

# Application of various machine learning algorithms in view of predicting the CO<sub>2</sub> emissions in the transportation sector

Gökalp Çınar<sup>1</sup>, Murat Kadir Yeşilyurt<sup>2</sup>, Ümit Ağbulut<sup>3</sup>, Zeki Yılbaşı<sup>4,\*</sup>, and Kazım Kılıç<sup>5</sup>

<sup>1</sup> Department of Computer Engineering, Faculty of Engineering-Architecture, Yozgat Bozok University, Yozgat, Türkiye

<sup>2</sup> Department of Mechanical Engineering, Faculty of Engineering-Architecture, Yozgat Bozok University, Yozgat, Türkiye

<sup>3</sup> Department of Mechanical Engineering, Faculty of Engineering, Düzce University, Düzce, Türkiye

<sup>4</sup> Department of Automotive Technology, Yozgat Vocational School, Yozgat Bozok University, Yozgat, Türkiye

<sup>5</sup> Department of Computer Technology, Yozgat Vocational School, Yozgat Bozok University, Yozgat, Türkiye

Received: 11 January 2024 / Accepted: 8 February 2024

**Abstract.** This study applies three different artificial intelligence algorithms (Multi-layer Perceptron (MLP), Extreme Gradient Boosting (XGBoost), and Support Vector Machine (SVM)) to estimate CO<sub>2</sub> emissions in Türkiye's transportation sector. The input parameters considered are Energy consumption (ENERGY), Vehicle Kilometers (VK), POPulation (POP), Year (Y), and Gross Domestic Product Per Capita (GDP). Strong correlations are observed, with ENERGY having the highest correlation followed by VK, POP, Y, and GDP. Four scenarios are designed based on the correlation effect: scenario 1 (ENERGY/VK/POP/Y/GDP), scenario 2 (ENERGY/VK/POP/Y), scenario 3 (ENERGY/VK/POP), and scenario 4 (ENERGY/VK). Experiments compare their effects on CO<sub>2</sub> emissions using statistical indicators ( $R^2$ , RMSE, MSE, and MAE). Across all scenarios and algorithms,  $R^2$  values range from 0.8969 to 0.9886, and RMSE values range from 0.0333 to 0.1007. The XGBoost algorithm performs best in scenario 4. Artificial intelligence algorithms prove successful in estimating CO<sub>2</sub> emissions. This study has significant implications for policymakers and stakeholders. It highlights the need to review energy investments in transportation and implement regulations, restrictions, legislation, and obligations to reduce emissions. Artificial intelligence algorithms offer the potential for developing effective strategies. Policymakers can use these insights to prioritize sustainable energy investments. In conclusion, this study provides insights into the relationship between input parameters and CO<sub>2</sub> emissions in the transportation sector. It emphasizes the importance of proactive measures and policies to address the sector's environmental impact. It also contributes to the understanding of AI-assisted CO<sub>2</sub> emissions forecasting in the transport sector, potentially informing future policy decisions aimed at emission reduction and sustainable transport development.

**Keywords:** CO<sub>2</sub> emissions, Transportation sector, Artificial intelligence, Machine learning algorithm, Statistical indicators.

## Nomenclature

ABC	Artificial Bee Colony
AE	Absolute Error
AI	Artificial Intelligence
ANN	Artificial Neural Network
APE	Absolute Percentage Error
ARIMA	Autoregressive Integrated Moving Average
BPNN	Back Propagation Neural Network
BSA	Backtracking Search Algorithm
CKC	Carbon Kuznets Curve

CO <sub>2</sub>	Carbon dioxide
CNN	Convolutional Neural Network
DGM	Discrete Grey Modeling
DL	Deep Learning
E	Error
EO	Equilibrium Optimizer
GBR	Gradient Boosting Regression
GDP	Gross Domestic Product
GHG	GreenHouse Gas
GHGs	GreenHouse Gas Emissions
GM	Grey Model
GWO	Grey Wolf Optimizer
H-W	Holt-Winters

\* Corresponding author: [zeki.yilbasi@bozok.edu.tr](mailto:zeki.yilbasi@bozok.edu.tr)

LMDI	Logarithmic Mean Divisia Index
LSA	Lightning Search Algorithm
MAE	Mean Absolute Error
MAPE	Mean Absolute Percentage Error
MENR	Ministry of Energy and Natural Resources
ML	Machine Learning
MNGM	Metabolic Nonlinear Grey Model
MPA	Marine Predators Algorithm
NSE	Nash–Sutcliffe Efficiency
OECD	Organization for Economic Cooperation and Development
OLS	Ordinary Least Squares Regression
PSO	Particle Swarm Optimization
RE	Relative Error
RFR	Random Forest Regressor
RMSE	Root Mean Squared Error
SMAPE	Symmetric Mean Absolute Percentage Error
SOS	Symbiotic Organisms Search
SVM	Support Vector Machine
SVR	Support Vector Regression

## 1 Introduction

The transportation sector is of great importance for the globalization of nations, and the sustainability of daily life for human beings [1, 2]. It provides a reliable service for the transfer of people, goods, and materials from one point to another. In line with this, automobile use and density in the transportation network have exponentially increased over the past decades due to its immense necessity and enhancement of living standards [3, 4]. However, this increment has triggered enormous energy consumption for the limited non-renewable fossil fuels [5]. Correspondingly, this case creates a negative effect, particularly on the countries that do not have their own fossil fuel reserves. From the beginning, a large portion of the transportation sector has been dominated by the burning of fossil-based fuels in internal combustion engines. Even today, it is estimated that the usage percentage of fossil fuels in the transportation sector is more than 99% [6]. From another perspective, the transportation sector alone consumes a quarter of the world's energy [7] and more than half of the oil production of the world [8]. Numerically, nearly 11 billion liters per day of diesel, gasoline, heavy fuel oil, and jet fuel are burned to sustain the transportation sector [9]. With this huge energy consumption, there is a rapid depletion of fossil fuel reserves, resulting in an upward fluctuation in their prices day by day. Today, many countries have suffered from the price increase of fossil fuels since their economies are adversely affected by these increments. The problem encountered with the use of fossil fuels is not only the rise in their prices. Such high depletion of fossil fuels is also the main trigger of many serious problems for the environment and living organisms. In other words, high fossil fuel consumption in this sector is also responsible for a growing percentage of pollutants in the atmosphere, which

dramatically worsens the air quality on a local and global scale. Consequently, the effect of GreenHouse Gas emissions (GHGs) has nowadays become more visible to all of humanity as well as to nature. Today, it is an accepted reality that the transportation sector is responsible for many human deaths and causes many serious diseases [10–12]. More and more, GHGs are, on the other hand, the most serious leading reason for climate change that the world has witnessed nowadays. Numerically, fossil fuel consumption in all sectors has cumulatively accounted for nearly 70% of total GHGs all across the world [13] and approximately 80–90% of the World carbon emissions (CO<sub>2</sub>) [14]. Among GHG emissions, CO<sub>2</sub> emissions are located in a significant position due to their excessive amount. It accounts for 76% of total GHG emissions, followed by methane (16%), nitrous oxide (6%), and fluorinated gases (2%) [15]. Therefore, it has been observed in the literature that the published works have focused in general on estimating only CO<sub>2</sub> emissions, and the researchers have set a nexus between the energetic and economic indicators with carbon emissions.

Türkiye is a country whose transportation network is constantly increasing year by year. In the country, the annual growth rate of registered motor vehicles is about 9%, and it has nearly doubled in the past two decades [10]. This is a very critical problem for sustainable economic development. Therefore, the government is seriously looking for solutions to prevent energy consumption and carbon emissions in the transportation sector. Another significant issue is that Türkiye is a country that is highly dependent on energy. Considering this issue numerically, Türkiye is a country that imports 3/4 of its primary energy needs and more than 90% of its petroleum needs [10, 16, 17]. On the other hand, the transportation sector in Türkiye accounted for more than 1/4 of Türkiye's total energy consumption in 2017 [18, 19]. All of these clearly demonstrated that the transportation sector is very critical for the sustainable development targets of the governments, the improvement of living standards, and the achievement of a healthy environment. In this concept, the ongoing attempts available in the literature can be classified in two ways. On one hand, the researchers have strongly relied on the idea that electric vehicles will be a robust solution to carbon mitigation arising from the transportation sector [20–22]. However, even if there are significant initiatives for alternative fuels and propulsion systems in today's world, the projections for 2050 demonstrate the key role of fossil fuels as well as the high share thereof in the transportation sector [9]. Therefore, the combustion community is dedicated to mitigating the carbon emissions arising from the transportation sector. In this direction, fuel researchers have, on the other hand, tried to improve the alternative fuels to the conventional fuels [23–25], which have a high energy density, low cost, and fewer exhaust pollutants. In both scenarios, the researchers were aware of the problem of the transportation sector and agreed that a quick solution to the problem was required as soon as possible. The solution to the problem must be found in advance, and the steps must be taken starting today. As it is well-known, yesterday's decisions will affect today, and today's decisions

will affect the world of tomorrow. That is why it is vital to forecast transportation-based energy consumption as well as transportation-based carbon emissions in any state. With this viewpoint, decision-makers and policymakers can revise their future investments in all sectors, including the transportation sector so that they can reach their carbon mitigation and energy-saving goals. All of which converts the issue into a very attractive and hot discussion around the global community. Many researchers have used various soft computing approaches, time series, and empirical models to forecast the carbon footprint and energy consumption arising from the transportation sector with minimum error. [Table 1](#) gives a summary of the recent works dealing with the forecasting of energy consumption and CO<sub>2</sub> emissions from different sectors.

Isik *et al.* argue that economic expansion, population growth, and emission intensity all contribute to Türkiye's transportation sector's CO<sub>2</sub> emissions, with fleet efficiency and fuel switching incentives showing encouraging trends from 2000 to 2010, but SUV popularity poses a challenge to emission reductions from 2010 to 2017. Influential factors on CO<sub>2</sub> emissions in the transportation sector, emission reduction potential, effects of fuel switching incentives, and fleet efficiency on emission mitigation. It was also reported that reducing CO<sub>2</sub> emissions in the transportation sector requires strategic measures, including: (i) meticulous planning of freight transportation demand management, encompassing comprehensive designs of production sites, material flows, and demand points, along with economic activities; (ii) implementation of passenger transportation strategies to decrease car travel, such as zoning for public transit corridors and enhancement of the public transportation system; and (iii) implementation of effective incentives for energy-efficient vehicles and clean energy technologies like electric cars to persuade individuals [19]. Bozdağ *et al.* conducted training in Machine Learning (ML) algorithms (Artificial Neural Network (ANN), *k*-Nearest Neighbors (kNN), LASSO, RF, Support Vector Regression (SVR), and Extreme Gradient Boosting (XGBoost)) using PM10 concentrations collected from seven stations located in Ankara, Türkiye. The input consisted of PM10 concentrations from 2009 to 2017 at six stations in Ankara. The goal was to forecast the PM10 concentrations for the seventh site in the year 2018. For each station, the model development stage was iterated, and the algorithms' performance and error rates were assessed by comparing their results with the actual outcomes. The most optimal outcomes were achieved using Artificial Neural Network (ANN) with a coefficient of determination ( $R^2$ ) of 0.58, a Root Mean Square Error (RMSE) of 20.8, and a Mean Absolute Error (MAE) of 14.4 [26]. Ozturk and Ozturk used data spanning from 1970 to 2015 about the utilization of coal, oil, natural gas, renewable energy sources, and overall energy to predict Türkiye's energy consumption over the next 25 years using the Autoregressive Integrated Moving Average (ARIMA) model. The ARIMA models have been identified as ARIMA (1, 1, 1) for coal consumption, ARIMA (0, 0, 0) for natural gas consumption, ARIMA (0, 1, 0) for oil consumption, ARIMA (1, 1, 0) for renewable energy consumption, and ARIMA (0, 1, 2) for total energy

consumption. The findings suggest that Türkiye's energy consumption is projected to see a sustained upward trend until the conclusion of 2040. The consumption of coal, natural gas, oil, renewable energy, and total energy is projected to grow at average annual rates of 4.87%, 4.39%, 3.92%, 1.64%, and 4.20% accordingly during the next 25 years [27].

Based on the previous literature given in [Table 1](#), upon analyzing the energy demand and carbon emissions arising from the transportation sector, it is well noticed that the researchers are trying to forecast the carbon emissions and energy consumption of the states with minimum error, depending on various input parameters for different sectors. In the literature, it is seen that ANNs are also used in different fields. Predictions of concrete compressive strength with ANNs [28], regression-based estimation of tunnel convergence [29], estimation of the safety factor of retaining walls with ANNs and genetic algorithms [30], and estimation of the foundation period of infilled reinforced concrete structures with ANNs [31] have been carried out. Accordingly, it is seen that the researchers have used various forecasting techniques, including ML algorithms, time series, empirical mathematical equations, and hybrid models, in recent years. In line with this, the present research focuses on the forecasting of transportation-related carbon emissions as well as energy consumption in Türkiye. Accordingly, years, gross domestic product, vehicle kilometers, transportation-based CO<sub>2</sub> emissions, and energy consumption between 1970 and 2016 in Türkiye are used as input and output parameters. In the forecasting of CO<sub>2</sub> emissions in Türkiye, the MLP, XGBoost, and SVM algorithms are used, taking into account several scenarios as follows: scenario 1: ENERGY/VK/POP/Y/GDP; scenario 2: ENERGY/VK/POP/Y; scenario 3: ENERGY/VK/POP; and scenario 4: ENERGY/VK. Furthermore, the performance successes of the algorithms are discussed with  $R^2$ , MAE, MAE, MAPE, MSE, RMSE, rRMSE, and MBE statistical metrics in the present research.

