



The Research Analytics

(A Peer Reviewed and Open Access Journal)

5

Machine Learning Driven Carbon Emission Detection – A Review

Suchana goswami¹, Dr. Usha Sharma², Sukanya Ghosh Neogi³, Minakshi Bag⁴

- 1) Research Scholar, Department Of IT, Gauhati University
- 2) Assistant professor, Department of Computer Science, AKS University Satna MP, usha28383@gmail.com
- 3) Software Development Engineer, Accenture, Mexico, sukanyaresearchwork@gmail.com
- 4) Assistant Professor, Department of Computer Science and Engineering, MIT, Bishnupur.

Abstract

With global warming and the energy crisis, The China government has made a target to control and decrease carbon releases, planning to maximize emissions by 2030 and become carbon neutral by 2060. This paper studies how leveraging machine learning is used to predict carbon emissions by reviewing research papers. It found that machine learning is a popular topic in carbon emission prediction. There are three machine learning models—Random Forest, Gradient Boosting and Support Vector Machine were used. The accuracy of these models improved by using a method called differencing, which helps make data more stable and creates extra features for the model to learn from. Machine Learning algorithms can also predict environmental conditions and warn if pollution levels go beyond safe limits.

Keywords: CO₂ emission, Deep Learning (DL), Machine Learning (ML), Model Prediction, Statistical.

1.INTRODUCTION

Energy plays a crucial role in financial growth and environmental sustainability. It is recognized as the primary driver of daily activities, with its availability and accessibility being essential for both industries and households. However, carbon emissions from fuels such as coal, paraffin, gas and other natural resources have significantly contributed to universal environmental degradation. This has led to additional repeated organic calamities, drinkable watershortages, and power insecurity, while a certain level of carbon dioxide (CO₂) is essential for maintaining ecological balance, excessive emissions have serious environmental consequences. Firstly, carbon dioxide (CO₂) discharges are a main provider to climate crisis. Secondly, understanding the carbon dioxide (CO₂) emission footprint is necessary for developing effective policies to action environmental shift. Thirdly, it is crucial to address the environmental and financial losses caused by carbon dioxide (CO₂) emissions. Additionally, analysing the impact of carbon dioxide (CO₂) emissions can help assess their effects on Gross domestic product (GDP) fluctuations, stock market trends, the emergence of new and existing diseases, and disruptions in air quality. Addressing these issues is critical for sustainable development and environmental protection.

As per the globe's biggest finance, the Chinese administration seeks to meet natural durability via aset ofrules and various approaches. It also targets to have carbon dioxide (CO₂) releases peak before 2030 and stable carbon objectivity before 2060 (Li. et al 2021).

Nevertheless, challenges related to data availability and quality persist, as satellite-based observations are frequently impacted by cloud cover, limitations in spatial resolution, and atmospheric distortions.(Li, Siu, & Zhao, 2021)

2. KEYWORDS

CO₂ emission, Deep Learning (DL), Machine Learning (ML), Model Prediction, Statistical.

3. LITERATURE REVIEW

To approximate these emissions, this study presents a Machine Learning Emissions Calculator (MLEC), a tool designed to help communities' various approaches to utilizing a Machine Learning Emissions Calculator enable a deeper understanding of the environmental footprint associated with training machine learning models. Despite its hard works, the calculator remains an estimate of true discharge for several reasons. One key challenge is global load balancing, which affects the accuracy of emissions estimation. Other elements, such as changes in energy sources and hardware efficiency, further contribute to the complexity of precisely measuring the natural effect of ML training (Lacoste et al. 2019). This paper aims to forecast greenhouse gas emissions, including carbon dioxide, methane, nitrogen dioxide using artificial neural networks, deep learning and support vector neural networks. The study focuses on releases from the electricity production sector in Turkey. The researchers are committed to developing accurate forecasting models to improve the understanding and prediction of GHG discharges in this sector (Bakay et al. 2020). In this work, the authors present Carbon tracker, the tool for monitoring and forecasting energy usage and carbon emissions training deep learning models.

The environmental result of deep learning can be relieved by examining ways to enhance energy efficiency. Additionally, by increasing awareness among practitioners about their power consumption and carbon balance, they can diligently take action to decrease it whenever it's possible (Anthony et al. 2020). This paper applies a non-parametric kernel prediction algorithm in machine learning for forecasting Carbon dioxide emissions.

