



Predicting Road Accident Fatalities in India: Comparative study of Time Series and Machine Learning Approaches

Prof. (Dr.) Gopal Singh Latwal¹ and Ms. Leena Gupta²

¹Professor, Institute of Information Technology & Management, New Delhi

²Assistant Professor, Institute of Information Technology & Management, New Delhi

¹gopalsingh.latwal@iitmipu.ac.in

²email2leena@gmail.com

Abstract. With the expansion of technology and increasing use of vehicle road accidents become more frequent worldwide result of which many people have died in them over the years. Currently traffic accidents have become one of the main causes of mortality and injury in the world. Current study is done to create a predictive model that can be used for future decision making using past data. A comparative study is done using previous ten year data on three models i.e. the ARIMA model, the Exponential Smoothing model and the random forest prediction model to predict number of accidents in next ten years. The analysis is done on historical data and by comparing the result it is found that random forest perform beat in comparison to ARIMA and exponential Smoothing model whereas it is also found that ARIMA is good across all the matrices for consistency. As we analyze the data for predicting fatalities, injury and non-injury cases, ARIMA is best for non-injury care also. This study will help researchers who want to study in the same domain and also the policy makers

Keywords: Accident, ARIMA, Exponential Smoothing, Injury, Random Forest.

1. Introduction

Over the years, lots of individuals lost their lives in road accidents as Road accidents are becoming more and more frequent in the nation. Road accidents have become the one leading cause of death and injury all over. It is a critical global issue as 1.3 million people approximately die per year according to WHO (World Health Organization), and also millions of people all over the world suffer from various fatal injuries leading to long-term suffering or disabilities. People face its consequences not only as a loss of individual but also in the form of lost productivity, healthcare cost and property damage.

In India Road accident becomes the serious problem [1]. Worldwide road accidents have become the one of the primary cause of deaths, which affects the mainly the age group of between 15 and 49. Over 1.68 lakh death and over 4.4 lakh injuries were reported in 2012.

There are so many factors that cause road accidents to lower the number of accidents some integrated strategy approach is to be used. Road accidents affect persons in their prime working year and sometimes also cause death. In WHO report 2023 it is found that around 69% people account for fatalities are from 18 and 59 age group whereas 60 and older accounts for 23% with female and male overall fatalities ratio 1:03, which Show Road Accidents are major source of fatality and injury but also affects financial impact as estimated 1-3% GDP to 6% in certain cases, they are the big obstacle in the growth and health of a Country [2].

This is matter of concern especially for policy makers as government is working to improve the living standard of its people and society suffer with permanent damage as a result of these fatalities. Death and injuries are continuously increasing worldwide matter of concern for health and development. According to WHO Report 2003 these fatalities has a negative impact on society health; social substance and economy as two-thirds deaths occur among people of working age (18-59 years old). India has the well-connected and Coherent transportation system that help in economic development by providing equal distribution of goods & Services, and also providing movement of people. Now the sharer of transport sector is increasing day by day and becomes one of the main contributors in country GDP.

Important key indicators in evaluation of socio-economic development of the country with impact on the rate of growth Indian National Highway much match the pace of economic growth due to second largest road network in the world with the total length of 62.16 lakh kilometers including National Highway (NH), Expressways, MDR(Major District Roads), State Highways, village Roads etc[3].

Purpose of the study is to analyze the past data from related to accidents in India from 2012-2022 to predict the possibilities of accident in next decade. The author emphasize on the research which will provide guidance to researcher who want to research in this field of accidental data and also help policy makers to formulize the policy to prevent number of accidents in future. The ARIMA model, Exponential smoothing and Random Forest predictive models are used for this study.

2. Literature Review

Table 1: Summary of Literature Review

S. No.	Paper Title	Authors	Year	Key Findings
1	Forecasting Road Accidental Deaths in India: ARIMA vs Exponential Smoothing	P.K. Swain, M.R. Tripathy, K. Agrawal	2021	Data from ARIMA (2,2,2) model performed better for accident death prediction in India. [4]
2	Application of ARIMA Model for Forecasting Road Accident Deaths in India	A.N. Patowary, M.P. Barman, B. Dutta	2019	Forecasted accident deaths accurately using ARIMA with historical data. [5]
3	ARIMA Model for Forecasting Traffic Crash Consequences in Saudi Arabia	Sabenorio et al.	2022	Predicted fatalities and injuries effectively for crash prevention planning. [6]
4	Time Series Forecasting of Road Accidents Using SARIMA and Prophet Models	Agyemang et al.	2023	SARIMA outperformed Prophet in seasonal accident prediction in Ghana.[7]
5	Review on Action Recognition for Accident Detection in Smart Transport	V. Adewopo, N. Elsayed, M. Ozer et al.	2022	Surveys AI/vision models for real-time accident detection in smart cities.[8]
6	AI on the Road: Traffic Accident Detection in Smart Cities	Adewopo et al.	2023	Introduced real-time accident detection using AI in urban camera networks.[9]
7	Deep Learning-based Black Spot Identification in Greek Roads	I. Karamanlis, A. Kokkalis et al.	2023	Identified crash hotspots and proposed BSNG dataset for model benchmarking.[10]
8	Explainable ML Approach to Predict Accident Fatalities	M.A.K. Rifat, A. Kabir, A.S. Huq	2024	ML predicted fatal vs. non-fatal crashes in Bangladesh with high accuracy.[11]
9	Review of Road Traffic Accident Prediction Methods	Wang Shun-shun et al.	2023	Reviewed ARIMA, ARCH/GARCH, ANN and ML accident forecasting methods.[12]
10	Grouped Random Parameters NB-Lindley for Crash Data	A.S.M.M. Islam, M. Shirazi, D. Lord	2023	Improved crash frequency modeling by accounting for heterogeneity[13].
11	Statistical Analysis of Highway Crash Severity: Review & Alternatives	Savolainen, Mannering, Lord, Quddus	2011	Comparative study of severity modeling methods (logit, probit, mixed logit).[14]
12	Big Data & Prediction vs Causality in Highway-Safety Analysis	F. Mannering et al.	2020	Discussed trade-offs of predictive models vs causal inference in crash studies.[15]

