Association Rule Mining of Road Traffic Accidents in Thailand

Suthasinee Kuptabut
Department of Computer, Sakon Nakhon Rajabhat University, Thailand
suthasinee@snru.ac.th

Sutisa Songleknok
Department of Computer, Sakon Nakhon Rajabhat University, Thailand
a_sutisa@snru.ac.th (corresponding author)

Received: 25 June 2025 | Revised: 24 August 2025 and 3 September 2025 | Accepted: 7 September 2025

 DOI: https://doi.org/10.48084/etasr.12933

ABSTRACT

Road traffic accidents have a substantial impact on Thailand's human capital, affecting both public health and economic productivity. Association rule mining can assist relevant agencies in identifying patterns of accident occurrences based on severity levels, thereby informing policy development and preventive measures. This study aims to analyze the support and confidence values consistent with real-world accident data and uncover association rules related to accident severity levels. Three provinces with the highest accident rates, representing diverse geographical contexts, were examined. Data were sourced from the Ministry of Transport's Accident Management System (TRAMS) between 2019 and 2024. The proposed system comprises four modules: dataset, data preprocessing, data imbalance handling, and association rule mining. The findings indicated that the optimal support and confidence values are 0.01 and 0.4, respectively. The association rules for the three provinces, categorized by severity levels—fatalities, serious injuries, and minor injuries—revealed distinct patterns for the first two severity levels, while similar patterns were observed for minor injuries in all three provinces. The results of this study are valuable for traffic and road transportation agencies in designing policies and guidelines to prevent or reduce accidents in alignment with their root causes.

Keywords-road traffic accidents; association rule mining; Apriori algorithm; accident patterns

I. INTRODUCTION

Road traffic accidents are a global concern, as they are one of the leading causes of mortality worldwide [1, 2]. Children and young adults aged 5 to 29 years are particularly affected, constituting the demographic with the highest mortality rates due to road traffic accidents. According to the World Health Organization (WHO), the global rate of road accident mortality decreased slightly to 1.19 million in 2021 compared to 2018. Most of these fatalities occurred in low- and middle-income countries. This situation significantly impacts human resources on a global scale, leading governments to allocate substantial budgets for the care of accident victims and investments in preventive measures, such as acquiring safety equipment and improving road infrastructure to enhance safety. Consequently, safe mobility is recognized as a fundamental human right, integral to the universal right to health [3].

In 2023, Thailand reported 18,218 road traffic fatalities, equivalent to 25.4 deaths per 100,000 people [3]. Thailand also ranks 18th globally in road traffic deaths, compared to its 2018 ranking of 9th, and has the highest road traffic fatality rate in Southeast Asia. The situation of road traffic accidents and fatalities in Thailand has shown a slight overall improvement, with a decrease in the number of fatalities, but Thailand remains one of the countries with the highest number of road traffic deaths. Motorcyclists account for the majority of these fatalities, accounting for 74.4%. When adjusted for population size, Thailand ranks first globally in motorcycle-related fatality rate. The age group most affected comprises youth aged 15-19 years, with a higher incidence among males. Overall, the majority of road traffic fatalities occur among males in the working-age group (15-59 years) [4].

In response to the challenges posed by road traffic accidents, countries around the world analyze the causative factors to formulate preventive strategies aimed at reducing accidents and fatalities. Collected accident data are examined to inform safety planning, such as designing and constructing proper and safe roadways and installing signals in appropriate locations. These preventive measures contribute to the reduction of accidents and mortality rates [3]. To address road traffic accidents, various countries are employing Artificial Intelligence (AI) methodologies, such as Machine Learning (ML) and Deep Learning (DL), to prevent and decrease road traffic accidents in the long term. Studies have shown that these approaches facilitate the development of intelligent systems capable of detecting helmet use [5], predicting vehicle movement through collective learning [6], and using predictive modeling to classify and forecast the severity of traffic accidents by applying ML algorithms such as Random Forest (RF), Naïve Bayes (NB), Logistic Regression (LR), and Artificial Neural Network (ANN) [7, 8].

Several studies have employed association rule mining and the Apriori algorithm to analyze traffic accident patterns. For instance, in [9], the traffic accident patterns associated with the levels of traffic accidents in Palembang from 2015 to 2020 were analyzed. Similarly, in [2], the Apriori algorithm was improved to discover associations between risk factors and explore the root causes of traffic accidents on urban roads. Additionally, in [10], association rule analysis was applied to identify interconnected traffic accident variables that increase the likelihood of fatalities. An association rule analysis underscores the increased risk of injury or death.

Previous studies [2, 9, 10] have applied association rule mining to the analysis of road traffic accidents at the country level, during specific festivals, or within individual provinces. The objectives of these studies were to identify patterns and contributing features to accidents and describe the effects on association rules by establishing distinct support confidence values. A higher minimum support threshold tends to yield fewer rules and involves fewer contributing features, potentially limiting the predictive capacity to only a subset of accident scenarios. Conversely, a lower minimum support threshold generates a larger number of rules and encompasses a wider range of contributing features, which may reduce the reliability of the prediction.

To address the issue of imbalanced data, various studies have employed different techniques, such as the Synthetic Minority Over-sampling Technique (SMOTE) family, to improve the performance of classifiers. In [11], the potential of integrating SMOTE-Tomek preprocessing with modern ML models was demonstrated to address class imbalance in requirements classification. The findings revealed that SMOTE-Tomek with absolute coefficient values reduced variability and noise in the model. In [12], SMOTE-family approaches were used to address the issue of data imbalance, finding that SMOTE-Tomek was better than other resampling approaches.

This study employs association rule mining to generate association rules for road traffic accidents at a provincial level, with the intention of comparing accident patterns across different locations based on accident severity. To address the issue of imbalanced input data, the SMOTE-Tomek technique was applied to enhance data quality, contributing to the reliability of the resulting association rules. The main contributions of this study include: (i) the determination of appropriate support and confidence values tailored to road traffic accident data, and (ii) the analysis of accident patterns in each province. The aim is to analyze support confidence values that are suitable for real accident data and to extract association rules on accident severity in high-incident provinces. The results can support policymakers in designing and developing context-specific strategies to reduce road accidents.

