Hindawi
Mathematical Problems in Engineering
Volume 2023, Article ID 6536768, 9 pages
https://doi.org/10.1155/2023/6536768

Research Article

Analysis and Detection of Road Traffic Accident Severity via Data Mining Techniques: Case Study Addis Ababa, Ethiopia

Demeke Endalie [1] and Wondmagegn Taye Abebe [2]

[1] *Faculty of Computing and Informatics, Jimma Institute of Technology, Jimma, Ethiopia*
[2] *Faculty of Civil and Environmental Engineering, Jimma Institute of Technology, Jimma, Ethiopia*

Correspondence should be addressed to Demeke Endalie; demeke.endalie@ju.edu.et

Received 4 March 2023; Revised 23 July 2023; Accepted 31 August 2023; Published 15 September 2023

Academic Editor: Ana C. Teodoro

Around the world, road traffic accidents are the leading cause of serious injuries and deaths. Ethiopia is one of the countries that suffer the most from traffic accidents. Every government in every country wants to keep its citizens safe from accidents. To keep people safe from accidents, it is necessary to conduct a detailed analysis of the factors that contribute to high-severity accidents and deaths. As a result, we developed a data mining algorithm-based road traffic accident severity analysis for the Addis Ababa subcity in this study. The longest frequent factors in the dataset were generated using the Apriori algorithm. The Apriori algorithm generates the most frequent factors as sex: male, driver–vehicle relationship: employee, weather condition: normal, pedestrian movement: not a pedestrian, road surface type: asphalt, and accident severity: high severity, with 42.21% and 84.35% support and confidence, respectively. In addition, we created an accident severity level predictive model using a support vector machine. The predictive model has an accuracy of 85%. The proposed predictive model outperforms other well-known predictive models, such as K-nearest neighbors, decision trees, and random forests. As a result, when making decisions or policies in Ethiopia, the government or private organizations should consider the association of factors that lead to serious severe accidents.

1. Introduction

A road traffic accident is an accident that occurs on a public road or street, killing or injuring people and involving at least one moving vehicle. Road traffic accidents present the greatest threat to personal security globally [1]. Every year, a traffic accident cuts approximately 1.25 million people's lives short [2]. In many countries, 3% of their gross domestic product (GDP) is lost due to road traffic accidents [3]. Road accident analysis aims to investigate the main factors that characterize an accident, understand patterns or behaviors, and identify appropriate countermeasures to avoid the accident.

Ethiopia has one of the world's worst records in road traffic accidents, ranking second among East African countries [4]. Road traffic accidents pose a huge development and health problem in Ethiopia. Although road traffic accidents are a major global public health problem, most occur in low- and middle-income countries, including Ethiopia. The Ethiopian government needs to address road traffic accident holistically, which requires involvement from multiple sectors (transport, police, health, education) that addresses the safety of roads, vehicles, and road users.

Data-driven decision-making is becoming increasingly popular in various fields [5]. It assists the user in determining the relationship that exists between items or attributes in the dataset. Learning from the pattern also gives a clue about what will happen in the future. To the best of the author's knowledge, no research on data-driven road traffic accident analysis has been conducted in Ethiopia. As a result, in this paper, we intend to use a data mining algorithm to examine the relationship between accident severity level and factors that cause the accident. The following are some studies related to our proposed road traffic analysis to extract the factors that cause high-severity accidents.

The work of Abegaz and Gebremedhin [6] examined the magnitude of road traffic accident-related injuries and fatalities using secondary data from a nationally representative survey conducted in Ethiopia in 2016. The study's findings conclude that road accident-related injuries and deaths are common and affect the productive sector of the population. In addition, males, individuals from better-off households, and vulnerable road users, including motorcyclists, pedestrians, and cyclists, are at increased risk of road traffic accidents.

Abdullah and Sipos [7] examine the severity of crashes by analyzing driver behavior and socioeconomic characteristics using a decision tree (DT) algorithm. Results show that the number of lanes, time of the accident, and human attitudes are the primary causes of accidents with victims. The Duhok city people participated in their survey, which was conducted in the Kurdistan area of northern Iraq. According to their study, over 30% of drivers who tend to drive faster than the speed limit are at risk of crashing.

Nidhi and Kanchana [8] proposed a road accident pattern prediction using Apriori and Naïve Bayesian techniques. Road accidents are an all-inclusive disaster with consistently rising patterns. There are different categories of vehicle accidents like rear end, head, and rollover accidents. Their statistical result shows that lower cities' rural mortality rate is higher.

