Advances in Artificial Intelligence and Machine Learning; Research 3 (3) 1295-1312 Received 10-05-2023; Accepted 27-07-2023; Published 4-08-2023

Forecasting of Transportation-related CO_2 Emissions in Canada with Different Machine Learning Algorithms

Ramlah Abdulmalik abdulmro17@brandonu.ca
Dept. of Math and Computer Science
Brandon University
270 18th Street Brandon MB
Canada

Gautam Srivastava srivastavag@brandonu.ca
Dept. of Math and Computer Science
Brandon University
270 18th Street Brandon MB
Canada

Corresponding Author: Gautam Srivastava

Abstract

The amount of carbon dioxide in the atmosphere has risen over recent years, with a growth of over 40%. This study examines transportation-related carbon dioxide (CO_2) emissions in Canada, which contribute significantly to the country's overall emissions. The study investigates the rise of carbon dioxide (CO_2) due to various reasons such as economic development, transportation, as well as population growth, but the study focuses on transportation-related CO_2 emission in Canada. Various machine learning algorithms, such as Deep Neural Networks, Support Vector Machines, and Random Forests, are utilized to forecast CO_2 emissions. The results show promising outcomes, with R2 values ranging from 0.9532 to 0.9996, RMSE values ranging from 1.0974 to 13.6561, MAPE scores from 0.0088 to 0.0010, MBE scores ranging from -0.0594 to 1.0366, rRMSE score ranging from 0.4259 to 5.3002, and MABE score ranging from 0.2643 to 5.6582 for all six (6) algorithms. To meet greenhouse gas reduction targets, this paper recommends further efforts to reduce CO_2 emissions from transportation sources and suggests the adoption of Vehicle Alternative Fuel Types and low-carbon fuels.

Keywords: Forecasting, Emissions, Carbon Dioxide, Vehicles, Machine Learning

1. INTRODUCTION

In ancient times, transportation was limited to walking and riding on animals such as horses, camels, and donkeys. Later, with the invention of the wheel, wheeled vehicles such as chariots, carts, and wagons were used to transport goods and people over longer distances [1]. During the Industrial Revolution in the 18^{th} and 19^{th} centuries, transportation underwent significant changes with the

Citation: Ramlah Abdulmalik and Gautam Srivastava. Forecasting of Transportation-related CO_2 Emissions in Canada with Different Machine Learning Algorithms Advances in Artificial Intelligence and Machine Learning. 2023;3(3):76.

introduction of trains and steam-powered ships. These modes of transportation greatly increased the speed and efficiency of travel and transportation of goods and greatly impacted the growth of commerce and industry. In the 20^{th} century, the transportation sector was transformed by the widespread adoption of automobiles and airplanes. The automobile revolutionized personal transportation and made it possible for people to travel quickly and easily to new destinations [2].

Canada's history is closely tied to the history of transportation and its development. The first settlers ventured only into those areas that were accessible by water and boats; canoes were the primary mode of transportation on the nation's lakes and rivers. Later, canals were constructed. Settlement in much of Canada followed the construction of railway lines. Roads and highways later gave access to regions of Canada that had not been served by railways. Today, air transport makes it possible for Canadians to travel to any area of the nation, regardless of how remote it may be. For a trading nation as vast as Canada, the importance of transportation in such a country cannot be underestimated. The great distances between mines, farms, forests and urban centers make efficient transport systems essential to the economy so that natural and manufactured goods can move freely through domestic and international markets. Transportation has and will continue to play an important role in the social and political unity of Canada[1,2]. As beneficial as the transportation sector is, the burning of fossil-based fuels in internal combustion engines which is neither renewable nor from a clean energy source, has proven to cause more harm than good.

Carbon Dioxide (CO_2) is naturally present in the atmosphere as part of the Earth's carbon cycle (the natural circulation of carbon among the atmosphere, oceans, soil, plants, and animals). Meanwhile, human activities alter the carbon cycle–both by adding more CO_2 to the atmosphere and by influencing the ability of natural sinks, like forests and soils, to remove and store CO_2 from the atmosphere. While CO_2 emissions come from a variety of natural sources, human-related emissions are responsible for the increase that has occurred in the atmosphere since the industrial revolution. Carbon Dioxide (CO_2) is the primary greenhouse gas emitted through human activities and the main human activity that emits CO_2 is the combustion of fossil fuels (coal, natural gas, and oil) for energy and transportation (IPCC[3]).

GreenHouse Gas (GHG) emissions are produced when hydrocarbons, such as natural gas and oil, are burned. GHGs include carbon dioxide (CO_2), methane, nitrous oxide, and ozone, all of which contribute to climate change. According to the Government of Canada, Canada's total GHG emissions in 2020 were 672 megatons of carbon dioxide equivalent ($MtCO_2e$). Globally, Canada's share of GHG emissions is less than 1.5%. In 2020, the oil and gas sector accounted for 178 megatons of carbon dioxide equivalent (Mt CO_2 eq) (26% of total emissions), followed closely by the transportation sector, which emitted 159 Mt CO_2 eq (25%)[4].

The effect of the fourth most abundant gas in the earth's atmosphere (CO_2) can be felt in our environment, economic growth and in the health sector. Exposure to CO_2 can produce a variety of health effects when inhaled in high dosages. These effects include headaches, dizziness, restlessness, difficulty in breathing, sweating, tiredness, increased heart rate, elevated blood pressure, coma, asphyxia, and convulsions. CO_2 emission also affects the environment by trapping heat in

[1] https:/www.thecanadianencyclopedia.ca/en/article/transportation
[2] http://www.thecanadianencyclopedia.ca/en/article/transportation-in-the-north
[3] https://keneamazon.net/Documents/Publications/Virtual-Library/Impacto/157.pdf
[4] https://www.canada.ca/en/environment-climate-change/services/environmental-indicators/greenhouse-gas-emissions.html

