

Income-Based Disparities in Fatal Interstate Crashes: An Association Rule Analysis of Contributing Factors from the NHTSA FARS Dataset

Mohammad Reza Abbaszadeh Lima*

Independent Researcher

Ewing, NJ, 08628

Email: mr.abbaszadelima@gmail.com

ORCID ID: 0009-0008-7728-8811

Seyed Amirhossein Naeimaeimousavi

PhD Candidate, Business Analytics and Information Systems

Harbert College of Business, Auburn University

405 West Magnolia Ave, Auburn, AL 36849-5337

Phone: (334) 844-4030

Email: szn0084@auburn.edu

ORCID ID: 0009-0001-7213-4003

*Corresponding Author

ABSTRACT

Income-based disparities in fatal Interstate crashes across the United States were investigated using data from the NHTSA Fatality Analysis Reporting System (FARS) for 2019–2023. Machine learning and association rule mining techniques were employed to examine how crash characteristics vary across socioeconomic strata. An XGBoost model was used to identify 25 key variables predictive of fatality count, with crash severity indicators, infrastructure characteristics, and emergency response lag times ranking highest in importance.

To uncover deeper crash patterns, the Apriori algorithm was applied to crash data stratified by income quintiles, adjusted for regional price parity. Over 15,000 significant rules were generated, with rule complexity increasing in higher-income areas. Rule patterns highlighted income-related differences in crash types, particularly pedestrian-involved crashes, same-direction collisions, and front-to-rear impacts, as well as variations in emergency response times and infrastructure conditions.

The findings reveal that fatal crash patterns are not uniform across income groups on Interstate highways. Lower-income areas exhibited concentrated high-risk crash factors, while higher-income regions demonstrated more diverse crash scenarios. Emergency response lag times and infrastructure disparities further indicated systemic inequities that may influence crash outcomes.

The value of integrating socioeconomic context into traffic safety research was demonstrated. Association rule mining enabled interpretable, income-specific insights that can inform more equitable policy and resource allocation. Tailored safety interventions based on regional economic characteristics may enhance effectiveness and help address persistent disparities in fatal crash outcomes.

Keywords: Fatal Interstate crashes, Association rule mining, Crash contributing factors, Socioeconomic factors, Traffic safety

INTRODUCTION

Examining the factors that contribute to fatal traffic incidents is a vital area of transportation safety research, requiring a comprehensive approach that accounts for driver behavior, infrastructure, and socioeconomic influences (1, 2).

A thorough understanding of these interrelated factors is essential for developing and applying evidence-based strategies that effectively reduce roadway fatalities and improve overall transportation safety (2–4). The complexity of crash causation necessitates an in-depth examination, acknowledging the interplay of human factors, vehicular characteristics, and the external environment (2, 5).

Existing literature highlights the disproportionate crash risks in low-income areas (6). Studies reveal that individuals residing in economically disadvantaged areas face an increased likelihood of experiencing fatal outcomes in vehicular crashes (6).

Furthermore, spatial analyses have established a correlation between crash frequency, income levels, and urban design, highlighting the pivotal role of accessible transit infrastructure in diminishing crash risks within economically disadvantaged areas. Conversely, other investigations have proposed a strong association between minority status and motor vehicle crashes, emphasizing the convergence of socioeconomic disadvantage with an increased susceptibility to fatal incidents (7).

Regression models in previous studies have identified income, demographic characteristics, and technological factors as significant predictors of fatal motor vehicle crash rates, while additional research suggests an inverse relationship between education, income, and involvement in fatal crashes (8). Such findings highlight the variate factors that contribute to fatal crashes, underscoring the need for a comprehensive and integrated approach to road safety interventions. Understanding these relationships is crucial for informed decision-making, policy, and infrastructure improvements aimed at reducing crash fatalities and improving overall road safety outcomes.

This study investigates the disparities in factors contributing to fatal crashes on Interstates across areas with varying income levels in the United States, aiming to explain the relationship between socioeconomic status and road safety outcomes. By analyzing crash data, socioeconomic demographics, especially the income level of the area of crash location, and roadway conditions, this research seeks to determine whether income disparities affect the factors that cause fatal crashes. This study also examines driver behavior, infrastructure, and access to safety resources as potential contributing factors. This investigation recognizes the well-documented influence of demographic and environmental factors on road crash rates and incorporates regional income level as a meaningful indicator of socioeconomic context, highlighting the importance of including such variables in modeling (9).

Factors such as advanced age of drivers, non-compliance with seatbelt usage, the event of ejection from the vehicle during a crash, alcohol impairment, excessive speed, and the temporal occurrence of crashes during early morning hours are consistently documented as significant contributors to the elevated risks of severe injuries and fatalities in traffic incidents (10, 11). Such factors are further compounded by vehicle and driver-level variables, including road type, vehicle compatibility, driver age, vehicle type, drunk driving, driving an unregistered vehicle, and driving experience, all of which significantly affect injury severity (12). Gaining insight into the intricate relationships among these factors and their connection to income levels is crucial for designing effective, targeted strategies to reduce fatal crashes on Interstate highways. This study seeks to deepen the understanding of the socioeconomic factors of road safety, specifically those directly linked to the income level of crash locations, offering a basis for future research and the development of data-driven policies aimed at reducing disparities in fatal crash rates across different income groups.

LITERATURE REVIEW

A large body of research has examined the contributing factors to fatal crashes, including driver behavior, road conditions, roadway geometry, socioeconomic status, and demographic characteristics. A substantial body of research confirms that low-income communities experience disproportionately high crash risks. For example, the findings of study (13) found individuals living in lower-income neighborhoods are at greater risk of fatal outcomes in crashes, standing in contrast to a wide range of

studies. Complementing this, study (8) found that crash frequency exhibits spatial patterns linked to both income levels and urban design, emphasizing that access to transit infrastructure plays a key role in reducing crash risks in economically disadvantaged areas. Conversely, several studies have emphasized a strong link between minority status and motor vehicle crashes, noting that socioeconomic disadvantage is closely associated with a higher likelihood of fatal outcomes (14). Also, study (15) utilizing a regression model, identified income, demographic characteristics, and technological factors as significant predictors of fatal motor vehicle crash rates. Findings in study (16) found that education and income are correlated inversely with fatal crash involvement.

Study (17) examines the impact of local socioeconomic conditions on driver speeding behavior through multilevel logistic regression models. The findings indicate that regional economic disparities contribute to variations in driving behavior. Broadly, these studies reveal significant differences in crash patterns across population groups and geographic areas with differing socioeconomic profiles. Study (18) revealed that income level plays a significant role in shaping risky driving behaviors. Through a multilevel analysis spanning different districts in Thailand, it identified both individual and environmental factors contributing to unsafe actions. On an individual level, lower income was closely linked to a higher likelihood of engaging in risky driving behavior. Study (19) further reinforces this point by noting that low-income areas in the Southeastern U.S. experience higher rates of fatal single-vehicle crashes, often associated with riskier behaviors like low seatbelt usage and limited safety awareness, particularly among young male drivers. Based on the findings of study (20), driver behavior is influenced by national income levels. In low-income countries, resource limitations often lead to more dangerous road conditions, leading to higher risk driving habits. Such constraints make unsafe actions like overtaking on undivided highways more common, raising the likelihood of crashes. Study (21) suggests that the national income level also affects driver behavior by shaping infrastructure quality and law enforcement. Studies have shown that expatriates from a given country living in high-income nations report fewer emotional violations, while domestic drivers exhibit higher rates due to weaker traffic systems. This highlights the role of stronger enforcement and better infrastructure in promoting safer driving in wealthier countries.

