Transportation Research Part D 135 (2024) 104389

Contents lists available at ScienceDirect

Transportation Research Part D

journal homepage: www.elsevier.com/locate/trd

Urban transport emission prediction analysis through machine learning and deep learning techniques

Tianbo Ji [a], Kechen Li [a], Quanwei Sun [a], Zexia Duan [b,*]

[a] School of Transportation and Civil Engineering, Nantong University, Nantong 226019, Jiangsu, China
[b] School of Electrical Engineering and Automation, Nantong University, Nantong 226019, Jiangsu, China

ARTICLE INFO

Keywords:
Machine learning
Transport emissions
Deep learning
Air pollution

ABSTRACT

About 6.6 million people die every year from air pollution diseases globally. Transportation industry is considered one of the leading contributors in air pollution. This research utilizes deep learning and machine learning techniques to predict China's transport-related CO_2 emissions and energy needs by utilizing variables like population, car kilometers, year and GDP per capita. The outcomes have been analyzed using six analytical measures: determination coefficient, RMSE, relative RMSE, mean absolute percentage error, mean bias error and mean absolute bias error. Findings indicate that yearly increase in transport-related CO_2 emissions in China will be 3.66%, and transport energy consumption will increase by 3.8%. Energy consumption and transport CO_2 emissions are projected to rise by roughly 3.5 times by 2050 as compared to current levels. Therefore, government should re-evaluate its energy investment plans for the future and institute new rules, and standards regarding transport-related energy consumption and pollution reduction.

1. Introduction

New concepts in urban development are focusing on smart infrastructure upgrades. City leaders are always looking for novel approaches to transport management and air pollution reduction. However, so far, smart solutions have only partially solved the problem of environmental effects. If the economies are serious about reducing pollution in big urban areas, one of the most pressing issues that needs to be addressed is the road transport emissions (Lakhouit et al., 2023). According to the Global Energy Agency, cities are responsible for more than 70 % of the world's greenhouse gas emissions from energy use. If not properly managed, the predicted increase in urban population implies that the environmental effect of cities will also increase. Some researches, however, challenge these numbers, arguing that modern agriculture, forest destruction and smaller cities alongwith rural regions are more responsible for the world's greenhouse gas emissions than big cities (Lakhouit et al., 2023). The transportation sector significantly contributes to the rapid exhaustion of fossil fuel resources, accounting for almost 50 % of global oil consumption. As a result of this, fuel prices have been steadily rising (Lan et al., 2023). Growth in the transportation industry has been observed on a global scale. Accordingly, about 25 % of all non-natural CO_2 emissions come from the transportation industry. Vehicles in transport industry are classified into various categories. Vehicles with lighter loads are part of the transportation category; vehicle that carries people and cargo and has a total vehicle weight of 10,000 lb or less is considered a light-duty automobile. Examples include automobiles, vans, Sport Utility Vehicles (SUVs) and pickup trucks etc. In this backdrop, the objectives of the paper are as under:

* Corresponding author.
E-mail addresses: jitianbo@ntu.edu.cn (T. Ji), kechen.li@stmail.ntu.edu.cn (K. Li), zexiad@ntu.edu.cn (Z. Duan).

https://doi.org/10.1016/j.trd.2024.104389
Received 15 July 2024; Received in revised form 12 August 2024; Accepted 26 August 2024
Available online 31 August 2024

(a) Employ deep learning and machine learning techniques to accurately predict transport-related CO_2 emissions and energy needs in China.
(b) Analyze how variables such as population, car kilometers, year, and GDP per capita influence CO_2 emissions and energy consumption.
(c) Provide a detailed forecast indicating that China's yearly increase in CO_2 emissions.

Government should reevaluate its energy investment plans and establish new regulations and standards to manage transport-related energy consumption and reduce pollution effectively.

The impact of energy consumption on national GDP growth and how carbon footprints impact people's health is quite evident. There are immediate and secondary effects of carbon emissions on human well-being. Firstly, carbon emissions produce direct impact on the respiratory system of humans, when inhaled in sufficient dosages, resulting in a state of headaches, dizziness, weakness and exhaustion. Furthermore, they indirectly impact individuals due to their contribution to significant global issues, including acid rain, increasing temperatures and changing climates (Maria et al., 2023). All of these problems are very significant for individuals and the environment, and restricting these emissions is the only effective way to lessen their adverse impacts. Transportation industries release essential air pollutants, such as carbon dioxide and other greenhouse gases (Chang et al., 2023). Nearly 6.6 million people die each year as a result of air pollution-related illnesses; this number exceeds the total mortality toll from accidents on the roads, TB, and HIV/AIDS put together (Araya & Ghezzehei, 2019). The contributions of the study are as under:

Cogent in addressing the perpetual rise in carbon dioxide and energy use connected to transport, this article fosters a novel application of deep learning and machine learning to forecast these variables in China. Its predictors include population, year, car kilometers, and GDP per capita. The work uses advanced predictive models to provide a comprehensive, fact-based analysis of possible transport-related carbon dioxide emissions and energy consumption. This makes the work different from traditional approaches, which primarily rely on analyzing the past without using a deep and critical analysis of the future. The benefit of this work is that the authors use six measures to test the accuracy of the predictions made by the models, ensuring the thoroughness and reliability of the findings. The value of this work is that it offers data to prove that there will be a rise in transport energy needs and carbon dioxide emissions, which will force the Chinese government to reconsider its policy and develop a profound strategy to address this. This work is valuable because it not only offers the forecast that makes it Cogent, but it also explains the areas where the government needs to take action to facilitate positive changes in the Chinese transportation sector.

The rest of the study is structured as under: Section 2 presents review of relevant literature; Section 3 provides the specifics of the dataset utilized and methodology adopted in this paper; Section 4 provides the definitions, calculations, explanations and requirements of the statistical criteria; Section 5 includes the results and analysis; Section 6 concludes the paper by laying out essential findings, policy recommendations and future work.

2. Literature review

It is difficult to compare various strategies for the GHG emissions and established methodologies (Nguyen et al., 2021) owing to a number of factors. Inconsistencies may result from difference in sources, including divergent understandings of the concept of "urbanization", and the inherent challenges in accurately gauging its effects on the environment. Magazzino et al. (2020) argue that the sustainable development agenda has mostly taken a backseat due overconsumption and usage of resources occurring in tandem with urbanization. Wu et al. (2013) highlight the critical relevance of studying the mechanisms of primary energy consumption and the causes of air pollutants on both a global and regional level, as well as the nexus among them. Usage of fossil fuels has been recognized as the primary cause of air pollution on a global scale.

According to Magazzino et al. (2021), the combustion of fossil fuels accounts for a remarkable 86 % of the world's leading sources of energy demand. This makes up the majority of the world's energy consumption. According to Lu et al. (2021), worldwide GHGs in 2022 were nearly 50 billion metric tons of carbon dioxide equivalents. Shahzad et al. (2023) reported that China's emissions reached approximately half a billion metric tons of carbon dioxide equivalent. The majority of China's GHG emissions come from carbon dioxide. According to Sun (2022), the percentage of GHG emissions attributable to CO_2 in China increased from 70 % in 1990 to 80.6 % in 2019. The transportation industry is one of China's most significant contributors to greenhouse gas emissions, accounting for 23.3 % of total carbon dioxide emissions. However, globally this number is around 15 % (Morellos et al., 2016). Therefore, it becomes evident that China's transportation industry produces significantly more greenhouse gas emissions than the global average. Consequently, it is important for China to reduce carbon emissions from the transport industry. Nevertheless, the assumptions that this drop in emissions will be significantly cut in a short period is unreasonable; in any country, plans must be made for the near, medium and distant future. China must reconsider its financial investments and intentions, keeping in consideration the increase in transportation-based emissions in the following years, wherein, policymakers must concentrate on reducing greenhouse gas emissions from the power sector. According to Guo et al. (2023), the energy sector contributes to 74 % of all greenhouse gas emissions across borders and domestically.