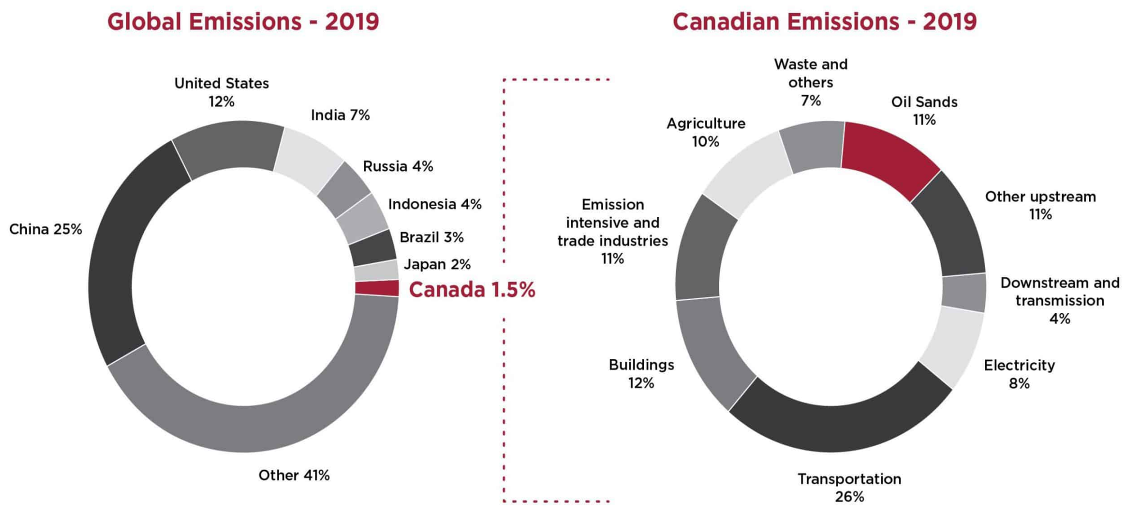

Figure 1: Carbon Canada's footprint in the year 2019 (The statistics were taken from the ref (Environment and Climate Change Canada, 2020; World Resources Institute, 2019).

the atmosphere and warming up the Earth. The burning of fossil fuels releases large amounts of carbon dioxide, a greenhouse gas, into the air causing global warming FIGURE 1. Approximately 6.5 million deaths annually occur due to air pollution-related diseases worldwide [3].

A study published by a Berlin think-tank says Canadians must cut their carbon footprints by 95 percent to help the world limit global warming to the 1.5 C goal set by the 2015 Paris Agreement. According to findings, the average person in Canada produces an equivalent of 14.2 tons of CO_2 as of 2019[5].

FIGURE 2 above is data gotten from the World Bank Data and our visualization shows that Canada produces the seventh highest per-capital carbon footprint in the world as of 2019. According to the World Energy Balances report published by International Energy Agency in the year 2020, the largest final energy consumption shares among the sectors belong to the industry with a rate of 38%, transportation with a rate of 28%, and residential with a rate of 21% in the year 2018 [4]. The transportation sector is the 2^{nd} largest source of Canada's GHG emissions and this accounted for approximately 25% of total GHG emissions in 2016.

The challenge of reducing GHG emissions while the energy demand is growing has lured the Government of Canada in 2015 to announce a climate target to reduce Canada's GHG emissions by 30% below 2005 levels by 2030. Furthermore, in 2020, the Government of Canada introduced 'A Healthy Environment and a Healthy Economy' which includes measures to exceed the 2030 target and advance Canada's emissions to net zero by 2050.

In recent years, various policies, and regulations, on both energy usage and emission mitigate issues have led to an accelerating increase in the popularity of these studies. Many researchers are dedicated to understanding the relationship between energy consumption and emission trends in

[5] https://www.cbc.ca/news/science/how-canadians-can-cut-carbon-footprints-1.6202194

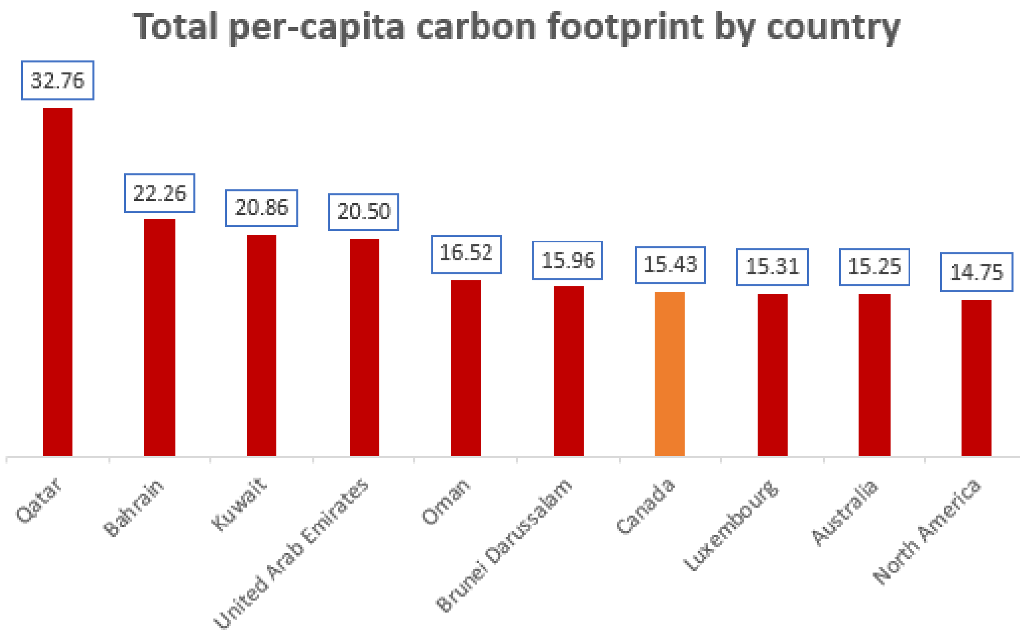

Figure 2: World Total per Capital Carbon Footprint

different countries [5]. Several methods including various algorithms, theories, and mathematical models have been studied for the modeling and forecasting of CO_2 emissions. Although some of these studies exist, the majority only focus on the overall energy consumption dataset of the relevant countries. There are very limited studies focusing on a fraction of CO_2 emission. [6].

This study aims to fill that gap by forecasting transportation-related CO_2 emissions in Canada by using Artificial Neural Network (ANN), Support Vector Machine (SVM), Deep Neural Network (DNN), Decision Tree Regressor and Gradient Boost Regressor (GBR). The datasets used are from the Government of Canada and Statistics Canada between the years 1995 – 2021. Some significant statistical metrics including R2, RMSE, MAPE, MBE, rRMSE, and MABE are used to discuss the performance of the algorithms for the 2022 – 2050 transportation-based CO_2 emission forecast.

Section 2 covers data collection and preprocessing, algorithm selection, model training and evaluation, hyperparameter optimization, model deployment and performance. Section 3 defines the statistical benchmarks, formulas, descriptions, and performance criteria. Section 4 presents the interpretation of the results and discussion achieved from the algorithms and mathematical models. The results are classified to show the success ability of the algorithms according to the classification used in the literature. Finally, the paper ends in Section 5 with the concluding remarks, a summary of the findings, implications for practice and policy and some suggestions on the study of the field of transportation and the environment.

2. METHODOLOGY

The flowchart of our research methods and algorithms used in this research paper can be seen in FIGURE 3, emission inventory method was applied by collecting relevant data from various sources then a statistical technique (Regression analysis) was used to analyze the relationship between one or more independent variables and a dependent variable, such as CO_2 emissions to help identify the factors that influence CO_2 emissions and predict the impact of changes in these factors. Several machine learning algorithms such as neural networks, decision trees, and random forests were used to analyze our large datasets and identify patterns in the data. These algorithms predicted future CO_2 emissions based on historical data and identified the factors that have the greatest impact on CO_2 emissions.