## 2 Methodology

In this section, where the data is obtained from, ML algorithms and scenarios created are explained in detail. [Section 2.1](#) specifies what was done during the preparation of the data set. In [Section 2.2](#), ML techniques and detailed mathematical explanations of the ML algorithms preferred in this study are presented under sub-titles.

### 2.1 Dataset

The data used in this study are transportation energy consumption/demand – Mtoe (ENERGY), vehicle kilometer-km (VK), POPulation – million (POP), Year (Y), Gross Domestic Product Per Capita (GDP), and CO<sub>2</sub> emissions from the transportation sector between 1970 and 2016. Among the parameters mentioned above, VK, GDP, POP, and ENERGY were used as input parameters in the analysis, and CO<sub>2</sub> emissions from the transportation sector were chosen as output parameters. The change in the raw

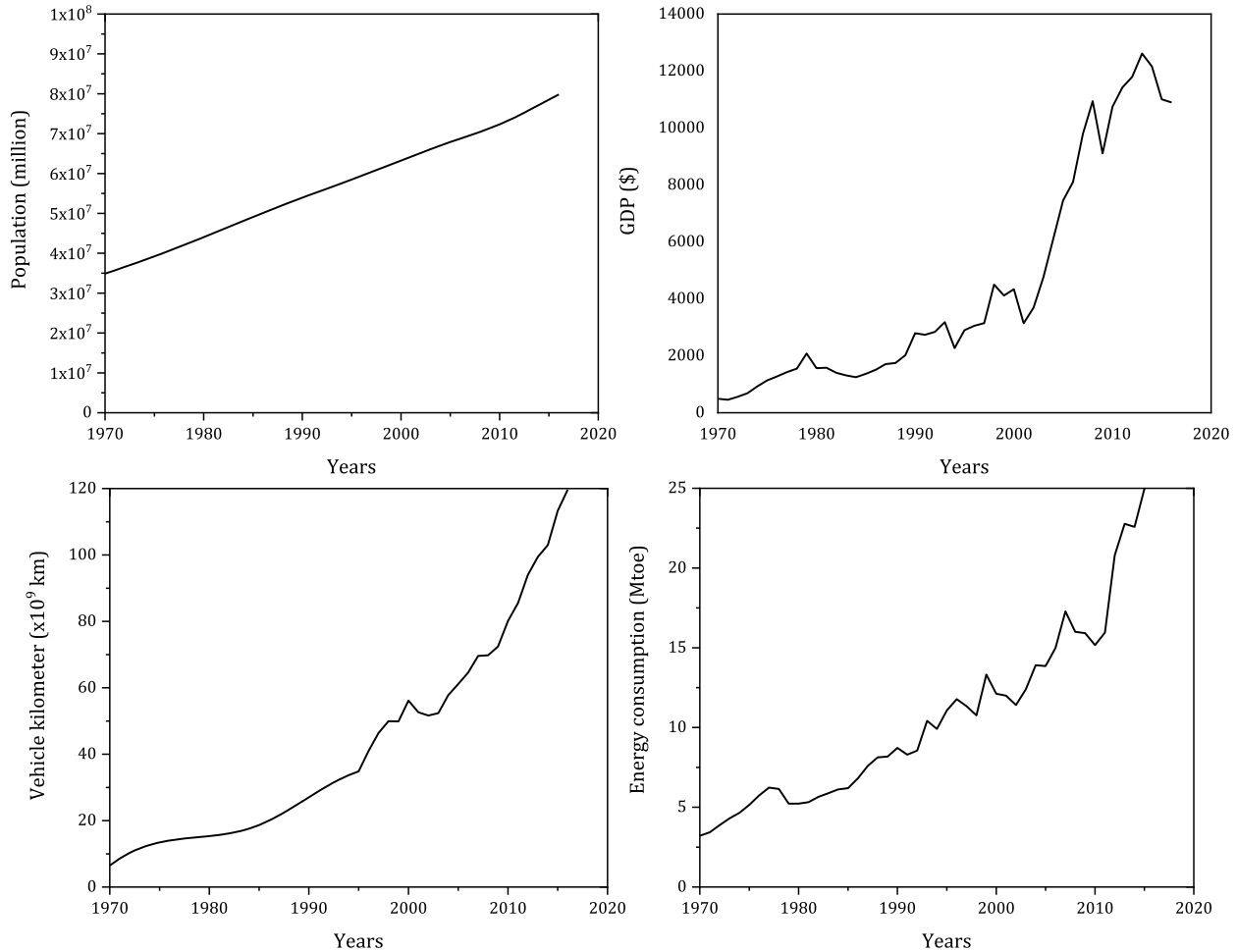
**Table 1.** Previous literature summary dealing with the forecasting of transportation-based energy and carbon emissions.

Reference	Region	Sector	Model	Dataset scale	Input	Output	Evaluation metrics
[32]	China	Transportation	PSO+GM	1990–2017	Historical CO <sub>2</sub> emissions	CO <sub>2</sub>	APE and MAPE
[33]	China	Transportation	MFO-ELM, PSO-ELM, CSO-SVM, PSO-SVM, PSO-BPNN, ELM, SVM, BPNN, and LSTM	2000–2017	13 different inputs	CO <sub>2</sub>	MAE, MAPE, RMSE, and SMAPE
[34]	Thailand	Transportation	Log-linear regression, path analysis, time series, and curve estimation	1989–2011	Population, GDP, and registered vehicles	CO <sub>2</sub>	MAPE
[35]	Different countries	Transportation	OLS, SVM, and GBR	1960–2020	Historical CO <sub>2</sub> emissions	CO <sub>2</sub>	$R^2$ , MAE, rRMSE, and MAPE
[10]	Türkiye	Transportation	ANN, DL, SVM, and curve estimation	1970–2016	GDP, population, vehicle kilometers, and year	CO <sub>2</sub> and energy demand	$R^2$ , RMSE, MAPE, MBE, rRMSE, and MABE
[36]	Iranian	All sectors	ARIMA and GM	1965–2010	Historical CO <sub>2</sub> emissions	CO <sub>2</sub>	RMSE, MAE, and MAPE
[37]	China	Cement industry	Hybrid GM	2018–2030	Historical CO <sub>2</sub> emissions	CO <sub>2</sub>	MAPE, RMSE, MAE
[38]	China	Industrial, building, transport, and agricultural sectors	CKC, and LMDI	1998–2015	Energy resources and GDP	CO <sub>2</sub>	$R^2$
[39]	China and India	Transportation, Building, Waste, and Manufacturing/Construction	DGM	2009–2017	Historical CO <sub>2</sub> emissions	CO <sub>2</sub>	MAPE, NMAPE, and RMSE
[40]	India	All sectors	EO, MPA, LSA, SOS, and BSA	1980–2018	Energy and economic indexes	All GHG emissions	$R^2$ , MBE, rRMSE, and MAPE
[41]	China, the United States, and India	–	ARIMA, BPNN, and MNGM	1990–2018	Historical CO <sub>2</sub> emissions	CO <sub>2</sub>	MAPE, MSE, and MSPE
[42]	India	Agricultural	SARIMAX, SVR, and H-W	1961–2018	Historical CO <sub>2</sub> emissions	CO <sub>2</sub>	MAPE, and MSE
[15]	Türkiye	Energy	DL, SVM, and ANN	1990–2018	Electricity production-based inputs	All GHG emissions	MBE, RMSE, rRMSE, $R^2$ , and MAPE

(Continued on next page)

**Table 1.** (Continued)

Reference	Region	Sector	Model	Dataset scale	Input	Output	Evaluation metrics
[43]	Türkiye	Energy	ARIMA	1970–2015	Energy production-based inputs	CO <sub>2</sub>	–
[44]	Türkiye	Energy	Empirical model	2001–2008	Electricity production-based inputs	CO <sub>2</sub>	$r$
[45]	Taiwan	Transportation	GM	1990–2006	Motor vehicle numbers	Energy consumption, and CO <sub>2</sub>	$E$
[46]	Türkiye	All sectors	GM	1999–2009	Historical CO <sub>2</sub> emissions	CO <sub>2</sub>	$E$
[27]	Türkiye	Transportation	ABC	1970–2013	GDP, vehicle number, population,	Energy demand	$R^2$ , MAE, RMS, and AE
[47]	Türkiye	Tourism	Empirical models	1960–2014	Tourist arrivals, energy consumption, and economic growth	CO <sub>2</sub>	$E$
[48]	Türkiye	All sectors	GWO-ANN	1990–2011	GDP, renewable energy production, energy consumption, population, and urbanization rate	GHG	RE, RMSE, MAE, $U$ -statistic, $R$ , and NSE
[49]	China	Transportation	SVR	1973–2017	GDP, energy intensity, energy consumption, population, urbanization rate, and proportion of the secondary industry	CO <sub>2</sub>	MSE



**Fig. 1.** The data used in this study.

data used in this study according to the years is shown in Figure 1. When Figure 1 is examined, it can be said that all input parameters have generally increased over the years. In addition, different scenarios for input parameters (scenario 1: ENERGY/VK/POP/Y/GDP, scenario 2: ENERGY/VK/POP/Y, scenario 3: ENERGY/VK/POP, and scenario 4: ENERGY/VK) were tested in this study and the results were compared. In this framework, gross domestic product per capita, population, energy consumption based on the transportation sector, and CO<sub>2</sub> emissions data were obtained from the World Bank [50]. The number of vehicles and kilometers in Türkiye were obtained from the Turkish General Directorate of Highways [51] and the Turkish Statistical Institute [52]. The input parameters, the data source used in this study, and their effects on the output parameters are presented in detail in Table 2. In addition, some important descriptive information about the data sets used in this study is given in Table 3.

Considering the differences between the size scales given in Table 3 for each parameter used in this study, it is not correct to present the data graphically. Therefore, it is more accurate to use a scaled form to make the data mathematically meaningful and comparable. Therefore, in this study, each parameter was used by normalizing between 0 and 1.

In this context, the equation presented below was used in the normalization of the data [53].

$$X_{\text{normalized}} = \frac{X - X_{\min}}{X_{\max} - X_{\min}}. \quad (1)$$

Here,  $X_{\text{normalized}}$  denotes  $X$  values normalized between 0 and 1.  $X_{\min}$  and  $X_{\max}$  indicate the minimum and maximum values, respectively, within the values in the dataset.

The act of standardizing data into a format that allows for fair feature comparisons and prevents problems with disparate scales is known as data normalization. Data normalization can reduce data redundancy, enhance data integrity, and improve data analysis. There are various methods of data normalization, such as rescaling, mean normalization, and standardization. Min-max normalization is one of the most popular methods of data normalization. It scales the range of features to a fixed interval, usually [0, 1] or [-1, 1], depending on the nature of the data. When the distribution's form and the precise values of the minimum and maximum are significant, min-max normalization is advantageous. It may not function effectively, however, if the dataset contains extreme values since it is vulnerable to outliers.

**Table 2.** Input parameters and their effects on the output parameters, along with data source used in this study.

Item	Effects on the output parameters	Source
Year	In light of historical data, it is understood that parameters such as energy demand and the number of vehicles will increase for countries.	–
Vehicle kilometer <sup>1</sup>	This parameter depends on the number of registered vehicles in the country. This value has increased over the years. For this reason, it is predicted that both energy consumption and CO <sub>2</sub> emissions will increase, depending on the number of vehicles in the transportation sector.	[52]
Population	It is a well-known fact that with the increase in population, the demand of countries for energy will increase. Therefore, it can be stated that, as a result of the increase in population, there will be an increase in the number of vehicles in the transportation sector and, accordingly, in the emission level.	[50]
Gross domestic product per capita	It is predicted that there is a high correlation between the CO <sub>2</sub> emissions originating from the transportation sector of the countries and the amount of production. It is also known that changes in the GDP parameter have a very significant effect on emissions.	[50]
Transportation energy consumption/demand	The amount of energy used in the transportation sector means the fuel used in vehicles. The increase in this value is directly related to the CO <sub>2</sub> emissions that occur as a result of combustion in the engines. The increase in the amount of energy originating from the transportation sector over the years means an increase in CO <sub>2</sub> emissions.	[50]

<sup>1</sup> Trucks (tankers and road tractors are also included), small trucks, buses, minibusses, passenger cars, and motorcycles are included. Administrative records collected from the odometers of vehicle inspection stations are computed.

