Ordinal Random Tree with Rank-Oriented Feature Selection (ORT-ROFS): A Novel Approach for the Prediction of Road Traffic Accident Severity

Bita Ghasemkhani [1] , **Kadriye Filiz Balbal** [2] , **Kokten Ulas Birant** [3,4] **and Derya Birant** [3,*]

1 Graduate School of Natural and Applied Sciences, Dokuz Eylul University, Izmir 35390, Turkey; bita.ghasemkhani@ogr.deu.edu.tr
2 Department of Computer Science, Dokuz Eylul University, Izmir 35390, Turkey; kadriyefiliz.balbal@deu.edu.tr
3 Department of Computer Engineering, Dokuz Eylul University, Izmir 35390, Turkey; ulas@cs.deu.edu.tr
4 Information Technologies Research and Application Center (DEBTAM), Dokuz Eylul University, Izmir 35390, Turkey
* Correspondence: derya@cs.deu.edu.tr

Abstract: Road traffic accident severity prediction is crucial for implementing effective safety measures and proactive traffic management strategies. Existing methods often treat this as a nominal classification problem and use traditional feature selection techniques. However, ordinal classification methods that account for the ordered nature of accident severity (e.g., slight < serious < fatal injuries) in feature selection still need to be investigated thoroughly. In this study, we propose a novel approach, the Ordinal Random Tree with Rank-Oriented Feature Selection (ORT-ROFS), which utilizes the inherent ordering of class labels both in the feature selection and prediction stages for accident severity classification. The proposed approach enhances the model performance by separately determining feature importance based on severity levels. The experiments demonstrated the effectiveness of ORT-ROFS with an accuracy of 87.19%. According to the results, the proposed method improved prediction accuracy by 10.81% over state-of-the-art studies on average on different train–test split ratios. In addition, it achieved an average improvement of 4.58% in accuracy over traditional methods. These findings suggest that ORT-ROFS is a promising approach for accurate accident severity prediction, supporting road safety planning and intervention strategies.

1. Introduction

Road traffic accidents are being recognized as one of the most serious problems in the world as they result in deaths, injuries, and disabilities after treatment. Moreover, they cause public and economic losses each year. Accident severity is one of the major issues related to road safety that requires further research. Prediction of injury severity is one of the crucial and challenging problems in traffic safety management and control. The problem is formed as a classification task in which class labels represent the severity of the crash (slight, serious, or fatal). This information is vital for the development of road

safety policies and the implementation of preventive measures aimed at reducing both the occurrence and impact of accidents. As such, accurate severity prediction models can help enhance the agency's decision-making ability, contributing to safer road environments and efficient emergency response systems. They can also assist in targeted interventions, ensuring that high-risk areas receive more focused attention, while also aiding in the design of infrastructure that better accommodates varying traffic conditions and safety needs [1].

This work aims to develop a machine learning (ML) model using historical accident data that can accurately predict the severity level of a crash according to a set of influential factors. If important factors that are responsible for leading to traffic accidents can be better understood and forecasted, it can be possible to provide useful information about the damages and their severity [2]. Forecasting potential road traffic accidents with the aid of artificial-intelligence-based approaches can help to prevent them, warn drivers of potential dangers, perform effective crash management methods, or improve the emergency management process [3,4]. For instance, if the injury severity of a crash is predicted as serious, emergency response personnel might prepare the required equipment to elevate the efficiency of their response. However, despite the importance of this task, existing methods often face challenges in effectively handling the complexity of traffic accident data, particularly in accounting for the ordered nature of crash severity. Current models that typically rely on nominal classification approaches do not exploit the inherent ranking of severity levels (slight < serious < fatal), limiting their predictive accuracy.

Developing an accurate machine learning model for the prediction of road traffic accident injury severity is a challenging task. First, incident severity datasets are generally imbalanced, with fewer fatal classes than non-fatal ones. Overlooking this imbalance often leads to weak or biased classifiers that struggle to predict the minority class (high-severity crashes). Second, accidents are the result of non-linear and complex interactions between different factors such as human characteristics, road conditions, vehicle properties, and environmental elements [5]. Mathematically, accident data are typically characterized by high dimensionality, multicollinearity, and nonlinearity. The dataset includes numerous features such as driver-specific attributes (age, sex, education, years of experience, etc.), vehicle type and age, weather conditions, road surface type, light conditions (daylight or darkness), casualty class (driver, rider, pedestrian, passenger), road geometry, vehicle movement, junction types, and the area where the accident occurred (e.g., residential, industrial, rural areas) [6]. This study aims to determine the important factors that influence accident severity in order to refine road safety and the effectiveness of accident prevention strategies.

In some previous studies, traditional classification methods such as naive Bayes, random forest, and logistic regression have been employed for classifying accident seriousness [7]. However, these nominal classification methods ignore the inherent ordering of class labels that reflect the severity of accidents, such as slight injury < serious injury < fatal injury. This oversimplification may lead to misclassifications since it treats the severity levels just as distinct categories rather than recognizing the ordinal relationship between them. Ordinal classification (OC) [8] is a special type of supervised classification in machine learning in which an inherent ordering exists among the classes, such as low, medium, and high. For instance, class labels for a machine component can have ranking values such as healthy, low-risk, medium-risk, high-risk, and critical failure arranged from the most favorable situation to the most severe one. Similarly, traffic accident injury categories are ordered in nature from most to least severe; therefore, ordinal classification is necessary to increase the accuracy and effectiveness of severity estimation models. By incorporating ordinal relationships, these models can better capture the nuances of accident severity, providing more reliable outcomes for proactive traffic management and intervention.

One promising technique for handling ordinal classification problems is binary decomposition (BD) [9], which converts a multi-class ordinal problem into a series of binary classification tasks. In this technique, data dependencies are extended with binary sets for expressing the order relationships between class values. BD is a valuable approach for applications where understanding the gradation of outcomes is essential for accurate predictions. While BD simplifies ordinal classification tasks, the effectiveness of such models can be further boosted when combined with techniques like the synthetic minority oversampling technique (SMOTE) [10] to handle class imbalance, which is common in real-world ordinal datasets. By generating synthetic examples for underrepresented severity levels, SMOTE helps balance the class distribution and enriches machine learning models to project minority classes without bias.

In addition to addressing class imbalance, the relevance of input features plays a significant role in improving model proficiency. Feature selection (FS) [11,12] techniques are necessary for enhancing the precision of machine learning models by identifying and retaining the most relevant input features while eliminating irrelevant or redundant ones. Proper feature selection upgrades model accuracy, reduces overfitting, and speeds up computation. Traditional feature selection techniques may not fully capture the subtleties of relationships in ordinal classification problems, leading to inefficient outcomes. This is because the relationships between features and class labels are more complex and require preserving the intrinsic sequence of the target variable. Therefore, an advanced feature selection approach that aligns with ordinal classification is essential for performance optimization, enabling the model to better handle the complexities of forecasting ordered targets.

In this study, we propose a rank-oriented feature selection (ROFS) approach which is specifically tailored for ordinal classification tasks, selecting features whose values correspond to the ordered nature of the class labels. Unlike traditional feature selection that might overlook this ordering, ROFS mathematically affirms that chosen features reflect progressive or consistent trends across severity levels. For instance, in accident severity prediction, ROFS might prioritize features like weather conditions or driver age according to their ordered relationship with severity. For weather conditions, an increasingly severe pattern (e.g., from light rain to heavy snow) might correlate with higher accident severity levels. Similarly, driver age may follow an ordinal trend with severity, as very young or elderly drivers might show increased risk due to factors like inexperience or diminished reflexes. Conversely, ROFS might de-emphasize features like the ownership type of the vehicle if its impact on severity fluctuates without a clear progression, such as in cases where both personal ownership and firm ownership might correlate with varying severity levels. By focusing on features that capture the ordinal structure, ROFS develops the model's estimation reliability, yielding more reliable insights for traffic safety interventions.

The random tree (RT) [13] algorithm is well-suited for classification tasks, particularly when working with high-dimensional datasets that are common in traffic accident data. RT excels in capturing complex patterns and interactions between various features, which makes it particularly useful in the context of traffic accident severity prediction. Its advantages include robustness to overfitting and high interpretability, which is essential when domain experts need to understand and trust model functionalities. These properties are particularly beneficial when dealing with often noisy variables present in accident data, such as road conditions, weather, and driver behavior. However, RT faces some limitations, such as inefficiency with sparse data and challenges in handling imbalanced datasets, which can affect its performance in real-world scenarios. Additionally, its space complexity and reliance on larger datasets for optimal performance may pose constraints when resources or data are limited. To mitigate these issues, we employed techniques like SMOTE for balancing datasets and optimized model implementation on appropriate

hardware, ensuring practical feasibility. In this study, we adapted the random tree for ordinal tasks, where class labels are ordered in nature (e.g., slight, serious, fatal injury); thus, the random tree's ability to model non-linear relationships becomes even more valuable. The proposed approach (ordinal random tree) allows it to capture the ordered nature of accident severity and tends to provide correct predictions.

In this study, we propose a novel method, called ordinal random tree with rank-oriented feature selection (ORT-ROFS), to improve prediction accuracy for ordinal classification tasks. The problem in this study is to develop a machine learning model that accurately predicts traffic accident severity levels according to a set of influential factors such as weather conditions, driver characteristics, road variables, vehicle properties, and environmental elements. As a solution, we propose a method (ORT-ROFS) that utilizes the ordered nature of accident severity levels—ranging from slight to serious to fatal—both in the feature selection and prediction stages. In the ORT-ROFS method, the influential features, which demonstrate a clear relationship with the ordered classes, are selected, while irrelevant features are eliminated. This method aims to accurately predict accident severity levels, providing effective predictions for traffic safety management and policy-making. This study analyzes daily data on road traffic accidents (RTA) that occurred between 2017 and 2020 in Addis Ababa City, Ethiopia, serving as the real-world context for the application of the ORT-ROFS method.