Traditional parametric modelling proximity and the gaussian process regression algorithm were also used, and their calculation performance was evaluated. Additionally, a Bayesian non-parametric kernel prediction algorithm was applied to improve Carbon dioxide (CO₂) emission forecasting. The discussion on prediction performance focuses on six primary factors controlling carbon emissions. Upcoming research must aim to further review the entirety of the set of driving factors and assess the benefit of the model forecasts by comparing them with other similarly used models (Meng et al. 2021). Some existing models exhibit noticeable limitations in forecasting carbon dioxide (CO₂) emissions accurately. This study executes four forecast models using SARIMA, which is based on ARIMA. The primary objective is to compare these models and recommend the most effective one for predicting future carbon dioxide (CO₂) emissions. Additionally, this paper focuses on developing an upgraded carbon dioxide (CO₂) emission prediction model. Future work could extend this research by creating a comprehensive web or mobile application, allowing users to track Carbon dioxide emissions with just a touch of a fingertip (Meng et al. 2022). In this paper, a carbon dioxide (CO₂) emission prediction model stem from Bi-LSTM (Bidirectional Long short-term Memory) is constructed. The findings indicate that carbon emissions pose a significant challenge, which is likely to deteriorate in future due to the significant emissions from China and India. This study targets to analyze the key factors influencing carbon emissions in the ONOR countries, particularly in South Asia and to provide prognostication of their development trends from 2020 to 2035 (Amir et al. 2022). This research proposes a novel technique for Unmanned Aerial Vehicle based.

Atmospheric condition analysis and carbon emission detection. It utilizes deep learning on UAV-collected weather report, which is analysed for noise reduction and smoothing. The processed data is then extracted and categorized using a Gaussian Belief Deep Neural Network and a Spatial Convolution Q-Swarm metaheuristic optimization approach. However, further refinement and development are still required to enhance the model's effectiveness (Mohanty et al. 2024). According to researchers, the Random Forest algorithm outperforms multiple traditional models, authenticate that visual characteristics play a significant role in explaining carbon emissions. Their findings support the usefulness of using street view images as a single data source to effectively forecast neighborhood-level residential carbon emissions. One of the key challenges in carbon emission (CE) modelling is the limitation of data source detailed energy utilization data is often unavailable in multiple cities and smaller cities, in particular, lack such data due to insufficient funding for carbon emissions data collection. Additionally, fine grained sociodemographic data is not universally accessible. Modelling accuracy is also constrained by the increasing complexity of urban form variables. Ideally, concrete detail on the periods when street view images were gathered should be obtained to account for seasonal variations in street environment. More advance studies could address this by accumulating time-series data and developing separate models for different seasons (Shi et al. 2024). This study examines the present state of machine learning applications in carbon emission forecast through a detailed review of research papers. Findings indicate that machine learning has become a key topic in the development of carbon emission forecast models. The primary models used for carbon emissions prediction are based on Back Propagation Neural Networks, Random Forests and Extreme Learning Machines. However, given the rapid advancement of machine learning, the field also experiences a high rate of model obsolescence.

To enhance research in carbon emission prediction, scholars should stay informed about the latest developments in machine learning (ML), including improvements and optimization techniques for these algorithms (Zhao et al. 2024). Many research studies have focused on improving the prediction of Carbon dioxide (CO₂) emissions. This study evaluates the outcomes of 14 in forecasting regularly Carbon dioxide emission from January 1, 2022, to September 30, 2023, across the four most polluting regions. The models include statistical approaches such as ARMA, ARIMA, and SARMA, as well as three machine learning models which are gradient boosting, random forest, super vector machine because deep learning models needs significant computational power, they have high processing demands, this study recommends machine learning models utilizing ensemble techniques such as voting and bagging for daily carbon dioxide (CO₂) emission prediction (Ajala et al. 2025)

4. DIFFERENT TECHNIQUES USED IN CARBON DIOXIDE (CO₂) EMISSIONS

Different Machine Learning Techniques (MLT) have revolutionized in ground of CO₂ emissions analysis next to making it possible to automatically process data, recognize patterns and predict with high accuracy. SLA like Gradient Boosting, Random Forest, Super Vector Machine are used to forecast CO₂ based on past data & geospatial factors. Deep learning methods (DL), specifically Convolution Neural Networks (CNN), have shown to be very effective in analyzing satellite imagery to find emission hotspots and detect changes in the environment (Magadam et al. 2025).

5: TECHNIQUES

5.1 Machine Learning :- It's a field of AI that algorithms to discover hidden patterns in data. This enables to forecast novel, comparable information devoid of requiring specific indoctrination intended for each task. Its impact extends to autonomous vehicles, drones, and robots, improving their ability to adapt to changing environments. ML can be classified into three types.

