Predicting Carbon Dioxide Emissions using Machine Learning Models

Aarav Chhabra
University of Washington
Seattle, WA

ABSTRACT
In India, carbon dioxide (CO_2) concentrations have been progressively rising, with a striking increase of 40% over the past decade, outpacing the global average. The per capita CO_2 emissions in India is equal to 0.41 tons per person, which is an increase by 1.49 above the figure of 1.90 tons per person in 2022. This concerning development is largely attributed to swift industrial growth, urban expansion, and rising energy demands. The demand for precise forecasting and reduction of CO_2 emissions has become more urgent, as elevated CO_2 levels are contributing significantly to changing climate. While the shift to renewable energy sources is vital for decreasing global CO_2 emissions, India's dependence on non-renewable persists. In relevance to this, it is proposed for a novel approach to forecast CO_2 levels in India using machine learning models. This study utilized a variety of machine learning models, comprising of support vector machines, linear regressions, and polynomial regressions, for analyzing historical data concerning CO_2 emissions and energy usage. This study's findings indicate that the threshold for critical CO_2 levels, set at 5000 ppm, is projected to be reached by 2082. This study's results show that these models can effectively predict CO_2 levels in India with high accuracy, providing valuable insights for future policy changes. By identifying patterns and trends in CO_2 emissions, this study can develop strategies aimed at mitigating climate change and fostering sustainable energy practices. This research highlights the importance of machine learning-based forecasting in supporting India's shift to a carbon neutral economy and achieving its ambitious carbon reduction goals.

Keywords
Machine Learning, CO_2 emissions, Prediction, India

1. INTRODUCTION
Climate change pertains to alterations in temperatures and weather patterns that occur over extended periods of time. Natural phenomenon like fluctuations in solar radiation or tectonic shifts can affect planetary warming patterns. For more than 200 years, Human activities have been the leading cause of global warming, mainly because of the combustion of carbon-based energy sources.

Heat-trapping gases are released into the air through the burning of hydrocarbon fuels, creating a barrier that retains solar heat, resulting in a rise in global heat levels. Heat-trapping gases are released into the air through the burning of hydrocarbon fuels, creating a barrier that retains solar heat, resulting in a rise in global heat levels. Carbon dioxide and other gases like methane are the primary heat-trapping gases as they capture infrared energy, accelerating the Earth's warming process. These emissions stem from activities like gasoline-powered transportation and the heating of buildings with coal. Similarly, deforestation and land use changes also release significant amounts of carbon dioxide. Methane emissions, on the other hand, largely originate from sectors such as agriculture and oil and gas production. Energy, industry, transportation, construction, agriculture, mining, and forestry are major contributors to greenhouse gas emissions.

In 2020, there was a decrease in carbon dioxide emissions due to the global health crisis caused by Corona Virus, but worldwide energy-based emissions still reached 31.5 gigatons. This sustained the concentration of carbon dioxide in the atmosphere at record levels, reaching an average of 421.08 ppm in 2023. Projections for 2024 indicate that energy-related CO_2 emissions will rebound by 6.52%, driven by the recovery in coal, oil, and gas demand as the global economy strengthens.

The recently published IPCC report on August 9, 2021, presents new forecasts regarding the probability of exceeding the 1.5°C limit for planetary heating in the near future. The report emphasizes that achieving the target of constraining warming to approximately One-point-five degrees to two degrees of warming will likely be unachievable without immediate, significant, and far-reaching cuts in heat-trapping gas emissions.

The document finds that since 1850-1900, human activities have contributed to about 1.1°C of global warming through greenhouse gas emissions. It further projects that within the next two decades, global temperatures are likely to reach or surpass the 1.5°C threshold. These conclusions are based on improved observational datasets for analyzing past warming trends and advances in understanding how the climate system responds to emissions from man-made gases that trap heat in the atmosphere.

The impact of planetary warming patterns on our planet is becoming more and more apparent as each year goes by. India, being the world's most populated country, is particularly affected by this crisis. Unpredictable monsoon seasons, droughts, floods, and extreme temperatures are among the consequences that the Indian population is currently facing and will continue to confront in the coming decades. Despite India's commendable efforts in combatting climate change, more extensive measures will be necessary to shield the country from the impending changes.

The monsoon season is arguably the most significant meteorological event in India. Each year, the summer monsoons provide essential irrigation for crops and replenish aquifers, playing a crucial role in supporting India's economy, which heavily relies on rainfall-dependent activities like agriculture.