The main contributions of the proposed method (ORT-ROFS) are summarized as follows:

- Novel ordinal classification method: Ordinal random tree (ORT) is introduced as an innovative approach for traffic accident severity prediction for the first time in the literature, addressing the limitations of traditional nominal classification methods.
- Handling ordinal complexity with binary decomposition: ORT uses binary decomposition (BD) to transform the multi-class ordinal task into simpler two-class problems, making it easier to model the progression in accident severity levels and boosting classification accuracy.
- Selecting features based on class orderings: It incorporates rank-oriented feature selection (ROFS), a new technique that chooses features based on the ordered progression of accident severity levels. Such selection advances the model's ability to differentiate between severity classes appropriately.
- Addressing class imbalance: Incident severity datasets often exhibit a class imbalance, where the number of fatal incidents is substantially lower than non-fatal ones. ORT-ROFS uses the synthetic minority over-sampling technique (SMOTE) to address this imbalance by augmenting the minority fatal classes, achieving more robust predictions.
- Providing explainability with a tree structure: ORT-ROFS builds a tree-based model, which can be easily interpretable, explainable, and understandable by humans while maintaining high predictive accuracy. Since the tree model is like a flowchart, ORT-ROFS can be seen as an explainable artificial intelligence (XAI) method.
- Enhanced prediction accuracy: Experimental results showed that ORT-ROFS achieved an average improvement of 10.81% over state-of-the-art methods. In addition, it demonstrated an average improvement of 4.58% in accuracy over traditional methods by considering orders among class labels both in the feature selection and prediction stages.

The organization of this paper is structured into six sections. Section 2 provides an overview of existing research on predicting traffic accident severity. In Section 3, we introduce the fundamental concepts behind the proposed method. Section 4 details the experimental studies, including information about the traffic accident severity data. Section 5 compares the functionality of ORT-ROFS with other state-of-the-art techniques. Finally,

Section 6 concludes the study by stating the main findings and suggesting potential works for future research to develop the field.

2. Related Works

In this section, the literature on road traffic accident severity is reviewed from different perspectives. The studies [14–28] are analyzed by categorizing them based on the regions where they were conducted, the specific tasks they addressed, the machine learning methods employed, and the evaluation metrics utilized.

Road traffic accident severity has been studied across various regions, including Ecuador [14], Canada [15], the United Kingdom [16,25,26], the Republic of Serbia and the Republic of Srpska [17], Saudi Arabia [18,20], the United States [19], Pakistan [21,22], Portugal [23], Korea [24,27], and Australia [28]. In Ecuador, data from the national public transport agency's 2023 records shows that 20,994 traffic accidents occurred nationwide, particularly in major cities like Guayaquil and Quito. Traffic accident risks were analyzed using data from Quito and its surroundings [14]. The dataset includes information on environmental conditions, traffic incidents, vehicles, and drivers. In Canada, traffic accident severity in Montreal was investigated using data collected between 2012 and 2021 [15]. In the United Kingdom, traffic accidents have been examined extensively, including the analysis of accidents from 2005 to 2014 [16], a review of 6515 accidents from 2005 to 2018 [25], and an examination of 10,000 accidents between 2011 and 2016 [26]. In the Republic of Serbia and Republic of Srpska, factors such as road type, speed limits, average daily traffic volume, and terrain type were considered in analyzing traffic accidents [17]. In Saudi Arabia, data from Qassim province (2017–2019) was used to analyze accident trends [18], while accidents in Al-Ahsa province (2016–2018) were examined to identify contributing factors [20]. In the United States, approximately 2.25 million traffic accidents on urban roads were analyzed for severity using data collected between 2016 and 2019 [19]. Accidents along Pakistan's national highway N-5 (2015–2019) were analyzed to identify causes and trends [21,22]. In Portugal, traffic accident data from the Setúbal region (2016–2019) were studied to determine contributing factors [23]. In Korea, hospital data was used to determine mortality rates from road traffic accidents [24], while accident severity on the Naebu highway in Seoul was analyzed to understand contributing factors [27]. Finally, in Australia, crash severity data from 74,909 traffic accidents in Victoria (2014–2019) was examined to uncover trends and underlying causes [28].

Road traffic accident severity has been analyzed for various objectives, depending on the dataset characteristics and study goals. Some analyses focused on regression tasks [17,27], while others addressed classification tasks [14–16,18–26,28]. Regression methods, both linear and nonlinear, were employed to identify influential factors and uncover relationships between them, mathematically modeling the underlying patterns in accident severity. They typically investigated the most significant factors affecting the number and severity of traffic accidents. Studies highlighted the importance of identifying key factors to mitigate accidents and enhance road safety [17,18,27].

In classification-focused research, the aim was to categorize complex traffic accident data into meaningful groups and predict future accident severity. Binary classification techniques, such as distinguishing between fatal and non-fatal accidents, were commonly applied [16,21–24,28]. Multi-class classification approaches were also utilized to differentiate between severity levels (e.g., slight-serious-fatal or low-medium-high-extreme) [14,15,19,20,25,26]. Some studies combined both binary and multi-class classification methods to deliver a comprehensive assessment of accident severity [18].

Traditional machine learning methods have been widely employed to analyze road traffic accident severity, including support vector machine (SVM) [14,19,24,26,28], naive Bayes (NB) [16,23], decision tree (DT) [23,27], random forest (RF) [14–16,18–21,23–25,27], k-nearest neighbors (KNN) [19,25], logistic regression (LR) [16,18,20,23–25], artificial neural networks (ANN) [14,16,17,19,26–28], and boosting algorithms [15,18,20–22]. Deep learning techniques, such as convolutional neural networks (CNN), were also applied in some studies [14].

Among ensemble learning methods, the RF algorithm was the most frequently used due to its capacity to handle high-dimensional data and deliver high accuracy. Boosting techniques like extreme gradient boosting (XGB) [15,18,20], categorical boosting (Cat-Boost) [15,22], adaptive boosting (AdaBoost) [21,22], and light gradient boosting machine (LGBM) [22] were employed in several studies. SVM and LR were also popular choices, as well as other standard classification algorithms such as DT, ANN, NB, and KNN. In terms of deep learning, CNN was specifically utilized to analyze accident severity in a study [14].

To evaluate model performance in regression tasks, mathematical metrics such as the coefficient of determination (r^2) [17], root mean square error (RMSE) [17,27], and mean squared error (MSE) [17,27] were commonly used. Lower RMSE and MSE values indicate better model performance, while a higher r^2 value signifies greater predictive accuracy. For classification tasks, frequently used evaluation metrics included accuracy [14–16,18,20–26,28], precision [15,16,18–26,28], recall [15,16,18–24], F-measure [15,16,18–22,24–26,28], specificity [14,20], sensitivity [14,20,26,28], area under the curve (AUC) [18,19,22,23], and receiver operating characteristic (ROC) analysis [18,24]. Among these, accuracy, precision, recall, and F-measure were the most commonly applied metrics for evaluating classification model performance. The AUC metric was particularly effective in representing a model's ability to distinguish between classes. Studies [14,20] also used specificity and sensitivity to assess model effectiveness, affording additional insights into classification performance. Furthermore, the ROC metric, employed in [14,20], was used to analyze the relationship between the false positive rate (FPR) and the true positive rate (TPR). Improved model performance was demonstrated by a larger area under the ROC curve, mathematically quantifying the model's discriminatory power.

Table 1 provides a comprehensive overview of related works, summarizing key aspects of the reviewed studies. The table includes columns for the year of publication (Year), geographic regions where the studies were conducted (Region), methods employed (Method), task types—classification (C) and regression (R), dataset time intervals (Period), performance metrics of the applied methods (Metric), and whether the study addressed ordinal classification tasks (Ordinal Classification). The metric column encompasses evaluation measures such as accuracy (ACC), precision (P), recall (R), F-measure (F), specificity (SPE), sensitivity (SEN), elapsed time (ET), coefficient of determination (r^2), and Chi-square (x^2), as applicable. Classification, regression, and ordinal classification tasks are denoted with a checkmark ($\sqrt{}$). This tabular representation makes available a concise yet thorough summary of methodologies, study regions, and evaluation criteria in road traffic accident severity research, enabling easy comparison and identification of trends across the literature.

Unlike the studies discussed earlier, this research adopts an ordinal classification approach, which is particularly suited for tasks like traffic accident severity prediction, where the predicted classes exhibit a natural order. Traditional classification methods, commonly used in prior studies, often overlook this inherent ordering, potentially leading to suboptimal predictions. To address this gap, the current study introduces the ordinal random tree with rank-oriented feature selection (ORT-ROFS) method. This novel approach leverages the ordinal nature of the problem to enhance prediction performance across multiple evaluation metrics, offering a significant contribution to the field.

Table 1. Summary of related works on road traffic accidents severity.

Reference	Year	Region	Method	C	R	Period	Metric	Ordinal Classification
Arciniegas-Ayala et al. [14]	2024	Ecuador	CNN, CNN-RF, GPC-RBF, SVM-RBF, ANN	√ *		Unspecified	ACC, SPE, SEN, ET	X
Muktar and Fono [15]	2024	Canada	XGB, CatBoost, RF, GB	√		2012–2021	ACC, P, R, F	X
Obasi and Benson [16]	2023	UK	NB, RF, LR, ANN	√		2005–2014	ACC, P, R, F	X
Gatarić et al. [17]	2023	Serbia Srpska	ANN		√	Unspecified	RMSE, MBE, MPE, x^2, r^2	X
Aldhari et al. [18]	2022	Saudi Arabia	XGB, RF, LR	√		2017–2019	ACC, AUC, ROC, P, R, F	X
Yan and Shen [19]	2022	USA	ANN, KNN, SVM, RF	√		2016–2019	AUC, P, R, F	X
Islam et al. [20]	2022	Saudi Arabia	LR, RF, XGB	√		2009–2016	ACC, SPE, SEN, P, R, F	X
Khattak et al. [21]	2022	Pakistan	RF, CART, LR, AdaBoost	√		2015–2019	ACC, P, R, F	X
Dong et al. [22]	2022	Pakistan	Natural GB, CatBoost, LGBM, AdaBoost	√		2015–2019	ACC, AUC, P, R, F	X
Santos et al. [23]	2021	Portugal	DT, RF, LR, NB	√		2016–2019	ACC, AUC, P, R	X
Boo and Choi [24]	2021	Korea	LR, RF, SVM	√		2013–2017	ACC, ROC, P, R, F	X
Fiorentini and Losa [25]	2020	UK	RT, KNN, LR, RF	√		Unspecified	ACC, TPR, FPR, TNR, F, P	X
Assi et al. [26]	2020	UK	ANN, FCM, SVM	√		2011–2016	ACC, SEN, F, P	X
Lee et al. [27]	2020	Korea	RF, ANN, DT		√	2007–2015	MSE, RMSE	X
Assi et al. [28]	2020	Australia	ANN, SVM	√		2014–2019	ACC, SEN, P, F	X
Proposed Method		Ethiopia	ORT-ROFS	√		2017–2020	ACC, P, R, F	√

* Classification, regression, and ordinal classification tasks are denoted with a checkmark (√), otherwise with (X).

