**PREDICTING ROAD CRASHES AND FATALITY TRENDS IN**

**ILOCOS SUR: A RANDOM FOREST MODEL**

**A Requirements Specification Document**

**Presented to**

**The Faculty of the College of Information**

**And Computing Sciences**

**Mindanao State University**

**In Partial Fulfillment**

**Of the Requirement for the course**

**ISY115**

**By:**

**Jossana Marie J. Busquit**

**Pearly Jane C. Cruz**

**Mark Benedict Acap**

**Instructor**

**Ma’am Nornina Dia**

**December 12, 2025**

**Abstract**

Road crashes remain a major public safety concern in the Philippines, resulting in injuries, fatalities, and significant economic losses. Localized studies are essential for understanding community-specific crash patterns and developing targeted safety interventions. This study presents a comprehensive analysis of road accident incidents in Magsingal, Ilocos Sur, spanning from 2019 to 2025. The dataset comprises 185 recorded cases, capturing critical information such as exact location, date and time, road user type (pedestrians, non-motorized vehicles, and motorized vehicles), collision type (head-on, rear-end, side-swipe, hit-and-run, and self-accidents), causation factors, and the resulting damages or injuries. Through data preprocessing, exploratory analysis, and descriptive and predictive modeling, this research identifies patterns and trends in road accidents, highlighting the primary factors contributing to crashes. Findings indicate that human error, drunk driving, overspeeding, lane encroachment, loss of control, and sudden environmental conditions are the most prevalent causes, with pedestrians and motorcycle users being particularly vulnerable. Using a Random Forest model, the study achieves strong predictive performance in classifying crash severity and identifying high-risk conditions, with time of day, collision type, and causation factors emerging as the most important severity predictors.

The results emphasize the value of integrating machine learning and locally sourced accident records into municipal-level traffic monitoring systems. Insights from this study can inform public awareness campaigns, guide stricter traffic law enforcement, support infrastructure improvements, and enable proactive road safety planning, ultimately reducing accidents and enhancing the safety of all road users in Magsingal.

To discuss this pressing issue, this research forecasts the frequency and severity of road incidents in the Philippine context by evaluating the effectiveness of environmental and road-condition variables through modern ML methods. Previous research has demonstrated that machine learning approaches are effective for predicting accident likelihood and severity using environmental and infrastructural variables (Abdel-Aty & Haleem, 2011; Xu et al., 2017; Pour et al., 2020; Chang et al., 2019). In this study, the best-performing model achieved an accuracy of 84.7%, indicating the strong potential of data-driven approaches in identifying hazardous road conditions. The findings highlight the importance of integrating environmental, temporal, and road condition data into predictive analytics systems, enabling policymakers and traffic authorities to implement proactive safety interventions, prioritize infrastructure improvements, and reduce road-related injuries and fatalities in the Philippines.



**Keywords:** Road Accident Prediction, Road Condition, Machine Learning, Traffic, Random Forest Model.

1. **Introduction**

According to the Asian Transport Observatory (2025), road accidents represent one of the most pressing public safety challenges globally, resulting in millions of injuries, fatalities, and significant economic losses annually. Worldwide, over 1.3 million people die each year from road traffic crashes, with an estimated 20–50 million sustaining non-fatal injuries. Vulnerable road users, such as pedestrians, cyclists, and motorcyclists, are particularly affected (World Health Organization [WHO], 2023). In the Philippines, road traffic incidents have escalated into a major concern. National data show a rise in traffic fatalities from 7,938 in 2011 to 11,096 in 2021, with most victims being male and young adults aged 15–29 disproportionately impacted (WHO Philippines, 2023). These injuries also impose a substantial economic burden, costing approximately 2.6% of the country's gross domestic product each year (WHO Philippines, 2023).

Key contributors to this high crash incidence include human errors like speeding, driving under the influence, and failure to use safety measures such as seat belts and helmets. Additional exacerbating factors are infrastructure deficiencies, weak traffic law enforcement, and fragmented data collection (BusinessMirror, 2024). While regulations like the Land Transportation and Traffic Code (Republic Act No. 4136) aim to promote road safety, inconsistent enforcement undermines their impact. Effective road safety demands a transition from reactive measures to proactive, evidence-based approaches. In collaboration with the WHO, the Department of Transportation (DOTr) introduced the Philippine Road Safety Action Plan 2023–2028, targeting at least a 35% reduction in road traffic deaths by 2028 through improvements in roads, vehicles, user behavior, and data systems (WHO Philippines, 2023).

Despite these efforts, localized crash pattern analysis remains limited. By integrating predictive analytics and machine learning techniques, such as Random Forest classifiers, it is possible to identify critical risk factors, forecast crash severity, and enable data-driven safety planning. This study bridges that gap by examining road accidents in Ilocos Sur from 2019 to 2025, classifying crash severity, and pinpointing high-risk conditions to guide municipal-level, evidence-based interventions.

**1.1 Background of the Study**

The Philippines, a rapidly developing country in Southeast Asia, faces significant and ongoing challenges in public road safety. The combined pressures of increasing population and infrastructure development have intensified traffic congestion and created complex traffic patterns, making road users more vulnerable to accidents. Road traffic accidents remain a leading cause of injury and death, with human factors such as reckless driving, speeding, and non-compliance with rules being highly influential. Furthermore, infrastructural deficits, including poorly maintained roads, inadequate signage, and exposure to extreme weather, exacerbate the risk, particularly in provincial areas.

