Global Journal of Mathematics and Mathematical Sciences (GJMMS).
ISSN 0972-9836 Volume 15, Number 2 (2025), pp. 91-99

http://www.ripublication.com/gjmms.htm

Predictive Analysis of Traffic and Safety: Integrating Time Series and Deep Learning Techniques

Smita Verma, Monika Prajapat
Department of Applied Mathematics and Computational Science, SGSITS, Indore 452003, MP, India
Email: smitaverma.sgsita@gmail.com

Abstract

Road safety and traffic management have become major issues for cities worldwide as urbanization picks up speed. In order to promote traffic optimization and safety measures, Madhya Pradesh needs sophisticated forecasting technologies because of the state's rising traffic congestion and traffic accidents. This study forecasts and analyses traffic accident patterns by combining Long Short-Term Memory (LSTM) models with Seasonal Auto-Regressive Integrated Moving Average (SARIMA) models. The algorithms' capacity to forecast traffic patterns, spot high-risk times, and recommend road safety measures was assessed using historical accident data from 2001 to 2022. Policymakers and urban planners can use the integrated approach's statistics and deep learning capabilities to get useful information.The results highlight how useful hybrid techniques are for managing traffic and open the door for data-driven policy change.

Keywords: Long Short-Term Memory (LSTM), Seasonal Autoregressive Integrated Moving Average (SARIMA), Seasonal Autoregressive Integrated Moving Average-Long Short-Term Memory, Predictive Accuracy.

Introduction

On our highways, thousands of people are murdered and wounded every day. Road accidents can leave behind broken families and communities when men, women, or children are riding, bicycling, or walking to work or school, playing in the streets, or embarking on lengthy journeys. Millions of individuals are hospitalized for extended periods of time each year following serious accidents, and many of them are unable to live, work, or play as they once did. The pain that victims and their families endure from injuries sustained in automobile accidents is immeasurable. The World Health Organization (WHO) estimates that there are 1.35 million traffic accidents worldwide each year, gravely injuring 20 to 50 million people. It currently ranks as the eighth most common cause of death worldwide, and if present trends continue, it might rank

seventh by 2030[1].
15% of disability-adjusted life years (DALYs) and 10% of all deaths worldwide are caused by injuries. According to recent studies, injuries account for 13% to 18% of all fatalities in India. Unintentional injuries include road traffic injuries (RTIs). Road traffic fatalities are defined as "any person killed immediately or dying within 30 days as a result of an injury or accident," albeit this definition differs by country. RTIs account for a higher percentage of hospitalizations, fatalities, disabilities, and socioeconomic losses in young and middle-aged populations, making them the sixth most common cause of mortality in India, according to the WHO. The health sector is also heavily burdened by RTIs in terms of acute and prehospital care as well as rehabilitation.
India's road transport is the most cost-effective mode of transportation for both freight and passengers, but its high exposure to adverse traffic environments due to rapid motorization and urbanization, coupled with economic growth, has led to high incidents of road accidents, injuries, and fatalities.
Road accidents are a leading cause of death globally, primarily affecting the age group of 15 to 49. In 2022, road crashes in India claimed 1.68 lakh lives and injured over 4.4 lakh people [2].

Literature Review

2.1 Statistical Models

Mutangi studied traffic accidents in Zimbabwe in order to develop a predicting model for similar incidents in the future. Using a variety of techniques, the study evaluated the model's performance and came to the conclusion that ARIMA(0,1,0) correctly predicted the yearly rate of traffic accidents [3].

The Hubballi-Dharwad Bus Rapid Transit System (HDBRTS) used automatic fare collecting data to model short-term passenger demand using the ARIMA and SARIMA models. Based on MAPE, the study demonstrated that SARIMA performed better than ARIMA [4].

In order to forecast the frequency of accidents on multilane roads, Caliendo used the Poisson, Negative Binomial, and Negative Multinomial regression models, and he discovered that they were useful [5].