A study stated by Comi et al. [9] uses data mining and clustering approaches to analyze accident data of the 15 districts of Rome municipality, collected from 2016 to 2019. Results show that such analyses can be a powerful tool to plan suitable measures to reduce accidents and to forecast the areas to be pointed out in advance.

The study by John and Shaiba [10] analyzes traffic accident data in Dubai for the year 2017 using data mining techniques and the Apriori algorithm. It finds that most accidents involve vehicle collisions due to inadequate space between vehicles. Youth is involved in the majority of accidents. The peak time for accidents is late at night, with most drivers intoxicated. Weekends have the highest number of accidents due to intoxication, while weekdays have the highest number due to inadequate space between vehicles. Recommendations to reduce accident rates are proposed based on the findings.

A fuzzy nonlinear programing study attempts to improve road traffic collision warning systems by developing a safety distance model to avoid rear-end collisions [11]. The method considers external environmental elements such as weather, road conditions, and vehicle speed to develop a mathematical model for safe distance overtaking. The simulation model is tested using fuzzy inference techniques to ensure that the model and parameter settings are reasonable. This method effectively eliminates false alarms and improves collision warning systems.

Multiagent systems are increasingly used to solve complicated problems with smaller task subdivisions [12]. Existing task planning strategies are inefficient and challenging to get optimal solutions. This work provides a multiagent control structure model that takes advantage of their advantages to complete complicated tasks. The technique enhances convergence and flexibility over existing strategies, achieving lower objective function values and better convergence. Regarding function value and obtained function values, it surpasses hierarchical task network planning (HTN) and time preference HTN.

The findings of the prior studies indicate that data-driven analysis of road traffic accidents is critical for forecasting accidents and identifying the most common factors in road accidents. As a result, in this study, we plan to analyze the relationship between the factors that cause road traffic accidents and accident severity levels and develop a severity level predictive model for the Addis Ababa subcity. The paper contributes two ideas. First, we generate the frequently occurring road traffic accident factors that coexist and result in high accident severity. Second, we created a model for predicting the severity of road traffic accidents in the Addis Ababa subcity.

The following describes the overall structure of the paper. Section 2 presents the state-of-the-art learning models. Section 3 describes the materials and methods used in this research. Section 4 presents the experiments, the results, and a discussion of them. Finally, Section 5 contains the study's conclusion.

2. Learning Model

Machine learning is an artificial intelligence offshoot that analyzes data to automate analytical model building. Machine learning suggests that, if properly trained, systems can identify patterns, learn from data, and make decisions with little or no human intervention [13]. The support vector machine, K-nearest neighbors (KNN), DT, and random forest (RF) are the most widely used machine-learning algorithms [14].

2.1. Support Vector Machine. Support vector machine, or SVM, is a popular supervised learning algorithm for classification and regression problems [15]. The SVM finds the best line or decision boundary for categorizing n-dimensional space to easily place new data points in the correct category in the future [16]. A hyperplane is the best decision boundary. SVM selects the extreme vectors that aid in the formation of the hyperplane. These extreme cases are known as support vectors. Consider the following diagram (Figure 1), which uses a decision boundary or hyperplane to classify two distinct categories and shows how the SVM algorithm determines the support vectors. The x-axis is a matrix of predictor data, with each row representing one observation and each column representing one predictor. The y-axis is an array of class labels, each denoting the value of the associated x-axis row.

2.2. KNN. KNN is a simple machine-learning algorithm that uses the supervised learning technique [17]. The KNN algorithm stores all available data and uses similarity to classify new data points. This means that when new data arrives, it can be quickly classified into a good suite category using the KNN algorithm. The KNN algorithm can be used for regression and classification, but it is most commonly used for classification problems. KNN is a nonparametric algorithm, which means it makes no assumptions about the underlying data. It is also known as a lazy learner algorithm because it does not immediately learn from the training set; instead, it stores the dataset and then acts on it during classification [18]. One of the difficulties with this technique is

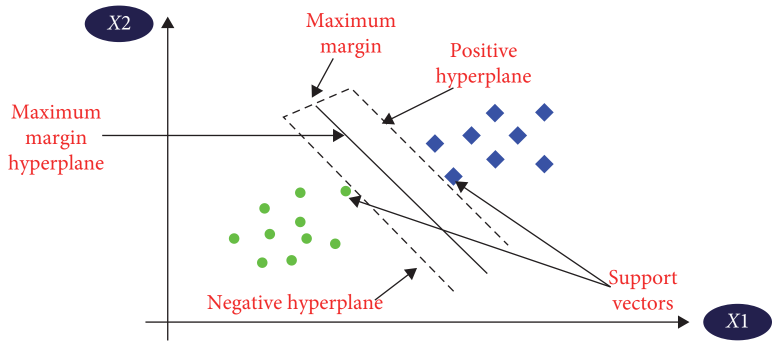

Figure 1: Decision boundary in SVM [15].