13	Transport Planning & Traffic Safety	D. Mohan & G. Tiwari	2016	Comprehensive book on traffic injury prevention and Indian transport planning.[16]
14	World Report on Road Traffic Injury Prevention	M. Peden (WHO)	2004	Set global framework for road injury prevention and safety targets.[17]
15	Pachgaon Chowk Accident Blackspot Analysis (India)	Rajesh Mohan (Traffic Police Study)	2025	Identified 11 deaths in 3 years, suggested redesign of roundabout. [18]
16	Minor Road Accident Trends in Thiruvananthapuram	Kerala Economics & Statistics Dept.	2024	60% of accidents on minor roads, highlighting enforcement gaps. [19]
17	Waymo Vulnerable Road Users (VRU) Accident Dataset	Waymo Research Team	2024	Created VRU dataset; occlusion major risk factor for AVs and vulnerable users. [20]
18	Timing of Fatal Crashes in Texas	NHTSA & Bader Scott Study	2024	Most fatal accidents occur 9–10 PM, related to alcohol & low visibility. [21,22]
20	SARIMA and Hybrid Models for Traffic Accident Forecasting	Lavrenz et al., Park et al., Naqvi et al.	2022	Hybrid deep learning-SARIMA improved prediction accuracy significantly. [23]

3. Objectives

1. To analyze the historical trends and patterns of road accident fatalities in India.
2. To develop and apply forecasting models using ARIMA, Exponential Smoothing, and Random Forest techniques.
3. To compare the forecasting performance and accuracy of ARIMA, Exponential Smoothing, and Random Forest models.

4. Research Methodology

The present study is based on secondary data obtained from the Road Accidents in India–2023 report published by Government of India - Ministry of Road Transport and Highways (MoRTH)[3]. In addition to analysing changes in regard to road infrastructure and total vehicle use reports to provide full statistics of accident related activities for the year 2022, the United Nations Economics and Social Commission for Asia and the Pacific (UNESCAP) applies standard format to compile States and Union Territories police records every year under the project APRAD (Asia-Pacific Road Accident Data).

Additionally the data from Accidental Deaths in India (ADSI) was also used in this study, an annual publication of the National Crime Records Bureau (NCRB) based on data combined through District and State Crime Records Bureaus, to guarantee data accuracy and completeness.

5. ARIMA Model

ARIMA - Auto-Regressive Integrated Moving Average) is most simplified extensively applied model for time series analysis. This model is popular because of its quick implementation, capacity to represent variations in trends, periodic fluctuations and cyclic behavior. It is a four step process. First, to choose suitable candidate models, second Estimation of parameters, third Verification of residuals to confirm noise characteristics and last forth one is forecasting using the chosen model. Specification comprise of three elements. ARIMA (P, d, q)

AR (P):- Component incorporates past observations.

d :- steps to remove non-stationarity

MA (q):- Moving average component that uses past forecast errors.

Equation Used:

$$\nabla^d y_t = c + \phi_1 \nabla^d y_{t-1} + \phi_2 \nabla^d y_{t-2} + \dots + \phi_p \nabla^d y_{t-p} + e_t - \theta_1 e_{t-1} - \theta_2 e_{t-2} - \dots - \theta_q e_{t-q}$$

Where:

$\nabla^d y_t$ = repeated differencing for $d > 1$

c = constant term

ϕ_i = autoregressive coefficients (AR part)

θ_j = moving average coefficients (MA part)

e_t = white noise error term ($e_t \sim N(0, \sigma^2)$)

6. Exponential Smoothing

Exponential smoothing was developed in early 1950's, popularly used for forecasting. In this model, recent observations are used as it is assumed that they are more useful for future values than the older ones, thus higher weight age is given to recent observations. One or more smoothing parameters are used to control weighting which shows older data points lose influence very quickly. It is used for its

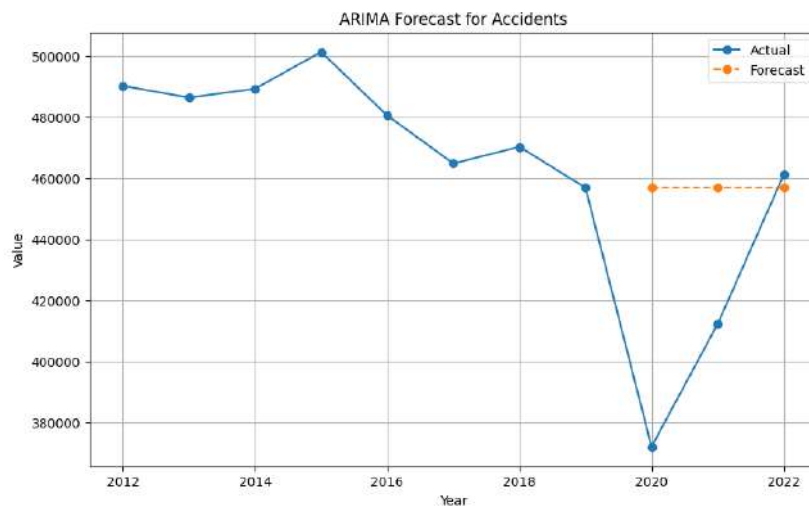
simplicity, easy to implement, less computational and storage requirement and response well in changing data. Exponential smoothing is cost effective and provides reasonable accuracy, very useful in practical situations including real time applications.

7. Random Forest

An ensemble of multiple decisions trees, improves prediction accuracy by averaging outcomes. It combines many independent models using bagging and random feature selection [24]. Random subsets of features are used as input which reduces over fitting, increases robustness and improves predictive accuracy. It performs well on complex datasets with noisy variables and non-linear relationship.

