TUAD, 2025, 8(1), 47–57

Trafik ve Ulaşım Araştırmaları Dergisi

Dergi Ana Sayfa: http://dergipark.org.tr/tuad

Safety Research Unit

Araştırma Makalesi | Research Article

Analysis Of Traffic Accidents Using Machine Learning Under Pandemic Conditions

Ülviye Gülsüm Haşıloğlu Aras*

University of Osmaniye Korkut Ata, Faculty of Engineering, Department of Civil Engineering, 80010, Osmaniye, Türkiye.

Abstract

The COVID-19 pandemic that emerged in 2019 affected all aspects of life, including spiritual, psychological, social, economic, health and transportation aspects. Despite its negative consequences, however, the COVID-19 pandemic also produced some positive results. This study investigated the effect of COVID-19 lockdowns on killed-and-injured traffic accidents in metropolitan cities and Zonguldak Province in Turkey from 2012–2019 using the Extreme Gradient Boost (XGBoost) algorithm. Nonlinear regression analyses were performed using machine learning in Python programming language on the Google Colab platform. The analysis provided an estimated number of accidents for 2020, which was compared with the real killed-and-injured accidents data from metropolitan cities and in Zonguldak in 2020. The comparison showed that COVID-19 lockdowns caused a decrease in traffic accidents in metropolitan cities and Zonguldak Province, except in Diyarbakır and Ordu. It has been revealed that the number of traffic accidents predicted by machine learning algorithms in metropolitan areas for 2020 is 18.3% higher than the number of traffic accidents in 2020. Therefore, although accurate predictions can be made with machine learning, it has been observed that there may be a margin of error in extraordinary situations such as earthquakes, wars and pandemics.

Keywords: Lockdown, traffic accidents, accident prediction model, machine learning, extreme gradient boost

Makine Öğrenmesi Algoritmalarını Kullanılarak Pandemi Şartları Altında Trafik Kazalarının Analizi

Öz

2019 yılında ortaya çıkan COVID-19 pandemisi, ruhsal, psikolojik, sosyal, ekonomik, sağlık ve ulaşım olmak üzere hayatın tüm alanlarına etki ettiği gözlemlenmiştir. COVID-19 ‘dan kaynaklı pandemi süreci birçok olumsuz sonuçlarına rağmen, bazı olumlu sonuçlar da doğurmuştur. Bu çalışmada, pandemi sürecinin olumlu sonuçlarından biri detaylı olarak ele alınmıştır. Çalışma, Türkiye’de Zonguldak ili de dahil olmak üzere tüm büyükşehirleri kapsamaktadır. Bu çalışma ile, COVID-19 karantinalarının 2012-2019 yılları arasında gerçekleşen ölümlü ve yaralanmalı trafik kazaları üzerindeki etkisi, Extreme Gradient Boost (XGBoost) algoritması kullanılarak detaylı olarak araştırılmıştır. Google Colab platformunda Python programlama dilinde makine öğrenmesi kullanılarak doğrusal olmayan regresyon analizleri yapılmıştır. Analiz sonucunda 2020 yılı için tahmini kaza sayısı elde edilmiş ve bu sayı 2020 yılında büyükşehirlerde ve Zonguldak'ta meydana gelen gerçek ölümlü ve yaralanmalı kaza verileri ile karşılaştırılmıştır. Bu analiz, COVID-19 karantinalarının Zonguldak, Diyarbakır ve Ordu illeri hariç tüm büyükşehirlerde trafik kazalarında azalmaya neden olduğunu göstermiştir. Makine öğrenimi algoritmaları ile 2020 yılı için büyükşehirlerde tahmin edilen trafik kaza sayılarının, 2020 yılında gerçekleşen trafik kaza sayılarına göre %18,3 oranında daha yüksek olduğu ortaya çıkmıştır. Dolayısıyla Makine öğrenmesi ile doğru tahminler yapılabilse de deprem, savaş ve pandemi gibi olağanüstü durumlarda hata payı olabileceği gözlemlenmiştir.

Anahtar Kelimeler: Karantina, trafik kazaları, kaza tahmin modeli, makine öğrenmesi, extreme gradient boost

* *İletişim / Contact: Ülviye Gülsüm HAŞILOGLU ARAS, Faculty of Engineering, Department of Civil Engineering, Osmaniye, Türkiye., 80010; e-posta:* uhasilogluaras@osmaniye.edu.tr

Gönderildiği tarihi / Date submitted: 30.10.2024, Kabul edildiği tarih / Date accepted: 28.03.2025

Alıntı / Citation: Haşıloğlu Aras, Ü., G. (20205). Analysis of accidents using machine learning under pandemic conditions. *Trafik ve Ulaşım Araştırmaları Dergisi, 8*(1), 47–57. https://doi.org/10.38002/tuad.1572607

Analysis of Traffic Accidents Using Machine Learning Under Pandemic Conditions

1. Introduction

Worldwide innovations and developments directly affect freight and passenger mobility. Increasing population, developing technology and increasing trade and communication opportunities have augmented the importance of transportation activities each day (Gökdağ et al., 2004; Jovic Vranes et al., 2018). Competition in the world markets has accelerated the development of transportation activities and companies that want to increase their market share need to provide increasingly more transportation services (Bayraktutan & Özbilgin, 2013).