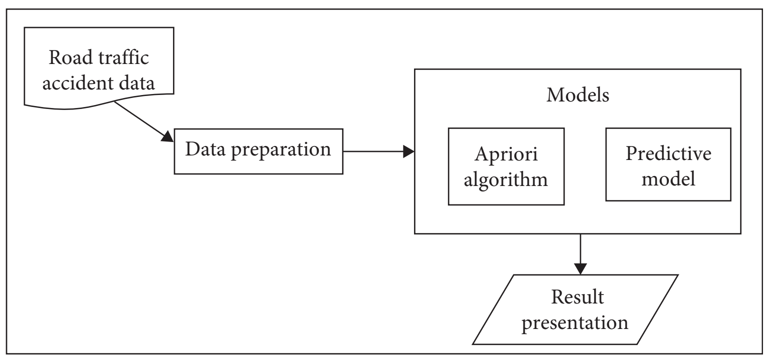

Figure 2: Proposed system architecture.

determining the "correct" value of K to go with a given labeled dataset [19].

2.3. DT. A DT is a supervised learning technique that can be used for both classification and regression problems, but it is most commonly used for classification. It is a tree-structured classifier in which internal nodes represent dataset features, branches represent decision rules, and each leaf node represents the result [20]. In a DT, the algorithm begins at the root node and works its way up to predict the class of a given dataset. This algorithm compares the values of the root attribute with the importance of the record (real dataset) attribute and then follows the branch and jumps to the next node based on the comparison. The algorithm compares the attribute value with the other subnodes and moves on to the next node. It repeats the process until it reaches the tree's leaf node.

2.4. RF. An RF algorithm is a supervised machine-learning algorithm widely used in classification and regression problems [21]. It combines the output of multiple DTs to reach a single result. In an RF tree, the greater the number of trees, the more robust it gets [22]. Similarly, the more trees in the RF algorithm, the more accurate and problem-solving capability it has. RF is a classifier that improves its predictive accuracy by taking the average of several DTs on different subsets of a given dataset. It is based on the concept of ensemble learning, which is the process of combining multiple classifiers to solve a complex problem and improve the model's performance.

3. Materials and Methods

Figure 2 depicts a high-level description of the methodology employed in this study. The proposed road traffic accident severity analysis consists of the following steps: dataset description, data preparation, association mining (Apriori algorithm), predictive model using SVM, and result evaluation.

The processed dataset is fed into the Apriori approach, as shown in Figure 2, to compute the most often coexisting road traffic accident risk factors, and then valid candidate association rules are constructed using the given minimum support and confidence levels. The Apriori algorithm generates the most frequently occurring accident risk factors in highly severe traffic incidents. Then, two significant outputs can be retrieved from this frequent set. The first output is a potential candidate association rule that meets the minimum support and confidence thresholds. The second will be the potential majority influential accident risk factors that cause extremely serious accidents. These risk indicators are then used to build the predictive model using the SVM learning model. The output from the two output modules can be used to judge road traffic incidents and the primary risk variables involved. Each component in the flow chart diagram depicted in Figure 2 is detailed below.

3.1. Data Used in the Study. The collection and preparation of the dataset utilized in this investigation were done by Bedane [23]. This dataset is collected from Addis Ababa subcity police departments. The dataset has been prepared from manual records of road traffic accidents. All the

Table 1: Predicates and the number of different values of the predicate.

Predicates	Number of different values of the predicate
Driver's age range	5
Sex	3
Educational level	7
Vehicle driver relation (vehicle ownership)	4
Driving experience	7
Type of junction	7
Road surface type	6
Light condition	4
Weather condition	9
Type of collision	10
Pedestrian movement	7
Vehicle movement	13
Cause of accident	20
Accident severity level	3

sensitive information has been excluded during data encoding. The dataset is available in the link: Road Traffic Accident Dataset of Addis Ababa City—Mendeley Data. The dataset was recorded in the Addis Ababa subcity from the year 2017 to 2020.

The dataset includes details about the drivers, weather conditions, infrastructure, and other events observed, and the accident's severity level. The collection consists of 12,316 instances of road traffic accidents and 15 factors of accidents in the study area. The dataset includes the history of 158 low-severity traffic accidents, 1,743 moderate (medium) severity traffic accidents, and 10,415 highly severe incidents. The dataset contains 11,437 traffic accidents involving male drivers, 702 involving female drivers, and 178 involving drivers of undetermined gender. Fourteen are risk factors (independent variables) that are substantially connected with the severity of a road traffic accident, and the 15th is the dependent variable (accident severity level). These include the driver's age range, sex, educational status, vehicle–driver relationship, driving experience, type of junction, type of road surface, light and weather conditions, type of collision, vehicle and pedestrian movement, cause of the accident, and level of accident severity. All of the above road traffic accident factors have two or more than two subcategories. For example, the driver's age range consists of five age groups: under 18, 18–30, 31–50, above 50, and unknown. Table 1 shows the basic predicate in the dataset and the number of categories in them.