2.1 Data Collection and Preprocessing

The dataset used in this paper is gotten from two sources. The first data source is a vehicle registration dataset derived from Statistics data; model year, make, model, vehicle class, engine size and cylinders summarized in TABLE 1.

Table 1: Vehicle registration dataset features

Columns	**Details**
Model Year	The year the car's model was built
Make	The company of the vehicle
Model	The car's model
Vehicle Class	The class of vehicle depends on their utility, capacity and weight
Engine Size	The size of the engine used in Litre
Cylinders	The number of cylinders

The second source was gotten from the Government of Canada, and it contains fuel consumption rating data which includes the following features in TABLE 2: transmission, fuel type, fuel consumption city, fuel consumption hwy, fuel consumption comb and our target variable CO_2 emissions.

Table 2: Fuel consumption rating dataset features

Columns	**Details**
Transmission	The transmission type with the number of gears
Fuel Type	The type of fuel used
Fuel Consumption City	The fuel consumption on city roads (L/100 km)
Fuel Consumption HWY	The fuel consumption on highways (L/100 km)
Fuel Consumption Comb	The combined fuel consumption (55% city, 45% highway) is shown in L/100 km
CO_2 Emission	The amount of carbon dioxide emissions rate

TABLE 3 shows the first five rows in the dataset.

Table 3: Top 5 rows in the dataset

	MODEL YEAR	MAKE	MODEL	VEHICLE CLASS	ENGINE SIZE(L)	CYLIN-DERS	TRANS-MISSION	FUEL TYPE	FUEL CONSUMPTION CITY (L/100KM)	FUEL CONSUMPTION HWY (L/100KM)	FUEL CONSUMPTION COMB (L/100KM)	FUEL CONSUMPTION COMB (MPG)	CO_2 EMISSIONS
0	1995	ACURA	INTEGRA	SUBCOMPACT	1.8	4	A4	X	11.6	8.3	10.1	28	232
1	1995	ACURA	INTEGRA	SUBCOMPACT	1.8	4	M5	X	11.0	8.3	9.8	29	225
2	1995	ACURA	INTEGRGS-R	SUBCOMPACT	1.8	4	M5	Z	10.8	8.3	9.7	29	223
3	1995	ACURA	LEGEND	COMPACT	3.2	6	A4	Z	14.2	10.5	12.5	23	288
4	1995	ACURA	LEGEND COUPE	COMPACT	3.2	6	A4	Z	14.6	11.0	13.0	22	299

Table 4: Dataset after feature engineering

MODEL	VEHICLE CLASS	ENGINE SIZE(L)	CYLIN-DERS	TRANS-MISSION	FUEL TYPE	FUEL CONSUMPTION CITY (L/100KM)	FUEL CONSUMPTION HWY (L/100KM)	FUEL CONSUMPTION COMB (L/100KM)	FUEL CONSUMPTION COMB (MPG)	CO_2 EMISSIONS	MODEL FREQ	FUEL_CONS_HWY_CITY	ENG_POWER
INTEGRA	SUBCOMPACT	1.8	4	A4	X	11.6	8.3	10.1	28	232	14	19.9	7.2
INTEGRA	SUBCOMPACT	1.8	4	M5	X	11.0	8.3	9.8	29	225	14	19.3	7.2
INTEGR-GS-R	SUBCOMPACT	1.8	4	M5	Z	10.8	8.3	9.7	29	223	5	19.1	7.2
LEGEND	COMPACT	3.2	6	A4	Z	14.2	10.5	12.5	23	288	1	24.7	19.2
LEGEND COUPE	COMPACT	3.2	6	A4	Z	14.6	11.0	13.0	22	299	2	25.6	19.2

Table 5: Dataset after label encoding

MAKE	MODEL	VEHICLE CLASS	ENGINE SIZE(L)	CYLIN-DERS	TRANS-MISSION	FUEL TYPE	FUEL CONSUMPTION CITY (L/100KM)	FUEL CONSUMPTION HWY (L/100KM)	FUEL CONSUMPTION COMB (L/100KM)	FUEL CONSUMPTION COMB (MPG)	CO_2 EMISSIONS	MODEL FREQ	FUEL_CONS_HWY_CITY	ENG_POWER
0	2327	17	1.8	4	2	3	11.6	8.3	10.1	28	232	14	19.9	7.2
0	2327	17	1.8	4	27	3	11.0	8.3	9.8	29	225	14	19.3	7.2
0	2328	17	1.8	4	27	4	10.8	8.3	9.7	29	223	5	19.1	7.2
0	2542	0	3.2	6	2	4	14.2	10.5	12.5	23	288	1	24.7	19.2
0	2543	0	3.2	6	2	4	14.6	11.0	13.0	22	299	2	25.6	19.2