Global climate change has increased the environmental concerns related to transportation presented by rapid urbanization (Liao et al., 2019). Carbon dioxide emissions from transport vehicles constitute 75 percent of the total emissions, whereas passenger cars are also responsible for 20 percent of the total Carbon dioxide emissions and greenhouse gases (GHG) worldwide (Lakhouit et al., 2023). Secondhand cars are also increasing despite rigorous regulations concerning greenhouse gas emissions and fuel efficiency (Picoli et al., 2018); excessive air pollution from ageing vehicles are some of the consequences. The transportation business plays an essential role in many aspects of daily life, such as the movement of passengers and the steady flow of goods. Therefore, the business has managed to stay sustainable by utilizing combustion engines, which use fossil fuels (Mele & Magazzino, 2020).

While most sectors have had relatively flat growth in total energy consumption over the past few years, the transport industry has seen an increase of over 20 %. Transportation industry being a significant contributor to air pollution; hence, the increase is significant. As a primary source of both air pollution and emissions of greenhouse gases, the transport industry depends extensively on liquid fuel derived from carbon (Aemmer & MacKenzie, 2022). According to multiple sources, the transportation industry explained nearly twenty percent of all greenhouse gas emissions in 2018. Indeed, the vehicles have served their transportation network well, and the vast majority of them have been driven by petrol and diesel engines, resulting in significant global emissions of greenhouse gases (Alova, 2020). However, fossil fuels are presently used for energy production in the transport sector worldwide at an average of more than 99 percent, with 95 percent of this supply coming from liquid petroleum-based energy sources (Singh & Goyal, 2023). Based on an analysis by Kokkinos et al. (2021), if a vehicle with an internal combustion engine consumes just one liter of oil, diesel or LPG, the resulting emissions of greenhouse gases are 2.3, 2.8, and 1.7 kg respectively. The predominant use of fossil fuels in the transport sector is to blame for both short and long-term global greenhouse gas emissions increases.

Furthermore, as the global population and level of lifestyle keeps growing, the percentage of transportation on the road is also expanding exponentially. Data shows that global automobile ownership will have tripled by 2050 compared to the present (Solano Meza et al., 2019). Over the last 20 years, China's inhabitants and standard of living have also increased dramatically, significantly increasing the country's vehicle usage. During the coming years, there is very little doubt that the usage will increase further. This increase in the total number of vehicles, which has increased the total fuel consumed, has also raised the amount of energy required and the amount of carbon dioxide (CO_2) released into the environment. Since China relies on foreign sources for approximately 75 percent of its energy needs, this has also negatively impacted the national economy and led to a significant annual rise in greenhouse gas emissions (Shahzad et al., 2023). Consequently, China has rapidly joined the countries affected by high energy use and air pollution and policy makers in the country need to take some actual steps.

In this regard, energy efficiency and reducing the country's greenhouse gas emissions are two interrelated issues that need simultaneous attention. During this process, it is critical to be aware of the country's historical data, assessment of its culture, learning how different factors interact and anticipation of its future needs, so that transportation-related energy consumption policies may be updated. National energy-emissions forecasting requires continuous study and updating (Chang et al., 2023). In addition, these studies have become increasingly popular in recent years due to multiple rules, regulations, standards, limitations, laws and obstacles about energy use and measures to reduce emissions. For this reason, many academics are devoted to studying the correlations between national energy use and pollutants; numerous approaches have been investigated in the relevant research to study trends, simulate them and predict CO_2 emissions. These approaches include different theories, mathematical frameworks and methods. Studies that have tried to forecast energy demand and carbon dioxide emissions for China and other nations have primarily relied on aggregate datasets showing total energy usage. However, a few articles attempt to predict China's CO_2 and other greenhouse gas (GHG) emissions utilizing transport details; however, the quantity of this research is minimal.

Literature reviews indicate that numerous methods have been used to predict future energy use and carbon dioxide emissions reliably. Several studies have shown a robust relationship between the need for energy and carbon dioxide emissions for specific nations indicating focus areas and using several variables such as population, vehicle kilometers driven, energy exports and imports, GDP, price of oil, passenger kilometers travelled, yearly vehicle kilometers travelled, past CO_2 emissions and past energy patterns.

3. Methods

3.1. Data collection

This research utilizes a database that includes variables such as GDP per capita, population in millions, automobile kilometers travelled per year, energy consumption/ demand in transportation (Mtoe) and CO_2 emissions from transportation from 2009 to 2022. The input variables include GDP, population, automobile number, and year, whereas, the output variables include energy consumption (EC) and CO_2 emissions from the transport industry. GDP, CO_2, population and energy consumption statistic are taken from World Bank data sources. The kilometers driven by vehicles were sourced from the China's General Directorate of Highways and the China's Statistical Institute. Table 1 provides relationship between input variables and data sources, whereas Table 2 provides a detailed breakdown of the statistical information utilized in this research.

This information, however, cannot be used for graph-making since significant disparities exist in the magnitude scales of the variables. Therefore, it is necessary to scale every value. Consequently, this study normalizes all parameters to values between 0 and 1.

Table 1
Relationships between input variables and sources of data.

Label	Factors that influence pre-selection?
Gross Domestic Product	The relationship between gross domestic product (GDP) and carbon dioxide emissions is highly correlated with the value of energy use. Changing GDP significantly impacts how energy and emissions will be used in the future
Population	Each country's desire is certain to go up as its population grows. Similarly, it has a direct impact on the level of emissions.
Vehicle kilometers	According to the country's transportation records, this amount depends on the vehicle's registration number. The transportation sector's energy usage and greenhouse gas emissions should increase with this yearly value.
Annual increase	According to past data, countries may expect their energy demand and emissions to continue to increase over time.

Source: China's General Directorate of Highways and the China's statistical institute.

Equation (1) demonstrates transforming the raw data into a normalized form.

$$X_{1(\text{normalized})} = \frac{X_1 - X_{1(\min)}}{X_{1(\max)} - X_{1(\min)}} \tag{1}$$

(X_1) is the lowest value among all values in all of these values, and ($X_{1(\min)}$) is an actual data, whereas (X_1) relates to the actual data. Similarly, ($X_{1(\max)}$) signifies the highest value in the collection. This means that ($X_{1(\text{normalized})}$) displays the average (X_1) number and ranges from 0 to 1. Although all of the variables have variations over the entire period, however, an overall upward trajectory can be seen. Table 4 shows the relationship between the value of every input variable and each output variable, providing insight into the connection between the two.

It can be observed that the correlation values of population are varying from 0.8994 to 0.9998 in Table 3. These correlation values can be utilized to determine the relationships between inputs and output parameters. According to the academic works cited (Araya & Ghezzehei, 2019), the correlation value can be understood in the following way:

- The correlation is compelling, with a value of $|r| \geq 0.9$.
- Solid relationship with $0.6 \leq |r| < 0.9$
- Moderate relationship with $0.5 \leq |r| < 0.7$.
- Weak correlation, with $0.3 \leq |r| < 0.5$.
- Absolute value of $|r|$ less than 0.3 indicates a fragile relationship.

This categorization affords the opportunity to deduce that level of correlation between every input and output parameter. Consequently, the MLA in China is trained with these inputs to estimate transportation-based CO_2 emissions and energy consumption.