While the population in Turkey increased at an annual average rate of 1.4% between 2012 and 2019, it increased by only 0.6% in 2020 compared to 2019. This rate of increase was the lowest annual rate of the last 60 years (World Health Organization [WHO], 2019)). However, the annual average increase in the number of motor vehicles has been 4.4% since 2012. The increasing population, number of registered vehicles and people with driver's licenses and speed and competition of development in the transportation sector, as well as the trend toward a more mobile lifestyle, have increased traffic density greatly. Thus, the safety of traffic has also been significantly impacted. According to the WHO, one person dies every 24 seconds on the road worldwide. Traffic accidents, which are the eighth leading cause of death (Selimoğlu, 2014; WHO, 2019), cause annually the deaths of 1.35 million worldwide, and between 20–50 million people experience non-fatal injuries.

The COVID-19 pandemic that emerged in 2019 and was announced on 11 March 2020 in Turkey (WHO, 2019) affected the whole world in all aspects of life, including material, spiritual, psychological, social, economic, health and transportation aspects. In Turkey, all schools, including universities, began conducting education on online platforms as of 16 March. Public local and intercity transportation began operating at 50% capacity on 24 March, and international flights were banned on 27 March. The first home lockdown period extended from 11–12 April 2020 for 30 metropolitan cities and Zonguldak for the weekend. The lockdown continued for the same cities from 18–19 April 2020. As of 23–26 April 2020, the home lockdown was applied throughout Turkey. Subsequently, a home lockdown was applied to all of Turkey again on 16–26 May 2020.

As a result of the pandemic, traffic mobility has changed; people's trust in public transportation has decreased, and the number of private vehicles has increased (Korkmaz, 2022). Additionally, the compulsory demand for online shopping due to the pandemic created a surge in the number of cargo and transportation vehicles in traffic. However, the lockdown measures applied to control the COVID-19 pandemic also had positive results (Aloi et al., 2020; Brodeur et al., 2021; Shilling & Waetjen, 2020), such as reduced traffic accidents (Christey et al., 2020; Nuñez et al., 2020; Oguzoglu, 2020; Saladié et al., 2020a). While 283,234 people were injured and 5,473 people lost their lives as a result of traffic accidents in Turkey in 2019, the number of people injured in traffic accidents decreased to 228,565 and the number of people who lost their lives decreased to 2,197 in 2020, when the pandemic bans started (Emniyet Genel Müdürlüğü [EGM], 2020; Türkiye İstatistik Kurumu [TUİK], 2020).

The interest in using statistical models to understand the traffic accidents that occur and to take the necessary precautions has grown, and many traffic accident prediction models using various methods have been described in the literature. These studies were carried out using statistical methods, and the speed of these studies has increased due to developments in the field of statistics and in computer programmes. Recently, these studies have used machine learning to make predictions about traffic accidents (Ma et al., 2019; Yavuz et al., 2021).

Machine learning is a category of algorithms that increase software programmes' ability to predict results accurately without explicit programming. In machine learning, algorithms can take input data and use statistical analysis to predict an output while updating the output as new data emerge (Envepo A.Ş., 2018). For example, Yavuz et al. (2021) made a classification of traffic accidents with fatalities and injuries in Antalya Province and its districts between 2012 and 2016 using the Naive Bayes (NB) machine learning method. Ma et al. (2019) carried out a spatial grid analysis of traffic accidents in Los Angeles with the Extreme Gradient Boost (XGBoost) algorithm and Geographical Information System (GIS), which are machine-learning methods. (Chong et al., 2005) proposed a new model using a combination of artificial neural networks (ANN) and regression trees

(RT) methods to estimate the degree of injury resulting from traffic accidents. Using the decision tree (DT) model, Chang and Wang (2006) concluded that the vehicle type is the most important variable affecting accident severity. Sohn and Shin (2001) determined the variables affecting the severity of traffic accidents in Korea using ANN, logistic regression analysis (LRA) and DT methods. Kwon et al. (2015) ranked the variables that cause traffic accidents according to their relative importance using NB and DT methods. Muhammed et al. (2017) tried to predict traffic accidents on highways using DT algorithms.

This study aims to conduct data-driven research using machine learning algorithms to analyse the impact of pandemic conditions on traffic accidents. The effects of changing traffic density and COVID-19 restriction policies on traffic accidents during the pandemic will be evaluated. In the study, the answers to the following main research questions were investigated. What are the differences between the number of traffic accidents before and during the pandemic? What are the most appropriate machine learning algorithms for analysing and predicting traffic accidents?

Therefore, within the scope of the hypotheses of the study, firstly, the effects of changes in traffic density on the number and type of traffic accidents during the pandemic period were investigated. Accordingly, the success of machine learning models in numerically predicting traffic accidents occurring during the pandemic period was examined to analyse the changes.

It is not known how the curfews imposed in Zonguldak province, and 30 metropolitan cities affect the number of traffic accidents. For this purpose, the number of fatal and injury traffic accidents occurring in Turkey between 2012-2019 was used and a prediction table was created with Python programming language on the Google Colab platform. Nonlinear regression analyses were performed using the XGBoost algorithm, which is a controlled machine learning algorithm in Python programming language. Tableau 2020.4 Desktop software was used to visualise the result values. The estimated number of accidents in 2020 was found and compared with the actual fatal and injury accident data from metropolitan cities and Zonguldak in 2020. Thus, the effect of COVID-19-induced curfews on the number of fatal and injury traffic accidents in 2020 was observed.

Using data from previous years, an answer to the hypothesis ‘How successful will machine learning models be in predicting the number of traffic accidents that may occur in the coming years?’ was sought. In addition, it has been investigated how changes in traffic density during the pandemic affect the number and type of traffic accidents.