Study (22) discovered that families with lower incomes are more likely to own older vehicles lacking modern safety features, which heightens their risk of traffic accidents. This highlights how socioeconomic status significantly influences access to safer transportation options and associated safety outcomes. The type of vehicle individuals drive is closely tied to their income level. Higher-income individuals tend to own newer sedans or SUVs, while those with lower incomes are more likely to drive older vehicles that offer less crash protection (23). Additionally, demographic factors such as income, gender, and household size play a major role in shaping vehicle preferences, influencing choices between compact cars, pickup trucks, and SUVs (24).

Focusing crash analysis on Interstate highways presents a methodological advantage in traffic safety research by controlling for variations in roadway facility quality, thereby enabling a more precise examination of other contributing factors. Study (25–27) conducted research that investigates fatal, nonfatal injury, and non-injury traffic crashes on both urban and rural Interstates, employing principal component analysis to identify critical influences beyond roadway design. Since Interstates are built to uniformly high standards, this focus reduces the confounding effects of infrastructure variability. Research article (28) further supports this methodology by highlighting the importance of identifying specific crash types based on roadway facility. This enhances the ability to analyze crash dynamics while minimizing the influence of roadway quality variations.

The literature review synthesizes a growing body of research examining the relationship between contributing factors to fatal crashes and socioeconomic variables. These studies have investigated motor vehicle crashes using a range of analytical methods, including negative binomial models, spatial regression analysis, geographically weighted regression (GWR), nonlinear regression, logistic regression, Lasso regression, multinomial logit models, spatial crash frequency analysis, geographic-demographic overlays, and umbrella reviews(8). Many of these studies emphasize the significance of both driver-related and environmental factors in analyzing crash contributors and developing effective safety policies. These reviews underscore the need to utilize these considerations in planning and the design of safer

transportation systems. However, it also reveals a gap in the literature: limited attention has been given to how income disparities influence contributing factors, specifically, on Interstate highways in the U.S., a consistent setting that minimizes the influence of infrastructure variability, allowing for a more isolated evaluation of income effects on crash causation. This creates a common ground that helps eliminate confounding factors, enabling a more accurate and isolated assessment of income's impact on crash causation.

METHODS

Association Rule Mining (ARM) is a data-mining technique that identifies attribute combinations that routinely appear together within the same observation (2, 4).

Because it explores relationships without prespecified hypotheses, ARM has proven to be a potent tool in traffic-safety investigations (2, 4, 29).

In this study, the Apriori algorithm was applied to extract association rules that help uncover the underlying contributing factors associated with fatal crashes across different income levels. By analyzing patterns within the data, the method enables the identification of specific variables that are more prevalent in crashes occurring in certain income level areas. This approach facilitates a deeper understanding of how socioeconomic disparities may influence the risk profiles and causal mechanisms behind fatal traffic incidents.

During the past few years, numerous researchers have leveraged ARM as a decision-support method to derive rules from high-dimensional crash datasets, like FARS, targeting distinct variable categories (2, 4, 30, 31).

Assume $T = \{t_1, t_2, t_3, \dots, t_n\}$ represents the set of selected records, each of which contains a subset of features taken from the itemset $Y = \{y_1, y_2, y_3, \dots, y\}$ (2, 4).

An association rule is expressed as $A \rightarrow B$, where $A, B \subseteq T$ and $A \cap B = \emptyset$ (2, 4).

In this notation, A denotes the antecedent (left-hand side, LHS), while B denotes the consequent (right-hand side, RHS). The antecedent may comprise several items within a single rule (2, 4).

For instance, consider the three-item rule $\{\text{REL_ROADNAME} = \text{On Roadside}, \text{TIME_BLOCK} = 12 \text{ P.M.} - 6 \text{ P.M.}\} \rightarrow \{\text{crash severity} = \text{property_damage}\}$, where $A = \{\text{REL_ROADNAME} = \text{On Roadside}, \text{TIME_BLOCK} = 12 \text{ P.M.} - 6 \text{ P.M.}\}$ and $B = \{\text{crash severity} = \text{property_damage}\}$. The number of items included in such rules can vary depending on the objectives of the study. The resulting rules reveal interdependencies among variables but must not be seen as proof of direct causation (2, 4, 32–34).

To evaluate each rule, key metrics such as support (S), confidence (C), and lift (L), complemented by auxiliary measures like coverage, together furnish a comprehensive view of rule strength and relevance within traffic-safety analyses (2, 33, 35).

Support quantifies the frequency with which the pattern ($A \rightarrow B$) appears across the dataset, whereas confidence represents the share of those occurrences relative to the total instances in which A arises (2, 4, 36).

Lift, the third measure, indicates how often the items co-occur within an independent crash record (2).

Conversely, count denotes the raw number of records that contain both the antecedent and the consequent simultaneously. Unlike support, count is unnormalized with respect to the total transactions and merely reports the absolute frequency of the complete rule (2, 37).

The relevant formulas are listed below in **Equations 1 to 5**:

$$\text{Support}(A) = \frac{A'}{N} \quad (1)$$

$$\text{Support}(B) = \frac{B'}{N} \quad (2)$$

$$\text{Support}(A \rightarrow B) = \frac{A' \cap B'}{N} \quad (3)$$

$$\text{Confidence}(A \rightarrow B) = \frac{\text{Support}(A \rightarrow B)}{\text{Support}(A)} \quad (4)$$

$$\text{Lift}(A \rightarrow B) = \frac{\text{Support}(A \rightarrow B)}{\text{Support}(A) \times \text{Support}(B)} \quad (5)$$

In these expressions, N denotes the total number of selected records, A' is the count of occurrences containing A , B' is the count containing B , and $(A' \cap B')$ is the count containing both A and B .

Lift is vital because it quantifies how much more often the antecedent and consequent co-occur than would be anticipated under statistical independence (2, 38).

A lift exceeding 1 signals a positive correlation between A and B , whereas a value lower than 1 implies a negative association (2, 39). A lift near 1 indicates that A is independent of the likelihood of B (2, 4, 40).

ARM Implementation Details

Association Rule Mining (ARM) was conducted separately for each income quintile to identify crash pattern differences across socioeconomic strata. The analysis utilized the Apriori algorithm with a minimum support threshold of 0.1 to ensure statistical reliability of identified patterns. Only rules with lift values exceeding 3.0 and confidence levels of at least 0.6 were retained for analysis, ensuring that identified associations were both statistically significant and practically meaningful.

For each income quintile, crash records were transformed into transactional format, where each crash represented a transaction containing relevant feature-value combinations. Then, ARM was applied to generate rules that met the minimum support and lift thresholds. Rules were filtered to focus on those predicting fatality categories (single versus multiple fatalities), as these outcomes represent the most severe crash consequences.