**Table 3.** Some of the statistical parameters of the dataset used in this study.

Item	Year	Population (million)	GDP (\$)	Vehicle kilometer ( $\times 10^9$ km)	Energy consumption (Mtoe)	CO <sub>2</sub> ( $\times 10^6$ ton)
Observation	47	47	47	47	47	47
Maximum	2016	79821724	12614.48	6.48	25.88	68.619
Median	1993	56605264	4375.17	42.33	10.75	34.12
Minimum	1970	34876303	455.10	119.67	3.21	11.503
Difference	46	44945421	12159.38	113.19	22.67	57.116
Skewness	0.00	0.02	1.00	0.91	0.95	0.62
Kurtosis	−1.20	−1.15	−0.52	−0.09	0.31	−0.20
Standard deviation	13.71	1313 5678.4	3888.46	30.76	5.87	15.27
Standard error	2	1916 035.6	567.19	4.49	0.86	2.23

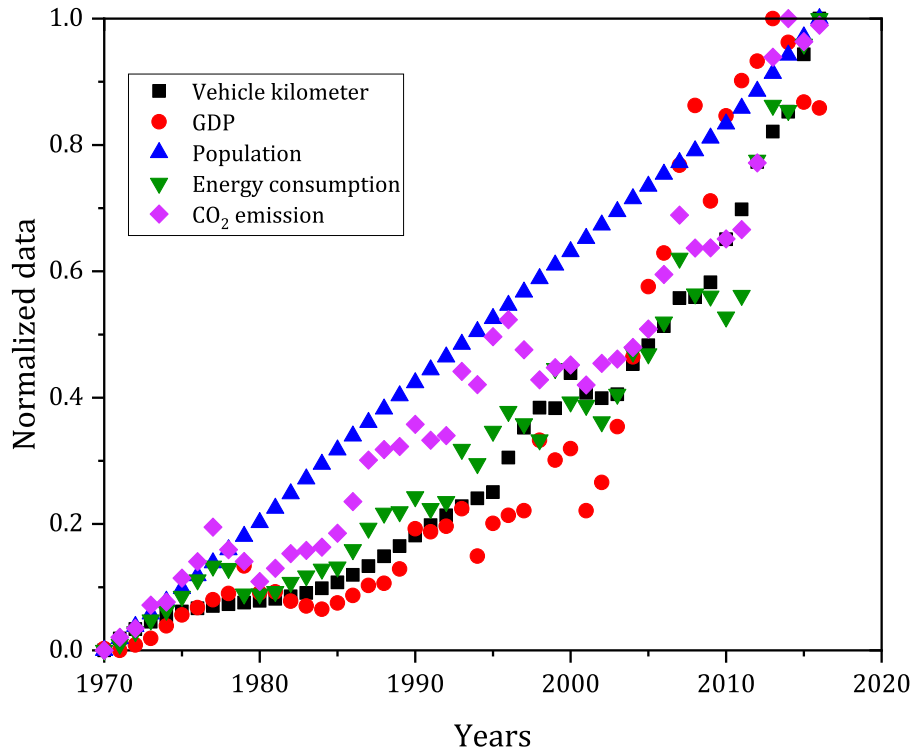
Table 4 shows the values that give the correlation indicators of each input parameter to the output parameter. The purpose of this was to provide the relationship between input and output parameters. When the table is examined, the minimum correlation value was found to be 0.89928, while the highest value was found to be 0.99971. In this context, the relationship between output and input parameters can be classified according to correlation coefficients. In other words, taking into account some studies in the literature [15, 54, 55], the comments related to the correlation coefficients can be expressed as follows:  $|r| < 0.2$  very weak relationship,  $0.2 \leq |r| < 0.4$  weak relationship,  $0.4 \leq |r| < 0.6$  moderate relationship,  $0.6 \leq |r| < 0.8$  strong

relationship, and  $|r| \geq 0.8$  very strong relationship. As a result, when the data and the relationship between each other are evaluated, it can be noted that there is a very strong connection between each input and output parameter. This situation states that it can be used to train ML algorithms to detect CO<sub>2</sub> emissions values originating from the transportation sector in Türkiye's current conditions.

The normalized values of the input and output parameters are shown in Figure 2. As can be seen from Figure 2, when the parameters are evaluated on the basis of years, it is understood that there are some decreases and increases. However, looking at the general trend, it can be said that the analyzed parameters are on an upward trend.

**Table 4.** Correlation indicators of variables with respect to each other used in the present study.

Item	CO <sub>2</sub> emission	Years	Vehicle kilometer	GDP	Population	Energy consumption
CO <sub>2</sub> emission	1	–	–	–	–	–
Years	0.95643	1	–	–	–	–
Vehicle kilometer	0.96875	0.94666	1	–	–	–
GDP	0.92063	0.90058	0.95352	1	–	–
Population	0.96002	0.99971	0.94980	0.89928	1	–
Energy consumption	0.98563	0.94117	0.98639	0.93364	0.94556	1

**Fig. 2.** Normalized data used in this study.

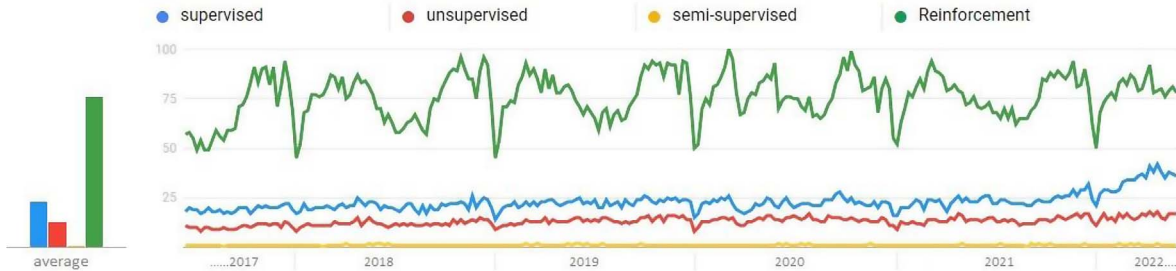
## 2.2 Artificial intelligence techniques

Artificial Intelligence (AI) is a technique used in areas such as ML, Natural Language Processing (NLP), optimization, image processing, etc. to enable computers to mimic human behavior, while ML is a subset of AI techniques that enable computer systems to learn from data observations [56]. In recent years, there has been an increase in the use of AI and ML techniques for data analysis and intelligent processing of information [57]. ML works in integration with smart hardware and automated systems that emerged with the Industry 4.0 revolution. Considering all these, it is very important to determine the basic factors affecting the quality of life, which is the most important feature of the systems and life cycle, with these algorithms. ML algorithms are grouped into four main categories. These are shown as supervised, unsupervised, semi-supervised, and reinforcement learning in the field, respectively. These algorithms are often used to classify data and apply regressions.

In Figure 3, the popularity values of these algorithms obtained from Google trends [58] in the last five years around the world are given.

The interest in AI algorithms continues to increase at a certain intensity every year with the development of technology. This is an indication that ML algorithms will not disappear from our lives in a short time. Non-traditional methods in complex engineering applications have represented an important class in recent years [59]. ANNs are very common and useful models used for classification, clustering, and prediction in different multidisciplinary studies. ANNs are competitive and useful models that are more preferred in ML than classical regression and statistical methods [60]. ML methods for generating effective and successful models in different application areas play an important role in classifying and predicting data. Common supervised neural networks are Single Layer Perceptron (SLP) [61], Multi-Layer Perceptron (MLP) [62], and linear classifiers [63]. Besides, Support Vector Machines (SVM)





**Fig. 3.** Popularity scores of ML algorithms worldwide over the past 5 years.

[64], kNN [65], naive bayes [66], and decision trees [67] are also examples of popular ANNs. In the literature, many algorithms that can be used in the field of ML have been proposed.

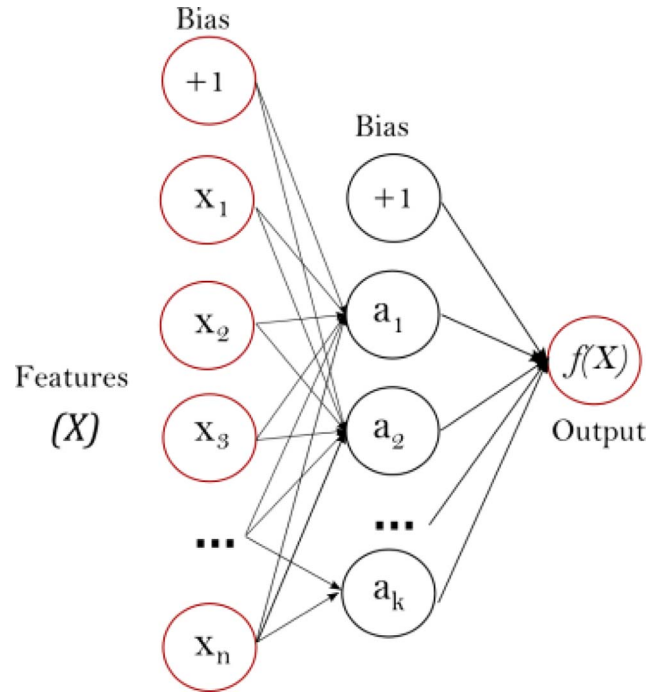
The characteristics of the algorithms used in the data set and the methods applied by these algorithms in the study are given below in detail.

### 2.2.1 MLP algorithm

MLPs are a common type of neural network used in different fields such as system load calculation, prediction of functional approaches, and analysis of complex systems [68]. An MLP can link multiple hidden layers between the input and output layers. In addition to being part of the MLP supervised learning method, it is a deep learning method. When distinguishing nonlinear data, the algorithm uses activation functions such as sigmoid and logistic functions [69]. In this way, it differs from logistic regression. Starting from the input layer, the data is propagated forward to the output layer. The error value is calculated based on the output obtained, and the error value is propagated back to minimize the error. The error value is calculated based on the difference between the known result and the predicted value, and the algorithm runs until this result is minimized. Figure 4 shows an MLP with an input, a scalar output, and a hidden layer. There are neurons in the input layer.  $\{x_i|x_1, x_2, x_3, \dots, x_m\}$  can be defined as input properties to the input layer. The neurons in the hidden layer weigh the values in the previous layer. In the output layer, the values from the hidden layer are converted to the output result. The MLP artificial neural network model is not prone to overfitting like deep neural networks [70]. In the experiments, the number of parameters (neurons) was kept low to prevent overfitting.

### 2.2.2 XGBoost algorithm

Gradient Boosting Machines (GBM) is a model of augmented ML in which a strong right learner uses a set of weak and base learners [71]. In boosting tree algorithms, the regression tree is created by the user as a weak learner and continues by dividing into two subsets at each level until the tree reaches the maximum depth. It is generally preferred to minimize the gradient loss function. The XGBoost algorithm is seen as the best of the decision tree-based algorithms. The XGBoost algorithm is the upper segment of the gradient boosting algorithm, which has been



**Fig. 4.** One hidden layer MLP.

improved with different arrangements, giving better results. XGBoost also calculates the quadratic gradients of the loss functions to minimize the loss in determining the best model, thus increasing performance by preventing overfitting [72]. In this study, the XGBoost algorithm was chosen for estimating CO<sub>2</sub> emissions due to its two main properties. First, XGBoost is one of the most popular boosting tree algorithms for GBM [73]. It is widely used in solving problems due to its maximum performance with minimum requirements in the industrial field [74]. The second main feature of the XGBoost algorithm is that it parallelizes the entire tree structure, especially during training over the CPU, and performs cache optimization of data structures and algorithms.

The XGBoost algorithm is more resistant to overfitting than deep learning algorithms. However, in order to prevent overfitting, data pre-processing and the use of fewer attributes are beneficial. Hyperparameter settings in the XGBoost algorithm also prevent overfitting. For this reason, all these factors were taken into consideration to prevent overfitting.

### 2.2.3 SVM algorithm

SVM is an important technique widely used in the classification and regression fields of ML. Because of this feature, SVM was also used in this study. With the hyperplanes it creates in the high or infinite-dimensional plane, SVM achieves low generalization error at high margins, thus demonstrating a strong distinction in each class [75]. It also has a property corresponding to a convex optimization problem in model determination, which is important for achieving optimal hyperplane parameters. Thus, the classes to which the new data belongs are separated by the optimal boundary line. Another important feature of SVM is that data sets cannot be separated by a line or plane. Classes are separated from each other by determining the optimal lines for each class to handle more complex data sets, such as non-separable and nonlinear datasets.

Of the 47 data used for each algorithm from 1970 to 2016, 85% were for educational purposes; and 15% were used for testing purposes. Test sets of algorithms are randomly selected and transferred to the classification network. The maximum number of iterations in the MLP network is set at 500. The model uses the backpropagation function, where the weight values are updated according to the behavior of the error function. In the study, while the information in the input cell is transferred to the intermediate layer, the relevant weights are multiplied, and the net input of each intermediate layer is calculated according to the weighted sum function. The output of the intermediate layer cells was calculated by passing the net input obtained through the activation function. Weights were updated by backward calculation, and the error rate was tried to be minimized. The flowchart of the processes carried out in this study is shown in Figure 5.

The SVM algorithm is resistant to overfitting. Overfitting is not as common as in artificial neural networks. But data normalization – adjusting the C value and the seed value – helps prevent overfitting. Additionally, using fewer features also prevents overfitting.

## 3 Statistical indicators

The accuracy of a model gives an idea of how well it can predict, but that does not necessarily mean how well the model will perform in real-world conditions. The fit (suitability) of a model, on the other hand, is about whether the model fits the characteristics of the data. Accuracy and suitability are important for model selection and performance evaluation. Therefore, in the current study, the performance success of the prediction results generated by ML algorithms (MLA) is discussed extensively with the following statistical parameters that are frequently used in academic studies: In equations (2)–(8), the indicator results can be calculated. The variables in these equations are:  $n$  represents the number of data points,  $\sum$  the sum of all data points,  $x_i$  the actual measured value,  $y_i$  the predicted value, and  $\bar{x}_i$  the mean of the partition results.