3.2. Machine learning algorithms (MLA)

Numerous fields, such as developing health, are benefitting from recent advancements in machine learning, an AI technique that allows for categorization, analysis and decision-making (Liao et al., 2019). Machine learning techniques guarantee a highly satisfactory level of reliability. Using Rapid Miner Studio Version 9.6, this research examines the performance of three MLAs, including DL, ANN, and SVM in predicting transportation-related utilization of energy and carbon dioxide emissions.

3.2.1. ANN algorithms

ANNs are one of the most popular AI algorithms, replicating the structure and function of the brain's neural networks, while making them easier to use (Wang et al., 2022). It duplicates the way the human nervous system functions. It attempts to learn the systems before applying the outcomes. As a result of its exceptional predictive power for complicated and nonlinear systems, it finds extensive application. The current study used Multilayer Perceptron (MLP) ANN, or Multilayered Feed-Forward Neural Networks. One popular approach to training MLP is the return propagation algorithm, which is widely recognized. To put it scientifically, it works by adjusting the MLP's weighting values and addressing the intricate nexus of input and output data. Equations (2) and (3) (Guo et al., 2023) provide the overall output and error functions of MLP.

$$\mathrm{y_i} = f\Big(\sum_{\mathrm{i},j=1}^{\mathrm{L,M}} (\mathrm{w_{ij}} . \mathrm{x_i})\Big) \tag{2}$$

$$E = \frac{1}{\mathrm{N}} \sum_{i=1}^{\mathrm{N}} (\mathrm{D_i} - \mathrm{y_i})^2 \tag{3}$$

With x_i depicting the input data and w_{ij} indicates the weight value, the total output and predicted output values are represented by z_i and D_i, respectively.

Table 2
Brief summary of the data collection.

	Duration	Vehicle KMs	GDP (US$)	Population	EC (Mtoe)	Carbon dioxide (10^6 ton)
Mean	1994	42.34	4375.18	57,605,265	10.76	34.13
Standard Error	3	4.50	567.20	2,916,035.7	0.87	2.24
Standard Deviation	13.72	30.77	3888.47	14,135,678.5	5.88	15.28
Kurtosis	−1.21	−0.08	−0.53	−1.16	0.32	−0.21
Skewness	0.00	0.92	1.01	0.03	0.96	0.63
Diversity	47	113.20	13,159.39	45,945,422	22.68	57.13
Minimum	1971	6.49	455.11	35,876,304	3.22	11.51
Maximum	2017	119.68	13,614.49	80,821,725	25.89	68.63
Calculation	48	48	48	48	48	48

Source: China's General Directorate of Highways and the China's statistical institute.

Table 3
Variables' correlation factors concern one another.

	Duration	Vehicle Kms	GDP	Population	EC	Carbon Dioxide
Duration	1	–	–	–	–	–
Vehicle kilometer	0.9468	1	–	–	–	–
Gross Domestic Product	0.9007	0.9536	1	–	–	–
Population	0.9998	0.9499	0.8994	1	–	–
Energy consumption	0.9413	0.9865	0.9337	0.9457	1	–
Carbon Dioxide	0.9566	0.9865	0.9208	0.97	0.9857	1

The standard structure of a back propagation algorithm includes an input layer, a hidden layer and a final layer, and the algorithm itself has two phases: testing and learning. The model structure seen in this research has input, output and secret layer configurations of *5–16-1* for carbon dioxide estimation and *5–29-1* for EC estimation yield. For both results, it was found that the training cycle, learning rate and momentum were 251, 0.2, and 0.8 respectively.

3.2.2. SVM

Support Vector Machines can use several methods for supervised learning, including those for categorization and regression (Nguyen et al., 2021). As a result of its remarkable efficacy and practicability as a machine learning method for categorization and regression, it is attracting the attention of researchers from a wide range of subject areas. However, when compared to MLA, the SVM algorithm displays some significant benefits. Statistical learning theory and the concepts of structural risk minimization are the basis of SVM's operation. It follows that instead of reducing the error during local training, it tries to do so during generalization. Compared to other MLAs, this is the most visible difference between SVM and its competitors. SVM generates a hyper plane using the data values provided as $P = \{(x_i, d_i)\}_i^n$ and uses it to find linear functions f in the same way as in Equation (4).

In this context, n is the investigation, x_i is the input vector, and d_i is the output variable. All points should be approached with the same function $f(x)$ (Maria et al., 2023).

$$\mathrm{f}(\mathrm{x}) = \omega\varphi(\mathrm{x}) + b \tag{4}$$

There must be a relationship between the bias and the weight vector of the goal function. The same equation describes an object in three-dimensional space. The projection from x , a low-dimensional space, is non-linear.

Comparison with the linear regression model reveals its striking similarity to Equation (4). The original regression issue is transformed into the minimization issue of the regularized risk function to calculate w and b. This transformation determines the model difficulty and experimental error under ε-*insensitive* loss. According to Lan et al. (2023), this can be expressed in the following equations (5)–(6):

$$\text{Minimize}\frac{1}{2}w^2 + C\sum_{i=1}^{n}(\xi_i - \xi_i^*) \tag{5}$$

$$Subjectto\left\{\begin{array}{c} d_i - w\varphi(x) + b_i \le \varepsilon + \xi_i \\ w\varphi(x) + b_i - d_i \le \varepsilon + \xi_i^* \\ \xi_i, \xi_i^* \ge 0, i = 1, \cdots .l \end{array}\right. \tag{6}$$

The regularity phrase is denoted by $\frac{1}{2}w^2$, with indicators for the loss function and regularization coefficient $(C > 0)$ and ξ_i, respectively.

Accordingly, ξ_i^* displays the beneficial empty variables that lower and raise the additional deviations. According to Mohammedi et al. (2015), an overall function can be expressed as in Eq. (7) by applying the LaGrange multipliers and optimality requirements.

$$f(x, a_i a_i^*) = \sum_{i=1}^{n}(a_i - a_i^*)K(x, x_i) + b \tag{7}$$

The kernel function K is expressed in equation (8) using the Mercer condition as follows, wherein (a_i) and (a_i^*) are LaGrange multipliers.

$$K(x, x_i) = \varphi(x_i)\varphi(x_j) \tag{8}$$

The appropriate selection of the kernel function, *(C)*, and $\boldsymbol{\varepsilon}$ variables is essential for achieving high precision and excellent productivity in the SVM method. The grid search technique is used in this paper to choose these variables. In light of this, the optimal outcomes for carbon dioxide prediction are observed when the kernel function type is a dot, the kernel storage is 221, the integration is 1.76, the C value is 0.167, the stages are 100,000, the positive and negative parameters are 4.4358 and 0.32, and accordingly, the $\boldsymbol{\varepsilon}$ value is 0. Subsequently, it was seen that the best predictions for energy use came from using the kernel function type dot.

3.2.3. DL algorithm

The DL algorithm is a novel type of MLA that requires massive data sets to solve complicated issues on which the training phase can occur in an unattended manner. It is linked to ANN. Thus, key performance indicators significantly affecting the learning stage and label reliability include the secret layer size, activating work, epsilon and the number of epochs (Sun, 2022). More significantly, according to Dargan et al. (2020), the deep learning system can automatically extract features from the collection of data to analyze. In this paper, the optimal CO_2-prediction outcomes were observed when the activation function was set to converter, the period number was 45.91, and the epsilon value was 10,200. However, when the activation function is rectified, the period number equals 28.8, and the epsilon is 10,200; the results were similarly observed for energy needs predictions. Furthermore, the sizes of the layers are chosen for both prediction layers as (50 × 50).