The analysis framework examined rules with a maximum itemset length of 6 to balance comprehensiveness with interpretability. Rules predicting fatality outcomes were prioritized, with particular attention to patterns that distinguished between single and multiple fatality crashes across different income strata.

Dataset Description

The Fatality Analysis Reporting System (FARS) is a publicly accessible database that is managed by the National Highway Traffic Safety Administration (NHTSA) and offers detailed records of motor vehicle crashes in the United States, including the 50 States, the District of Columbia, and Puerto Rico, that resulted in at least one fatality (4, 41). This subset was selected to examine crash patterns within the Interstate highway system, which includes freeways and is characterized by high-speed travel, limited access points, and connectivity across diverse socioeconomic regions. Figure 1 displays the geographic distribution of fatal Interstate crashes by state over the study period. States such as Texas and California exhibited higher fatal crash counts, which may reflect their larger populations, extensive Interstate mileage, and higher vehicle miles traveled. Interstate highways were selected for evaluating the effect of income level because they adhere to consistently high national standards for lighting, signage, pavement quality, and geometric design. As noted in the research study (42), these facilities are governed by well-

defined geometric control criteria such as stopping sight distance, lane width, and shoulder width, which reduce design variability and allow for more precise evaluations of socioeconomic disparities in crash outcomes. Additionally, their limited presence of traffic control devices minimizes external interruptions to traffic flow, allowing us to isolate the influence of socioeconomic factors without confounding effects from localized traffic management features.

This dataset captures a wide range of variables, including crash-level related factors, environmental and situational details, involved parties (e.g., pedestrians, drivers, and vehicles), and crash dynamics (such as the initial harmful event and type of collision) (41). Given the depth and breadth of its variables, FARS serves as a valuable resource for researchers and policymakers aiming to analyze and understand the patterns, causes, and contexts surrounding fatal traffic incidents. Its comprehensive structure makes it particularly effective for exploring risk factors and informing safety interventions (41). This study focuses on Interstate crashes that occurred between 2019 and 2023. Since the FARS dataset includes fatal crashes across all roadway types, a filtering process was applied to isolate only those crashes that took place on Interstate facilities, resulting in a total of 23,232 incidents. The annual distribution presented below in Figure 2 shows a modest increase in the number of crashes during the first three years, followed by a decline in the subsequent years. A summary of annual fatal Interstate crashes from 2019 to 2023 is presented in Figure 2.

In this study, individual CSV files from the FARS dataset were obtained for each year and subsequently merged to create a comprehensive database. Since FARS data is structured across multiple levels, including crash-level, vehicle-level, and person-level files, the datasets were integrated in a manner that preserved the most granular information, namely at the person level. During the merging process, duplicate columns arising from overlapping information across files were identified and removed to ensure consistency and eliminate redundancy.

Because the analysis required income data at a regional level, the median household income for each crash location was obtained using the ZIP code of the crash site, with data sourced from the American Community Survey. However, because the cost of living varies significantly from state to state, raw median income values are not directly comparable across regions. To address this, the incomes were adjusted using the Regional Price Parity (RPP) index, which normalizes income levels based on local purchasing power. After adjustment, the income values were grouped into five categories to facilitate comparative analysis.

Feature Selection and Engineering

The initial dataset contained variables across multiple domains, including vehicle characteristics, crash severity indicators, safety equipment usage, demographic factors, substance use, driver behaviors, geographic locations, infrastructure conditions, temporal factors, environmental conditions, emergency response metrics, and crash circumstances. Several engineered features were created to enhance the analytical framework. Age was categorized into four groups: Young (less than 25), Middle Adult (25-45), Middle Aged (46-65), and Older (65+). Temporal variables were binned to capture crash patterns across time periods, with hours categorized into four six-hour blocks and days grouped into weekday versus weekend categories. A new variable was created to capture the severity of crashes based on fatality count, categorizing cases into single-fatality and multiple-fatality crashes. Emergency response lag time was calculated by combining lag hours and minutes, then categorized into four intervals: less than 1 hour, 1-3 hours, 3-6 hours, and greater than 6 hours. Seasonal patterns were captured by grouping months into traditional four-season categories.

Economic Data Integration

To examine income-based disparities, the FARS data was enriched with economic indicators from multiple sources. ZIP code-level median household income data for each year were obtained from the American Community Survey and were merged with the crash records using road ZIP codes. Regional Price Parity (RPP) data from the Bureau of Economic Analysis was incorporated to adjust income values

for local cost of living variations, providing a more accurate representation of purchasing power across different geographic regions.

The adjusted income variable was calculated as:

$$\text{Adjusted Income} = \text{Median Household Income} / (\text{Regional Price Parity} \times 0.01).$$

This adjustment accounts for geographic variations in cost of living, ensuring that income comparisons reflect real economic conditions rather than nominal dollar amounts. The adjusted income values were then divided into quintiles (Q1-Q5), with Q1 representing the lowest income quintile and Q5 the highest.

In this study, we began by compiling a broad set of variables covering various aspects of the crash, including driver characteristics, distractions, vehicle and roadway features, and environmental conditions, based on insights from previous studies (2, 3). We then conducted a feature importance ranking to evaluate the influence of these variables on the outcome variable FATALS, which represents the number of fatalities in each crash. For analysis purposes, we categorized this variable into 'single fatality' and 'multiple fatalities'.

Feature Importance Analysis

To identify the most predictive variables for crash fatality outcomes, an XGBoost regression model was employed. The model was trained using an 80%-20% train-test split with fatality count as the target variable. Categorical variables were encoded using label encoding to maintain compatibility with the gradient boosting framework. The model utilized 100 estimators with a fixed random state for reproducibility. Feature importance scores were calculated and ranked to identify the most influential variables in predicting fatality outcomes; To facilitate the primary objective of this study, which is to examine how disparities in income level are associated with variations in contributing factors to fatal crashes occurring on Interstate roadways.

Interstate Fatal Crashes, 2019 to 2023 (FARS)