Furthermore, road accidents are one of the most concerning public safety issues globally. According to Norona, M. et al.,(2021, April). The total number of vehicular accidents in the Philippines has been on a rising trend, doubling from 63,072 incidents in 2007 to 116,906 in 2018. As of 2018, Road Traffic Accident Deaths reached 10,624, accounting for 1.74% of total deaths in the country.The road traffic accidents have been a burning national concern regarding the population welfare problem in developing nations, such as the Philippines, where swift urbanization, rising considerably in vehicle ownership, and heavy traffic congestion are choking an already poor road network. The surfaces of many roads particularly in urban and semi urban settings are irregular, lack adequate safety measures and pedestrian and cyclist facilities. Additionally, there are factors that cause road accidents (crash) -- Human factor, Mechanical Defect, Road Condition, and etc. (Scribd, 2024).

Moreover, the Municipality of Magsingal in Ilocos Sur serves as a vital case study, as it shares these national challenges at the local level. This study utilizes a localized dataset of 185 recorded road accident incidents, capturing specific details on collision type (Head On, Rear End, Hit & Run, etc.) and causation factors (Drunk driving, Human error, Lost control, etc.). Initial descriptive analysis of this data shows that human error, drunk driving, loss of control, and overspeeding are among the most prevalent causes of crashes in the area, with motorcycle users and pedestrians identified as particularly vulnerable groups.

Recognizing that many road safety efforts are still reactive, this research is driven by the need to utilize the power of predictive analytics to move toward a preventive paradigm. By employing a Random Forest model, this research provides a comprehensive method to not only classify accident severity but also to objectively rank the key predictors. The findings will underscore which specific factors such as collision type and causation emerge as the most statistically important severity predictors, enabling local authorities in Magsingal to implement targeted interventions, guide stricter traffic law enforcement, and prioritize infrastructure improvements based on data, thereby enhancing the overall safety of road users.

Additional risks of accidents include the environmental issues, including having heavy rainfall, floods, and poor visibility in the cases of typhoons which are most pronounced in the rural and provincial roads (PIDS, 2025; Asian Transport Observatory, 2025).

In addition, this challenge is compounded by the absence of a systematic government strategic plan for road safety, which is gradually implemented. Along with other factors, accidents and complaints have often been the reasons for the government to intervene in terms of road construction or repair, rather than being guided by a systematic risk assessment or predictive analytics (PW 1489, Philippine Road Safety Data 2018). Hence, dangerous conditions on roads still exist, and preventive actions are scarce. The human and economic costs are significant. About 12,000 Filipinos die each year in road-related incidents. In 2024, the Highway Patrol Group reported 2,747 deaths and more than 31,000 road crash incidents, which is a 35% increase from the previous year (Philstar, 2025). Only 5% of roads meet safe pedestrian standards. Additionally, 14% meet safe bicyclist standards, showing serious gaps in infrastructure (Asian Transport Observatory, 2025).

By with this study demonstrate the potential by using Machine Learning (ML) for predicting the road accidents by pulling together things like weather, road conditions, and timing (Pourroostaei Ardakani et al., 2023; Aleksić et al., 2023; Rifat et al., 2024; Liu, 2025). Moreover, the data collection process for road crashes, relying primarily on manually recorded police blotters, is difficult to conduct meaningful analysis. This study aims to contribute to more proactive and evidence-based traffic management by demonstrating the potential of Machine Learning (ML) for predicting road accidents.

1. **Review of Related Literature**

This study provides a review of related literature, past studies as a foundation for road accident prediction, and Predicting Road Accidents and Analyzing Road Conditions in the Philippines Using Machine Learning: An Environmental Analysis. This study concept for local and foreign studies contributes to road safety planning and accident prevention measures in the Philippines.

**2.1 Review of Related Studies**

**2.1.1 Global Perspective on Road Accidents**

Road traffic accidents continue to be one of the most serious global public health problems, taking millions of lives and hurting communities in every region. According to the World Health Organization, over 1.25 million people die each year, with nearly 3,400 deaths occurring daily and up to 50 million people suffering non-fatal injuries (Wegman, 2017). These injuries frequently result in long-term disability, significantly altering the lives of victims and their families. However, the impact of road crashes is unevenly spread, with about 90% of all traffic-related deaths occurring in low- and middle-income countries (LMIC), where road systems, enforcement, and emergency care are often underdeveloped. Mortality rates vary greatly between nations, ranging from fewer than 3 to over 40 fatalities per 100,000 people, averaging less than 9 in high-income countries (HIC) and around 20 in low- and middle-income countries (LMIC), with Africa having the highest rate at 26.6 deaths per 100,000 (Wegman, 2017).

More recent reports have echoed this issue. According to Ahmed et al. (2023), 1.35 million people are killed on the road each year, with 20 to 50 million injured making road accidents the eighth greatest cause of mortality globally, and perhaps seventh by 2030 if current trends continue. OECD data also reveal conflicting results. While most member nations have reduced road fatalities over the last decade, the United States has seen a modest increase, despite a 13% decrease in total miles driven (Ahmed et al., 2023). This difference highlights that even high-income countries are not immune to road safety issues.

