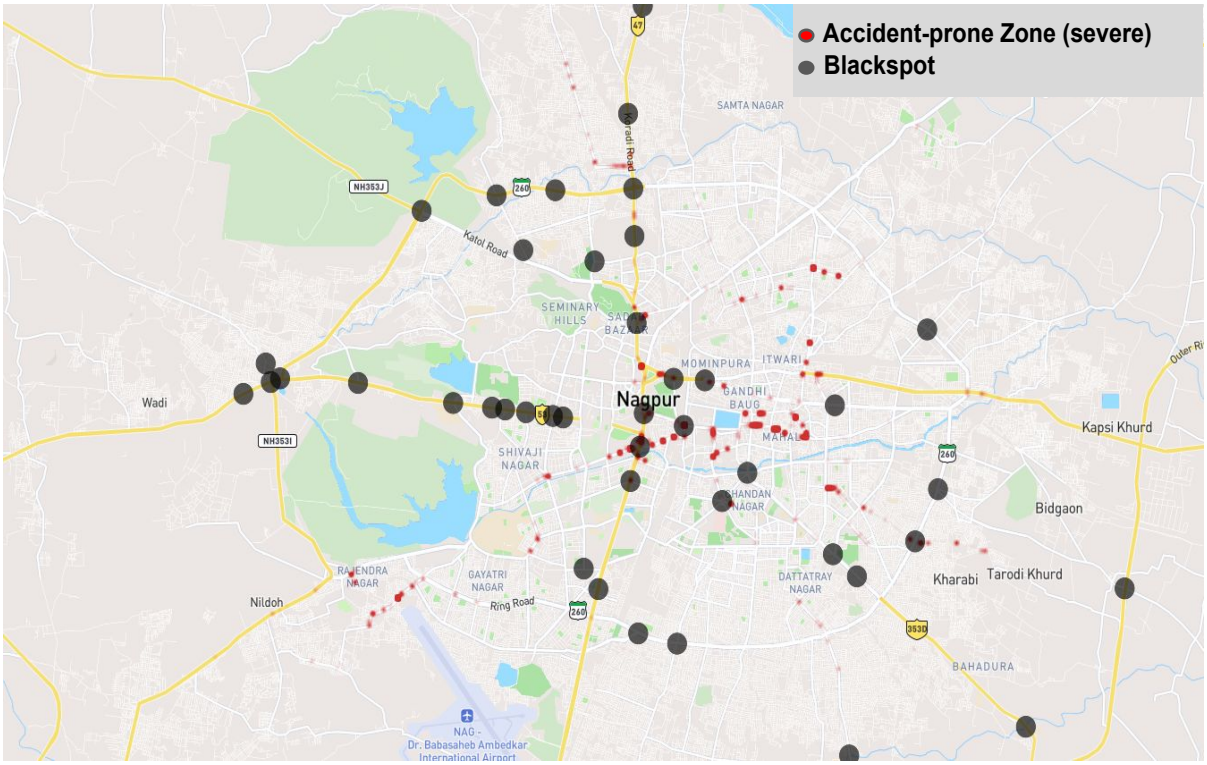

Figure 2.10 Accident-prone zones (severe) during the time interval 20:00-23:59 hours, depicting late-night traffic.

$$
\begin{aligned}
f_{ij} &\geqslant 0; \qquad 1 \leqslant i \leqslant c,\ 1 \leqslant j \leqslant c \\
\sum_{i=1}^{c}\sum_{j=1}^{c} f_{ij} &= min\left(\sum_{i=1}^{c} n_{a_i}, \sum_{j=1}^{c} n_{h_j}\right)
\end{aligned}
$$

Once the optimal optimal flow $\mathbf{F}$ is determined, the Earth Mover Distance is obtained as:

$$\text{EMD}(H, A) = \frac{\sum_{i=1}^{c}\sum_{j=1}^{c} f_{ij} d_{ij}}{\sum_{i=1}^{c}\sum_{j=1}^{c} f_{ij}} \tag{2.4}$$

2.4.5 Implementation Details

The bandwidth of the Gaussian Kernel in KDE is set to 10^{-6}. Given the relatively small spatial extent of geographical coordinates compared to the earth's curvature, the Haversine metric and ball tree algorithm is used for modeling large spatial extent densities. The density values are transformed using a negative logarithm to output values in the range $[5, 17]$. The density threshold is determined statistically, and the Q3 value of kernel density estimated.

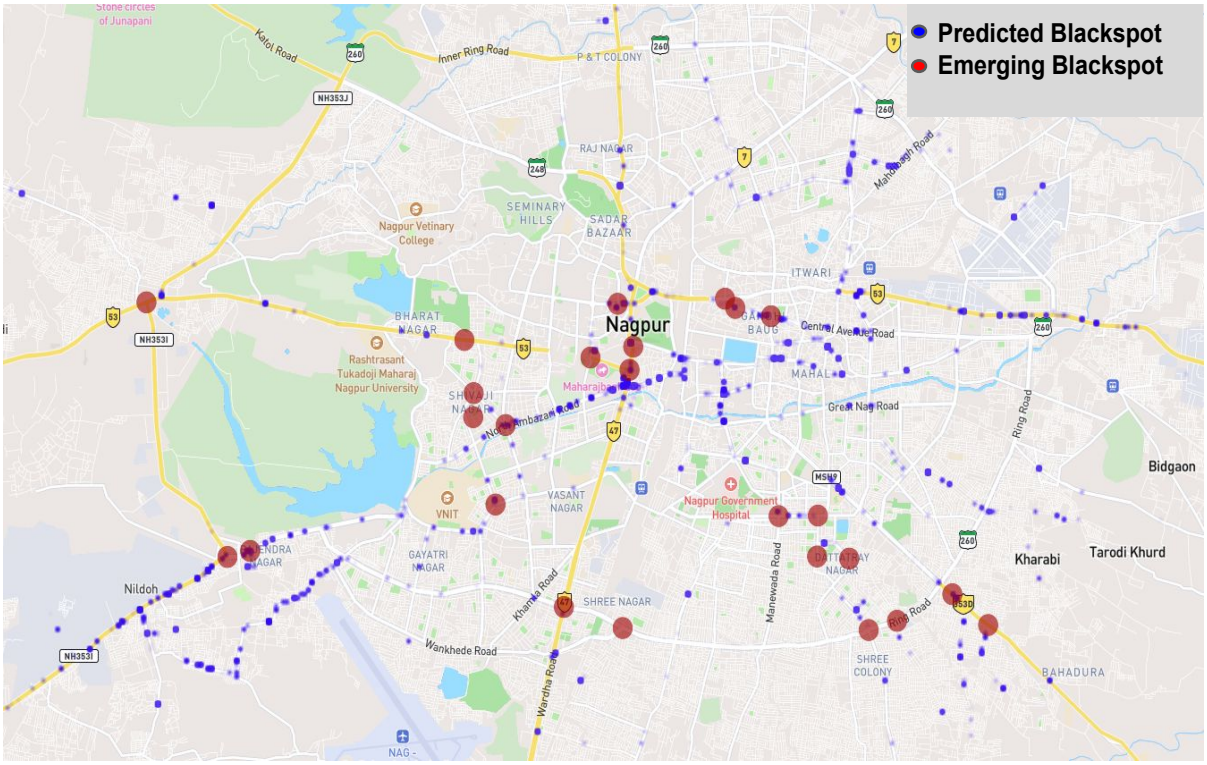

Figure 2.11 Comparison of predicted blackspot locations by our novel method (in blue) with recently identified emerging blackspot locations (in red) by civic authorities.

For blackspot prediction (Sec. 2.4.4), the top $c = 15$ 'blackspots' with the highest KDE values are chosen as reference. To find the 2D histogram for blackspot prediction, a half-width of 0.1 kilometers is considered, and the grid resolution is set to $P = Q = 10$. The base distance for EMD is Euclidean distance. The threshold for EMD distance determines the number of locations for manual verification by civic authorities. Given the range in alert densities, the threshold for various time intervals is empirically set between 20 and 103.

2.4.6 Methods for comparison

In this section, a concise overview of established methodologies used for modeling road accident data is provided. These methodologies serve as essential reference points against which the efficacy of the proposed approach centered around Kernel Density Estimation (KDE) is evaluated. The evaluation is based on the previously introduced metric *Recall-d*, which quantifies the coverage of high-risk areas, also known as blackspots.

1. **Crash Frequency:** The Crash Frequency method stands as a conventional and widely adopted technique. It involves partitioning the target road into discrete segments and gauging the safety performance of each segment based on the count of reported accidents during a specified time window.

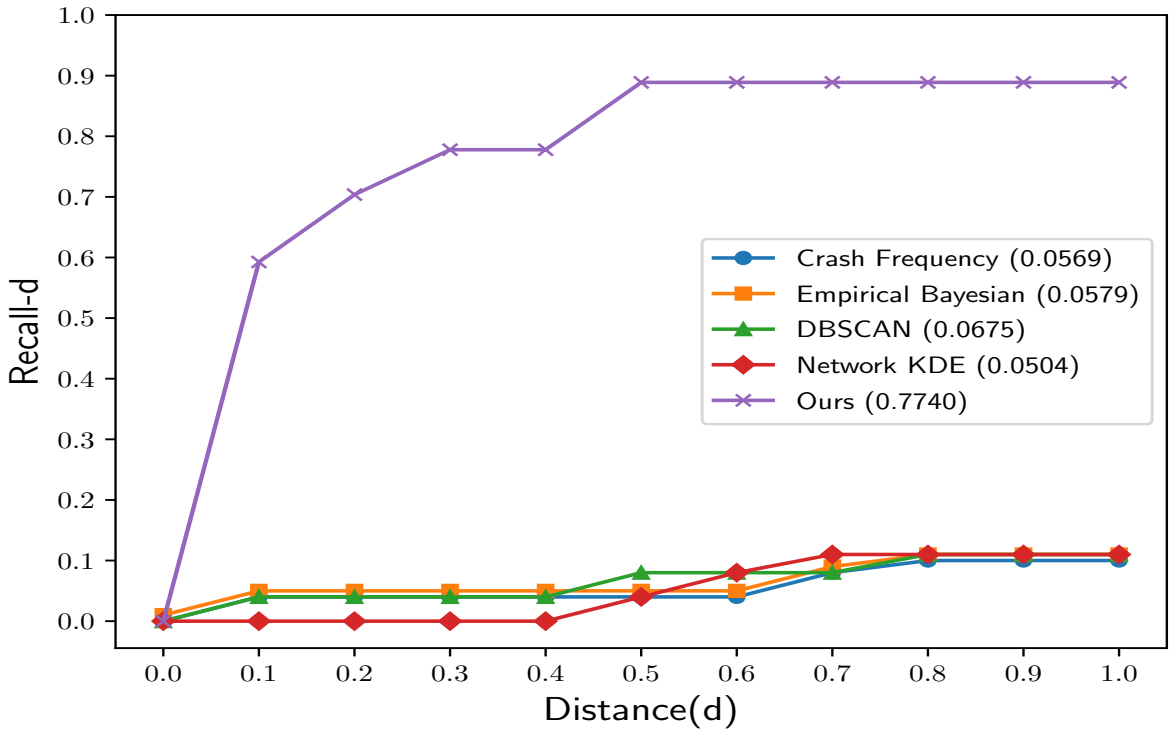

Figure 2.12 Recall-d plot for the 8:00-11:59 hours interval.