Mean absolute error (MAE) is an indicator that is particularly useful when evaluating models that are susceptible to systematic errors. It offers a simple measure to evaluate

the overall accuracy of the model by measuring the average magnitude of the deviation between the predicted values and the actual values [76]. In the model success comparison, a model with a smaller MAE value is considered more successful.

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^n |y_i - x_i|. \quad (2)$$

Mean absolute percentage error (MAPE) is calculated as the average of absolute percentage errors between actual and predicted values, expressed as a percentage. MAPE is a useful metric that provides a percentage measure of the accuracy of a forecast. In this way, datasets can be easily compared and interpreted. However, it is critical to interpret the results carefully and use MAPE in conjunction with other metrics because of its limitations, such as being misleading when the actual values are close to zero [77].

$$\text{MAPE, \%} = \frac{1}{n} \sum_{i=1}^n \left| \frac{x_i - y_i}{x_i} \right| \times 100. \quad (3)$$

Mean Bias Error (MBE) is a metric whose statistical analysis is used to evaluate the performance of a model. A negative MBE result indicates that the estimates are consistently lower than the actual values, while a positive result indicates that the estimates are consistently higher than the actual values [78, 79].

$$\text{MBE} = \frac{1}{n} \sum_{i=1}^n (y_i - x_i). \quad (4)$$

The Mean Square Error (MSE) statistical indicator is the square of the difference between the actual and predicted values, and it is required to be minimal by training algorithms in prediction studies [80].

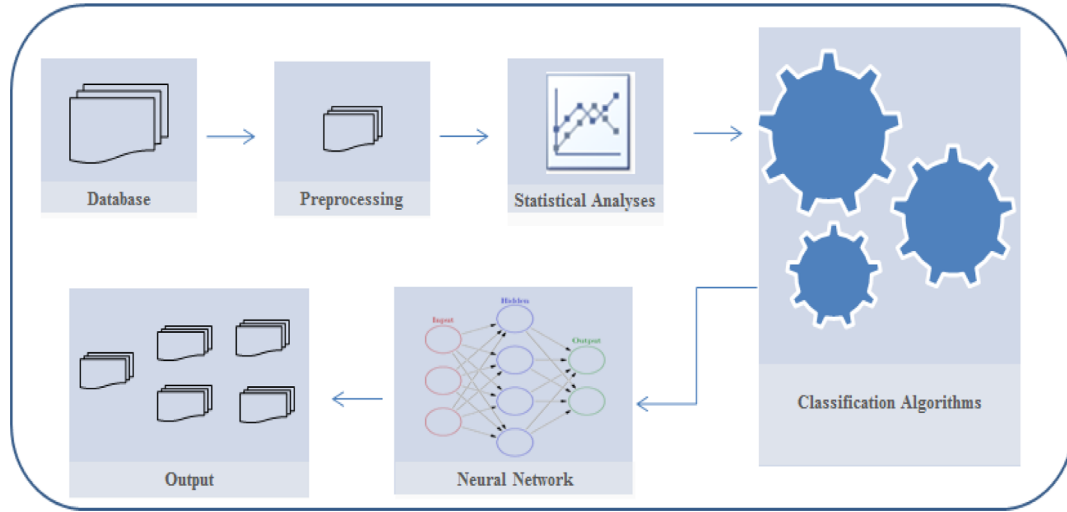
$$\text{MSE} = \frac{1}{n} \sum_{i=1}^n (y_i - x_i)^2. \quad (5)$$

The coefficient of determination ( $R^2$ ) is a statistical measure used to evaluate how fit a regression model is.  $R^2$  represents the percentage of the independent variables' explanatory power for the dependent variable.  $R^2$  takes values between 0 and 1 [81]. The closer the value of the model is to 1, the better the fit [82].

$$R^2 = 1 - \frac{\sum (y_i - x_i)^2}{\sum (x_i - \bar{x}_i)^2}. \quad (6)$$

Root Mean Square Error (RMSE) is a performance indicator used to measure how far a model's predictions are from the actual values. The RMSE indicates that (i) the predicted values are closer to the actual values; (ii) the "0" ideal value; and (iii) the model with the lower value performs better. Also, the results must be positive [83]. The fact that RMSE is sensitive to larger errors as well as outliers makes it a good measure of the model's performance success [84].

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - x_i)^2}. \quad (7)$$



**Fig. 5.** Flowchart of the study.

Relative root mean square error (rRMSE) is a modified version of RMSE that takes into account the scale of the measured data. The rRMSE is calculated as the ratio of the RMSE to the mean of the measurement results [85]. rRMSE normalizes RMSE by data range to allow a more meaningful comparison of the accuracy of different models.

$$\text{rRMSE, \%} = \frac{\sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - x_i)^2}}{\bar{x}_i} \times 100. \quad (8)$$

The a20 metric quantifies the percentage of instances when the absolute deviation between actual and predicted values is equal to or less than 20% of the actual value. A higher a20 score indicates better prediction accuracy, as the model can make more accurate predictions that are closer to the true values. The best possible score is 1.0 [86, 87].

$$\text{a20 index, \%} = \frac{\sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - x_i)^2}}{\bar{x}_i} \times 100 \quad (9)$$

$$\text{a20 index, \%} = \frac{1}{n} \sum_{i=1}^n \begin{cases} 1, & \text{if } \frac{|\hat{y}_i - y_i|}{y_i} \leq 0.2 \\ 0, & \text{otherwise} \end{cases}. \quad (10)$$

## 4 Result and discussion

In this section, the findings and discussions obtained within the scope of the study are presented under sub-headings. First, a general year-by-year discussion of CO<sub>2</sub> emissions from the transportation sector is planned to be presented. In this context, the reasons behind the increase in CO<sub>2</sub> emissions were tried to be explained. Thus, it will be possible to understand how the analysis is followed. Then, the results of analyses performed in light of 3 different ML algorithms and 3 different scenarios described in the above

sections are shared. The results of this analysis were discussed by taking into account the most commonly used statistical indicators in the literature, and the most appropriate algorithm and scenario were determined for this research study.

### 4.1 The outlook of transportation-based CO<sub>2</sub> emissions in Türkiye

In addition to the effects of the transportation sector on the economic development of countries, it also has negative effects on air quality. This is particularly problematic for developing countries. Today, the basic energy needs of the transportation sector are supplied by fossil-based fuels. For this reason, it is a known fact that many deaths occur worldwide due to diseases associated with air pollution. It can be clearly stated that knowing the CO<sub>2</sub> emissions values, which depend on many parameters along with energy demand, is important for countries to turn and update their policies and investments in this direction in the coming years. One of the most effective techniques within the scope of combating climate change is setting a price for CO<sub>2</sub>. It can be pointed out that this case appears to be the most appropriate theory. Unfortunately, carbon prices are still relatively low in many countries in order to struggle with climate change, as stated above. In other words, carbon prices have an inefficient level. On the other hand, it is stated that this is not the end, based on the report published by the Organization for Economic Cooperation and Development (OECD). The energy taxes are not designed in light of the carbon price. But they can contribute to the influence of the emissions trading systems as well as CO<sub>2</sub> emissions. Meanwhile, the OECD has computed effective carbon prices considering the aforesaid impacts for more than four major countries worldwide. In addition, the effective CO<sub>2</sub> price is mapped for the purpose of presenting which countries have the highest effective carbon price [88]. The EU carbon permits data chart is plotted in Figure 6. As observed, the highest EU carbon permit was €104.80 in

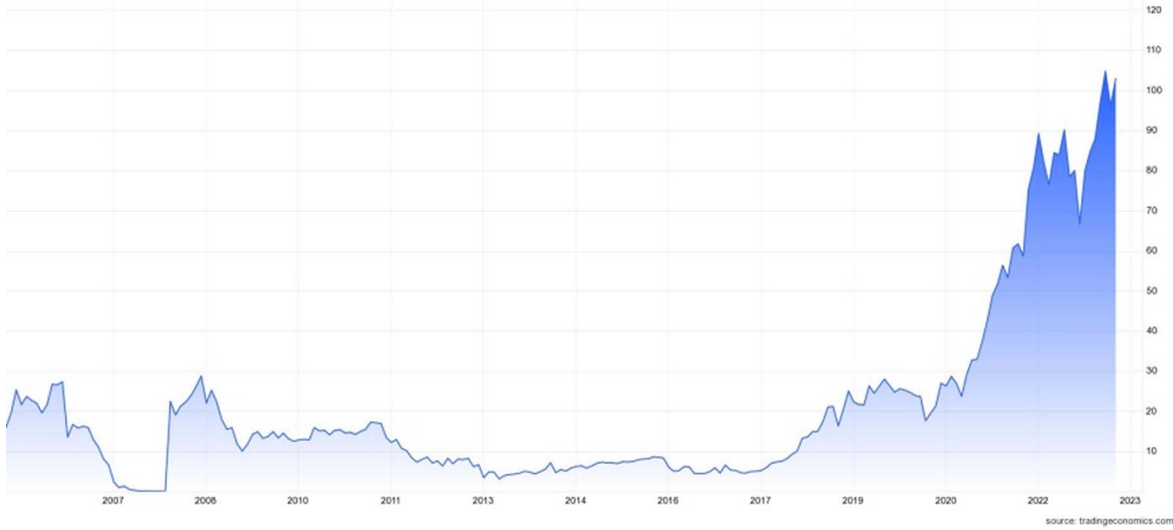


Fig. 6. EU carbon permits data chart [89].

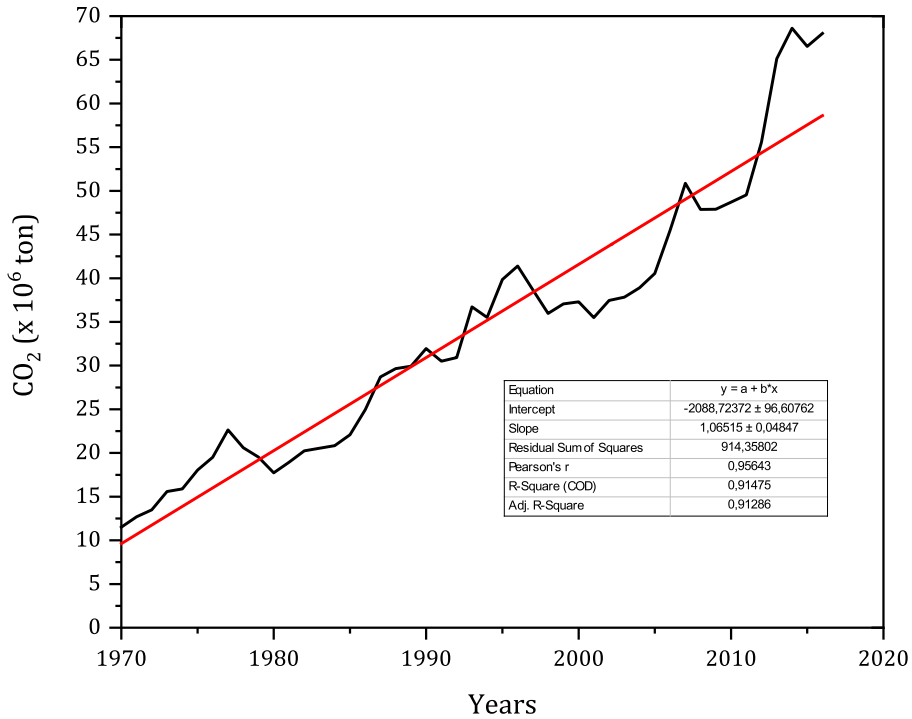
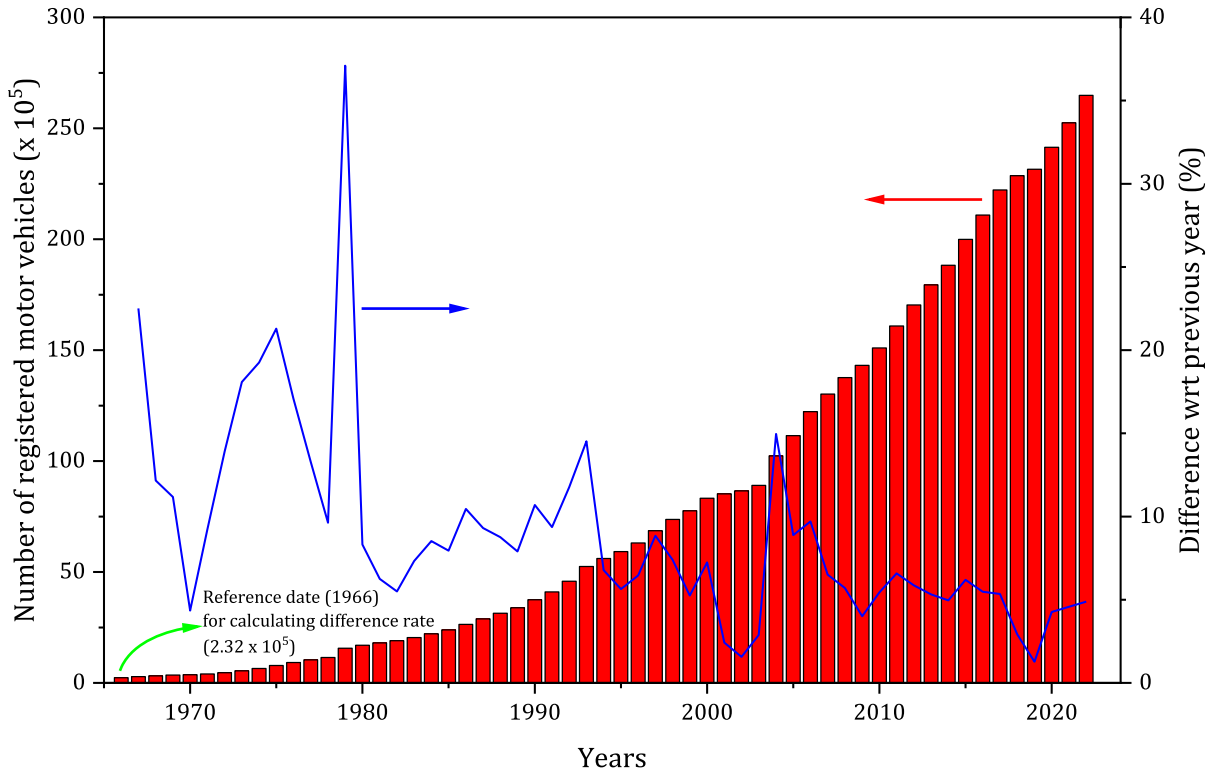


Fig. 7. Change of CO<sub>2</sub> emissions from transportation between 1970 and 2016 in Türkiye.