An enhanced seasonal autoregressive integrated moving average (ISARIMA) model and a multi-input autoregressive (AR) model using genetic algorithm (GA) optimization are combined in the hybrid prediction technique put forth by Luo, Niu, and Zhang (2018). According on test findings utilizing real traffic data provided by TDRL, the recommended approach performs better in terms of forecast accuracy than alternative approaches [6].

Statistical modeling methods have been widely used to forecast high-risk areas and examine accident trends. Assuming that accident occurrences follow a Poisson distribution, Poisson regression is a fundamental approach for modeling accident counts among the popular models. Negative binomial regression has been presented as a more adaptable substitute, enabling improved modeling of accident rates, in response to overdispersion in accident data, where the variance surpasses the mean (Lord & Mannering, 2018)[7].

2.2 Deep Learning Models

Since the output of CNN layers is regarded as input for LSTM, the CNN+LSTM combination has shown promise for time series forecasting. The hybrid neural network model has a high level of efficiency and can be used to predict the price of gold, stocks, energy use, crashes, and other things[8-9] .

By eliminating the vanishing gradient issue and cutting down on training time, Transformer, a more recent deep learning model designed to address the shortcomings of RNN, LSTM, and GRU, is gradually gaining traction in time series forecasting [10].

For short-term traffic flow prediction, Mou, Zhao, and Chen (2019) primarily focus on temporal information. The LSTM model's prediction ability is enhanced by the temporal information, which includes day-by-day data for traffic flow prediction [11].

Data and methodology

3.1 Data Collection

The dataset consist of traffic accident data of Madhya Pradesh from 2001 to 2022. This includes [12-15]:

2001 : 26,239 accidents, 2002 : 26,929 accidents, 2003 : 30,264 accidents, 2004 : 32,445 accidents, 2005 :35,123 accidents, 2006 : 38,041 accidents, 2007 : 41,981 accidents, 2008 : 43,852 accidents, 2009 : 47,267 accidents, 2010 : 50,023 accidents, 2011 : 49,406 accidents, 2012 : 51,210 accidents, 2013 :51,810 accidents, 2014 :53,472 accidents, 2015 :54,947 accidents, 2016 : 53,972 accidents, 2017 : 53,399 accidents, 2018 : 51,397 accidents, 2019 : 50,669 accidents, 2020 : 45,266 accidents, 2021 : 48,877 accidents, 2022 : 54,432 accidents.

3.2 Predictive Models

3.2.1 SARIMA Model

A time series forecasting technique that builds upon the ARIMA model by adding seasonal components is called the Seasonal Autoregressive Integrated Moving Average (SARIMA) model. The notation for the general SARIMA model is SARIMA (p,d,q) (P,D,Q)s.

Where:

- p,d,q : Non-seasonal parameters for auto regression (AR), differencing (I), and moving average (MA).
- P,D,Q : Seasonal counterparts of the ARIMA parameters.
- s : Length of the seasonal cycle.

  Where we take,

$$p = 1, d = 1, q = 1, P = 1, D = 1, Q = 1, s = 4$$

The SARIMA equation is expressed as:

$$\Phi_P(B^s)\phi_p(B)(1-B)^d(1-B^s)^d Y_t = \theta_Q(B^s)\theta_q(B)\epsilon_t$$

Where:

- B: Backward shift operator, $BY_t = Y_{t-1}$
- $\Phi_P(B^s)$: Seasonal AR polynomial.
- $\phi_p(B)$: Non-seasonal AR polynomial.
- $\theta_Q(B^s)$: Seasonal MA polynomial.
- $\theta_q(B)$: Non-seasonal MA polynomial.
- ϵ_t: White noise error term.