The delicate monsoon pattern is changing due to climate change and other climatic phenomena like El Niño, making the rainy season more unpredictable. This results in both prolonged dry spells and unexpected, intense bursts of rain, which produces flooding in some places and drought in others. For example, there were notable variations in the 2023 monsoon season, resulting in a nationwide rain deficit [1]. The areas most impacted were the East and North-East, which saw over 20% less rain than usual. The drought emergency declared by India in August of that year coincided with the monsoon's poorest month. As of February 2024, over 25% of India's land was impacted by at least a moderate drought, more than twice the amount of area affected in the same month the previous year. The dry conditions do not appear to be granting a truce.

India which is positioned third in the list of countries emitting greenhouse gases globally, faces unique challenge of safeguarding its population from the adverse impact of planetary-warming and swiftly adopting a greenhouse gas-neutral environment. In late 2021, at a global climate meeting, India announced its goal of reducing its carbon footprint to zero by the year 2070, in accordance with Paris Agreement (Article 4- Para 19) aimed at combating climate change [2]. India has shared its plan to build a green economy and cut carbon pollution with United Nations Framework Convention on Climate Change. This plan shows India's dedication to reaching zero carbon emissions by the year 2070. The principles of fairness & climate justice, as well as the concept of Shared Duties based on Varying Circumstances and Capabilities, act as the base of India's objective of long-term low-carbon development strategy which is underpinned by seven crucial shifts [3]. These encompass:

(1) the promotion of carbon-neutral electrical systems that align with growth objectives,

(2) the establishment of an inclusive, efficient, and unified transportation system,

(3) the promotion of changes in city development, resource and energy efficiency in structures, and eco-friendly urban growth,

(4) the progress towards decoupling economic expansion from pollution and the creation of rational, cutting-edge, and eco-friendly manufacturing sector,

(5) the creation of methods for removing greenhouse gases and related technological innovations,

(6) the enhancement of green cover in line with eco-friendly and socioeconomic contemplations,

(7) the identification of economic and financial requirements for eco sustained growth.

The strategy outlines the global and domestic commitments relevant to each transition, existing policies and programs, various key elements, probable advantages, and challenges adversities with each transition. In accordance with the 1988 National Forest Policy (NFP), which aims to bring more than a quarter of the nation's land under forest cover and ensure 75% of such coverage in mountainous regions, several afforestation projects are managed by the Ministry of Environment, Forest, and Climate Change (MoEFCC). The MoEFCC offers assistance to states and union territories for implementing tree-planting efforts through government programs like the National Mission for a Green India (GIM), the Nagar Van Yojana, and compensatory afforestation, which is regulated by the Compensatory Afforestation Fund Management and Planning Authority (CAMPA). In addition, states and union territories carry out their own afforestation projects, including initiatives funded by the Mahatma Gandhi National Rural Employment Guarantee Act (MGNREGA).

The Ministry of Environment, Forest, and Climate Change (MoEFCC), Government of India, in collaboration with the U.S. Agency for International Development (USAID), launched the Trees Beyond Forests in India (TOFI) initiative, which is being implemented across seven states— Rajasthan, Uttar Pradesh, Andhra Pradesh, Tamil Nadu, Odisha, Assam and Haryana [4]. This program is allocated $25 million for a duration of five years and aims to expand tree planting outside forests to enhance ecological services, particularly carbon sequestration, and create more eco sustained livelihoods and socioeconomic prospects for rural communities.

Previous research on climate change has highlighted the irreversible impacts of carbon dioxide emissions and offered precise calculations for the necessary reduction rates to achieve safe atmospheric CO_2 levels in the United States. Expanding on this work, the study also explores the historical trajectory of CO_2 emissions in India and projects when the country's carbon dioxide levels may reach 5000 ppm.

However, there is a research gap in providing a comprehensive analysis of Carbon emissions in India. The aim of this research is to fill this void by considering emissions from different sources. Accurate predictions of greenhouse gases like CO_2 emissions from various sources can assist establish specific timelines for necessary changes in these levels in India.

CO_2 predictions are particularly important for industries with high energy consumption or production. Anticipating energy consumption and carbon emissions is crucial for supporting a company's initiatives to improve energy usage efficiency and cut down on carbon pollution. Typically, historical data is necessary for these purposes while quantitative and data-driven techniques such as ARIMA (Auto-Regressive Integrated Moving Average) can be used, machine learning and AI models are best suited for making predictions.