February 2023. The EU carbon permits data chart is plotted in Figure 6. The highest EU carbon permit was €104.80 in February 2023. As observed, carbon prices have increased sharply in recent years. Growing environmental concerns and imposing strict restrictions are very important. In this perspective, there is a need for restrictions, regulations, norms, legislations, and wide-ranging policies that are strongly recommended to be supported to reduce

fossil-based fuel consumption in the transportation sector and thus reduce GHGs.

The variation in CO<sub>2</sub> emissions from transportation in Türkiye between 1970 and 2016 is given in Figure 7. As can be seen from the graph, although CO<sub>2</sub> emissions have increased and decreased over the years, it should be said that they have increased when a general evaluation is made. This upward trend in CO<sub>2</sub> emissions originating from the



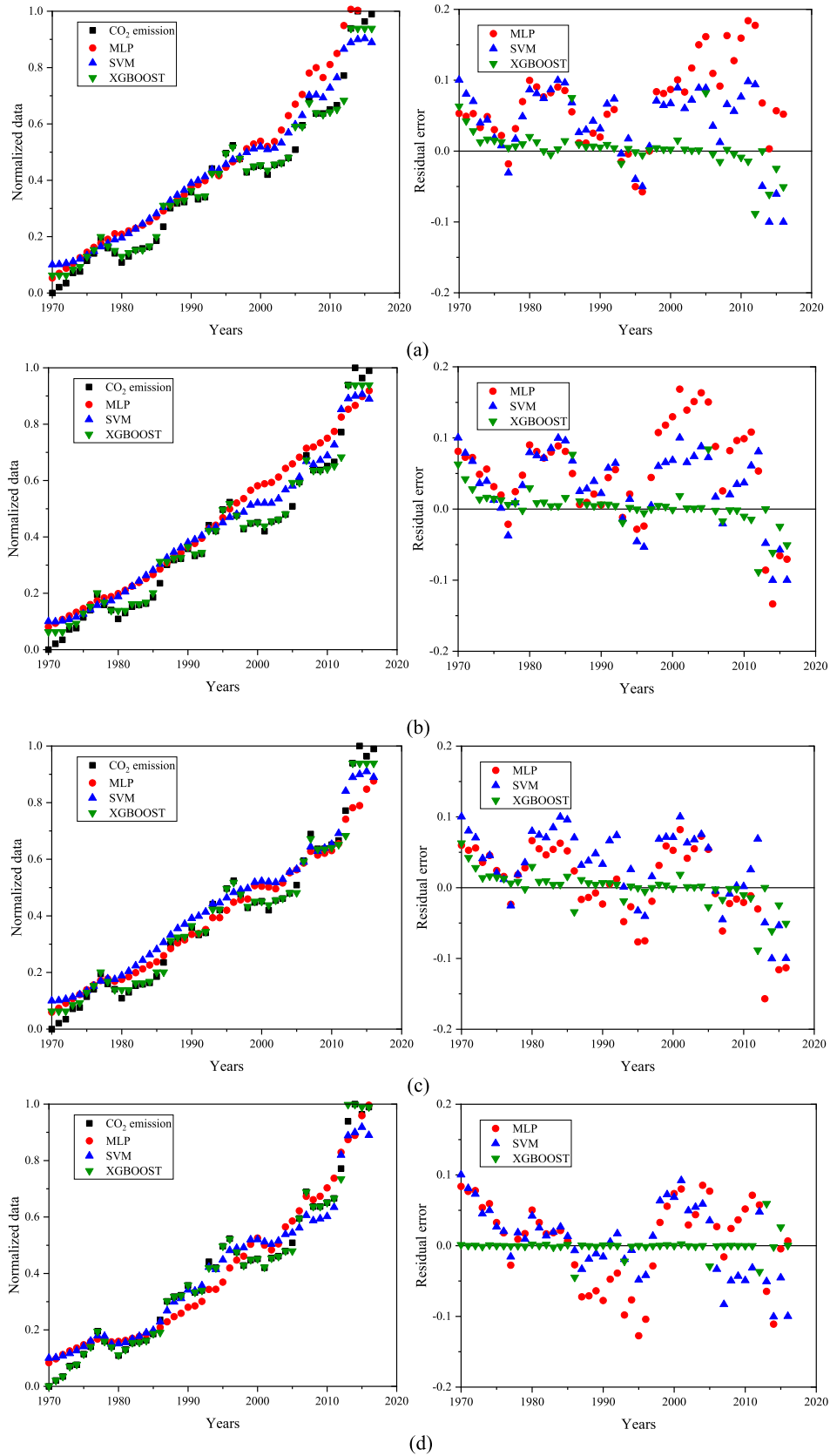
**Fig. 8.** The variation of number of registered motor vehicles in Türkiye according to the year (The graph was plotted based on the data coming from the ref [52]).

transportation sector, of course, depends on many parameters. It can be said that these are population, energy consumed in the transportation sector, an increase in the number of vehicles, etc. This is very important for understanding the relationship between the parameters. In addition, when Table 4 is examined, it is understood that vehicle kilometer information, population, GDP, and energy consumption values used as input parameters provide a very strong correlation with CO<sub>2</sub> emissions. In other words, vehicle kilometers were correlated at 0.96875, GDP at 0.92063, population at 0.96002, and energy consumption at 0.98563.

When the CO<sub>2</sub> emissions originating from the transportation sector in Türkiye are examined, the value that was  $11.503 \times 10^6$  tons in 1970 reached  $68.016 \times 10^6$  tons at the end of 2016 [50]. From 1970 to 2016, vehicle kilometers increased by 18.47 times, GDP by 22.24 times, population by 2.29 times, and energy consumption by 8.06 times. In general, when the change in CO<sub>2</sub> emissions over the years is evaluated, it can be stated that there is an approximate linear trend. A linear curve is drawn in Figure 4, and the equation of the curve is also presented to show this situation more clearly. On a yearly basis, CO<sub>2</sub> emissions increased by approximately 18.778% in 1993 compared to the previous year. In the same year, energy consumption increased by 21.871%, population by 1.623%, GDP by 11.885%, and vehicle kilometers by 5.282%. In contrast, the largest reduction occurred in 1980, with approximately 9.221%. In the same year, energy consumption did not change; the population increased by 2.284%, GDP decreased by 24.768%, and

vehicle kilometers increased by 1.927%. These results show that GDP and energy consumption have significant effects on CO<sub>2</sub> emissions. Due to the rapidly increasing population, socioeconomic developments, and modernization have increased the burden on the transportation sector. This situation causes an increase in CO<sub>2</sub> emissions. According to the data of the Turkish Statistical Institute, as of December 31, 2022, the population of Türkiye increased by 599,280 people compared to the previous year and increased to over 85 million [52]. In addition, it is stated that a large part of this population consists of people of working age. As is known, Türkiye is a developing country. Considering the GDP data, there was a 3.5% growth in the last quarter of 2022 and a 5.6% growth in the whole of 2022 compared to 2021. While the per capita income was \$9,693 in 2018, it is expected to exceed \$12,000 in 2023.

The increase in the data described above directly affects the number of road vehicles in Türkiye. The number of motor vehicles registered to traffic in Türkiye over the years is presented in Figure 8. Based on the data obtained from TurkStat, the number of registered motor land vehicles in traffic in Türkiye, which was over 230 thousand in 1966, will increase to over 26 million by the end of 2022. When the data from the last decade is taken into consideration, the number of registered motor land vehicles in Türkiye has increased by 47.62%. However, the number of registered motor land vehicles in Türkiye grew by an average of 8.98% annually. While the highest annual increase was in 1979, the lowest increase was in 2019 with 1.27%. In addition, 53.88% of the total number of registered vehicles in Türkiye



**Fig. 9.** Prediction of transportation-based CO<sub>2</sub> emissions using various ML algorithms (a) scenario 1, b) scenario 2, c) scenario 3, d) scenario 4).

**Table 5.** Statistical indicators of XGBoost, SVM, and MLP algorithms for scenario 1 to predict CO<sub>2</sub> emissions.

Features	Algorithms	Statistical indicators	Results
ENERGY/ VK/ POP/Y/GDP	XGBoost	$R^2$	0.9731
		MAE	0.0366
		MAPE	0.0866
		MSE	0.0026
		RMSE	0.0514
		rRMSE	7.0159
		MBE	0.0042
		a20	0.8937
		SVM	$R^2$
	MAE		0.0657
	MAPE		0.2179
	MSE		0.0050
	RMSE		0.0711
	rRMSE		15.0994
	MBE		0.0408
	a20		0.6809
	MLP		$R^2$
		MAE	0.0865
		MAPE	0.2457
		MSE	0.0101
		RMSE	0.1007
		rRMSE	18.6294
		MBE	0.0651
		a20	0.5107

**Table 6.** Statistical indicators of XGBoost, SVM, and MLP algorithms for scenario 2 to predict CO<sub>2</sub> emissions.

Features	Algorithms	Statistical indicators	Results
ENERGY/ VK/ POP/Y	XGBoost	$R^2$	0.9722
		MAE	0.0383
		MAPE	0.0953
		MSE	0.0027
		RMSE	0.0522
		rRMSE	7.1228
		MBE	0.0045
		a20	0.8724
		SVM	$R^2$
	MAE		0.0621
	MAPE		0.2101
	MSE		0.0043
	RMSE		0.0656
	rRMSE		14.2356
	MBE		0.0335
	a20		0.7022
	MLP		$R^2$
		MAE	0.0711
		MAPE	0.2171
		MSE	0.0064
		RMSE	0.0801
		rRMSE	18.5797
		MBE	0.0519
		a20	0.5320

consists of automobiles. When the distribution of automobiles by fuel type is examined, it is reported that 26.8% are gasoline, 36.9% are diesel, 35.1% are LPG and the rest are hybrid, electric, etc.

Fossil-based fuels like coal, oil, and natural gas constitute the base item of energy consumed in basic industries such as trade, industry, housing, and transportation, and oil ranks first among the above-mentioned fuel types. When the global primary energy consumption rates of 2021 are evaluated, fossil fuels constitute 83.4% of total consumption. When the foreign dependency data of Türkiye between 1990 and 2020 were evaluated, the foreign dependency ratio, which was 51.6% in 1990, reached 70.1% in 2020. When Türkiye's primary energy consumption is evaluated on the basis of sectors, it is stated that 18.3% is used in the transportation sector [90]. Oil consumption, which was 313.2 thousand b/d in 1980, nearly tripled and reached 922.6 thousand b/d in 2020. As can be seen from Figure 5, it can be predicted that there will be an increase in the number of registered motor land vehicles in Türkiye over the years. This shows that Türkiye's oil consumption is increasing very rapidly. In other words, the energy demand of the transportation sector will undoubtedly boost with the augmentation in the number of vehicles. Having realized 50,127 b/d of crude oil production in 2012, Türkiye

produced 69,332 b/d of crude oil in 2021. On the other hand, while Türkiye imported 385 thousand b/d of crude oil in 2012, this value increased to 631 thousand b/d in 2021. The fact that a large part of the energy need is met from oil and its derivatives and the oil reserves are constantly decreasing leads humanity to long-lasting, non-depleting, clean, and environmental sources. Considering environmental factors such as acid rain, global warming, and climate change due to the effects of carbon dioxide and GHG formed as a result of the use of fossil fuels, the importance of alternative energy sources whose effects on the environment can be controlled is gradually increasing. In particular, it can be stated that Türkiye, which meets more than 90% of its oil needs from other countries through imports, is a net importer in terms of energy resources. This situation not only hinders the economic growth of the country but also raises environmental issues. Another significant point is the emergence of the necessity of increasing the tendency towards alternative and renewable energy sources to meet Türkiye's needs in terms of energy resources [90, 91].