3.2.2 LSTM Model

Designed for sequential data, LSTM introducing memory cells and gates.
The LSTM cell is governed by the following equations:

- **Forget Gate**: It filters out the no relevant information and define as
$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f)$$
- **Input Gate**: It determines the impact of new information and written as
$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i)$$
$$\tilde{C}_t = tanh(W_C \cdot [h_{t-1}, x_t] + b_C)$$
- **Cell State Update**: It stores long term memory
$$C_t = f_t \cdot C_{t-1} + i_t \cdot \tilde{C}_t$$
- **Output Gate**: Generates the final prediction
$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o)$$
$$h_t = o_t \cdot tanh(C_t)$$

Where W, b andσdenote weight matrices, biases & the sigmoid activation function respectively.

3.2.3 Hybrid SARIMA-LSTM Models

In hybrid models, the linear seasonal prediction skills of SARIMA are combined with the real-time flexibility and non-linear trend handling capabilities of LSTM.

Workflow of SARIMA-LSTM Hybrid Model:

- **SARIMA for Seasonal Component**:

Use SARIMA to break down traffic data into residual and seasonal components. Seasonal trends are eliminated from the data by modelling them.

- **LSTM for Residual Component**:

Input residual data, which is non-linear, into an LSTM model.
Predict changes in the actual world, like weather or accidents.

- **Combining Predictions**:

The seasonal forecast from SARIMA and the residual prediction from LSTM are combined to get the final result.

The combined prediction can be expressed as:

$$\widetilde{Y}_t = \hat{Y}_{SARIMA,t} + \hat{Y}_{LSTM,t}$$

Where $\widetilde{Y}_t$= Final Traffic Prediction

Implementation and Results

4.1 Data Pre-processing

*Data Imputation***:** Missing values were filled using statistical techniques like interpolation or mean imputation to ensure completeness in the dataset.

Time-Series Decomposition: The dataset was decomposed into trend, seasonal, and residual components to identify underlying patterns and periodic fluctuations.

Feature Scaling: Normalization techniques such as Min-Max scaling were applied to ensure that different variables are within the same range, improving model efficiency and convergence.

4.2 Model Evaluation

Mean Absolute Error (MAE): Measures the average absolute difference between predicted and actual values. Lower MAE indicates higher accuracy.

Root Mean Square Error (RMSE): Evaluates prediction deviations by giving higher weight to larger errors. A lower RMSE implies a better-performing model.

Mean Absolute Percentage Error (MAPE): Expresses the prediction error as a percentage of actual values, making it useful for evaluating relative performance.

4.3 Results Comparison

The goal is to predict the amount of traffic incidents that will occur between 2019 and 2022 by utilizing historical accident data from 2001 to 2018. We employ three models for forecasting:

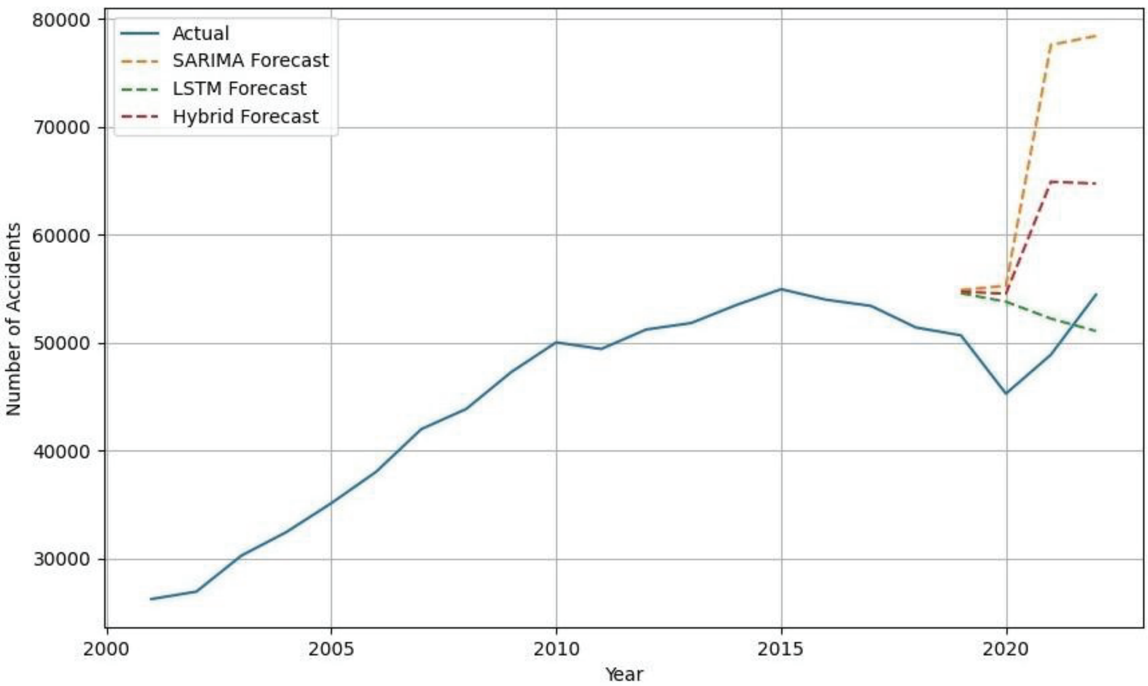

Figure 1. Road Accident Forecasting

4.4 Model Performance Comparison

A statistical model for forecasting time series is called SARIMA. A neural network model based on deep learning is called LSTM. Predictions from both models are averaged in hybrid SARIMA + LSTM. The accuracy of each model is then assessed using measures derived from real accident data from 2019 to 2022: MeanAbsolute Error (MAE), Root Mean Squared Error (RMSE),Mean Absolute Percentage Error (MAPE) are calculated:

Table 1

	MAE	RMSE	MAPE
SARIMA	16719.04	21532.06	35.004%
LSTM	52913.04	52930.51	65.74%
Hybrid SARIMA-LSTM	9910.95	12404.48	20.81%

Using data from 2001 to 2022, we use three modelling approaches—SARIMA, LSTM, and a hybrid SARIMA-LSTM model—to predict road traffic accidents in Madhya Pradesh from 2023 to 2026.

Table 2

Models	2023	2024	2025	2026
SARIMA	54,924	54,990	54,999	55,001
LSTM	52,262	50,928	50,113	49,617
Hybrid SARIMA-LSTM	53,707	53,322	53,148	53,101