Count of crashes on Interstate facilities by state

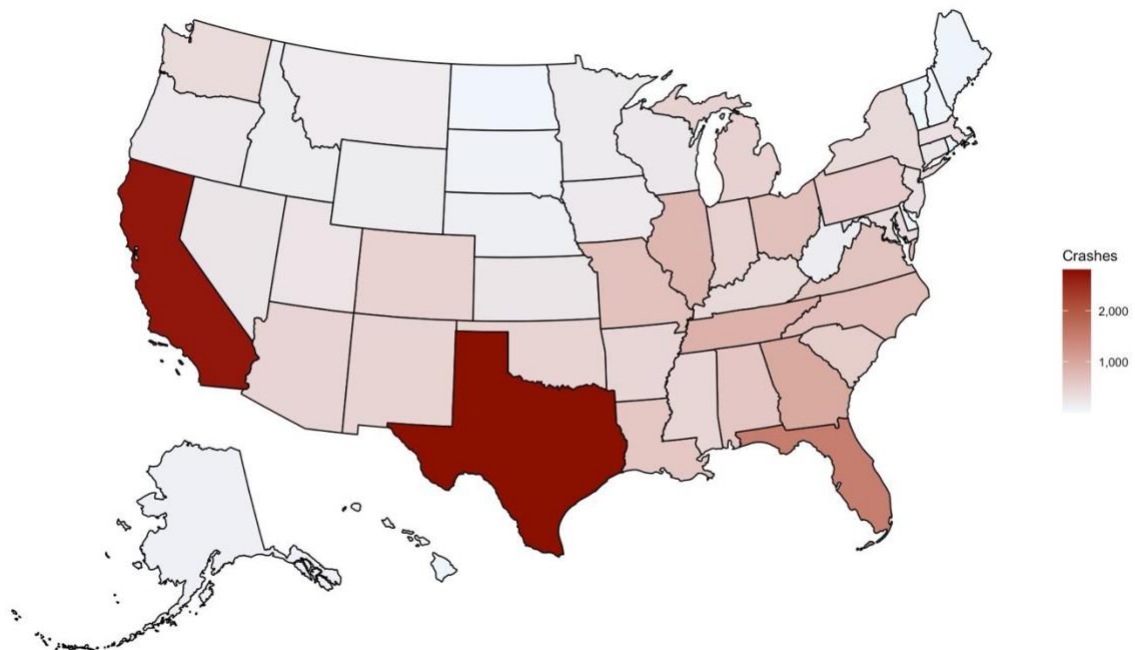


Figure 1 Choropleth Map of Fatal Interstate Crashes by State, 2019–2023 (Based on NHTSA FARS Data)

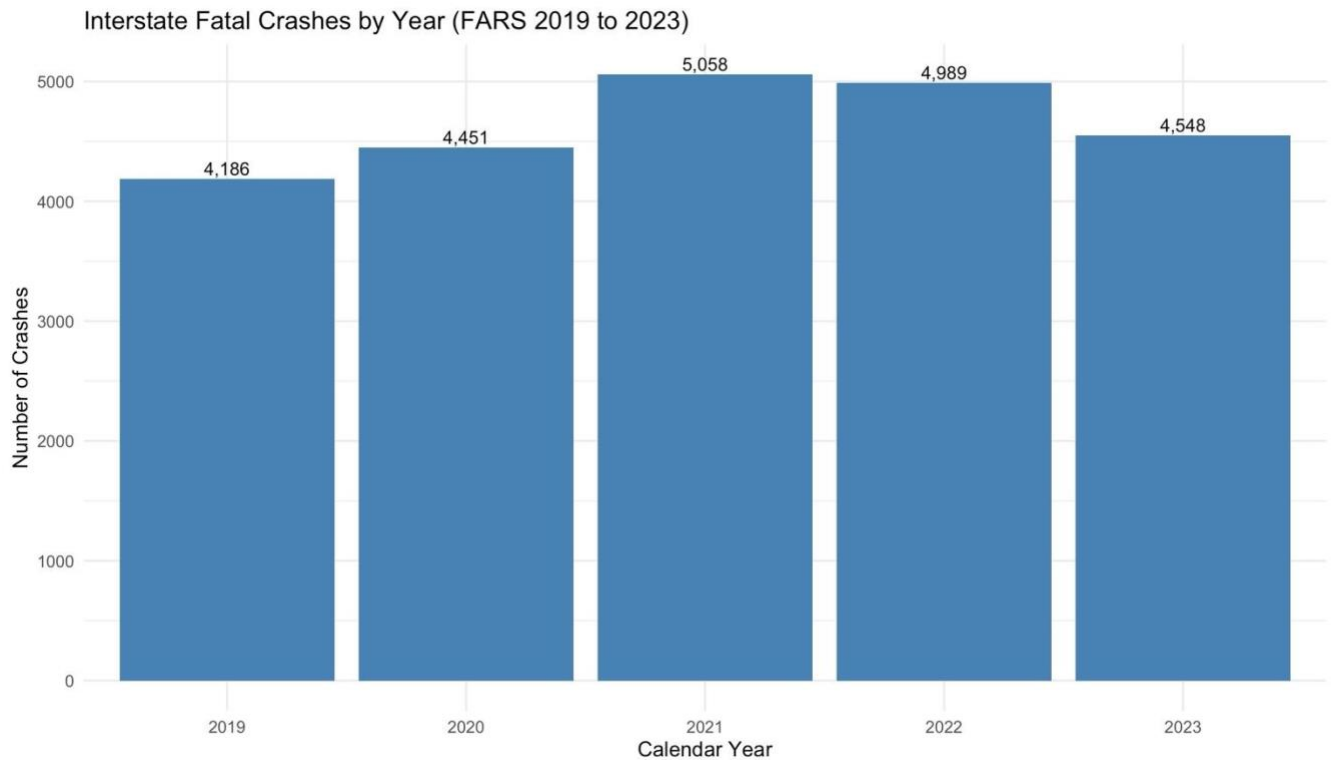


Figure 2 Annual Count of Fatal Interstate Crashes in the United States, 2019–2023 (Based on NHTSA FARS Data)

RESULTS

Feature Importance Analysis

The XGBoost model identified 25 features most predictive of fatality count in Interstate crashes. The feature importance analysis revealed that crash severity indicators dominated the top predictors, with harmful event types, fire explosions, and most harmful event classifications (M_HARM) comprising several of the highest-ranking variables. Infrastructure and emergency response factors also featured prominently, suggesting that both road characteristics and post-crash response capabilities significantly influence fatality outcomes. As shown in Figure 3, the XGBoost analysis reveals that crash severity, emergency response timing, and roadway context are the strongest predictors of fatality count. Demographic and safety equipment factors also contribute, highlighting the multifaceted nature of fatal crash outcomes.

Top 25 Features Predicting Fatality Count (XGBoost)