II. METHODOLOGY

Figure 1 offers an overview of the process of constructing traffic accident rules, which consists of four modules: dataset, data preprocessing, data imbalance handling, and association rule mining.

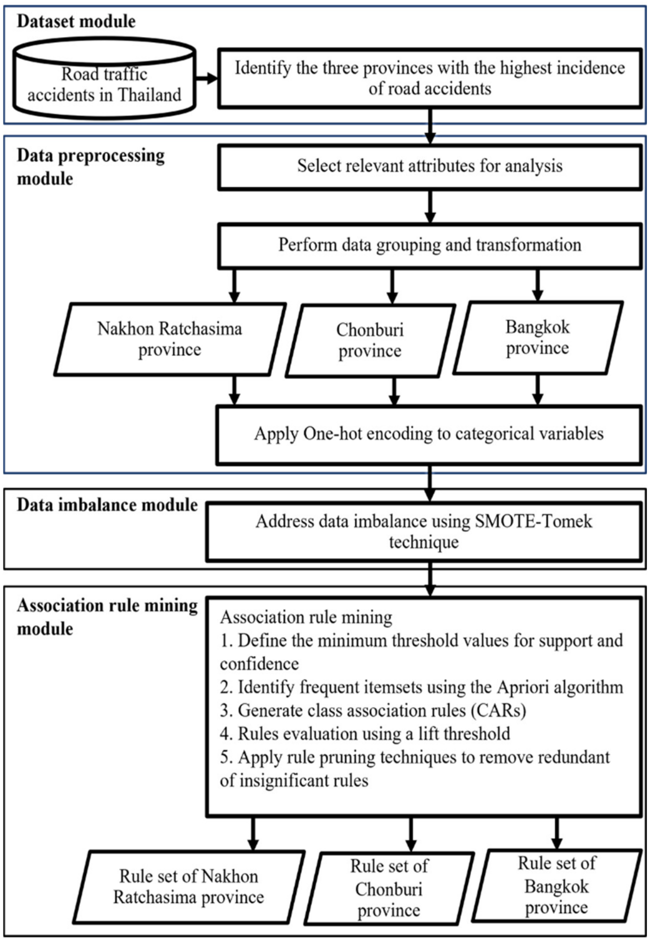

Fig. 1. Research framework.

A. Dataset

The road traffic accident dataset encompasses incidents that occurred on highways, rural roads, and expressways. The dataset was obtained from the Ministry of Transport's Accident Management System (TRAMS), administered by the Information and Communication Technology Center, Office of the Permanent Secretary, Ministry of Transport, Thailand. The dataset comprises accident records collected between 2019 and 2024, consisting of 129,033 items with 37 attributes, and was retrieved from the MOT Data Catalog [13].

Thailand comprises 76 provinces, each characterized by a distinct geographic, environmental, and cultural context. Consequently, accident patterns are expected to vary across different regions. In this study, the three provinces with the highest number of reported accidents were selected for analysis:

- Bangkok province, the capital and the most populous city in Thailand, serves as the country's primary center of economic, social, and infrastructure development, contributing approximately 25% of the national Gross Domestic Product (GDP). The road infrastructure consists of expressways, elevated highways, interchanges, and ring roads, which are characterized by traffic congestion and high vehicle density. Primary roads are connected to numerous minor streets and intersections, and the majority of vehicles using them are cars and motorcycles.
- Nakhon Ratchasima province, located in the northeastern region, is the largest province by area. Its terrain includes mountains, plains, shallow and deep undulating lands. It is a major transportation hub for the Northeast, with extensive road networks connecting to various regions. The province has the highest Gross Provincial Product (GPP) in the Northeast and is second nationally. The road infrastructure consists of highways with four to ten lanes. The majority of the vehicles that use these roads are four-wheeled or larger, including freight trucks utilized for goods transport. Traffic density increases significantly during festival periods.
- Chonburi province, located in the central region, has one of the most robust economies in the country. It is a prominent coastal tourist destination, with the highest GPP in the Eastern region and ranks third nationally. The province has a diverse landscape, including rolling hills, coastal plains, river valleys, and steep areas. Road infrastructure comprises highways, four- to ten-lane highways, and coastal roads. Traffic congestion is pronounced during holidays and festivals, while heavy truck transportation is frequently associated with the movement of goods to the Eastern Economic Corridor (EEC), industrial estates, and major seaports.

To facilitate a more detailed examination, accident cases were classified into three severity levels: Fatalities, Serious injuries, and Minor injuries, as presented in Table I.

TABLE I. ROAD TRAFFIC ACCIDENT SEVERITY LEVELS

Provinces	**Number of individuals affected by severity level**			**Total**
	Fatalities	**Serious injuries**	**Minor injuries**	
Nakhon Ratchasima	549 (10.05%)	312 (5.71%)	4,604 (84.25%)	5,465
Chonburi	292 (4.23%)	266 (3.85%)	6,344 (91.92%)	6,902
Bangkok	208 (1.78%)	255 (2.19%)	11,207 (96.03%)	11,670

B. Data Preprocessing

The data preprocessing module consists of attribute selection, data grouping and transformation, one-hot encoding, and data imbalance handling. Data preprocessing preserved the accuracy and reliability of the results. Consequently, the dataset was narrowed down to focus on three accident severity levels or classes: Fatalities, Serious injuries, and Minor injuries.

1) Attribute Selection

Based on a review of the relevant literature, attributes were selected and refined to ensure their relevance to the objectives of the study. 28 irrelevant attributes—such as the reported date, reported time, geographic coordinates (latitude and longitude), route number, and accident code—were excluded from the analysis. As a result, 10 attributes were retained, as summarized in Table II. These attributes were selected for their potential influence on traffic accident patterns.

2) Data Grouping and Transformation

To reduce data sparsity and improve the efficiency of rule generation, the values within each attribute were grouped based on similar or related characteristics. This grouping aimed to minimize excessive variability within the dataset, which could otherwise hinder the derivation of meaningful patterns or rules. For instance, the "time of the accident" attribute was categorized into periods, and the "accident location" attribute—originally comprising 63 distinct values—was consolidated into 25 representative groups. Similarly, the "accident characteristics" attribute, which initially included 92 unique values, was grouped into 30 categories to improve interpretability. The severity of the accident was classified into three levels: Fatalities, if at least one death was reported; Serious injuries; and Minor injuries, in cases involving only minor injuries or no injuries at all. Table II details these grouped and recoded attributes. The dataset was divided by province to facilitate localized analysis.