The field of machine learning (ML) has seen significant advancements and is now widely applied across various industries, including healthcare, manufacturing, logistics, and even climate change. Unlike traditional statistical models, ML techniques leverage multiple features and algorithms to provide more advanced and accurate predictions.

ML and AI techniques involve creating computational and statistical models that learn from provided data and make predictions about projected values [5][6]. When appropriately modelled, ML algorithms can generate highly precise results. Many proven algorithms have been employed to predict emission values of CO_2.

For example, Support Vector Regression (SVR) algorithms are commonly applied in both regression and classification tasks, utilizing labelled datasets during the training process [7]. These algorithms operate by establishing decision boundaries using the concept of decision planes. One application involved using SVR to forecast the year when CO_2 emissions would reach critical levels, using different input variables related to carbon emissions [8].

Additionally, linear regression model is utilized to forecast the value of one variable using another [9]. This method calculates the coefficients of a linear equation with one or more predictors to accurately estimate the outcome variable's value. By generating a line or surface that minimizes the gap between predicted and actual results, linear regression was used to project the year when carbon emissions would reach critical levels based on historical CO_2 emission data.

2. RESEARCH OBJECTIVES

In light of the aforementioned circumstances, the following topics are the focus of this research, which seeks to identify, examine, and provide significant findings on:

(1) The year that CO_2 emissions surpass the threshold of 5000 parts per million.

(2) The necessary decrease and turnaround of the present emission levels to lower them to the 316-ppm non-toxic level.

(3) How the various social and economic sectors within a nation are interconnected and the effect of these connections on carbon dioxide emissions.

This study presents research on an Indian dataset that examined the country's annual carbon emissions over time. The many characteristics are listed, examined, and their connection to carbon emissions is discussed. The important elements are noted and identified. Machine learning-based models were created and applied to train and forecast potential carbon emission levels in the future. The models were then used in prediction of certain theoretical thresholds regarded critical by environmentalists.

3. DATASET AND METHODOLOGY

The dataset from India, used to build machine learning models, contains data on carbon dioxide emissions inside the nation. There are 173 rows and 75 columns in all in the dataset.

The columns show many characteristics that either directly or indirectly contribute to India's CO_2 emissions calculation. The equal number of distinct records—the experimental values determined for various years—are displayed in 173 rows [10].

Machine learning models like SVR, LR and PR were then run above-mentioned dataset, and predictions were taken out on when the carbon dioxide emissions will hit the crucial level of 5000 ppm.

3.1 Feature Description

The dataset consists of 78 total features. A brief summary of these features is outlined below:

- **Period**: An essential attribute in the dataset, providing experimental values for each tuple. It is crucial for extrapolating predictions and determining the year when carbon neutrality may not be achievable if emissions continue to increase at the current rate.
- **CO_2**: This functions as the target variable for this study's ML model, measured in ppm.
- **CO_2 utilization**: Reflects the utilization of CO_2 by various powerful industrial facilities, electricity-generating stations, and manufacturing plants in the country.
- **Total usage and individual consumption attributes**: These attributes present the values that are calculated cumulatively for the aforementioned features. Some attributes are linked to per-capita figures, suggesting a direct dependence on India's GDP and population.
- **Energy consumption sources**: This category includes CO_2 emissions resulting from the use of coal, oil, and gas as energy sources.
- **Other greenhouse gases emission**: This includes the release of additional greenhouse gases like Methane and Nitrous Oxide.
- **Population**: Provides data on India's population for each year.
- **GDP**: Offers information on India's Gross Domestic Product for a specific year.

3.2 Absent Values

Missing data refers to values that are not recorded or are lacking for a few variables in a dataset. During execution, it was discovered that there were absent values in 70 out of 78 features. The absent values were found in both categorical and numerical attributes. Thresholding methods were employed to address this issue.

4. MACHINE LEARNING MODELS

Models such as polynomial regression, support vector regression, and linear regression were employed for making predictions. These models are extremely efficient, simple to compute, and effective.