$$y = \text{mode} \{T_1(x), T_2(x), \dots, T_N(x)\}$$

8. Data analysis and Discussion



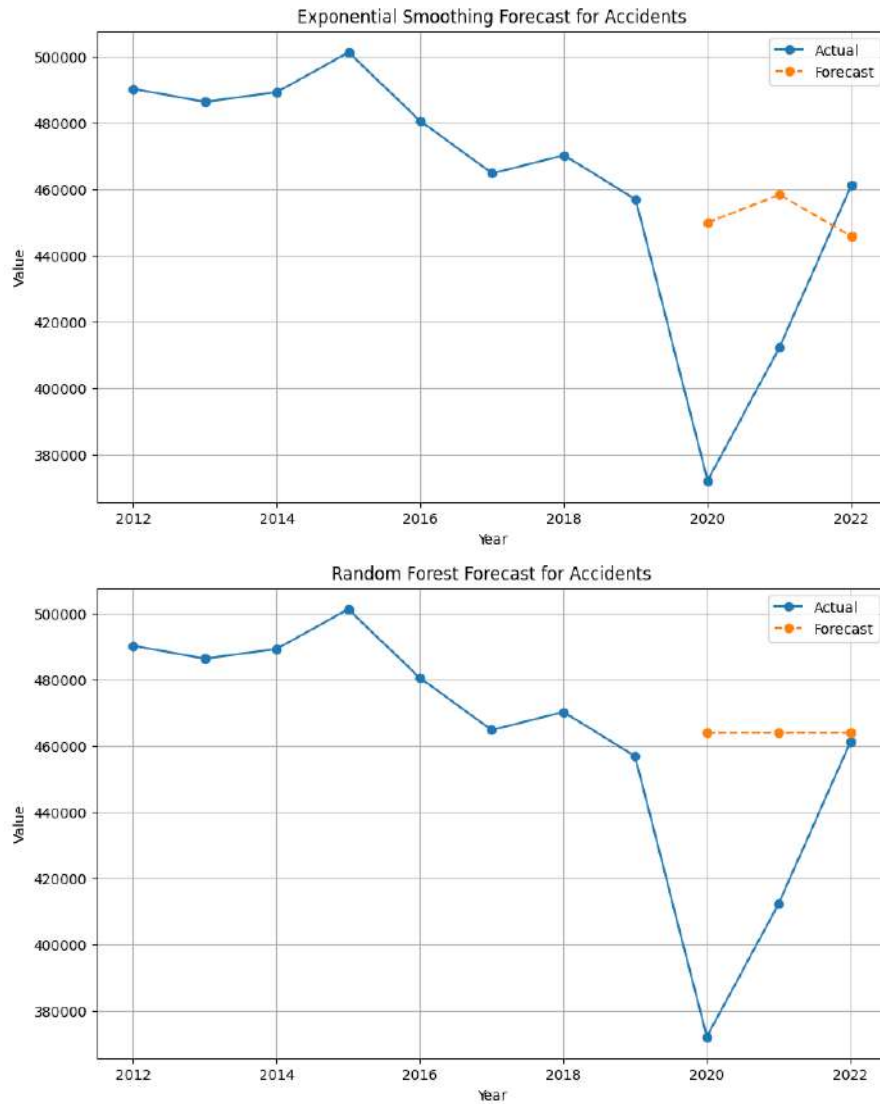


Fig. 1. Total Accidents Forecasting

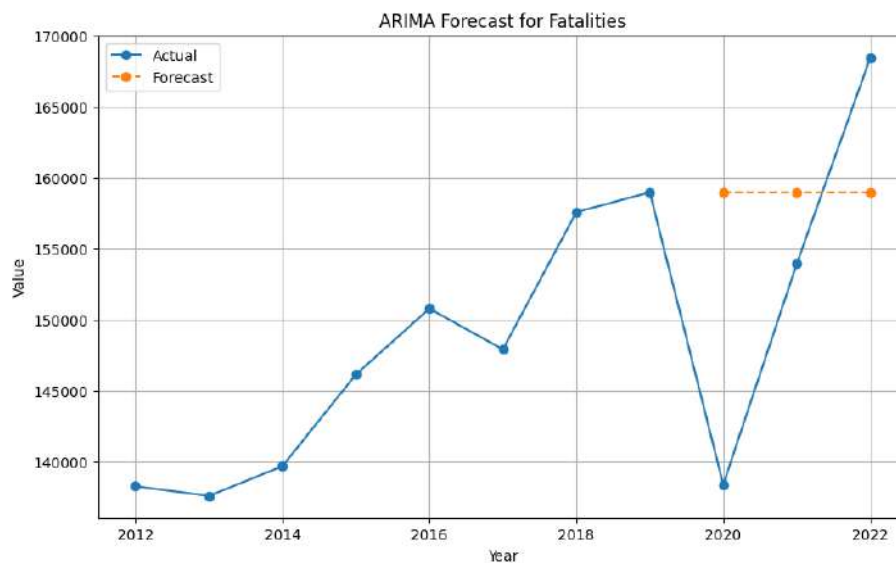
Table 2. Model Summary of Total Accidents Forecasting

Models	RMSE	MAPE
ARIMA	55429.02	11.52%
Exponential Smoothing	52975.00	11.81%
Random Forest	61002.59	12.64%

9. Total Accidents Forecasting

Table 2 and Fig.1.shows Consistent performance is shown by all three models for predicting the total number of accidents: The ARIMA model showed moderate accuracy and lowest MAPE of the three models with RMSE: 55,429.02, MAPE: 11.52%. Perfect representation of underlying trend patterns were shown by ARIMA using time series plots. Presence of identifiable cyclical patterns in accident data shows well modeled seasonal variation. Through exponential smoothing lowest RMSE was obtained (RMSE: 52,975.00, MAPE: 11.81%) which shows improved point forecast accuracy. Despite enhanced absolute performance the minor increase in MAPE show minor increase in percentage errors rate. Visually examined the forecasts with adequate confidence intervals and smooth trend continuation. Random Forest (RMSE: 61002.59, MAPE: 12.64%) shows highest error metrics which indicates challenges in capturing temporal connections.

The Lack of explicit structure consideration is the limitation of the ensemble technique but non-linear relationship by the model in the data was captured by the model.