TABLE II. DATA DESCRIPTION

Attributes	**Description**	**Ref.**
time of the accident	T1: 00:00-03:00, T2: 03:01-06:00 T3: 06:01-09:00, T4: 09:01-12:00 T5: 12:01-15:00, T6: 15:01-18:00 T7: 18:01-21:00, T8: 21:01-24:00	[1, 2, 7, 8, 10]
road type	HIG: Highway, National Highway under the jurisdiction of the Department of Highways, RUR: Rural road under the jurisdiction of the Department of Rural Roads, EXP: Expressway, Expressway under the jurisdiction of the Expressway Authority of Thailand	[9, 10]
month of the accident	M1…M12	[8]
types of vehicles involved in the accident	For example: MOT: motorcycle, PIT: pickup truck, CAR: car, 6WT: 6-wheel truck, TRA: trailer	[1, 8, 10]
accident location	Characteristics of the route in the area where the accident occurred, for example: ROA: road, SLR: slope road, WIC: wide curve	[9]
accident characteristics	For example: REA: rear-end, ROL: rollover, CWO: collision with obstacles (on the roadway), HEC: head-on collision, CWP: collision with pedestrians	[2, 9]
accident causes	The main cause of the accident, for example: ESL: exceeding the speed limit, PVA: pedestrian/vehicle/ananym suddenly crossing in front, DRO: drowsy driving, VED: vehicle defect	[2, 9]
weather	Weather conditions on the day of the accident: CLE: clear sunny, FOG: foggy, RAI: rainy, CLO: cloudy, OTH: other	[2, 9, 10]
accident severity levels	Accident severity level: FATAL: Fatalities, MAJOR: Serious injuries, MINOR: Minor injuries	[2, 9]

The characteristics of the dataset, after grouping the values of each attribute as illustrated in Figures 2 and 3, reveal that significant disparities exist in the frequency distribution of certain attribute values. Despite the efforts to consolidate similar categories, some attributes continue to exhibit substantial variation in the number of observations per category. This uneven distribution indicates the presence of data imbalance.

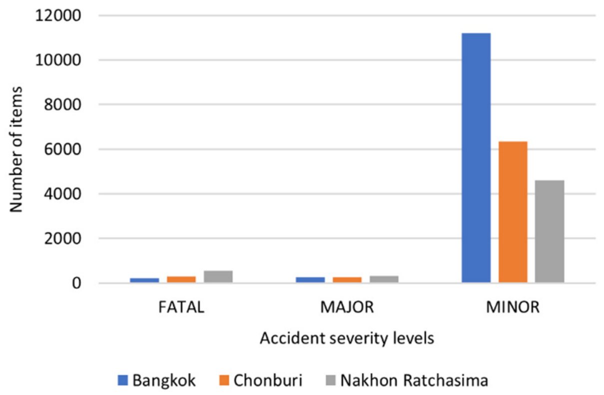

Fig. 2. Relationship between provinces and accident severity levels.

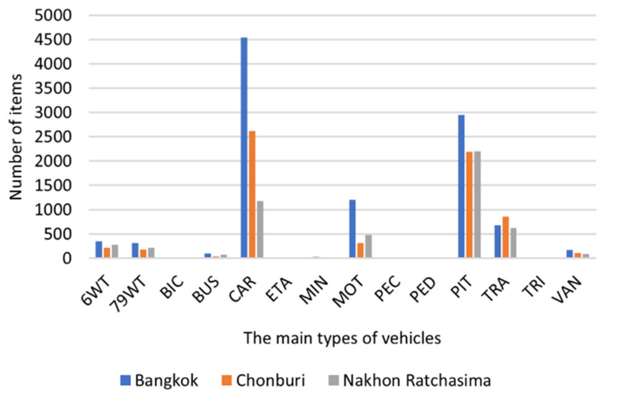

Fig. 3. Relationship between provinces and the main types of vehicles involved in accidents.

3) One-Hot Encoding

One-hot encoding is a widely used technique for transforming categorical variables into a binary numerical format, facilitating their use in ML models. This method generates separate binary columns for each unique category within a categorical attribute, where a value of '1' indicates the presence of that category and '0' denotes its absence. For example, the 'road type' attribute is transformed into three distinct binary columns: 'HIG' (highway), 'RUR' (rural road), and 'EXP' (expressway). During this transformation process, columns corresponding to NULL or meaningless values (other values) were also removed to ensure data quality and consistency. As a result, the dataset was expanded to a total of 132 binary attributes, which were then used for subsequent modeling and analysis.

C. Data Imbalance

Imbalanced datasets can lead to biased ML models that disproportionately favor majority classes, thus compromising the predictive accuracy and generalizability of the model, particularly for minority classes [14]. To address this issue, it is essential to implement resampling techniques during the model development process to ensure a more balanced and equitable distribution of all classes. Among various resampling strategies, techniques from the SMOTE family have shown considerable effectiveness in mitigating class imbalance [15]. In particular, the SMOTE-Tomek method has demonstrated superior performance compared to traditional oversampling or undersampling approaches [11, 12]. The SMOTE-Tomek technique is a hybrid approach that combines oversampling of minority classes using SMOTE with the removal of noisy instances through Tomek links. Specifically, SMOTE generates synthetic samples for minority classes to equalize class distribution, while Tomek link removal eliminates pairs of samples from different classes [11]. For implementation, this study utilized the imbalanced-learn Python package to handle imbalanced datasets [16], and the results are shown in Table III. The standard procedure for SMOTE-Tomek can be summarized as follows:

- Step 1: Given an imbalanced dataset, the SMOTE algorithm is applied to generate an extended dataset by synthesizing new samples for the minority class.
- Step 2: Tomek link removal is then performed to eliminate overlapping or ambiguous sample pairs, resulting in a cleaner and more structured dataset suitable for training.

