International Journal of Community Medicine and Public Health
Divyashree N et al. Int J Community Med Public Health. 2023 Dec;10(12):4862-4868
http://www.ijcmph.com
pISSN 2394-6032 | eISSN 2394-6040

DOI: https://dx.doi.org/10.18203/2394-6040.ijcmph20233791

Original Research Article

Analysis of road accident mortality based on time of occurrence for Kerala, India

Divyashree N.[1], Stavelin Abhinandithe K.[2]*, Vidya Raju[3], Sridhar Ramachandran[4], Prerana M.[1]

[1]Division of Medical Statistics, School of Life Sciences, Mysuru, Karnataka, India
[2]Department of Medical Statistics, School of Life Sciences, JSS AHER, Mysuru, Karnataka, India
[3]Department of Statistics, Yuvaraja's College, University of Mysore, Karnataka, India
[4]Ramakrishna Mission Vidyalaya, Coimbatore, Tamil Nadu, India

Received: 05 October 2023
Revised: 21 November 2023
Accepted: 22 November 2023

***Correspondence:**
Dr. Stavelin Abhinandithe K.,
E-mail: stavelin.ak@jssuni.edu.in

Copyright: © the author(s), publisher and licensee Medip Academy. This is an open-access article distributed under the terms of the Creative Commons Attribution Non-Commercial License, which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

ABSTRACT

Background: Road traffic accidents (RTA) pose a significant socio-economic burden and global public health concern. Monitoring road safety initiatives' efficacy necessitates analysing RTA incidence. This study examines time zone-specific RTA mortality in Kerala state, India, from 2016 to 2021.
Methods: Utilizing compiled secondary-level time series data, the study encompasses total RTA fatalities in Kerala from 2016 to 2021. Data includes fatalities per year in nine consecutive three-hour time periods. Exploratory data analysis, time series regression, and exponential smoothing were employed for analysis.
Results: Data reveals fluctuating trends in road accident (RA) fatalities, peaking in 2018 with a notable decrease in 2020. 18:00 to 21:00 recorded the highest and lowest fatalities, total 901 deaths. Disproportionate RA fatalities occurred from 06:00 to 09:00 (527 deaths) and 15:00 to 18:00 (697.5 deaths). The study employs Holt-Winters exponential smoothing for short-term forecasting, with a mean absolute scaled error (MASE) less than 1 signifying accurate predictions.
Conclusions: The analysis highlights temporal patterns, emphasizing 18:00 to 21:00 as critical. Holt-Winters exponential smoothing proves vital for accurate short-term forecasting, with MASE reflecting precision. Urgency is stressed in adopting targeted measures for time-specific road accidents. Government intervention is pivotal, advocating for improved infrastructure, enhanced driver education, efficient vehicle management, and sustained traffic enforcement. Tailoring traffic laws to time zones, coupled with forecasting techniques, aligns with the overarching goal of enhancing road safety and reducing RA mortality rates.

Keywords: Road accident, Time series regression, Prophet, Deep learning, Seasonality

INTRODUCTION

Road accident has always been a concern in India, America and the world at large. Road accidents account for more than 1.3 million deaths per year in the world. The number of vehicles purchased by people has increased annually due to the dollarization of the economy, which has made it easy for individuals to purchase second-hand vehicles mainly from other countries. Road traffic accidents were listed as one of the world's top ten causes of death. Due to the significant cost associated with them, which results in social and economic problems, road accidents have been identified as one of the unfavourable factors that contribute to the stifling of economic progress in developing nations. These significant costs connected with traffic accidents include human costs (e.g., desire to pay to prevent pain, sadness, and suffering), the direct economic costs of missed productivity, and the medical expenditures related to traffic accident injuries.

Increasing the amount of road traffic has benefited society in terms of accessibility and mobility, but it has consequences as well. These costs include not only the direct costs associated with providing transportation services, such as infrastructure, labour, and equipment costs, but also a number of indirect costs associated with the adverse effects on the environment, such as noise and air pollution, travel delays caused by traffic congestion, and the loss of life and property damage due to accidents on the road.[1]

In terms of traffic volume and economic impact on the country, road transportation is India's most prevalent means of transportation. The road transportation network in India has seen significant change over time due to high rates of population expansion, rising mobility, and an increase in the number of cars. The expansion of traffic accidents is a drawback of the upgrading of the road network (RTA). Every day, a large number of individuals are killed and injured in road accidents as a result of RTA, shattering families and communities and imposing significant socioeconomic costs on society. More than 1.55 lakh people died in road accidents in India in 2021, according to data recently provided by the National Crime Records Bureau (NCRB).[2]