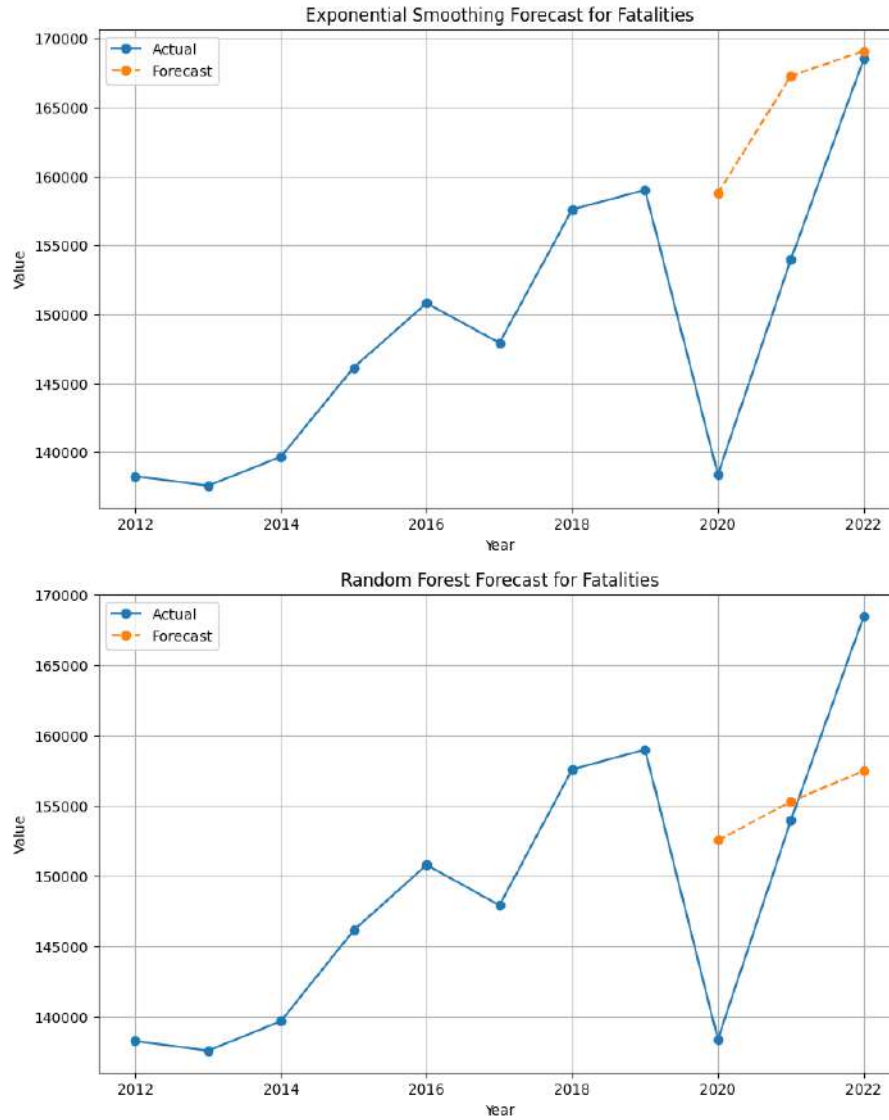


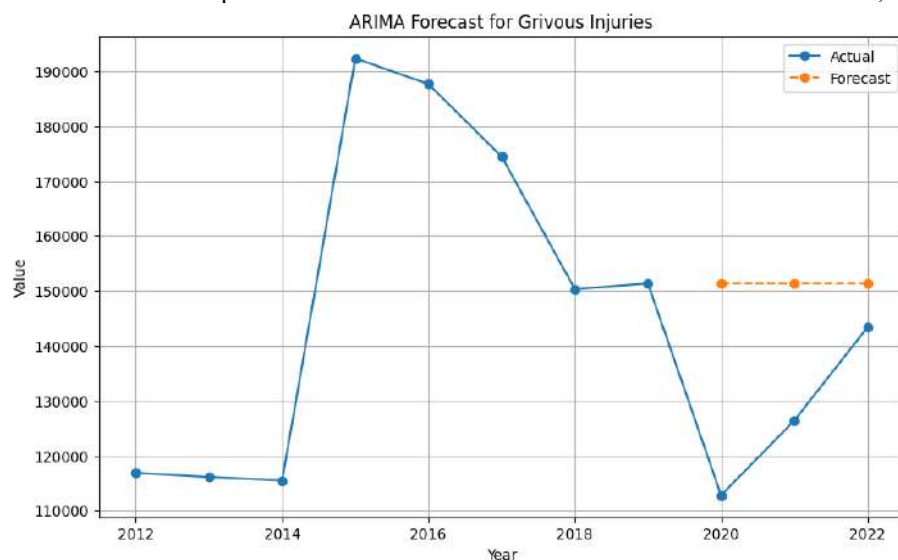
Fig. 2. Fatalities Forecasting

Table 3. Model Summary of Fatalities

Models	RMSE	MAPE
ARIMA	13416.99	7.93%
Exponential Smoothing	14045.98	7.90%
Random Forest	10378.99	5.87%

10. Fatalities Forecasting

Table 3 and Fig.2.shows Interesting model dynamics were demonstrated by Fatality prediction & ARIMA (RMSE: 13,416.99, MAPE: 7.95 shows competitive performance with satisfactory accuracy. The higher RMSE than Random Forest indicates that linear autoregressive techniques may not correctly collect some temporal patterns over the time even having higher RMSE. Exponential Smoothing (RMSE: 14,045.98, MAPE: 7.90%) perform better than ARIMA in respect of MAPE. This discrepancy shows the presence of a few considerable forecast errors that increase RMSE while maintaining acceptable percentage accuracy. By achieving lowest error rate Random forest (RMSE: 19378.99, MAPE: 5.87%) shows outstanding performance across both metrics. The higher efficiency indicates that complex, non-linear factors captured by Random forest have an impact on Fatality patterns. The result shows that feature-based approaches predict better Fatalities in comparison to pure time-series methods,