This direct assessment of accident occurrences serves as a fundamental benchmark for evaluating road safety[1]. We selected the top 20 segments to check the support from the existing blackspot data. The recall-d value is very low for small distances. This shows these methods are not capturing the alert data pattern.

2. **Empirical Bayesian:** The Empirical Bayesian approach takes a statistical perspective by predicting the likelihood of future alerts. By amalgamating historical accident data with exposure information, it furnishes a probabilistic estimate of potential safety issues. Noteworthy for its adaptability to fluctuations in traffic volume, this method is frequently employed in safety analyses[23]. For both the Crash Frequency and Empirical Bayesian methods, road segmentation is pivotal.

3.**DBSCAN (Density-Based Spatial Clustering of Applications with Noise):** DBSCAN represents a clustering algorithm that discerns clusters of dense data points while discerning sparser regions as noise. In the context of road safety, DBSCAN can be deployed to detect clusters of high-alert density, signifying potential accident-prone zones[39]. In our study, we meticulously explored different values of DBSCAN's hyperparameters to determine the optimal configuration. we kept the hyperparameters same for all the time intervals.

4. **Network KDE (Kernel Density Estimation):** Network KDE employs a statistical technique to infer the probability density function of alert occurrences across the road network. This method furnishes a spatially nuanced depiction of alert density, allowing for pinpointing areas of heightened

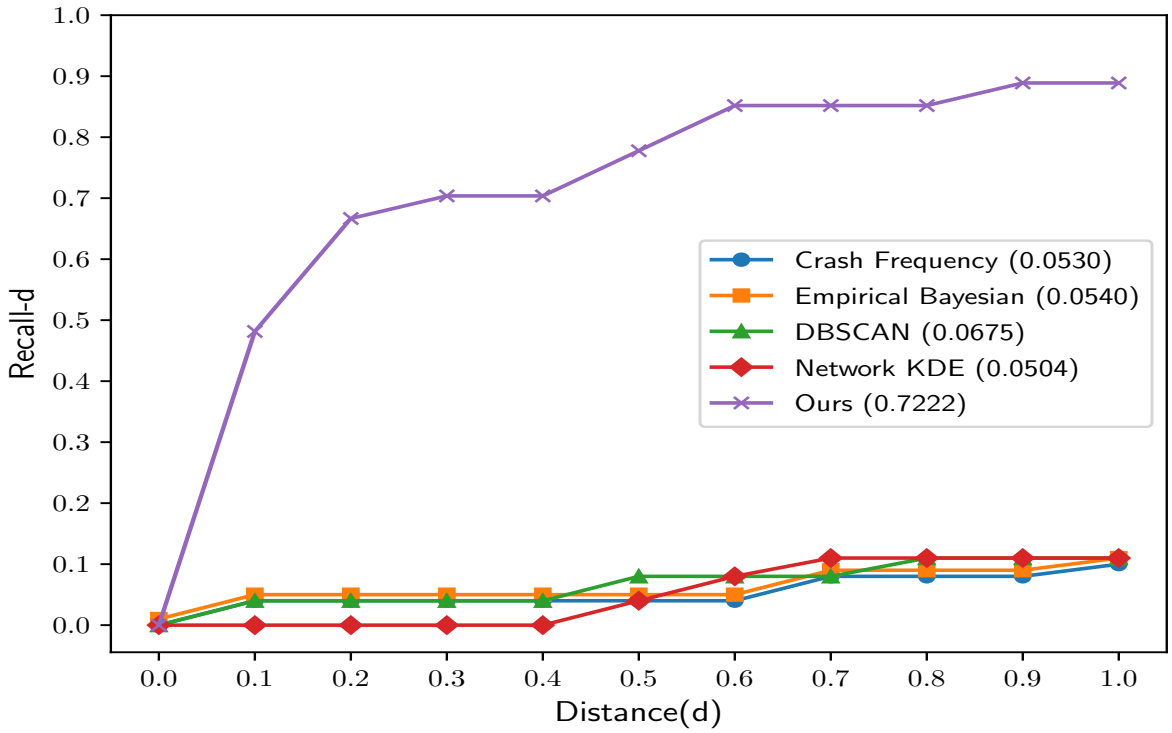

Figure 2.13 Recall-d plot for the 12:00-15:59 hours interval.

safety concern[50]. In this approach, density calculations are confined to individual road segments of standardized areas. It's worth noting that the KDE parameter remains uniform for all segments. The bandwidth of the Gaussian Kernel in KDE is set to 10^{-4}. The KDE estimate provides insights into areas with elevated safety concerns. Road segments with higher density values are indicative of zones where alerts are clustered, suggesting potential risk-prone areas. While past research predominantly utilized accident data for modeling, we extrapolate this method to alert data. An inherent limitation of Network KDE is its constraint in modeling data within individual segments.

All the aforementioned methods are comparable, as they leverage historical accident or alert data to assess road safety. Each method has its strengths and limitations, which make them suitable for different scenarios and research objectives. Importantly, these methods share similarities with our approach, which utilizes alert data to predict potential safety issues. Figure 2.12, 2.13, 2.14, 2.15, demonstrates that our method achieves the highest recall-d values for small distances across all time intervals, with Network KDE following as the next best performer. However, there is a notable difference in the values. On the contrary, the other methods exhibit significantly lower recall-d values, indicating their inadequacy in modeling the alert data accurately.

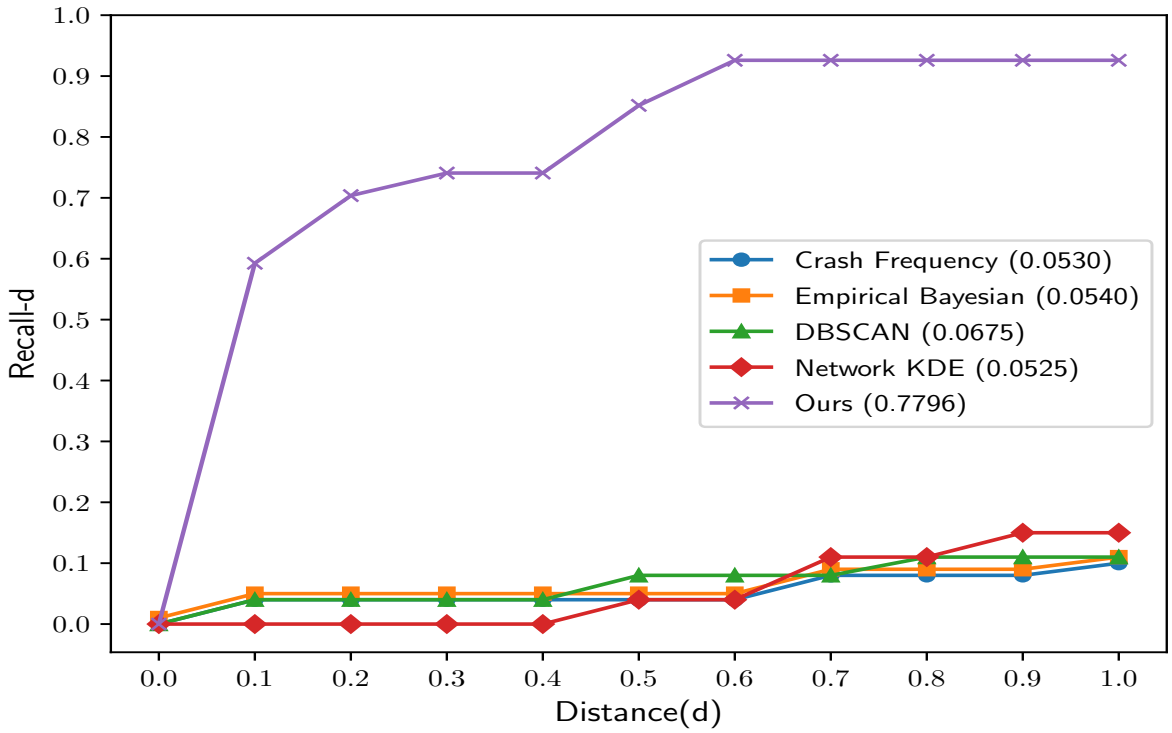

Figure 2.14 Recall-d plot for the 16:00-19:59 hours interval.

2.5 Results and Analysis

2.5.1 Primary data

Figure 2.3 shows the distribution of all alerts. PCW (pedestrian) alert dominates at 58.8 %, followed by LDW (lane departure) and FCW (forward collision) alerts, which are generated at around 31.5 % and 9.5 % respectively. Further, these alerts are examined hourly. Figure 2.5 demonstrates that alerts are not evenly dispersed throughout the day. Primarily, PCW alerts are generated between the hours of 7:00-12:00 (morning) and 17:00-20:00 (evening), and more LDW alerts are observed at time intervals of 5:00-8:00 (morning), indicating frequent lane changes at high-speed during the lean hours of traffic movement in the road network of Nagpur. The generation of FCW alerts is almost similar throughout the day, and there has been no significant pattern depicted.

Subsequent to the preliminary analyses, PCW and FCW alerts were selected for further analysis. As Nagpur is an urban area, LDW alerts are more commonly generated in the region because of the heterogeneous movement of vehicles.