The majority of passenger and freight traffic in a growing nation like India is transported via the road network. The main contributors of RTA injury are the rise in motorised traffic on highways with high rates of acceleration and speed. Pedestrians, cyclists, and motorcyclists are the most vulnerable road users and account for more than half of all traffic-related fatalities. The researchers forecast traffic accidents using a variety of methodologies. Here, in this study, we have used exponential smoothing method.[3]

Exponential smoothing

Exponential smoothing is a general technique for combining time series data using the exponential window function. In contrast to the basic moving average, which equally weights earlier data, exponential functions are used to apply weights that decrease exponentially over time. It is a simple procedure that can be learned and applied to decision-making based on user assumptions, including seasonality. Exponential smoothing is often used in time-series data analysis.[4]

Time series forecasting techniques using exponential smoothing fall into three categories. An extension that specifically addresses trends, a straightforward technique that makes no assumptions about systematic structure, and the most sophisticated strategy that incorporates seasonality support are all possible approaches.[5]

Smoothing by a single exponential

Single exponential smoothing, or SES, is a time series forecasting method for univariate data without a trend or seasonality.

Smoothing using two exponentials

Double exponential smoothing is an extension of exponential smoothing with the explicit addition of trend support in the univariate time series. Double exponential smoothing with an additive trend is commonly referred to as Holt's linear trend model, named for Charles Holt, the method's inventor.

Triple exponential smoothing

Triple exponential smoothing is an extension of exponential smoothing that explicitly supports seasonality in the univariate time series. This method is also known as Holt-Winters exponential smoothing in recognition of two individuals who made significant contributions to the technology, Peter Winters and Charles Holt. Similar to the trend, a change in the seasonality that is either linear or exponential can be described as an additive or multiplicative process.

Additive seasonality

Triple exponential smoothing with a linear seasonality.

Multiplicative seasonality

Triple exponential smoothing with an exponential seasonality.

Current effort intends to estimate the uncertainty associated with the incidence of total number of persons killed (NPK) due to RTA in Kerala during 2016–2021. The findings of this study may assist policymakers in creating and implementing further measures to reduce future traffic accidents. We also go through the data's primary attributes and its source. Also briefly outlines the statistical approaches taken to conduct this investigation.[6]

METHODS

The data for this study were compiled secondary level time series data that were published at the Kerala Police's official website (accessed on 2022) at https://old.keralapolice.gov.in/public-information/crime-statistics/road-accident. The data, consisting of the total number of people killed by RTA in Kerala, India, between 2016 and 2021. The information includes the total number of fatalities per year for nine consecutive three-hour time periods, including 06.00–9.00 (day), 09.00–12.00 (day), 12.00–15.00 (day), 18.00–21.00 (night), 21.00–24.00 (night), 00.00–3.00 (night), and 03.00–6.00 (night). The data contained records of unknown.

Data setup

Data was setup as a csv file and imported to R (version 4.2.3) where we had two columns, in which the first column consists of years from 2016-2021 and the second column consists of mortality numbers.

Converting raw data into a time series format

The data was converted into time series. There are two critical inputs which gives the function frequency and start timets= ts(data$Mortality number, start=2016,end=2021, frequency=11).

Decomposing in time series

Three components make up the time series: trend, seasonal and random. We have used plot (decompose(timets)).

Fitting with Holt-Winters

Following decomposition, data must be fitted for further prediction; in this case, we employed the Holt-Winters function, in which R determines the tuning parameters on its own. The three tuning parameters are alpha, beta, and gamma are considered. Alpha is the basic value, and a higher alpha value delays the most recent observations more. Higher beta values, which are considered trend values, indicate that the trend slope is more dependent on recent trend slopes, while higher gamma values, which are considered seasonal components, give the most recent seasonal components more weight. timets_ts=HoltWinters (timets,alpha=0.2,beta=0.1,gamma=0.1)timets_ts.