3.2. Data Preparation. Data preparation is a critical step in the data analysis process. Because all standard machine learning and deep learning models operate on numerical values, we must modify each of the 14 road traffic accident factors. We identify the number of predicates identified under the 14 independent variables as subcategories before converting the given cleaned dataset to a numeric vector, as shown in Table 1. Three nested loops are used in the procedure to convert the dataset to a numeric vector. The first regulates the number of independent variables in the dataset, the second the number of instances, and the third the number of subcategories under each attribute. The program receives cleaned data as input and outputs numeric matrices that may be fed to machine-learning algorithms. This is demonstrated in Algorithm 1.

3.3. Apriori Algorithm. The Apriori algorithm is designed to work on transactional databases and generates association rules from frequent item sets [24]. It determines how strongly or weakly two objects are connected using these association rules. All possible association combinations were formed for each large itemset, and those with calculated confidence values greater than a predefined threshold were output as the association rule. This can be summarized as follows:

(i) To determine the minimum threshold values of support and confidence.

(ii) To find large item sets iteratively. The number of occurrences of the largest item set must be greater than or equal to the minimum support value determined at *i*.

(iii) To create the largest item set's association rules that meet the minimum support and confidence value.

The three steps outlined above are used to obtain the frequent itemsets and generate association rules from the dataset's frequent itemsets.

3.4. Evaluation Metrics. Here are some evaluation metrics we used to see how the proposed method performed with the dataset used in this study, as stated in Sharma et al. [25] and Li [26].

(1) *Support*: It is defined as the percentage of transactions in the dataset that comprises the itemset. Support calculates the frequency of association or how many times a specific item appears in a dataset. A frequent or large itemset obtains high support in the data. It can be expressed in probability theory as follows: $P(A\,B)$ = several transactions containing both A and B divided by the total number of transactions.

(2) *Confidence*: Confidence measures the strength of the association's rules. It is the ratio of transactions containing all items from a specific frequent item set to transactions containing all items from the subset. It determines how frequently item B appears in a transaction that includes item A. Confidence expresses an item's conditional probability. Confidence $= P(A \mid B)$.

(3) *Accuracy* (A): Accuracy measures how well the model fits the training samples [27]. Formally, accuracy is defined as Equation (1) follows:

$$\text{Accuracy} = \frac{\text{Number of correct prediction}}{\text{Total number of prediction}}. \tag{1}$$

(4) *Precision* (P) is defined as the percentage of correct positive predictions to total positive predictions. It is

```
List of categories
Began
for i in the range of zero to the total number of attributes
  for counter_1 in the range of zero to the total number of instances in the dataset
    for counter_2 in the range of zero to the total number of categories attribute_i
      if(attribute_i[counter_1] is not found in subcategories attribute_i)
        categories attribute_i[counter_2] = attribute_i[counter_1]
      endif
      if (attribute_i[counter_1] == subcategories attribute_i[counter_2])
        attribute_i[counter_1] = index of the name of the category in subcategories attribute_i
      endif
    endfor
  endfor
endfor
End
```

Algorithm 1: Conversion of each column of the dataset numeric.

sometimes referred to as positive predictive value. Mathematically, it can be computed as defined in Equation (2).

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}}. \quad (2)$$

(5) *Recall* (*R*): A recall represents the proportion of correctly categorized positive samples to total positive samples. Specificity is defined similarly as the fraction of correctly categorized negative samples compared to total negative samples and computed as shown in Equation (3).

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{Fn}}. \quad (3)$$

(6) *F1-score* (*F*): The F1-score combines a classifier's precision and recall into a single metric by taking their harmonic mean, and mathematically, it can be defined as shown in Equation (4).

$$\text{F1-score} = \frac{2\text{PR}}{P + R}, \quad (4)$$

where TP is true positives: when the actual and expected classes of a data point are both one. TN is true negatives: when a data point's real and forecasted classes are zero. False positives (FP) arise when a data point's actual class is zero and its predicted class is one. False negatives (FN) occur when a data point's real class is one, but its projected class is zero [28].