Overall, global patterns show that road traffic injuries are a neglected but critical health issue that requires concerted action, context-sensitive solutions, and long-term commitment from both national governments and the international community. HIC strategies cannot be simply replicated in LMIC contexts, but the guiding principles safe systems, proactive preventive, and evidence-based policies provide a solid foundation for countries looking to improve their road safety regimes.

**2.1.2 Local Studies on Road Accidents in the Philippines**

Road accidents in the Philippines are the result of a complex combination of economic, topographical, and infrastructure difficulties. Because of the country's archipelagic structure, main economic activities and population centers are concentrated on the islands of Luzon, Visayas, and Mindanao, resulting in high vehicle traffic on primary national roadways (Villoria et al., 2000; Velasco et al., 2021). As Metro Manila struggles with significant congestion and limited area for expansion, industry and traffic have shifted to neighboring regions, increasing mobility demands. Despite the introduction of national measures like as speed limit limits, helmet and seatbelt rules, and anti-drunk and drug-driving laws, road crashes continue to occur.

In addition, regional evaluations in the Philippines, such as Velasco et al. (2021), show that fast urbanization, shifting economic zones, and increasing traffic density all contribute to increased accident risk, particularly on overburdened roads. These findings highlight the necessity of considering both human behavior and physical road conditions for estimating accident risk.

**2.1.3 Road Traffic Accident Correlates in Barangay Sampaloc**

According to (Dela Cruz, J. M., & Reyes, A. P. 2020) local study titled “Correlation of Factors Affecting Road Traffic Accidents in MacArthur Highway, Barangay Sampaloc, Apalit, and Pampanga using Multinomial Logit Model” examined the relationship between various factors and the likelihood of road accidents along a major national highway. Using a multinomial logit model, the study identified that factors such as road curvature, traffic volume, road surface condition, and time of day significantly influenced accident occurrence. The authors emphasized that poor roadway conditions and inadequate traffic control measures contributed to the frequency and severity of crashes in the area. Although the study applied statistical modeling rather than machine learning, it provides important local evidence that environmental and road-condition factors play a crucial role in road accident risk. This supports the relevance of integrating similar variables into predictive models, such as those using machine learning techniques, which the present study aims to implement.

**2.1.4 Predicting Road Traffic Accidents in CALABARZON Using Machine Learning Algorithms.**

Torres and Asor (2021) conducted a study titled “Predicting Road Traffic Accidents in CALABARZON Using Machine Learning Algorithms.” The researchers analyzed regional accident records provided by the Philippine National Police – Highway Patrol Group (PNP-HPG) and applied several machine learning models, including Decision Trees, Random Forest, Naïve Bayes, and Neural Networks. Their results showed that Neural Network and Random Forest models achieved the highest prediction accuracy, reaching up to 87.6%, demonstrating that machine learning can effectively classify accident-prone areas and identify significant contributors such as road type, lighting conditions, weather, and time of day. The study emphasized the importance of integrating environmental and roadway data into predictive systems to support data-driven policy formulation. Although their work focused only on CALABARZON, it shows the potential of ML approaches in improving road safety across different regions of the Philippines, thereby supporting the relevance of the current study.

**2.1.5 Factors Influencing Motorcycle Accidents in Urban Roads of Metro Manila**

Seva et al. (2013) examined motorcycle crash risks in the Philippines through the study “Factors Influencing Motorcycle Accidents in Urban Roads of Metro Manila.” Using logistic regression analysis, the authors identified that environmental factors—such as road surface condition, weather, lighting, and traffic density—significantly influenced crash likelihood. They also found that poor visibility and wet road surfaces sharply increased accident probability. Their research highlights the importance of incorporating environmental and roadway variables into risk prediction models. Although the study used statistical analysis rather than machine learning, it supports the present study’s goal of analyzing environmental and road-condition factors to predict road accidents.

**2.2 Research Questions**

1. Which human and vehicular factors are the most significant predictors of crash severity in road accidents in Magsingal, Ilocos Sur?
2. Can the Random Forest Classifier accurately predict the severity outcome (Fatal vs. Non-Fatal) of road accidents using historical crash data from Magsingal, Ilocos Sur, and how does it perform in classifying the minority (Fatal) class?
3. Based on the Random Forest model's results, what specific high-risk factors should local government units (LGUs) in Magsingal, Ilocos Sur, prioritize to implement data-driven and targeted road safety interventions?

**2.3 Research Hypotheses**

H1: Human and vehicular factors available in the dataset have no statistically significant predictive influence on the severity outcome (Fatal vs. Non-Fatal) of road accidents recorded in Magsingal, Ilocos Sur.

H2: The Random Forest Classifier is not capable of accurately classifying the severity outcome (Fatal vs. Non-Fatal) of road accidents from the Magsingal, Ilocos Sur dataset, performing no better than a random baseline predictor.

H3: The factors identified as having the highest predictive importance by the Random Forest model do not offer actionable or meaningful insights that can effectively inform and prioritize targeted local road safety interventions in Magsingal, Ilocos Sur.