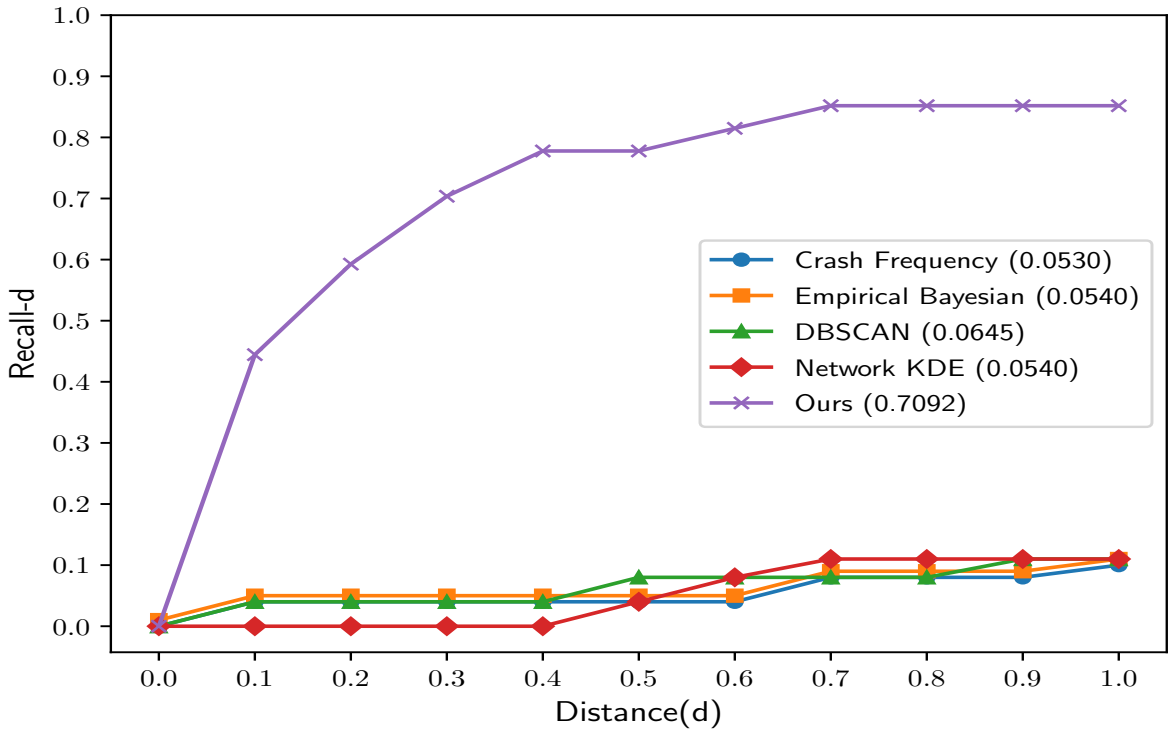

Figure 2.15 Recall-d plot for the 20:00-23:59 hours interval.

2.5.2 Support for blackspots from alert data

The plots of Recall-d (Sec. 2.4.3) values for various distance-to-blackspot thresholds d can viewed in Figure 2.16. From the plots, it can be seen that the values quickly rise to a large (0.8 and above) recall. This indicates strong support from the alert data for existing manually determined blackspots. In the figure, each plot line corresponds to a specific time interval. The area under the respective curve (AUC) is computed to identify the time interval with the best overall recall. For severe alerts, the highest AUC is for 12:00 - 15:59 (orange curve), suggesting that the support for blackspots is strongest during afternoon peak hour traffic. This is followed by 16:00 - 19:59, the evening peak hour traffic (green curve).

Figure **??** shows the manually determined blackspot locations by civic authorities (black circles) and 'severe' alert locations (red small circles), overlaid on the city map. The overlapping presence of alerts at blackspot locations is consistent with our recall-based metrics from the previous section.

2.5.3 Predicting Blackspots from Alert Data

Figure 2.11 illustrates the alert-based predictions of blackspots generated by our novel method (Section 2.4.4) in blue. To validate our blackspot predictions, we collaborated with the Nagpur civic authorities, who furnished us with a list of 25 emerging blackspots, highlighted in red. Out of these, 16

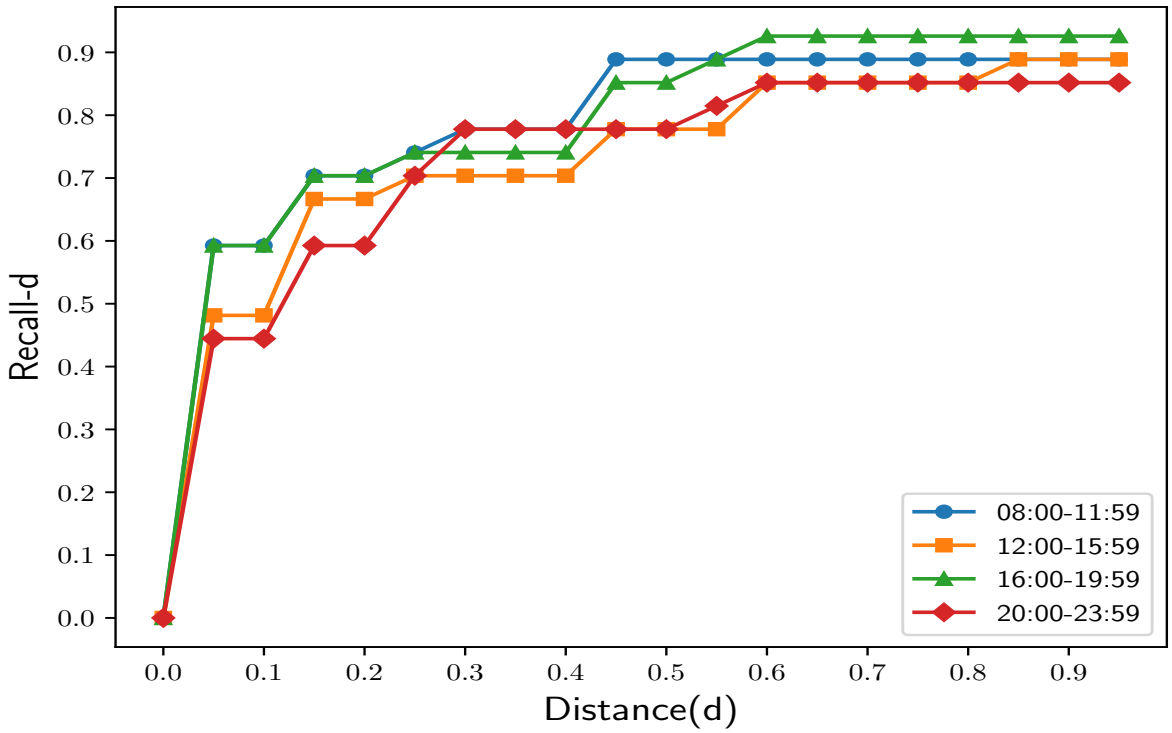

Figure 2.16 Recall-d plots for severe alert locations. The x-axis represents the distance-from-blackspot threshold d. It's noteworthy that high recall values are achieved at smaller distance values, indicating strong support for blackspots from alert data.

emerging blackspots align closely or in very close proximity to our predicted locations, resulting in a recall rate of 0.64 for our prediction model. The advantage of such predictive models is that they provide a ranked priority list which can be used to reduce the number of locations visited by civic authorities to confirm its accident-prone.

Also, the overlay diagrams are a useful way for civic authorities to determine the trends in alert locations as a function of various time intervals. For example, the persistent presence of alerts regardless of time interval could suggest that the location needs traffic-calming measures on a priority basis.

The use of KDE with ADAS alert data makes our method robust to changing external factors that can affect road accident risk, such as climate changes, road expansion, etc. Because changes in any of these factors can lead to an increase or decrease in alerts in that location, KDE can incorporate these changes effectively. While previous work has to use these changes as features for their model. This is an improvement over previous methods that are more dependent on these factors, and their use of past accident data as a basis for predictions was only sometimes accurate as road networks and the number of vehicles changed over time. By using real-time data and considering the dynamic nature of traffic patterns, the proposed method can provide more accurate results of accident risk zones on the road.

2.6 CONCLUSIONS

In conclusion, this study has demonstrated the application of the KDE method in identifying accident-prone zones in Nagpur, India, using ADAS alerts. By evaluating the correlation between vehicle flow and road parameters with ADAS alerts and evaluating the effectiveness of these alerts in predicting accident-prone zones, the KDE approach is proven to be highly accurate in identifying accident-prone zones. The results of this study have significant implications for the field of road safety, as the ability to accurately identify the accident-prone zone will aid in solving the accident-prone zone. The study methodologies used in this study are generic and can be applied to predict accident-prone zones on various types of roadways and regions.

In terms of future scope, further research can be done to validate the results of this work across different regions, road types, and weather conditions. Additionally, different approaches could be explored to improve the accuracy of the predictions made using the KDE method.

Chapter 3

IDD-CRS: A Comprehensive Video Dataset for Critical Road Scenarios in Unstructured Environments

3.1 INTRODUCTION

Road safety is an increasingly critical issue around the globe, especially on unstructured roads. The growing number of vehicles on the road, coupled with the complexity of modern transportation systems, has led to a surge in accidents and near-miss incidents. Road safety efforts focus on preventing accidents and mitigating risks for all road users, including pedestrians, drivers, cyclists, and motorcyclists.

While numerous studies and datasets have been available to address road safety, most have prioritized pedestrian safety [35] [34] or ego-driver behavior [33]. These datasets typically focus on capturing actions related to the ego vehicle like right/left lane change, U-turn, etc., or on the behavior of road agents concerning the ego-vehicle like yielding, cutting, overspeeding etc. [4] This narrow focus limits the understanding of the broader interactions that occur on the road, particularly the risky behaviors of vehicles. Observing a vehicle changing lanes, a pedestrian appearing, or a car in front does not automatically indicate a safety issue. The real risk arises when these road agents are close to the ego-vehicle. Existing datasets fail to capture this crucial aspect. Human judgment naturally assesses safety by evaluating the distance between road agents and the ego vehicle. To address this gap, we introduce a dataset IDD-CRS designed to capture critical road scenarios where accidents can happen if the driver is not precocious. This dataset emphasizes pedestrian safety while also addressing complex ego-vehicle behaviors, such as high-speed lane changes, close encounters with other road agents, and instances of unsafe following distances. Additionally, it includes normal driving classes. To the best of our knowledge, our dataset is the first to provide a comprehensive class for pedestrian, vehicle, normal driving, and ego vehicle behavior, specifically incorporating unsafe distances.

We used an ADAS to pinpoint and measure the exact timing of important road events. Unlike existing datasets that rely on manual annotations—which can be inconsistent and inaccurate—ADAS gives precise start and end times for these actions. This helps ensure our dataset accurately captures critical safety moments on the road. We have established benchmarks for action recognition and long-tail ac-

Figure 3.1 Inside view from our car installed with a DDpaiX2 RGB Dash-cam and an Advanced Driver Assistance System (ADAS). The ADAS system comprises (i) a camera installed inside on a windshield and monitoring the road ahead of the vehicle, and (ii) a small display with a buzzer to provide audio and visual alerts to the driver.

tion recognition on the IDD-CRS dataset using existing popular models for these tasks. Additionally, we have identified the limitations of current methods and provided insights for future improvements.