3.3. Mathematical models

The study employs two commonly used mathematical models; exponential and linear regression to predict future utilization of energy and Carbon dioxide emissions from China's transportation sector as a function of time. Calculations are therefore based on information gathered from 1971 to 2004.

Yearly factors have been examined to forecast energy consumption and carbon dioxide emissions in China's transport sector. The main objective is to forecast future changes in demands for energy and carbon dioxide emissions in China. As a result, subsequent formulas for linear and nonlinear regression were obtained. These equations (9–12) give the numerical computation for each year of China's predicted energy and carbon dioxide consumption.

$$ModelI\_CO_2 = 0.8981y - 1757.2 \tag{9}$$

$$ModelII\_CO_2 == 9E - 31e0.0364y \tag{10}$$

$$Energy\_demand = 0.2886y - 565.54 \tag{11}$$

$$ModelII\_Energy\_demand == 6E - 34e0.0395y \tag{12}$$

This is where Models I and II, CO_2's linear and logarithmic regression equations come into effect. These equations are constructed by analyzing carbon dioxide data collected during 2009–2022. The two essential metrics used to measure their performance achievement are R2 and RMSE.

Table 4
Metrics alongwith descriptions, equations and evaluations.

Metrics	Formulas	Explanation	Performance Criteria	Equation No
R^2	$1-\frac{\sum(y_i-x_i)^2}{\sum(x_i-\overline{x}_i)^2}$	One of the most common ways to evaluate the accuracy of estimation is by looking at its coefficient of determination (R^2). It indicates how accurately the trends in the model can duplicate the trends in the actual data.	Larger R^2 values are preferred, and R^2 can vary from 0 to 1.	13
RMSE	$\sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i-x_i)^2}$	RMSE is a reasonable metric for analyzing short-term test data. It provides a metric for evaluating the predicted values' correctness.	RMSE ranges from zero to infinitely, with smaller RMSE values being preferred.	14
MAPE%	$\frac{1}{n}\sum_{i=1}^{n}\left\|\frac{x_i-y_i}{x_i}\right\| \times 100$	MAPE is among the most commonly used metrics. In statistics, it measures how well a model comprehends the future.	A less significant MAPE value is preferred for the outcomes of the predictions.	15
MBE	$\frac{1}{n}\sum_{i=1}^{n}(y_i-x_i)$	This parameter is often brought up while discussing the outcomes of long-term prediction performance evaluations.	The average of the predicted outcomes is larger than the actual when the sign of the MBE metric is negative. When the MBE metric has a positive value, the average of the forecast outcomes is smaller than the actual.	16
rRMSE %	$\frac{\sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i-x_i)^2}}{\overline{x}_i} \times 100$	Dividing the RMSE and the mean actual values yields the rRMSE.	Its value ranges from 0 to 1	17
MABE	$\frac{1}{n}\sum_{i=1}^{n}\|x_i-y_i\|$	The MABE provides a numerical representation of the bias errors.	A smaller MABE number indicates that the applicable model would produce more reliable results, as MABE takes values between zero and infinity.	18

4. Statistical standards

In this research, six commonly used statistical standards have been analyzed to evaluate the accuracy of forecasts made by machine learning systems. These include: R2, RMSE, MAPE, MBE, rRMSE, and MABE. All of these measures are summarized in Table 4, alongwith their definitions, performance standards and formulae.

5. Results and discussion

5.1. China's transport-based ED and carbon dioxide emission trends

Fig. 1 shows the fluctuation of ED and carbon dioxide emissions driven by transportation in China from a period of 1977 to 2022. There is a clear upward trend exhibited by both variables over time, as seen in the figure. When looking over a specific period, the peaks of both results fall into years that are otherwise quite comparable. As a result, the link between energy use and transportation-related CO_2 emissions is quite substantial.

In 1971, China's energy demand from transportation was 13.21 Mtoe, while the country's CO_2 emissions from this sector were 11.51×10^1 tons. According to the World Bank (2022), both categories have progressively grown since that year, with energy utilization reaching 25.89 Mtoe and CO_2 emissions reaching 23 tons in 2016. The average increase rate for energy demand and CO_2 emissions from 1977 to 2022 was determined to be about 6 % and 4.3 %, respectively. Rising living standards, a thriving economy and a fast-expanding population contribute to China's increasing energy demand and CO_2 emissions. According to Mohsin and Jamaani (2023), the national income per capita was $8,812 in 2020 and projected to reach $12,485 in 2024. With such indicators, country's road traffic is bound to increase. The number of registrations peaked at 25,144.858 in 2020, compared to an initial value of 9,655.171 in 2003 (the earliest year for which data is available). China's total number of registered road motor vehicles climbed by 180 % in 2021 compared to 2003. Moreover, the number of registered motor vehicles in China is estimated to grow by around 10 % annually, thus, the number of motor vehicles in the nation will likely rise annually. Abdul Rajak et al. (2019) state that China itself caters for approximately 40 % of its petroleum needs and approximately 60 % of its primary energy requirements. The nation's energy consumption in the transportation sector will undoubtedly rise as the number of automobiles on the road rises. This is likely to cause significant environmental issues and difficulties in developing the nation's economy. It is important to note that just one liter of petrol or diesel fuel used in a combustion engine releases almost three kilograms of CO_2 into the environment. All of this information indicates that transportation energy use and its effects on the economy and the environment need serious consideration by policy formulators and decision makers alike to develop a future course of action about what to do in the near, middle and long term.

4.2. Artificial intelligence-based CO_2 emission forecasting

Using three machine learning techniques, China's CO_2 emissions from transportation have been projected from 2009 to 2022. Fig. 2

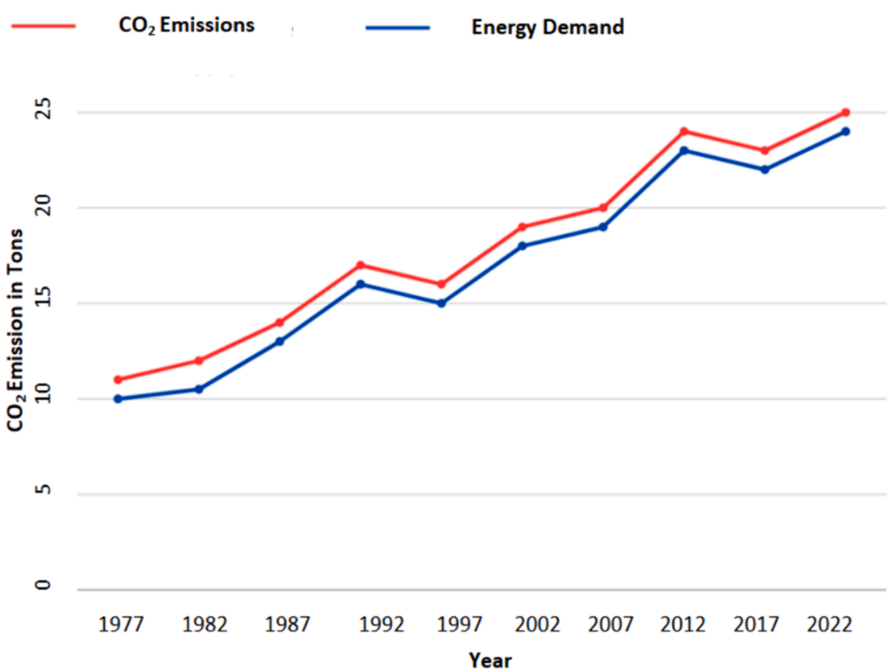

Fig. 1. Trends in China's Transport- related CO_2 Emissions and Energy Demand from 1977 to 2022.

shows the historical and projected CO_2 emissions.