3. Material and Methods

3.1. General Description of the Proposed Method

The problem in this study lies in accurately predicting traffic accident severity levels, which is challenging due to the complexity and imbalance of the data. The solution we proposed integrates the ordinal random tree (ORT) model with the rank-oriented feature selection (ROFS) approach, explicitly considering the ordered relationships among class labels.

Although standard classification methods have been used for predicting accident severity, they suffer from an important limitation: they ignore the relationship among class labels. Road traffic accident injury severity ranges from no or slight injury to severe injury and, lastly, to fatal injury. This ordered structure makes the classification problem appropriate to be treated as an ordinal task. Taking this into consideration, our study proposes a novel method, named Ordinal Random Tree with Rank-Oriented Feature Selection (ORT-ROFS), which is specifically designed to use order relationships among classes during both feature selection and prediction. To address the ordinal nature of the problem, ORT-ROFS transforms the multi-class classification problem into a set of binary classification problems. For a dataset with k ordinal classes, ORT decomposes the problem into $k - 1$ binary sub-problems. Each binary classifier is trained to differentiate samples above and below a particular class threshold. This decomposition allows the model to capture the ordering among classes, as each binary classifier addresses whether an instance falls within or above a certain severity level. This transformation is essential for maintaining the ordinal relationships and achieving improved prediction accuracy for ordered outcomes.

Figure 1 illustrates the overall architecture of the ORT-ROFS method, highlighting how the ordinal classification problem is handled through binary decomposition. The

process begins with addressing the imbalanced nature of the ordinal road traffic accidents dataset, where the class distribution for different severity levels is often skewed, particularly with fewer instances of more severe accidents (e.g., fatal injuries). Minor severity injury classes are identified by analyzing the dataset, and these classes are targeted for synthetic sample generation. The synthetic minority over-sampling technique (SMOTE) is applied to generate synthetic samples for the minority classes, balancing the dataset to guarantee that the model can learn from all severity levels. After that, binary decomposition is performed on the balanced ordinal dataset. This decomposition breaks down the problem into multiple binary sub-problems, each corresponding to a distinct severity threshold. By addressing the binary decisions at each level, the method is able to capture the ordered nature of the severity levels, confirming that the classification process respects the inherent hierarchy from slight to fatal injuries.
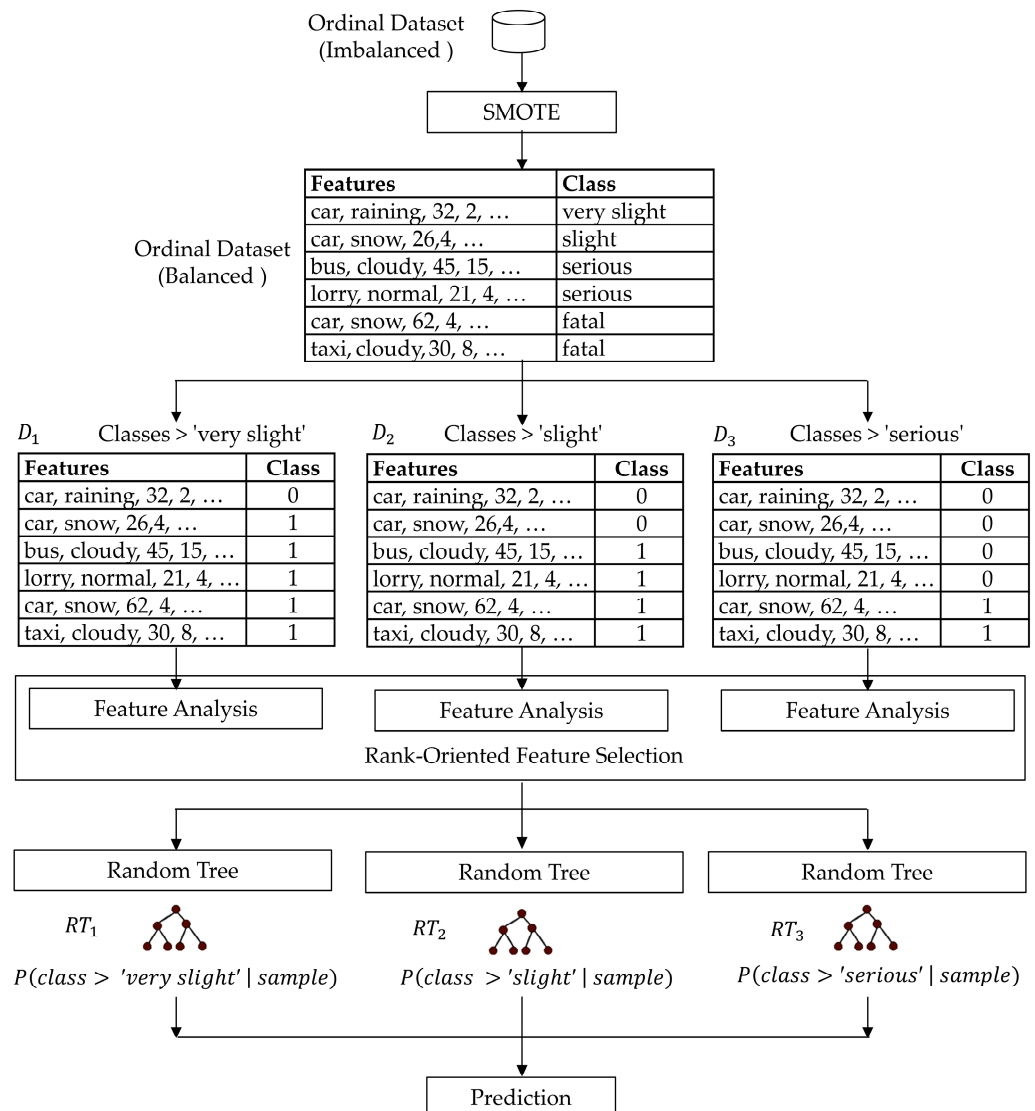

Figure 1. The general architecture of the proposed ORT-ROFS method.

The architecture in Figure 1 visually demonstrates how the model progressively distinguishes between classes at each threshold, thereby advancing predictive accuracy through a step-by-step evaluation of the severity levels. Assume that each sample in the dataset has numerous features (vehicle type, weather condition, age of driver, driving experience in years, etc.) and is assigned into one of four categories on an ordinal scale (very slight < slight < serious < fatal). The dataset is transformed by creating separate binary

datasets, including D_1, D_2, and D_3—each designed to capture the ordering information among the class labels by corresponding to a threshold in injury severity (e.g., classes greater than "very slight", "slight", and "serious"). For each threshold, a binary dataset is generated to distinguish whether a sample falls above or below that level of severity. For example, in the case of the classes > "very slight" threshold, samples with a severity of "very slight" are labeled as 0, while samples with "slight", "serious", or "fatal" injuries are labeled as 1. This binary labeling promotes that the classifier focuses on determining whether a sample surpasses the "very slight" injury level. Similarly, for subsequent thresholds, the process is repeated, with samples either being labeled as 0 or 1 based on whether their severity exceeds the next defined threshold.

After the decomposition step, each binary dataset undergoes rank-oriented feature selection to refine the feature set according to its respective severity level, strengthening the model to consider only the most relevant features at each threshold. In other words, the algorithm separately determines feature importance based on each severity level. Thus, features are selected by taking into account the ranking of class values. Here, the square root of the number of features is considered in each selection process.

After feature selection, each binary sub-problem is separately solved by constructing a random tree classifier, denoted as RT_1, RT_2, and RT_3. In the prediction phase, an unseen query instance is processed by each of the random tree classifiers, the probability of each on ordinal class labels is calculated, and finally the class with maximum probability is chosen as the result of the classification. For example, the second random tree (RT_2) mathematically calculates the probability P(class > "slight"|sample) that indicates the likelihood of the sample's severity being greater than "slight", including "serious" and "fatal". Such probabilities reflect the binary decisions made by the classifiers at each severity level. The maximum probability is then chosen, yielding the final severity classification for a particular instance. This multi-step process, as displayed in Figure 1, allows the ORT-ROFS method to leverage both the ordered nature of the labels and the strength of binary classifiers, enhancing prediction accuracy for road traffic accident severity levels.

3.2. Formal Description of the Proposed Method

3.2.1. Class Imbalance Handling with SMOTE

Road traffic accidents data are acutely imbalanced since the number of observations categorized in the slight injuries class (majority) is much higher than those categorized in the serious or fatal injuries class (minority). Therefore, there is a significant difference in the number of instances between the three classes that makes modeling difficult. When an imbalanced dataset is utilized to implement a classification task, the majority class dominates the classifier creation process, resulting in unsatisfactory prediction effectiveness. On the other hand, the minority class also provides noteworthy information that must be utilized. To handle such an unbalanced dataset, the synthetic minority oversampling technique (SMOTE) was applied in this study due to its simple procedure. The application of SMOTE involves identifying the subset of the dataset D that contains samples belonging to the minority class c_i, as mathematically presented in Equation (1):

$$D_{minor} = \left\{ (x_j, y_j) \in D \mid y_j = c_i \text{ and } |c_i| < threshold \right\} \tag{1}$$

Here, D_{minor} represents the subset of the dataset D containing samples from the minority class c_i. The condition $|c_i| < threshold$ guarantees that only underrepresented classes are targeted, where $|c_i|$ denotes the cardinality of class c_i (i.e., the number of samples in c_i). By balancing the dataset, this step establishes that the method effectively utilizes information from all classes, considerably improving prediction accuracy for minority classes. SMOTE works by generating synthetic instances for the minority class. Given a

minority instance $x \in D_{minor}$, SMOTE creates a synthetic instance by interpolating between x and one of its k-nearest neighbors as x_{NN}. The synthetic instance, $x_{synthetic}$, is computed through Equation (2) as follows:

$$x_{synthetic} = x + \sigma \times (x_{NN} - x) \quad \sigma \in [0, 1] \tag{2}$$

In this equation, x denotes a feature vector from the minority class and x_{NN} denotes one of its k-nearest neighbors. The synthetic instance, $x_{synthetic}$, is generated by interpolating between x and x_{NN}, with the interpolation factor controlled by σ, a random value between 0 and 1. This ascertains that the synthetic instance lies somewhere between the original instance and its neighbor in the feature space, with the degree of interpolation determined by σ. The technique balances the dataset by increasing the representation of the minority class, enabling the classifier to learn more effectively from both the majority and minority classes. This step heightens the productivity of the method, allowing it to more accurately predict traffic accident severity levels, especially for the underrepresented serious and fatal injury classes.