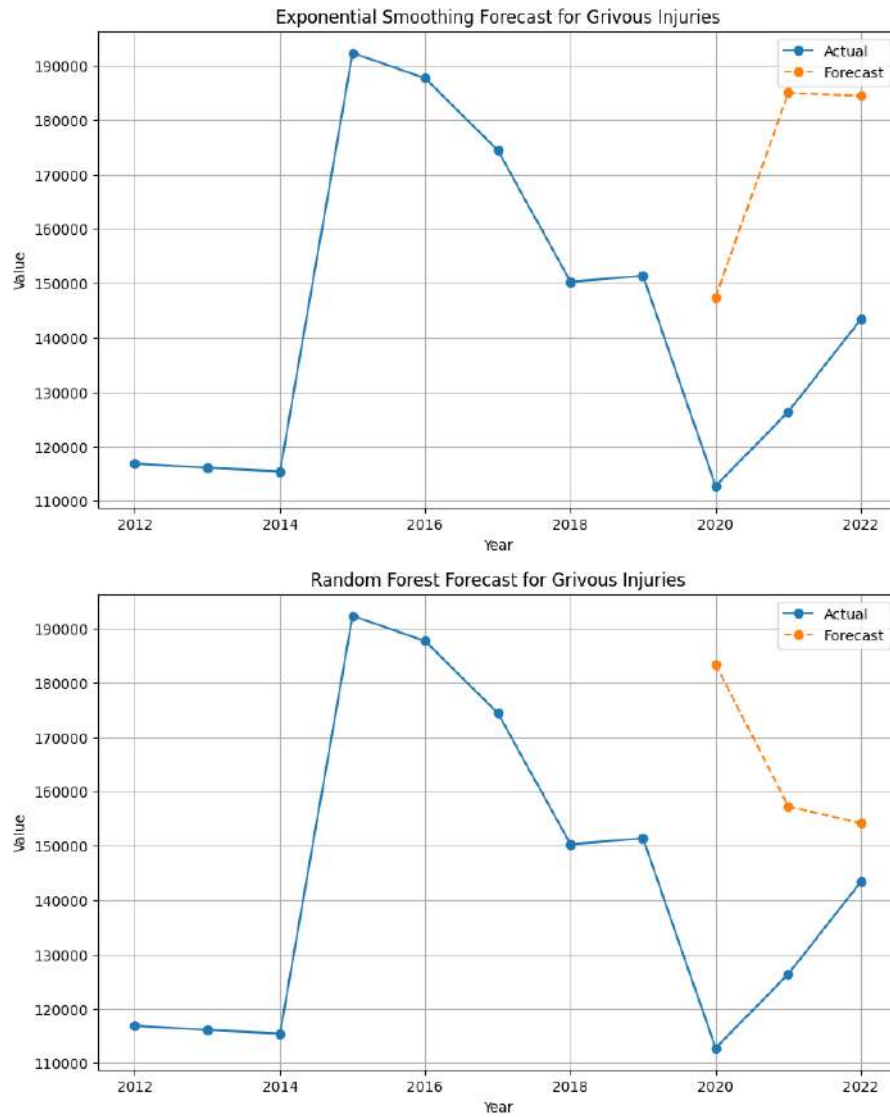


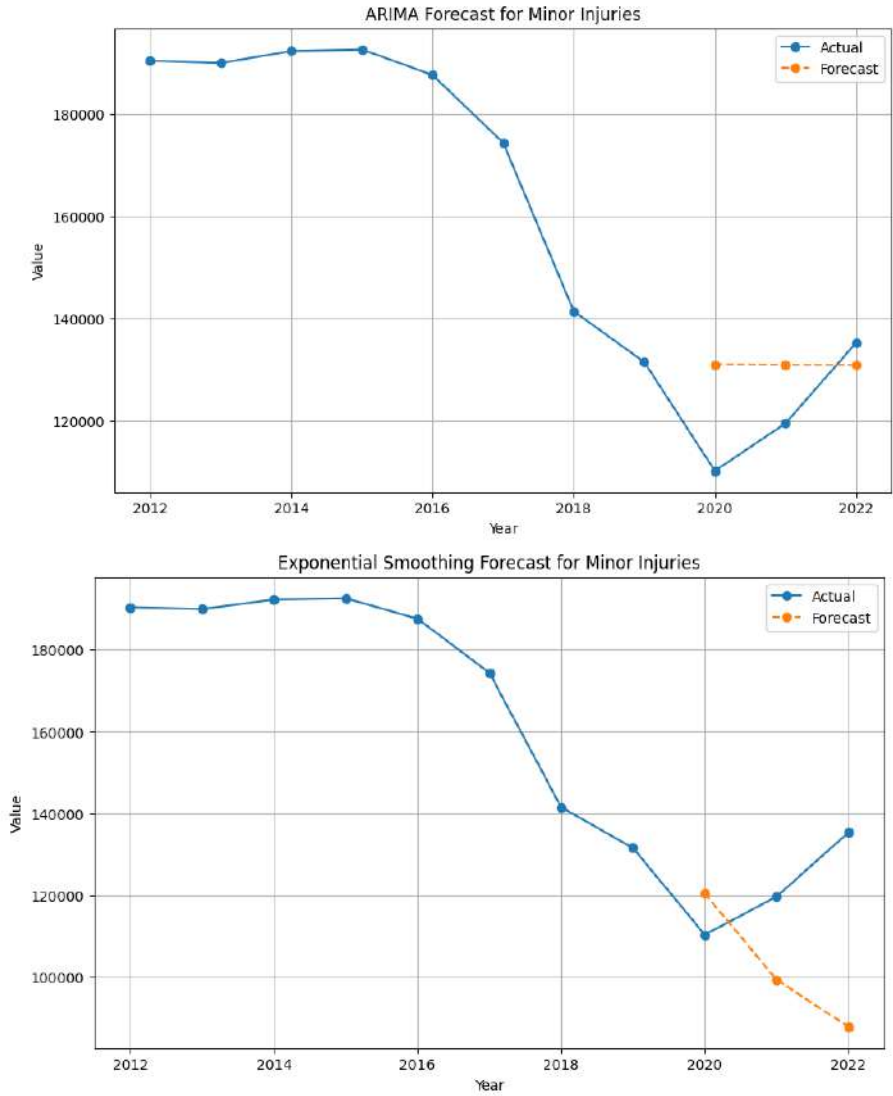
Fig.3. Grievous Injuries Forecasting

Table 4 Model Summary of Grievous Injuries Forecasting

Models	RMSE	MAPE
ARIMA	26912.48	19.83%
Exponential Smoothing	45883.71	35.23%
Random Forest	44926.46	31.51%

11. Grievous Injuries Forecasting

Table 4 and Fig. 3. shows that In grievous injury prediction most significant performance difference were observed where ARIMA (RMSE: 26,912, MAPE: 19.83%) outperformed which indicates that grievous injury patterns follow identifiable autoregressive patterns. Strong historical temporal dependencies were shown by grievous injury data by giving superior performance. Poorest performance was shown by Exponential smoothing which indicates that simple trend smoothing may not be suitable for grievous injury data with RMSE: 45,883.71, MAPE: 35.23%). Systemic forecasting bias is appeared by the high MAPE in this category. Random Forest with moderate performance (RMSE: 44,926.46, MAPE: 31.51%) indicates that feature based methods are able to explain grievous injury patterns.