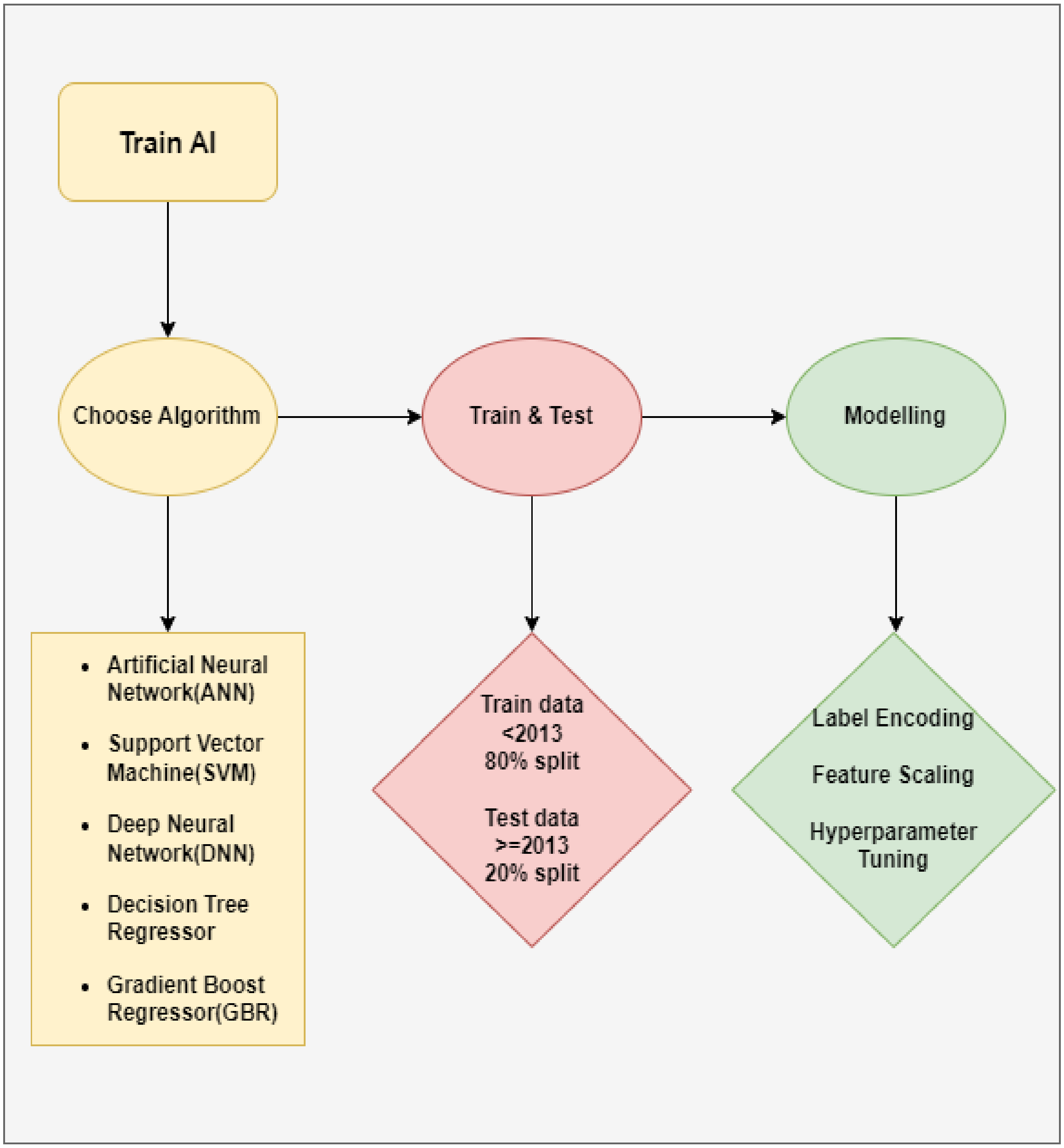

Figure 3: ML Algorithm flowchart

The feature engineering done on the dataset created three additional columns. The first new column named MODEL FREQ, encodes and counts the frequency of occurrences of the vehicle model. The second new column, FUEL_CONS_HWY_CITY sums up the fuel consumption of both city and highway and finally the last new column, ENG_POWER computes the product of 'Number of Cylinders' and the' engine size' to determine the engine power and fuel consumption. TABLE 4 shows how the data set looks after feature engineering.

Label encoding on the categorical features transformed the following categorical variables in the dataset ['MAKE', 'MODEL', 'VEHICLE CLASS', 'TRANSMISSION', 'FUEL TYPE'] into numerical values. The final table output can be seen in TABLE 5. Label encoding makes the data ready for scaling and modelling.

To examine the performance success of the machine learning algorithms in terms of CO_2 emissions forecasting, the dataset with model years between 1995 to 2012 is used to train the algorithms, and then model years within the last 9 years (2013–2021) are forecasted with different algorithms.

2.2 Machine Learning Algorithms

The train test split method comes next after deciding the X independent variables which in this data set are ['MODEL YEAR', 'MAKE', 'MODEL', 'VEHICLE CLASS', 'ENGINE SIZE(L)', 'CYLINDERS', 'TRANSMISSION', 'FUEL TYPE', 'FUEL CONSUMPTION CITY (L/100 km)', 'FUEL CONSUMPTION HWY (L/100 km)', 'FUEL CONSUMPTION COMB (L/100 km)', 'FUEL CONSUMPTION COMB (mpg)', 'MODEL FREQ', 'FUEL_CONS_HWY_CITY', 'ENG_POWER'] and the Y variable or target variable which is ['CO_2 EMISSIONS ']. TABLE 3 shows the proportion of the train test split on the dataset. The modelling process includes Label encoding, feature scaling and hyperparameter tuning after which the best-performing algorithm is used to forecast the CO_2 emission from 2022 – 2050 on retail sales in Canada.

In this research study, different machine learning algorithms were used – Artificial Neural Network (ANN), Support Vector Machine (SVM), Deep Neural Network (DNN), Decision Tree Regressor and Gradient Boost Regressor (GBR) are studied by using the tool Google Colab in the prediction of transportation-related CO_2 emissions in Canada.

2.2.1 Artificial Neural Network (ANN)

In our study, we employ an artificial neural network (ANN) as a computational model for tasks such as prediction, classification, and decision-making. The ANN consists of artificial neurons, like those found in the human brain. Neurons within the network are connected, and the strength of these connections, known as weights, modifies the inputs. Each neuron processes inputs and produces an output using a non-linear activation function [7].

For our supervised learning study, both inputs and outputs are provided to the network. The network compares its outputs to the desired outputs, and errors are propagated back through the system to adjust the weights. This iterative process continues as the weights are continuously adjusted. The structure of the ANN model algorithm is depicted in FIGURE 4.

2.2.2 Support Vector Machine (SVM)

Support Vector Machine (SVM) is one of the most popular Supervised Learning algorithms, which is used for Classification or Regression in Machine Learning. The goal of the SVM algorithm is to create the best line or decision boundary called a hyperplane, that can segregate n-dimensional