3.2.2. Ordinal Classification

Ordinal classification addresses problems where the target variable has a natural order. In this study, we predict the ordered severity levels of traffic accidents, characterized by class labels $Y = \{C_1, C_2, \ldots, C_k\}$, where $C_1 < C_2 < \ldots < C_k$. Given an input vector, the objective is to assign one of these ordinal labels by calculating probabilities for each class using Equations (3)–(5) as follows:

$$P(C_1) = 1 - P(Class > C_1) \tag{3}$$

$$P(C_i) = P(Class > C_{i-1}) \times (1 - P(Class > C_i)) \quad i = 2, 3, \ldots, k-1 \tag{4}$$

$$P(C_k) = P(Class > C_{k-1}) \tag{5}$$

Here, $P(C_1)$ represents the probability of an instance belonging to the first class, while $P(C_i)$ with $i = 2, 3, \ldots, k-1$ calculates probabilities for intermediate classes using threshold differences, and $P(C_k)$ provides the probability of the final class. The term $P(Class > C_i)$ denotes the probability that the instance belongs to a class higher than C_i. These equations affirm that the class with the highest probability is assigned to the instance, capturing the ordinal nature of the classification task. In other words, binary classifiers are used for each threshold between adjacent classes, preserving the ordinal structure of severity levels in traffic accidents. The ORT predicts the accident severity by selecting the class label with the highest probability for a given input instance, ensuring that the natural order of the severity levels is respected. To illustrate, consider three severity levels: C_1 = "Very Slight", C_2 = "Slight", and C_3 = "Serious". The estimation of the probability for the first ordinal class label depends on a single classifier: $1 - P(class > C_1) = 0.05$, as well as for the last ordinal class: $P(class > C_2) = 0.7$. Here, for class labels in the middle of the range, the probability depends on a pair of classifiers and is given by $P(class > C_1) \times (1 - P(class > C_2)) \cong 0.25$. The algorithm assigns the class with the highest probability, which in this case is C_3 as "Serious", demonstrating how the ordinal relationships are preserved while selecting the most probable class label.

3.2.3. Binary Decomposition

Binary decomposition is a powerful technique used to simplify complex classification problems by converting a multi-class ordinal classification problem into a series of binary classification tasks. This approach allows the model to focus on distinguishing between ordered categories, effectively breaking down the problem into manageable sub-tasks while preserving the ordinal relationships among the classes. For a given ordinal dataset

$D = \{ (x_i, y_j) \mid j = 1, 2, \ldots, n \}$ with n instances, where x_i signifies the input features, and y_j denotes the original ordinal class labels, the decomposition process constructs a binary dataset for each class threshold C_i. Each instance is assigned a binary label y_j', mathematically defined in Equation (6) as follows:

$$D_i = \left\{ (x_i, y_j') : y_j' = 0 \; if \; y_j \leq C_i, \; else \; y_j' = 1 \right\} \tag{6}$$

This approach divides the original problem into $k-1$ binary classification tasks, where k is the total number of ordinal classes. By reducing the number of classes in each sub-problem, binary decomposition simplifies the classification process and enables the use of binary classification algorithms. Binary decomposition has a major advantage because many well-established machine learning algorithms, such as decision trees, are inherently designed for binary classification, enabling their effective application to ordinal tasks. Each binary classifier focuses on distinguishing instances across a specific threshold C_i, validating that the ordinal structure of the data is preserved. Additionally, by simplifying the original multi-class problem into two-class sub-problems, binary decomposition reduces computational complexity and facilitates more precise modeling. After solving all $k-1$ binary classification tasks independently, the output of the model with maximum probability is chosen to predict the final severity level for each input, maintaining consistency with the ordinal structure of the target variable.

3.2.4. Rank-Oriented Feature Selection with Pearson Correlation

Rank-oriented feature selection (ROFS) aims to identify the most relevant features for each binary classification problem within the proposed framework. In the context of ordinal classification, where the target variable has a natural order (e.g., severity levels of traffic accidents), the objective of ROFS is to select features that capture meaningful relationships between the features and the target variable while maintaining the ordered structure of the problem. To achieve this, Pearson correlation [29] is used to assess the linear relationship between each feature and the ordinal target attribute. The Pearson correlation coefficient quantifies the strength and direction of the linear relationship between two variables, providing a measure of how well each feature correlates with the target. In mathematical terms, the calculation is performed through Equations (7)–(9) as follows:

$$r_{x,y} = \frac{cov(x,y)}{\sqrt{var(x) \times var(y)}} \tag{7}$$

$$cov(x,y) = \frac{1}{N} \sum_{i=1}^{N} (x_i - \overline{x}) \times (y_i - \overline{y}) \tag{8}$$

$$var(x) = \frac{1}{N} \sum_{i=1}^{N} (x_i - \overline{x})^2 \tag{9}$$

The Pearson correlation coefficient $r_{x,y}$ quantifies the linear relationship between a feature x and the target variable y by combining the covariance $cov(x,y)$, which measures how x and y vary together, and the variances $var(x)$ and $var(y)$, which capture the spread of each variable. The covariance is calculated as the average product of the deviations of x and y from their respective means $\overline{x}$ and $\overline{y}$ over N instances.

In the context of ROFS, features with high absolute Pearson correlation values are deemed the most relevant for each binary classification task. This indicates that features demonstrating a strong linear relationship with the target variable are selected. Pearson correlation offers several advantages, namely that it reliably measures dependency, effectively handles continuous data, and confirms both the strength and direction of relationships.

Pearson correlation is used to optimize the feature set for each binary decomposition of the ordinal classification task. This supports the idea that ROFS not only respects the natural ordering of the labels but also significantly increases the accuracy of resulting classifiers.

3.2.5. Random Tree Classifier

The random tree serves as the backbone for handling the binary classification tasks created during the binary decomposition process in the proposed ORT-ROFS method. For each binary dataset constructed based on specific class thresholds, a separate random tree classifier is trained to model the decision boundaries relevant to that threshold. The random tree algorithm is characterized by its hierarchical structure and inherent randomness, making it a powerful tool for classification tasks. It grows a tree recursively by splitting the dataset at each node based on an impurity reduction criterion, such as Entropy or Gini indices [30], which measure the purity of the data after each split. From a mathematical standpoint, these criteria are represented in Equations (10) and (11), respectively:

$$H(D) = -\sum_{j=1}^{k} p_j \times \log_2 (p_j) \tag{10}$$

$$G(D) = 1 - \sum_{j=1}^{k} p_j^2 \tag{11}$$

where p_j denotes the proportion of instances in the dataset that belong to the j-th class and k represents the total number of classes. The Entropy function in Equation (10) measures the uncertainty or disorder within the dataset, while the Gini index in Equation (11) quantifies the degree of impurity or misclassification. Lower values of either measure indicate purer splits, guiding the tree to better separate the classes.

In the ORT-ROFS method, random trees are built after the feature selection process, confirming that only the most relevant features are considered during tree construction. This approach combines the benefits of feature selection with the diversity and flexibility of random trees. The random tree classifier is chosen for its ability to capture nonlinear relationships between features and class labels, making it particularly effective for complex tasks like predicting traffic accident severity. Additionally, it offers a highly interpretable and understandable model, which is important for humans to rely on and manage decision results in the context of explainable artificial intelligence (XAI). By combining rank-oriented feature selection with the modeling power of random trees, ORT-ROFS provides an efficient approach to dealing with complex ordinal classification problems.

3.3. ORT-ROFS Algorithm

The ordinal random tree with rank-oriented feature selection (ORT-ROFS) algorithm (Algorithm 1) is a structured approach to solving ordinal classification problems by combining synthetic oversampling, binary decomposition, feature selection, and random tree modeling. The algorithm begins by handling class imbalance utilizing the synthetic minority oversampling technique (SMOTE). Minority classes are identified by analyzing the class distribution through counts, which records the number of instances for each class $c_i \in \{c_1, c_2, \ldots, c_k\}$. Classes with fewer instances than a predefined threshold are included in the D_{Minor}, which refers to minority classes. Synthetic samples are then generated for these underrepresented classes, and the resulting samples are added back to the dataset D to achieve balance. This step ensures that the model learns successfully from all classes, including the minority ones. Given an ordinal dataset $D = \{(x_1,y_1), (x_2,y_2), \ldots, (x_n,y_n)\}$ with n instances and ordered class labels $y \in \{c_1, c_2, \ldots, c_k\}$, the algorithm then decomposes the dataset into $k-1$ binary datasets D_i, where each dataset represents a threshold c_i separating adjacent classes. Instances in D_i are labeled

as 0 if their class $y_j <= c_i$, and 1 otherwise. Next, rank-oriented feature selection identifies relevant features for each binary dataset D_i using Pearson correlation, and these feature subsets are merged into a unified set *ROFS*, which is applied to the entire dataset. The next step involves training a random tree classifier for each binary dataset D_i, with the resulting models M_i aggregated into a collective model M^*. Finally, for predicting the class label of new instances in T, the algorithm evaluates the probability $P(c_j)$ of belonging to each class c_j based on ordinal-specific probability equations presented in the "Ordinal Classification" section. These probabilities capture the ordered relationships among the classes, ensuring that the ordinal nature of the problem is respected. The class with the highest calculated probability is then assigned as the predicted label for the instance, reflecting the most likely severity level based on the model's evaluation. This approach effectively encodes the ordering information among class labels in both feature selection and prediction to offer accurate predictions on ordinal data.