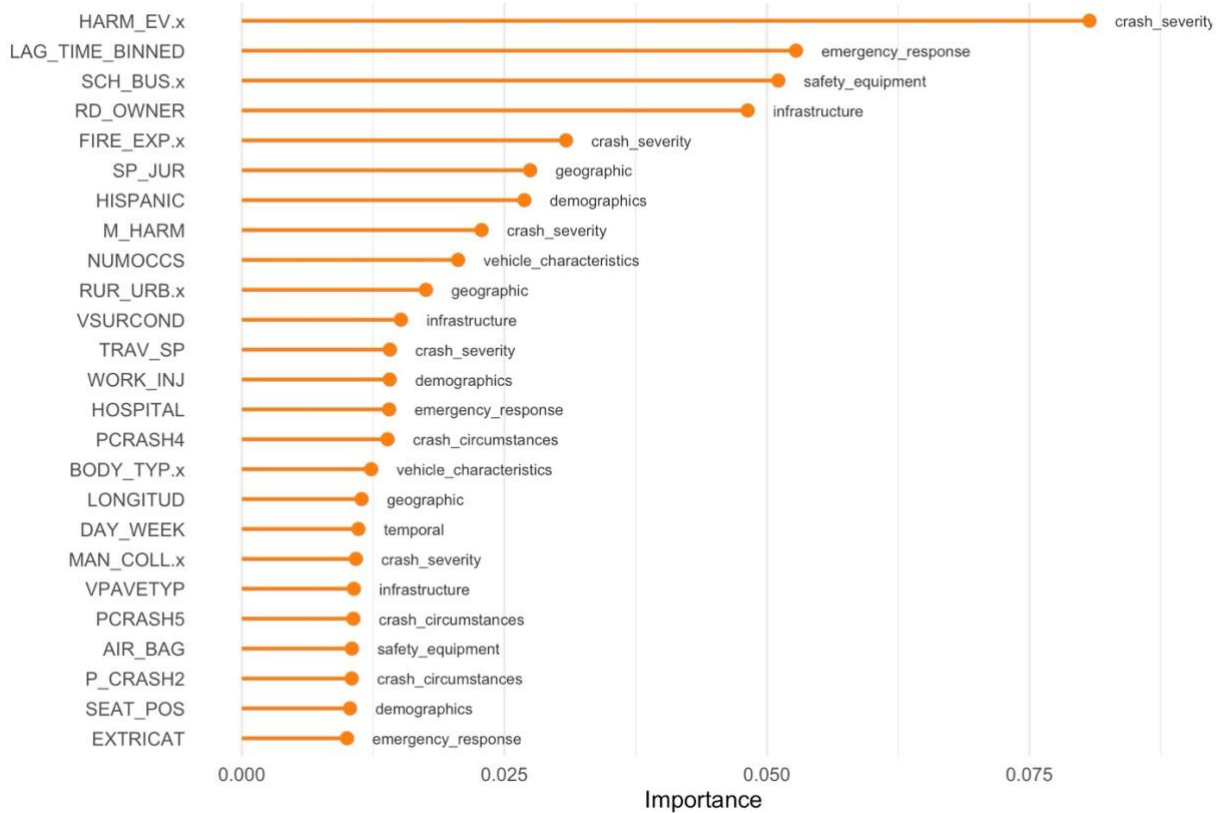


Figure 3 Top 25 features predicting fatality count in Interstate crashes (NHTSA FARS Dataset 2019–2023)

Economic Data Integration Outcomes

Based on the previous studies, the income distribution analysis revealed substantial variation in crash frequency across economic strata (8, 43). The adjusted income calculation, incorporating regional price parity adjustments, provided a more nuanced view of economic disparities than nominal income values alone. Income quintile distributions showed varying crash frequencies, with Q1 (lowest income) and Q5 (highest income) demonstrating distinct pattern characteristics.

Association Rule Mining Results

The ARM analysis successfully identified significant crash patterns across all income quintiles, generating a total of 15,258 rules meeting the specified criteria. The distribution of significant rules varied considerably across income levels: Q5 (highest income) generated 4,851 rules, Q4 produced 3,433 rules, Q2 yielded 2,752 rules, Q3 resulted in 2,257 rules, and Q1 (lowest income) generated 1,965 rules. After further processing, the number of significant rules were reduced to 619 rules. From this refined set, we retained only the top five rules with the highest lift values for each of the five income quintiles. These rules are presented in TABLE 1. The definitions of each itemset of each rule are briefly described in Figure 3.

TABLE 1 Top Five Association Rules with Highest Lift for Each Income Quintile

Quintile	Antecedents	Consequents	Support	Confidence	Lift
----------	-------------	-------------	---------	------------	------

Q1	({'PCRASH5=4.0', 'HISPANIC=7'})	({'MAN_COLL.x=0', 'FATALS_CAT=Multiple', 'WORK_INJ=0'})	0.114	0.859	4.814
Q1	({'MAN_COLL.x=0', 'WORK_INJ=0'})	({'PCRASH5=4.0', 'FATALS_CAT=Multiple', 'HISPANIC=7'})	0.114	0.639	4.814
Q1	({'PCRASH5=4.0', 'WORK_INJ=0'})	({'MAN_COLL.x=0', 'FATALS_CAT=Multiple', 'HISPANIC=7'})	0.114	0.696	4.805
Q2	({'PCRASH5=4.0', 'WORK_INJ=0'})	({'MAN_COLL.x=0', 'FATALS_CAT=Multiple', 'HISPANIC=7'})	0.111	0.709	5.048
Q2	({'MAN_COLL.x=0', 'HISPANIC=7'})	({'FATALS_CAT=Multiple', 'PCRASH5=4.0', 'WORK_INJ=0'})	0.111	0.789	5.048
Q2	({'PCRASH5=4.0', 'HISPANIC=7'})	({'MAN_COLL.x=0', 'FATALS_CAT=Multiple', 'WORK_INJ=0'})	0.111	0.857	5.044
Q3	({'MAN_COLL.x=0', 'WORK_INJ=0'})	({'FATALS_CAT=Multiple', 'PCRASH5=4.0', 'HISPANIC=7'})	0.112	0.663	5.119
Q3	({'PCRASH5=4.0', 'HISPANIC=7'})	({'MAN_COLL.x=0', 'FATALS_CAT=Multiple', 'WORK_INJ=0'})	0.112	0.868	5.119
Q3	({'PCRASH5=4.0', 'WORK_INJ=0'})	({'MAN_COLL.x=0', 'FATALS_CAT=Multiple', 'HISPANIC=7'})	0.112	0.740	5.103
Q4	({'PCRASH5=4.0', 'HISPANIC=7'})	({'MAN_COLL.x=0', 'FATALS_CAT=Multiple', 'WORK_INJ=0'})	0.106	0.860	5.347
Q4	({'MAN_COLL.x=0', 'WORK_INJ=0'})	({'PCRASH5=4.0', 'FATALS_CAT=Multiple', 'HISPANIC=7'})	0.106	0.662	5.347
Q4	({'MAN_COLL.x=0', 'HISPANIC=7'})	({'PCRASH5=4.0', 'FATALS_CAT=Multiple', 'WORK_INJ=0'})	0.106	0.769	5.336
Q5	({'PCRASH5=4.0', 'WORK_INJ=0'})	({'MAN_COLL.x=0', 'FATALS_CAT=Multiple', 'HISPANIC=7'})	0.106	0.747	5.646
Q5	({'MAN_COLL.x=0', 'HISPANIC=7'})	({'PCRASH5=4.0', 'FATALS_CAT=Multiple', 'WORK_INJ=0'})	0.106	0.798	5.646
Q5	({'MAN_COLL.x=0', 'WORK_INJ=0'})	({'PCRASH5=4.0', 'FATALS_CAT=Multiple', 'HISPANIC=7'})	0.106	0.673	5.638

Key patterns emerged from the rule analysis. Harmful event types showed variation across income levels, with pedestrian-involved crashes (HARM_EV.x = 8.0 Pedestrian) featuring prominently in certain quintiles. Motor vehicle in-transport events (HARM_EV.x = 12 Motor Vehicle In-Transport) represented another significant pattern type. Crash circumstances also varied, with same-direction travel at higher speeds (P_CRASH2 = 53.0) and front-to-rear collisions (MAN_COLL.x = 1 Front-to-Rear) appearing as frequent antecedents in crash scenarios.

The most harmful event classifications revealed socioeconomic disparities, with pedestrian involvement (M_HARM = 8.0 Pedestrian) showing differential patterns across income quintiles. These findings suggest that crash types and severities may be systematically related to the economic characteristics of crash locations.