TABLE III. DATA AFTER SMOTE-TOMEK

Provinces	**Number of individuals affected by severity levels**		
	Fatalities	**Serious injuries**	**Minor injuries**
Nakhon Ratchasima	4,582	4,585	4,569
Chonburi	6,334	6,336	6,330
Bangkok	11,194	11,195	11,184

D. Association Rules Mining

Once data preprocessing was completed, the next step involved the mining of association rules using the Apriori algorithm on the dataset of each selected province to extract meaningful association rules. Association rule mining aims to discover interesting relationships, patterns, or co-occurrences among attributes within datasets. The association rule mining process involves the following steps.

1) Step 1: Define Minimum Support and Confidence Thresholds

Support is used to identify frequent itemsets within the dataset, as it measures how often a particular itemset appears in the data. Confidence evaluates the reliability of an association rule, indicating how often items in the consequent appear in transactions that contain the antecedent. Setting appropriate minimum support and confidence values is crucial, as these thresholds directly influence both the quality and the number of rules generated. Choosing values that are too high may lead to missing important patterns, while values that are too low may produce a large number of less meaningful rules [17].

2) Step 2: Generate Frequent Items

The goal is to identify frequent item sets—combinations of items that occur together in the dataset with a frequency of more than the minimum support threshold. The Apriori algorithm is commonly used for this step, as it systematically explores item combinations, beginning with individual items and expanding to larger item sets. This step follows a similar approach to [9].

$$support\ A \Rightarrow B = support\ (A \cup B) = \frac{|set\ of\ records\ containing\ A \cup set\ of\ records\ containing\ B|}{total\ number\ of\ records} \quad (1)$$

3) Step 3: Generate Class Association Rules (CARs)

CARs are generated from the previously identified frequent item sets, as in [18, 19], who utilized CARs' mining for analyzing accident-related data. These rules are filtered such that the class label, in this case accident severity, appears as the consequent or Right-Hand Side (RHS) of the rule. The association rules are presented in the form $A \Rightarrow B$, where A is the antecedent or Left-Hand Side (LHS) and B is the consequent or RHS. The rules are evaluated based on a minimum confidence threshold. Only rules with confidence levels exceeding the specified threshold are considered for selection. Increased confidence values correspond to greater reliability. However, association rule mining can generate a large number of rules, making it necessary to apply filtering and evaluation techniques to extract the most relevant patterns, as noted in [20].

$$confidence(A \Rightarrow B) = \frac{support(A \cup B)}{support(A)} \quad (2)$$

4) Step 4: Rule Evaluation

The rules are evaluated based on a lift threshold. Lift ratio is the ratio between the support value of the rule with the antecedent and the consequent support value.

$$lift(A \Rightarrow B) = \frac{support(A \cup B)}{support(A) \times support(B)} \quad (3)$$

A lift value greater than one indicates a positive association between the antecedent and consequent, suggesting that the rule is both interesting and potentially useful. The higher the lift value, the stronger the association. Rules with lift values above this threshold are often referred to as strong or interesting rules [20].

5) Step 5: Prune Rules

After generating CARs, a pruning process is necessary to eliminate redundant, weak, or less meaningful rules. This step helps in reducing the number of rules and retaining only those that are most relevant and informative. According to [19], pruning enhances the interpretability and efficiency of the resulting rule set. The pruning process sorts the rules based on descending order of confidence, followed by the antecedent and consequent. Rules that share the same antecedent are then analyzed, and only the more representative rule (the superset) is selected. The final rule set is both concise and informative. For example:

- Rule1: {time of accident = 0:00-03:00, type of vehicle = motorcycle, accident characteristics = rear-end collision} ⇒ {weather = clear, accident level = fatalities, road type = highway} confidence = 0.89 lift = 3.24
- Rule2: {type of vehicle = motorcycle, time of accident = 0:00-03:00, accident characteristics = rear-end collision} ⇒ {accident level = fatalities, road type = highway} confidence = 0.89 lift 2.90

Since Rule 1 is a superset of Rule 2 (i.e., it includes all the consequent elements of Rule 2 plus more) and has a higher lift, Rule 2 is considered redundant and is removed. The final pruned rule set is therefore more concise, non-redundant, and easier to interpret.

III. RESULTS AND DISCUSSION

This study analyzes the relationship of rules by defining the rules on the right as severity level classes, analyzed by province (the rules are based on provinces).

A. Set Support and Confidence

To determine appropriate support and confidence thresholds for mining association rules from accident data, a series of experiments was conducted by varying the minimum support values from 0.01 to 0.04 and the confidence values from 0.4 to 0.7 for the three provinces. Figure 4 shows sample results from the Chonburi province. Notably, when support and confidence were set to 0.01 and 0.4, respectively, the maximum number of association rules (460) was generated. Among these, 18 distinct LHS attribute combinations were identified, aligning with the eight key accident-related attributes defined in this study. In contrast, increasing the support or confidence thresholds resulted in a significant decrease in the number of attributes, as summarized in Table IV.

TABLE IV. NUMBER OF ATTRIBUTES OF FATALITIES IN CHONBURI PROVINCE CLASSIFIED BY SUPPORT AND CONFIDENCE THRESHOLDS

Support thresholds	Confidence thresholds			
	0.4	0.5	0.6	0.7
0.01	18	10	10	10
0.02	9	8	8	7
0.03	7	7	7	6
0.04	6	5	5	5

The appropriate support and confidence thresholds in Bangkok and Nakhon Ratchasima use principles similar to those in Chonburi. The results revealed that Bangkok had support = 0.01 and confidence = 0.4 thresholds to construct the rules for predicting fatalities (322), serious injuries (543), and minor injuries (5,288). Nakhon Ratchasima had support and confidence thresholds of 0.01 and 0.4, respectively, to construct the rules for predicting fatalities (987), serious injuries (258), and minor injuries (3,485). The minimum support value is important in the association rule formation process [17]. Errors obtained when determining this resulting rule are not desired. The determination of this value is challenging, especially if the user is ignorant of the dataset characteristics, as a feature that was used as a reference in determining the value of minimum support.