1. **Materials and Methods**

**Research Design**

This study uses Applied Machine Learning focused on a Supervised Classification approach. The objective of this research is to use the model relationship between key Human and Vehicular Factors (the independent variables) and the resulting Crash Severity (the dependent variable) in road accidents in Magsingal, Ilocos Sur. The study employs the Random Forest Classifier to build a robust model that classifies crash severity as either Fatal or Non-Fatal. This design emphasizes extracting actionable insights through Feature Importance Analysis to identify the most critical human and vehicular behaviors (e.g., Accident Causation and Collision Type) driving severe accident outcomes, supporting evidence-based, proactive policy interventions at the local government level.

**3.1 Data Source and Acquisition**

The primary data source for this study was the "ANNEX A: ROAD CRASH DATA SHEET INVOLVING AT USERS", a form prepared by the Department of Transportation (DOTr) in conjunction with agencies like the Philippine National Police (PNP). This dataset was utilized for our analysis, covering accident records from 2019 to 2021 and extending up to 2024 in some records. While the specific dataset used was a compilation of approximately 50,000 entries across major Philippine regions, the foundational data is derived from the official forms submitted by local police units and traffic investigators.

The structure of this data sheet is constrained by the current national reporting practices, which historically involve manually recorded reports transferred from Police Traffic Precinct blotter books. Key features extracted from the Road Crash Data Sheet include Time, Date, Exact Location (Street, LGU), Vehicle Type involved, Collision Type (e.g., Head On, Side Swipe), Causation (e.g., Human error, Drunk driving, Lost control), and Accident Damage/Injury (Severity: fatal, non-fatal, damage to property).

**3.2 Exploratory Data Analysis (EDA)**

Exploratory Data Analysis (EDA) was employed to examine the distribution, trends, and correlations among the primary variables consistently used throughout the study: crash severity, accident causation, collision type, and year. These variables were chosen since they are the primary predictors employed in the Random Forest model and subsequent SHAP interpretation.

The initial study focused on the distribution of crash severity, which revealed a significant class imbalance, with non-fatal incidents accounting for the vast majority of documented cases, while fatal collisions constituted a small but essential fraction of the dataset (Figure 1). This disparity justified the use of stratified sampling and robust classification approaches in subsequent modeling stages.

Temporal analysis using the Year variable revealed an increasing trend in recorded road crashes, particularly in 2023 and 2024, indicating heightened road safety concerns in Magsingal, Ilocos Sur. This tendency supports the use of temporal characteristics in predictive models.

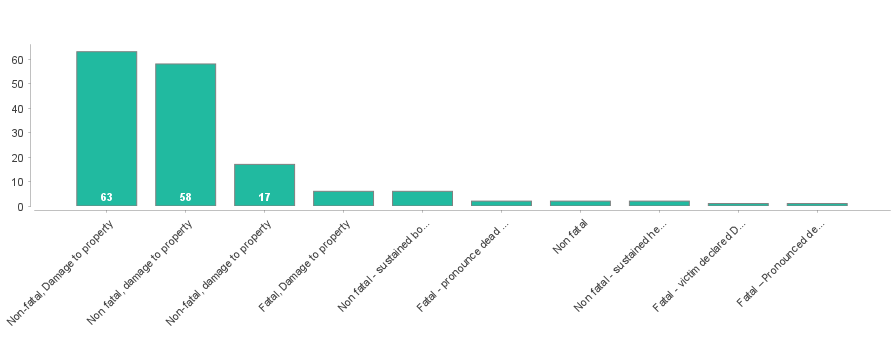


Figure 1. Distribution of Crash Severity (Fatal vs. Non-Fatal)

This figure illustrates the imbalance in the target variable used throughout the EDA, correlation analysis, and Random Forest modeling.

EDA also investigated the frequency distribution of collision types and accident causation factors. The findings revealed that rear-end collisions and self-accidents were the most common, but head-on and pedestrian-related incidents, though less common, were more significantly connected with fatal outcomes. Similarly, human-related causes such as human mistakes, intoxicated driving, and loss of control dominated the causation categories, demonstrating their importance as severity predictors.

Overall, these basic variables were chosen to provide methodological consistency throughout exploratory research, correlation testing, and machine learning modeling. The EDA indicates that temporal, behavioral, and collision-related aspects influence accident severity, validating their consistent usage in correlation analysis and machine learning modeling.

**3.2.1 Relationship Between Collision Type, Causation, and Severity**

Correlation analysis was intentionally limited to the core categorical variables such as Crash Severity, Collision Type, and Accident Causation, as these factors were consistently used in subsequent modeling and SHAP interpretation.