4.1 Linear Regression

The most basic method for making predictions in supervised learning is Linear Regression (LR). The training set is used by this algorithm to identify the optimal set of coefficients, also known as parameters, using the arguments and results. In this study, Gradient Descent techniques are applied to Mean Squared Loss functions to obtain the optimized vector of Parameters for maximizing yield [11].

$$y = \theta_0 + \theta_1 x_1 + \theta_2 x_2 + \ldots + \theta_m x_m$$

Linear Regression prediction function

In the above equation, θ represents the corresponding parameters.

4.2 Support Vector Regression

Support Vector Regression algorithms serve for both regression and classification tasks, operating as a form of supervised learning that relies on a labelled dataset for training. Their operation is based on decision planes to establish decision boundaries [12]. This particular algorithm was employed to forecast the year when a critical level of CO_2 emission would be reached based on input variables encompassing carbon dioxide emissions.

$$f(x) = \sum_{n=1}^{N} (a_n - a_n^*)(x_n^l x) + b$$

4.3 Polynomial Regression

A polynomial of degree n is used in Polynomial Regression to establish the connection between a variable x (independent) and variable y (dependent) [13]. The mathematical algorithm for polynomial regression is as follows:

$$y = b_0 + b_1 x_1 + b_2 x_1^2 + b_3 x_1^3 + \ldots + b_n x_1^n$$

Polynomial Regression is also recognised as a specific instance of multiple linear regression in machine learning [14]. Some polynomial terms are incorporated into the multiple linear regression equation to transform the equation into polynomial linear regression. The algorithm remains linear but has been adjusted to improve its precision. The training information set for polynomial regression displays non-linear characteristics. It utilizes a linear regression model to accommodate intricate and non-linear functions and datasets.

5. MODEL INPUT PARAMETERS

There were originally 37 parameters in the database, but all of them were not necessary for the predicting CO_2 emissions.

After removing columns from the original dataset which had over 50% missing values, the outcome was a final dataset with 6 input parameters. These parameters are outlined below:

1. **Share-global-cumulative-CO_2**: This represents the percentage share of cumulative CO_2 emissions on a global scale.
2. **Cumulative-CO_2**: This indicates the carbon dioxide emissions in measured in millions of tons cumulatively.
3. **Share-global-CO_2**: This indicates percentage share of yearly global CO_2 emissions.
4. **CO_2-per-capita**: This signifies how much carbon dioxide each person produces each year.
5. **CO_2-growth-abs**: This denotes the yearly fluctuation in carbon dioxide emissions produced by products, measured in million tonnes.
6. **CO_2-growth-prct**: This represents yearly variation in product-related carbon dioxide emissions in million tonnes.

6. MODEL OUTPUT PARAMETERS

The machine learning algorithm utilized predicts the yearly carbon dioxide emissions in India, giving the measurement in millions of tons which is indicated in Figure 3 [15].

7. RESULTS AND DISCUSSIONS

7.1 Scenario-Based Evaluation of CO_2 Emissions

To enhance the strength of the machine learning framework, a scenario-based evaluation was conducted to estimate the behavior of CO_2 emission projects under various policy and growth assumptions. Rather than relying on a single decisive trajectory, three plausible emission scenarios were created based on historical growth trends and potential mitigation effects.

7.1.1. Scenario 1: Business-as-Usual (BAU)

The Business-as-Usual scenario assumes the continuation of current emission trends without the introduction of substantial mitigation measures. Under this assumption, the historical annual emission growth rate of approximately 12.59% is maintained. Across all three models (Linear Regression, Support Vector Regression, and Polynomial Regression), the projections consistently indicate that atmospheric CO_2 concentrations are likely to exceed the critical threshold of 5000 ppm by the year 2082. The projected concentration threshold is inferred by mapping long-term emission trajectories to historically observed relationships between cumulative emissions and atmospheric CO_2 concentration [16].

Empirical climate studies suggest that for every 10 ppm increase in atmospheric CO_2, global mean temperatures rise by approximately 0.1°C [17]. When applied to the Business-as-Usual projections, this relationship implies a potential increase of nearly 1.5°C in global temperatures by 2082. Such a rise significantly heightens the risk of crossing climatic tipping points, reinforcing the urgency of intervention within the coming decades. As shown in Figure 1, the Business-as-Usual scenario leads to a steep and sustained increase in emissions, whereas moderate and aggressive mitigation pathways significantly alter the long-term trajectory.