4. Experiment

4.1. Association Rule. Before applying an Apriori algorithm to our dataset to select the longest frequent item set coexist, we must determine the appropriate minimum support and confidence thresholds for the dataset. Determining a wrong value of support and confidence leads to the failure of the association rule to obtain the required rule [29]. The threshold points significantly impact the frequent itemsets and the association rule generated from those itemsets. A low value results in the inclusion of too many items in the rule-formation process. When an excessively high value is used, fewer items are involved, resulting in much data loss. Based on the algorithm described by Lee et al. [30], we adjust the minimum levels of support and confidence used in this study to 40% and 60%, respectively. Even though the dataset contains 105 predicate values, only 10 of them meet the minimum support we set for this study. Table 2 lists the 10 predicate values that satisfy the minimum support requirements and their predicate names.

Table 2: Predicate values that meet the minimum support requirements.

Predicate name	Value
Sex	Male
Educational level	Junior high school
Vehicle driver relation	Employee
Light condition	Daylight
Weather condition	Normal
Road surface type	Asphalt
Vehicle movement	Straight
Pedestrian movement	Not a pedestrian
Type of collision	A vehicle with the vehicle
Accident severity	High severity (2)

To find the longest frequent itemset from the above predicate values, the number of items in the itemset is increased by one at each iteration, and every combination of predicate values should be checked with minimum support. The following steps will use the combination of predicate values that

Table 3: The number of items in the frequent itemset with the total number of rules at each iteration.

Iteration number	Number of items in the itemset	Number combination of items
1	Two	45 item combinations
2	Three	33 item combinations
3	Four	30 item combinations
4	Five	11 item combinations
5	Six	1 item combination

Table 4: Rules generated from the frequent itemset and their confidence.

Rules	Confidence (%)
(*Gender: male*) ^ (*driver–vehicle relation: employee*) ^ (*weather condition: normal*) ^ (*pedestrian movement: not a pedestrian*) ^ (*road surface type: asphalt*) ⟶ (*accident severity: high*)	84.35
(*Gender: male*) ^ (*driver–vehicle relation: employee*) ^ (*weather condition: normal*) ⟶ (*pedestrian movement: not a pedestrian*) ^ (*road surface type: asphalt*) ^ (*accident severity: high*)	71.34
(*Gender: male*) ^ (*driver–vehicle relation: employee*) ^ (*weather condition: normal*) ^ (*pedestrian movement: not a pedestrian*) ⟶ (*road surface type: asphalt*) ^ (*accident severity: high*)	77.42

Table 5: Performance evaluation of the proposed model using several evaluation measures.

Learning model	Evaluation metrics			
	Accuracy (A)	Precision (P)	Recall (R)	F1-score (F)
SVM	85%	86.6%	84%	85.28%

SVM, support vector machine.

meet the defined minimum support. Table 3 presents the number of items in the frequent itemset and the number of possible item combinations with a support value greater than or equal to the threshold support value.

There is only one possible combination of predicate values with six items, as shown in Table 3. The longest combinations of road traffic accident factors are sex: male, driver–vehicle relation: employee, weather condition: normal, pedestrian movement: not a pedestrian, road surface type: asphalt, and accident severity: high severity, with 42.21% support. The following step is to generate rules from the aforementioned factors by the defined confidence value. Table 4 shows rules generated from the frequent itemset and their confidence value.

Finally, we can conclude that an employed male driver driving in normal weather conditions on an asphalt road and not a pedestrian is the cause of the high-severity accident in the Addis Ababa subcity. The correlation discovered in this study backs up some of the findings of prior studies, such as [10, 31]. This data-driven information shows that drivers, pedestrians, property owners, and traffic officers in the study area should be aware of the key risk variables that cause highly severe traffic accidents to mitigate the accident. In addition, when making decisions or policies in Ethiopia, the government or any other organization should consider the association of factors that lead to higher severity accidents.

4.2. Predictive Model. In addition to generating the most frequently occurring factors in high-severity road traffic accidents, we develop a road traffic accident severity level predictive model. SVM was used to create the predictive model. All experiments are conducted in a Windows 10 environment on a machine equipped with a Core i7 processor and 16 GB of RAM. The train–test split module is used to assess the performance of the suggested learning algorithm. It involves dividing the dataset into two subsets. The first subset fits the model, whereas the second is fed into the model after the prediction and comparison to the expected value. There is no commonly accepted splitting ratio in machine learning. However, the most commonly used dividing ratio is train: 80%, test: 20% [32]. As a result, we used an 80/20 train–test split ratio throughout the studies. Eighty percent of the dataset is used to train the model, while 20% is used to test the learned model. The predictive model based on SVM is built using the specified train–test splitting ratio. Table 5 shows the performance of the proposed predictive model in terms of accuracy, precision, recall, and F1-score.