5.1.1 Supervised Machine Learning: It is trained on the considered information. They learn to chart contribution feature to targets based on considered education information. Supervised Machine Learning (SML) can be categorized in the direction of various type, every by means of separate explanation and submission. Here are some of the most common types of algorithms.

- *Liner Regression:* It is SLA used for regression everyday jobs, which predicts a continuous output based on contribution description
- *Random Forest:* RF another time has completed up of multiple decision tresses that work together to make prediction. Every hierarchy in afforest is built by means of a different specimen of input data and a specific set of features.
- *Gradient Boosting:* Boosting builds an additional authoritative replica by combine multiple feeble learners, such as verdict foliage. It keeps improving by adding new models that mend the mistakes of the previous.
- *Gaussian Process Regression:* It is a supervised learning technique. It is a collection of random variables, where any finite subset follows a various normal distribution. Instead of learning a fixed parametric function; GPs define a distribution over functions, making them highly flexible.

5.1.2 Super Vector Machine: This algorithm creates a hyperplane To divide N-dimensional space into groups and determine the correct category for a new data point.

5.1.3 Artificial Neural Network: An ANN consists of artificial neurons, known as units, which are organized into layers. These layers together form the complete ANN system. An ANN typically consist of three types of layer of three. First input layer, second output layer, third one & more hidden layers. First layer receive statistics information from different external sources, which the NN analyses. This data is after that processed from side to side one to more hidden layers, where it is distorted into carrying great weight in sequence. At long last, the production coating produces the concluding consequence based on the processed records.

5.1.4 Convolution Neural Networks (CNN): It is a specialized deep learning (DL) structural design primarily used for workstation hallucination errands. It has superior feature of an ANN designed to haul out advance character & well-suited images and videos recognition task, where patterns are crucial. This CNNs are designed to analyze data that comes in the form of multiple arrays (like RGB images), preserving the spatial relationships between pixels.

CNNs are composed of several key layers:-

- Input layer,
- Convolution layers,
- Pooling layers
- Fully Connected layers

Each one layer plays a specific role in processing and analysing the data to identify important features and patterns. The strength of CNNs lies in their potentiality to learn automatically about spatial hierarchies of features from low-level edges to high-level object parts with minimal pre-processing.

6. METHODOLOGY

6.1 Data Collection:

The dataset contains day by day synchronized CO² emissions statistics recorded as of (<https://carbonmonitor.org>). It is source as of the Co² monitoring scheme and was specifically selected to Regular economic activities after the world started getting better on or after covid pandemic. Over the past decade, CO² emissions, measured in million tons per day, have shown a fluctuating trend, influenced by economic cycles, policy changes, and technological advancements in energy production.

It represents the overall output from six key factors:

- Domestic and international aviation
- Ground transportation,
- Power generation,
- Industrial production,
- Residential consumption.

The data covers emissions from five major regions: China, India, the United States, the European Union (EU27) and the United Kingdom (Ajala et al. 2025).

6.2 Data Pre-processing:

No accusation designed formisplacedstatisticsberequisiteseeing that the dataset haverefusalgone astraystandards. The explanationstatisticsdispensationstepladderconsist of the enlargedDickey-Fuller (DF) analysis was wornto establish whether informationbemotionless. Ensuring stationary is essential as it guarantees that signifyalong withtouchstonedeparturestay behindconstantlarger thanmoment in time. It'spredominantlyimperative for geometricsculptso as to depend resting onassumption to make reliable predictions (Ajala et al. 2025).