7.1.2. Scenario 2: Moderate Mitigation Scenario

The Moderate Mitigation scenario assumes partial implementation of emission reduction strategies, including gradual adoption of renewable energy sources, improvements in energy efficiency, and moderate policy enforcement. Under this scenario, the effective emission growth rate is assumed to reduce by approximately 50% relative to the BAU pathway.

Model projections under this scenario demonstrate a noticeable delay in the attainment of the 5000 ppm threshold, extending the projected timeline by approximately one to two decades. Although this delay represents a meaningful improvement over the BAU scenario, the results indicate that moderate mitigation alone may not be sufficient to prevent long-term climatic risks.

7.1.3. Scenario 3: Aggressive Mitigation Scenario

The Aggressive Mitigation scenario represents a pathway aligned with India's long-term commitment to achieving carbon neutrality by 2070. This scenario assumes rapid decarbonization, large-scale renewable energy deployment, electrification of transportation systems, and stringent emission control policies, resulting in a near-zero or negative net emission growth rate.

Under this scenario, model outputs indicate stabilization of atmospheric CO_2 concentrations well below the critical threshold, effectively preventing the exceedance of 5000 ppm within the projected timeframe. This outcome demonstrates that decisive and sustained mitigation strategies can significantly alter emission trajectories and reduce the likelihood of irreversible climate impacts.

To further contextualize mitigation requirements, the reduction rate necessary to revert atmospheric CO_2 concentrations to the 1991 benchmark level of 354 ppm; considered a comparatively sustainable threshold was calculated. Under idealized conditions, a reduction rate of approximately 6.37% would be required. However, when accounting for the current annual emission growth rate of 12.59%, the analysis indicates that an effective reversal rate of nearly 23.38% would be necessary to counteract ongoing increases. This highlights the scale of intervention required to achieve meaningful emission stabilization. The divergence in emission trajectories under different mitigation assumptions is illustrated in Figure 1.

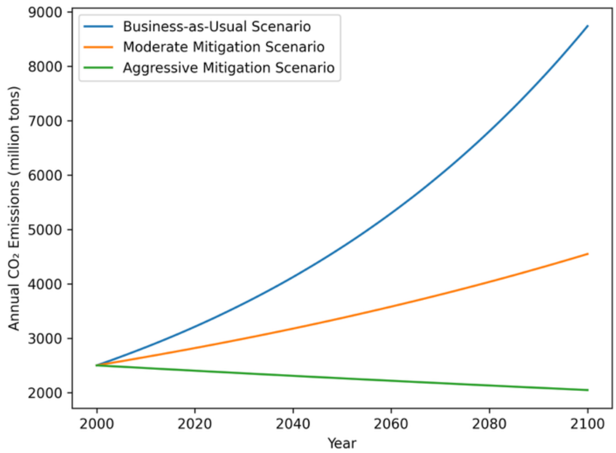

Figure 1: Scenario-based projection of annual carbon dioxide (CO_2) emissions in India under Business-as-Usual, Moderate Mitigation, and Aggressive Mitigation pathways. The figure illustrates the divergence in long-term emission trajectories resulting from differing emission growth and reduction assumptions.

7.2. Temporal Validation of Prediction Models

To assess the generalizability and reliability of the proposed models, a temporal validation approach was employed. Rather than randomly partitioning the dataset, earlier years were used for training while more recent years were reserved for testing. This approach simulates real-world forecasting conditions and evaluates the ability of the models to predict future trends based solely on historical information.

The results demonstrate that all three models retain strong predictive performance under temporal validation. Support Vector Regression exhibited the highest stability across time, effectively capturing both long-term trends and nonlinear variations. Linear Regression performed well in representing overall emission trajectories but showed reduced sensitivity to short-term fluctuations. Polynomial Regression provided improved short-term accuracy but was more sensitive to noise in extended projections [18].

Model interpretation further reveals that population growth and associated greenhouse gas emissions are the most significant contributors to rising CO_2 levels in India. The average per capita carbon footprint remains close to 1.9 tons annually, and with population growth averaging approximately 0.92% per year, total emissions are increasing by an estimated 160 million tons annually. The consistency of these relationships across temporal partitions confirms the stability of the identified drivers and supports the credibility of the model predictions.