Algorithm 1: Ordinal Random Tree with Rank-Oriented Feature Selection (ORT-ROFS)

Inputs:
  D: the ordinal dataset with n instances such that $D=\{(x_1,y_1), (x_2,y_2), \ldots, (x_n,y_n)\}$
  Y: ordinal class labels $y \in \{c_1,c_2,\ldots,c_k\}$ with an order $c_1 < c_2 < \ldots < c_k$
  D_{minor}: the minority class(es)
  T: new instances to be predicted
Outputs:
  $\hat{Y}$: predicted class labels for the inputs in T

//Step 1—Synthetic Minority Oversampling Technique (SMOTE)
for $i = 1$ **to** k **do**
    foreach (x_j,y_j) **in** D
        counts$[y_i]+ = 1$
    end foreach
    if counts$[i] < threshold$
        Minor.Add(i)
    end if
end for
foreach class **in** D_{Minor}
    syntheticSamples = SMOTE (class)
    D.Add (syntheticSamples)
end foreach
//Step 2—Construction of Binary Datasets
for $i = 1$ **to** $k − 1$ **do**
    foreach (x_j,y_j) **in** D
        if $(y_j <= c_i)$
            D_i.Add $(x_j,0)$
        else
            D_i.Add $(x_j,1)$
        end if
    end foreach
end for
//Step 3—Rank-Oriented Feature Selection (ROFS)
ROFS $= \varnothing$
for $i = 1$ **to** $k − 1$ **do**
    FS = FeatureAnalysis (D_i)
    ROFS = ROFS U FS
end for
D.Apply (ROFS)
//Step 4—Construction of Models
for $i = 1$ **to** $k − 1$ **do**
    M_i =RandomTree (D_i)
    $M^* = M^*$ U M_i
end for
//Step 5—Classification of New Samples in T
foreach x **in** T
    $y = M^*(x) = $ MAX
    $P(c_1) = 1 − P(A_x \succ c_1)$
    for $i = 2$ **to** $k−1$ **do**
        $P(c_i) = P(A_x \succ c_{i-1}) \times (1^- P(A_x \succ c_i))$
    end for
    $P(c_k) = P(A_x \succ c_{k-1}))$
    $\hat{Y} = \hat{Y} \cup y$
end foreach
End Algorithm

The time complexity of the ORT-ROFS method is $O((n\log n) \times (k - 1))$, where n is the number of instances and k is the number of classes. This complexity aligns with the base random tree algorithm, reflecting the efficiency of our proposed method.

3.4. Dataset Description

This study utilizes the Road Traffic Accident dataset, a comprehensive resource including information about traffic accidents that occurred in Addis Ababa, Ethiopia, between 2017 and 2020. The dataset is publicly accessible via the Mendeley Data [31] and Kaggle [32] repositories. The dataset contains 12,316 records with 32 features describing various aspects of each accident. These features deliver detailed insights into factors influencing accidents and their outcomes.

Table 2 summarizes the dataset attributes and their respective categories, providing an overview of the key features used in the analysis. Key attributes include driver-related information such as age, gender, education level, and driving experience, as well as details about the accident site, such as road type, surface conditions, lighting, and weather conditions. Additionally, the dataset records the number and types of vehicles involved, the number of injuries, and the severity of injuries. This study specifically focuses on injury severity, classified into three ordered categories: slight, serious, and fatal.

Table 2. Summary of dataset attributes and their respective categories.

No	Feature Name	Description	Values
1	Hour	The time the accident occurred	Numeric
2	Day_of_week	The day the accident occurred	Monday, Sunday, Friday, Wednesday, Saturday, Thursday, Tuesday
3	Age_band_of_driver	The age range of the driver	'Under 18', 18–30, 31–50, 'Over 51'
4	Sex_of_driver	The gender of the driver	Female, Male
5	Educational_level	The education level of the driver	'Junior high school', 'Above high school', 'Elementary school', 'High school', Illiterate, 'Writing and reading'
6	Vehicle_driver_relation	The driver's relationship to the vehicle involved in the crash	Employee, Owner, Other
7	Driving_experience	The driving experience of the driver involved in the accident	1–2 year, 'Above 10 yr', 5–10 yr, 2–5 yr, 'No License', 'Below 1 yr'
8	Type_of_vehicle	Type of vehicle involved in the accident	Automobile, 'Public (>45 seats)', 'Lorry (41–100 Q)','Public (13–45 seats)', 'Lorry (11–40 Q)','Long lorry', 'Public (12 seats)', Taxi, 'Ridden horse', 'Pick up to 10 Q', Station wagon, Turbo, Bajaj, Motorcycle, 'Special vehicle', Bicycle, Other
9	Owner_of_vehicle	The ownership type of vehicle	Owner, Governmental, Organization, Other
10	Service_year_of_vehicle	The time elapsed since the vehicle's last service before the accident	'Above 10 yr', 5–10 yrs, 1–2 yr, 2–5 yrs, 'Below 1 yr'
11	Defect_of_vehicle	The defect status of the vehicle before the accident	'No defect', 7, 5
12	Area_accident_occured	The area where the accident occurred	'Office areas', 'Residential areas', 'Recreational areas',' Industrial areas', 'Industrial areas', 'Church areas', 'Market areas', 'Rural village areas', 'Hospital areas', 'Outside rural areas', 'School areas', 'Recreational areas', 'Rural village areas Office areas', Other
13	Lanes_or_Medians	The type of lane in which the vehicle was traveling at the time of the accident	'Double carriageway (median)', 'Undivided Two way', 'One way', 'Two-way (divided with broken lines road marking)', 'Two-way (divided with solid lines road marking)', Other
14	Road_allignment	The terrain of the road where the accident occurred	'Tangent road with flat terrain', 'Tangent road with mild grade and flat terrain', Escarpments, 'Tangent road with rolling terrain', 'Gentle horizontal curve', 'Tangent road with mountainous terrain and', 'Steep grade downward with mountainous terrain', 'Sharp reverse curve', 'Steep grade upward with mountainous terrain'
15	Types_of_Junction	The type of road junction where the accident occurred	'No junction', 'Y Shape', Crossing, 'O Shape', 'T Shape', 'X Shape', Other
16	Road_surface_type	The type of road surface on which the accident occurred	'Earth roads', 'Asphalt roads', 'Gravel roads', 'Asphalt roads with some distress', Other
17	Road_surface_conditions	The condition of the road surface	Dry, Snow, 'Wet or damp', 'Flood over 3 cm. deep'
18	Light_conditions	The lighting conditions when the accident occurred	Daylight, 'Darkness—lights lit', 'Darkness—no lighting', 'Darkness—lights unlit'

Table 2. *Cont.*

No	Feature Name	Description	Values
19	Weather_conditions	Weather conditions at the time of the accident	Normal, Raining, Cloudy, 'Raining and Windy', Windy, Snow, 'Fog or mist', Other
20	Type_of_collision	The manner in which the vehicles collided	'Collision with roadside-parked vehicles', 'Vehicle with vehicle collision', 'Collision with roadside objects', 'Collision with animals', Rollover, 'Fall from vehicles', 'Collision with pedestrians', 'With Train', Other
21	Number_of_vehicles_involved	The number of vehicles involved in the accident	Numeric
22	Number_of_casualties	The number of accident-related deaths	Numeric
23	Vehicle_movement	The driver's behavior just before the accident	'Going straight', U-Turn, 'Moving Backward', Turnover, 'Waiting to go', 'Getting off', Reversing, Parked, Stopping, Overtaking, 'Entering a junction', Other
24	Casualty_class	The classification of the person injured or killed in the accident	'Driver or rider', Pedestrian, Passenger
25	Sex_of_casualty	Gender of the person injured or killed in the accident	Male, Female
26	Age_band_of_casualty	The age range of the person injured or killed in the accident	31–50, 18–30, 'Under 18', 'Over 51', 5
27	Casualty_severity	The numeric value that indicates seriousness of the injury	Numeric
28	Work_of_casualty	The employment status of the person injured or killed in the accident	Driver, Unemployed, Employee, Self-employed, Student, Other
29	Fitness_of_casualty	The health condition of the person injured or killed in the accident before the accident	Normal, Deaf, Blind, Other
30	Pedestrian_movement	If a pedestrian was involved, the pedestrian's movement and location at the time of the accident	'Not a Pedestrian', 'Crossing from drivers nearside', 'Crossing from nearside—masked by parked or stationary vehicle', 'Unknown or other', 'Crossing from offside—masked by parked or stationary vehicle', 'In carriageway, stationary—not crossing (standing or playing)', 'Walking along in carriageway, back to traffic', 'Walking along in carriageway, facing traffic', 'In carriageway, stationary—not crossing (standing or playing)—masked by parked or stationary vehicle'
31	Cause_of_accident	Cause of the accident	'Moving Backward', Overtaking, 'Changing Lane to the left', 'Changing Lane to the right', Overloading, 'No priority to vehicle', 'No priority to pedestrian', 'No distancing', 'Getting off the vehicle improperly', 'Improper parking', Overspeed, 'Driving carelessly', 'Driving at high speed', 'Driving to the left', Overturning, Turnover, 'Driving under the influence of drugs', 'Drunk driving', Other
32	Accident_severity	Severity of the accident	'Slight Injury', 'Serious Injury', 'Fatal Injury'

4. Experimental Studies

In the experimental studies, the results were obtained using a real-world dataset of road traffic accidents (RTA), illustrating the practical relevance of the method to traffic safety management. When studying crash severity using historical traffic accident records, severity level is regarded as a dependent (class) variable, whereas other features are referred to as independent variables (predictors). The data was organized into injury severity levels, which are classified into slight injuries, serious injuries, and fatalities.

To evaluate the performance of ORT-ROFS, we used 10-fold cross-validation, which is a robust technique for estimating the model's generalization ability. In 10-fold cross-validation, the dataset is randomly divided into 10 equal-sized folds. The model is trained on nine of these folds and tested on the remaining one. This process is repeated 10 times, with each fold serving as the test set once. The final performance metrics are averaged across all 10 folds, providing a reliable estimate of how the model will perform on unseen data.