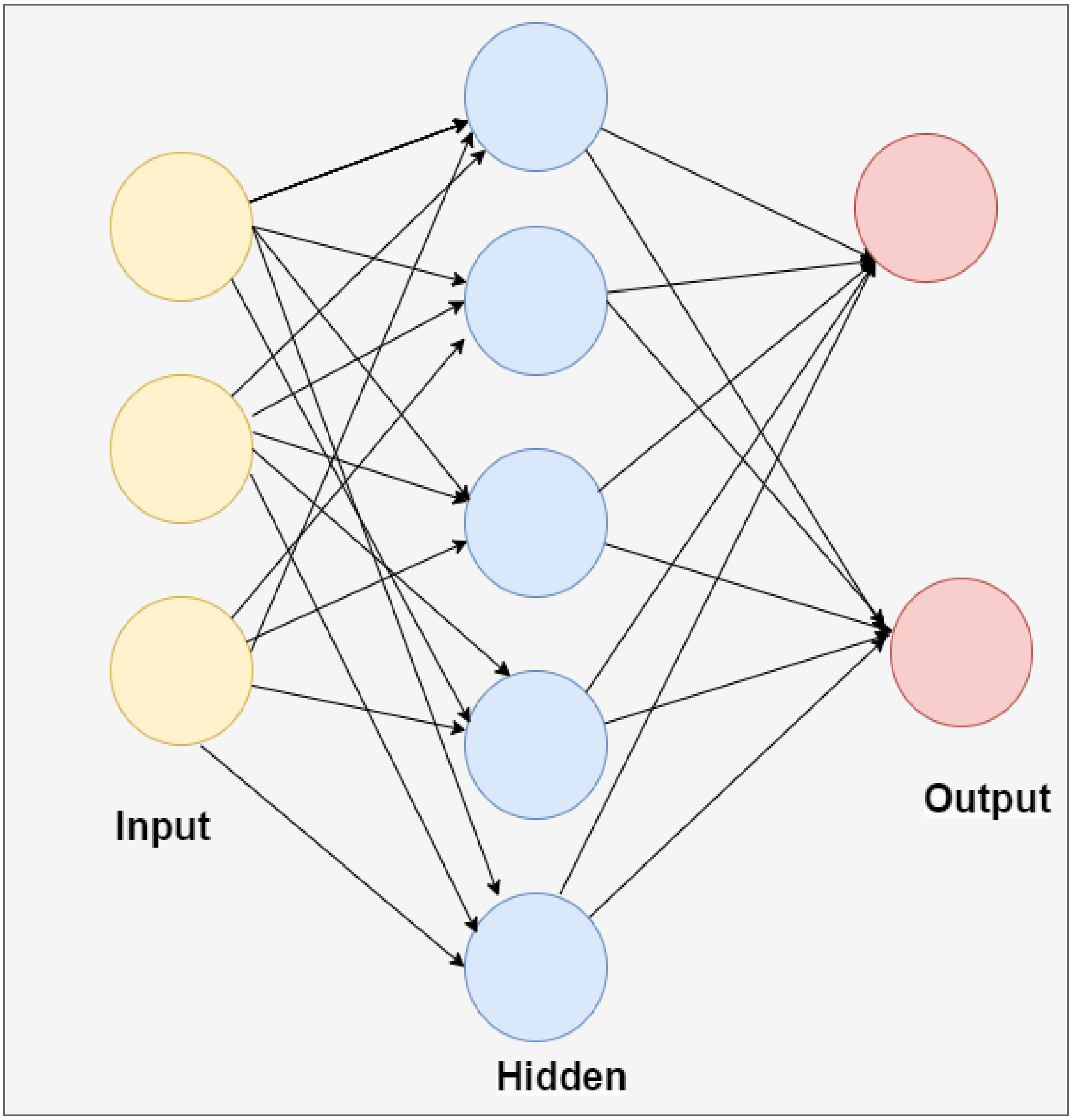

Figure 4: Architecture of ANN

space into classes so that we can easily put the new data point in the correct category in the future [8]. SVM chooses the extreme points/vectors that help in creating the hyperplane called support vectors [9].

2.2.3 Deep Neural Network (DNN)

A deep neural network (DNN) is an ANN with multiple layers of interconnected nodes between the input and output layers. Neural networks are widely used in supervised learning and reinforcement learning problems. These networks are based on a set of layers connected. The main purpose of a neural network is to receive a set of inputs, perform progressively complex calculations on them, and give output to solve real-world problems.

2.2.4 Decision Tree Regressor

A decision tree is a non-parametric supervised learning algorithm used for classification and regression. It consists of a hierarchical tree structure with nodes and branches FIGURE 5. shows a hierarchical, tree structure, which consists of a root node, branches, internal nodes, and leaf nodes. The algorithm selects the best feature to split the data into different classes or categories and the process continues until the tree is fully grown or a stopping criterion is met. The decision tree can then be used for predictions on new data by following the tree's path based on input feature values [10].

2.2.5 Gradient Boost Regressor (GBR)

Gradient boost is a machine learning algorithm that works on the ensemble technique called 'Boosting'. It can be used for predicting continuous target variables and categorical target variables (as a Classifier). When it is used as a regressor, the cost function is Mean Square Error (MSE) and when it is used as a classifier then the cost function is Log loss. Gradient boosting is one of the variants of ensemble methods where you create multiple weak models and combine them to get better performance in general.

2.2.6 Extreme Gradient Boosting (XGBoost)

XGBoost, or Extreme Gradient Boosting, is a powerful algorithm known for its efficient construction of boosted trees and ability to handle both classification and regression problems. It optimizes the objective function's value within the gradient boosting framework. XGBoost integrates multiple decision tree models to reduce variance and bias, resulting in improved performance. Its parallel tree boosting capability enables fast and accurate solutions for various data science problems [11].

2.2.7 Random Forest

Random Forest algorithm functions by randomly selecting a portion of the training data and creating a decision tree from it. This procedure is repeated multiple times, with each tree developed from a distinct subset of the data. The result of the algorithm is the prediction or most frequently occurring class among all the individual trees. Compared to merely forecasting summary metrics from