As a result of the burning of fossil fuels in internal combustion engines, many harmful pollutants are released into the atmosphere, and therefore, these gases seriously threaten the health of humans and other living things [92]. In addition, the use of petroleum-based fuels significantly

**Table 7.** Statistical indicators of XGBoost, SVM, and MLP algorithms for scenario 3 to predict CO<sub>2</sub> emissions.

Features	Algorithms	Statistical indicators	Results
ENERGY/ VK/POP	XGBoost	$R^2$	0.9862
		MAE	0.0259
		MAPE	0.0588
		MSE	0.0013
		RMSE	0.0367
		rRMSE	6.0759
		MBE	-0.0002
	SVM	a20	0.8937
		$R^2$	0.9614
		MAE	0.0569
		MAPE	0.2009
		MSE	0.0038
		RMSE	0.0616
		rRMSE	14.0604
		MBE	0.0329
		a20	0.6596
		MLP	$R^2$
	MAE		0.0661
	MAPE		0.1606
	MSE		0.0062
	RMSE		0.0789
	rRMSE		15.5542
	MBE		0.0013
	a20		0.7235

**Table 8.** Statistical indicators of XGBoost, SVM, and MLP algorithms for scenario 4 to predict CO<sub>2</sub> emissions.

Features	Algorithms	Statistical indicators	Results
ENERGY/ VK	XGBoost	$R^2$	0.9886
		MAE	0.0278
		MAPE	0.0576
		MSE	0.0011
		RMSE	0.0333
		rRMSE	3.4950
		MBE	-0.0010
	SVM	a20	0.9788
		$R^2$	0.9883
		MAE	0.0298
		MAPE	0.0648
		MSE	0.0011
		RMSE	0.0338
		rRMSE	12.2321
		MBE	0.0072
		a20	0.8298
		MLP	$R^2$
	MAE		0.0455
	MAPE		0.1076
	MSE		0.0030
	RMSE		0.0552
	rRMSE		14.6818
	MBE		0.0057
	a20		0.7235

increases the effects of GHG, leading to global warming [93]. To overcome global warming, one of the biggest environmental concerns of the century, CO<sub>2</sub> emissions need to be reduced [94]. In light of all the situations and information described in this section, it is understood that the energy used in the transportation sector has a great impact on both the environment and the country's economy. As a result, this should be taken into account for future investments. In addition, the above-mentioned problems can be overcome by concrete steps to be taken in the short, medium, and long term.

#### 4.2 Prediction of transport-based CO<sub>2</sub> emissions using several ML algorithms

In this study, supervised learning models were used; this is because algorithms or models learn from labeled data. It is an important ML technique used, especially in classification and regression studies. In this study, the XGBoost ensemble learning algorithm, which is the most advanced of the boosting algorithms used in both classification and regression studies, was used. To see the response of the system to the new samples, the MLP algorithm was chosen. In addition, the SVM was included in our analysis due to its

high accuracy in regression studies with ML. A comprehensive study is presented in which a data-driven system was tested with ML algorithms. In the study, the data were analyzed using the regression technique. The regression technique is a very popular method among ML modeling techniques. The relationship between the dependent variable and the independent variables was examined by using the most appropriate straight line obtained in the applied regression method. In classification, a linear line is considered a boundary separating two classes, while the line created in regression reflects the linear relationship between two variables. MLP, on the other hand, is the most popular type of artificial neural network used in solving nonlinear problems [95].

In this analysis, four different scenarios were created with three different ML methods. These are scenario 1: ENERGY/VK/POP/Y/GDP, scenario 2: ENERGY/VK/POP/Y, scenario 3: ENERGY/VK/POP, and scenario 4: ENERGY/VK conditions of the input parameters as reported in the above sections. Within the framework of these scenarios, CO<sub>2</sub> emissions from the transportation sector were estimated between 1970 and 2016. Figure 9 shows the estimated CO<sub>2</sub> emissions over the operating range, the actual values, and the estimation error that



occurred. As observed from the graph, CO<sub>2</sub> emissions from the transportation sector are generally increasing in Türkiye. It is not difficult to say that the estimated CO<sub>2</sub> emissions using ML techniques are very close to the actual values. In other words, the estimated and actual CO<sub>2</sub> emissions are very close to each other.

The ML algorithms used in this study and the statistical indicators obtained in line with the planned scenarios are presented in detail in Tables 5–8, respectively, for better comparison. Prediction success was calculated using the target value and the obtained result value. To measure the success of the predictions, the success prediction ratio,  $R^2$ , MAE, MAPE, MSE, RMSE, rRMSE, and MBE methods were used. The equations of the relevant statistical indicators are explained in detail in Chapter 3. When the tables are evaluated, each statistical indicator should be found at certain intervals. For example,  $R^2$  should be in the range of 0–1, RMSE 0–5, rRMSE 0–15%, MAPE 0–15%, MABE 0–5, and MBE 0–3. Considering all the results, each ML algorithm gave very satisfactory results in the estimation of CO<sub>2</sub> emissions from transportation in Türkiye.

The average MSE value in the study is ideal because it is very close to 0. It is desired that the RMSE value be close to zero (in the range of 0.03–0.1 in this study), and it is seen that it is very close to the desired value. It is also desired to have the smallest values for MAE and MAPE, and in this study, it is seen that the values are very close to the desired values in the two-featured XGBoost algorithm with 0.0278 and 0.0576, respectively. When the scenarios created were examined, it was seen that the XGBoost algorithm was the most successful. The highest  $R^2$  value (0.9886) of the XGBoost algorithm was reached when the VK and ENERGY features were used in the 4th scenario. It was observed that the MLP algorithm reached the highest value of 0.9689 and the SVM algorithm reached the highest  $R^2$  value of 0.9883 in the 4th scenario. For the limit values stated above, while the  $R^2$  value of the XGBoost algorithm in the 4th scenario gave the closest result to 1, it showed that it was the most appropriate algorithm with the values of RMSE 0.0333, rRMSE 3.4950%, MAPE 0.0576, MAE 0.0278, and MBE –0.0010.

## 5 Conclusion

Within the scope of this study, is aimed at estimating the CO<sub>2</sub> emissions originating from the transportation sector in Türkiye with different artificial intelligence algorithms. Three different ML algorithms (XGBoost, SVM, and MLP) were used to estimate CO<sub>2</sub> emissions. In the training of the algorithms, the correlation method was preferred in determining the input parameters, and the parameters that provided the most correlation with the output parameter were determined. These are ENERGY > VK > POP > Y > GDP, respectively. Accordingly, four different scenarios were prepared, and the effects of parameter groups on CO<sub>2</sub> emissions were tried to be determined using various statistical indicators. In other words, seven different statistical measurements were used to determine the prediction

performance of the algorithms. In conclusion, based on this research, the following important findings can be listed:

- It was determined that there was a strong correlation between energy consumption from transportation (0.98563), vehicle kilometers (0.96875), population (0.96002), year (0.95643), economic indicators (0.92063), and CO<sub>2</sub> emissions.
- Compared to today, when the metrics mentioned above and selected as input parameters within the scope of this study are evaluated, it is observed that there will be increases in the coming years.
- When all the results obtained were evaluated, it was observed that the XGBoost algorithm came to the forefront in estimating the CO<sub>2</sub> emissions from the transportation sector, while the other algorithms generally gave lower results. However, when all metrics are considered, there is no doubt that each ML algorithm offers very successful results in predicting CO<sub>2</sub> emissions outputs.
- It was seen that the best scenario for estimating CO<sub>2</sub> emissions according to four different scenarios planned is number 4 and all ML algorithms produce very successful results when all parameters are accepted as inputs.
- It should be said that all algorithms give successful results in the estimation of CO<sub>2</sub> emissions according to  $R^2$ , which is one of the statistical indicators widely used in the literature.
- According to the MAPE indicator, it was determined that all algorithms other than the XGBoost algorithm (MAPE = 0.0576) gave low prediction accuracy in CO<sub>2</sub> emissions estimation.
- It has been seen that Artificial Intelligence Technologies can be used in the estimation of CO<sub>2</sub> emissions as in every other field.

In summary, it is clear that to sustain the momentum in Türkiye's economic development in recent years, decision-makers need to review energy policy to a large extent. Otherwise, the biggest obstacle to economic development will be the energy problem. In addition, due to the limited energy resources, the energy demand method is necessary for the more efficient use of energy resources in the future. This will continue to be an important argument for CO<sub>2</sub> emissions. As a result, in the continuation of this study, the future of parameters such as future energy demand, number of vehicles, and CO<sub>2</sub> emissions can be determined by making predictions with forward-looking time-based algorithms, and measures can be taken within this scope. Although there are estimation studies for the future in the literature, it is not difficult to say that they are insufficient. As a result, the different ML algorithms preferred in this paper have provided very good results in estimating CO<sub>2</sub> emissions from the transportation sector in Türkiye. However, the forecasted results for the future displayed that some serious initiatives have to be taken by policymakers in the near future to determine and mitigate the growth rate of the increases in the above-mentioned parameters. It is thought that the results obtained from this study will contribute to the development of investment plans for public

institutions. In addition, it is planned to estimate CO<sub>2</sub> emissions with different deep learning architectures and features in future studies.

## 6 Limitations and future work

This study has some limitations that should be acknowledged and addressed in future work. First, this study applied three artificial intelligence algorithms (MLP, XGBoost, and SVM) to estimate CO<sub>2</sub> emissions in Türkiye's transportation sector. Other algorithms (Random Forest, K-Nearest Neighbors, etc.) could also be explored and compared for their performance and suitability for this task. Second, this study considered five input parameters (energy consumption, vehicle kilometers, population, year, and gross domestic product per capita) that were assumed to have a linear relationship with CO<sub>2</sub> emissions. However, there could be other factors, such as fuel types, traffic patterns, weather conditions, or vehicle types, that could have a nonlinear or interactive effect on CO<sub>2</sub> emissions. Future research could incorporate these additional data sources to further refine CO<sub>2</sub> emissions predictions and potentially identify specific emission hotspots. Third, this study designed four scenarios based on the correlation effect of the input parameters. However, there could be other scenarios that reflect different assumptions, projections, or policy interventions that could affect CO<sub>2</sub> emissions in the transportation sector. Future research could develop scenario-based models that simulate the effects of various policy interventions, such as carbon taxes, subsidies, regulations, or incentives, on CO<sub>2</sub> emissions and inform the design of effective emission mitigation strategies. Fourth, this study was restricted to Türkiye, which has a specific geographic, economic, and social context that could influence CO<sub>2</sub> emissions in the transportation sector. Therefore, the findings of this study may not be generalizable or applicable to other regions or countries that have different characteristics or challenges. Future research could extend the scope of this study to other regions or countries and compare the results and implications for CO<sub>2</sub> emissions in the transportation sector.

### Acknowledgments

The authors would like to thank the Editors and anonymous reviewers for helping us to present a balanced account of our paper.

### Conflict of interest

The authors declared that there is no competing financial interest in this research.

### Data availability statement

The data used and/or analyzed throughout the present study are available from the authors upon reasonable request.

### Authors' contributions statement

Gökalp Çınarlar and Kazım Kılıç: Methodology, Formal analysis, Visualization, Writing – original draft, Writing – review & editing. Murat Kadir Yeşilyurt and Ümit Ağbulut: Conceptualization, Investigation, Methodology, Data curation, Validation, Resources, Formal analysis, Visualization, Writing – original draft, Writing – review & editing. Zeki Yılbaş:

Investigation, Visualization, Writing – original draft, Writing – review & editing.

### Ethics approval

The authors declared that no animal and human studies are presented in this manuscript and no potentially identifiable human images or data are given in this research.