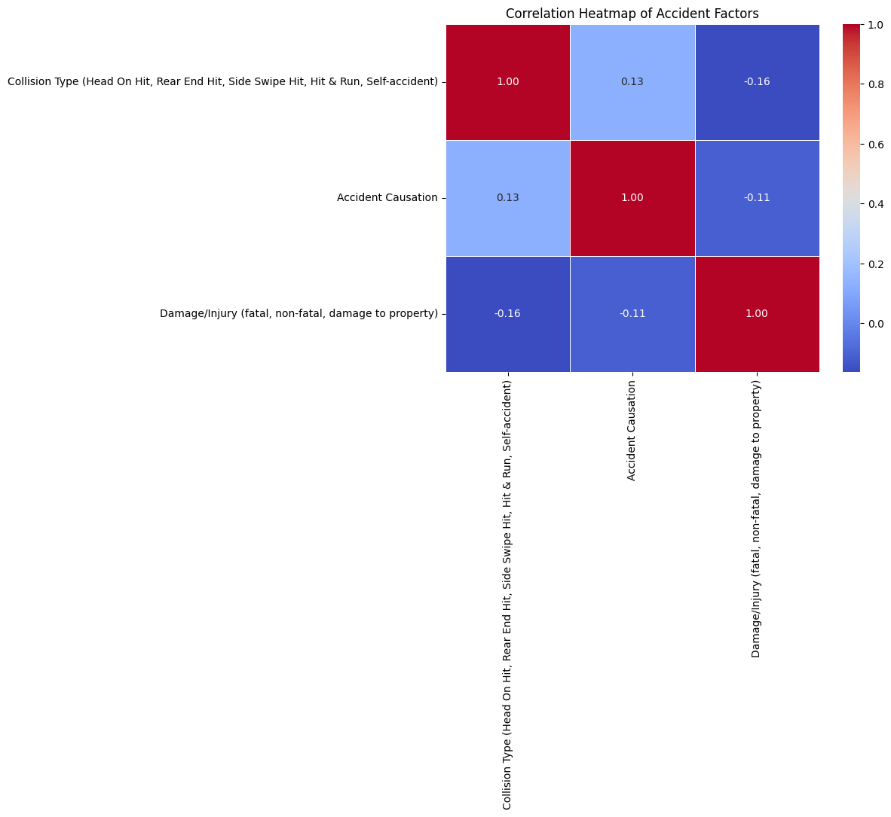
****

Figure 2. Correlation Analysis of Core Crash Factors (Severity, Collision Type, and Accident Causation

Figure 2 depicts a correlation heatmap of the links between Collision Type, Accident Causation, and Damage/Injury Severity. The findings show weak correlations across all factors, implying that no single factor has a strong linear association with crash severity.

The results show a **weak positive correlation** between Collision Type and Accident Causation, indicating that certain causation factors are more likely to occur with specific collision mechanisms:

Weak **negative correlations** were observed between Collision Type and Crash Severity, as well as between Accident Causation and Crash Severity, suggesting that these factors alone do not have a strong linear relationship with fatal outcomes:

Overall, the heatmap emphasizes the multifaceted nature of road crashes and the limits of straightforward correlation analysis for categorical event data. These findings suggest the adoption of machine learning methods like the Random Forest classifier, which can better capture non-linear interactions and combination influences that influence crash severity.

**3.3 Data Preprocessing**

The data pre-processing phase was critical in converting the raw, unstructured accident log data to a format acceptable for the Random Forest Classifier. The procedure began with data cleansing, which included standardizing textual entries and managing any missing values. The primary phase was feature engineering, which involved converting existing date and time stamps into relevant predictive variables such as Time of Day categories, Day of Week, and a Seasonal/Half-Year Factor, which were critical given their later prominence in the model (Figure 4). Simultaneously, the multi-class Damage/Injury outcome was simplified and converted to the needed binary target variable ('Fatal' vs. 'Non-Fatal') for categorization. Finally, all remaining categorical features, such as Collision Type and Barangay, were converted into numerical format using techniques such as One-Hot Encoding before the entire dataset was divided into training and testing sets using a Stratified Split, ensuring that the critical, rare "Fatal" outcome was proportionately represented in both subsets.

**3.4 Machine Learning Model(s)**

The machine learning task chosen for this study is Supervised Binary Classification. The objective is to predict the discrete, categorical outcome of Crash Severity specifically classifying a recorded accident as either Fatal or Non-Fatal based on the available accident attributes (Accident Causation, Collision Type, and AT User Involvement). This classification task is justified as it directly supports proactive safety planning by identifying and mitigating high-risk factors that drive the most severe outcomes in Magsingal, Ilocos Sur. We utilized the Random Forest Classifier, an ensemble learning method selected for its robustness, ability to handle the study's high-dimensional categorical features, and its inherent interpretability via feature importance analysis. A Random Forest is constructed from an ensemble of individual Decision Trees, where each tree is built using a randomly sampled subset of both the data and the features. The final prediction for an accident is determined by a majority vote across the entire forest, aggregating individual tree results

The model was trained on an 80% training set and evaluated on a 20% testing set, with stratified sampling applied to maintain the true proportion of the rare "Fatal" class across both subsets, ensuring a more reliable evaluation of the model's performance on the critical, minority outcome.

**3.5 Mathematical Formulation of the Random Forest Model**

Let the dataset be defined as

where:

* represents the vector of predictor variables (Accident Causation, Collision Type, Time of Day, Barangay, etc.),
* denotes the binary crash severity outcome  
  (0 = Non-Fatal, 1 = Fatal).

A **Random Forest** consists of an ensemble of decision trees:

Each decision tree is trained on a bootstrap sample of the dataset and selects a random subset of features at each split to reduce correlation among trees.