2. Method

A retrospective number of fatal and injured accidents, number of people killed and injured, total number of vehicles and the population data for provinces in Turkey from 2012–2019 were taken from the reports of the Turkish Statistical Institute (TUİK, 2020). The data for 2020 were obtained from the annual report published by the Police Department (EGM, 2020).

2.1. Machine Learning

Statistical analysis in the study was carried out with machine learning, a sub-branch of artificial intelligence. With machine learning, a system that creates predictions by making inferences from the data using mathematical and statistical operations is created. Many different machine learning methods have emerged for this inference process. If the output of the method is categorical, then classification is performed. If the output is numerical, regression analysis is performed. Since the output of the present study was numerical, regression analysis was performed.

Machine learning has two different learning situations: unsupervised and supervised. The main purpose of unsupervised learning is training the system with algorithms to create a model or pattern from the available data. In unsupervised learning, examples of events and situations are known in advance, but the corresponding results (i.e., classifications) are not known in advance. In contrast, supervised learning predicts the result for future situations by using examples of past events whose inputs and outputs are discovered upon analysis. In the present study, previous data were investigated with multivariate regression analysis (the Supervised Machine Learning Method) using Python programming language on the Google Colab platform.

In supervised learning, regression analysis and linear and nonlinear algorithm types are available. In the current study, many linear and nonlinear regression analyses were performed to obtain the best estimation result. For the LRA, the following were applied: simple linear regression, ridge regression, lasso regression, elastic net regression, Bayesian linear regression and support vector regression (SVR) algorithms. For the non-linear regression analyses, gradient boosting regression, XGBoost regression, K-nearest neighbours' regression, AdaBoost regression, CatBoost regression and Light GBM regression algorithms were used as comparison metrics. The XGBoost algorithm provided the most significant results and is explained in detail below.

2.2. Python Libraries

In the Python programming language, libraries store the files belonging to the most-used code pieces. These libraries are stored in pre-written files that can be used repeatedly. The current study utilized Numpy, Pandas, MatPlotLib, SciKit-Learn, XGBoost, Light GBM and CatBoost libraries for the various algorithms mentioned above (Table 1).

Table 1. Libraries and algorithms used in the study's model

Names of Libraries	Algorithms
SciKit-Learn	Simple Linear Regression
SciKit-Learn	Ridge Regression
SciKit-Learn	Lasso Regression
SciKit-Learn	Elastic Net Regression
SciKit-Learn	Bayesian Linear Regression
SciKit-Learn	Support Vector Regression
SciKit-Learn	Gradient Boosting Regression
SciKit-Learn & XGBoost	XGBoost Regression
SciKit-Learn	K-Nearest Neighbours Regression
SciKit-Learn	AdaBoost Regression
CatBoost	CatBoost Regression
Light GBM & SciKit-Learn	Light GBM Regression

2.3. Comparison of The Used Algorithms

The above-mentioned algorithms were analysed to determine the algorithm with the best performance. The XGBoost algorithm had the best performance.

The evaluation of the algorithms considered the R-squared (R2), mean absolute error (MAE) and root mean square deviation (RMSE) values. These values are defined as follows:

- R2 *(coefficient of determination)* represents how well the values fit relative to the original values. The value of R2 varies between 0 and 1, and the closer it is to 1, the more meaningful it is (Deok, 2019). The values are interpreted as percentages, and the higher the value is, the better the model.

- MAE represents the difference between the original and predicted values and is subtracted by averaging the absolute difference over the dataset.

- RMSE is the error rate determined by the square root of the mean squared error (Deok, 2019).

Of these, lower MAE and RMSE values indicate a higher reliability of the analysis (Y. Lin & Li, 2020; Tang et al., 2020). The applied models are charted below in Table 2. As seen in the table, the algorithm that provided the best results for the metropolitan cities and Zonguldak Province was XGBoost.

Table 2. Comparison of algorithms used for metropolitan cities and Zonguldak Province

Algorithms	MAE	R^2	MSE	RMSE
Simple Linear Regression	846.42	0.88	1370630.30	1170.74
Ridge Regression	846.42	0.88	1370630.30	1170.74
Lasso Regression	846.42	0.88	1370630.30	1170.74
Elastic Net Regression	846.42	0.88	1370630.30	1170.74
Bayesian Linear Regression	845.96	0.88	1372605.33	1171.58
Support Vector Regression	945.66	0.43	6693012.49	2587.09
Gradient Boosting Regression	319.85	0.98	227619.54	477.09
XGBoost Regression	**289.41**	**0.99**	**146562.80**	**382.83**
K-Nearest Neighbours Regression	357.11	0.98	238257.71	488.12
AdaBoost Regression	410.46	0.98	272379.37	521.90
CatBoost Regression	311.39	0.98	187875.45	433.45
Light GBM Regression	518.33	0.94	696565.61	834.60

As seen in the Figure 1, using the XGBoost algorithm the measured and estimated numbers of fatalities and injuries for 2020 very similarly to one another overlaps.