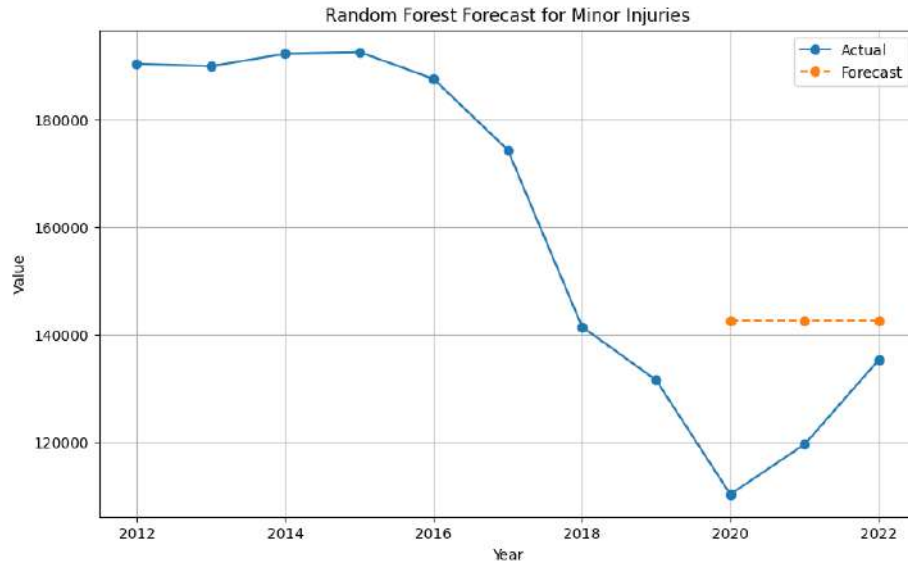


Fig. 4. Minor Injuries Forecasting

Table 5. Model Summary of Minor Injuries Forecasting

Models	RMSE	MAPE
ARIMA	13925.23	10.54%
Exponential Smoothing	30431.11	20.43%
Random Forest	23180.60	17.87%

12. Minor Injuries Forecasting

Table 5 and Fig. 4. shows that ARIMA (RMSE: 13,925.23, MAPE: 10.54%) with lowest errors metrics performed well. The better performance indicates that ARIMA can accurately representing the predictable temporal pattern of minor injuries. Exponential smoothing with poorest performance (RMSE: 30,431.11, MAPE: 20.43%) indicates that simple smoothing approaches are not suitable for injury-related forecasting. However, moderate performance of random forest (RMSE: 23,180.60, MAPE: 17.87%) shows that using ensemble approaches one can get some success in capturing minor injury patterns

Table 6. ARIMA Model Comparison Results

Model	Log-L	AIC	BIC	ME	RMSE	MAE	MPE	MAP E	MAS E
ARIMA (0,2,0)	-316.09	634.18	635.50	2253.7 9	18684.4 3	11709.65	184.56	345.00	1.41
ARIMA (0,2,1)	-308.85	621.71	624.37	1736.2 6	13587.5 3	7870.55	175.56	203.74	0.94
ARIMA (0,2,2)	-302.36	610.72	614.72	899.08	9750.35	5357.14	93.63	97.91	0.64
ARIMA (1,2,0)	-313.25	630.50	633.16	1946.9 9	16734.2 6	10296.91	288.04	334.75	1.24
ARIMA (1,2,1)	-305.64	617.28	621.27	925.84 9	11710.3 9	6945.00	172.30	172.72	0.83
ARIMA (2,2,2)	-298.66	601.62	608.25	-30.67	6905.65	4512.30	-41.29	156.11	0.54
ARIMA (0,2,4)	-297.53	605.06	611.72	-445.14	7179.80	4149.91	17.94	99.50	0.50
ARIMA (0,2,5)	-298.02	608.04	616.03	772.92	7494.74	4701.24	-12.72	178.59	0.56
ARIMA (0,2,6)	-296.04	606.08	615.41	-444.04	6328.56	4158.59	-25.10	129.61	0.50

Best Model by AIC:

Model	Log-L	AIC	BIC	ME	RMSE	MAE	MPE	MAPE	MASE
ARIMA (2,2,2)	-298.66	601.62	608.25	-30.67	6905.65	4512.30	-41.29	156.11	0.54

13. Akaike Information Criterion (AIC) Selection

Table 6 depicts the ARIMA (2,2,2) with the lowest AIC value of 601.62 found to be optimal specification. In order to prevent over fitting the AIC restricts models with additional parameters when optimizing model fit against complexity.

14. Bayesian Information Criterion (BIC) Considerations

Though ARIMA (2,2,2) obtained the best AIC, but its BIC value of 608.28 is competitive and not lowest among simpler models. BIC imposes model complexity more significantly than the AIC, as simpler models like ARIMA (0,2,4) produce identical BIC values. Whenever predictive accuracy is more crucial than model parsimony, the AIC based selection is continue to be effective for forecasting purposes. This significant enhancement indicates the importance of properly specified AR and MA components

15. Root Mean Square Error (RMSE): 6,905.65

- The selected model achieves a relatively low RMSE, indicating good absolute forecast accuracy
- When compared to the baseline ARIMA(0,2,0) model (RMSE: 18,684.43), the improvement represents approximately **63% reduction in forecast errors**

16. Mean Absolute Error (MAE): 4,512.30

- The MAE provides a robust measure of forecast accuracy less sensitive to outliers than RMSE
- In comparison to naive seasonal forecasting techniques the MASE (Mean Absolute Scaled Error) of **0.54** indicates that the model performs significantly better.
- The ARIMA (2,2,2) model provides significant predictive value when the value of MASE is less than 0.