Crash Severity Patterns

The analysis of fatality categories across income quintiles demonstrated varying propensities for single versus multiple fatality crashes. The association rules revealed that certain combinations of factors were more strongly associated with multiple fatality outcomes in specific income strata, suggesting that economic factors may influence both crash occurrence patterns and severity outcomes.

Emergency response characteristics, particularly lag time categories, showed income-related variations that may contribute to differential fatality outcomes. These patterns suggest potential disparities in emergency services accessibility or response capabilities across different socioeconomic areas along Interstate corridors.

DISCUSSION

Socioeconomic Disparities in Crash Patterns

The findings reveal substantial disparities in fatal crash patterns across income quintiles on Interstate highways, suggesting that economic factors play a significant role in traffic safety outcomes. The variation in association rule counts across income levels indicates that crash patterns become increasingly complex in higher-income areas, potentially reflecting greater diversity in travel behaviors, vehicle types, and infrastructure characteristics in these regions.

The presence of pedestrian-involved crashes on Interstate facilities in certain income quintiles raises important research questions, given that pedestrian access is typically prohibited on Interstate highways.

While Interstates are designed as controlled-access highways with limited pedestrian exposure, the presence of pedestrian fatalities suggests potential gaps in safety measures near Interstate interchanges or in areas where alternative transportation options are limited.

Infrastructure and Emergency Response Implications

The high predictive importance of road ownership characteristics and emergency response lag times highlights the critical role of infrastructure management and emergency services in determining crash outcomes. The observed income-related variations in these factors suggest potential disparities in road maintenance standards and emergency response capabilities across different socioeconomic areas (44).

The prominence of harmful event types in the predictive model underscores the importance of crash circumstances in determining fatality outcomes. The differential patterns across income quintiles suggest that crash types may be systematically related to local economic conditions, potentially through mechanisms such as traffic volume, vehicle mix, infrastructure quality variations, enforcement intensity, or access to vehicle safety technologies.

Demographic Factors

The inclusion of Hispanic ethnicity among the top predictive features raises important considerations about demographic disparities in traffic safety outcomes. This finding, combined with the geographic variables' predictive importance, suggests that cultural, linguistic, or structural factors may contribute to differential crash risks and outcomes across communities (45).

Methodological Considerations

The ARM methodology proved effective in identifying distinct crash patterns across income quintiles, though the substantial variation in rule counts suggests that different analytical approaches may be needed

for different income strata. The higher rule counts in affluent areas may indicate greater diversity in crash scenarios, while lower counts in less affluent areas might reflect more concentrated risk factors.

Policy Implications

The identified disparities suggest several policy intervention opportunities. The income-related variations in crash patterns indicate that safety interventions may need to be tailored to local socioeconomic contexts rather than applying uniform approaches across all Interstate segments. Areas serving lower-income populations may benefit from focused interventions addressing the specific crash types identified in their association rules.

The emergency response lag time variations across income quintiles suggest potential needs for enhanced emergency services coverage or capabilities in certain areas. Infrastructure improvements targeting the specific harmful event types identified in each income stratum could provide more effective safety benefits than generic safety measures.

CONCLUSIONS

This study provides compelling evidence of systematic disparities in fatal crash patterns across income quintiles on Interstate highways, demonstrating that socioeconomic factors significantly influence both crash characteristics and outcomes. The association rule mining methodology successfully identified distinct crash patterns for each income stratum, revealing 15,258 significant rules that illuminate the complex and hidden relationships between economic status and traffic safety outcomes.

The analysis identified 25 key predictive features for crash fatalities, with harmful event types, emergency response characteristics, and infrastructure factors emerging as primary determinants of crash severity. The substantial variation in association rule patterns across income quintiles suggests that crash complexity and diversity increase with economic status, potentially reflecting greater variety in travel patterns, vehicle types, and infrastructure characteristics in more affluent areas.

Key findings include the prominence of pedestrian-involved crashes and motor vehicle in-transport events across different income strata, with varying patterns of same-direction collisions and front-to-rear impacts. The differential emergency response lag times and infrastructure characteristics across income quintiles highlight potential disparities in safety resources and capabilities that may contribute to varying fatality outcomes.

The study's identification of demographic factors, particularly Hispanic ethnicity, among top predictive features underscores the intersection of socioeconomic and cultural factors in traffic safety outcomes. Combined with geographic variables such as rural-urban classification, these findings suggest that effective traffic safety interventions must consider the complex interplay of economic, demographic, and geographic factors. The integration of economic indicators with crash records provided crucial context for understanding disparities.

These findings have significant implications for traffic safety policy and resource allocation. The identified income-based disparities suggest that uniform safety interventions may be insufficient to address the diverse crash patterns across socioeconomic strata. Instead, targeted interventions addressing the specific risk patterns identified for each income quintile may prove more effective in reducing fatal crash rates.

Future research should explore the causal mechanisms underlying these disparities, investigate temporal trends in income-related crash patterns, and evaluate the effectiveness of targeted interventions designed to address the specific risk profiles identified for each socioeconomic stratum. Additionally, expanding the analysis to include non-fatal crashes could provide insights into the full spectrum of traffic safety disparities across income levels.

ACKNOWLEDGMENTS

We would like to thank NsongErnest Asiedu for his valuable input during the initial variable selection phase of the NHTSA FARS dataset cleaning process.

AUTHOR CONTRIBUTIONS

The authors confirm their contribution to the paper as follows:

Study conception and design: Seyed Amirhossein Naeimaeimousavi, Mohammad Reza Abbaszadeh Lima;

Data collection: Mohammad Reza Abbaszadeh Lima, Seyed Amirhossein Naeimaeimousavi;

Analysis and interpretation of results: Seyed Amirhossein Naeimaeimousavi;

Draft manuscript preparation: Mohammad Reza Abbaszadeh Lima, Seyed Amirhossein Naeimaeimousavi

All authors reviewed the results and approved the final version of the manuscript.

DECLARATION OF CONFLICTING INTERESTS

The authors declared no potential conflicts of interest with respect to the research, authorship, and publication of this article.

FUNDING

The authors disclosed no financial support for the research, authorship, and publication of this article.