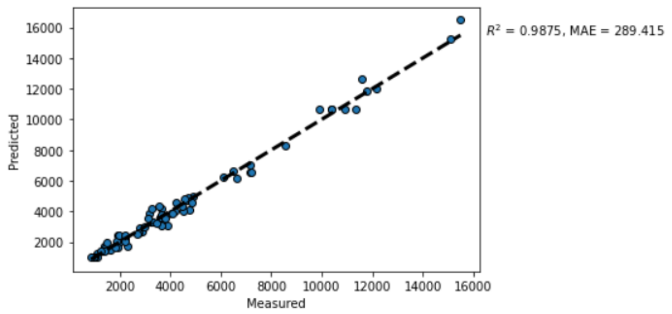

Figure 1. The measured and estimated numbers of fatalities and injuries for 2020 using the XGBoost algorithm

2.4. Extreme Gradient Boosting

XGBoost is a high-performance version of the Gradient Boosting algorithm that has been optimized with various arrangements. It can be used for regression and classification problems (Ibrahem Ahmed Osman et al., 2021). The following equations illustrate how XGBoost works in a dataset with m attributes and n samples (Dataset, [DS]), DS={(xa,ya) : a=1….n, xa ∈ Rm, ya∈R }).

Let $\hat{y}_a$ be the prediction data of an ensemble tree model produced from the following equations. K represents the number of trees in the model, as found by f_k (k-th tree).

$$\text{Å.}_a = \phi(x_a) + \sum_{k=1}^{K} f_k(x_a), \mathrm{f_k}) \in \mathcal{F} \quad (1)$$

To solve the above equation, the best set of functions must be determined by minimizing the losses and the arrangement objects, as follows:

$$L(\Phi) = \sum_a \mathrm{l}\,(\mathrm{y_a}, \text{Å.}_a) + \sum_k \Omega\ (\mathrm{f_k}) \quad (2)$$

In this equation, l represents the missing function, which is the difference between the actual data (y_a) and the predicted data ($\hat{y}_a$). Ω indicates that the model is quite confusing; this variable helps avoid overfitting the model. It is calculated by the formula below:

$$\Omega(\mathrm{f_k}) = \gamma \mathrm{T} + 0.5\lambda\,||\mathrm{w}||^2 \quad (3)$$

Here w is the weight of each leaf, and T is the number of leaves of the tree. DTs work by adding a new function f as the model continues its education. In this process, it is used to minimize the increase of this function in the training of the model. Therefore, in the t^{th} iteration, a new tree is added, as follows:

$$L^t = \sum_{a=1}^{n} l\left(y_a\, \text{Å.}_a^{(t-1)} + f_t(x_a)\right) + \Omega(f_t) \quad (4)$$

$$L_{split} = 0.5 \left[\frac{(\sum_{a\in l_L} g_a)^2}{\sum_{a\in l_L} h_a + \lambda} + \frac{(\sum_{a\in l_R} g_a)^2}{\sum_{a\in l_R} h_a + \lambda} - \frac{(\sum_{a\in l} g_a)^2}{\sum_{a\in l} h_a + \lambda}\right] - \gamma \quad (5)$$

$$g_a = \partial_{\text{Å.}^{t-1}} \mathrm{l}\,(\mathrm{y_a}, \text{Å.}^{(t-1)}) \quad (6)$$

$$g_a = \partial^2_{\text{Å.}^{t-1}} \mathrm{l}\,(\mathrm{y_a}, \text{Å.}^{(t-1)}) \quad (7)$$

2.5. Model Parameters of XGboost

Several parameter values are needed to obtain the best model, and the parameter setting is particularly important for using multiple parameters for the XGBoost algorithm. XGBoost is used to prevent overfitting and prevents excessive complexity (Parsa et al., 2020). Overfitting occurs when the model is too complex (i.e., has too many features/variables compared to the number of observations). If the model has begun to memorize and work too hard on the dataset used for training, it will have very high prediction accuracy for the training data but will likely not be able to predict untrained or new data accurately (Bayraktutan & Özbilgin, 2013). Table 3 shows the parameters that were tested on the model with the XGBoost algorithm. Table 4 shows the best parameter values of the analysed model.

Table 3. Parameter values for the XGBoost algorithm

Parameters	Tried Parameter Values				
Learning rate	0.01	0.1	0.5	0.05	-
Max depth	5	8	10	10	15
Iterations	100	500	1000	1500	2000
Colsample by tree	1	0.5	0.1	-	-

Table 4. The best parameter values for the XGBoost algorithm

Parameters	The Best Parameter Values
Learning rate	0.01
Max depth	5
Number of iterations	500
Colsample by tree	1

3. Results

Due to the COVID-19 outbreak, lockdown restrictions were enforced in metropolitan cities and Zonguldak Province in Turkey in 2020. To determine the impact of the lockdown on fatality and injury traffic accidents in these cities, nonlinear multivariate regression analysis was compared to the results of other algorithms. The XGBoost algorithm, which provided the best results, was used, and the results were analysed. Of the data from 2012–2019, 25% were chosen randomly as test data, and the remaining 75% were chosen as training data. The data from Istanbul, which had much larger values than the data from other cities, were considered outlier. However, data cleaning was not performed to remove the Istanbul data from the model since cleaning of data was regarded as a limitation to the model.

3.1. Model Results of Metropolitan Cities and Zonguldak

Zonguldak and all metropolitan cities in Turkey were examined in the study as the first group. As seen in Table 5, the year, population, number of fatal and injury accidents, number of vehicles by provinces and number of accidents involving fatalities and injuries were predicted for 2020 with the XGBoost algorithm using the data from 2012–2019. Zonguldak was included in the same group as the metropolitan cities because lung diseases are common there due to the coal mining industry; therefore, the COVID-19 lockdown was imposed in Zonguldak as well as the metropolitan cities. Zonguldak province has very patients with chronic lung diseases. So, the management of COVID-19 in patients with chronic lung diseases requires a separate effort (Metintas, 2020). Therefore, Zonguldak was evaluated as metropolitan city and the provincial data from Zonguldak were also included in the study. In general, the accidents were fewer than predicted, except in two cities (Diyarbakır and Ordu). Thus, the lockdowns decreased the traffic accidents in most metropolitan cities. According to the predicted number of traffic accidents, the rate of decrease in metropolitan cities was 18.3%. Despite the decrease in metropolitan cities across Turkey, an increase was found in Diyarbakır and Ordu, which may have been due to quarantine restrictions not being employed strictly enough.