3.2 Related Work

3.2.1 Exisiting Datasets

In recent years, the study of driver and pedestrian behavior [25] has gained significant attention due to its role in collision prevention [49] and road safety [26]. Behavior prediction [33] focuses on anticipating driving actions like turns, acceleration, merging, and braking, as well as driver behaviors [4] such as overspeeding, overtaking, cut-ins, and rule violations. While much of the research has focused on pedestrian-vehicle interactions [35, 29], vehicle-to-vehicle interactions are equally important for ensuring overall road safety. Multiple datasets exist that cater to pedestrian systems [34], including both real-world and synthetic data [21], with some using simulators [28, 17] for scenario generation. Given

the critical nature of traffic safety events, collecting real data is challenging and resource-intensive, leading to various methods for scenario generation by editing existing videos—such as introducing new agents or modifying the trajectories of existing ones [47].

Some studies have resorted to collecting accident data from sources like YouTube [9], while others focus on specific driving agents, primarily pedestrians, at locations like intersections. However, these datasets often lack detailed temporal annotations that indicate when these agents are actually in danger. Instead, they assume that the presence of any traffic agent in the frame demands precautionary action. In reality, any road user, including vehicles, can pose safety risks, and danger is not constant throughout a scene. ADAS addresses this gap by issuing alerts when traffic agents are genuinely at risk.

3.2.2 Action recognition

Action recognition in video has garnered significant attention, driven by its wide range of applications in surveillance, autonomous driving, and human-computer interaction. Traditional approaches often rely on extracting spatio-temporal features using 3D convolutional neural networks (C3D) [46] or Inflated 3D Convolutional Networks (I3D) [3] to capture motion patterns across frames. More recent methods have explored using architectures like SlowFast [11] and X3D [10], which effectively balance the trade-off between accuracy and computational efficiency by processing videos at different temporal resolutions. Attention mechanisms, including Transformer-based models such as Motionformer [30], have also been integrated to capture long-range dependencies and improve action recognition in complex scenes. Despite these advancements, challenges remain in recognizing actions in real-world, long-tail scenarios, where certain activities are rare and models need to generalize effectively across varied and dynamic environments.

3.2.3 Long-tail Methods

Addressing long-tail recognition typically involves two strategies: re-weighting and re-balancing. Re-weighting methods focus on penalizing the misclassification of tail class samples by adjusting logits [18] or weighting [43] the loss according to class size or sample difficulty. Other techniques like label smoothing [55], enforcing separation between class embeddings [20], and combining predictions from experts specialized in tail classes are also employed. These methods aim to improve model performance on underrepresented classes by adjusting how errors are penalized or how class predictions are handled.

Re-balancing approaches, however, focus on adjusting the training data distribution rather than the loss function. This is often done through class-equalizing feature banks [22] or equal sampling from each class, with a standard practice of first using instance-balanced sampling followed by class-balanced sampling [13]. Augmentations further enhance tail class diversity by combining samples with nearby class prototypes, expanding tail classes through feature clouds, or pasting tail objects onto head class backgrounds [19]. Additionally, contrastive learning [6] improves representations, while video-specific techniques like LMR [31], Mixup [53] and Framestack [54] mix up samples temporally during training.

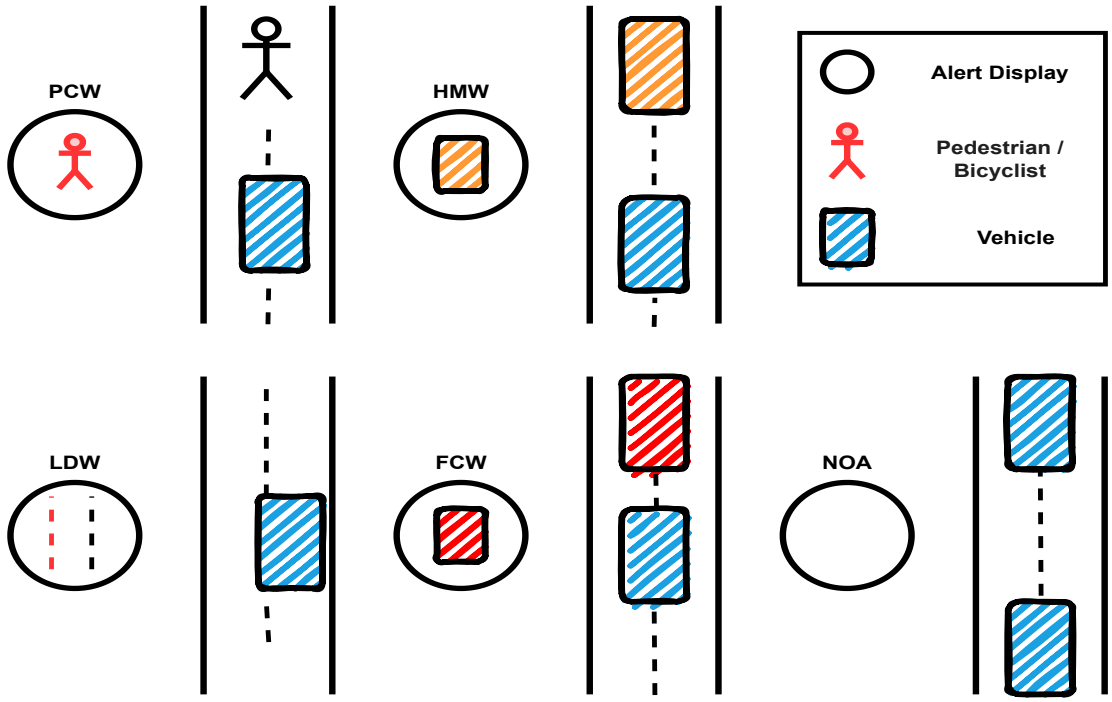

Figure 3.2 Alerts triggered by ADAS: **Pedestrian Collision Warning (PCW)** alerts the driver to potential collisions with pedestrians/bicyclists; **Forward Collision Warning (FCW)** indicates when the vehicle is too close to the one in front; **Lane Departure Warning (LDW)** notifies if the vehicle drifts out of its lane; **Headway Monitoring Warning (HMW)** warns of possible collisions with vehicles ahead; **No Obstacle Alert (NOA)** signifies no detected critical events. These alerts are crucial for enhancing driving safety by identifying and mitigating potential hazards.

3.3 Proposed Dataset

3.3.1 Sensors

3.3.1.1 Advance Driving Assistance System (ADAS)

In this work, the camera-based proprietary ADAS was utilized. The system is capable of detecting the presence of objects (stationary as well as moving) with type as well as their distance, including GPS coordinates around the vehicle, and accordingly sends visual and audio alarms. These alerts are given to the driver in the output unit if the vehicle is detected to be on an unsafe path (like lane departure), unsafely close to another vehicle/pedestrian/bicycle, etc. Based on these visual or audio alerts, the driver can potentially take corrective actions in driving to prevent or avoid an impending collision. Figure 3.1

Table 3.1 Comparisons of existing datasets based on action categories with respect to the ego vehicle, **where the IDD-CRS dataset stands out for having precise temporal annotations from ADAS. Clip lengths in IDD-CRS are determined by the speed of the ego vehicle at the time of alert triggers. Unlike other datasets, IDD-CRS clips are distance-aware, as they are formed based on ADAS alerts.**

Dataset	Action Categories				Method of Clip Extraction		
	Pedestrian	Vehicle	Ego Vehicle	Normal Driving	Temporal Boundary	Distance Aware	Speed Aware
JAAD [35]	✓	✗	✗	✗			
PIE [34]	✓	✗	✗	✗			
ROAD [41]	✓	✓	✗	✗	Manual	✗	✗
DADA [9]	✓	✓	✗	✗			
HDD [33]	✗	✓	✓	✗			
METEOR [4]	✓	✓	✓	✗			
IDD-CRS	✓	✓	✓	✓	**ADAS**	✓	✓

shows the device installed in the car from our study. The device has one AI-enabled camera (input) fitted on the dashboard of the car and is focused toward the road at an optimum angle to detect various features such as pedestrians, cyclists, lane departure, chances of a collision, road features, and has a display unit (output) which gives visual as well as audio alerts to the driver while driving. To store the huge geotagged data coming from the ADAS-equipped car, a centralized server is used. Figure 3.2 shows the visual scenarios where ADAS trigger alerts.

3.3.1.2 Camera

We used the DDpai X2S Pro RGB camera to record video, positioning it next to the ADAS device as shown in Figure 3.1. The camera captures front-facing footage at a resolution of 2560x1440 with a frame rate of 25fps. It features a lens system consisting of five optical lenses and one infrared filter lens, providing a 140-degree field of view and an F1.8 aperture. This setup ensures clear, wide-angle video, delivering high-quality performance that meets the requirements of our task.

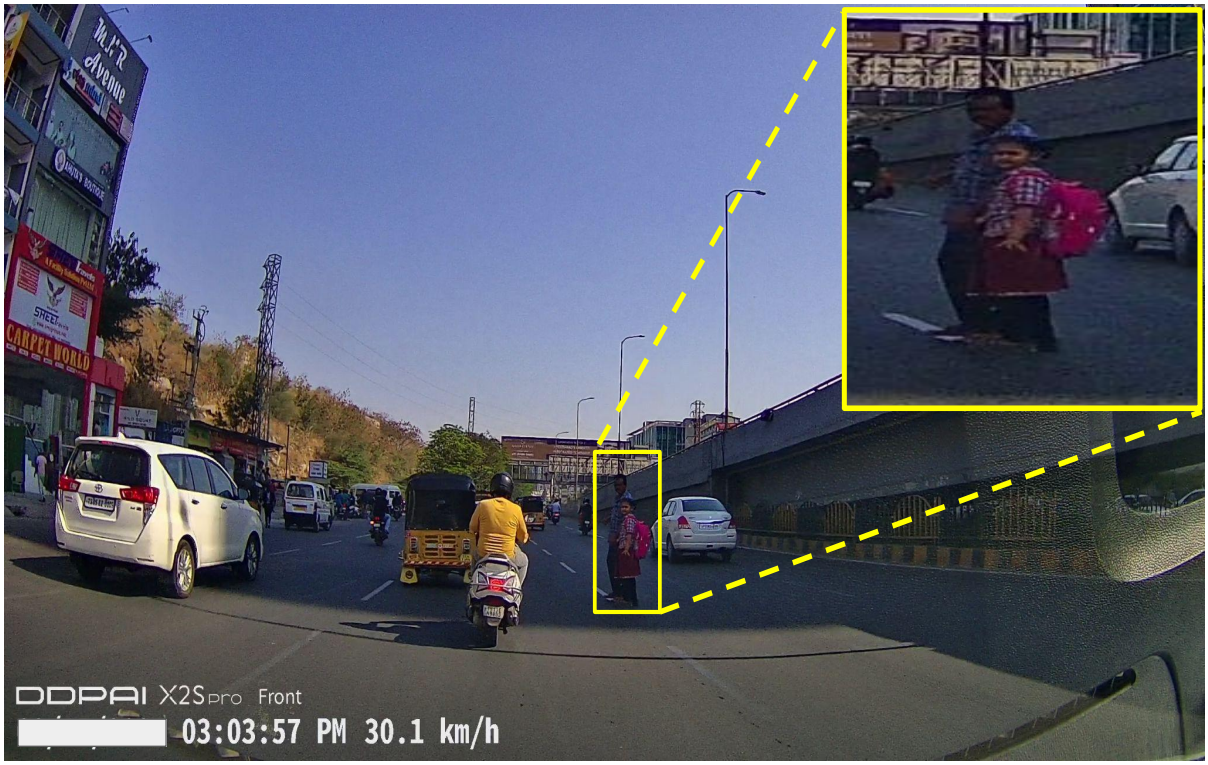

Figure 3.3 PCW scenario from IDD-CRS, with a zoomed-in section highlighting the agent that triggered the alert. The reason for this alert is detailed in Figure 3.2.