The figure shows that CO_2 emissions from China's transportation industry have increased over time, with little variation between predicted and actual emissions. Table 5 provides numerical findings of statistical variables which have been used to compare the methods accurately.

According to the estimated indicators, the R^2 value for CO_2 emissions from transportation ranges from 0.8639 to 0.9117. The accuracy of predictions can be based on the coefficient of determination (R^2), which provides a sense of how closely the forecasting lines track the actual information. This is especially true after 2014 when the gap between real and predicted data started increasing, and almost no algorithms could account for the actual data's lines. This instance causes a decrease in R^2 metrics for all algorithms. Table 5 shows that numerically, ANN and DL have R^2 values that are relatively close to one another. Regarding CO_2 emissions, the DL algorithm shows a highly satisfying outcome with an R^2 of 0.9117, whereas SVM offers the worst R^2 result at 0.8639.

This research also considers the MBE metric, which is another statistic. Assuming all algorithms achieve numerical outcomes near 0 for the computed MBE measure, it would be preferable to use this metric. Here, the SVM method outperforms the others with an MBE score of 0.8277; the DL and ANN algorithms trail closely behind in terms of CO_2 emission prediction. Fig. 3 shows that after 2014, all algorithms predicted CO_2 emission data at amounts significantly lower than the actual levels. On the other hand, out of the three algorithms, SVM produces the most optimistic CO_2 emission predictions beyond that year. This means that while ANN and DL generate lower-than-actual CO_2 emission predictions, the appropriate algorithm produces higher-than-actual predictions.

Table 6 displays the MABE outcomes, which allow for an opportunity to conclude that the algorithms are statistically identical. Three algorithms have MABE metrics with different values: ANN has 3.2428, SVM has 3.8444 and DL has 3.7104. Therefore, the disparity between the MABE metric values of other methods is much lower, even if the most significant outcome is computed for the ANN algorithm. It is important to discuss the outcomes of various criteria in determining the performance of algorithms for CO_2 prediction. An additional significant statistic covered in this article is MAPE. This statistic shows how varied the predictions are in terms of percentage. Previous researches recommended four distinct approaches to categorize the performance of the MAPE metric. In other words, if the MAPE is less than or equal to 10 %, the prediction findings can be characterized as having a high level of accuracy.

- If the MPE is between 10 % and 20 %, the forecast findings can be categorized as having acceptable precision.
- The estimated results can be categorized as "reasonably accurate" if the MAPE is 20 % to 50 %.
- An "inaccurate prediction" classification can be applied to the prediction results if the MAPE exceeds 50 % (Zhang et al., 2023).

Each algorithm's prediction findings can be classified as having "high predictive reliability" according to this frequently utilized classification in the literature. None of the three MLAs employed in this study achieved MAPE values of less than 10 % when predicting transportation CO_2 emissions. The MAPE metric yielded 8.58 % for ANN, 7.25 % for SVM and 9.82 % for DL algorithms. Significant variations are also evident among any two algorithms, even if they all provide accurate forecasts according to the MAPE measure. Therefore, a more sensitive categorization is required to compare algorithms more accurately and that research should use a more specific categorization.

In discussing measuring the algorithms' performance, it is important also touch on rRMSE. This measure is used to scale dimensions from 0 to 100. A popular categorization for gaining insight into algorithm performance from rRMSE findings is present in the existing literature. Therefore, the following categorization provides insight into how an algorithm displays the superior outcome of the rRMSE

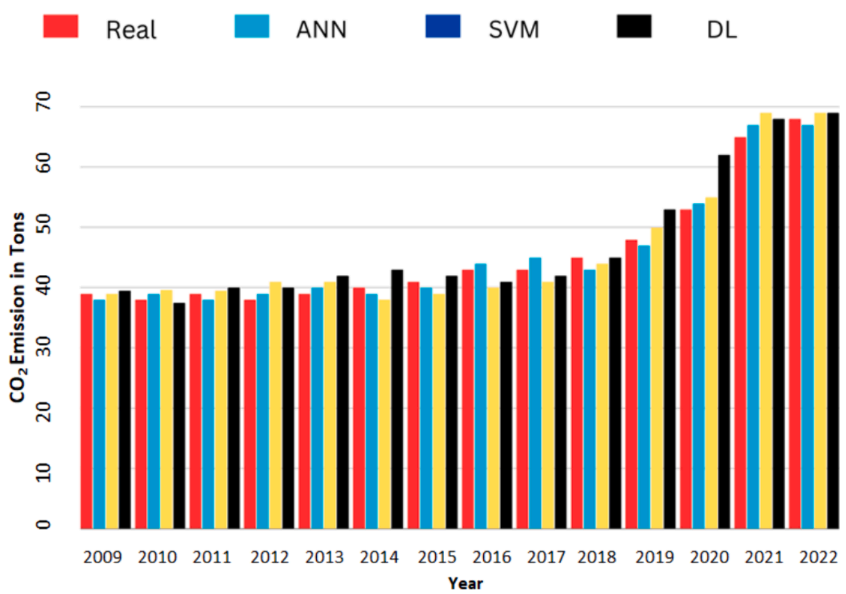

Fig. 2. CO_2 emissions data and predictions from transportation systems using various MLAs.

Table 5
Data analysis and statistical outcomes for CO_2 emission predictions in China's transportation sector using MLA.

Response	Metrics	ANN	SVM	DL
Carbon Dioxide	R^2	0.9089	0.8639	0.9117
	MBE, 106*ton*	−1.4734	0.8277	−1.4442
	MABE, 106*ton*	3.2428	3.8443	3.7104
	MAPE, %	8.58	7.25	9.82
	rRMSE, %	7.43	7.25	9.26
	RMSE, 106*ton*	3.8775	4.4016	4.8308

Table 6
China's energy consumption forecast using MLA in the transportation sector and related statistics.

Response	*Metrics*	*ANN*	*SVM*	*DL*
Energy Demand	*R2*	0.9236	0.8707	0.9020
	MBE, Mtoe	−0.3936	−0.3250	0.4276
	MABE, Mtoe	1.05	1.6106	1.5855
	MAPE, %	8.40	8.39	12.80
	rRMSE, %	8.39	11.06	9.997
	RMSE, Mtoe	1.2680	1.9924	1.8027

measure. If the root-mean-squared error (RSE) is less than 10 %, the prediction results can be deemed "exceptional" in this classification.

- Prediction results are "good" if the root-mean-squared error is less than 20 %.
- Prediction results can be reasonably accurate if the relative deviation from the mean (rRMSE) is between 20 % and 30 %.
- When the root-mean-squared error (RSE) is greater than 30 %, the prediction results can be categorized as "poor".