As seen in Figure 2, the actual data and predicted values were compared using the XGBoost algorithm. As seen in Figure 3 (created with Tableau 2020.4 Desktop program), the highest number of accidents was observed in Ankara, Izmir, Antalya, Bursa, Mersin and İstanbul. The population and number of vehicles in these cities are dense.

As seen in Figure 4, the relationship between population and the number of accidents involving fatal and injuries between 2012 and 2019 were observed. Population and severe accidents are directly proportional.

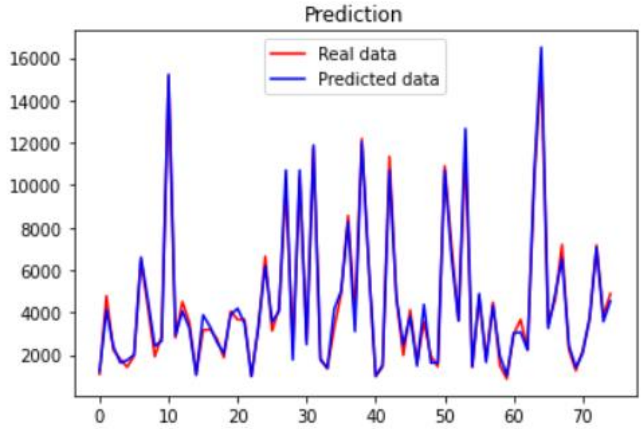

Figure 2. Comparison of the measured and predicted number of fatal and injury accidents in 2020

Figure 3. Average cases of accidents in metropolitan cities and Zonguldak

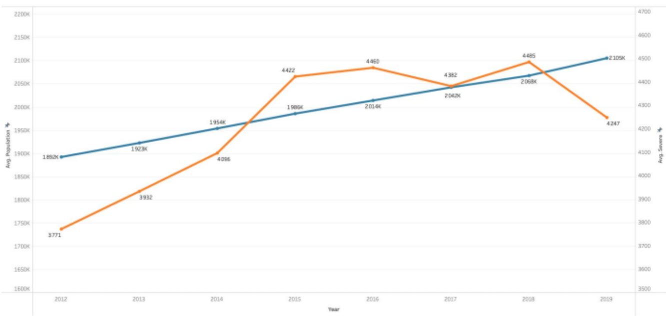

Figure 4. Population and the number of accidents involving fatal and injuries between 2012 and 2019

3.2. Normality Test Results

To evaluate the result of fatality and injury accidents in 2020, evaluating whether the difference between the predicted value and the actual value was statistically significant was important. First, a Shapiro Wilk test was performed as a normality test to determine whether the measurements had a normal distribution. P-values lower than 0.05 would indicate that the measurements did not have a normal distribution and that non-parametric tests should be applied, while p-values greater than 0.05 would show a normal distribution, in which case parametric tests should be applied (Bazlamit et al., 2020; Rezapour Mashhadi et al., 2017). The p-values were $p < .001$.

As they were quite smaller than 0.05 and the distribution was therefore not normal, non-parametric tests were employed. Thus, a Wilcoxon signed ranks test was performed. The predicted value and the actual value were compared to determine whether the difference obtained was significant. For the Wilcoxon signed ranks test statistic = 7, $p < .001$. In health sciences, if the p-value obtained as the result of this test is less than 0.001, the difference between the predicted and the actual number of accidents is regarded as statistically significant (Jayavel & Lizy, 2014). In social and physical sciences, significance is assumed when the p-value is less than 0.05 (Liu et al., 2019). As the p-value in this study was smaller than either of the criteria ($p < .001$), the difference between values was regarded as statistically significant.

One of the positive impacts of the measures implemented to control the spread of COVID-19 was the decrease in traffic accidents on both urban and interurban roads (Aloi et al., 2020; Brodeur et al., 2021; Shilling & Waetjen, 2020), which resulted in a marked fall in the number of traffic-related injuries and fatalities (Saladié et al., 2020). However, interestingly Lin et al. (2020) found that the number of non-fatal accidents decreased while the number of severe and fatal traffic accidents remained the same during the pandemic in two cities in the U.S., Los Angeles and New York City. Additionally, Qureshi et al. (2020) found a significant reduction in road traffic accidents resulting in minor or no injuries (M_{before}=14.5, vs. M_{after}=10.8, p < 0.0001) but not in accidents resulting in serious or fatal injuries (M_{before}=3.4 vs. M_{after}=3.7, p = 0.42) after mandated societal lockdown. The authors in both articles could not clarify the reasons behind lack of reduction in road traffic accidents resulting in serious or fatal injuries during the COVID-19 pandemic. Thus, as in other countries around the world, the COVID-19 pandemic led to a significant decrease in the number of fatal and injury traffic accidents in Turkey, especially in metropolitan cities where quarantine was enforced in 2020. Furthermore, this study showed that machine learning can make mistakes in extraordinary situations such as earthquake, war and pandemic, etc., no matter the accuracy of its predictions.