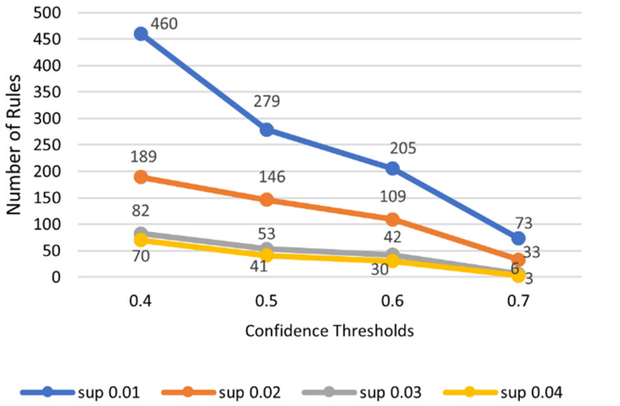

Fig. 4. Number of association rules categorized by support and confidence thresholds in Chonburi province.

B. Association Rules

The association rules were generated to predict accident patterns at three levels of severity across three provinces. Some rules were redundant. Rule pruning was performed to enhance predictive performance. As noted in [19], the rules are obtained from the rule mining algorithm, but some are redundant. Rule pruning is essential and reduces the number of rules. As a result of the rule pruning process, the number of association rules for each province, classified by severity levels, was reduced as follows.

1) Association Rules on Chonburi

A total of 460, 470, and 4,684 association rules were created corresponding to Fatalities, Serious injuries, and Minor injuries. After pruning, 302, 248, and 3,481 rules remained, respectively. Table V shows the example rules for each severity level. The association rules are described below.

- Fatalities: Rule (1) - Fatal injuries frequently occur between 12:01 and 3:00, with most incidents involving motorcycles. The predominant causes are rear-end collisions and speeding, particularly on highways. Rule (2) - A motorcycle accident occurred while driving over the speed limit on a straight road in clear weather conditions on a highway at midnight. Rule (3) - An accident occurred involving a pickup truck that was driving over the speed limit on a highway during rainy conditions, tooking place on a straight road and resulting in the vehicle overturning. Rule (4) - An accident involving a motorcycle occurred on a highway. The motorcycle was speeding, and although the road was straight, the accident happened while navigating a curve.
- Serious injuries: Rule (5) - The accident occurred in August on a highway. The accident occurred on a straight road. The accident occurred when the vehicle overturned and fell off the road on a curve on a clear day. Rule (6) - The accident occurred between 18:01 and 21:00. The accident happened in a car on a highway. Rule (7) - The accident occurred between 18:01 and 21:00 on a straight road. It was caused by speeding and resulted in a rear-end collision.
- Minor injuries: Rule (8) - Accidents caused by cars driving faster than the speed limit usually occur between 12:01 and 15:00 on straight roads on highways, with the most common accident type being rear-end collisions. Rule (9) - Accidents involving trucks with more than 10 wheels or trailers that overturn or fall off the road on a curve on a highway. Rule (10) - Accidents caused by defective vehicle equipment, and the result is a rollover or a fall off the road on a curve.

TABLE V. ASSOCIATION RULES CLASSIFIED BY ACCIDENT SEVERITY LEVELS IN CHONBURI PROVINCE

Antecedents	Consequents	Support	Confidence	Lift
1) MOT, REA, T1, CLE, HIG	FATAL	0.010316	0.890909	2.672446
2) ESL, ROA, MOT, T1, CLE, HIG	FATAL	0.016263	0.882857	2.648293
3) MOT, HIG, ROA, ESL, ROL	FATAL	0.017474	0.757991	2.273733
4) PIT, HIG, RAI, ROA, ESL, ROL	FATAL	0.010158	0.419565	1.258563
5) T8, ROA, ROL, CLE, HIG	MAJOR	0.011000	0.641104	1.922503
6) T7, CAR, HIG, ROA, CLE	MAJOR	0.012842	0.520256	1.560111
7) T7, HIG, ROA, CLE, ESL, REA	MAJOR	0.011526	0.496599	1.489169
8) ESL, T5, ROA, CAR, REA, HIG	MINOR	0.017000	0.952802	2.859912
9) ROL, CLE, TRA, HIG	MINOR	0.015105	0.950331	2.852495
10) VED, ROL, HIG	MINOR	0.011789	0.949153	2.848957

2) Association Rules on Bangkok

The accident patterns in Bangkok were extracted into 322, 543, and 5,288 rules for Fatalities, Serious injuries, and Minor injuries, respectively. After pruning, 246, 341, and 4,472 rules remained, respectively. Table IV summarizes example rules. The association rules are explained as follows.

- Fatalities: Rule (1) - The accident was a collision with a pedestrian on a clear day. Rule (2) - The accident occurred between 00:01 and 03:00 with a motorcycle on a rural highway. Rule (3) - The accident involved a motorcycle on a straight, steep road under clear weather conditions, located on a rural road.
- Serious injuries: Rule (4) - Accidents occur between 00:00 and 03:00 on a clear day on a rural road in a straight and steep area. Rule (5) - Accidents occur with motorcycles driving faster than the specified speed limit between 18:01 and 21:00. Rule (6) - Accidents involve pickup trucks driving faster than the specified speed limit on straight roads.
- Minor injuries: Rule (7) - An accident occurred in October involving a car on a straight road. Rule (8) - A car-overturned accident was caused by driving faster than the speed limit on the highway during rainy conditions. Rule (9) - An accident occurred involving a car driving faster than the specified speed limit in October on a clear day. Rule (10) - An accident occurred between 06:01 and 09:00. It was a rear-end collision on a straight road under clear weather conditions.