REFERENCES

1. Chen, L., S. Huang, C. Yang, and Q. Chen. Analyzing Factors That Influence Expressway Traffic Crashes Based on Association Rules: Using the Shaoyang–Xinhuang Section of the Shanghai–Kunming Expressway as an Example. *Journal of Transportation Engineering, Part A: Systems*, Vol. 146, No. 9, 2020. <https://doi.org/10.1061/JTEPBS.0000425>.
2. Hossain, M. M., M. R. Abbaszadeh Lima, and H. Zhou. Severity Analysis of Secondary Crashes on High-Speed Roadways: Pattern Recognition Using Association Rule Mining. *Transportation Research Record: Journal of the Transportation Research Board*, Vol. 2678, No. 8, 2024, pp. 919–931. <https://doi.org/10.1177/03611981231223194>.
3. Adanu, E. K., P. Penmetsa, D. Wood, and S. L. Jones. Incorporating Systems Thinking Approach in a Multilevel Framework for Human-Centered Crash Analysis. *Transportation Research Interdisciplinary Perspectives*, Vol. 2, 2019, p. 100031. <https://doi.org/10.1016/j.trip.2019.100031>.
4. Abbaszadeh Lima, M. R., M. M. Hossain, H. Zhou, and Y. Song. Data Mining Approach to Explore the Contributing Factors to Fatal Wrong-Way Crashes by Local and Non-Local Drivers. *Future Transportation*, Vol. 4, No. 3, 2024, pp. 985–999. <https://doi.org/10.3390/futuretransp4030047>.
5. Chakraborty, M., T. J. Gates, and S. Sinha. Causal Analysis and Classification of Traffic Crash Injury Severity Using Machine Learning Algorithms. *Data Science for Transportation*, Vol. 5, No. 2, 2023, p. 12. <https://doi.org/10.1007/s42421-023-00076-9>.
6. Sagar, S., N. Stamatiadis, S. Wright, and A. Cambron. Identifying High-Risk Commercial Vehicle Drivers Using Sociodemographic Characteristics. *Accident Analysis & Prevention*, Vol. 143, 2020, p. 105582. <https://doi.org/10.1016/j.aap.2020.105582>.
7. Verma, S., F. Wilson, H. Wang, L. Smith, and H. J. Tak. Impact of Community Socioeconomic Characteristics on Emergency Medical Service Delays in Responding to Fatal Vehicle Crashes. *AJPM Focus*, Vol. 2, No. 4, 2023, p. 100129. <https://doi.org/10.1016/j.focus.2023.100129>.
8. Liu, J., S. Das, and M. N. Khan. Decoding the Impacts of Contributory Factors and Addressing Social Disparities in Crash Frequency Analysis. *Accident Analysis & Prevention*, Vol. 194, 2024, p. 107375. <https://doi.org/10.1016/j.aap.2023.107375>.
9. McCarty, D., and H. W. Kim. Risky Behaviors and Road Safety: An Exploration of Age and Gender Influences on Road Accident Rates. *PLOS ONE*, Vol. 19, No. 1, 2024, p. e0296663. <https://doi.org/10.1371/journal.pone.0296663>.
10. Alharbi, R. J., V. Lewis, and C. Miller. A State-of-the-Art Review of Factors That Predict Mortality among Traumatic Injury Patients Following a Road Traffic Crash. *Australasian Emergency Care*, Vol. 25, No. 1, 2022, pp. 13–22. <https://doi.org/10.1016/j.auec.2021.01.005>.
11. Peura, C., J. A. Kilch, and D. E. Clark. Evaluating Adverse Rural Crash Outcomes Using the NHTSA State Data System. *Accident Analysis & Prevention*, Vol. 82, 2015, pp. 257–262. <https://doi.org/10.1016/j.aap.2015.06.005>.
12. Wang, L., R. Li, C. Wang, and Z. Liu. Driver Injury Severity Analysis of Crashes in a Western China’s Rural Mountainous County: Taking Crash Compatibility Difference into Consideration.

- Journal of Traffic and Transportation Engineering (English Edition)*, Vol. 8, No. 5, 2021, pp. 703–714. <https://doi.org/10.1016/j.jtte.2020.12.002>.
13. Noland, R. B., and M. L. Laham. Are Low Income and Minority Households More Likely to Die from Traffic-Related Crashes? *Accident Analysis & Prevention*, Vol. 120, 2018, pp. 233–238. <https://doi.org/10.1016/j.aap.2018.07.033>.
 14. Li, X., S. Yu, X. Huang, B. Dadashova, W. Cui, and Z. Zhang. Do Underserved and Socially Vulnerable Communities Observe More Crashes? A Spatial Examination of Social Vulnerability and Crash Risks in Texas. *Accident Analysis & Prevention*, Vol. 173, 2022, p. 106721. <https://doi.org/10.1016/j.aap.2022.106721>.
 15. Sohaee, N., and S. Bohluli. Nonlinear Analysis of the Effects of Socioeconomic, Demographic, and Technological Factors on the Number of Fatal Traffic Accidents. *Safety*, Vol. 10, No. 1, 2024, p. 11. <https://doi.org/10.3390/safety10010011>.
 16. Sagar, S., N. Stamatiadis, and A. Stromberg. Effect of Socioeconomic and Demographic Factors on Crash Occurrence. *Transportation Research Record: Journal of the Transportation Research Board*, Vol. 2675, No. 12, 2021, pp. 80–91. <https://doi.org/10.1177/03611981211027887>.
 17. Ghasemzadeh, A., and M. M. Ahmed. Quantifying Regional Heterogeneity Effect on Drivers' Speeding Behavior Using SHRP2 Naturalistic Driving Data: A Multilevel Modeling Approach. *Transportation Research Part C: Emerging Technologies*, Vol. 106, 2019, pp. 29–40. <https://doi.org/10.1016/j.trc.2019.06.017>.
 18. Jomnonkwo, S., S. Uttra, B. Khampirat, and V. Ratanavaraha. Applying Multilevel Analysis and the Driver Behavior Questionnaire (DBQ) on Unsafe Actions under a Road Safety Policy. *PLOS ONE*, Vol. 17, No. 11, 2022, p. e0277750. <https://doi.org/10.1371/journal.pone.0277750>.
 19. Stamatiadis, N., and G. Puccini. Fatal Crash Rates in the Southeastern United States: Why Are They Higher? *Transportation Research Record: Journal of the Transportation Research Board*, Vol. 1665, No. 1, 1999, pp. 118–124. <https://doi.org/10.3141/1665-16>.
 20. Mahmud, S. M. S., L. Ferreira, Md. S. Hoque, and A. Tavassoli. Overtaking Risk Modeling in Two-Lane Two-Way Highway with Heterogeneous Traffic Environment of a Low-Income Country Using Naturalistic Driving Dataset. *Journal of Safety Research*, Vol. 80, 2022, pp. 380–390. <https://doi.org/10.1016/j.jsr.2021.12.019>.
 21. Nordfjærn, T., A. M. Hezaveh, and A. R. Mamdoohi. An Analysis of Reported Driver Behaviour in Samples of Domestic and Expatriate Iranians. *Journal of Risk Research*, Vol. 18, No. 5, 2015, pp. 566–580. <https://doi.org/10.1080/13669877.2014.910684>.
 22. Girasek, D. C., and B. Taylor. An Exploratory Study of the Relationship Between Socioeconomic Status and Motor Vehicle Safety Features. *Traffic Injury Prevention*, Vol. 11, No. 2, 2010, pp. 151–155. <https://doi.org/10.1080/15389580903531598>.
 23. Bhat, C. R., S. Sen, and N. Eluru. The Impact of Demographics, Built Environment Attributes, Vehicle Characteristics, and Gasoline Prices on Household Vehicle Holdings and Use. *Transportation Research Part B: Methodological*, Vol. 43, No. 1, 2009, pp. 1–18. <https://doi.org/10.1016/j.trb.2008.06.009>.