Experimental studies were conducted to evaluate the performance of the proposed ORT-ROFS method through various metrics, including accuracy, precision, recall, and F-measure, by comparing it with alternative classification approaches. Additionally, we present a detailed analysis using a confusion matrix derived from ORT-ROFS classifications to establish a comprehensive evaluation of the method's performance. The metrics are defined using common terms such as true positives (TP), true negatives (TN), false positives

(FP), and false negatives (FN), which are utilized to calculate the values for each metric. The mathematical expressions for these metrics are included in Equations (12)–(15), as follows:

- Accuracy: It measures the proportion of correctly classified instances among the total instances. It is a simple and widely used metric to evaluate the overall performance of a machine-learning model.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{12}$$

- Precision: It evaluates the proportion of true positive predictions among all positive predictions made by the model. It is particularly useful in imbalanced datasets.

$$Precision = \frac{TP}{TP + FP} \tag{13}$$

- Recall: It is also known as sensitivity or true positive rate, and measures the proportion of actual positive instances correctly identified.

$$Recall = \frac{TP}{TP + FN} \tag{14}$$

- F-measure: It is the harmonic mean of precision and recall metrics, providing a balance between the two. It is particularly useful when the dataset is imbalanced.

$$Fmeasure = 2 \times \frac{Precision \times Recall}{Precision + Recall} \tag{15}$$

The implementation of the ORT-ROFS method was developed in C# using the Weka library [33]. All our experiments were conducted under consistent computational settings on a standard computer machine (Intel® Core™ i7, 1.90 GHz, 8.00 GB RAM, manufactured by Dell Inc., Round Rock, Texas, United States). In our experiments, the hyperparameters for the techniques employed—SMOTE, ordinal classification, feature selection, and the random tree classifier—were systematically tested across different ranges at each stage to determine their optimal values for the given dataset. These ranges were carefully selected based on prior research and experimental validation, ensuring that the final configurations provided the best performance. The details of the hyperparameter tuning process and the selected values are described below:

- SMOTE was used to address the class imbalance inherent in the road traffic accidents dataset, particularly for the "serious injury" and "fatal injury" classes. SMOTE generates new samples according to the neighborhood strategy, where the key parameter "nearestNeighbors" determines the number of neighbors considered when creating synthetic examples. To determine the optimal parameter value, we systematically tested a range of "nearestNeighbors" values (k = 5, 6, 7, 8, 9, 10). The results demonstrated that k = 5 achieved the highest accuracy for the ORT-ROFS method, shown in Table 3. Therefore, in our experiments, this parameter was set to 5. The other key parameters of SMOTE include "ClassValue", which specifies the target class for oversampling, and "Percentage", which defines the percentage increase in instances for that class. Here, SMOTE was applied twice: the first time with "ClassValue" set to 2 and "Percentage" set to 400 to augment the "serious injury" class, and the second time with "ClassValue" set to 3 and "Percentage" set to 200 to augment the "fatal injury" class. These configurations were selected to balance the dataset while maintaining meaningful relationships between the classes.

Table 3. Accuracies of the ORT-ROFS method across different "nearestNeighbors" (k) values.

	Accuracy (%)
k	**ORT-ROFS**
5	87.19
6	86.95
7	86.90
8	86.53
9	86.10
10	85.87

- For the process of ordinal classification, the "batchSize" was set to 100, which is the default value. We tested various values for the "batchSize" parameter and observed no significant changes in the results. The classifier was configured as "Random Tree", leveraging its ability to model nonlinear relationships in the data competently. Here, the binary decomposition was employed to transform the ordinal problem into multiple binary classification problems. This decomposition allows the random tree classifier to handle ordinal relationships effectively.

- ROFS was employed to select the most relevant features while preserving the ordinal relationships in the dataset. In our method, the attribute evaluator was set to "CorrelationAttributeEval" to measure the correlation between each feature and the target class, and the search method was configured as "Ranker" to rank and select attributes based on their correlation scores. The evaluator calculates the Pearson correlation coefficient for numeric attributes, while nominal attributes are evaluated by treating each value as an indicator variable and computing an overall correlation via a weighted average. Under these settings, key features including Hour, Day_of_week, Age_band_of_driver, Types_of_junction, Light_conditions, Weather_conditions, Number_of_vehicles_involved, and Number_of_casualties were chosen as the most influential in predicting traffic accident severity. These selected features were then used for classification, leading to effective prediction results.

- Random tree classifier served as the base classifier for the proposed ORT-ROFS method due to its capability to model complex relationships. The "maxDepth", determining the maximum depth of the tree, was tested for values ranging from 1 to 20, as illustrated in Figure 2. Setting "maxDepth" to "NaN" allows the tree to grow without restrictions, effectively making it limitless. The results showed that accuracy steadily increased and reached 87.08% at 20. Notably, the unrestricted setting of "maxDepth" achieved an accuracy of 87.19%, which was slightly higher than the best accuracy observed within the tested range. This indicates that while deeper trees generally lead to better accuracy, allowing the tree to grow without restrictions can still yield competitive results, demonstrating the flexibility of the model.

The "minNum", representing the minimum number of instances per leaf, was tested for values ranging from 1 to 100 in increments of 5, as represented in Figure 3. Initially set to 1, this configuration allowed every leaf node to have at least one instance, achieving the highest accuracy of 87.19%. However, as the "minNum" value increased, a steady decline in accuracy was observed. For example, accuracy dropped to 85.37% at 5, 83.74% at 10, and continued decreasing, reaching 73.90% at 100. This trend indicates that increasing the minimum number of instances per leaf restricts the tree's ability to capture finer-grained patterns, leading to reduced predictive performance. Therefore, a lower "minNum" value proves crucial for achieving optimal accuracy in this context.
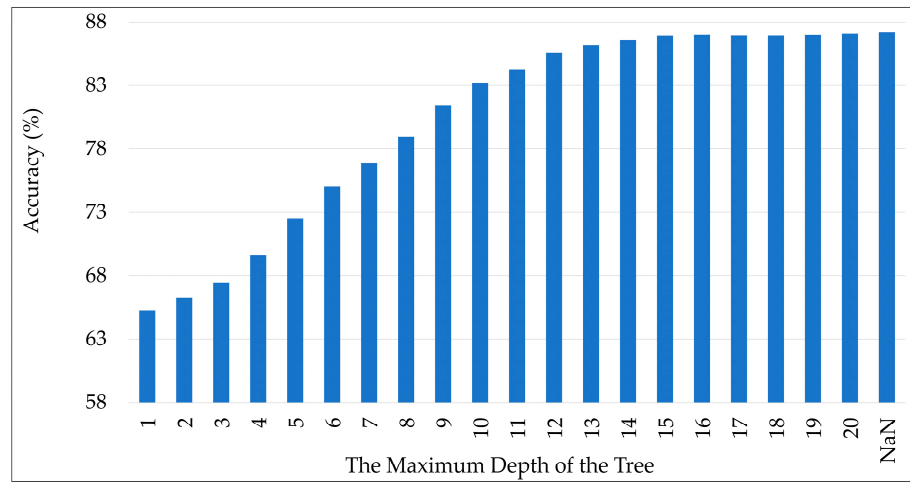
Figure 2. Accuracies of the ORT-ROFS method across different "maxDepth" values.
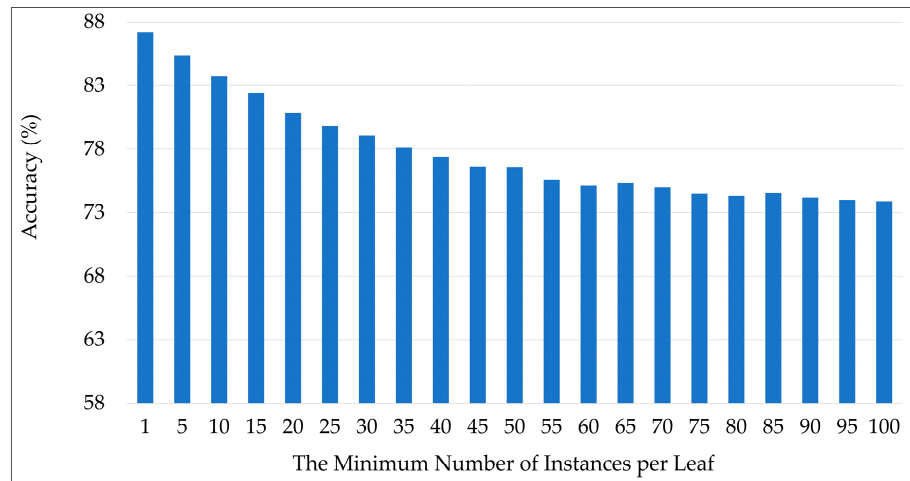

Figure 3. Accuracies of the ORT-ROFS method across different "minNum" values.

Additionally, the "minVarianceProp" hyperparameter, defining the minimum variance proportion required to split a node, was tested for values ranging from 0.0001 to 0.3. No changes in accuracy were observed across this range, with the default value of 0.001 providing consistent results. The "KValue", which specifies the number of random attributes to consider for splitting at each node, was tested with values of 0, 1, 3, 5, 7, and 9. Among these, the best accuracy was achieved when "KValue" was set to 0, meaning $int(\log_2(\#predictors) + 1)$ attributes were considered. Other parameters include "batchSize", which was set to 100, and seed for reproducibility, set to 1. These parameter values were chosen to mathematically balance tree complexity and computational competency while maintaining model accuracy.

The performance of four methods—random tree (RT), ordinal random tree (ORT), ordinal random tree with feature selection (ORT-FS), and ordinal random tree with rank-oriented feature selection (ORT-ROFS)—was evaluated and compared in terms of accuracy. These methods were tested on the same dataset under identical conditions, ensuring consistent evaluation. The accuracy results for these methods are summarized in Table 4, where the ORT-ROFS method demonstrated an advancement, achieving an average improvement of 4.58% in accuracy.

Table 4. Comparison of methods based on the accuracy performance metric.

Dataset	Accuracy (%)			
	RT	ORT	ORT-FS	ORT-ROFS (Proposed)
Road Traffic Accident (RTA)	84.09	84.65	81.44	87.19

In addition to the accuracy measure, the performance of ORT-ROFS was validated using a range of evaluation metrics, confirming a complete assessment of its predictive capability. As shown in Figure 4, ORT-ROFS achieved improvements of 4.71%, 4.58%, and 4.78% on average in precision, recall, and F-measure, respectively. These results highlight the effectiveness of the proposed method in handling ordinal classification tasks while utilizing its rank-oriented feature selection mechanism.
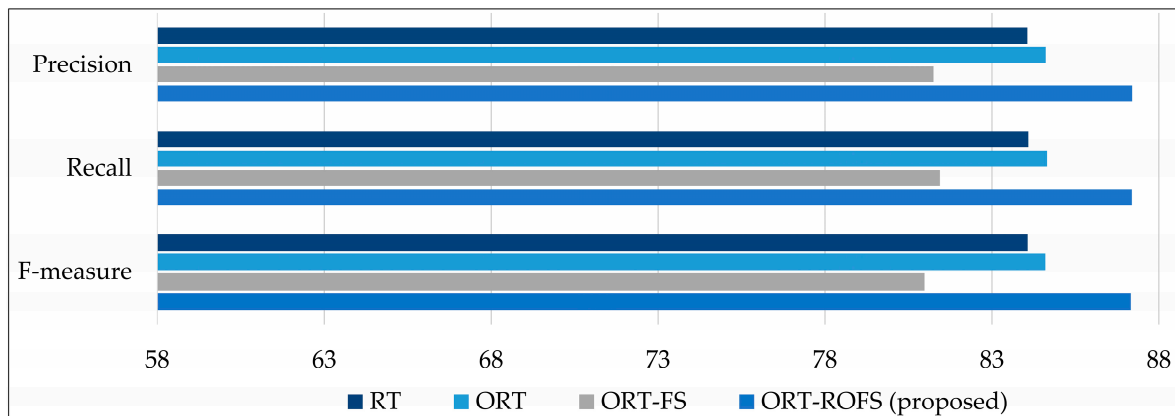

Figure 4. Comparison of methods based on precision, recall, and F-measure performance metrics.