Table 6 depicts the corresponding rRMSE values for the ANN, SVM and DL algorithms as 7.43 %, 7.25 % and 9.26 %. It can be concluded that, when looking at the rRMSE measure, all of the algorithms' prediction outcomes are "excellent" based on the previous designation. Alternatively, according to Adedeji and Wang (2019), the appropriate algorithm achieves higher forecasting outcomes when the rRMSE value becomes 0. Table 6 shows that the SVM algorithm produced the good forecasting result out of the three algorithms when evaluating rRMSE. Next, when it came to CO_2 emission predictions using the rRMSE measure, ANN algorithm performance was second only to SVM algorithm. In addition to the rRMSE metric, the RMSE metric produces the following results for the ANN, SVM and DL algorithms: 3.8775, 4.4016, and 4.8308 respectively. Applying the same reasoning, it is found that ANN produces the best related root mean squared error metric outcome, followed by SVM and other methods. When looking at transportation-related

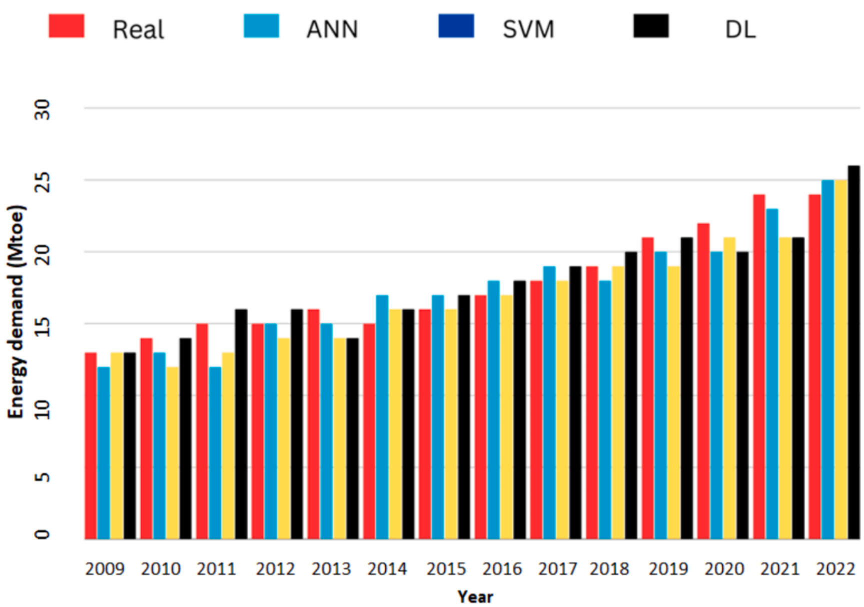

Fig. 3. Methods for predicting energy demand from transportation systems using various MLA (2009–2022).

CO_2 emissions forecasts in China, the SVM algorithm is the most suitable match, taking into consideration all statistical variables. But according to other statistical measures, the DL algorithm performs poorly and produces the worst CO_2 forecast of all the algorithms tested. This is despite the fact that it has the highest R^2 value for CO_2 prediction.

The results produced by each MLA for predicting CO_2 emissions in China's transportation sector are highly encouraging. Out of three machine learning methods, support vector machines (SVM), artificial neural networks (ANN) and deep learning (DL) have achieved the most favorable outcomes three times, twice and once respectively.

5.3. Travel energy demand forecasting with machine learning

MLA is used to predict the energy demand in China's transportation sector from 2009 to 2022. Fig. 3 shows China's historical and projected energy demand trend.

The figure shows that transportation accounts for a growing share of China's actual energy demand, which is growing year over year and is typically relatively close to predicted levels. The algorithms can be compared more accurately with the numerical results of statistical measures provided in Table 6.

Results show that the R^2 value of energy demand ranges from 0.8707 to 0.9236 (Table 6). Consequently, the ANN algorithm now has the optimal curve that predicts the changes in the curves of the data collected. Even though the ANN algorithm produced the highest R^2 value of 0.9236, the DL and SVM algorithms achieved satisfactory results 0.8707 and 0.9020, respectively. Considering the MBE metric for all methods combined, the values of this metric are minimal and closely spaced. To further complicate the problem, no other algorithm can predict ED with a positive MBE value other than the DL method. In other words, the DL algorithm is the only one that overestimates the ED value in its predictions. When looking at the MABE metric, the ANN algorithm has the best result at 1.05, while the SVM algorithm has the least favorable result at 1.62. However, as shown in Table 6, the MAPE values of the ANN and SVM algorithms for consumption prediction are shown to be almost identical. According to the previous categorization (Shao et al., 2019), these two algorithms have been categorized as "reliable" as they predict the real data with an error is no more than 10 %.

On the other hand, DL methods that exceed this 10 % limit are categorized as "similar" in the same category. For the RMS metric, a separate categorization is assigned according to the RMS metric's categorization; only the categorization is considered "good" since it is above the 10 % limit; however, the ANN and DL algorithms are considered "exceptional" because they are very near to this limit. However, as is evident from the findings, the statistical measures use various criteria to determine the predictors' effectiveness. As a result, the significance of analyzing various statistical variables, when discussing the accuracy and achievement of an indicator stands out. However, the findings also showed extensive categories employed in the previous researches. If algorithms using MAPE and rRMSE measurements are compared, they are classified as having achieved the same level of effectiveness, whether their values are 1 % or 9 %. This throws off the researchers' performance discussion of the predictors to a certain degree. Therefore, going over several statistical indicators is essential. This study also discusses RMSE, which is another statistical metric. It is found that for every method, this metric is less than 2. Using this metric for ED prediction; therefore, it can be concluded that the ANN algorithm consistently produces the most accurate ED reaction predictions (RMSE=1.2680).

5.4. Estimating transportation-based carbon dioxide and ED with mathematical techniques

To achieve a state of energy equilibrium, it is crucial to understand the nation's potential energy demand. International organizations, governments and policy makers use this information for planning and introducing new energy initiatives (Xia et al., 2022). In addition, in the modern world, a nation's trends in CO_2 emissions can be deduced based on its energy expenditure plans for the future. Because of this, researchers are constantly interested in studying energy demand and emission collection forecasts. Equations (9)–(12) provide the foundation for the two mathematical structures used in this research. Data on energy consumption and CO_2 emissions from 1971 to 2003 has been utilized in the model and also for creating mathematical equations, much like machine learning techniques. Data from 2009 to 2022 has subsequently been used to test the models. Using the data from the testing phase of both models for transportation-related carbon dioxide emissions and energy consumption in China, the effectiveness of the models can be determined by calculating the R^2 and root mean square error metrics using Equations 13 and 14, respectively. Making forecasts using the suggested models is possible after considering that the study's selected projection period does not include EC and CO_2 emissions reaching the pivot point. Statistics for these statistically significant indicators are presented in Table 7 and Table 8.

Table 7 shows that the mathematical models' outcomes for the process's training and testing stages are very similar and fall outside the permitted range for metric findings. According to the numbers, Model II produces lower CO_2 emissions and lower energy demands than Model I. As a result, Table 7 shows China's projected CO_2 emissions and energy consumption for transportation between now and

Table 7
Calculations of R^2 and root mean squared error for mathematical model assessment.

Results	Carbon Dioxide				ED			
Phase	Training		Testing		Training		Testing	
Metric	R^2	$RMSE, 10^6$ ton	R^2	$RMSE, 10^6$ ton	R^2	$RMSE, Mtoe$	R^2	$RMSE, Mtoe$
Model I	0.9134	2.64	0.8967	8.54	0.9286	0.7637	0.8509	4.88
Model II	0.9141	4.08	0.9076	3.96	0.9361	0.7323	0.8786	1.90

Table 8
Forecasted transportation-based CO_2 emissions and energy consumption.

Years	$CO_2(106ton)$		ED (Mtoe)	
	Model I	Model II	Model I	Model II
2025	60.56	89.19	18.60	31.57
2026	61.46	92.49	18.89	32.84
2027	62.36	95.92	19.17	34.16
2028	63.26	99.48	19.46	35.54
2029	64.16	103.17	19.75	36.97
2030	65.05	106.99	20.04	38.46
2031	65.95	110.96	20.33	40.00
2032	66.85	115.07	20.62	41.62
2033	67.75	119.33	20.91	43.30
2034	68.65	123.76	21.19	45.04
2035	69.55	128.34	21.48	46.85
2036	70.44	133.08	21.77	48.74
2037	71.34	138.04	22.06	50.70
2038	72.24	143.15	22.35	52.75
2039	73.14	148.46	22.64	54.87
2040	74.04	153.96	22.93	57.08
2041	74.93	159.67	23.21	59.38
2042	75.83	165.59	23.50	61.77
2043	76.73	171.73	23.79	64.26
2044	77.63	178.09	24.08	66.85
2045	78.53	184.69	24.37	69.54
2046	79.42	191.54	24.66	72.35
2047	80.32	198.64	24.95	75.26
2048	81.22	205.98	25.23	78.29
2049	82.12	213.64	25.52	81.45
2050	83.02	221.56	25.81	84.73

2050.