## References

- Nundy S., Ghosh A., Mesloub A., Albaqawy G.A., Alnaim M.M. (2021) Impact of COVID-19 pandemic on socio-economic, energy-environment and transport sector globally and sustainable development goal (SDG), *J. Clean. Prod.* **312**, 127705. <https://doi.org/10.1016/j.jclepro.2021.127705>.
- Gulseven O., Al Harmoodi F., Al Falasi M., Alshomali I. (2020) *How the COVID-19 pandemic will affect the UN sustainable development goals?* SSRN, Available at <https://doi.org/10.2139/ssrn.3592933>.
- Sun L., Zhang T., Liu S., Wang K., Rogers T., Yao L., Zhao P. (2021) Reducing energy consumption and pollution in the urban transportation sector: a review of policies and regulations in Beijing, *J. Clean. Prod.* **285**, 125339. <https://doi.org/10.1016/j.jclepro.2020.125339>.
- Zhang M., Li H., Zhou M., Mu H. (2011) Decomposition analysis of energy consumption in Chinese transportation sector, *Appl. Energy* **88**, 6, 2279–2285. <https://doi.org/10.1016/j.apenergy.2010.12.077>.
- Singh A.P., Kumar D., Agarwal A.K. (2021) Introduction to Alternative Fuels and Advanced Combustion Techniques as Sustainable Solutions for Internal Combustion Engines, Singh A.P., Kumar D., Agarwal A.K. (eds.), *Alternative fuels and advanced combustion techniques as sustainable solutions for internal combustion engines*, Springer, Singapore, pp. 3–7. [https://doi.org/10.1007/978-981-16-1513-9\\_1](https://doi.org/10.1007/978-981-16-1513-9_1).
- Ağbulut Ü., Bakır H. (2019) The investigation on economic and ecological impacts of tendency to electric vehicles instead of internal combustion engines, *Duzce Univ. Saglik. Bilim. Enst. Dergisi.* **7**, 1, 25–36. <https://doi.org/10.29130/dubited.457914>.
- Mohsin M., Abbas Q., Zhang J., Ikram M., Iqbal N. (2019) Integrated effect of energy consumption, economic development, and population growth on CO<sub>2</sub> based environmental degradation: a case of transport sector, *Environ. Sci. Pollut. Res.* **26**, 32, 32824–32835. <https://doi.org/10.1007/s11356-019-06372-8>.
- Kodjak D. (2015) Policies to reduce fuel consumption, air pollution, and carbon emissions from vehicles in G20 nations, The International Council on Clean Transportation, Available at [https://theicct.org/sites/default/files/publications/ICCT\\_G20-briefing-paper\\_Jun2015\\_updated.pdf](https://theicct.org/sites/default/files/publications/ICCT_G20-briefing-paper_Jun2015_updated.pdf).
- Kalghatgi G., Levinsky H., Colket M. (2018) Future transportation fuels, *Prog. Energy Combust. Sci.* **69**, 103–105. <https://doi.org/10.1016/j.pecs.2018.06.003>.
- Ağbulut Ü. (2022) Forecasting of transportation-related energy demand and CO<sub>2</sub> emissions in Turkey with different machine learning algorithms, *Sustain. Prod. Consum.* **29**, 141–157. <https://doi.org/10.1016/j.spc.2021.10.001>.
- Sekar M., Kumar T.P., Kumar M.S.G., Vanířková R., Maroušek J. (2021) Techno-economic review on short-term anthropogenic emissions of air pollutants and particulate matter, *Fuel* **305**, 121544. <https://doi.org/10.1016/j.fuel.2021.121544>.
- Liaquat A.M., Kalam M.A., Masjuki H.H., Jayed M.H. (2010) Potential emissions reduction in road transport sector using biofuel in developing countries, *Atmos. Environ.* **44**, 32, 3869–3877. <https://doi.org/10.1016/j.atmosenv.2010.07.003>.
- Fan J.L., Yu H., Wei Y.M. (2015) Residential energy-related carbon emissions in urban and rural China during 1996–2012: From the perspective of five end-use activities, *Energy Build.* **96**, 201–209. <https://doi.org/10.1016/j.enbuild.2015.03.026>.
- Ren F., Long D. (2021) Carbon emission forecasting and scenario analysis in Guangdong Province based on optimized fast learning network, *J. Clean. Prod.* **317**, 128408. <https://doi.org/10.1016/j.jclepro.2021.128408>.

- 15 Bakay M.S., Ağbulut Ü. (2021) Electricity production based forecasting of greenhouse gas emissions in Turkey with deep learning, support vector machine and artificial neural network algorithms, *J. Clean. Prod.* **285**, 125324. <https://doi.org/10.1016/j.jclepro.2020.125324>.
- 16 Gürel A.E., Ağbulut Ü., Ergün A., Ceylan I. (2020) Environmental and economic assessment of a low energy consumption household refrigerator, *Eng. Sci. Technol. Int. J.* **23**, 2, 365–372. <https://doi.org/10.1016/j.jestch.2019.06.003>.
- 17 Solak A.O. (2013) Türkiye’de ulaştırma sektörü enerji tüketiminin azaltılması: bir senaryo yaklaşımı, *Ekonomik ve Sosyal Araştırmalar Dergisi*. 125–140. <https://dergipark.org.tr/en/pub/esad/issue/6061/81512> (in Turkish).
- 18 MENR (2020). *Energy balance tables*. Available at <https://www.eigm.gov.tr/tr-TR/Denge-Tabloları/Denge-Tabloları> (accessed May 21, 2020).
- 19 Isik M., Sarica K., Ari I. (2020) Driving forces of Turkey’s transportation sector CO2 emissions: An LMDI approach, *Transp. Policy* **97**, 210–219. <https://doi.org/10.1016/j.tranpol.2020.07.006>.
- 20 Hill G., Heidrich O., Creutzig F., Blythe P. (2019) The role of electric vehicles in near-term mitigation pathways and achieving the UK’s carbon budget, *Appl. Energy* **251**, 113111. <https://doi.org/10.1016/j.apenergy.2019.04.107>.
- 21 Lander L., Kallitsis E., Hales A., Edge J.S., Korre A., Offer G. (2021) Cost and carbon footprint reduction of electric vehicle lithium-ion batteries through efficient thermal management, *Appl. Energy* **289**, 116737. <https://doi.org/10.1016/j.apenergy.2021.116737>.
- 22 Gryparis E., Papadopoulos P., Leligou H.C., Psomopoulos C.S. (2020) Electricity demand and carbon emission in power generation under high penetration of electric vehicles. A European Union perspective, *Energy Rep.* **6**, 475–486. <https://doi.org/10.1016/j.egyr.2020.09.025>.
- 23 Yesilyurt M.K., Aydın M., Yilbasi Z., Arslan M. (2020) Investigation on the structural effects of the addition of alcohols having various chain lengths into the vegetable oil-biodiesel-diesel fuel blends: An attempt for improving the performance, combustion, and exhaust emission characteristics of a compression ignition engine, *Fuel* **269**, 117455. <https://doi.org/10.1016/j.fuel.2020.117455>.
- 24 Yaman H., Doğan B., Yesilyurt M.K., Erol D. (2021) Application of higher-order alcohols (1-Hexanol-C6 and 1-Heptanol-C7) in a spark-ignition engine: analysis and assessment, *Arab. J. Sci. Eng.* **46**, 11937–11961. <https://doi.org/10.1007/s13369-021-05765-7>.
- 25 Güllüm M., Yesilyurt M.K., Bilgin A. (2020) The modeling and analysis of transesterification reaction conditions in the selection of optimal biodiesel yield and viscosity, *Environ. Sci. Pollut. Res.* **27**, 10, 10351–10366. <https://doi.org/10.1007/s11356-019-07473-0>.
- 26 Bozdağ A., Dokuz Y., Gökçek Ö.B. (2020) Spatial prediction of PM10 concentration using machine learning algorithms in Ankara, *Turkey. Environ. Pollut.* **263**, 114635. <https://doi.org/10.1016/j.envpol.2020.114635>.
- 27 Ozturk S., Ozturk F. (2018) Forecasting energy consumption of Turkey by Arima model, *J. Asian Sci. Res.* **8**, 2, 52–60. <https://doi.org/10.18488/journal.2.2018.82.52.60>.
- 28 Asteris P.G., Skentou A.D., Bardhan A., Samui P., Lourenço P.B. (2021) Soft computing techniques for the prediction of concrete compressive strength using non-destructive tests, *Const. Build. Mater.* **303**, 124450. <https://doi.org/10.1016/j.conbuildmat.2021.124450>.
- 29 Hajihassani M., Abdullah S.S., Asteris P.G., Armaghani D.J. (2019) A gene expression programming model for predicting tunnel convergence, *Appl. Sci.* **9**, 21, 4650. <https://doi.org/10.3390/app9214650>.
- 30 Chen H., Asteris P.G., Jahed Armaghani D., Gordan B., Pham B.T. (2019) Assessing dynamic conditions of the retaining wall: developing two hybrid intelligent models, *Appl. Sci.* **9**, 6, 1042. <https://doi.org/10.3390/app9061042>.
- 31 Asteris P.G., Tsaris A.K., Cavaleri L., Repapis C.C., Papalou A., Di Trapani F., Karypidis D.F. (2016) Prediction of the fundamental period of infilled RC frame structures using artificial neural networks, *Comput. Intell. Neurosci.* **2016**, 5104907. <https://doi.org/10.1155/2016/5104907>.
- 32 Ye L., Xie N., Hu A. (2021) A novel time-delay multivariate grey model for impact analysis of CO2 emissions from China’s transportation sectors, *Appl. Math. Model.* **91**, 493–507. <https://doi.org/10.1016/j.apm.2020.09.045>.
- 33 Wang W., Wang J. (2021) Determinants investigation and peak prediction of CO2 emissions in China’s transport sector utilizing bio-inspired extreme learning machine, *Environ. Sci. Pollut. Res.* **28**, 39, 55535–55553. <https://doi.org/10.1007/s11356-021-14852-z>.
- 34 Ratanavaraha V., Jomnonkwa S. (2015) Trends in Thailand CO2 emissions in the transportation sector and policy mitigation, *Transp. Policy* **41**, 136–146. <https://doi.org/10.1016/j.tranpol.2015.01.007>.
- 35 Li X., Ren A., Li Q. (2022) Exploring patterns of transportation-related CO2 emissions using machine learning methods, *Sustainability* **14**, 8, 4588. <https://doi.org/10.3390/su14084588>.
- 36 Lotfalipour M.R., Falahi M.A., Bastam M. (2013) Prediction of CO2 emissions in Iran using grey and ARIMA models, *Int. J. Energy Econ. Policy* **3**, 3, 229–237. <https://dergipark.org.tr/en/download/article-file/361229>.
- 37 Ofosu-Adarkwa J., Xie N., Javed S.A. (2020) Forecasting CO2 emissions of China’s cement industry using a hybrid Verhulst-GM (1, N) model and emissions’ technical conversion, *Renew. Sust. Energy Rev.* **130**, 109945. <https://doi.org/10.1016/j.rser.2020.109945>.
- 38 Chen X., Shuai C., Wu Y., Zhang Y. (2020) Analysis on the carbon emission peaks of China’s industrial, building, transport, and agricultural sectors, *Sci. Total Environ.* **709**, 135768. <https://doi.org/10.1016/j.scitotenv.2019.135768>.
- 39 Javed S.A., Cudjoe D. (2022) A novel grey forecasting of greenhouse gas emissions from four industries of China and India, *Sustain. Prod. Consum.* **29**, 777–790. <https://doi.org/10.1016/j.spc.2021.11.017>.
- 40 Bakır H., Ağbulut Ü., Gürel A.E., Yıldız G., Güvenç U., Soudagar M. E.M., Hoang A.T., Deepanraj B., Saini G., Afzal A. (2022) Forecasting of future greenhouse gas emission trajectory for India using energy and economic indexes with various metaheuristic algorithms, *J. Clean. Prod.* **360**, 131946. <https://doi.org/10.1016/j.jclepro.2022.131946>.
- 41 Wang Q., Li S., Pisarenko Z. (2020) Modeling carbon emission trajectory of China, US and India, *J. Clean. Prod.* **258**, 120723. <https://doi.org/10.1016/j.jclepro.2020.120723>.
- 42 Singh P.K., Pandey A.K., Ahuja S., Kiran R. (2022) Multiple forecasting approach: a prediction of CO2 emission from the paddy crop in India, *Environ. Sci. Pollut. Res.* **29**, 17, 25461–25472. <https://doi.org/10.1007/s11356-021-17487-2>.
- 43 Ari I., Koksall M.A. (2011) Carbon dioxide emission from the Turkish electricity sector and its mitigation options, *Energy Policy* **39**, 10, 6120–6135. <https://doi.org/10.1016/j.enpol.2011.07.012>.
- 44 Lu I.J., Lewis C., Lin S.J. (2009) The forecast of motor vehicle, energy demand and CO2 emission from Taiwan’s road transportation sector, *Energy Policy* **37**, 8, 2952–2961. <https://doi.org/10.1016/j.enpol.2009.03.039>.
- 45 Yılmaz H., Yılmaz M. (2013) Forecasting CO2 emissions for Turkey by using the grey prediction method, *Sigma* **31**, 141–148. <https://sigma.yildiz.edu.tr/storage/upload/pdfs/1636020414-tr.pdf>.
- 46 Sonmez M., Akgüngör A.P., Bektaş S. (2017) Estimating transportation energy demand in Turkey using the artificial bee colony algorithm, *Energy* **122**, 301–310. <https://doi.org/10.1016/j.energy.2017.01.074>.
- 47 Eyuboglu K., Uzar U. (2020) The impact of tourism on CO2 emission in Turkey, *Curr. Issues in Tour.* **23**, 13, 1631–1645. <https://doi.org/10.1080/13683500.2019.1636006>.
- 48 Uzlu E. (2021) Estimates of greenhouse gas emission in Turkey with grey wolf optimizer algorithm-optimized artificial neural networks, *Neural. Comput. Appl.* **33**, 20, 13567–13585. <https://doi.org/10.1007/s00521-021-05980-1>.
- 49 Zhu C., Wang M., Du W. (2020) Prediction on peak values of carbon dioxide emissions from the Chinese transportation industry based on the SVR model and scenario analysis, *J. Adv. Transp.* **2020**, 8848149. <https://doi.org/10.1155/2020/8848149>.
- 50 World Bank(2020) Turkey, Available at <https://www.worlddata.info/asia/turkey/index.php>.
- 51 TGDH (2020) *Turkish general directorate of highways, 2020*, Highways Transportation Statistics, Available at <http://www.eigm.gov.tr/en-US/>.