Forecast evaluation

Further forecast calculates the quality of prediction by taking the differences between observed values and the predicted values for each data points which are added as residuals to a forecast model. To evaluate the smoothing functions, correlation was checked between forecast errors in order to achieve this we use the ACF function to evaluate the correlation of fit-residuals between points of various temporal separations in the time series (lags).[7]

```
# install.packages("devtools")
# devtools::install_github("RamiKrispin/TSstudio")
#setwd("C:/Users/hp/Desktop/stav")
library(tidyverse)
```

```
## ── Attaching core tidyverse packages ──────────────────── tidyverse 2.0.0 ──
## ✓ dplyr     1.1.1     ✓ readr     2.1.4
## ✓ forcats   1.0.0     ✓ stringr   1.5.0
## ✓ ggplot2   3.4.2     ✓ tibble    3.2.1
## ✓ lubridate 1.9.2     ✓ tidyr     1.3.0
## ✓ purrr     1.0.1
## ── Conflicts ────────────────────────────────────── tidyverse_conflicts() ──
## ✕ dplyr::filter() masks stats::filter()
## ✕ dplyr::lag()    masks stats::lag()
## ℹ Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors
```

```
library(plotly)
```

```
##
## Attaching package: 'plotly'
##
## The following object is masked from 'package:ggplot2':
##
##     last_plot
##
## The following object is masked from 'package:stats':
##
##     filter
##
## The following object is masked from 'package:graphics':
##
##     layout
```

```
library(TSstudio)
library(readr)
library(timetk)
library(janitor)
```

Figure 1: Forecast evaluation.

HW1.pred=predict(HW1,48,prediction.interval=TRUE,level=0.95)
plot(timets,xlim=c(2016,2025))
lines(HW1$fitted[,1],lty=2,col="blue")
lines(HW1.pred[,1],col="orange")
lines(HW1.pred[,2],lty=2,col="red")
lines(HW1.pred[,3],lty=2,col="orange")

When p value is more than 0.05, the lung-Box test, which detects correlations, is used to determine whether the residuals are independent. The residuals' histogram was constructed to confirm that they have a normal distribution. Our model will continuously overshoot in one direction if the residuals are highly skewed.

library(forecast)
HW1_for <- forecast(HW1, h=24, level=c(80,95))
plot(HW1_for, xlim=c(2008.5, 2020))
lines(HW1_for$fitted, lty=2, col="purple")
acf(HW1_for$residuals, lag.max=20, na.action=na.pass)
Box.test(HW1_for$residuals, lag=20, type="Ljung-Box")
hist(HW1_for$residuals)

We have used MASE values to evaluate the forecast error as forecasting metric.

Time series regression analysis

Time series method was used to analyse a set of data points gathered over a period of time. Instead of just capturing the data points intermittently or arbitrarily, time series analysis records the data points at regular intervals over a predetermined length of time. Depending on the specific application industry, it may be used in a wide range of ways once it has been developed. Exponential smoothing techniques, Box-Jenkins methodology, and time series regression (TSR) analysis are a few of the more well-liked ones. An observed time series model may have an additive decomposition. The decomposition in the time series means the 3 main components that make-up the time series: trend, seasonal and random. These 3 components are extracted and plotted with the command.

plot(decompose(timets))

Model diagnostics

The L-jung-Box test can be used to confirm the residuals' white noise assumption after fitting the times series model along with the sample autocorrelation function (ACF) plot and normal density histogram plot.

Regression analysis with time series

In order to conduct regression analysis for time series data, we have used R and installed package devtools along with tidyverse library. We have plotted the graph using ggplot2. In order to convert the data to a time series object, prophet library was imported from Facebook. Seasonality were checked and further prediction and forecasting was done for different time zones.

RESULTS

The key findings of the study employing the approaches described are presented in this section. Using R software, all statistical analyses were carried out on a two-sided basis with a 5% level of significance.

Exploratory data analysis

Time plot for the NPK time-series data from 2016-2021 is displayed in Figure 2. The time plot supports the seasonal volatility in the dataset. According to NPK, the number of fatalities from road accidents increased from 2017-2018, with the lowest number being reported in 2020. The biggest and lowest numbers of RA fatalities occurred between 18:00 and 21:00 (901 deaths) and at an unknown period (7 deaths). Table 1 displays the data's summary statistics. The statistics of the data for nine different time zones are shown in summary form in Table 2. The overall number of NPKs in each time zone has increased over time, as shown in Table 2.[8]