The energy consumption in China related to transportation was 17.80 Mtoe in 2022; however, according to Model I, it will be 25.81 Mtoe in 2050 and according to Model II, it will be 84.73 Mtoe. Energy consumption in China for transportation has been on the rise and will likely continue in the years to come. The results of the mathematical models indicate that, compared to 2022, China's ED for transportation will rise significantly by the year 2050. Similarly, emissions of CO_2 from transportation are projected to surge by over 242 % by the year 2050.

In addition, Model I's accumulated annual growth rate in energy demand is 0.4 %, while Model II's is 3.8 %. Conversely, CO_2 emissions are rising at 0.8 % and 3.66 % annually. As a result, these results show that the transportation industry will likely continue to be a source of environmental problems and a barrier to the country's economic progress. Given China's heavy reliance on foreign energy sources, policymakers must take aggressive steps and adjust the country's energy strategy accordingly. Reduced energy consumption and more effective use of resources are necessary for China. The country purchases around 76 % of its total energy to maintain economic progress (Agajanian et al., 2019).

6. Conclusion and policy implications

Combined energy consumption and carbon dioxide emissions from China's transport industry have been evaluated in this article. Three MLAs: DL, ANN, and SVMs are first used to predict these results. Considerations such as GDP per individual, population, vehicle kilometers travelled and the year are utilized in developing the MLA. Six indicators of the algorithm's effectiveness in the prediction have been used. Then, two statistical models have been utilized. This mathematical framework predicts the CO_2 emissions and energy consumption in China's transport industry up to the year 2050. The findings indicate that economics, population, vehicle kilometers travelled, energy consumption for transport and carbon dioxide emissions are all strongly correlated. Moreover, DL algorithm typically produces the lowest results out of the three algorithms for predicting transportation-related carbon dioxide emissions and energy needs. In contrast, ANN and SVM algorithms have emerged as contenders by combining all measurements. However, the outcomes show that all machine learning algorithms have done a considerably good task of forecasting future CO2 emissions and energy use. Based on the MAPE measure, all of the algorithms demonstrated "high forecasting efficiency" for CO2 emissions and energy consumption, with the possible exception of the Deep Learning algorithm in the energy-demand domain.

6.1. Policy recommendations

The anticipated substantial rise in carbon dioxide emissions and energy usage associated with mobility indicates that the government should re-evaluate its present and future energy investment programmes to ensure they are in line with sustainable development objectives. It is crucial to establish strict laws and standards that explicitly address the energy consumption and emissions associated with transportation. Possible rules may encompass more stringent pollution thresholds, higher fuel economy requirements, and incentive for the use of cleaner technologies. The results emphasize the significance of supporting and allocating resources to clean energy technologies, such as electric and hydrogen-powered vehicles. Measures should be developed to promote the utilization of public transport, cycling, and walking. Investments in infrastructure that facilitate various modes of transportation have the potential to greatly decrease dependence on private vehicles and decrease overall emissions. Utilizing sophisticated machine learning models to accurately predict emissions establishes a solid basis for policy making that relies on data. Policymakers should utilize these predicted insights to create and execute focused actions with the goal of decreasing emissions connected to transportation.

6.2. Limitations

The machine learning models employed in this study rely on specific assumptions and simplifications that may not fully account for the intricacies of actual transportation systems and emission dynamics. These assumptions may induce biases and affect the generalizability of the findings. Although the study incorporates important variables such as population, car kilometers, year, and GDP per capita, it fails to consider additional influential factors such as technological advancements, policy changes, and shifts in consumer behaviour. These factors can have a substantial impact on emission trends. Moreover, this study specifically examines China and its findings may not be immediately relevant to other countries or regions that have distinct transportation infrastructures, policies, and socioeconomic conditions. This restricts the worldwide applicability of the findings. Temporal limitations refer to the fact that projections are derived from historical data and present patterns, which may not comprehensively encompass future developments. Unanticipated occurrences, such as financial downturns, global health crises, or significant advancements in technology, have the potential to change the course of transportation emissions and energy usage. The Health Impact Assessment does not establish a direct correlation between transport emissions and health effects. Although it emphasises the significance of decreasing emissions, there is a lack of thorough examination regarding the effects on public health, which might offer a more full comprehension of the advantages of reducing emissions.

6.3. Future work

Future research may investigate other advanced machine learning techniques, such as deep reinforcement learning, support vector machines, and hybrid models, to enhance the prediction accuracy and robustness of emission forecasts. Future studies could model various policy intervention scenarios to predict their impact on transport-related emissions and energy consumption. This can help policymakers understand the potential effectiveness of different strategies before implementation. Moreover, researchers may evaluate the impact of emerging technologies, such as autonomous vehicles, smart transportation systems, and next-generation biofuels, on future transport emissions. This assessment can guide investment in research and development of these technologies.

Another area of future research may be to expand the study to include multiple countries or regions to compare emission trends and policy impacts globally. Additionally, extending the temporal scope to include extended future projections can provide deeper insights into long-term trends. Incorporation of behavioral and societal factors into the models, such as changes in public transportation usage patterns, shifts in work-from-home trends, and consumer preferences for sustainable transport options is another interesting area to be explored in future. Climate feedback mechanisms may be incorporated into the models to understand how changes in transport emissions could influence broader climate patterns and, conversely, how evolving climate conditions could impact transportation systems.

CRediT authorship contribution statement

Tianbo Ji: Writing – review & editing, Writing – original draft, Visualization, Supervision. **Kechen Li:** Writing – review & editing, Writing – original draft. **Quanwei Sun:** Writing – review & editing, Writing – original draft, Methodology, Conceptualization. **Zexia Duan:** Writing – review & editing, Writing – original draft, Visualization, Validation, Supervision, Project administration, Methodology, Formal analysis, Conceptualization.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

References

Abdul Rajak, A.R., Hasan, S., Mahmood, B., 2019. Automatic waste detection by deep learning and disposal system design. J. Environ. Eng. Sci. 15 (1), 38–44. https://doi.org/10.1680/JENES.19.00023.

Adedeji, O., Wang, Z., 2019. Intelligent waste classification system using deep learning convolutional neural network. Procedia Manuf. 35, 607–612. https://doi.org/10.1016/j.promfg.2019.05.086.

Aemmer, Z., MacKenzie, D., 2022. Generative population synthesis for joint household and individual characteristics. Comput. Environ. Urban Syst. 96 https://doi.org/10.1016/j.compenvurbsys.2022.101852.

Agajanian, S., Oluyemi, O., Verkhivker, G.M., 2019. Integration of random forest classifiers and deep convolutional neural networks for classification and biomolecular modeling of cancer driver mutations. Front. Mol. Biosci. 6 (JUN) https://doi.org/10.3389/FMOLB.2019.00044.

Alova, G., 2020. A global analysis of the progress and failure of electric utilities to adapt their portfolios of power-generation assets to the energy transition. Nat. Energy 5 (11), 920–927. https://doi.org/10.1038/S41560-020-00686-5.