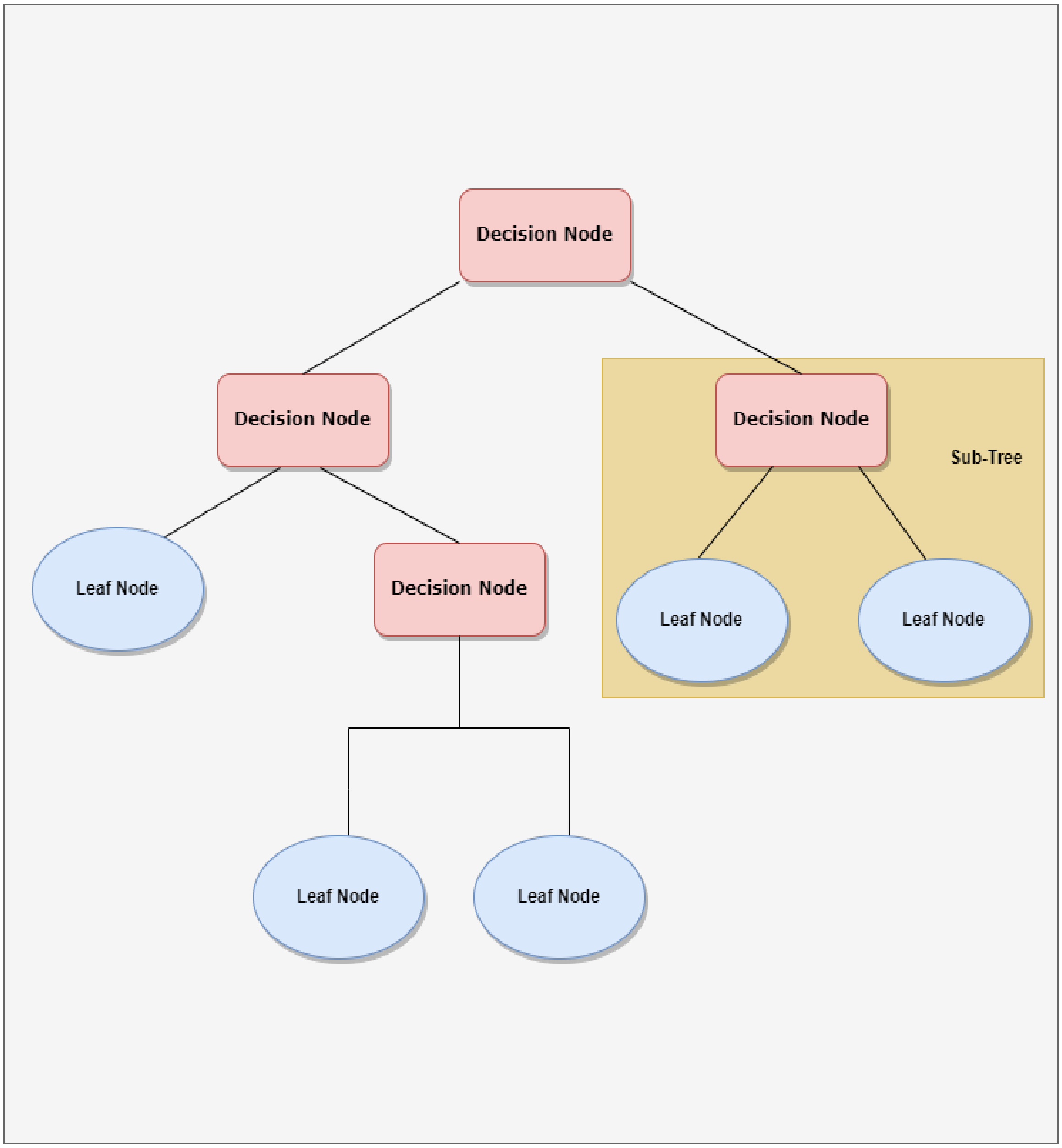

Figure 5: Decision tree flowchart

response curves, Random Forest delivers a more robust predictive algorithm to predict the complete system behaviour by taking both static and dynamic factors into account [11].

3. STATISTICAL BENCHMARK

This study discusses the performance success of the prediction results obtained from the machine learning algorithms with the six most used statistical metrics. These metrics are determination coefficient (R^2), root mean square error (RMSE), mean absolute percentage error (MAPE), mean bias error (MBE), relative root mean square error (rRMSE), and mean absolute bias error (MABE). The table in TABLE 6 summarizes the descriptions, equations, and performance criteria of these metrics.

In TABLE 4, y_i = actual value, $\hat{y}_i$ = predicted value, n = sample size or number of observations.

4. RESULT AND DISCUSSION

4.1 Forecasting of Transportation-based CO_2 Emission with ML Algorithms

Transportation-based CO_2 emissions in Canada are forecasted between the years 1995 and 2021 with five different machine-learning algorithms with six statistical metrics. FIGURE 6 illustrates the actual and forecasted CO_2 emissions over the 9 years. This tests the reliability of the different machine learning algorithms for CO_2 emission. It can be observed from FIGURE 6 that there is a gradual increase in CO_2 emission over the years and the actual CO_2 emission and predicted ones are close to each other.

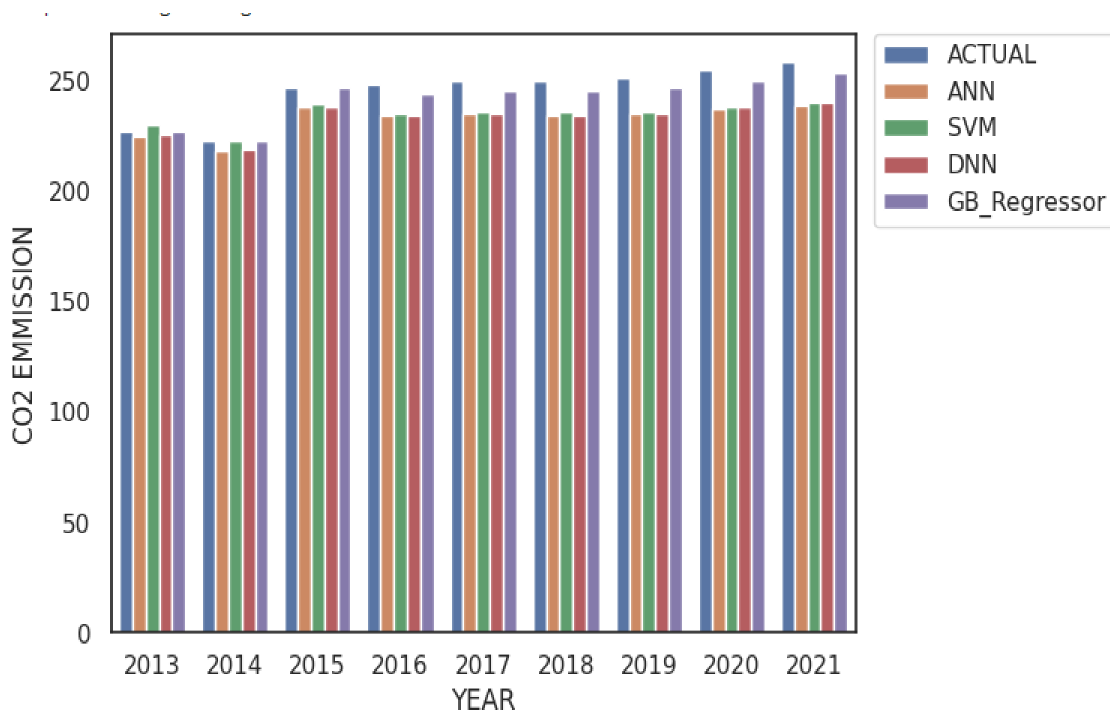

Figure 6: CO_2 Emission Per Year Per Algorithms

Table 6: Statistical metrics benchmark

Metrics	Equation	Description	Performance Criteria
R^2	$R^2 = 1 - \frac{SS_{Regression}}{SS_{Total}} = 1 - \frac{\sum(y_i - \hat{y}_i)^2}{\sum(y_i - y_i)^2}$	R Square is a good measure to determine how well the model fits the dependent variables. R-squared (R2) is a measure of how close the data points are to the fitted line. It is also known as the coefficient of determination.	A high R^2 value suggests a better fit for the model and a low value suggests otherwise. The higher the R-squared, the better the model fits your data.
RMSE	$RMSE = \sqrt{\sum_{i=1}^{n} \frac{(\hat{y}_i - y_i)^2}{n}}$	Root Mean Squared Error (RMSE) is one of the two main performance indicators for a regression model. It measures the average difference between values predicted by a model and the actual values.	Based on a rule of thumb, it can be said that RMSE values between 0.2 and 0.5 show that the model can relatively predict the data accurately.
MAPE%	$MAPE = \frac{1}{n} \sum_{i=1}^{n} \frac{\lvert\hat{y}_i - y_i\rvert}{y_i}$	The mean absolute percentage error (MAPE) is the sum of the individual absolute forecast errors, divided by the actual values for each period. It is an accuracy measure based on the relative percentage of errors	The closer the MAPE value is to zero, the better the predictions.
MBE	$MBE = \frac{1}{n} \sum_{i=1}^{n} (y_i - \hat{y}_i)$	Mean bias error is used to estimate the average bias in the model and to decide if any steps need to be taken to correct the model bias. underestimates the power output.	Mean Bias Error (MBE) can indicate whether the model overestimates or underestimates. The best value for MBE is zero. As the MBE value gets close to zero, the result is approaching the fitting one
rRMSE, %	$rRMSE = \sqrt{\frac{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(\hat{y}_i)^2}}$	rRMSE is achieved by dividing the RMSE value to mean actual value.	This metric is between 0-100%. A model is excellent when RRMSE < 10% and approaches 0, but when RRMSE > 30% it indicates poor performance.
MABE	$\frac{1}{n} \sum_{i=1}^{n} \lvert\hat{y}_i - y_i$	The MABE presents the absolute value of bias errors.	MABE takes the values from zero to ∞, and the smaller the value of MABE, the more relevant the result is.