1. **Results**

**4.1 Model Performance**

The Random Forest Classifier was successfully used to carry out the classification task, which was designed to predict the binary outcome of crash severity (Fatal vs. Non-fatal). The model performed well in the classification job, reaching a robust overall predictive accuracy of 84.7% on the hold-out test set, confirming its usefulness for assessing and predicting crash risk severity in the Magsingal dataset. However, a deeper assessment utilizing the Confusion Matrix was required to evaluate the model's performance on the crucially essential but extremely imbalanced "Fatal" class.

The matrix indicates that while the model was highly effective at properly classifying the majority (Non-Fatal) class (34 True Negatives), it struggled with the minority class, successfully recognizing only one out of two genuine Fatal cases (True Positive = 1). The solitary False Negative, which represents a fatal crash that was wrongly projected as non-fatal, highlights the intrinsic difficulty of achieving perfect recall for uncommon, high-impact occurrences, even with high overall accuracy.

**4. 2** **Model Interpretation**

The Random Forest model's robust success is due to its ability to identify and harness critical human and environmental elements that influence accident outcomes. The feature importance analysis found that Accident Causation, Collision Type, and Barangay (location) are the most important predictors of crash severity. Crucially, Figure 4 depicts the decision tree structure within the classifier, which gives direct confirmation of the model's logic. This depiction indicates that the initial and most decisive splits leading to a "Fatal" outcome are primarily dependant on human-controlled variables, notably extreme causation factors such as lost control or drunk driving when combined with extremely damaging collision types such as head-on.

This interpretability emphasizes the importance of targeted policy interventions, suggesting that the most effective method for reducing fatalities must prioritize tougher enforcement against high-risk, behavioral elements over purely reactive infrastructure changes.

Because of the dataset's class imbalance, several evaluation metrics beyond accuracy were employed to provide a thorough assessment of the model's performance.

A chart of different colored squares

AI-generated content may be incorrect.

Figure 4. Colored Confusion Matrix

TP (True Positive) = 1

TN (True Negative) = 34

FP (False Positive) = 1

FN (False Negative) = 1

**Accuracy**

**Result:** Accuracy = **84.7%**

**The Random Forest model was able to classify the road crash incidences correctly by 84.7%. The model performed very well despite the nature of the problem having class imbalance.**

**Precision**

**Result:** Precision = **0.50 (50%)**

When the predicted crash is fatal, the probability of a correct prediction is 50%.

**Recall (Sensitivity)**

**Result:** Recall = **0.50 (50%)**

The model identifies 50% of actual fatal crashes, reflecting the difficulty in detecting events that are rare.

**F1-score**

**Result: F1-Score = 0.50 (50%)**

The F1-score means a balanced but moderate performance in predicting fatal crash outcomes.

**Specificity (Non-Fatal Class**

**Result:** Specificity **= 0.971 (97.1%)**

It turns out to do an extremely good job of classifying non-fatal crashes, which are the majority of the dataset.

The Random Forest model produced the following results for this study: 84.7% accuracy. High accuracy in non-fatal situations. Moderate recall for fatal cases, indicating the difficulty in forecasting uncommon but crucial results. The significance of recall-oriented optimization in subsequent work is highlighted by these findings.

**4.2.1 SHAP-Based Feature Importance Analysis for Road Crash Fatality Prediction**

**A blue and white bar graph

AI-generated content may be incorrect.**

Figure 4. Road Crash Fatality Prediction

The relative significance of the variables affecting the Random Forest model's crash severity prediction (Fatal vs. Non-Fatal) is depicted in the SHAP bar plot. The most significant factor is year, suggesting that temporal trends have a substantial impact on the probability of fatalities, which is in line with the recent rise in traffic accidents.  
  
 Collision Type and Accident Causation have the greatest influence on mortality outcomes among behavioral and crash-related factors. Predictions of fatalities are significantly influenced by high-risk collision mechanisms, including self-accidents, head-on collisions, side-swipes, rear-end collisions, and crashes involving pedestrians. Similar to this, driver behavior plays a major role in serious crashes, as seen by the emergence of causation variables including human error, intoxicated driving, and loss of control as important predictors.

The severity grade is mostly unaffected by lower-ranked causes like mechanical failure and abrupt animal or pedestrian crossings. Overall, the SHAP analysis supports the application of focused behavioral and enforcement-based road safety measures by confirming that temporal trends, collision type, and human-related causation are the main causes of fatal traffic crashes in Magsingal, Ilocos Sur.

**V. Discussion and Conclusion**

**5.1 Interpretation of Findings**

The findings successfully address the research question by identifying and prioritizing the key factors that influence road crash severity in Magsingal, Ilocos Sur. What was learned is that while accident frequency is significant—with a sharp increase noted in 2023 and 2024 (Figures 1 and 2)—the primary problem is not infrastructure but human behavior. The strong performance of the Random Forest Classifier, achieving an 84.7% accuracy, confirms that crash severity is highly predictable using localized data attributes. Critically, the model interpretation (Figure 4) revealed that factors related to driver decision-making, such as Accident Causation (e.g., drunk driving, loss of control, and reckless driving), are the most decisive variables separating fatal outcomes from non-fatal ones. This strongly suggests that a strategy focused on rigorous traffic law enforcement and targeted public awareness campaigns against impaired and reckless driving will yield the greatest reduction in fatal incidents, a finding consistent with broader road safety literature that attributes the majority of crashes to human error.