Araya, S.N., Ghezzehei, T.A., 2019. Using Machine Learning for Prediction of Saturated Hydraulic Conductivity and Its Sensitivity to Soil Structural Perturbations. Water Resour. Res. 55 (7), 5715–5737. https://doi.org/10.1029/2018WR024357.

Chang, L., Mohsin, M., Hasnaoui, A., Taghizadeh-Hesary, F., 2023. Exploring carbon dioxide emissions forecasting in China: A policy-oriented perspective using projection pursuit regression and machine learning models. Technol. Forecast. Soc. Chang. 197, 122872.

Dargan, S., Kumar, M., Ayyagari, M.R., Kumar, G., 2020. A Survey of Deep Learning and Its Applications: A New Paradigm to Machine Learning. Arch. Comput. Meth. Eng. 27 (4), 1071–1092. https://doi.org/10.1007/S11831-019-09344-W.

Guo, X., Huang, D., Li, X., Liang, C., 2023. Are categorical EPU indices predictable for carbon futures volatility? Evidence from the machine learning method. Int. Rev. Econ. Financ. 83, 672–693. https://doi.org/10.1016/j.iref.2022.10.011.

Kokkinos, K., Karayannis, V., Nathanail, E., Moustakas, K., 2021. A comparative analysis of Statistical and Computational Intelligence methodologies for the prediction of traffic-induced fine particulate matter and NO2. J. Clean. Prod. 328 https://doi.org/10.1016/j.jclepro.2021.129500.

Lakhouit, A., Shaban, M., Alatawi, A., Abbas, S.Y.H., Asiri, E., Al Juhni, T., Elsawy, M., 2023. Machine-learning approaches in geo-environmental engineering: Exploring smart solid waste management. J. Environ. Manage. 330, 117174 https://doi.org/10.1016/J.JENVMAN.2022.117174.

Lan, D.Y., He, P.J., Qi, Y.P., Wu, T.W., Xian, H.Y., Wang, R.H., Lü, F., Zhang, H., 2023. Optimizing the Quality of Machine Learning for Identifying the Share of Biogenic and Fossil Carbon in Solid Waste. Anal. Chem. 95 (9), 4412–4420. https://doi.org/10.1021/ACS.ANALCHEM.2C04940.

Liao, M., Kelley, S., Yao, Y., 2019. Generating energy and greenhouse gas inventory data of activated carbon production using machine learning and kinetic based process simulation. ACS Sustain. Chem. Eng. 8 (2), 1252–1261.

Lu, W., Lou, J., Webster, C., Xue, F., Bao, Z., Chi, B., 2021. Estimating construction waste generation in the Greater Bay Area, China using machine learning. Waste Manag. 134, 78–88. https://doi.org/10.1016/j.wasman.2021.08.012.

Magazzino, C., Mele, M., Schneider, N., 2020. The relationship between municipal solid waste and greenhouse gas emissions: Evidence from Switzerland. Waste Manag. 113, 508–520. https://doi.org/10.1016/j.wasman.2020.05.033.

Magazzino, C., Mele, M., Schneider, N., 2021. A machine learning approach on the relationship among solar and wind energy production, coal consumption, GDP, and CO2 emissions. Renew. Energy 167, 99–115. https://doi.org/10.1016/j.renene.2020.11.050.

Maria, M.R., Ballini, R., Souza, R.F., 2023. Evolution of Green Finance: A Bibliometric Analysis through Complex Networks and Machine Learning. Sustainability (Switzerland) 15 (2). https://doi.org/10.3390/SU15020967.

Mele, M., Magazzino, C., 2020. A Machine Learning analysis of the relationship among iron and steel industries, air pollution, and economic growth in China. J. Clean. Prod. 277, 123293 https://doi.org/10.1016/j.jclepro.2020.123293.

Mohsin, M., Jamaani, F., 2023. A novel deep-learning technique for forecasting oil price volatility using historical prices of five precious metals in context of green financing–A comparison of deep learning, machine learning, and statistical models. Resour. Policy 86, 104216.

Morellos, A., Pantazi, X.E., Moshou, D., Alexandridis, T., Whetton, R., Tziotzios, G., Wiebensohn, J., Bill, R., Mouazen, A.M., 2016. Machine learning based prediction of soil total nitrogen, organic carbon and moisture content by using VIS-NIR spectroscopy. Biosyst. Eng. 152, 104–116. https://doi.org/10.1016/j.biosystemseng.2016.04.018.

Nguyen, Q., Diaz-Rainey, I., Kuruppuarachchi, D., 2021. Predicting corporate carbon footprints for climate finance risk analyses: A machine learning approach. Energy Econ. 95 https://doi.org/10.1016/j.eneco.2021.105129.

Picoli, M.C.A., Camara, G., Sanches, I., Simões, R., Carvalho, A., Maciel, A., Coutinho, A., Esquerdo, J., Antunes, J., Begotti, R.A., Arvor, D., Almeida, C., 2018. Big earth observation time series analysis for monitoring Brazilian agriculture. ISPRS J. Photogramm. Remote Sens. 145, 328–339. https://doi.org/10.1016/j.isprsjprs.2018.08.007.

Shahzad, U., Sengupta, T., Rao, A., Cui, L., 2023. Forecasting carbon emissions future prices using the machine learning methods. Ann. Oper. Res. https://doi.org/10.1007/S10479-023-05188-7.

Shao, W., Ge, Z., Song, Z., Wang, K., 2019. Nonlinear industrial soft sensor development based on semi-supervised probabilistic mixture of extreme learning machines. Control Eng. Pract. 91.

Singh, S., Goyal, M.K., 2023. An innovative approach to predict atmospheric rivers: Exploring convolutional autoencoder. Atmos. Res. 289 https://doi.org/10.1016/j.atmosres.2023.106754.

Solano Meza, J.K., Orjuela Yepes, D., Rodrigo-Ilarri, J., Cassiraga, E., 2019. Predictive analysis of urban waste generation for the city of Bogotá, Colombia, through the implementation of decision trees-based machine learning, support vector machines and artificial neural networks. Heliyon 5 (11). https://doi.org/10.1016/J.HELIYON.2019.E02810.

Sun, C., 2022. The correlation between green finance and carbon emissions based on improved neural network. Neural Comput. & Applic. 34 (15), 12399–12413. https://doi.org/10.1007/S00521-021-06514-5/METRICS.

Wang, X., Li, J., Zheng, Y., Li, J., 2022. Smart systems engineering contributing to an intelligent carbon-neutral future: opportunities, challenges, and prospects. Frontiers of Chemical Science and Engineering 16 (6), 1023–1029. https://doi.org/10.1007/S11705-022-2142-6.

Wu, G., Kechavarzi, C., Li, X., Wu, S., Pollard, S.J.T., Sui, H., Coulon, F., 2013. Machine learning models for predicting PAHs bioavailability in compost amended soils. Chem. Eng. J. 223, 747–754. https://doi.org/10.1016/j.cej.2013.02.122.

Xia, W., Jiang, Y., Chen, X., Zhao, R., 2022. Application of machine learning algorithms in municipal solid waste management: A mini review. Waste Manag. Res. 40 (6), 609–624. https://doi.org/10.1177/0734242X211033716.

Zhang, S., Omar, A.H., Hashim, A.S., Alam, T., Khalifa, H.A.E.W., Elkotb, M.A., 2023. Enhancing waste management and prediction of water quality in the sustainable urban environment using optimized algorithm of least square support vector machine and deep learning techniques. Urban Clim. 49, 101487 https://doi.org/10.1016/J.UCLIM.2023.101487.