Table 5. Model results for metropolitan cities and Zonguldak

City	Year	Killed or Severely Injured Accidents	Population	Sum of Vehicles	Predicted Value	Percentage (%)
Adana	2020	4146	2258718	680645	6043.59	31.398
Ankara	2020	9572	5663322	2125454	12658.00	24.380
Antalya	2020	6901	2548308	1137476	8140.00	15.221
Aydın	2020	3103	1119084	474254	3684.82	15.790
Balıkesir	2020	3319	1240285	502080	4104.24	19.132
Bursa	2020	5405	3101833	945412	7365.61	26.618
Denizli	2020	2643	1040915	423574	3292.60	19.729
Diyarbakır	2020	2033	1783431	133568	1795.47	-13.229
Erzurum	2020	1083	758279	124761	1542.37	29.784
Eskişehir	2020	1838	888828	296918	2764.30	33.509
Gaziantep	2020	3231	2101157	539568	5643.75	42.751
Hatay	2020	3540	1659320	509371	3910.05	9.464
İstanbul	2020	15421	15462452	4565985	16737.00	7.863
İzmir	2020	8868	4394694	1487675	10703.00	17.145
Kahramanmaraş	2020	2215	1168163	243540	2713.28	18.364
Kayseri	2020	2954	1421455	390057	3537.43	16.493
Kocaeli	2020	3127	1997258	419821	3497.92	10.604
Konya	2020	4718	2250020	745076	5993.65	21.283
Malatya	2020	1459	806156	182560	1886.92	22.678
Manisa	2020	3608	1450616	610155	4594.24	21.467
Mardin	2020	966	854716	81820	1166.79	17.209
Mersin	2020	5085	1868757	640566	6096.63	16.593
Muğla	2020	3872	1000773	527163	4048.23	04.353
Ordu	2020	1618	761400	141047	1504.79	-7.523
Sakarya	2020	2166	1042649	299025	3253.17	33.419
Samsun	2020	2829	1356079	371919	3021.44	06.369
Şanlıurfa	2020	2806	2115256	269153	2981.42	05.884
Tekirdağ	2020	1870	1081065	280955	3177.05	41.140
Trabzon	2020	1491	811901	206640	1755.80	15.082
Van	2020	1145	1149342	82740	1397.82	18.087
Zonguldak	2020	875	591204	161323	1210.24	27.700

4. Discussion

The findings of this study reveal a significant decrease in the number of fatal and injury traffic accidents in Turkey during the COVID-19 pandemic, particularly in metropolitan cities where strict lockdown measures were enforced. These results align with previous studies conducted in various countries, which also reported a reduction in traffic accidents due to mobility restrictions and reduced vehicular traffic (Aloi et al., 2020; Brodeur et al., 2021; Shilling & Waetjen, 2020). For instance, Saladié et al. (2020b) found that mobility restrictions led to a sharp decline in road accidents and associated injuries in Spain, supporting the argument that limited movement directly correlates with fewer traffic incidents.

However, contrary to the overall trend of reduced accidents, an increase was observed in Diyarbakır and Ordu. This anomaly may be attributed to the inadequate enforcement of lockdown measures in these cities. Similar findings have been reported in other studies, where cities with less stringent quarantine policies experienced a lesser reduction in traffic accidents compared to regions with stricter measures (Qureshi et al., 2020). Furthermore, socio-economic and cultural factors may have played a role in shaping driver behaviors in these regions, necessitating further investigation into localized impacts of lockdown policies.

An interesting aspect of this study is the analysis of severe and fatal traffic accidents during the pandemic. While the overall number of accidents decreased, some studies indicate that the proportion of severe accidents remained stable or even increased. L. Lin et al. (2020) found that the number of non-fatal accidents decreased while the number of severe and fatal traffic accidents remained the same during the pandemic in two cities in the U.S., Los Angeles and New York City. A similar pattern was noted by Qureshi et al. (2020), who reported a reduction in minor accidents (M_{before}=14.5 vs. M_{after}=10.8, $p < 0.0001$) but found no significant change in fatal accidents (M_{before}=3.4 vs. M_{after}=3.7, $p = 0.42$) during the COVID-19 pandemic. These findings suggest that although there were fewer vehicles on the road, factors such as increased speeding on empty roads, riskier driving behaviors, and reduced law enforcement might have contributed to the unchanged fatality rates.

Moreover, the present study highlights the limitations of machine learning models in predicting outcomes during extraordinary circumstances such as pandemics, natural disasters, or war. Despite XGBoost providing highly accurate predictions under normal conditions, the unforeseen impact of lockdowns introduced inconsistencies between the predicted and actual accident rates. This observation aligns with the argument that machine learning models, despite their sophistication, can struggle to account for abrupt, unprecedented societal changes (Jayavel & Lizy, 2014; Liu et al., 2019). The results underscore the necessity of incorporating external factors, such as real-time government policies and human behavioral changes, to enhance predictive accuracy in such extraordinary scenarios.

This study has certain limitations that should be acknowledged. One key limitation is that the analysis is based solely on data from Türkiye, which may limit the applicability of the findings to other regions. Additionally, the study utilizes accident data from the period between 2012 and 2019, meaning that more recent trends and potential changes in accident patterns are not considered.

Overall, this study contributes to the growing body of literature examining the impact of the COVID-19 pandemic on road safety. It confirms the effectiveness of lockdown measures in reducing traffic accidents while also emphasizing the challenges in predicting accident patterns during global crises. Future research should focus on integrating adaptive machine learning models that can dynamically adjust to extraordinary events and investigate long-term trends in post-pandemic traffic behavior to develop more resilient traffic management policies.