The results presented in Table 4 and Figure 4 provide a comprehensive comparison of the four methods. Additionally, we detail the performance of each method as follows:

- Random tree (RT) is the baseline method, showing an accuracy of 84.09%. This method does not apply any specific handling of ordinal relationships or feature selection, offering a solid but unspecialized approach to the dataset.
- Ordinal random tree (ORT), which incorporates handling for ordinal data through binary decomposition of the target classes, achieved a slight improvement in accuracy of 84.65% over RT. By explicitly considering the ordered nature of the target variable, ORT demonstrates that using the class ordinality can enhance classification performance, even without feature selection.
- Ordinal random tree with feature selection (ORT-FS), which applies Pearson correlation for feature selection on multi-class targets, achieved an accuracy of 81.44%. However, despite the application of ordinal classification and feature selection, the approach may not fully capture the nuances of ordinal relationships as effectively as ORT and ORT-ROFS, resulting in lower performance.
- Ordinal random tree with rank-oriented feature selection (ORT-ROFS) mathematically expressed the best performance across all metrics, with an accuracy of 87.19%. This method, which applies the ROFS technique on binary class targets derived from the dataset, yielded the highest precision of 87.20, recall of 87.19, and F-measure of 87.16. It showcased a superior ability to handle the ordinal nature of the data while also selecting the most relevant features, leading to a noticeable improvement in classification accuracy.

Figure 5 illustrates the confusion matrix for the ORT-ROFS classification across the road traffic accident dataset with class labels, including slight injury, serious injury, and fatal injury. The diagonal elements indicate the number of instances accurately classified for each injury category. For example, 9160 instances of slight injury were correctly classified from the total of 10,415 slight injury instances, 7623 instances of serious injury were correctly identified from the total of 8715 serious injury instances, and 310 instances of fatal injury were accurately recognized from the total of 474 fatal injury instances. Off-diagonal elements display misclassifications: e.g., 15 instances of serious injury were misclassified as fatal injury. The confusion matrix offers a complete overview of the model's performance across different classes, highlighting both correct classifications and misclassifications. The background colors in the confusion matrix represent the distribution of values, where darker shades indicate higher numbers of correct instances, and lighter shades represent lower numbers of correct ones for each row. This color gradient visually emphasizes the classification results, making it easier to distinguish between correct classifications and misclassifications. Here, it provides valuable evaluations of the model's ability to predict the severity level of injuries in road traffic accidents.

Figure 5. Confusion matrix for the ORT-ROFS method over the road traffic accident severities.

5. Discussion

In this section, we compared the performance of the proposed ORT-ROFS method with state-of-the-art methods reported in previous studies [34–40] for the same road traffic accidents (RTA) dataset. Table 5 provides a detailed comparison, including accuracy values and split ratios within these studies. In the split ratio column, "NA" indicates that the split strategy was not available in the related reference paper. The results show that ORT-ROFS consistently outperformed competing methods across different split strategies for training and testing sets, achieving accuracies of 86.69% (80:20 split), 86.88% (70:30 split), and 87.19% (10-fold cross-validation). These values are higher than the accuracies obtained by state-of-the-art methods under each corresponding split ratio, highlighting its robustness and adaptability to varying data partitions. To clarify the methodology, we calculated our method's improvement relative to the state-of-the-art methods within each group of split strategies. On average, ORT-ROFS demonstrated an improvement of 10.81% over the previous studies in terms of accuracy. The proposed method mathematically excelled over decision-tree-based methods [39], such as the random tree (84.51%), the J48 pruned tree (85.47%), and the logistic model tree (LMT) (84.74%). Additionally, ORT-ROFS outperformed ensemble-learning-based methods, including extra trees (81.35%) [40], random forest (RF) (84.49%) [35], adaptive boost classifier (85.30%) [38], light gradient boosting machine (LGBM) (84%) [34], and gradient boosting (GB) (77.75%) [40].

These results underscore the superiority of ORT-ROFS in handling complex ordinal data. As seen in the results, standard classification methods like k-nearest neighbors are less effective in capturing the order in the class labels reflecting the crash severity (slight < serious < fatal injury). In summary, the consistently superior performance of ORT-ROFS over other competitive methods underscored its capability of classifying ordinal road traffic accident severity data.

Table 5. The comparison of the performance of the proposed method with already reported outputs of the state-of-the-art methods on the same dataset in terms of accuracy (%).

Reference	Year	Method	Split Ratio	Accuracy (%)
Xiao and Duan [34]	2024	Light Gradient Boosting Machine + SMOTE	NA	84.00
Obaid [35]	2024	Random Forest Decision Tree	80:20	84.49 83.06
Ramya and Eswari [36]	2024	Logistic Regression Extreme Gradient Boosting Decision Tree Random Forest	NA	87.00 86.00 74.00 84.00
Endalie and Abebe [37]	2023	Support Vector Machines	80:20	85.00
Kodepogu et al. [38]	2023	Decision Tree Random Forest K-Nearest Neighbors Naive Bayes Adaptive Boost Classifier	80:20	83.30 77.40 82.20 85.30 85.30
Adeliyi et al. [39]	2023	J48 Pruned Tree Naive Bayes Bagging K-Nearest Neighbors Logistic Model Tree Decision Tree Random Tree Logistic Regression	10-fold-cross-validation	85.47 83.53 84.29 77.58 84.74 84.52 84.51 84.51
Alhosani [40]	2022	Gradient Boosting Random Forest Logistic Regression Decision Tree Support Vector Classifier Extra Trees	70:30	77.75 79.78 69.47 53.10 56.67 81.35
Proposed Approach		ORT-ROFS	80:20 70:30 10-fold-cross-validation	86.69 86.88 87.19

6. Conclusions and Future Works

Traffic accident severity prediction is a crucial task in attaining better transportation safety and management, as it informs responsible authorities and the public about ways to mitigate adverse effects. For this purpose, our paper presents the ordinal random tree with rank-oriented feature selection (ORT-ROFS) method, which emphasizes the ordered progression of severity levels and integrates innovative rank-oriented feature selection. Leveraging the explainability of tree-based structures, the proposed method not only enhances prediction accuracy but also provides valuable mathematical standpoints into the contributing factors behind accident severity. By utilizing the ROFS technique, the model identifies key features that expressively impact accident severity prediction. These features include the time of the incident, day of the week, driver's age group, type of junction, light and weather conditions, the number of vehicles involved, and the number of casualties.

When evaluated on a real-world dataset of road traffic accidents in Addis Ababa, Ethiopia (2017–2020), ORT-ROFS achieved substantial advances with an accuracy of 87.19%. It demonstrated improvements of 4.58%, 4.71%, 4.58%, and 4.78% in accuracy, precision, recall, and F-measure, respectively, when compared to its counterparts—random tree (RT), ordinal random tree (ORT), and ordinal random tree with feature selection (ORT-FS). Furthermore, the method surpassed state-of-the-art techniques, achieving a 10.81% improvement in predictive accuracy on average. Traffic accident severity prediction is a critical task to enable the efficiency of proactive traffic crash management. These results

underscore the potential of ORT-ROFS to advance road safety strategies by enabling more informed policy-making.

While the ORT-ROFS method has shown substantial promise, several future directions can be explored. The integration of technological tools, such as web or mobile applications enhanced with cloud computing, could further extend the applicability of the ORT-ROFS method for both transportation authorities and individual drivers. Future research could focus on the social implications of key features identified by the rank-oriented feature selection mechanism, enabling targeted actions such as public awareness campaigns or interventions to mitigate accident severity for specific regions or demographics. Additionally, it can be applied to an expanded dataset that includes diverse regions to adapt it to a wide range of environments and scenarios.

Author Contributions: Conceptualization, B.G. and K.F.B.; methodology, B.G. and K.F.B.; software, B.G. and D.B.; validation, K.F.B.; formal analysis, B.G.; investigation, K.F.B.; resources, K.U.B. and D.B.; data curation, K.U.B.; writing—original draft preparation, B.G. and K.F.B.; writing—review and editing, K.U.B. and D.B.; visualization, B.G.; supervision, D.B.; project administration, K.U.B.; funding acquisition, K.F.B. and K.U.B. All authors have read and agreed to the published version of the manuscript.

Data Availability Statement: The road traffic accidents (RTA) in Addis Ababa, Ethiopia (2017–2020) dataset is publicly available through the Mendeley Data [31] (https://data.mendeley.com/datasets/xytv86278f/1, accessed on 22 November 2024) and Kaggle [32] (https://www.kaggle.com/datasets/saurabhshahane/road-traffic-accidents, accessed on 22 November 2024) repositories.