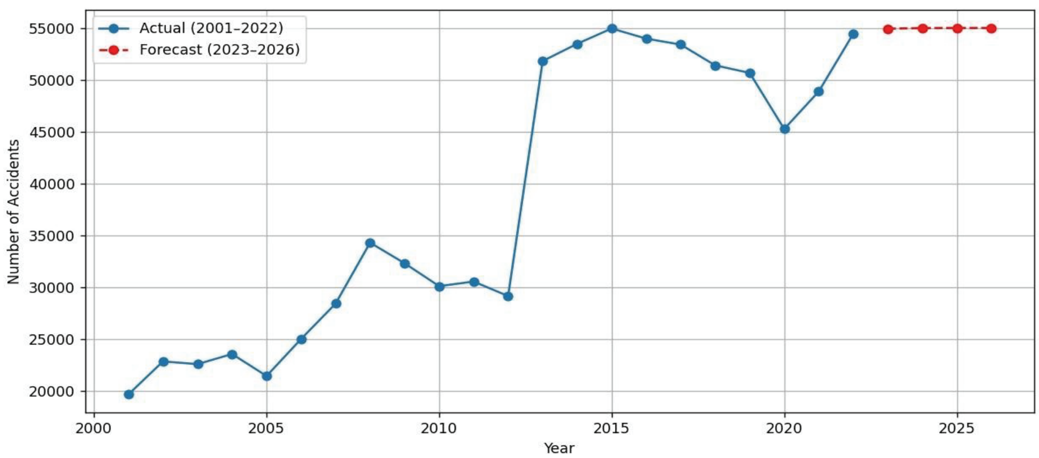

Figure 2:. Road accident Forecast (2023 – 2026) Using SARIMA

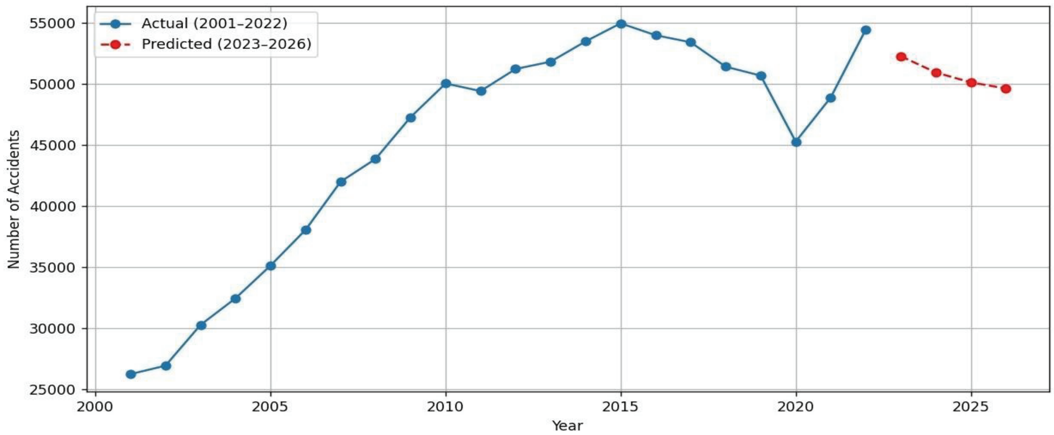

Figure 3: Road accident Forecast (2023 – 2026) Using LSTM

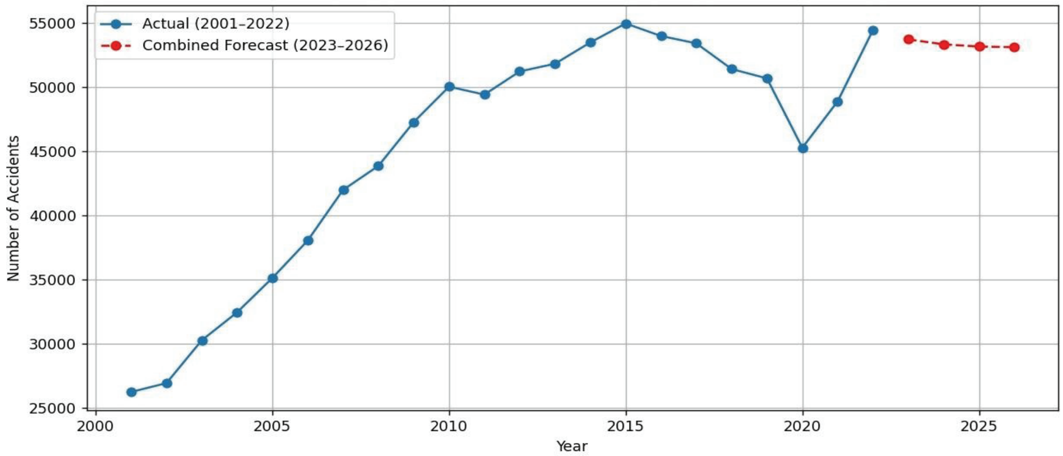

Figure 4: Hybrid Road accident Forecast (2023 – 2026) Using SARIMA and LSTM

Conclusion

This study used a combination of conventional statistical techniques and machine learning methodologies to examine the statistical modelling of road traffic accident (RTA) trends and the prediction of future accident hotspots.In order to anticipate traffic accidents, this study suggests a hybrid forecasting model that combines Long Short-Term Memory (LSTM) neural networks with Seasonal Autoregressive Integrated Moving Average (SARIMA). By identifying seasonal trends and temporal relationships in historical data, the hybrid technique improves accuracy and resilience. The technique entails gathering historical data on traffic accidents, preprocessing it, fitting SARIMA models, obtaining residuals, and using the residuals to train LSTM neural networks. Accurate forecasts and long-term relationships are captured by these models, which also capture complicated and nonlinear data components. Following training, the predictions from both models are combined to generate the hybrid SARIMA-LSTM model. Improved forecast accuracy results from this integration, which makes sure that both temporal and nonlinear trends are taken into account.