Ethics Committee Approval Statement

Since the study was conducted with data obtained from secondary data sources, ethics committee approval is not required.

Acknowledgement

I would like to thank The Turkish Statistical Institute for sharing the data used in this study.

Author Contributions

Percentages of the author(s) contributions is present below. All authors reviewed and approved final version of the manuscript.

	First author
C	100
D	100
S	-
DCP	-
DAI	-
L	100
W	100
CR	100
SR	100
PM	100
FA	100

Note. C= concept, D= design, S= supervision, DCP= data collection and/or processing, DAI= data analysis and/or interpretation, L= literature search, W= writing, CR= critical review, SR= submission and revision, PM= project management, FA= funding acquisition.

References

Aloi, A., Alonso, B., Benavente, J., Cordera, R., Echániz, E., González, F., Ladisa, C., Lezama-Romanelli, R., López-Parra, Á., Mazzei, V., Perrucci, L., Prieto-Quintana, D., Rodríguez, A., & Sañudo, R. (2020). Effects of the COVID-19 Lockdown on Urban Mobility: Empirical Evidence from the City of Santander (Spain). *Sustainability*, *12*(9), 3870. https://doi.org/10.3390/su12093870

Bayraktutan, Y., & Özbilgin, M. (2013). Türkiye'de İller Düzeyinde Karayolu Yük Trafiği Dağılımının Analizi. *Çukurova Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi*, *17*(2), 81-92.

Bazlamit, S. M., Eid, A., & Al-mofleh, H. (2020). Impact of fasting on traffic accidents. *Jordan Journal of Civil Engineering*, *14*(2), 431–442.

Brodeur, A., Cook, N., & Wright, T. (2021). On the effects of COVID-19 safer-at-home policies on social distancing, car crashes and pollution. *Journal of Environmental Economics and Management*, *106*, 102427. https://doi.org/10.1016/j.jeem.2021.102427

Chang, L.-Y., & Wang, H.-W. (2006). Analysis of traffic injury severity: An application of non-parametric classification tree techniques. *Accident Analysis & Prevention*, *38*(5), 1019–1027. https://doi.org/10.1016/j.aap.2006.04.009

Chong, M., Abraham, A., & Paprzycki, M. (2005). No TitleTraffic accident analysis using machine learning paradigms. *Informatica*, *29*(1), 89–98.

Christey, G., Amey, J., Campbell, A., & Smith, A. (2020). Variation in volumes and characteristics of trauma patients admitted to a level one trauma centre during national level 4 lockdown for COVID-19 in New Zealand. *The New Zealand Medical Journal*, *133*(1513), 81-88.

Deok, K. C. (2019). Data-Driven Materials Modeling with XGBoost Algorithm and Statistical Inference Analysis for Prediction of Fatigue Strength of Steels. *International Journal of Precision Engineering and Manufacturing*, *20*(1), 129–138. https://doi.org/10.1007/s12541-019-00048-6

Envepo A.Ş. (2018) Makine Öğrenmesi Nasıl Çalışır? Makine Öğrenme Algoritmaları. https://www.envepo.com/blog/makine-ogrenmesi-nasil-calisir-makine-ogrenme-algoritmalari

Emniyet Genel Müdürlüğü Trafik Başkanlığı (2020). *Trafik İstatistik Bülteni Ülke Geneli*. https://trafik.gov.tr/kurumlar/trafik.gov.tr/04-Istatistik/Aylik/202503/Mart-2025.pdf

Gökdağ, M., Kaya, M. D., Atalay, A., & Haşiloğlu, A. S. (2004). Injuries and fatalities in Turkish road traffic accidents. *Proceedings of the Institution of Civil Engineers: Transport*, *157*(4), 231–237. https://doi.org/10.1680/tran.2004.157.4.231

Ibrahem Ahmed Osman, A., Najah Ahmed, A., Chow, M. F., Feng Huang, Y., & El-Shafie, A. (2021). Extreme gradient boosting (Xgboost) model to predict the groundwater levels in Selangor Malaysia. *Ain Shams Engineering Journal*, *12*(2),1545-1556. https://doi.org/10.1016/j.asej.2020.11.011

Jayavel, M., & Lizy, M. L. (2014). Knowledge and Attitude Regarding prevention of Road Traffic Accidents among Adolescents. *Innovations in Pharmaceuticals and Pharmacotherapy (IPP)*, *2*(1), 328–339.

Jovic Vranes, A., Bjegovic Mikanovic, V., Milin Lazovic, J., & Kosanovic, V. (2018). Road traffic safety as a public health problem: Evidence from Serbia. *Journal of Transport and Health*, *8*, 55–62. https://doi.org/10.1016/j.jth.2017.12.005

Korkmaz, E. (2022). COVID-19 Salgın Sürecinin Toplu Taşıma Sistemlerine Etkisinin Anket Yöntemiyle İstanbul-Ankara İçin İncelenmesi. *Afet ve Risk Dergisi*, *5*(1), 247-260. https://doi.org/10.35341/afet.1060291

Kwon, O. H., Rhee, W., & Yoon, Y. (2015). Application of classification algorithms for analysis of road safety risk factor dependencies. *Accident Analysis & Prevention*, *75*, 1–15. https://doi.org/10.1016/j.aap.2014.11.005

Lin, L., Shi, F., & Li, W. (2021). Assessing inequality, irregularity, and severity regarding road traffic safety during COVID-19. *Scientific Reports, 11*(1), 1-7. https://doi.org/10.1038/s41598-021-91392-z