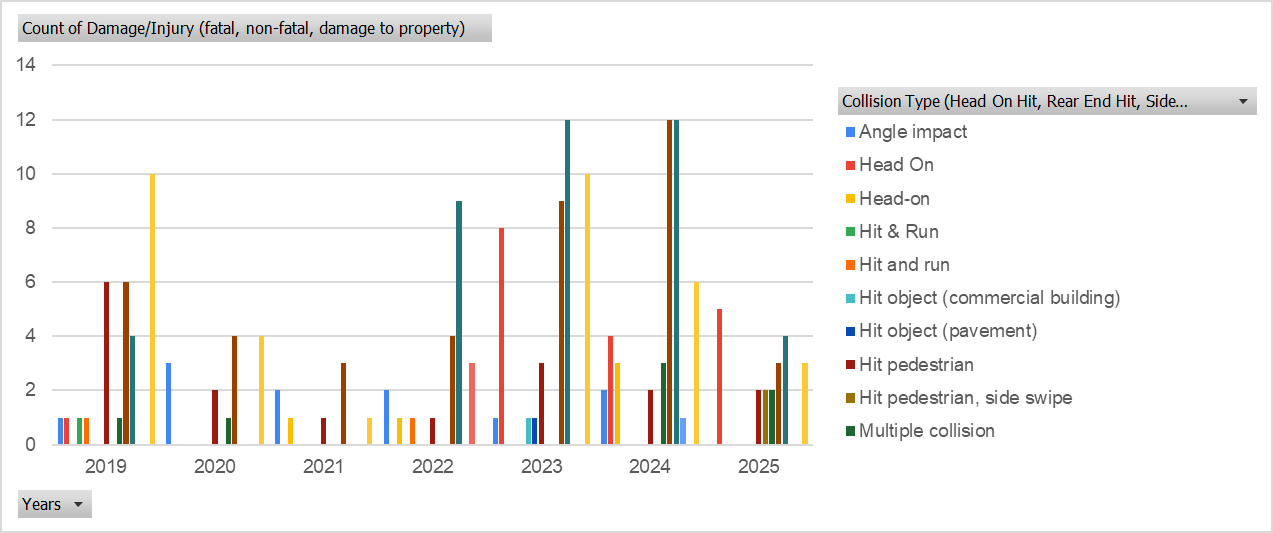


Figure 5. Collision Types by Year (2019–2025)

Figure 4 shows the temporal history of road crashes, emphasizing a significant increase in occurrence frequency over the research period, with a peak volume of accidents in 2023 and 2024. This demonstrates the rising nature of Magsingal's road safety concern and serves as the key basis for predictive modeling.

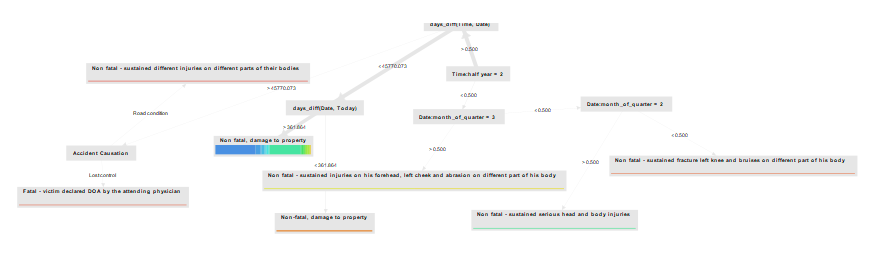


Figure 6. Accident Causation

As the core component of model interpretation, Figure 5 provides a visual map of the prediction logic. It explicitly demonstrates that specific human-controlled factors, such as Accident Causation (e.g., Lost Control and Drunk Driving), are the definitive, high-leverage splitting criteria in the model’s tree structure that lead to the classification of fatal crashes.

**5.2 Limitations**

Despite obtaining excellent overall accuracy, this study has numerous limitations in its data, methodology, and model. The most critical data limitation is the dataset's severe class imbalance (as shown in Figure 3), which fundamentally undermines the model's ability to achieve perfect Recall on the rare but critical "Fatal" class, as evidenced by the single False Negative in the Confusion Matrix. Furthermore, relying on police-reported data involves potential reporting bias and under-reporting, particularly for minor property damage cases, which may distort the true distribution of accident frequency.

Methodologically, the analysis was limited to data available in local records, omitting the incorporation of highly relevant external elements such as real-time traffic volume, specific road geometry, or transient weather conditions, all of which are established predictors of crash risk. Finally, while the Random Forest model is understandable (Figure 4), it reflects correlation rather than causation, which means that the model highlights which features precede a severe consequence but does not definitively prove a direct relationship.