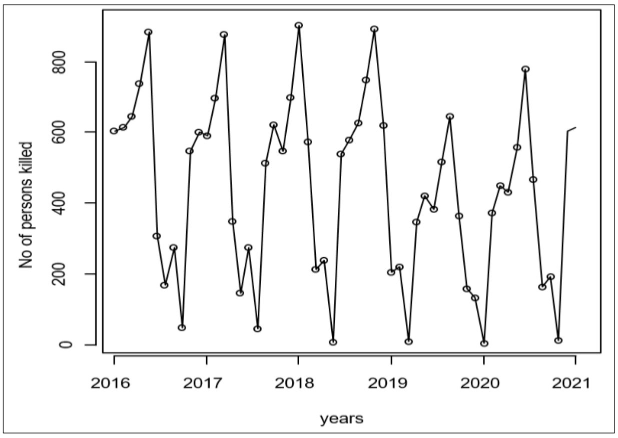

Figure 2: Time plot of NPK data.

Table 1: Summary statistics of NPK due to road accidents in Kerala during 2016-2021.

N	Mean	SD	P25	Median	P75	Min	Max
54	436.75	254.23	212.5	459	620.25	7	901

Table 2: Summary statistics of NPK regarding each time zone through 2016-2021.

Time-zone	Mean	SD	Median	P25	P75	Min	Max
06.00-9.00 hours (day)	487.6667	103.9782	527	366.2500	562.2500	346	605
09.00-12.00 hours (day)	547.8333	87.87358	590	443.7500	615.0000	422	621
12.00-15.00 hours (day)	537.1667	107.5945	569.5	418.7500	630.75	382	645
15.00-18.00 hours (day)	658.8333	97.34766	697.5	547.0000	739.0000	517	748
18.00-21.00 hours (night)	828.8333	100.6964	880	744.0000	892.7500	645	901
21.00-24.00 hours (night)	447.3333	127.9057	415.5	339.7500	585.5000	309	620
00.00-3.00 hours (night)	177.3333	26.83778	168.5	157.2500	207.5000	146	215
03.00-6.00 hours (night)	223.1667	53.3907	230	179.2500	275.0000	135	275
Unknown	22.66667	19.73491	12	8.5000	47.5000	7	49

A minimal number of RA-related fatalities happened within an unidentified time period. The greatest number of fatalities (880 (744,892.75)) occurred between 18 and 21. In comparison to other time zones, Kerala experiences a relatively high rate of RA fatalities during night and throughout the evening hours. The number of RA fatalities that occurred between the hours of 06:00 and 9.00 (day) (527(366.25, 562.25)) and between the hours of 15:00 and 18:00 (day) (697.5(547, 739)) was likewise disproportionately high. The seasonal plots of the NPK owing to RTA in Kerala for various time zones from 2016 to 2021 are shown in Figure 3. Additionally, it shows that the NPK demonstrates a declining character during the hours of 06.00-9.00 (Day) and 09.00-12.00 (Day) in comparison to 2021. Additionally, the time zones of 12 to 15 hours (Day), 15 to 18 hours (Day), 18 to 21 hours (Night), 21 to 24 hours (Night), 0.00 to 3 hours (Night), and 03 to 6 hours (Night) are unknown. The NPK exhibits a rise at an unidentified time zone among these.[8]

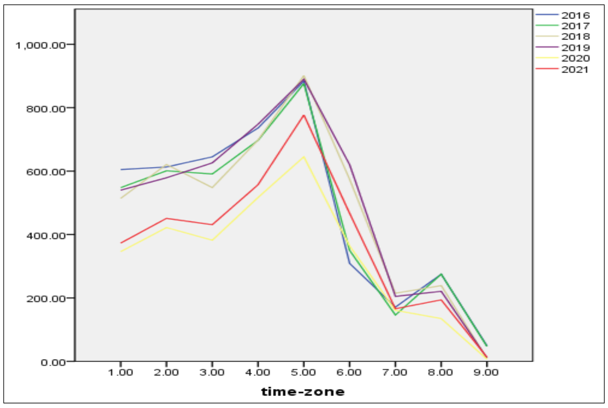

Figure 3: NPK due to RA in Kerala at different time zone during 2016-2021.