Future Work

The utilization of different time series models with weather and periodic features introduced as additional information will make the model more powerful in temporal feature extraction. Analyzing the impact of the additional information in normal regular days such as Weather, holidays, and events on the traffic flow could be our future work.

References

[1] [1]Ahmed, S., Hossain, M. A., Ray, S. K., Bhuiyan, M. M. I., &Sabuj, S. R. (2023). A study on road accident prediction and contributing factors using explainable machine learning models: analysis and performance. Transportation research interdisciplinary perspectives, 19, 100814.

[2] [2]Gururaj, G. (2008). Road traffic deaths, injuries and disabilities in India: current scenario. National Medical Journal of India, 21(1), 14.

[3] [3]Mutangi K. Time series analysis of road traffic accidents in Zimbabwe 2015.

[4] [4]Halyal S, Mulangi RH, Harsha MM. Short-term passenger demand modelling using automatic fare collection data: a case study of Hubli-Dharwad BRTS. AdvTransp Stud 2023;59.

[5] [5]Caliendo C, Guida M, Parisi A. A crash-prediction model for multilane roads. Accid Anal Prev 2007;39:657–70. https://doi.org/10.1016/j.aap.2006.10.012.

[6] [6]Luo, X., Niu, L., & Zhang, S. (2018). An algorithm for traffic flow prediction based on improved SARIMA and GA. KSCE Journal of Civil Engineering, 22, 4107-4115. doi: 10.1007/s12205-018-0429-4.

[7] [7]Lord, D., & Mannering, F. (2018). The statistical analysis of crash-frequency data: A review and assessment of methodological alternatives. Transportation Research Part A: Policy and Practice, 116, 20-36.

[8] [8]Livieris IE, Pintelas E, Pintelas P. A CNN–LSTM model for gold price time-series forecasting. Neural ComputAppl 2020;32:17351–60. https://doi.org/10.1007/s00521-020-04867-x.

[9] [9]Li P, Abdel-Aty M, Yuan J. Real-time crash risk prediction on arterials based on LSTM-CNN. Accid Anal Prev 2020;135:105371. https://doi.org/10.1016/j.aap.2019.105371.

[10] [10].Agga A, Abbou A, Labbadi M, Houm Y El, Ou Ali IH. CNN-LSTM: An efficient hybrid deep learning architecture for predicting short-term photovoltaic power production. Electr Power Syst Res 2022;208:107908. https://doi.org/10.1016/j.epsr.2022.107908.

[11] [11].Mou, L., Zhao, P., & Chen, Y. (2019). Short-term traffic flow prediction: A long short-term memory model enhanced by temporal information. In CICTP 2019 (pp. 2411-2422).doi:10.1061/9780784482292.210.

[12] [12].Road Accidents. Accessed:

[13] [13].Road Accidents in India 2013. Accessed:

[14] https://morth.nic.in/sites/default/files/Road_Accidents_in_India_2013%20(1).pdf

[15] [14].Road Accidents in India 2017. Accessed:
[16] https://morth.nic.in/sites/default/files/Road_Accidents_in_India_2017.pdf
[17] [15].Road Accidents in India 2022, Accessed:`
[18] https://morth.nic.in/sites/default/files/RA_2022_30_Oct.pdf .
[19] [16].Adeyeye, J. S., &Nkemnole, E. B. (2023). Predicting Malaria Incident Using Hybrid SARIMA-LSTM Model. International Journal Of Mathematical Sciences And Optimization: Theory And Applications, 9(1), 59-80