TABLE VI. ASSOCIATION RULES CLASSIFIED BY ACCIDENT SEVERITY LEVELS IN BANGKOK PROVINCE

Antecedents	Consequents	Support	Confidence	Lift
1) CWP, CLE	FATAL, HIG	0.019599	0.774118	2.886755
2) T1, MOT, RUR	FATAL	0.010514	0.769063	2.306571
3) CLE, MOT, SLR, RUR	FATAL	0.017067	0.506631	1.519487
4) T1, CLE, RUR, SLR	MAJOR	0.010157	0.736501	2.208714
5) ESL, MOT, T7	MAJOR	0.010008	0.724138	2.171638
6) CLE, PIT, ESL, T1, ROA	MAJOR	0.012838	0.550117	1.650751
7) ROA, M10, CAR	MINOR, HIG	0.011587	0.977387	3.320228
8) ROL, HIG, RAI, ESL	MINOR	0.011259	0.887324	2.663638
9) ESL, ROA, CAR, M10, HIG	MINOR	0.010246	0.974504	2.925343
10) CAR, ESL, REA, T3, ROA	MINOR	0.012391	0.941176	3.469652

3) *Association Rules on Nakhon Ratchasima*

The accident pattern of Nakhon Ratchasima was extracted into 987, 258, and 3,485 rules according to Fatalities, Serious injuries, and Minor injuries. After pruning, 688, 141, and 2,612 rules remained, respectively. Example rules are shown in Table VII. The association rules are explained as follows.

- Fatalities: Rule (1) - Accident involving a pickup truck driving faster than the specified speed limit. The accident was a collision in the opposite direction (not an overtaking). The location of the accident was a straight road on a clear day. Rule (2) - Accidents caused by driving faster than the specified speed limit. The accident was a collision with a pedestrian on the highway. Rule (3) - Accident during 03:01 and 06:00 on a clear day. The accident was a rear-end collision. The location of the accident was a straight road.
- Serious injuries: Rule (4) - A collision in the opposite direction (not an overtaking incident) caused by a person, vehicle, or animal cutting in front of the vehicle at close range on a highway. Rule (5) - A motorcycle accident, the location of the incident is a straight road, caused by a person, vehicle, or animal cutting in front of the vehicle at close range on a clear day on a highway. Rule (6) - An accident that occurs between 00:01 and 03:00 with a car, the incident is a rollover, and the location is a straight road.
- Minor injuries: Rule (7) - An accident occurred with a 6-wheel truck driving faster than the specified speed limit, resulting in an accident of overturning or falling off the road on a curve. Rule (8) - An accident occurred with a pickup truck between 12:01 and 15:00, caused by driving faster than the specified speed limit, resulting in overturning on a straight road. Rule (9) - An accident occurred on a rainy day between 12:01 and 15:00, caused by driving faster than the specified speed limit, resulting in an accident of overturning or falling off the road on a curve. Rule (10) - Accident on a clear day. The accident occurred on a straight highway because the driver fell asleep.

TABLE VII. ASSOCIATION RULES CLASSIFIED BY ACCIDENT SEVERITY LEVELS IN NAKHON RATCHASIMA PROVINCE

Antecedents	Consequents	Support	Confidence	Lift
1) ESL, PIT, ROA, HEC, CLE, HIG	FATAL	0.010119	0.891026	2.671132
2) ESL, ROA, CWP, HIG	FATAL	0.014779	0.886463	2.657454
3) CLE, T2, REA, HIG, ROA	FATAL	0.012449	0.766816	2.298775
4) HEC, PVA, HIG	CLE, MAJOR	0.010847	0.723301	2.427972
5) ROA, MOT, CLE, PVA, HIG	MAJOR	0.010629	0.648889	1.943978
6) ROL, CAR, ESL, HIG, T1, ROA	CLE, MAJOR	0.017036	0.479508	1.60961
7) ESL, 6WT, ROL	MINOR, HIG	0.011357	0.987342	3.038111
8) ESL, T5, PIT, ROA, ROL	MINOR, HIG	0.010920	0.980392	3.016729
9) ESL, RAI, T5, ROL	MINOR, HIG	0.011794	0.947368	2.915110
10) CLE, ROL, HIG, DRO, ROA	MINOR	0.013104	0.714286	2.147391

C. *Analysis of Traffic Accident Patterns*

The primary attributes contributing to road traffic accidents of all severity levels across the three provinces are as follows.

1. Driver-related attributes, particularly speeding, were identified as a major cause. This finding is consistent with [21], which stated that the most significant attributing risk factors were related to driver errors. Similarly, [8] noted that in most cases, fatal injuries happen due to harsh driving, inattentiveness, and speeding.
2. The time of accident occurrence, particularly between 00:00 and 06:00, was commonly associated with fatal and other severe accidents. When compared with [22], which analyzed the time of accidents in Bangkok, it was found that the time of fatal accidents was also at night. This observation corresponds with the findings of [23], who reported that accidents were more likely to occur at night due to inadequate street lighting or the absence of illumination on certain road segments, which reduced drivers' visibility of intersections and other hazards.
3. Accidents occurring on straight roads were also prominent. In [10], it was emphasized that straight roads without junctions pose a significant risk due to drivers often exceeding speed limits. This finding is consistent with [2], which stated that accidents frequently occurred on straight roads and often involve collisions with moving vehicles.

These key causal attributes are also associated with clear weather conditions, with moving vehicles, which tend to encourage careless driving behaviors and increase the risk of accidents.

In addition, a common pattern observed across all three provinces was that motorcycles were consistently involved in fatal accidents. In contrast, minor injury accidents occurred throughout the year, with vehicles involved including passenger cars, pickup trucks, and 6-10 wheel trucks.

The patterns of road traffic accidents across the three provinces differ in terms of fatalities and serious injuries. For fatal accidents in Chonburi, incidents involve primarily motorcycle riders, pickup trucks, and trucks with more than ten wheels. These accidents typically occur on highways, rural roads, and straight or wide curved roads, between 00:00 and 09:00, during April and December, under clear or rainy weather conditions. The common types of accidents included rear-end collisions, overturns, and incidents caused by speeding or sudden obstacles, such as people, animals, or vehicles crossing the road unexpectedly.

In Bangkok, fatal accidents almost exclusively involved motorcycle riders. These occurred on highways and rural roads, particularly straight roads, steep straight inclines, or wide straight roads, typically between 00:00 and 06:00, during October and January, and under clear weather conditions. The predominant accident types were rear-end collisions, overturns, and collisions with pedestrians. For Nakhon Ratchasima, fatal accidents involved both pickup trucks and motorcycles, occurring mainly on highways, with straight road accidents. These accidents typically occurred between 00:00 and 09:00, and between 15:01 and 21:00, during January and November, under clear weather conditions. Common accident types included head-on collisions with pedestrians, rear-end collisions, and overturns, often caused by speeding and sudden crossings by people, animals, or vehicles. These results are consistent with the findings of [24], which reported that the highest number of fatalities occurred among riders on dry straightways under clear daytime weather conditions.