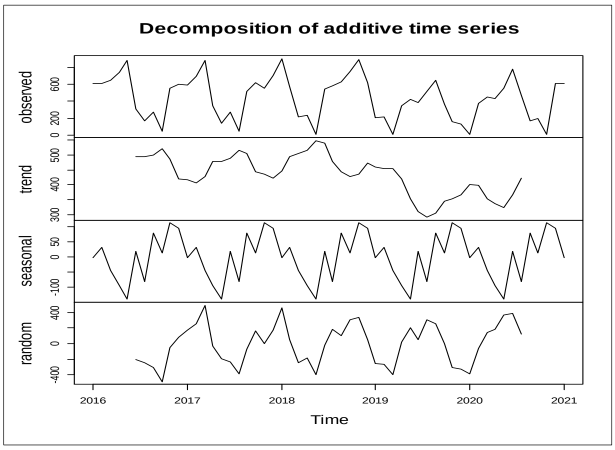

Figure 4: Decomposition time series plots of the time series.

We can see and understand how to sum up our observed values when we examine the decomposition components because it's crucial to examine the values of each part to determine which one is more prominent. A time series is decomposed by breaking it down into its constituent parts, such as its trend, irregular component, and, if it's a seasonal time series, its seasonal component.

Time series regression

It was further forecasted using the regression plot where regression plots show the variation in the time points.

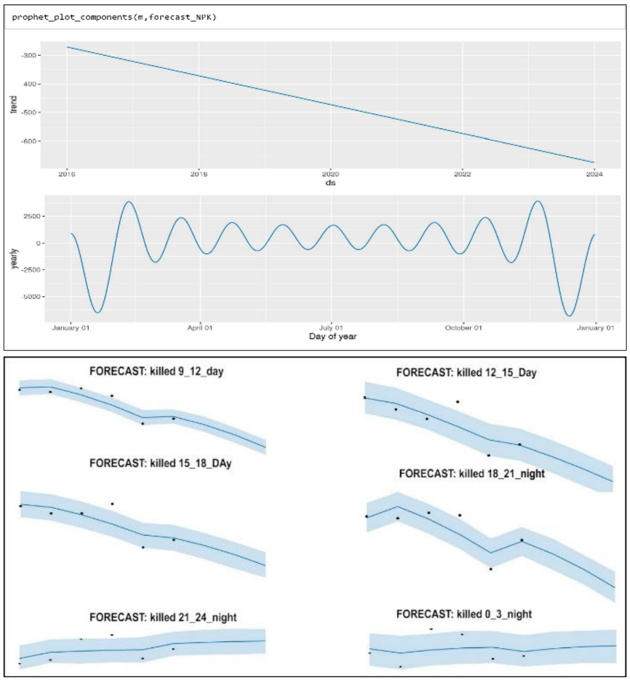

Figure 5: Forecasted values for different time zones.

Holt-Winters exponential smoothing

Holt-Winters exponential smoothing alters the Holt-exponential smoothing technique, so that it can be used in the presence of both seasonality and trend.

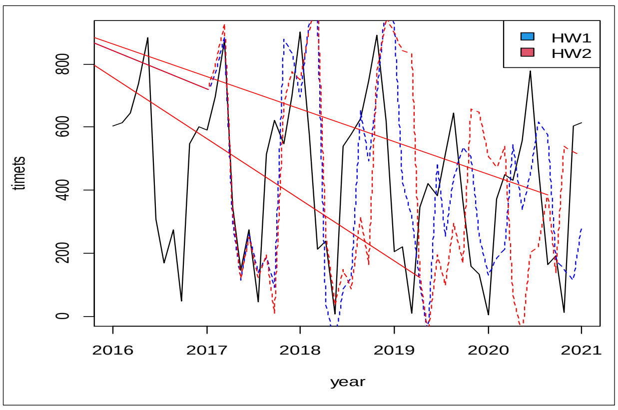

Figure 6: Holt-Winters fit.

The forecast and the actual data closely match each other in terms of the model's seasonal pattern and growing trend near the conclusion of the data (Figure 6). Our original time series covers the sum of squared errors for the in-sample forecast errors for a specific time period, which is a measure of prediction accuracy. With 10 degrees of freedom and a p value of 2.2e-16, the square sum in this case is 117.28.

After creating a fit data Holt-Winters function in R it will figure it out parameters by its own. By manually adjusting tuning factors like alpha, beta, and gamma, we may additionally fine-tune the fit. Here, we created two fits and plotted them in relation to the available raw data and the fit quality. To see the values how they change.