The corresponding diagram (Figure 3) depicts the actual and anticipated accident severity level using the selected machine-learning model for the testing set. The diagram shows that the actual accident severity level and severity levels predicted by the proposed algorithm overlap frequently. This means that for 85% of the 60 traffic accidents instance testing dataset, the real and expected levels are the same, while for the remaining 15%, the model's predicted level and the actual level diverge. The x-axis represents the values of the testing instances, while the y-axis represents the dependent (accident severity level).

We evaluate the proposed predictive model with the testing dataset (20% of the dataset) in terms of predictive

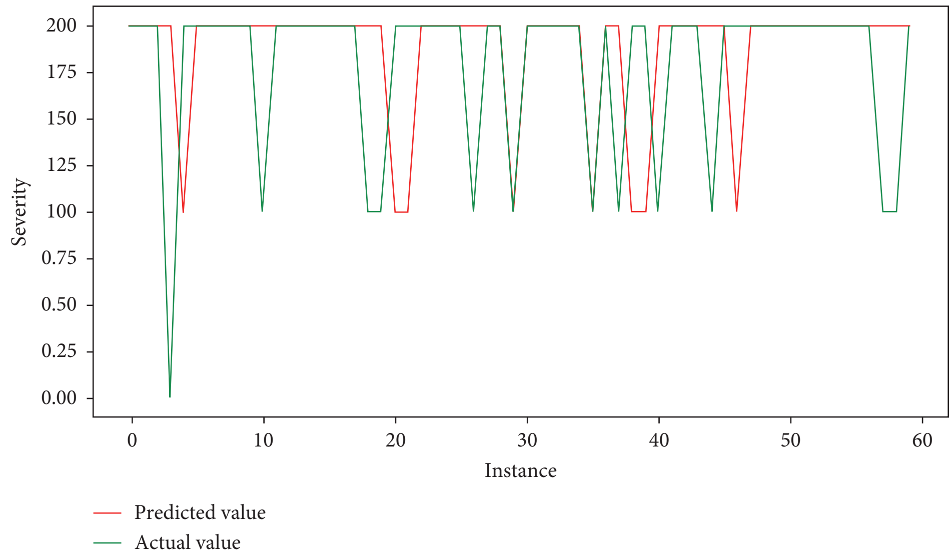

FIGURE 3: Evaluation of the proposed model using 60 instances of data.

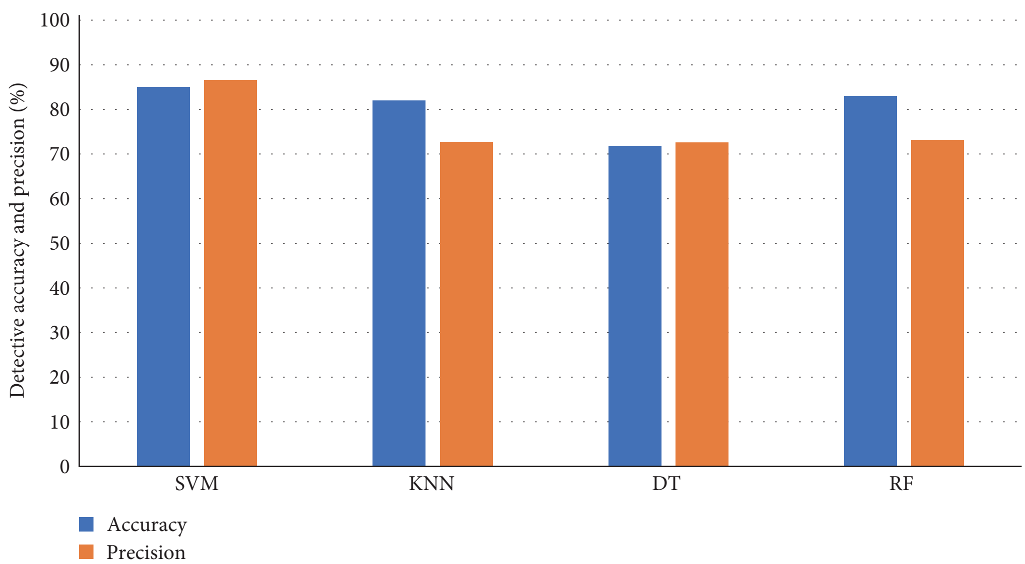

FIGURE 4: Comparison of predictive accuracy between SVM, KNN, DT, and RF.

accuracy. The developed model had a predictive accuracy of 85%. This implies that this model can make it easier for judges and traffic cops to determine the severity level of a traffic accident by observing the factors in that specific road traffic accident. To demonstrate that the proposed model is suitable for the severity level prediction task. We compared its accuracy to other state-of-the-art predictive models like KNN, DT, and RF. The hyperparameters of this predictive model are adjusted using a grid-searching strategy. Figure 4 shows the predictive accuracy of these selected models.