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

AdaBoost	Adaptive boosting
ANN	Artificial neural network
AUC	Area under curve
BD	Binary decomposition
CART	Classification and regression trees
CatBoost	Categorical boosting
CNN	Convolutional neural network
DT	Decision tree
FCM	Fuzzy c-means
FPR	False positive rate
FS	Feature selection
GB	Gradient boosting
GPC	Gaussian process classifier
KNN	K-nearest neighbors
LGBM	Light gradient boosting machine
LR	Logistic regression
MBE	Mean bias error
ML	Machine learning
MPE	Mean percentage error
MSE	Mean square error
NB	Naive Bayes
OC	Ordinal classification
ORT	Ordinal random tree
ORT-FS	Ordinal random tree with feature selection
ORT-ROFS	Ordinal random tree with rank-oriented feature selection

RBF	Radial basis function
RF	Random forest
RMSE	Root mean square error
ROC	Receiver operating characteristic
ROFS	Rank-oriented feature selection
RT	Random tree
RTA	Road traffic accidents
SMOTE	Synthetic minority oversampling technique
SVM	Support vector machines
TNR	True negative rate
TPR	True positive rate
XGB	Extreme gradient boosting

References

1. Ali, Y.; Hussain, F.; Haque, M.M. Advances, Challenges, and Future Research Needs in Machine Learning-Based Crash Prediction Models: A Systematic Review. *Accid. Anal. Prev.* **2024**, *194*, 107378. [CrossRef] [PubMed]

2. Hee, L.V.; Khamis, N.; Noor, R.M.; Abdul Karim, S.A.; Puspitasari, P. Predicting Fatality in Road Traffic Accidents: A Review on Techniques and Influential Factors. In *Intelligent Systems Modeling and Simulation III*; Abdul Karim, S.A., Ed.; Studies in Systems, Decision and Control; Springer: Cham, Switzerland, 2024; Volume 553. [CrossRef]

3. Chai, A.B.Z.; Lau, B.T.; Tee, M.K.T.; McCarthy, C. Enhancing Road Safety with Machine Learning: Current Advances and Future Directions in Accident Prediction Using Non-Visual Data. *Eng. Appl. Artif. Intell.* **2024**, *137*, 109086. [CrossRef]

4. Wen, X.; Xie, Y.; Jiang, L.; Pu, Z.; Ge, T. Applications of Machine Learning Methods in Traffic Crash Severity Modelling: Current Status and Future Directions. *Transp. Rev.* **2021**, *41*, 855–879. [CrossRef]

5. Wang, J.; Zhao, C.; Liu, Z. Can Historical Accident Data Improve Sustainable Urban Traffic Safety? A Predictive Modeling Study. *Sustainability* **2024**, *16*, 9642. [CrossRef]

6. Qi, Z.; Yao, J.; Zou, X.; Pu, K.; Qin, W.; Li, W. Investigating Factors Influencing Crash Severity on Mountainous Two-Lane Roads: Machine Learning Versus Statistical Models. *Sustainability* **2024**, *16*, 7903. [CrossRef]

7. Pourroostaei Ardakani, S.; Liang, X.; Mengistu, K.T.; So, R.S.; Wei, X.; He, B.; Cheshmehzangi, A. Road Car Accident Prediction Using a Machine-Learning-Enabled Data Analysis. *Sustainability* **2023**, *15*, 5939. [CrossRef]

8. Frank, E.; Hall, M. A Simple Approach to Ordinal Classification. In *Machine Learning: ECML 2001*; De Raedt, L., Flach, P., Eds.; Lecture Notes in Computer Science; Springer: Berlin/Heidelberg, Germany, 2001; Volume 2167. [CrossRef]

9. Fürnkranz, J.; Hüllermeier, E.; Vanderlooy, S. Binary Decomposition Methods for Multipartite Ranking. In *Machine Learning and Knowledge Discovery in Databases*; Buntine, W., Grobelnik, M., Mladenić, D., Shawe-Taylor, J., Eds.; Lecture Notes in Computer Science; Springer: Berlin/Heidelberg, Germany, 2009; Volume 5781. [CrossRef]

10. Chawla, N.V.; Bowyer, K.W.; Hall, L.O.; Kegelmeyer, W.P. SMOTE: Synthetic minority over-sampling technique. *J. Artif. Intell. Res.* **2002**, *16*, 321–357. [CrossRef]

11. Chandrashekar, G.; Sahin, F. A survey on feature selection methods. *Comput. Electr. Eng.* **2018**, *66*, 31–47. [CrossRef]

12. Dhal, P.; Azad, C. A comprehensive survey on feature selection in the various fields of machine learning. *Appl. Intell.* **2022**, *52*, 4543–4581. [CrossRef]

13. Frank, E.; Kirkby, R. Random Tree. Available online: http://weka.sourceforge.net/doc.dev/weka/classifiers/trees/RandomTree.html (accessed on 22 November 2024).

14. Arciniegas-Ayala, C.; Marcillo, P.; Valdivieso Caraguay, Á.L.; Hernández-Álvarez, M. Prediction of Accident Risk Levels in Traffic Accidents Using Deep Learning and Radial Basis Function Neural Networks Applied to a Dataset with Information on Driving Events. *Appl. Sci.* **2024**, *14*, 6248. [CrossRef]

15. Muktar, B.; Fono, V. Toward Safer Roads: Predicting the Severity of Traffic Accidents in Montreal Using Machine Learning. *Electronics* **2024**, *13*, 3036. [CrossRef]

16. Obasi, I.; Benson, C. Evaluating the effectiveness of machine learning techniques in forecasting the severity of traffic accidents. *Heliyon* **2023**, *9*, e18812. [CrossRef] [PubMed]

17. Gatarić, D.; Ruškić, N.; Aleksić, B.; Đurić, T.; Pezo, L.; Lončar, B.; Pezo, M. Predicting Road Traffic Accidents—Artificial Neural Network Approach. *Algorithms* **2023**, *16*, 257. [CrossRef]

18. Aldhari, I.; Almoshaogeh, M.; Jamal, A.; Alharbi, F.; Alinizzi, M.; Haider, H. Severity Prediction of Highway Crashes in Saudi Arabia Using Machine Learning Techniques. *Appl. Sci.* **2023**, *13*, 233. [CrossRef]

19. Yan, M.; Shen, Y. Traffic Accident Severity Prediction Based on Random Forest. *Sustainability* **2022**, *14*, 1729. [CrossRef]

20. Islam, M.K.; Reza, I.; Gazder, U.; Akter, R.; Arifuzzaman, M.; Rahman, M.M. Predicting Road Crash Severity Using Classifier Models and Crash Hotspots. *Appl. Sci.* **2022**, *12*, 11354. [CrossRef]

21. Khattak, A.; Almujibah, H.; Elamary, A.; Matara, C.M. Interpretable Dynamic Ensemble Selection Approach for the Prediction of Road Traffic Injury Severity: A Case Study of Pakistan's National Highway N-5. *Sustainability* **2022**, *14*, 12340. [CrossRef]

22. Dong, S.; Khattak, A.; Ullah, I.; Zhou, J.; Hussain, A. Predicting and Analyzing Road Traffic Injury Severity Using Boosting-Based Ensemble Learning Models with SHAPley Additive exPlanations. *Int. J. Environ. Res. Public Health* **2022**, *19*, 2925. [CrossRef]

23. Santos, D.; Saias, J.; Quaresma, P.; Nogueira, V.B. Machine Learning Approaches to Traffic Accident Analysis and Hotspot Prediction. *Computers* **2021**, *10*, 157. [CrossRef]

24. Boo, Y.; Choi, Y. Comparison of Prediction Models for Mortality Related to Injuries from Road Traffic Accidents after Correcting for Undersampling. *Int. J. Environ. Res. Public Health* **2021**, *18*, 5604. [CrossRef]

25. Fiorentini, N.; Losa, M. Handling Imbalanced Data in Road Crash Severity Prediction by Machine Learning Algorithms. *Infrastructures* **2020**, *5*, 61. [CrossRef]

26. Assi, K.; Rahman, S.M.; Mansoor, U.; Ratrout, N. Predicting Crash Injury Severity with Machine Learning Algorithm Synergized with Clustering Technique: A Promising Protocol. *Int. J. Environ. Res. Public Health* **2020**, *17*, 5497. [CrossRef] [PubMed]

27. Lee, J.; Yoon, T.; Kwon, S.; Lee, J. Model Evaluation for Forecasting Traffic Accident Severity in Rainy Seasons Using Machine Learning Algorithms: Seoul City Study. *Appl. Sci.* **2020**, *10*, 129. [CrossRef]

28. Assi, K. Traffic Crash Severity Prediction—A Synergy by Hybrid Principal Component Analysis and Machine Learning Models. *Int. J. Environ. Res. Public Health* **2020**, *17*, 7598. [CrossRef]

29. Benesty, J.; Chen, J.; Huang, Y.; Cohen, I. Pearson Correlation Coefficient. In *Noise Reduction in Speech Processing*; Springer: Berlin/Heidelberg, Germany, 2009; pp. 1–4. [CrossRef]

30. Dagum, C. Decomposition and interpretation of Gini and the generalized entropy inequality measures. *Statistica* **1997**, *57*, 295–308.

31. Bedane, T.T. Road Traffic Accident Dataset of Addis Ababa City; Mendeley Data. Available online: https://data.mendeley.com/datasets/xytv86278f/1 (accessed on 22 November 2024).

32. Shahane, S. Road Traffic Accident Dataset of Addis Ababa City; Kaggle. Available online: https://www.kaggle.com/datasets/saurabhshahane/road-traffic-accidents (accessed on 22 November 2024).

33. Witten, I.H.; Frank, E.; Hall, M.A.; Pal, C.J. *Data Mining: Practical Machine Learning Tools and Techniques*, 4th ed.; Morgan Kaufmann: Cambridge, MA, USA, 2016; pp. 1–664.

34. Xiao, Y.; Duan, Z. Improving crash injury severity prediction via ensemble models and sampling techniques for addressing data imbalance. *SSRN-Soc. Sci. Res. Netw.* **2024**, 4730366. [CrossRef]

35. Obaid, M.A.A. Driver's Accident Behavioral Analytics Using AI. Master's Thesis, Rochester Institute of Technology, Dubai, United Arab Emirates, 2024. Available online: https://repository.rit.edu/theses/11748/ (accessed on 1 January 2025).

36. Ramya, A.; Eswari, M.S. Road accident severity in India: A machine learning approach. *Int. J. Progress. Res. Eng. Manag. Sci.* **2024**, *4*, 372–375.

37. Endalie, D.; Abebe, W.T. Analysis and detection of road traffic accident severity via data mining techniques: Case study Addis Ababa, Ethiopia. *Math. Probl. Eng.* **2023**, *2023*, 6536768. [CrossRef]

38. Kodepogu, K.; Manjeti, V.B.; Siriki, A.B. Machine learning for road accident severity prediction. *Mechatron. Intell. Transp. Syst.* **2023**, *2*, 211–226. [CrossRef]

39. Adeliyi, T.T.; Oluwadele, D.; Igwe, K.; Aroba, O.J. Analysis of road traffic accidents severity using a pruned tree-based model. *Int. J. Transp. Dev. Integr.* **2023**, *7*, 131–138. [CrossRef]

40. Alhosani, M. Traffic Accidents Analysis & Prediction in UAE. Master's Thesis, Rochester Institute of Technology, Dubai, United Arab Emirates, 2022. Available online: https://repository.rit.edu/theses/11380/ (accessed on 1 January 2025).