For serious injury accidents, distinct patterns were observed in the three provinces. In Chonburi, serious injuries typically involved pickup truck and passenger car drivers, occurring on highways with straight road segments, between 00:00 and 06:00, and 15:01 and 24:00, during March, April, and September, under clear weather conditions. The most common type was vehicle overturning, mainly caused by speeding above the legal limit. In Bangkok, serious injuries involved motorcycle riders, pickup trucks, passenger cars, and trucks with more than ten wheels. These accidents occurred on highways and rural roads, particularly on straight roads, between 00:00 and 03:00 and 18:01 and 24:00, predominantly in December, under clear weather conditions. Typical types included rear-end collisions and overturns, with speeding again identified as the primary cause. In Nakhon Ratchasima, vehicles involved in serious injury accidents included motorcycles, passenger cars, and pickup trucks, occurring on rural highways along straight road sections, between 0:00 and 03:00, and 12:01 and 18:00, in March and December, under clear weather conditions. Accident types included head-on collisions, collisions with traffic cones, overturns, and rear-end collisions. The leading causes were sudden crossings by people, animals, or vehicles, as well as excessive speeding.

Given the varying patterns of accident occurrences, identifying interesting relationships and predicting associative behavior is essential to understanding accidents. Decision-makers must make informed strategic choices to effectively respond to and adapt to existing conditions, characterized by volatility, uncertainty, complexity, and ambiguity [17, 25].

IV. CONCLUSION

This study presents a system for discovering association rules in imbalanced road accident data. The findings are consistent with the research objectives. This study experimented with different values of support and confidence, and it was found that the optimal values for all three provinces were the same: a minimum support threshold of 0.01 and a minimum confidence of 0.4. These values were effective in generating association rules that could comprehensively predict traffic accidents based on causal attributes. Following the conceptual framework, association rules were derived and classified by severity levels: Fatalities, Serious injuries, and Minor injuries. Association rules were generated and categorized by severity levels, with Chonburi yielding 302 fatalities, 248 serious injuries, and 3,481 minor injuries rules, Bangkok yielding 246 fatalities, 342 serious injuries, and 4,472 minor injuries rules, and Nakhon Ratchasima yielding 688 fatalities, 141 serious injuries, and 2,612 minor injuries rules.

The association rules derived from road traffic accident data illustrate the relationship between province and accident severity level. A comparative analysis of common causal patterns across the three provinces reveals the following shared characteristics: (i) Fatal accidents are typically associated with motorcycles, occurring on highways, particularly along straight road segments, during the early hours between 00:00 and 06:00, under clear weather conditions, often involving speeding and resulting in rear-end collisions and vehicle overturns; (ii) Serious injury accidents are commonly linked to pickup trucks and passenger cars, occurring on straight roads, under clear weather conditions, with vehicle overturning as a typical outcome, and the main contributing factor being driving at excessive speeds; (iii) Minor injury accidents tend to involve passenger cars, pickup trucks, and trucks with more than 10 wheels, occurring between 00:00 and 21:00, across all months of the year. These incidents predominantly occur under clear weather conditions, on highways with either straight or wide curved sections, and are primarily caused by speeding.

Analysis of association rules related to road traffic accidents revealed key attributes associated with different levels of accident severity. These findings can inform policymakers in developing targeted policies and strategies to prevent or reduce traffic accidents. For instance, the high involvement of motorcycles in fatal accidents, or driver behaviors, such as speeding and drowsy driving, as well as vehicle-rated factors such as mechanical failures, were identified as significant contributors.

These insights can guide relevant authorities in formulating precise regulations or directives, such as promoting or mandating regular vehicle inspections—especially before travel—to ensure vehicles are safe and roadworthy. In addition, stricter enforcement of speed limits, installation of speed-limiting devices, and smart signage systems that alert drivers when exceeding speed limits can be implemented. In addition, engineering measures, such as speed humps, rumble strips to encourage speed reduction, and the installation of appropriate signage and road lighting, can help enhance driver awareness and reduce accident risks, especially in high-risk areas.

In future work, the effectiveness of the discovered association rules will be evaluated using new accident datasets through Classification Based on Associations (CBA), specifically employing CARs for predictive modeling. The performance of this approach will be compared with other ML techniques to identify the most suitable and effective model for road accident data. In addition, future studies may expand the scope of analysis by integrating datasets from other relevant agencies. This would enable the development of more comprehensive and context-sensitive safety measures that better reflect the actual circumstances of road accidents.

The data used in this study is based on road traffic accidents involving fatalities per volume of travel on road networks under the jurisdiction of the Ministry of Transport of Thailand. These data primarily focus on road-related features and do not include detailed information about drivers, such as age, gender, or driver behaviors (e.g., helmet use or seatbelt compliance). Additionally, the dataset does not cover road conditions, such as the presence or adequacy of street lighting, which limits the ability to analyze more granular contributing factors relevant to the design of targeted safety measures.

REFERENCES

[1] M. D. Belete, G. K. Alitasb, S. Nibretu, and M. E. Dessie, "Road traffic accident determinant factor identification in case of East Gojjam, Ethiopia using wrapper feature selection algorithm," *African Transport Studies*, vol. 3, 2025, Art. no. 100018, https://doi.org/10.1016/j.aftran.2024.100018.

[2] Q. Cai, "Cause Analysis of Traffic Accidents on Urban Roads Based on an Improved Association Rule Mining Algorithm," *IEEE Access*, vol. 8, pp. 75607–75615, 2020, https://doi.org/10.1109/ACCESS.2020.2988288.

[3] "Global status report on road safety 2023," *World Health Organization*. https://www.who.int/publications/i/item/9789240086517.

[4] "Report on Road Accident Situation Analysis on the Ministry of Transport's Road Network," Office of Transport and Traffic Policy and Planning, Bureau of Safety Planning, Ministry of Transport, Bangkok, Thailand, 2021.