3.3.2 Data Acquisition and Statistics

We collected data using a custom setup in our vehicle, which was equipped with an ADAS system and an RGB camera installed at the front. An image of this setup from inside the vehicle is shown in Figure 3.1. The data collection took place in Hyderabad, India, a city with a diverse range of roads, from rural lanes to modern highways, providing diverse driving scenarios. To capture different driving conditions, we recorded data on various road types and at different times of the day, including morning, afternoon, evening, and night. Over 30 days, we accumulated approximately 90 hours of driving footage, ensuring that the dataset reflects natural and varied driving environments.

Each video recorded by the camera is one minute long, resulting in a total of 5,400 one-minute videos and 135,000 frames. However, not all videos contain alert scenes. We extracted specific clips where the ADAS triggered an alert. Figure 3.2 explains the scenarios in which ADAS triggers alerts, resulting in a total of 2,310 alert clips. These clips include various alert types, with 261 clips for FCW, 281 for PCW, 789 for HMW, 485 for LDW, and 489 for NOA. Figure 3.7 shows the distribution of alert clips, while Figure 3.8 illustrates the speed of the ego-vehicle when the alerts are triggered.

Figure 3.4 FCW scenario from IDD-CRS, with a zoomed-in section highlighting the agent that triggered the alert. The reason for this alert is detailed in Figure 3.2.

3.3.3 Clip Formation and Annotation

We extract clips based on the vehicle's speed at the time of the alert. For higher speeds, we capture longer distances, and for slower speeds, we use shorter distances. Most clips are 6 seconds long, including 3 seconds of footage before the alert, 1 second during the alert, and 2 seconds after the alert. The inclusion of 3 seconds of pre-alert footage is based on the vehicle's speed when the alert was triggered.

We collect data using ADAS detailing when each alert was triggered and the vehicle's speed at that time. This information helps us match the alerts with their corresponding timestamps in the video, allowing us to accurately extract the relevant clips. For the "No Obstacle Alert" (NOA) class, we randomly select clips from the video that do not contain any alerts. This method introduces hard negatives during training, helping the model avoid overfitting to just the critical action classes and improving its ability to effectively recognize critical actions. Figure **??** shows frames from IDD-CRS dataset clips

3.3.4 Comparsion with existing dataset

Road safety-related classes are typically categorized into vehicle, pedestrian, and ego-driver behavior. Existing datasets often fall short in providing precise temporal information needed to determine

Figure 3.5 HMW scenario from IDD-CRS, with a zoomed-in section highlighting the agent that triggered the alert. The reason for this alert is detailed in Figure 3.2.

when these elements are at risk, focusing primarily on pedestrian or ego-vehicle behavior with manually annotated clips that can be inaccurate. Our dataset addresses these gaps by leveraging ADAS for enhanced annotation accuracy. It provides a more reliable basis for safety studies by incorporating actions based on the distance between the ego vehicle and other road users and accounting for vehicle speed at the time of alerts. This approach ensures a comprehensive view of critical road scenarios and improves action recognition precision.

Unlike datasets such as JAAD [35] and PIE [34], which primarily focus on pedestrian safety with the assumption that road agents are always vulnerable, our dataset captures unsafe distances and triggers ADAS alerts, providing a more accurate reflection of real traffic scenarios. While datasets like ROAD [41], HDD [33], and METEOR [4] offer data for action recognition, they often focus on ego-driver behavior or interactions with other traffic agents. The DADA [9] dataset, on the other hand, consists of accident videos collected from YouTube. Our dataset fills this gap by emphasizing crucial aspects of road safety and leveraging ADAS for precise temporal annotations, resulting in greater accuracy and efficiency compared to manual methods. Additionally, we include a No Obstacle Alert / Normal driving action class, which serves as a hard negative to help models distinguish between normal actions and critical events. A detailed comparison of these aspects is shown in Table 3.1.

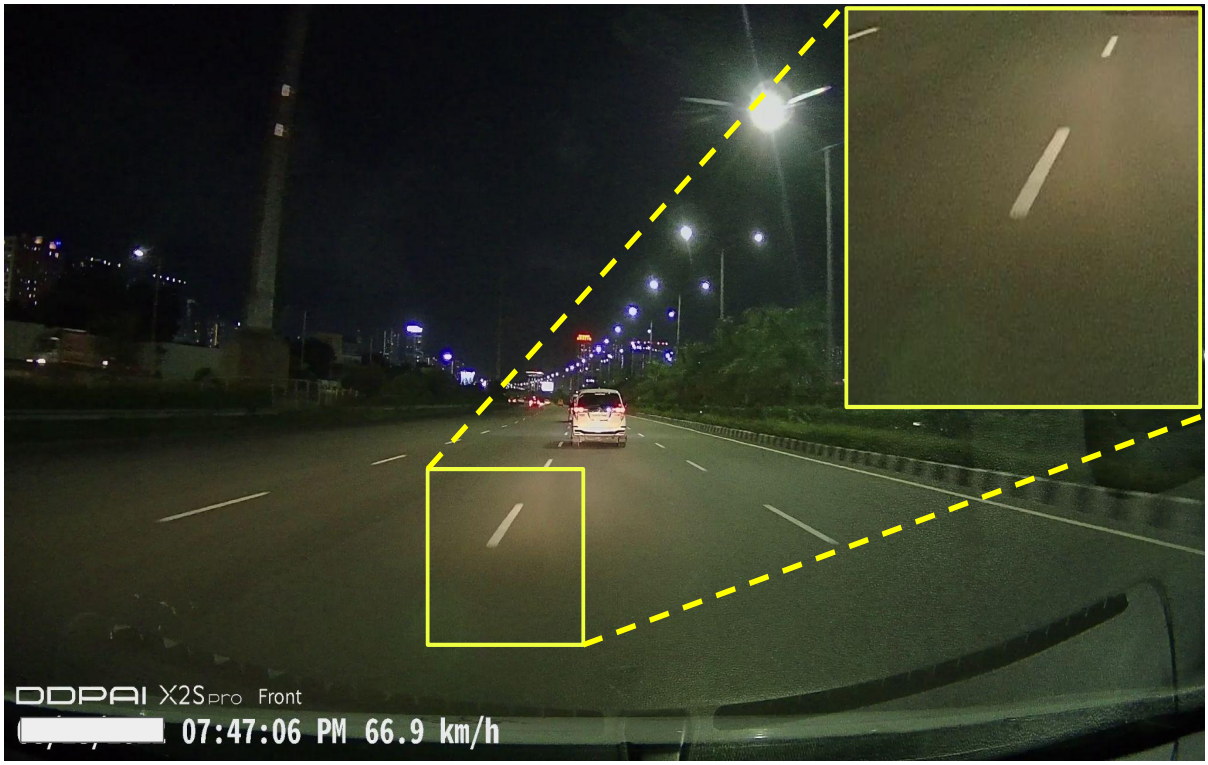

Figure 3.6 LDW scenario from IDD-CRS, with a zoomed-in section highlighting the agent that triggered the alert. The reason for this alert is detailed in Figure 3.2.

3.4 BENCHMARKS AND BASELINE RESULTS

We have discussed the dataset, the data collection process, and the annotation. In this section, we present an extensive analysis of IDD-CRS with existing methods to highlight the diversity and usefulness of data. We first discuss the experimental setup and then based on the evaluations, report the understanding about the dataset properties and behavior of different approaches.

3.4.1 Task on IDD-CRS dataset

Action Recognition: Given an action segment $A_i = [t_{si}, t_{ei}]$, we aim to classify the segment into its action class, where classes are defined as $C_a = \{(c_v \in C_V, c_n \in C_N)\}$, and c_n is the alert name. In IDD-CRS, we have five classes FCW, PCW, HMW, LDW and NOA.

Long-tail action recognition: refers to the challenge of classifying action classes that have a small number of clips compared to more common classes. In this context, HMW, LDW, and NOA have a large number of clips, making them frequent classes, whereas PCW and FCW have a relatively small number of clips, making them long-tail classes. Figure 3.7 shows the distribution of alert clips among different classes.