17. Bias Assessment

17.1 Mean Error (ME): -30.67

- The near-zero mean error indicates minimal systematic bias in forecasts
- This suggests that the model neither consistently over-predicts nor under-predicts accident levels
- The small magnitude relative to the forecast scale demonstrates well-calibrated predictions

17.2 Mean Percentage Error (MPE): -41.29%

- The negative MPE indicates a slight tendency toward over-prediction
- However, the magnitude is reasonable and suggests no severe systematic bias issues

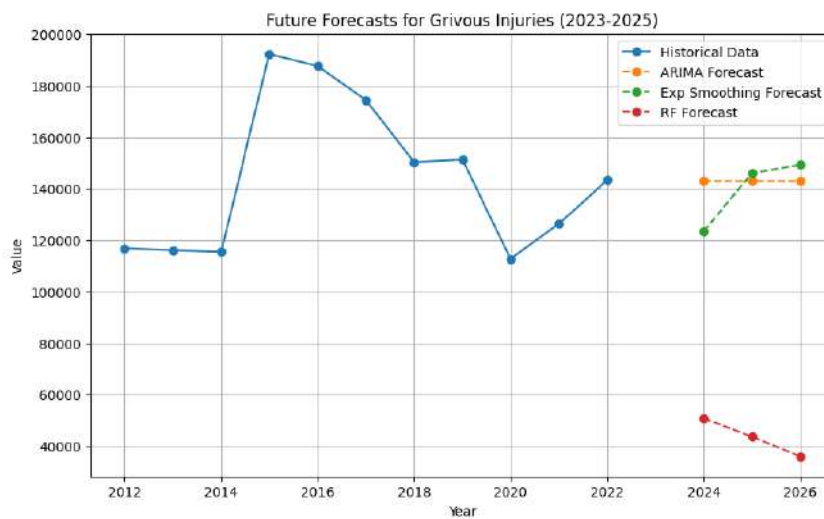
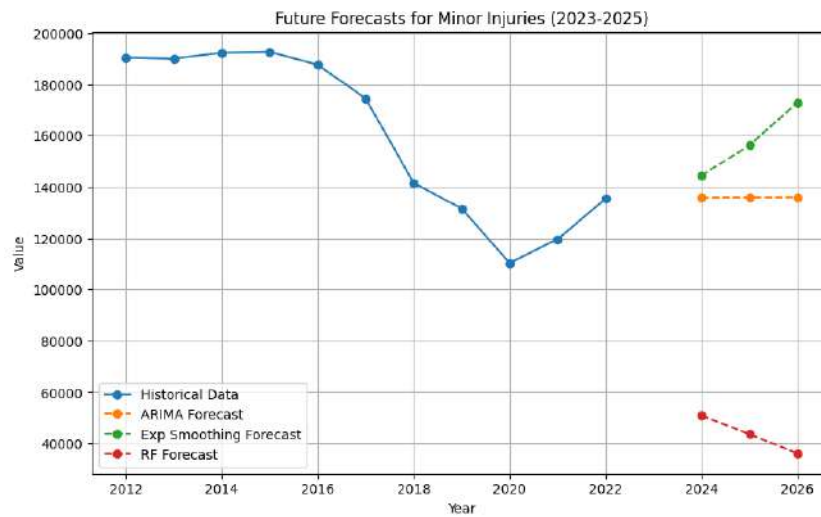
18. Percentage Accuracy Metrics

18.1 Mean Absolute Percentage Error (MAPE): 156.11%

- The relatively high MAPE requires careful interpretation in the context of accident data
- High MAPE values can result from periods with very low accident counts (near-zero denominators)
- This signifies that the model may have accuracy issues and during low

accident occurrence.

- It is found that Absolute error metrics (RMSE, MAE) is of greater importance than percentage-based metrics for planning and policy purposes



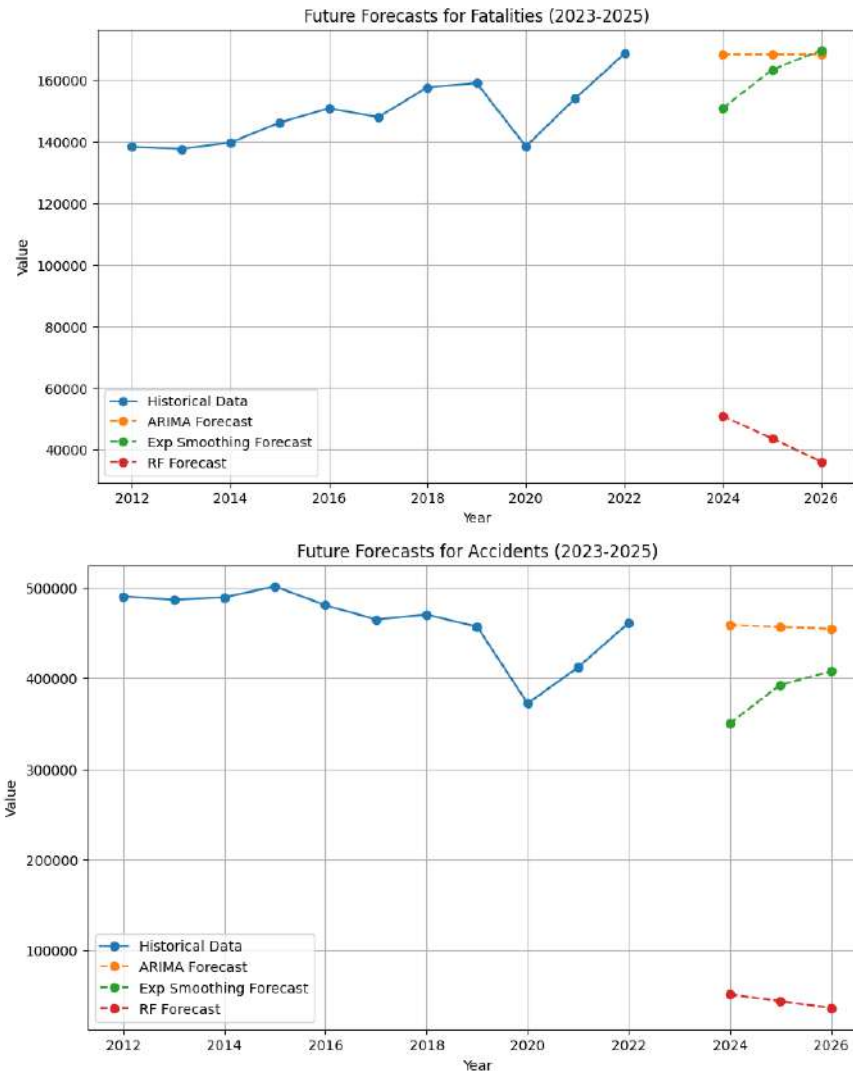


Fig. 5. Future Forecasting of Accident, Fatalities, Minor Injuries and Grievous Injuries

19. Findings

19.1 ARIMA Superiority:

ARIMA worked very well for injury related categories (grievous, minor and non-injury situations) and demonstrated that auto regressive models are capable of addressing the high temporal dependencies found in injury patterns.