[5] Y. Alassaf and Y. Said, "DPPNet: A Deformable-Perspective-Perception network for Safety Helmet Violation Detection," *Engineering, Technology & Applied Science Research*, vol. 14, no. 1, pp. 12659–12669, Feb. 2024, https://doi.org/10.48084/etasr.6633.

[6] A. S. Alkarim, A. S. Al-Malaise Al-Ghamdi, and M. Ragab, "Ensemble Learning-based Algorithms for Traffic Flow Prediction in Smart Traffic Systems," *Engineering, Technology & Applied Science Research*, vol. 14, no. 2, pp. 13090–13094, Apr. 2024, https://doi.org/10.48084/etasr.6767.

[7] I. C. Obasi and C. Benson, "Evaluating the effectiveness of machine learning techniques in forecasting the severity of traffic accidents," *Heliyon*, vol. 9, no. 8, Aug. 2023, Art. no. e18812, https://doi.org/10.1016/j.heliyon.2023.e18812.

[8] Shweta, J. Yadav, K. Batra, and A. K. Goel, "A Framework for Analyzing Road Accidents Using Machine Learning Paradigms," *Journal of Physics: Conference Series*, vol. 1950, no. 1, Dec. 2021, Art. no. 012072, https://doi.org/10.1088/1742-6596/1950/1/012072.

[9] Y. Pranata, T. B. Kurniawan, E. S. Negara, and A. H. Mirza, "Analysis of Traffic Accident Patterns Using Association Rule Mining," *Journal of Data Sciences*, vol. 2024, Dec. 2024, https://doi.org/10.61453/jods.v2024no63.

[10] I. Mohamad, R. Kasemsri, V. Ratanavaraha, and S. Jomnonkwao, "Application of the Apriori Algorithm for Traffic Crash Analysis in Thailand," *Safety*, vol. 9, no. 3, Aug. 2023, Art. no. 58, https://doi.org/10.3390/safety9030058.

[11] B. Or, "Improving Requirements Classification with SMOTE-Tomek Preprocessing." arXiv, 2025, https://doi.org/10.48550/ARXIV.2501.06491.

[12] M. H. Kotb and R. Ming, "Comparing SMOTE Family Techniques in Predicting Insurance Premium Defaulting using Machine Learning Models," *International Journal of Advanced Computer Science and Applications*, vol. 12, no. 9, 2021, https://doi.org/10.14569/IJACSA.2021.0120970.

[13] "Road accident data on the road network of the Ministry of Transport - Dataset – MOT data catalog." [Online]. Available: https://datagov.mot.go.th/dataset/roadaccident.

[14] A. Elsobky, A. Keshk, and M. Malhat, "A Comparative Study for Different Resampling Techniques for Imbalanced datasets," *International Journal of Computers and Information*, vol. 10, no. 3, pp. 147–156, Nov. 2023, https://doi.org/10.21608/ijci.2023.236287.1136.

[15] T. Tanantong and P. Loetwiphut, "Association rule mining framework for financial credit-risk analysis in peer-to-peer lending platforms," *Science, Engineering and Health Studies*, Dec. 2023, Art. no. 23020006, https://doi.org/10.69598/sehs.17.23020006.

[16] G. Lemaitre, F. Nogueira, and C. K. Aridas, "Imbalanced-learn: A Python Toolbox to Tackle the Curse of Imbalanced Datasets in Machine Learning." arXiv, 2016, https://doi.org/10.48550/ARXIV.1609.06570.

[17] E. Hikmawati, N. U. Maulidevi, and K. Surendro, "Minimum threshold determination method based on dataset characteristics in association rule mining," *Journal of Big Data*, vol. 8, no. 1, Dec. 2021, Art. no. 146, https://doi.org/10.1186/s40537-021-00538-3.

[18] M. Nasr, M. Hamdy, D. Hegazy, and K. Bahnasy, "An efficient algorithm for unique class association rule mining," *Expert Systems with Applications*, vol. 164, Feb. 2021, Art. no. 113978, https://doi.org/10.1016/j.eswa.2020.113978.

[19] J. Zhang, T. Feng, H. Timmermans, and Z. Lin, "Improved imputation of rule sets in class association rule modeling: application to transportation mode choice," *Transportation*, vol. 50, no. 1, pp. 63–106, Feb. 2023, https://doi.org/10.1007/s11116-021-10238-9.

[20] S. Katragadda *et al.*, "Association mining based approach to analyze COVID-19 response and case growth in the United States," *Scientific Reports*, vol. 11, no. 1, Sep. 2021, Art. no. 18635, https://doi.org/10.1038/s41598-021-96912-5.

[21] R. Abdulla, B. Qader, and K. Sdiq, "Traffic Accident Traits and Driver Characteristics Implication on Road Accidents using Descriptive Analysis: A Cross Sectional Study in Sulaymaniyah, Iraq," *Engineering, Technology & Applied Science Research*, vol. 13, no. 2, pp. 10372–10376, Apr. 2023, https://doi.org/10.48084/etasr.5669.

[22] I. Mohamad, S. Jomnonkwao, and V. Ratanavaraha, "Using a decision tree to compare rural versus highway motorcycle fatalities in Thailand," *Case Studies on Transport Policy*, vol. 10, no. 4, pp. 2165–2174, Dec. 2022, https://doi.org/10.1016/j.cstp.2022.09.016.

[23] S. Kumar and D. Toshniwal, "A data mining framework to analyze road accident data," *Journal of Big Data*, vol. 2, no. 1, Dec. 2015, Art. no. 26, https://doi.org/10.1186/s40537-015-0035-y.

[24] P. Iamtrakul and S. Chayphong, "GIS-based analysis of spatio-temporal clustering of road traffic accidents in Bangkok Metropolitan Region (BMR), Thailand from 2012 to 2021," *Transportation Research Interdisciplinary Perspectives*, vol. 31, May 2025, Art. no. 101489, https://doi.org/10.1016/j.trip.2025.101489.

[25] A. C. Amil, "Leadership decision-making in Vuca Bureaucracy: Global Turbulence, influence, challenges, and strategies," *International Journal of Multidisciplinary Research & Reviews*, vol. 3, no. 3, pp. 109–127, Jul. 2024, https://doi.org/10.56815/IJMRR.V3I3.2024/109-127.