As shown in Figure 4, the proposed predictive model (using SVM) outperforms the other models regarding predictive accuracy and precision. We evaluated the proposed predictive model to those learning methods in terms of precision, recall, and F1-score, in addition to detection accuracy. Table 6 compares the experimental results with different models.

SVM exceeds KNN, DT, and RF classifiers in terms of detection accuracy, precision, recall, and F1-score, as demonstrated in Table 6. The results show that SVM outperformed other models by 2% of the remaining best (RF) and 7% in precision over the one that produced the best among the others, RF. This is due to the fact that (i) there is a distinct line between classes and (ii) this study's dataset is high dimensional and memory efficient. The study did not test deep learning models to predict accident severity levels from factors because the dataset was insufficient.

Table 6: A comparison of the model's performance in terms of various quality metrics.

Learning models	Evaluation metrics			
	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)
KNN	82	72.73	82	77
DT	71.8	72.6	71.7	72.14
RF	83	73.15	83	77.76
SVM	85	86.6	84	85.28

DT, decision tree; KNN, K-nearest neighbors; RF, random forest; SVM, support vector machine.

5. Conclusion

Road traffic accidents are a worldwide problem that affects every country. The situation is significantly worse in developing and low-income countries like Ethiopia. The main goal of this study is to investigate traffic accidents to determine the major factors contributing to high-severity accidents. The Apriori algorithm is applied to data collected from the Addis Ababa subcity in Ethiopia. We discovered hidden patterns by mining association rules with the Apriori algorithm, and we discovered attribute relationships by extracting rules. Following data analysis, we identified male drivers employed and driving on asphalt road surfaces under normal conditions, causing high-severity accidents with no pedestrians. In addition to this result, the severity level predictive model based on the recorded factors yields an acceptable result. The findings of this research can be used as one module for traffic accident management solutions. The outcome helps judges and traffic officers in determining the severity degree of a traffic accident by observing the circumstances involved in that specific road traffic accident and deciding the sort of punishment associated with that violation. Future research will investigate rural areas with animals involved.

Data Availability

The data can be obtained from the corresponding author (email: demeke.endalie@ju.edu.et) upon request.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

The authors thank Tarikwa Tesfa Bedane [23] for preparing the dataset and making it available for future research.

References

[1] S. Gopalakrishnan, "A public health perspective of road traffic accidents," *Journal of Family Medicine and Primary Care*, vol. 1, no. 2, pp. 144–150, 2012.

[2] D. A. Sleet, G. Baldwin, A. Dellinger, and B. Dinh-Zarr, "The decade of action for global road safety," *Journal of Safety Research*, vol. 42, no. 2, pp. 147-148, 2011.

[3] R. Suphanchaimat, V. Sornsrivichai, S. Limwattananon, and P. Thammawijaya, "Economic development and road traffic injuries and fatalities in Thailand: an application of spatial panel data analysis, 2012–2016," *BMC Public Health*, vol. 19, Article ID 1449, 2019.

[4] A. Honelgn and T. Wuletaw, "Road traffic accident and associated factors among traumatized patients at the emergency department of University of Gondar Comprehensive Teaching and Referral Hospital," *PAMJ Clinical Medicine*, vol. 4, Article ID 9, 2020.

[5] B. Marcinkowski and B. Gawin, "Data-driven business model development—insights from the facility management industry," *Journal of Facilities Management*, vol. 19, no. 2, pp. 129–149, 2021.

[6] T. Abegaz, S. Gebremedhin, and S. A. Useche, "Magnitude of road traffic accident related injuries and fatalities in Ethiopia," *PLOS ONE*, vol. 14, no. 1, Article ID e0202240, 2019.

[7] P. Abdullah and T. Sipos, "Drivers' behavior and traffic accident analysis using decision tree method," *Sustainability*, vol. 14, no. 18, Article ID 11339, 2022.

[8] R. Nidhi and V. Kanchana, "Analysis of road accidents using data mining techniques," *International Journal of Engineering & Technology*, vol. 7, no. 3, pp. 40–44, 2018.

[9] A. Comi, A. Polimeni, and C. Balsamo, "Road accident analysis with data mining approach: evidence from Rome," *Transportation Research Procedia*, vol. 62, pp. 798–805, 2022.

[10] M. John and H. Shaiba, "Apriori-based algorithm for dubai road accident analysis," *Procedia Computer Science*, vol. 163, pp. 218–227, 2019.