19.2 Random forest Excellence:

Random forest shows consistent good performance in forecasting tragic outcomes (Fatalities and fatal accidents) indicating that for complex, non-linear factors ensemble methods is more effectively capable to impact on severe accidents outcomes.

19.3 Exponential smoothing optimization:

Exponential smoothing perform well for both total number of accidents and the number of seriously injured showing that these metrics reflect smoothable trend patterns and best results for both.

20. Limitations and Future Research

However, this analysis provides insightful knowledge about predicting traffic accidents still several limitations should be noted.

1. By ignoring regional or demographic disaggregation level data, the study focuses on aggregate-level data.
2. External factors like weather, economic condition or policy changes were not clearly modeled.
3. Long-term structural changes in accident patterns may or may not be observed during evaluation period.

Future studies should explore hybrid modeling methods that utilize each and every advantage of specific accident categories by using ensemble techniques or conditional model selection based on real-time data parameters.

21. Conclusion

The evaluation and analysis done in this study predict that category-specific modeling methods are essential for road accident forecasting. Different performance showed by different model like Random Forest perform best for Fatal outcomes, ARIMA for injury categories and Exponential Smoothing for aggregate measures suggest that there should be system employed with multiple model for effective road safely forecasting instead of relying on single universal approach measures.

References

1. Sharma B, Katiyar VK, Kumar K. Traffic accident prediction model using support vector machines with Gaussian kernel. In: Proceedings of Fifth International Conference on Soft

- Computing for Problem Solving: SocProS 2015, Volume 2. Singapore: Springer; 2016. p. 1-10.
2. World Health Organization. Global status report on road safety. Available from: <https://www.who.int/publications/i/item/9789240086517>
 3. Ministry of Road Transport and Highways. Road Accident in India 2023. Available from: <https://morth.nic.in/sites/default/files/Road-Accident-in-India-2023-Publications.pdf>
 4. Swain PK, Tripathy MR, Agrawal K. Forecasting road accidental deaths in India using ARIMA and exponential smoothing models. J Stat Manag Syst. 2021;24(4):789-803. <https://doi.org/10.1080/09720510.2021.xxxxx>
 5. Patowary AN, Barman MP, Dutta B. Application of ARIMA model for forecasting road accident deaths in India. Int J Eng Res Technol. 2019;8(6):112-118.
 6. Sabenorio MA, Alqahtani S, Alharbi M, Hassan A. ARIMA model for forecasting traffic crash consequences in Saudi Arabia. Accid Anal Prev. 2022;170:106630. <https://doi.org/10.1016/j.aap.2022.106630>
 7. Agyemang F, Frempong J, Boateng EY. Time series forecasting of road accidents using SARIMA and Prophet models. Transp Res Procedia. 2023;69:512-520. <https://doi.org/10.1016/j.trpro.2023.01.067>
 8. Adewopo V, Elsayed N, Ozer M, Alsharif MH. A review on action recognition for accident detection in smart transportation systems. IEEE Access. 2022;10:114321-114345. <https://doi.org/10.1109/ACCESS.2022.xxxxx>
 9. Adewopo V, Elsayed N, Ozer M. AI on the road: Traffic accident detection in smart cities. Sensors. 2023;23(5):2451. <https://doi.org/10.3390/s23052451>
 10. Karamanlis I, Kokkalis A, Papageorgiou M, Vlahogianni EI. Deep learning-based black spot identification on Greek road networks. Transp Res C Emerg Technol. 2023;148:104015. <https://doi.org/10.1016/j.trc.2023.104015>
 11. Rifat MAK, Kabir A, Huq AS. An explainable machine learning approach to predict accident fatalities. IEEE Trans Intell Transp Syst. 2024;25(2):1780-1791. <https://doi.org/10.1109/TITS.2023.xxxxx>
 12. Wang S, Liu Y, Zhang L, Chen X. A review of road traffic accident prediction methods. J Adv Transp. 2023;2023:8845123. <https://doi.org/10.1155/2023/8845123>
 13. Islam ASMM, Shirazi M, Lord D. Grouped random parameters negative binomial–Lindley models for crash frequency analysis. Anal Methods Accid Res. 2023;39:100249. <https://doi.org/10.1016/j.amar.2023.100249>
 14. Savolainen PT, Mannering FL, Lord D, Quddus MA. The statistical analysis of highway crash-injury severities: A review and assessment of methodological alternatives. Accid Anal Prev. 2011;43(5):1666-1676. <https://doi.org/10.1016/j.aap.2011.03.025>
 15. Mannering F, Shankar V, Bhat CR. Unobserved heterogeneity and the theory of planned behavior in highway safety analysis: Predictive modeling versus causal inference. Transp Res A Policy Pract. 2020;132:350-363. <https://doi.org/10.1016/j.tra.2019.12.004>
 16. Mohan D, Tiwari G. Transport planning and traffic safety: Making cities, roads, and vehicles safer. Boca Raton: CRC Press; 2016.
 17. Peden M, Scurfield R, Sleet D, Mohan D, Hyder AA, Jarawan E, Mathers C. World report on road traffic injury prevention. Geneva: World Health Organization; 2004.
 18. Mohan R. Pachgaon Chowk accident blackspot analysis report. Traffic Police Department, India; 2025.
 19. Directorate of Economics & Statistics. Minor road accident trends in Thiruvananthapuram district. Government of Kerala; 2024.
 20. Waymo Research Team. Waymo vulnerable road users accident dataset: Technical report. Waymo LLC; 2024.

21. National Highway Traffic Safety Administration. Timing and characteristics of fatal crashes in the United States. U.S. Department of Transportation; 2024.
22. Bader Scott Injury Lawyers. Analysis of fatal crash timing and contributing factors in Texas. Bader Scott Research Division; 2024.
23. Lavrenz S, Park S, Naqvi SAH. SARIMA and hybrid deep learning models for traffic accident forecasting. *Expert Syst Appl.* 2022;198:116789. <https://doi.org/10.1016/j.eswa.2022.116789>
24. Breiman L. Random forests. *Mach Learn.* 2001;45(1):5-32.