24. Choo, S., and P. L. Mokhtarian. What Type of Vehicle Do People Drive? The Role of Attitude and Lifestyle in Influencing Vehicle Type Choice. *Transportation Research Part A: Policy and Practice*, Vol. 38, No. 3, 2004, pp. 201–222. <https://doi.org/10.1016/j.tra.2003.10.005>.
25. Kassu, A., and M. Hasan. Identifying the Principal Factors Influencing Traffic Safety on Interstate Highways. *SN Applied Sciences*, Vol. 1, No. 12, 2019, p. 1729. <https://doi.org/10.1007/s42452-019-1796-2>.
26. Adeyemi, O. J., R. Paul, and A. Arif. An Assessment of the Rural-Urban Differences in the Crash Response Time and County-Level Crash Fatalities in the United States. *The Journal of rural health : official journal of the American Rural Health Association and the National Rural Health Care Association*, Vol. 38, No. 4, 2021, pp. 999–1010. <https://doi.org/10.1111/jrh.12627>.
27. Wu, Q., F. Chen, G. Zhang, X. C. Liu, H. Wang, and S. M. Bogus. Mixed Logit Model-Based Driver Injury Severity Investigations in Single- and Multi-Vehicle Crashes on Rural Two-Lane Highways. *Accident Analysis & Prevention*, Vol. 72, 2014, pp. 105–115. <https://doi.org/10.1016/j.aap.2014.06.014>.
28. Saleem, T., R. J. Porter, R. Srinivasan, D. Carter, S. Himes, and T. Le. *Contributing Factors for Focus Crash and Facility Types*. Federal Highway Administration, 2020.
29. Pande, A., and M. Abdel-Aty. Market Basket Analysis of Crash Data from Large Jurisdictions and Its Potential as a Decision Support Tool. *Safety Science*, Vol. 47, No. 1, 2009. <https://doi.org/10.1016/j.ssci.2007.12.001>.
30. Hossain, A., X. Sun, S. Islam, S. Alam, and M. Mahmud Hossain. Identifying Roadway Departure Crash Patterns on Rural Two-Lane Highways under Different Lighting Conditions: Association Knowledge Using Data Mining Approach. *Journal of Safety Research*, Vol. 85, 2023. <https://doi.org/10.1016/j.jsr.2023.01.006>.
31. Hossain, M. M., H. Zhou, and S. Das. Data Mining Approach to Explore Emergency Vehicle Crash Patterns: A Comparative Study of Crash Severity in Emergency and Non-Emergency Response Modes. *Accident Analysis & Prevention*, Vol. 191, 2023, p. 107217. <https://doi.org/10.1016/j.aap.2023.107217>.
32. Kumbhare, T. A., and S. V Chobe. An Overview of Association Rule Mining Algorithms. *Computer Science and Information Technologies*, Vol. 5, No. 1, 2014.
33. Dol, S. M., and P. M. Jawandhiya. Classification Technique and Its Combination with Clustering and Association Rule Mining in Educational Data Mining — A Survey. *Engineering Applications of Artificial Intelligence*, Vol. 122, 2023, p. 106071. <https://doi.org/10.1016/j.engappai.2023.106071>.
34. Hossain, M. M., H. Zhou, M. A. Rahman, S. Das, and X. Sun. Cellphone-Distracted Crashes of Novice Teen Drivers: Understanding Associations of Contributing Factors for Crash Severity Levels and Cellphone Usage Types. *Traffic Injury Prevention*, Vol. 23, No. 7, 2022. <https://doi.org/10.1080/15389588.2022.2097667>.
35. Du, L., F. Huang, H. Lu, S. Chen, and Q. Guo. An Association Rule Mining-Based Modeling Framework for Characterizing Urban Road Traffic Accidents. *Sustainability*, Vol. 16, No. 23, 2024, p. 10597. <https://doi.org/10.3390/su162310597>.

36. Das, S., R. Tamakloe, H. Zubaidi, I. Obaid, and A. Alnedawi. Fatal Pedestrian Crashes at Intersections: Trend Mining Using Association Rules. *Accident Analysis & Prevention*, Vol. 160, 2021, p. 106306. <https://doi.org/10.1016/j.aap.2021.106306>.
37. Yeon, K. T., H. KyungHyun, and H. S. Oun. 네트워크 침입 탐지를 위한 Coverage와 Exclusion 기반의 새로운 연관 규칙 마이닝. 사물인터넷융복합논문지, Vol. 9, No. 1, 2023, pp. 77–87. <https://doi.org/10.20465/KIOTS.2023.9.1.077>.
38. Kong, X., S. Das, K. Jha, and Y. Zhang. Understanding Speeding Behavior from Naturalistic Driving Data: Applying Classification Based Association Rule Mining. *Accident Analysis & Prevention*, Vol. 144, 2020, p. 105620. <https://doi.org/10.1016/j.aap.2020.105620>.
39. Montella, A. Identifying Crash Contributory Factors at Urban Roundabouts and Using Association Rules to Explore Their Relationships to Different Crash Types. *Accident Analysis and Prevention*, Vol. 43, No. 4, 2011. <https://doi.org/10.1016/j.aap.2011.02.023>.
40. Liu, P., Y. Guo, P. Liu, H. Ding, J. Cao, J. Zhou, and Z. Feng. What Can We Learn from the AV Crashes? – An Association Rule Analysis for Identifying the Contributing Risky Factors. *Accident Analysis & Prevention*, Vol. 199, 2024, p. 107492. <https://doi.org/10.1016/j.aap.2024.107492>.
41. Li, Y., M. M. Karim, R. Qin, Z. Sun, Z. Wang, and Z. Yin. Crash Report Data Analysis for Creating Scenario-Wise, Spatio-Temporal Attention Guidance to Support Computer Vision-Based Perception of Fatal Crash Risks. *Accident Analysis & Prevention*, Vol. 151, 2021, p. 105962. <https://doi.org/10.1016/j.aap.2020.105962>.
42. Hossain, M. M., M. R. A. Lima, and H. Zhou. Potential Impacts of Connected and Autonomous Vehicles on Controlling Criteria for Road Geometric Design: A Review. *Transportation Research Record: Journal of the Transportation Research Board*, 2024. <https://doi.org/10.1177/03611981241242764>.
43. Wang, L., M. Abdel-Aty, J. Lee, and Q. Shi. Analysis of Real-Time Crash Risk for Expressway Ramps Using Traffic, Geometric, Trip Generation, and Socio-Demographic Predictors. *Accident Analysis & Prevention*, Vol. 122, 2019, pp. 378–384. <https://doi.org/10.1016/j.aap.2017.06.003>.
44. Hamim, O. F., and S. V. Ukkusuri. Determining Prominent Factors across System Hierarchies to Improve Road Safety in LMICs: A Case Study of Bangladesh. *Safety Science*, Vol. 150, 2022, p. 105709. <https://doi.org/10.1016/j.ssci.2022.105709>.
45. Roll, J., and N. McNeil. Race and Income Disparities in Pedestrian Injuries: Factors Influencing Pedestrian Safety Inequity. *Transportation Research Part D: Transport and Environment*, Vol. 107, 2022, p. 103294. <https://doi.org/10.1016/j.trd.2022.103294>.