[11] F. Peng, Y. Wang, H. Xuan, and T. V. T. Nguyen, "Efficient road traffic anti-collision warning system based on fuzzy nonlinear programming," *International Journal of System Assurance Engineering and Management*, vol. 13, no. Suppl 1, pp. 456–461, 2022.

[12] M. Chen, A. Sharma, J. Bhola, T. V. T. Nguyen, and C. V. Truong, "Multi-agent task planning and resource apportionment in a smart grid," *International Journal of System Assurance Engineering and Management*, vol. 13, no. S1, pp. 444–455, 2022.

[13] I. H. Sarker, "Machine learning: algorithms, real-world applications and research directions," *SN Computer Science*, vol. 2, no. 3, Article ID 160, 2021.

[14] Demeke Endalie and Getamesay Haile, "Automated Amharic news categorization using deep learning models," *Computational Intelligence and Neuroscience*, vol. 2021, Article ID 3774607, 9 pages, 2021.

[15] D. A. Pisner and D. M. Schnyer, "Support vector machine," in *Machine Learning*, pp. 101–121, 2020.

[16] J. Cervantes, F. Garcia-Lamont, L. Rodríguez-Mazahua, and A. Lopez, "A comprehensive survey on support vector machine classification: applications, challenges and trends," *Neurocomputing*, vol. 408, pp. 189–215, 2020.

[17] S. Uddin, I. Haque, H. Lu, M. A. Moni, and E. Gide, "Comparative performance analysis of K-nearest neighbour (KNN) algorithm and its different variants for disease prediction," *Scientific Reports*, vol. 12, no. 1, Article ID 6256, 2022.
[18] M. A. jabbar, B. L. Deekshatulu, and P. Chandra, "Classification of heart disease using K-nearest neighbor and genetic algorithm," *Procedia Technology*, vol. 10, pp. 85–94, 2013.
[19] S. Zhang, "Challenges in KNN classification," *IEEE Transactions on Knowledge and Data Engineering*, vol. 34, no. 10, pp. 4663–4675, 2022.
[20] E. Szczerbicki, "Management of complexity and information flow," in *Agile Manufacturing: The 21st Century Competitive Strategy*, pp. 247–263, 2001.
[21] Y. Liu, Y. Wang, and J. Zhang, "New machine learning algorithm: random forest," *Information Computing and Applications*, vol. 7473, pp. 246–252, 2012.
[22] M. Savargiv, B. Masoumi, and M. R. Keyvanpour, "A new random forest algorithm based on learning automata," *Computational Intelligence and Neuroscience*, vol. 2021, Article ID 5572781, 19 pages, 2021.
[23] T. T. Bedane, *Road Traffic Accident Dataset of Addis Ababa City*, Addis Ababa Science and Technology University, Addis Ababa, 2020.
[24] N. Patil and D. Jagadale, "Analysis of road accidents using apriori, naive-bayes and K-means," *International Journal of Scientific & Engineering Research*, vol. 12, no. 3, pp. 181–185, 2021.
[25] M. Sharma, J. Choudhary, and G. Sharma, "Evaluating the performance of Apriori and predictive Apriori algorithm to find new association rules based on the statistical measures of datasets," *International Journal of Engineering Research & Technology (IJERT)*, vol. 1, no. 6, 2012.
[26] J. Li and Q. Zhang, "Assessing the accuracy of predictive models for numerical data: Not r nor r2, why not? Then what?" *PLOS ONE*, vol. 12, no. 8, Article ID e0183250, 2017.
[27] D. Endalie, T. Tegegne, and T. R. Gadekallu, "Designing a hybrid dimension reduction for improving the performance of Amharic news document classification," *PLOS ONE*, vol. 16, no. 5, Article ID e0251902, 2021.
[28] Z. Vujovic, "Classification model evaluation metrics," *International Journal of Advanced Computer Science and Applications*, vol. 12, no. 6, pp. 599–606, 2021.
[29] 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, 2021.
[30] Y.-C. Lee, T.-P. Hong, and W.-Y. Lin, "Mining association rules with multiple minimum supports using maximum constraints," *International Journal of Approximate Reasoning*, vol. 40, no. 1-2, pp. 44–54, 2005.
[31] J. Xi, Z. Zhao, W. Li, and Q. Wang, "A traffic accident causation analysis method based on AHP-Apriori," *Procedia Engineering*, vol. 137, pp. 680–687, 2016.
[32] V. R. Joseph, "Optimal ratio for data splitting," *Statistical Analysis and Data Mining: The ASA Data Science Journal*, vol. 15, no. 4, pp. 531–538, 2022.