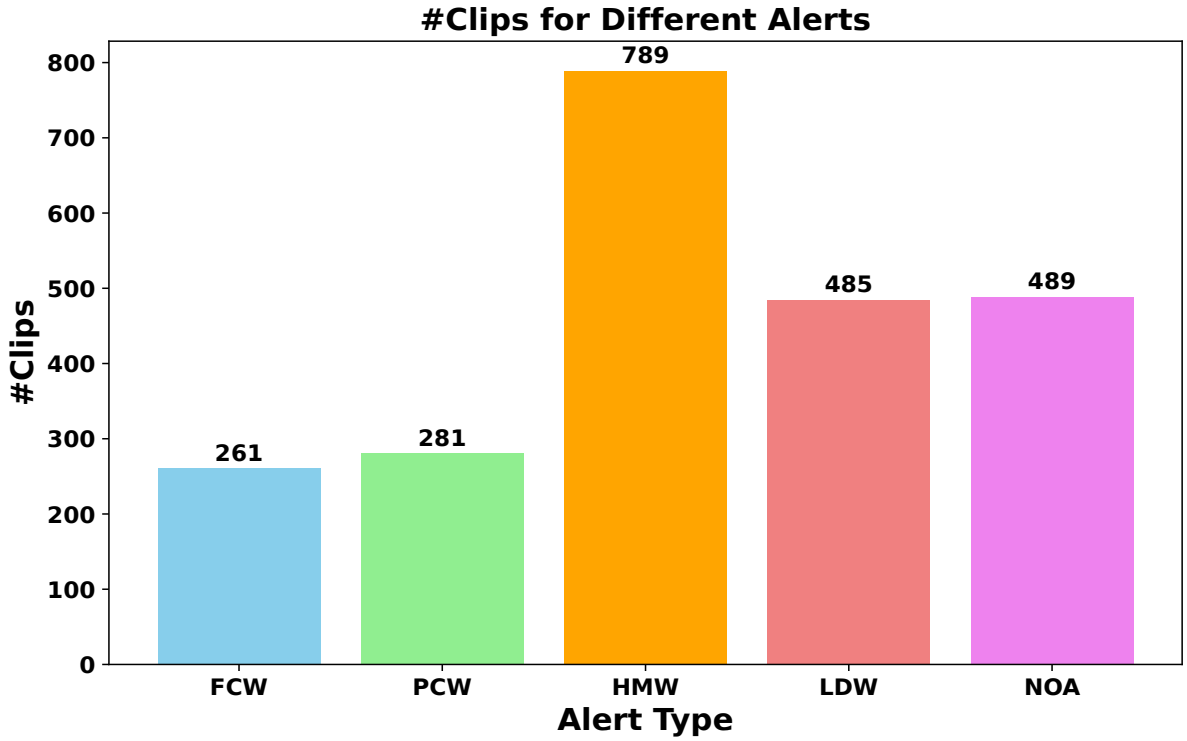

Figure 3.7 Distribution of video clips for the five different alerts in the IDD-CRS dataset. FCW and PCW have fewer clips compared to the other alerts, indicating a long-tail distribution of data in IDD-CRS.

3.4.2 Evaluation Metric

We use the mean Average Precision (mAP) as the evaluation metric. mAP is computed by averaging the Average Precision (AP) across all N action classes. For each class, AP is calculated as the area under the precision-recall curve, where precision is measured at different recall thresholds. The mAP formula is defined as:

$$\text{mAP} = \frac{1}{N} \sum_{i=1}^{N} \text{AP}_i$$

where N is the total number of classes, and AP_i represents the average precision for the i-th class. The higher the mAP score, the better the model's overall performance in distinguishing between different actions.

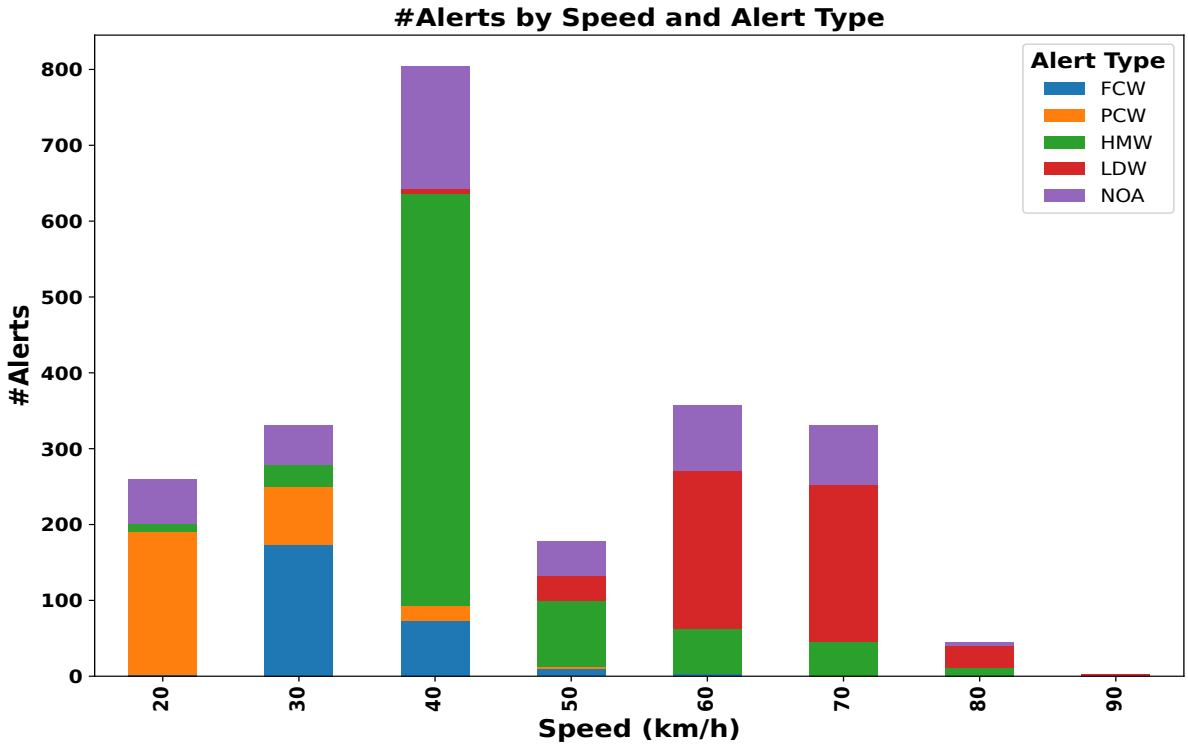

Figure 3.8 Speed distribution of the ego-vehicle at the moment alerts are triggered in the recorded clips. IDD-CRS captures critical scenarios across all speeds. Alerts are not considered for speeds less than 20 km/h, as no agents are in danger at such speeds. For speeds above 20 km/h, the speed is rounded up to the nearest integer divisible by 10 (for this plot). Most FCW and PCW alerts occur at speeds below 40 km/h, while LDW alerts trigger at speeds above 50 km/h. HMW and NOA alerts are present across all speeds.

3.4.3 Data Augmentation

We apply rectangular cropping as a data augmentation technique as shown in Figure 3.9. This process involves cutting out a central rectangular area from the original image. By focusing on a central portion of the image, this technique reduces the impact of less relevant areas around the edges. The result is a more focused dataset that can improve model accuracy and robustness. This method is particularly effective in emphasizing the central features of the image, which are often the most important for classification and analysis.

Figure 3.9 Augmented image: The height is reduced to 0.5 times the original height, while the left and right widths are each reduced to 0.12 times the original width.

3.4.4 Baseline and Implementation Details

3.4.4.1 Action Recognition

We experimented with several well-established video action recognition backbones using standard training methods and cross-entropy loss. The backbones included CNN-based architectures such as C3D [46], I3D [3], X3D [10], SlowFast [11], and the transformer-based backbone MotionFormer [30], all of which have shown considerable success in general human action recognition tasks. The C3D model was pre-trained on the Sports-1M and UCF101 datasets, while the I3D model with a ResNet-50 backbone was pre-trained on the Kinetics-400 human action dataset. Similarly, the X3D model (x3d-m) and the SlowFast model, both with ResNet-50 backbones, were pre-trained on Kinetics-400. Lastly, the MotionFormer model, with a ViT backbone, was pre-trained on the EpicKitchens-100 dataset. We fine-tuned these pre-trained models on our IDD-CRS dataset to evaluate their performance in recognizing complex driving behaviors.

3.4.4.2 Implementation Details

We conducted all experiments on a system with four NVIDIA 3080Ti GPUs using PyTorch. We utilized the Adam optimizer and tuned the training parameters for optimal performance while maintaining the backbone inputs as specified in their respective papers. The C3D model uses 16-frame clips at a 112×112 resolution, I3D uses 64-frame clips at a 224×224 resolution, X3D (xmd-m) uses 16-frame clips at a 256×256 resolution, SlowFast uses 32-frame clips at a 256×256 resolution with a SlowFast alpha of 4, and MotionFormer uses 16-frame clips at a 224×224 resolution. We applied the frame augmentation techniques described in Section 3.4.3 and depicted in Figure 3.9, resulting in performance improvements detailed in Section 3.4.5.

3.4.4.3 Action Recognition + Long tail Methods

Real-world data, particularly in the traffic domain, often exhibits a long-tail distribution. To address this characteristic, we conducted extensive experiments with existing methods for long-tail video classification. We used the best-performing backbone results as our baseline and applied this top-performing backbone to these methods.

- **CE (Cross-Entropy)**: The standard cross-entropy loss function, trained using instance-balanced sampling. Each instance in the dataset is treated equally during training, without any adjustments for class imbalances.
- **EQL (Equalization Loss)** [43]: Like CE, this method also uses instance-balanced sampling but introduces an Equalization Loss. This loss function reduces the penalties for incorrectly classifying head (frequent) classes as tail (rare) classes, addressing class imbalance.
- **cRT (Classifier Retraining)** [13]: Classifier Retraining is now a standard method in handling class imbalance. It first trains the model using instance-balanced sampling, then resets the classifier and re-trains it with class-balanced sampling. This ensures the model pays equal attention to both frequent and rare classes during classification.
- **Mixup** [53]: This technique combines pairs of training samples and their labels. Mixing up the input data, helps the model generalize better by introducing new training examples that are weighted combinations of existing ones.
- **Framestack** [54]: In this approach, video frames are mixed based on a running total of class average precision. It aims to improve the overall precision of action recognition tasks by giving weight to classes based on their performance during training.
- **LMR** [31]: A mixed reconstruction approach uses pairwise feature similarities to reconstruct video features, with few-shot samples excluded. Pairwise label mixing enhances feature diversity by combining video samples within a batch. The reconstructed and mixed features are passed to the classifier, improving the recognition of underrepresented classes.

3.4.5 Results and Analysis

Our experiments with various video backbones reveal that the performance of these models differs significantly under different conditions. As shown in Table 3.2, which presents results without data augmentation, the SlowFast backbone achieves the highest overall mAP of 67.7, excelling particularly in the LDW and NOA categories. This demonstrates its superior capability in handling complex scenarios, despite the lack of data augmentation. Other backbones like X3D and C3D also show strong performance but do not surpass SlowFast in overall effectiveness for the given categories.

Table 3.3 displays the performance of the same video backbones with data augmentation. Here, the SlowFast backbone again stands out, achieving the highest overall mAP of 70.9, which is a 3.3 gain from SlowFast without augmentation. This indicates that data augmentation significantly enhances model performance, allowing SlowFast to perform better across various action categories. All other backbones also benefits from data augmentation. I3D and X3D also benefit from data augmentation, but the SlowFast model consistently outperforms the others in both overall mAP and category-specific performance.

Further analysis of long-tail methods applied to the SlowFast backbone, as shown in Table 3.4, highlights several advancements. The SlowFast model with the LMR training method increases the mAP from 70.9 to 72.0, representing a gain of 1.1. However, there is no significant improvement in the long-tail classes FCW and PCW. The cRT method shows a gain of 1.8 in mAP for the PCW class but underperforms in other classes. Overall, while the LMR method boosts the model's overall performance, it does not significantly improve performance for the long-tail classes compared to the baseline.