Lin, Y., & Li, R. (2020). Real-time traffic accidents post-impact prediction: Based on crowdsourcing data. *Accident Analysis and Prevention*, *145*(2020), 1-11 105696. https://doi.org/10.1016/j.aap.2020.105696

Liu, W. Q., Hwang, B. G., Shan, M., & Looi, K. Y. (2019). Prefabricated Prefinished Volumetric Construction: Key Constraints and Mitigation Strategies. *IOP Conference Series: Earth and Environmental Science*, *385*(1), 1-7. https://doi.org/10.1088/1755-1315/385/1/012001

Ma, J., Ding, Y., Cheng, J. C. P., Tan, Y., Gan, V. J. L., & Zhang, J. (2019). Analyzing the Leading Causes of Traffic Fatalities Using XGBoost and Grid-Based Analysis: A City Management Perspective. *IEEE Access*, *7*, 148059–148072. https://doi.org/10.1109/ACCESS.2019.2946401

Metintas, S. (2020). Epidemiology of COVID-19. *Eurasian Journal of Pulmonology*, *22*(4), 2-7. https://doi.org/10.4103/ejop.ejop_55_20

Muhammad, L. J., Salisu, S., Yakubu, A., Malgwi, Y. M., Abdullahi, E. T., Mohammed, I. . A., & Muhammad, N. A. (2017). Using Decision Tree Data Mining Algorithm to Predict Causes of Road Traffic Accidents, its Prone Locations and Time along Kano –Wudil Highway. *International Journal of Database Theory and Application*, *10*(1), 197–206. https://doi.org/10.14257/ijdta.2017.10.1.18

Nuñez, J. H., Sallent, A., Lakhani, K., Guerra-Farfan, E., Vidal, N., Ekhtiari, S., & Minguell, J. (2020). Impact of the COVID-19 Pandemic on an Emergency Traumatology Service: Experience at a Tertiary Trauma Centre in Spain. *Injury*, *51*(7), 1414–1418. https://doi.org/10.1016/j.injury.2020.05.016

Oguzoglu, U. (2020). Covid-19 Lockdowns and Decline in Traffic Related Deaths and Injuries. *SSRN Electronic Journal. 13278,* 19*,* 1-17. https://doi.org/10.2139/ssrn.3608527

Parsa, A. B., Movahedi, A., Taghipour, H., Derrible, S., & Mohammadian, A. (Kouros). (2020). Toward safer highways, application of XGBoost and SHAP for real-time accident detection and feature analysis. *Accident Analysis and Prevention*, *136*(2020), 105405. https://doi.org/10.1016/j.aap.2019.105405

Qureshi, A. I., Huang, W., Khan, S., Lobanova, I., Siddiq, F., Gomez, C. R., & Suri, M. F. K. (2020). Mandated societal lockdown and road traffic accidents. *Accident Analysis and Prevention*, *146*(2020), 105747. https://doi.org/10.1016/j.aap.2020.105747

Rezapour Mashhadi, M. M., Saha, P., & Ksaibati, K. (2017). Impact of traffic Enforcement on Traffic Safety. *International Journal of Police Science & Management*, *19*(4), 238–246. https://doi.org/10.1177/1461355717730836

Saladié, Ò., Bustamante, E., & Gutiérrez, A. (2020). COVID-19 lockdown and reduction of traffic accidents in Tarragona province, Spain. *Transportation Research Interdisciplinary Perspectives*, *8*(2020), 100218. https://doi.org/10.1016/j.trip.2020.100218

Selimoğlu, Eda. (2014). Trafik Kazalarının Nedenleri , Sonuçları ve Kazaların Önlenmesine İlişkin Öneriler. *Ziraat Mühendisliği Dergisi*, *361*, 51–54.

Shilling, F., & Waetjen, D. (2020). Special Reports: Impact of COVID-19 on California Traffic Crashes. *Road Ecology Center* https://roadecology.ucdavis.edu/content/reports-covid-19-mitigation-and-traffic

Sohn, S. Y., & Shin, H. (2001). Pattern recognition for road traffic accident severity in Korea. *Ergonomics*, *44*(1), 107–117. https://doi.org/10.1080/00140130120928

Tang, J., Zheng, L., Han, C., Liu, F., & Cai, J. (2020). Traffic Incident Clearance Time Prediction and Influencing Factor Analysis Using Extreme Gradient Boosting Model. *Journal of Advanced Transportation*, 2020, 1-12. https://doi.org/10.1155/2020/6401082

Türkiye İstatistik Kurumu. (2020). *Türkiye İstatistik Kurumu.* Türkiye karayollarında 2012-2018 yılları arasında gerçekleşen trafik kazaları verileri. https://data.tuik.gov.tr/bulten/index?p=karayolu-trafik-kaza-istatistikleri-2020-37436

World Health Organization (2019). *World Health Organization, Statistics Annual* 2019: Geneva., WHO, Statistics Annual. https://iris.who.int/bitstream/handle/10665/324835/9789241565707-eng.pdf

Yavuz, A. A., Ergül, B., & Aşık, E. G. (2021). Trafik Kazalarının Makine Öğrenmesi Yöntemleri Kullanılarak Değerlendirilmesi Evaluation of Traffic Accidents Using Machine Learning Methods. *Uluslararası Mühendislik Araştırma ve Geliştirme Dergisi*, *13*(1), 66–73. https://doi.org/10.29137/umagd.705156