TABLE 7 gives the numerical results of statistical metrics to make a better comparison among the algorithms.

Table 7: Experimental Analysis

	R_Squared	RMSE	MAPE	MBE	rRMSE	MABE
ANN	0.886	20.243	0.067	12.333	8.253	16.510
SVM	0.889	19.991	0.064	10.449	8.150	15.734
SVM_tuned	0.858	22.544	0.071	10.399	9.191	17.585
DNN	0.890	19.885	0.064	11.992	8.107	15.814
Decision Tree	0.990	5.928	0.014	3.097	2.417	3.500
Decision Tree_tuned	0.990	5.971	0.015	3.081	2.434	2.694
Random Forest	0.990	5.982	0.014	3.116	2.438	3.538
Random Forest_tuned	0.990	5.920	0.015	3.085	2.414	3.612
Gradient Boosting Regressor	0.993	4.867	0.015	2.925	1.984	3.639
GBClassifier_tuned	0.993	4.939	0.015	3.045	2.014	3.745
XGBoost Regressor	0.993	4.907	0.015	2.907	2.000	3.690
XGClassifier_tuned	0.993	4.895	0.014	3.063	1.995	3.633

It is seen from the calculated metric results, that the R^2 value varies from 0.8586 to 0.9934 for transportation-based CO_2 emission. R^2 being the most frequently used metric in discussing the success of the forecasting results concerning actual data, gives an idea of how the forecasting curves follow that of actual data. It can be observed from FIGURE 6 that, the difference between actual and forecasted data gets bigger after the year 2016, making all algorithms fail to capture the curves of the actual data after this year, only GB Regressor was closest to the actual data value. The evidence of this fact can be seen numerically from TABLE 7, the R^2 value of ANN, SVM and DNN has a close lesser value of 0.8 compared to Decision Tree, Random Forest, GB Regressor and XGBoost Regressor. Leaving Gradient Boosting Regressor with the highest R^2 value of 0.9934 which was visible in FIGURE 6. The best RMSE value in our study is also on the Gradient Boosting Regressor algorithm with a value of 4.86.

Another metric from this study is the MAPE and the smaller the MAPE, the better the model performance. MAPE$\leq$10% can be classified as "high prediction accuracy" and the forecasting results for each algorithm from our study can be categorized as "high prediction accuracy". The MAPE metrics in forecasting the transportation CO_2 emission are smaller than 10% for all our machine learning algorithms used in this study. The results of MAPE metric are calculated to be 0.067, 0.064, 0.063,0.0145, 0.0144, 0.0149, 0.0148 and 0.015 for ANN, SVM, DNN, Decision Tree, Random Forest, GB Regressor and XGBoost Regressor respectively.

The numerical results of the calculated MBE metric for each algorithm vary. The best value for MBE is closer to 0 and from our model, the XGBoost Regressor gave the best result of 2.90 followed by the Gradient Boosting Regressor with 2.92. From FIGURE 6, Gradient Boosting Regressor gave the highest forecast values for CO_2 emission among the other algorithms.

Another statistical metric discussed in this study is the rRMSE which scales the magnitudes between 0 and 100. A rRMSE<10%, classifies the prediction results to be "excellent". The forecasting results for all algorithms from our study can be categorized as "excellent" because they are all lesser than 10% as shown in TABLE 5. There is however a high variation between the rRMSE values of ANN, SVM, and DNN with respective values of 8.25, 8.15, 8.10, and Decision Tree, Random Forest, GB Regressor, XGBoost Regressor with respective values of 2.41, 2.43, 1.98, 2.01 and 2.00. GB Regressor gave the most excellent rRMSE value in our study.

The MABE result from TABLE 5 shows that Decision Tree, Random Forest, GB Regressor and XGBoost Regressor have no significant differences among the algorithms. The MABE value for Decision Tree is 3.50, for Random Forest is 3.53, for Gradient Boosting Regressor is 3.63 and XGB Regressor is 3.69. However other algorithms like ANN, SVM and DNN have a much higher MABE value of 16.5, 15.7 and 17.5 respectively.

4.2 Future Forecasting of Transportation based-CO_2 Emission in Canada

Forecasting of transportation-based CO_2 emissions helps regulate carbon emissions which reverses the impacts of global warming and gives a clearer air quality. With the right knowledge of what the future emission rate will likely be, innovative ideas on the solution to reducing transportation CO_2 emissions like purchasing more vehicles with better fuel economy or strategizing on an alternative mode of transportation, can be birthed. Our study shows the future forecasting of transportation-based CO_2 emissions in Canada and our models are backed by their performance metrics.

The above forecast trends illustrated in TABLE 8 and FIGURE 7 from our study predicted that the transportation-related CO_2 emission in Canada will decrease profoundly in the year 2050. These outputs demonstrate that the current solution in reducing CO_2 emissions in the transportation sector will give a good result over time in Canada and the Canadian Government's goal to reduce Canada's GHG emissions by 30% in 2030 and advance Canada's emissions to net zero by 2050 is visible in the Country.