We also tested our baseline model on the lane change class (Right / Left Lane Change) of the HDD dataset, as this was the only class in the existing datasets that matched our labels. The Average Precision for lane changes in HDD was recorded as 76.3. Despite the differences in data distribution, as HDD was collected in a structured environment, our model performed well. This demonstrates that the proposed dataset and model can be effectively applied to any geographical area.

Table 3.2 Baseline results for action recognition **without data augmentation**

Video Backbone	FCW	PCW	HMW	LDW	NOA	Overall mAP
I3D [3]	39.2	66.5	70.1	83.3	48.9	61.6
Slowfast [11]	39.4	**77.6**	**76.9**	89.1	**55.7**	**67.7**
X3D [10]	45.5	72.2	73.3	**90.9**	52.7	66.0
C3D [46]	**56.6**	67.8	73.0	86.7	50.4	66.9
Motionformer [30]	46.9	62.4	64.1	62.6	29.3	53.0

Table 3.3 Baseline results for action recognition **with data augmentation**

Video Backbone	FCW	PCW	HMW	LDW	NOA	Overall mAP
I3D	50.8	**78.8**	69.8	90.2	57.8	69.5
Slowfast	51.2	77.2	75.8	**92.4**	**57.9**	**70.9**
X3D	56.0	76.7	**80.2**	87.3	45.1	69.1
C3D	**59.1**	68.5	77.3	86.6	41.0	66.5
Motionformer	45.6	67.4	61.5	67.7	34.8	55.4

Table 3.4 Performance of the best video backbone, enhanced with various **Long-tail Methods**

Backbone	Method	FCW	PCW	HMW	LDW	NOA	mAP
Slowfast	**CE**	51.2	77.2	75.8	92.4	57.9	70.9
Slowfast	**EQL**	48.4	76.8	69.2	91.2	53.4	67.8
	Framestack	50.6	73.1	76.0	92.4	58.2	70.1
	cRT	**53.0**	73.1	77.4	93.2	**60.0**	71.3
	Mixup	50.2	74.6	**78.5**	**94.4**	59.1	71.4
	LMR	51.3	**77.3**	78.0	93.6	59.9	**72.0**

3.5 Conclusion

In conclusion, the IDD-CRS dataset addresses critical gaps in road safety research by incorporating both vehicle and pedestrian behaviors in diverse, high-risk traffic scenarios. By utilizing ADAS for accurate temporal annotations, this dataset offers a more reliable foundation for safety analysis compared to manually annotated datasets. With 90 hours of video footage, consisting of 5400 one-minute videos and 135,000 frames, IDD-CRS provides a comprehensive view of road interactions, including newly introduced vehicle-related classes and hard negative examples to enhance model robustness. Our benchmarks on action recognition and long-tail methods highlight the current limitations of existing models, underscoring the need for continued improvements in road safety technology. This dataset sets the stage for future innovations aimed at mitigating risks for all road users.

Chapter 4

Conclusions

This thesis addresses the urgent issue of road safety by introducing innovative solutions that combine predictive modeling for accident-prone zones with the creation of a specialized dataset for action recognition in critical traffic scenarios. The first major contribution is the development of a predictive framework using geo-tagged alert data gathered from the Advanced Driver Assistance System (ADAS) device installed in a fleet of city buses. This data, collected over an extensive period, enables the identification of emerging accident-prone zones in the city of Nagpur, India. By employing a nonparametric Kernel Density Estimation (KDE) method along with a recall-based metric and Earth Mover Distance (EMD) analysis, this model offers a proactive, data-driven approach to predicting high-risk locations. Unlike traditional models that rely on post-incident accident data, this framework utilizes real-time alert data, identifying critical areas based on traffic behavior patterns. This approach supports early intervention strategies, equipping civic authorities with actionable insights for deploying traffic-calming measures and structural changes to reduce accidents before they occur. The framework's results demonstrate a significant improvement over conventional techniques, showing that KDE- and EMD-based approaches can detect emerging accident-prone areas, thereby enabling urban planners to address safety concerns dynamically and adaptively.

The second key contribution is the introduction of IDD-CRS (Indian Driving Dataset for Critical Road Scenarios), a dataset designed to capture a wide range of complex vehicle and pedestrian interactions in unstructured and high-risk driving environments. With over 90 hours of video data comprising 5400 individual one-minute-long videos and 135,000 frames, IDD-CRS is a large-scale, meticulously annotated dataset collected using ADAS and dash cameras. By incorporating ADAS technology, the dataset offers precisely defined temporal annotations that accurately capture the start and end points of critical events, improving the reliability of safety analyses. Unlike previous datasets that primarily focus on either pedestrian or ego-vehicle behaviors, IDD-CRS provides diverse scenarios that reflect real-world driving complexities, including high-speed lane changes, close proximity vehicle interactions, and abrupt pedestrian crossings. The dataset introduces new vehicle-related and hard negative classes, providing a richer context for training models to recognize and respond to both frequent and rare safety-critical

events. This unique structure also facilitates benchmarks for both action recognition and long-tail action recognition tasks, addressing the challenges posed by infrequent but hazardous actions that are crucial for road safety applications. Initial experiments on the IDD-CRS dataset using popular action recognition models reveal significant limitations, underscoring the need for more advanced and context-sensitive recognition techniques to accurately identify risky behaviors in diverse traffic scenarios.

In conclusion, this thesis delivers a novel framework that integrates predictive modeling and detailed action recognition data to advance proactive road safety measures. This work significantly contributes to both road safety management and action recognition research, providing tools and insights that can be adapted for broader applications in urban planning, autonomous driving, and intelligent transportation systems. While the current work focuses on a specific geographic context and data sources, future research could expand these methodologies to different regions, refine long-tail recognition capabilities, and incorporate real-time processing for integration into autonomous systems. This thesis not only lays the groundwork for innovative safety measures in traffic management but also presents a scalable approach for continuous learning and adaptation in rapidly evolving road environments, ultimately contributing to the global effort in accident prevention and traffic safety.

Related Publications

- **Ravi Shankar Mishra**, Dev Singh Thakur, Anbumani Subramanian, Mukti Advani, S Velmurugan, Juby Jose, C V Jawahar and Ravi Kiran Sarvadevabhatla, **"Enhancing Road Safety: Predictive Modeling of Accident-Prone Zones with ADAS-Equipped Vehicle Fleet Data"**, in proceedings of *IEEE Intelligent Vehicles Symposium (IV)*, 2024.
- **Ravi Shankar Mishra**, Anbumani Subramanian, C V Jawahar and Ravi Kiran Sarvadevabhatla **"IDD-CRS: A Comprehensive Video Dataset for Critical Road Scenarios in Unstructured Environments"**, in under review *IEEE International Conference on Robotics and Automation (ICRA)*, 2025.

Related co-author publications:

- Chirag Parikh, **Ravi Shankar Mishra**, Rohan Chandra and Ravi Kiran Sarvadevabhatla, **"Transfer-LMR: Heavy-Tail Driving Behavior Recognition in Diverse Traffic Scenarios"**, in proceedings of *arXiv preprint arXiv:2405.05354*, 2025.

Chapter 5

Future Work

- **Analysis of Alerts Across Cities:** Future work can focus on analyzing the alerts generated in different cities to identify common locations and the primary causes of alert generation. This information could be valuable for improving infrastructure and safety measures in various regions.
- **Development of Efficient Action Detection Model:** One of the key future directions is to leverage the IDD-CRS dataset to develop an efficient method capable of running on edge devices, such as the NVIDIA Jetson Nano, for real-time action detection. This would facilitate online action detection in constrained environments.
- **Incorporating Multimodal Data:** The performance of recent video backbones has not been fully optimized. Exploring the use of additional modalities, such as depth information or optical flow, in these networks could lead to better performance in action recognition tasks.
- **Addressing Long-Tail Distribution in Traffic Datasets:** Many traffic-related datasets exhibit a long-tail distribution, with rare events being underrepresented. To improve model performance, future work should explore techniques to address this imbalance, such as synthetic data generation or specialized loss functions.
- **Domain Adaptation for Pretraining:** Traffic-related datasets often suffer from domain shift when pretrained models are used. To mitigate the risk of poor performance due to pretraining on irrelevant datasets, future research could focus on fine-tuning models specifically on traffic-related data or exploring domain adaptation techniques.

Bibliography

[1] A. Abdulhafedh et al. Crash frequency analysis. *Journal of Transportation Technologies*, 6(04):169, 2016. 5, 6, 17

[2] M. Bíl, R. Andrášik, and Z. Janoška. Identification of hazardous road locations of traffic accidents by means of kernel density estimation and cluster significance evaluation. *Accident Analysis & Prevention*, 55:265–273, 2013. 6

[3] J. Carreira and A. Zisserman. Quo vadis, action recognition? a new model and the kinetics dataset. In *proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 6299–6308, 2017. 25, 34, 36

[4] R. Chandra, X. Wang, M. Mahajan, R. Kala, R. Palugulla, C. Naidu, A. Jain, and D. Manocha. Meteor: A dense, heterogeneous, and unstructured traffic dataset with rare behaviors. In *2023 IEEE International Conference on Robotics and Automation (ICRA)*, pages 9169–9175. IEEE, 2023. 23, 24, 27, 30

[5] Q. Chen, X. Song, H. Yamada, and R. Shibasaki. Learning deep representation from big and heterogeneous data for traffic accident inference. In *Thirtieth AAAI conference on artificial intelligence*, 2016. 6

[6] J. Cui, Z. Zhong, S. Liu, B. Yu, and J. Jia. Parametric contrastive learning. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 715–724, 2021. 25

[7] D. Eisenberg. The mixed effects of precipitation on traffic crashes. *Accident analysis & prevention*, 36(4):637–647, 2004. 5

[8] A. K. Erenler and B. Gümüş. Analysis of road traffic accidents in turkey between 2013 and 2017. *Medicina*, 55(10):679, 2019. 5

[9] J. Fang, D. Yan, J. Qiao, J. Xue, and H. Yu. Dada: Driver attention prediction in driving accident scenarios. *IEEE transactions on intelligent transportation systems*, 23(6):4959–4971, 2021. 25, 27, 30

[10] C. Feichtenhofer. X3d: Expanding architectures for efficient video recognition. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2020. 25, 34, 36

[11] C. Feichtenhofer, H. Fan, J. Malik, and K. He. Slowfast networks for video recognition. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 6202–6211, 2019. 25, 34, 36

[12] S. Hashimoto, S. Yoshiki, R. Saeki, Y. Mimura, R. Ando, and S. Nanba. Development and application of traffic accident density estimation models using kernel density estimation. *Journal of traffic and transportation engineering (English edition)*, 3(3):262–270, 2016. 6