- 52 TurkStat (2023) Vehicle-kilometer Statistics, Available at <https://data.tuik.gov.tr/Bulten/Index?p=49527&dil=2> (accessed 30 January 2023).
- 53 Eryilmaz T., Yesilyurt M.K., Taner A., Celik S.A. (2015) Prediction of kinematic viscosities of biodiesels derived from edible and non-edible vegetable oils by using artificial neural networks, *Arab. J. Sci. Eng.* **40**, 3745–3758. <https://doi.org/10.1007/s13369-015-1831-6>.
- 54 Swinscow T.D.V., Campbell M.J. (2002) *Statistics at square one*, BMJ, London, pp. 111–125. Available at [http://publish.uwo.ca/~gzou2/Stats\\_at\\_Square1.pdf](http://publish.uwo.ca/~gzou2/Stats_at_Square1.pdf).
- 55 Hidecker M.J.C., Ho N.T., Dodge N., Hurvitz E.A., Slaughter J., Workinger M.S., Kent R.D., Rosenbaum P., Lenski M., Messaros B. M., Vanderbeek S.B., Deroos S., Paneth N. (2012) Inter-relationships of functional status in cerebral palsy: analyzing gross motor function, manual ability, and communication function classification systems in children, *Dev. Med. Child Neurol.* **54**, 8, 737–742. <https://doi.org/10.1111/j.1469-8749.2012.04312.x>.
- 56 Nguyen G., Dlugolinsky S., Bobák M., Tran V., Lopez Garcia A., Heredia I., Malík P., Hluchý L. (2019) Machine learning and deep learning frameworks and libraries for large-scale data mining: a survey, *Artif. Intell. Rev.* **52**, 1, 77–124. <https://doi.org/10.1007/s10462-018-09679-z>.
- 57 Sarker I.H., Furhad M.H., Nowrozy R. (2021) Ai-driven cybersecurity: an overview, security intelligence modeling and research directions, *SN Comput. Sci.* **2**, 1–18. <https://doi.org/10.1007/s42979-021-00557-0>.
- 58 Google trends (2022) Available at <https://trends.google.com/trends/>.
- 59 Asteris P.G., Nozhati S., Nikoo M., Cavaleri L., Nikoo M. (2019) Krill herd algorithm-based neural network in structural seismic reliability evaluation, *Mech. Adv. Mater. Struct.* **26**, 13, 1146–1153. <https://doi.org/10.1080/15376494.2018.1430874>.
- 60 Dave V.S., Dutta K. (2014) Neural network based models for software effort estimation: a review, *Artif. Intell. Rev.* **42**, 2, 295–307. <https://doi.org/10.1007/s10462-012-9339-x>.
- 61 Tokar A.S., Johnson P.A. (1999) Rainfall-runoff modeling using artificial neural networks, *J. Hydrol. Eng.* **4**, 3, 232–239. [https://doi.org/10.1061/\(ASCE\)1084-0699\(1999\)4:3\(232\)](https://doi.org/10.1061/(ASCE)1084-0699(1999)4:3(232)).
- 62 Singh G., Sachan M. (2014) Multi-layer perceptron (MLP) neural network technique for offline handwritten Gurmukhi character recognition, in: Karthikeyan M. (ed.), *International Conference on Computational Intelligence and Computing Research (ICIC)*, IEEE, pp. 1–5. <https://doi.org/10.1109/ICIC.2014.7238334>.
- 63 Kolchinsky A., Lourenço A., Li L., Rocha L.M. (2013) Evaluation of linear classifiers on articles containing pharmacokinetic evidence of drug-drug interactions, in: R.B. Altman, L. Hunter, M.D. Ritchie, T. Murray, T.E. Klein (eds), *Biocomputing 2013*, World Scientific Publishing, pp. 409–420. [https://doi.org/10.1142/9789814447973\\_0040](https://doi.org/10.1142/9789814447973_0040).
- 64 Wen Z., Shi J., Li Q., He B., Chen J. (2018) ThunderSVM: A fast SVM library on GPUs and CPUs, *J. Mach. Learn. Res.* **19**, 1, 797–801. <https://www.jmlr.org/papers/volume19/17-740/17-740.pdf>.
- 65 Kohli S., Godwin G.T., Urolagin S. (2021) Sales prediction using linear and KNN regression, in: S. Patnaik, X.-S. Yang, I.K. Sethi (eds), *Advances in Machine Learning and Computational Intelligence, Proceedings of ICMLCI 2019*, Springer, Singapore, pp. 321–329. [https://doi.org/10.1007/978-981-15-5243-4\\_29](https://doi.org/10.1007/978-981-15-5243-4_29).
- 66 Pradeep K.R., Naveen N.C. (2018) Lung cancer survivability prediction based on performance using classification techniques of support vector machines, C4.5 and Naive Bayes algorithms for healthcare analytics, *Procedia Comput. Sci.* **132**, 412–420. <https://doi.org/10.1016/j.procs.2018.05.162>.
- 67 De Caigny A., Coussemont K., De Bock K.W. (2018) A new hybrid classification algorithm for customer churn prediction based on logistic regression and decision trees, *Eur. J. Oper. Res.* **269**, 2, 760–772. <https://doi.org/10.1016/j.ejor.2018.02.009>.
- 68 Popescu M.C., Balas V.E., Perescu-Popescu L., Mastorakis N. (2009) Multilayer perceptron and neural networks, *WSEAS Trans. Circuits Syst.* **8**, 7, 579–588. <https://www.academia.edu/download/69679997/29-485.pdf>.
- 69 Oquab M., Bottou L., Laptev I., Sivic J. (2015) Is object localization for free? Weakly-supervised learning with convolutional neural networks, in: LeCun Y. (ed), *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 685–694, <https://doi.org/10.1109/CVPR.2015.7298668>.
- 70 Armaghani D.J., Asteris P.G. (2021) A comparative study of ANN and ANFIS models for the prediction of cement-based mortar materials compressive strength, *Neural. Comput. Appl.* **33**, 9, 4501–4532. <https://doi.org/10.1007/s00521-020-05244-4>.
- 71 Friedman J.H. (2001) Greedy function approximation: a gradient boosting machine, *Ann. Stat.* **29**, 1189–1232. <https://www.jstor.org/stable/2699986>.
- 72 Pedregosa F., Varoquaux G., Gramfort A., Michel V., Thirion B., Grisel O., Blondel M., Prettenhofer P., Weiss R., Dubourg V., Vanderplas J., Passos A., Cournapeau D., Brucher M., Perrot M., Duchesnay É. (2011) Scikit-learn: machine learning in Python, *J. Mach. Learn. Res.* **12**, 2825–2830. <https://www.jmlr.org/papers/volume12/pedregosa11a/pedregosa11a.pdf?ref=https>
- 73 Dong W., Huang Y., Lehane B., Ma G. (2020) XGBoost algorithm-based prediction of concrete electrical resistivity for structural health monitoring, *Automat. Constr.* **114**, 103155. <https://doi.org/10.1016/j.autcon.2020.103155>.
- 74 Möller A., Ruhlmann-Kleider V., Leloup C., Neveu J., Palanque-Delabrouille N., Rich J., Carlberg R., Lidman C., Pritchett C. (2016) Photometric classification of type Ia supernovae in the SuperNova Legacy Survey with supervised learning, *J. Cosmol. Astropart. Phys.* **2016**, 12, 008. <https://doi.org/10.1088/1475-7516/2016/12/008>.
- 75 Sarker I.H. (2021) Machine learning: Algorithms, real-world applications and research directions, *SN Comput. Sci.* **2**, 3, 1–21. <https://doi.org/10.1007/s42979-021-00592-x>.
- 76 Quej V.H., Almorox J., Ibrakhimov M., Saito L. (2016) Empirical models for estimating daily global solar radiation in Yucatán Peninsula, Mexico, *Energy Convers. Manag.* **110**, 448–456. <https://doi.org/10.1016/j.enconman.2015.12.050>.
- 77 Notton G., Voyant C. (2018) Forecasting of intermittent solar energy resource, in: Yahyaoui I. (ed), *Advances in renewable energies and power technologies*, Elsevier, pp. 77–114. <https://doi.org/10.1016/B978-0-12-812959-3.00003-4>.
- 78 Fan J., Wang X., Wu L., Zhang F., Bai H., Lu X., Xiang Y. (2018) New combined models for estimating daily global solar radiation based on sunshine duration in humid regions: a case study in South China, *Energy Convers. Manag.* **156**, 618–625. <https://doi.org/10.1016/j.enconman.2017.11.085>.
- 79 Manju S., Sandeep M. (2019) Prediction and performance assessment of global solar radiation in Indian cities: A comparison of satellite and surface measured data, *J. Clean. Prod.* **230**, 116–128. <https://doi.org/10.1016/j.jclepro.2019.05.108>.
- 80 Voyant C., Notton G., Kalogirou S., Nivet M.L., Paoli C., Motte F., Fouilloy A. (2017) Machine learning methods for solar radiation forecasting: A review, *Renew. Energy* **105**, 569–582. <https://doi.org/10.1016/j.renene.2016.12.095>.
- 81 Gouda S.G., Hussein Z., Luo S., Yuan Q. (2019) Model selection for accurate daily global solar radiation prediction in China, *J. Clean. Prod.* **221**, 132–144. <https://doi.org/10.1016/j.jclepro.2019.02.211>.
- 82 Ağbulut Ü., Gürel A.E., Biçen Y. (2021) Prediction of daily global solar radiation using different machine learning algorithms: Evaluation and comparison, *Renew. Sust. Energ. Rev.* **135**, 110114. <https://doi.org/10.1016/j.rser.2020.110114>.
- 83 Mohamed Z.E., Farouk R.M., Saleh H.H. (2018) The Performance Evaluation of Mathematical Models for Predicting MAD-GSR in Egypt, *Glob. J. Pure Appl. Math.* **14**, 7, 897–918. [https://www.ripublication.com/gjpm18/gjpm14n7\\_01.pdf](https://www.ripublication.com/gjpm18/gjpm14n7_01.pdf).
- 84 Singla P., Duhann M., Saroha S. (2022) Different normalization techniques as data preprocessing for one step ahead forecasting of solar global horizontal irradiance, in: Dubey A.K., Narang S.K., Srivastav A.L., Kumar A., García-Díaz V. (eds), *Artificial intelligence for renewable energy systems*, Woodhead Publishing, pp. 209–230. <https://doi.org/10.1016/B978-0-323-90396-7.00004-3>.

- 85 Guermoui M., Rabehi A., Gairaa K., Benkacali S. (2018) Support vector regression methodology for estimating global solar radiation in Algeria, *Eur. Phys. J. Plus* **133**, 1–9. <https://doi.org/10.1140/epjp/i2018-11845-y>.
- 86 Apostolopoulou M., Asteris P.G., Armaghani D.J., Douvika M.G., Lourenço P.B., Cavaleri L., Bakolas A., Moropoulou A. (2020) Mapping and holistic design of natural hydraulic lime mortars, *Cem. Concr. Res.* **136**, 106167. <https://doi.org/10.1016/j.cemconres.2020.106167>.
- 87 Asteris P.G., Koopialipour M., Armaghani D.J., Kotsonis E.A., Lourenço P.B. (2021) Prediction of cement-based mortars compressive strength using machine learning techniques, *Neural. Comput. Appl.* **33**, 19, 13089–13121. <https://doi.org/10.1007/s00521-021-06004-8>.
- 88 Yesilyurt M.K. (2018) The evaluation of a direct injection diesel engine operating with waste cooking oil biodiesel in point of the environmental and enviroeconomic aspects, *Energy Sources A Recovery Util. Environ. Eff.* **40**, 6, 654–661. <https://doi.org/10.1080/15567036.2018.1454546>.
- 89 Trading Economics (2023) Available at <https://tradingeconomics.com/commodity/carbon> (accessed April 12, 2023).
- 90 TPAO (2023). *Türkiye Petrolleri Anonim Ortaklığı Genel Müdürlüğü. 2021 Ham Petrol ve Doğal Gaz Sektör Raporu*. <https://www.tpao.gov.tr/file/2206/2021-petrol-ve-dogal-gaz-sektor-raporu-861629db02eb5624.pdf> (accessed March 14, 2023).
- 91 DGEA (2021) Directorate general of energy affairs, Available at <https://enerji.gov.tr/eigm> (accessed June 18, 2021).
- 92 Gürel A.E., Daşdemir A. (2011) Economical and environmental effects of thermal insulation thickness in four different climatic regions of Turkey, *Int. J. Renew. Energy Res.* **1**, 1, 1–10. <https://dergipark.org.tr/en/pub/ijrer/issue/16088/168432>.
- 93 Efe Ş., Ceviz M.A., Temur H. (2018) Comparative engine characteristics of biodiesels from hazelnut, corn, soybean, canola and sunflower oils on DI diesel engine, *Renew. Energy* **119**, 142–151. <https://doi.org/10.1016/j.renene.2017.12.011>.
- 94 Kumar A., Subramanian K.A. (2017) Control of greenhouse gas emissions (CO<sub>2</sub>, CH<sub>4</sub> and N<sub>2</sub>O) of a biodiesel (B100) fueled automotive diesel engine using increased compression ratio, *Appl. Therm. Eng.* **127**, 95–105. <https://doi.org/10.1016/j.applthermaleng.2017.08.015>.
- 95 Kargı V.S.A. (2014) A comparison of artificial neural networks and multiple linear regression models as in predictors of fabric weft defects, *Text. Appar.* **24**, 3, 309–316. <https://dergipark.org.tr/en/pub/tektstilvekonfeksiyon/issue/23645/251871>.