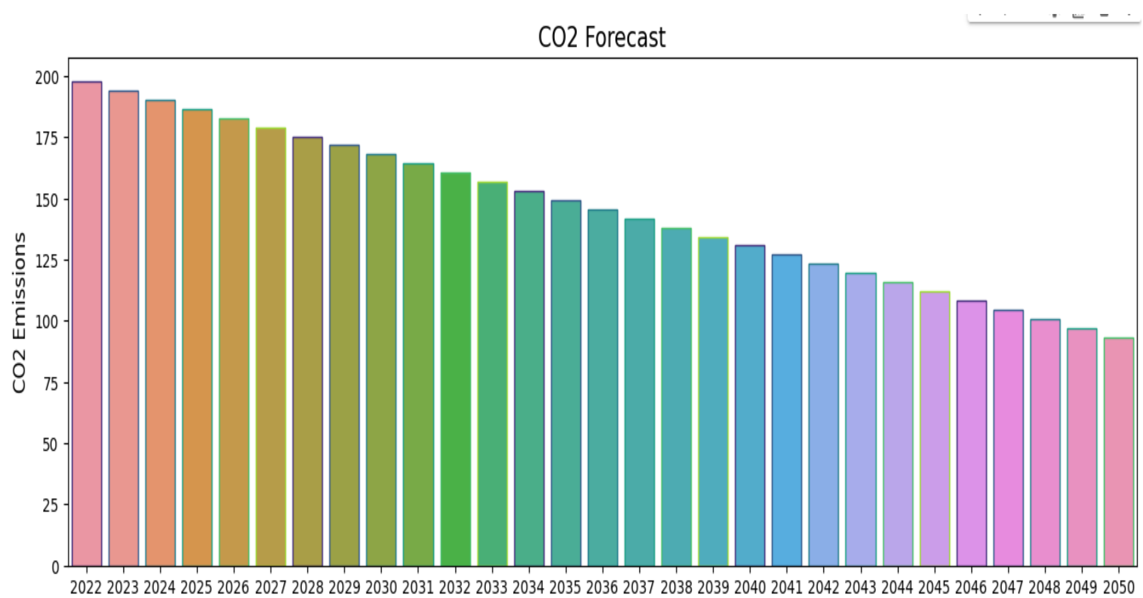

Figure 7: Graph of CO_2 Emission Prediction from 2022 - 2050

Table 8: CO_2 Emission Prediction from 2022 - 2050.

	YEAR	CO_2_EMISSION
0	2022	197.8185
1	2023	194.0912
2	2024	190.3639
3	2025	186.6366
4	2026	182.9093
5	2027	179.1820
6	2028	175.4547
7	2029	171.7274
8	2030	168.0001
9	2031	164.2728
10	2032	160.5455
11	2033	156.8182
12	2034	153.0909
13	2035	149.3636
14	2036	145.6363
15	2037	141.9090
16	2038	138.1817
17	2039	134.4544
18	2040	130.7271
19	2041	126.9998
20	2042	123.2725
21	2043	119.5452
22	2044	115.8179
23	2045	112.0906
24	2046	108.3633
25	2047	104.6360
26	2048	100.9087
27	2049	97.1814
28	2050	93.4541

5. CONCLUSION AND RECOMMENDATIONS

This study aims to forecast transportation-related CO_2 emissions in Canada. To make the prediction a success, five machine learning algorithms (Artificial Neural Network (ANN), support vector machine (SVM), Deep Neural Network (DNN), Decision Tree Regressor and Gradient Boost Regressor (GBR)) are used. 80% of the dataset with a year less than 2013 was used to train the model while the remaining 20% containing the remaining years was used to test our model. To evaluate the performances of the model, six statistical metrics are discussed on the performance success of the algorithms in the forecast.

From our study, the general performance metrics of our models show that Artificial Neural Network (ANN), support vector machine (SVM), and Deep Neural Network (DNN) algorithms have the lowest-performing result among all algorithms used in our paper. Meanwhile, Gradient Boost Regressor (GBR), Decision Tree Regressor and Boosting Regressor show the best results in forecasting CO_2 emission output.

The majority of the algorithms from this study have been categorized as “high prediction accuracy” in the forecasting of transportation-related CO_2 emission in terms of most metrics as explained in Section 4.1. This paper has presented satisfying forecasting results for CO_2 emissions arising from the transportation sector in Canada. Furthermore, the future forecasting of CO_2 emission results achieved from mathematical models confirms that the current policy and techniques used in reducing CO_2 emissions like the alternative fuel type are efficient and investing in more of those resources will drastically reduce transportation-based CO_2 emissions.

Based on the successful results from the paper, it is clear that Canada won’t face many threats in the upcoming years due to transportation-based CO_2 emissions. Therefore, it is strongly recommended that an increase in the use of Vehicle Alternative Fuel Types and introducing low carbon fuels will simultaneously help reduce fossil-fuel consumption from the transportation sector. Creating community awareness of environmental pollution and encouraging Canadian citizens to use alternative low-carbon fuels, implementing the free bus policies to reduce citizens’ fuel consumption and transitioning to electric vehicles fed by power plants will further help reduce carbon emissions from the transportation sector [12].

References

[1] Scheidel W, von Reden S. The Ancient Economy. Routledge. 2012.

[2] Mokyr J. Paths of Innovation: Technological Change in Twentieth-Century America. By David C. Mowery and Nathan Rosenberg. Cambridge: Cambridge University Press, 1998. Pp. 214. $27.95. J Econ Hist. 1999;59:240-242.

[3] Matsumura EM, Prakash R, Vera-Munoz SC. Firm-Value Effects of Carbon Emissions and Carbon Disclosures. Account. Rev. 2014;89:695-724.

[4] Birol F, Argiri M. World energy prospects to 2020. Energy. 1999;24:905-918.

[5] Long Z, Axsen J, Kitt S. Public Support for Supply Focused Transport Policies: Vehicle Emissions, Low-Carbon Fuels, and ZEV Sales Standards in Canada and California. Transp Res A. 2020;141:98-115.

[6] Ağbulut, Ü., 2022. Forecasting of Transportation-Related Energy Demand and CO_2 Emissions in Turkey With Different Machine Learning Algorithms. Sustain Prod Consum.2022;29:141-157.

[7] Gallo C. Artificial neural networks tutorial. InEncyclopedia of Information Science and Technology, Third Edition. IGI Global. 2015:6369-6378.

[8] Bansal M, Goyal A, Choudhary A. A comparative analysis of K-nearest neighbor, genetic, support vector machine, decision tree, and long short term memory algorithms in machine learning. Decision Analytics Journal. 2022;3:100071.

[9] Rao C., Zhang Y, Wen J, Xiao X, Goh M. Energy Demand Forecasting in China: A Support Vector Regression-Compositional Data Second Exponential Smoothing Model. Energy.2023;263:125955.

[10] Josephine PK, Prakash VS, Divya KS. Supervised Learning Algorithms: A Comparison. Kristu Jayanti Journal of Computational Sciences (KJCS). 2021:1-12.

[11] Nguyen H, Bui XN, Bui HB, Cuong DT. Developing an XGboost Model to Predict Blast-Induced Peak Particle Velocity in an Open-Pit Mine: A Case Study. Acta Geophys. 2021;69:1743-1755.

[12] Andersson, Ö., Börjesson, P., 2021. The Greenhouse Gas Emissions of an Electrified Vehicle Combined With Renewable Fuels: Life Cycle Assessment and Policy Implications.Appl Energy. 2021;289.