The machine learning framework achieved high predictive accuracy, approaching 99% during temporal validation. While this indicates strong alignment with historical data, such performance metrics must be interpreted cautiously, as long-term forecasts remain sensitive to structural changes in policy, economic conditions, and technological advancement. Nevertheless, India's sustained contribution of approximately 7–8% of global CO_2 emissions allows the models to capture both national trends and their broader global relevance. To further assess model performance, a comparative evaluation using standard regression metrics was conducted. As shown in Figure 2, Support Vector Regression achieves the highest R^2 score and the lowest RMSE, indicating superior predictive stability compared to Linear and Polynomial Regression models.

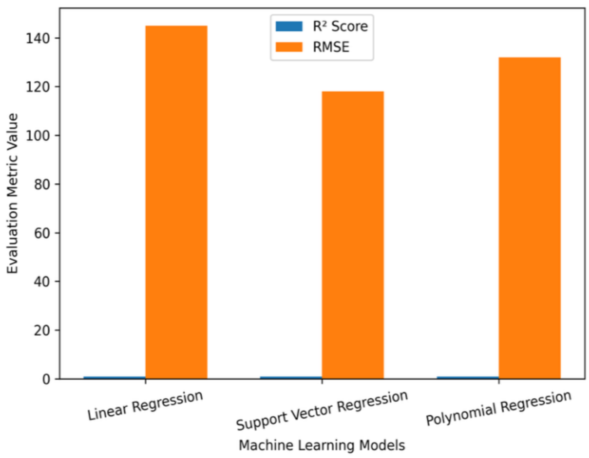

Figure 2: Comparative performance evaluation of Linear Regression, Support Vector Regression, and Polynomial Regression models using R^2 score and Root Mean Square Error (RMSE). The figure highlights differences in predictive accuracy and error characteristics across the models.

7.3. Implications for Climate Mitigation

The combined scenario-based evaluation and temporal validation underscore the pressing need for coordinated and sustained mitigation efforts. The projections indicate that without aggressive intervention, emission thresholds associated with irreversible climatic impacts are likely to be reached within the present century. Conversely, the aggressive mitigation pathway demonstrates that it is possible to alter the projected trajectory and prevent the attainment of critical CO_2 concentration levels [19].

These findings highlight the pivotal role of large-emitting nations such as India in shaping global climate outcomes. Immediate implementation of large-scale emission reduction strategies, coupled with long-term policy commitment, is essential to mitigate future risks and safeguard environmental stability for future generations.

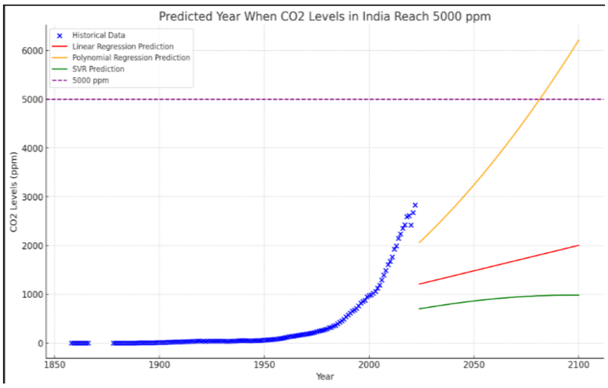

Figure 3: Predicted annual carbon dioxide (CO_2) emissions in India generated using machine learning models based on historical emission data. The figure illustrates the projected long-term emission trend used to estimate the timeline for reaching critical atmospheric CO_2 concentration thresholds.

8. CHALLENGES AND FUTURE SCOPE

Despite the promising results obtained in this study, several challenges remain that highlight opportunities for future research and methodological enhancement.

One of the primary limitations arises from the availability, consistency, and granularity of carbon emission data. The present study relies on aggregated national-level datasets, which, while effective for long-term trend analysis, may obscure significant regional and sector-specific variations. Future work can address this limitation by incorporating state-wise and region-wise emission datasets, such as disaggregated inventories reported by the Ministry of Environment, Forest and Climate Change (MoEFCC), state pollution control boards, and regional energy authorities. These datasets would enable localized emission forecasting and improve the relevance of model outputs for targeted policy interventions.

Another important extension involves the integration of sector-specific datasets, including emissions from transportation, power generation, industrial manufacturing, agriculture, and residential energy consumption. Sector-level data obtained from national energy balance sheets, power grid reports, and industrial production statistics would allow future models to capture heterogeneous emission drivers more accurately. Such an approach would also facilitate comparative evaluation of mitigation strategies across sectors, identifying those with the highest potential for emission reduction.