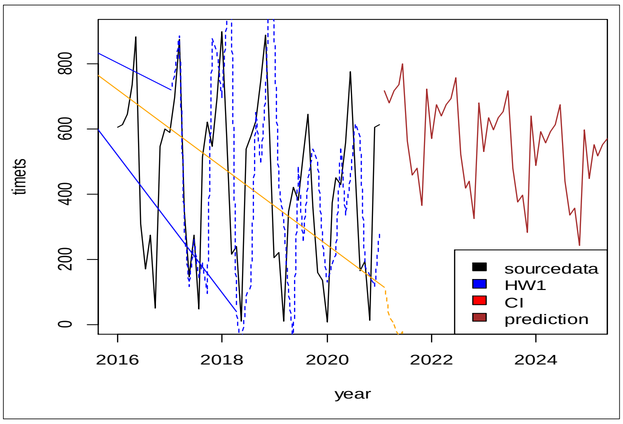

Figure 7: Predictions for HW1 seasonally-additive fit.

Plotting is simplified and the prediction comes with a 95% confidence interval when using time series forecasting with a forecast wrapper that supports greater confidence intervals. We looked for any possible association between forecasting mistakes. The L-Jung Box is also used to show the presence of correlation at 95% confidence intervals in time series to determine the correlation of fit residuals between distinct temporal separations. Additionally, make sure that the ACF bars are inside the blue area. To establish normality, residual histogram plot is employed.[10]

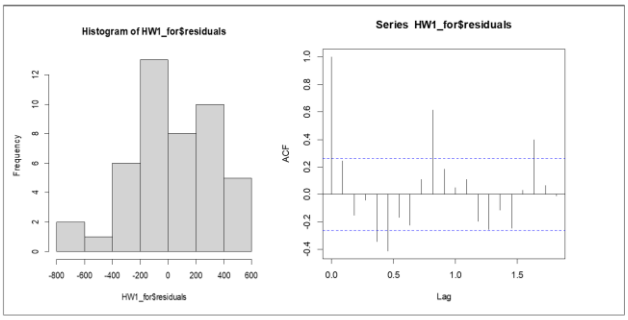

Figure 8: Residual application plots.

The R "acf()" function was used to create a correlogram for lag 10. Although it appears to be somewhat tilted left in comparison to a normal curve, the histogram in Figure 8 depicts the distribution of forecast errors and is roughly centred on zero.

DISCUSSION

Around the world, road accidents are a major source of injuries, hospitalisation, illnesses, and deaths, with significant socioeconomic costs to society. According to the information that is currently available, RTA deaths are a substantial public health concern in India and are largely preventable. Due to the vast, dense population and significant reliance on on-road traffic for the stability of its economy, a growing nation like India, an analytical study on the incidence of RTA is always a relevant research issue with social importance. In this paper, we have analysed the number of fatalities caused by RTA in Kerala State, India, from 2016 to 2021 across various time zones. We have used exploratory data analysis and time series regression models for predicting and forecasting the total cases reported of accidents and the number of deaths through this method.

The Holt-winters forecasting approach was utilised in addition to the exploratory data analysis to analyse the data. A strong method for predicting future data in a time series is Holt-Winters forecasting. In this study, the NPK was predicted for several particular time zones throughout the year. The study is hampered by a lack of data on the numerous RTA-related parameters in each daytime time zone. A greater understanding of the situation might be obtained by examining the link between the NPK in various time zones and associated risk variables, such as driving habits, car attributes, traffic conditions, and geographic and environmental effects.

In the current work, an additive time series model with growing or decreasing trend and seasonality was described utilising exploratory data analysis and Holts Exponential Smoothing. Short-term forecasts are made using Holt's exponential Smoothing, which estimates the level and slope at the current time point. The smoothing is governed by two parameters: Alpha for the level estimate at the current time point and beta for the slope estimate of the trend component at the current time point. Since MASE is less than 1, it indicates that predicting is more accurate for the current set of data. MASE is a measure of prediction accuracy proposed by (Koehler and Hyndman, 2006). To estimate the burden of deaths and non-fatal injuries brought on by traffic collisions, it is important to monitor and anticipate road traffic accidents. The study's findings support the use of forecasting techniques, which are also utilised to develop and track the implementation of road safety regulations pertaining to certain time zones for reducing RTA fatality rates. Therefore, the authorities must execute all essential measures to increase road safety, including safer roads, adequate driver training, efficient vehicle management, and ongoing traffic enforcement.