**5.3 Future Work**

To expand on these findings, future research should explore a variety of paths for improvement and expansion. To begin, further modeling attempts should explicitly address the class imbalance by using advanced approaches such as SMOTE (Synthetic Minority Oversampling Technique) or applying a cost-sensitive learning approach to prioritize strengthening the model's recall of the "Fatal" class. Second, the existing research should be improved by including Geospatial Information System (GIS) analysis to identify specific road segments or intersections (black spots) that pose the greatest risk, as this would be more valuable for infrastructure planning than broad Barangay categorization.Third, external dynamic data sources, such as hourly weather logs or traffic flow statistics, must be added to the predictive framework in order to create a more robust and responsive risk prediction model. Finally, the project's utility could be increased by creating an interactive, real-time dashboard for the Magsingal Local Government Unit (LGU), which would allow law enforcement agencies to enter current conditions and receive immediate, actionable risk scores to optimize patrol deployment and resource allocation.

**References**

Abdel-Aty, M., & Haleem, K. (2011). Analyzing crash injury severity using decision tree and random forest techniques. *Accident Analysis & Prevention, 43*(1), 119–128. <https://doi.org/10.1016/j.aap.2010.08.007>

Ahmed, S. K., Mohammed, M. G., Abdulqadir, S. O., El‐Kader, R. G. A., El‐Shall, N. A., Chandran, D., ... & Dhama, K. (2023). Road traffic accidental injuries and deaths: A neglected global health issue. *Health science reports*, *6*(5), e1240.

Asian Development Bank. (2022). *Philippines transport sector assessment, strategy, and road map*. Asian Development Bank. <https://www.adb.org>

Asian Transport Observatory. (2025). *Philippines Road Safety Profile 2025*. Asian Transport Observatory.<https://asiantransportobservatory.org/analytical-outputs/roadsafetyprofiles/philippines-road-safety-profile-2025/>

Chang, F., Xu, P., Zhou, H., Chan, A. H. S., & Huang, H. (2019). Investigating injury severities of motorcycle riders using advanced predictive modeling. *Accident Analysis & Prevention, 131*, 132–140. <https://doi.org/10.1016/j.aap.2019.06.014>

Fruelda, M. J., Fampulme, S. L., Fontamillas, F., Lilang, J., Iii, A. F., Madla, I., ... & Gacu, J. (2025). Road Infrastructure Assessment and Traffic Dynamics Using GIS: A Case Study in the Philippines. *Rev. Int. De Geomat*, *34*, 187-207.

Haque, M. M., Chin, H. C., & Huang, H. (2009). Modeling fault among motorcyclists involved in crashes. *Accident Analysis & Prevention, 41*(2), 327–335. <https://doi.org/10.1016/j.aap.2008.12.011>

Macalalag, E. R. A., & Lucia, M. A. S. (2021). 3A. 002 Rapid assessment of vulnerability of philippine local governments to road crash.

Norona, M., Rodriguez, R., & Villamaria, J. (2021, April). Analysis of Factors Affecting Road Traffic Accidents in the City of Makati Philippines. In *2nd South American International Conference on Industrial Engineering and Operations Management*, https://doi.org/10.46254/SA02.20210602.

Philippine National Police – Highway Patrol Group. (2024). *Annual road crash statistics report*. PNP-HPG.

Philippine Statistics Authority. (2023). *Road traffic accident statistics in the Philippines*. <https://psa.gov.ph>

Philstar. (2025). Road crashes surge by 35% in 2024 — HPG report. *Philstar News*. <https://philstar.com>

Pour, S. A., Shirazi, M. S., & Rahimi, A. (2020). Predicting traffic accident severity using machine learning techniques. *Journal of Advanced Transportation, 2020*, Article 5475281. https://doi.org/10.1155/2020/5475281

Seva, R. R., Alampay, R. B., & Cortez, J. M. (2013). Factors associated with motorcycle-related road crashes in Metro Manila. *Accident Analysis & Prevention, 59*, 452–457. <https://doi.org/10.1016/j.aap.2013.06.025>

Velasco, V. P., Robielos, R. A. C., & Gumasing, M. J. J. (2021). Analysis of road traffic accident criticality in CALABARZON Philippines. In *Proceedings of the 11th Annual International Conference on Industrial Engineering and Operations Management*.

Wegman, F. (2017). The future of road safety: A worldwide perspective. *IATSS research*, *40*(2), 66-71.

Tanseco, O. S. (2019, August 27). *Data system for road safety: DRIVERS implementation in the Philippines* [PowerPoint presentation]. Capacity Development Workshop on Road Safety for the Southeast Asia Subregion, Manila. [https://www.unescap.org/sites/default/files/2-%20Presentation%20on%20Data%20System%20f r%20Road%20Safety-Philippines-HPG.pdf](https://www.unescap.org/sites/default/files/2-%20Presentation%20on%20Data%20System%20for%20Road%20Safety-Philippines-HPG.pdf)

Torres, M. A., & Asor, J. R. (2021). Predicting road traffic accidents in CALABARZON using machine learning algorithms. *Philippine Computing Journal, 16*(2), 20–32.

Umer, S., Sadiq, S., & Ishaq, A. (2020). Machine learning models for predicting traffic accident severity: A comprehensive review. *International Journal of Advanced Computer Science and Applications, 11*(10), 1–15. https://doi.org/10.14569/IJACSA.2020.0111001

Xu, C., Wang, W., & Huang, H. (2017). Modeling crash data with random forest and Bayesian networks. *Accident Analysis & Prevention, 108*, 226–235. https://doi.org/10.1016/j.aap.2017.09.019