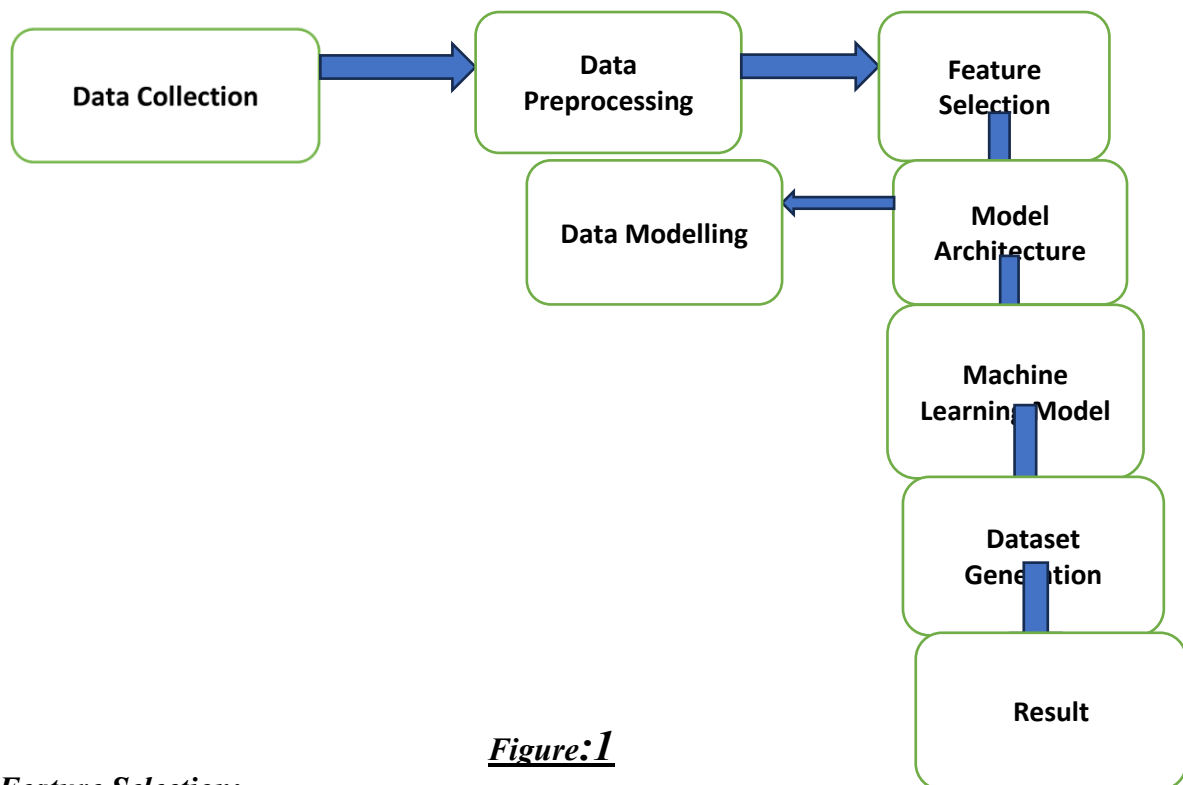


Figure:1

6.3 Feature Selection:

Feature selection is crucial in developing accurate and efficient machine learning models for CO₂ emissions prediction. It helps in reducing dimensionality, improving interpretability, and enhancing model performance. Some techniques of feature selection is

- ❖ *Correlation Analysis* – Checking if some sectors are highly correlated and removing redundant ones.
- ❖ *Random Forest Feature Importance* – Identifying which sectors contribute most to CO₂ emissions.
- ❖ *Lasso Regression* – Selecting key variables while reducing less significant ones.

6.4 Machine learning Model:

When we do not make strong assumptions about how the input and output are related, we use nonlinear models. As per the researchers (Li et al. 2021)these models are useful when

working with datasets that have many features, especially if some of them are related to each other.

- Here are some common examples of nonlinear models, often used in time series forecasting, regression, or machine learning tasks:

6.4.1 *Classification & Regression Tree (CART)*: Machine learning for in cooperation cataloguing (sorting into categories) furthermore deterioration (predicting numbers).

6.4.2 *Support Vector Regression (SVR)*: A regression model (used for predicting numbers). It is like chalk and cheese from SVM, Which is used, premeditated for classification. Main dissimilarity is that SVR uses two slack variables while SVM uses only one when finding the best decision boundary.

6.4.3 *K-Nearest Neighbours (KNN)*: Can be previously owned intended for in cooperation cataloguing and deterioration. In cataloguing, it predicts the category of a novel statistics aim looking at categories of its adjacent neighbours. In deterioration, it predicts a value by averaging the values of nearby data points.

One advantage of machine learning is that it can handle nonlinear, meaning data that does not follow a simple straight-line pattern. If a system, such as CO₂ emissions and its six input variables, is nonlinear, then nonlinear models are a better choice.

7. ADVANTAGE

- **Improved accuracy** – ML models be capable of analyze enormous amount of environmental and industrial facts to predict emissions more accurately than traditional methods.
- **Real-time Monitoring** – Machine learning enables continuous monitoring of CO₂ levels, helping industries and governments take immediate action when emissions exceed limits.
- **Predictive Insights** – ML can forecast future emissions trends based on historical data, aiding in better decision-making for sustainability policies.
- **Anomaly Detection** – Machine learning can detect irregularities in emissions, helping identify faulty equipment or inefficiencies in industrial processes.