Limitations

The study covers data only for the years 2016 to 2021. This relatively short time frame might not capture long-term trends or patterns in road accidents.

The study focuses exclusively on road accidents in Kerala, India. Findings may not be generalizable to other regions with different socio-economic, cultural, or infrastructural characteristics.

The data is categorized into nine three-hour time periods. While this granularity provides detailed information, it might oversimplify the complexities of road accidents, and certain patterns may be obscured.

The dataset appears to provide information only on the total number of fatalities and the time of day. Additional variables such as weather conditions, road conditions, driver behaviour, and vehicle types where not considered.

The data relies on official records, which may be subject to underreporting due to factors such as lack of reporting, misclassification, or administrative issues.

CONCLUSION

The analysis underscores the significance of temporal patterns, particularly emphasizing the critical timeframe from 18:00 to 21:00. The application of Holt-Winters exponential smoothing emerges as vital for achieving accurate short-term forecasting, as evidenced by the favourable mean absolute scaled error (MASE) reflecting precision in predictions. The urgency of implementing targeted measures during these time-specific periods to address road accidents is paramount. Government intervention plays a pivotal role in this endeavour, necessitating improvements in infrastructure, enhanced driver education, efficient vehicle management, and sustained traffic enforcement.

The recommendation to tailor traffic laws to specific time zones, complemented by the use of forecasting techniques, aligns seamlessly with the overarching goal of enhancing road safety and ultimately reducing road accident mortality rates. As stakeholders collaborate to implement these measures, there exists a tangible opportunity to create a safer road environment, emphasizing the importance of proactive and time-sensitive interventions to mitigate the impact of road accidents.

And also, to determine the burden of road accident injuries brought on by traffic crashes, it is essential to track and forecast these incidents. The research's conclusions emphasize the need to monitor and put into practice time zone-specific safety laws in order to reduce the number of RTA fatalities. Among the preventive strategies are tightening traffic rules, expanding post-crash care, and upgrading the safety of the roads and motorists. Even though the issue and its solutions are more understood, political resolve to take the required steps is crucial.

Funding: No funding sources
Conflict of interest: None declared
Ethical approval: The study was approved by the Institutional Ethics Committee

REFERENCES

1. Nazari SS, Moradi A, Rahmani K. A systematic review of the effect of various interventions on reducing fatigue and sleepiness while driving. Chinese J traumatol. 2017;20(5):249-58.
2. Sunny CM, Nithya S, Sinshi KS, Vinodini V, Kg AL, Anjana S, Manojkumar TK. Forecasting of road accident in Kerala: A case study. In: 2018 International Conference on Data Science and Engineering (ICDSE). IEEE. 2018;1-5.
3. Jayan KD, Ganeshkumar B. Identification of accident hot spots: AGIS based implementation for Kannur District, Kerala. Int J Geomat Geosci. 2010;1(1):51-9.
4. Mohan D. Road accidents in India. IATSS Research. 2009;33(1):75.
5. Behura A, Behura A. Road accident prediction and feature analysis by using deep learning. In2020 11th International Conference on Computing, Communication and Networking Technologies (ICCCNT). IEEE. 2020;1-7.
6. Sumit K, Ross V, Ruiter RA, Polders E, Wets G, Brijs K. An exploration of characteristics and time series forecast of fatal road crashes in Manipal, India. Sustainability. 2022;14(5):2851.
7. Abdulhafedh A. Road crash prediction models: different statistical modeling approaches. J Transport Technol. 2017;7(02):190.
8. Vipin N, Rahul T. Road traffic accident mortality analysis based on time of occurrence: Evidence from Kerala, India. Clin Epidemiol Global Health. 2021;11:100745.
9. Deretić N, Stanimirović D, Awadh MA, Vujanović N, Djukić A. SARIMA modelling approach for forecasting of traffic accidents. Sustainability. 2022;14(8):4403.
10. Milton JC, Shankar VN, Mannering FL. Highway accident severities and the mixed logit model: an exploratory empirical analysis. Accident Analysis & Prevention. 2008;40(1):260-6.

Cite this article as: Divyashree N, Abhinandithe SK, Raju V, Ramachandran S, Prerana M. Analysis of road accident mortality based on time of occurrence for Kerala, India. Int J Community Med Public Health 2023;10:4862-8.