[13] B. Kang, S. Xie, M. Rohrbach, Z. Yan, A. Gordo, J. Feng, and Y. Kalantidis. Decoupling representation and classifier for long-tailed recognition. *arXiv preprint arXiv:1910.09217*, 2019. 25, 35

[14] M. Karimi, J. Hedner, H. Häbel, O. Nerman, and L. Grote. Sleep apnea related risk of motor vehicle accidents is reduced by continuous positive airway pressure: Swedish traffic accident registry data. *Sleep*, 38(3):341–349, 2015. 5

[15] S. S. A. Kazmi, M. Ahmed, R. Mumtaz, and Z. Anwar. Spatiotemporal clustering and analysis of road accident hotspots by exploiting gis technology and kernel density estimation. *The Computer Journal*, 65(2):155–176, 2022. 6

[16] K. D. Kusano and H. C. Gabler. Safety benefits of forward collision warning, brake assist, and autonomous braking systems in rear-end collisions. *IEEE Transactions on Intelligent Transportation Systems*, 13(4):1546–1555, 2012. 6

[17] J. Li, L. Sun, J. Chen, M. Tomizuka, and W. Zhan. A safe hierarchical planning framework for complex driving scenarios based on reinforcement learning. In *2021 IEEE International Conference on Robotics and Automation (ICRA)*, pages 2660–2666. IEEE, 2021. 24

[18] M. Li, Y.-m. Cheung, and Y. Lu. Long-tailed visual recognition via gaussian clouded logit adjustment. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 6929–6938, 2022. 25

[19] S. Li, K. Gong, C. H. Liu, Y. Wang, F. Qiao, and X. Cheng. Metasaug: Meta semantic augmentation for long-tailed visual recognition. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 5212–5221, 2021. 25

[20] T. Li, P. Cao, Y. Yuan, L. Fan, Y. Yang, R. S. Feris, P. Indyk, and D. Katabi. Targeted supervised contrastive learning for long-tailed recognition. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 6918–6928, 2022. 25

[21] H. Liu, L. Zhang, S. K. S. Hari, and J. Zhao. Safety-critical scenario generation via reinforcement learning based editing. In *2024 IEEE International Conference on Robotics and Automation (ICRA)*, pages 14405–14412. IEEE, 2024. 24

[22] Z. Liu, Z. Miao, X. Zhan, J. Wang, B. Gong, and X. Y. Stella. Open long-tailed recognition in a dynamic world. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 46(3):1836–1851, 2022. 25

[23] D. Lord and P. Y.-J. Park. Investigating the effects of the fixed and varying dispersion parameters of poisson-gamma models on empirical bayes estimates. *Accident Analysis & Prevention*, 40(4):1441–1457, 2008. 5, 6, 17

[24] Y. Lv, S. Tang, and H. Zhao. Real-time highway traffic accident prediction based on the k-nearest neighbor method. In *2009 international conference on measuring technology and mechatronics automation*, volume 3, pages 547–550. IEEE, 2009. 6

[25] M. Lyssenko, P. Pimplikar, M. Bieshaar, F. Nozarian, and R. Triebel. A safety-adapted loss for pedestrian detection in autonomous driving. In *2024 IEEE International Conference on Robotics and Automation (ICRA)*, pages 4428–4434. IEEE, 2024. 24

[26] C. Min, W. Jiang, D. Zhao, J. Xu, L. Xiao, Y. Nie, and B. Dai. Orfd: A dataset and benchmark for off-road freespace detection. In *2022 international conference on robotics and automation (ICRA)*, pages 2532–2538. IEEE, 2022. 24

[27] S. Moosavi, M. H. Samavatian, S. Parthasarathy, R. Teodorescu, and R. Ramnath. Accident risk prediction based on heterogeneous sparse data: New dataset and insights. In *Proceedings of the 27th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*, pages 33–42, 2019. 5

[28] K. Mukoya, E. Weng, R. Choudhury, and K. Kitani. Jaywalkervr: A vr system for collecting safety-critical pedestrian-vehicle interactions. In *2024 IEEE International Conference on Robotics and Automation (ICRA)*, pages 9600–9607. IEEE, 2024. 24

[29] G. M. Muktadir and J. Whitehead. Adaptive pedestrian agent modeling for scenario-based testing of autonomous vehicles through behavior retargeting. In *IEEE Int. Conf. Robot. Automat.(ICRA)*, 2024. 24

[30] M. Patrick, D. Campbell, Y. Asano, I. Misra, F. Metze, C. Feichtenhofer, A. Vedaldi, and J. F. Henriques. Keeping your eye on the ball: Trajectory attention in video transformers. *Advances in neural information processing systems*, 34:12493–12506, 2021. 25, 34, 36

[31] T. Perrett, S. Sinha, T. Burghardt, M. Mirmehdi, and D. Damen. Use your head: Improving long-tail video recognition. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 2415–2425, 2023. 25, 35

[32] M.-H. Pham, A. Bhaskar, E. Chung, and A.-G. Dumont. Random forest models for identifying motorway rear-end crash risks using disaggregate data. In *13th International IEEE Conference on Intelligent Transportation Systems*, pages 468–473. IEEE, 2010. 6

[33] V. Ramanishka, Y.-T. Chen, T. Misu, and K. Saenko. Toward driving scene understanding: A dataset for learning driver behavior and causal reasoning. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 7699–7707, 2018. 23, 24, 27, 30

[34] A. Rasouli, I. Kotseruba, T. Kunic, and J. K. Tsotsos. Pie: A large-scale dataset and models for pedestrian intention estimation and trajectory prediction. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 6262–6271, 2019. 23, 24, 27, 30

[35] A. Rasouli, I. Kotseruba, and J. K. Tsotsos. Are they going to cross? a benchmark dataset and baseline for pedestrian crosswalk behavior. In *Proceedings of the IEEE International Conference on Computer Vision Workshops*, pages 206–213, 2017. 23, 24, 27, 30

[36] A. Reichenbach and J.-E. Navarro-B. A model for traffic incident prediction using emergency braking data. In *2021 IEEE Intelligent Vehicles Symposium (IV)*, pages 22–27. IEEE, 2021. 6

[37] H. Ren, Y. Song, J. Wang, Y. Hu, and J. Lei. A deep learning approach to the citywide traffic accident risk prediction. In *2018 21st International Conference on Intelligent Transportation Systems (ITSC)*, pages 3346–3351. IEEE, 2018. 6

[38] Y. Rubner, C. Tomasi, and L. J. Guibas. The earth mover's distance as a metric for image retrieval. *International journal of computer vision*, 40(2):99, 2000. 14

[39] B. Ryder, B. Gahr, P. Egolf, A. Dahlinger, and F. Wortmann. Preventing traffic accidents with in-vehicle decision support systems-the impact of accident hotspot warnings on driver behaviour. *Decision support systems*, 99:64–74, 2017. 6, 17

[40] B. Sharma, V. K. Katiyar, and K. Kumar. Traffic accident prediction model using support vector machines with gaussian kernel. In *Proceedings of fifth international conference on soft computing for problem solving*, pages 1–10. Springer, 2016. 6

[41] G. Singh, S. Akrigg, M. Di Maio, V. Fontana, R. J. Alitappeh, S. Khan, S. Saha, K. Jeddisaravi, F. Yousefi, J. Culley, et al. Road: The road event awareness dataset for autonomous driving. *IEEE transactions on pattern analysis and machine intelligence*, 45(1):1036–1054, 2022. 27, 30

[42] Z. Sun, D. Wang, X. Gu, M. Abdel-Aty, Y. Xing, J. Wang, H. Lu, and Y. Chen. A hybrid approach of random forest and random parameters logit model of injury severity modeling of vulnerable road users involved crashes. *Accident Analysis & Prevention*, 192:107235, 2023. 6

[43] J. Tan, X. Lu, G. Zhang, C. Yin, and Q. Li. Equalization loss v2: A new gradient balance approach for long-tailed object detection. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 1685–1694, 2021. 25, 35

[44] G. R. Terrell and D. W. Scott. Variable kernel density estimation. *The Annals of Statistics*, pages 1236–1265, 1992. 9

[45] L. Thakali, T. J. Kwon, and L. Fu. Identification of crash hotspots using kernel density estimation and kriging methods: a comparison. *Journal of Modern Transportation*, 23:93–106, 2015. 6

[46] D. Tran, L. Bourdev, R. Fergus, L. Torresani, and M. Paluri. Learning spatiotemporal features with 3d convolutional networks. In *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, December 2015. 25, 34, 36

[47] J. Wang, A. Pun, J. Tu, S. Manivasagam, A. Sadat, S. Casas, M. Ren, and R. Urtasun. Advsim: Generating safety-critical scenarios for self-driving vehicles. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 9909–9918, 2021. 25

[48] K. S. Wowo, R. Dadwal, T. Graen, A. Fiege, M. Nolting, W. Nejdl, E. Demidova, and T. Funke. Using vehicle data to enhance prediction of accident-prone areas. In *2022 IEEE 25th International Conference on Intelligent Transportation Systems (ITSC)*, pages 2450–2456. IEEE, 2022. 6

[49] X. Xie, C. Zhang, Y. Zhu, Y. N. Wu, and S.-C. Zhu. Congestion-aware multi-agent trajectory prediction for collision avoidance. In *2021 IEEE International Conference on Robotics and Automation (ICRA)*, pages 13693–13700. IEEE, 2021. 24

[50] Z. Xie and J. Yan. Kernel density estimation of traffic accidents in a network space. *Computers, environment and urban systems*, 32(5):396–406, 2008. 6, 18

[51] S. Yao, J. Wang, L. Fang, and J. Wu. Identification of vehicle-pedestrian collision hotspots at the micro-level using network kernel density estimation and random forests: A case study in shanghai, china. *Sustainability*, 10(12):4762, 2018. 6

[52] L. Zha, D. Lord, and Y. Zou. The poisson inverse gaussian (pig) generalized linear regression model for analyzing motor vehicle crash data. *Journal of Transportation Safety & Security*, 8(1):18–35, 2016. 5

[53] H. Zhang. mixup: Beyond empirical risk minimization. *arXiv preprint arXiv:1710.09412*, 2017. 25, 35

[54] Y. Zhang, B. Hooi, L. Hong, and J. Feng. Test-agnostic long-tailed recognition by test-time aggregating diverse experts with self-supervision. *arXiv preprint arXiv:2107.09249*, 2(5):6, 2021. 25, 35

[55] Z. Zhong, J. Cui, S. Liu, and J. Jia. Improving calibration for long-tailed recognition. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 16489–16498, 2021. 25