8. DISADVANTAGE

- **High Computational expenditure** – education composite ML models requires note worthy computational influence, which itself know how to supply on the road to CO₂ emissions particularly. If powered by non-renewable energy sources, the environmental impact may be significantly higher due to increased greenhouse gas emissions and resource depletion.
- **Data superiority & accessibility** – mechanism rudition models rely resting on huge datasets, but accurate along with comprehensive CO₂ emissions statistics may not always be available or may be inconsistent across regions.
- **Complexity & Interpretability** – countless ML models, specially deep learning-based ones, work as "black boxes," making it difficult to understand how they make decisions, which can reduce trust in their recommendations.
- **Dependence on Historical Data** – ML models learn from past data, which may not always reflect future trends accurately, especially in a rapidly changing climate and policy landscape.

8. DISCUSSION

Governmental efforts to reduce global warming depend on the precise forecasting of regularly real-time CO₂ emissions. In order to anticipate daily CO₂ emissions across different locations, this study compares several machine learning (ML) models. In environmental science, machine learning has become a revolutionary force, providing creative answers to enduring problems. Its uses in a variety of fields, including as pollution monitoring, biodiversity assessment, and

remote sensing, advance our knowledge of ecosystem. Continuous investigation and teamwork are opening up new ML possibilities, advancing environmental science and encouraging sustainable practices.

REFERENCES:

- 1) 1. Ajala, A. A., Adeoye, O. L., Salami, O. M., & Jimoh, A. Y. (2025). An examination of daily CO₂ through a comparative analysis of machine learning, deep learning and statistical models. *Environmental Science and Pollution Research* (2025) 32:2510–2535.
- 2) 2. L. Pham, T. H., Rafieizonooz, M., Han, S., & Lee, D. E. (2021). Sustainability 2021, 13, 13579. doi:<https://doi.org/10.3390/su132413579>
- 3) 3. Mohanty, S. N., Dash, B. B., hanmugasundar, G., MGM4, J., Aswani, I., Sundaram, A., & Varghese, I. k. (2024). UAV Databased Temperature Patterns Analysis with Carbon Emission Detection Using Deep Neural Network. *Remote Sensing in Earth Systems Sciences*.
- 4) 4. Aamir, M., Bhatti, M. A., Bazai, S. U., Marjan, S., Mirza, A. M., Wahid, A., . . . Bhatti, U. A. (2022). Predicting the Environmental Change of Carbon Emission Patterns in South Asia: A Deep Learning Approach Using BiLSTM. *Atmosphere* 2022, 13, 2011. doi:<https://doi.org/10.3390/atmos13122011>
- 5) 5. Anthony, L. F., Kanding, B., & Selvan, R. (2020). Carbontracker: Tracking and Predicting the Carbon Footprint of Training Deep Learning Models. *arXiv:2007.03051v1 [cs.CY]* 6 Jul 2020.
- 6) 6. Bakay, M. S., & Ağbulut, Ü. (2020). Electricity production based forecasting of greenhouse gas emissions in Turkey with deep learning, support vector machine and artificial neural network algorithms. *Journal of Cleaner Production*. doi:<https://doi.org/10.1016/j.jclepro.2020.125324>
- 7) 7. Lacoste, A., Luccioni, A., Schmidt, V., & Dandres, T. (2019). Quantifying the Carbon Emissions of Machine Learning. *arXiv:1910.09700v2 [cs.CY]* 4 Nov 2019.
- 8) 8. Li, S., Siu, Y. W., & Zhao, G. (2021). Driving Factors of CO₂ Emissions Further Study Based on Machine learning. *Frontiers in Environmental Science*. Retrieved from www.frontiersin.org
- 9) 9. Ma, N., Shum, W. Y., & Han, T. (2021). Can Machine Learning be Applied to Carbon Emissions Analysis: An Application to the CO₂ Emissions Analysis Using Gaussian Process Regression. *Frontiers in Energy Research*.
- 10) 10. Magadum, T., Garg, K., Murgod, S., Yadav, V., Mittal, H., & Kushwaha, O. (2025). Geospatial Analysis in Machine Learning for CO₂ Emissions Prediction Analysis in 2100: A Continent-Wise Analysis. *Preprints.org* (www.preprints.org). doi:[10.20944/preprints202502.0729.v1](https://doi.org/10.20944/preprints202502.0729.v1)
- 11) 11. Meng, Y., & Noman, H. (2022). Predicting CO₂ Emission Footprint Using AI through Machine Learning. *Atmosphere* 2022, 13, 1871. doi:<https://doi.org/10.3390/atmos13111871>
- 12) 12. Shi, W., Xiang, Y., Ying, Y., jiao, Y., Zhao, R., & Qiu, W. (2024). Predicting Neighborhood-Level Residential Carbon Emissions from Street View Images Using Computer Vision and Machine Learning Machine Learning. *Remote Sens.* 2024, 16, 1312. doi:<https://doi.org/10.3390/rs16081312>