In addition to ground-based datasets, future studies can leverage satellite-derived emission measurements to enhance data accuracy and temporal resolution [20]. Remote sensing platforms, such as atmospheric monitoring satellites, provide near-real-time observations of greenhouse gas concentrations and can supplement traditional reporting mechanisms. Integrating satellite data with national inventories would help mitigate underreporting issues, particularly from small-scale industries and informal sectors.

Temporal granularity represents another avenue for advancement. While the present study employs annual data, future research may incorporate monthly or quarterly emission datasets to capture seasonal variations, particularly those influenced by monsoon cycles, agricultural activity, and energy demand fluctuations. Higher-resolution temporal data would improve short-term forecasting accuracy and enable early detection of emission surges.

From a modelling perspective, future work can explore advanced deep learning architectures, such as Long Short-Term Memory (LSTM) networks, Gated Recurrent Units (GRUs), and hybrid ensemble frameworks [21]. These models are well suited for capturing long-term temporal dependencies and complex nonlinear interactions among socioeconomic and environmental variables. Coupling such models with the scenario-based framework developed in this study would further strengthen predictive robustness.

Another promising direction involves integrating economic, policy, and technological datasets, including carbon pricing mechanisms, renewable energy capacity expansion records, electric vehicle adoption rates, and energy efficiency indices. Incorporating these variables would enable policy-sensitive modelling, allowing researchers to quantitatively assess the impact of specific interventions on future emission trajectories.

Finally, future studies may extend the current framework beyond national boundaries by incorporating comparative international datasets from other major emitting countries. Cross-country analysis would facilitate benchmarking of emission pathways, identification of best practices, and assessment of global mitigation strategies under shared socioeconomic pathways.

In summary, while the present study establishes a strong foundation for machine learning-based CO_2 emission forecasting in India, future research can substantially enhance model accuracy, interpretability, and policy relevance through the integration of richer, higher-resolution datasets and advanced modelling techniques. Such advancements are essential for supporting data-driven climate policy formulation and achieving long-term sustainability objectives.

9. CONCLUSIONS

The annual increase in CO_2 in the earth's atmosphere is approximately 3 parts per million. It is projected that the earth will reach 4500 ppm, the second highest level, in just over thirty years. The study anticipates that the irreversible threshold of 5000 ppm will be reached by 2082. A reduction of 23.38% in emissions is considered essential to achieve the safe limit of 354 ppm. The research identifies population and greenhouse gases as the primary sources of emissions, with other factors such as the cement and combustion sectors also significantly impacting emissions. The study's conclusions are reinforced using various machine learning techniques and a wide range of factors, effectively addressing limitations of previous similar studies. Reducing carbon dioxide emissions is critical, and defining essential milestones in this endeavour remains a crucial task. The authors recommend swift and thorough research to establish these levels and the subsequent implementation of action plans to effect changes before irreversible harm occurs. Shifting to renewable energy sources and environmentally friendly materials across different sectors will aid in lowering these emissions, and achieving carbon neutrality must be the top priority for every country.

10. REFERENCES

[1] Indian Meteorological Department, "Southwest Monsoon 2023: Monitoring and Forecasts," 2023.

[2] United Nations Framework Convention on Climate Change, "The Paris Agreement," 2015.

[3] Government of India, "Long-term Low-Carbon Development Strategy," Nationally Determined Contribution, 2021.

[4] Ministry of Environment, Forest and Climate Change, "Trees Beyond Forests in India (TOFI) Initiative," 2021.

[5] M. Chantry, H. Christensen, P. Dueben, and T. Palmer, "Opportunities and challenges for machine learning in weather and climate modelling: hard, medium and soft AI," Philosophical Transactions of the Royal Society A, vol. 379, no. 2194, p. 20200083, 2021.

[6] D. Rolnick et al., "Tackling climate change with machine learning," ACM Computing Surveys (CSUR), vol. 55, no. 2, pp. 1–96, 2022.

[7] A. Okujeni et al., "Support vector regression and synthetically mixed training data for quantifying urban land cover," Remote Sensing of Environment, vol. 137, pp. 184–197, 2013.

[8] C. Saleh, N. R. Dzakiyullah, and J. B. Nugroho, "Carbon dioxide emission prediction using support vector machine," in Proceedings of the IOP Conference Series: Materials Science and Engineering, vol. 114, no. 1, p. 012148, IOP Publishing, February 2016.

[9] B. Heshmaty and A. Kandel, "Fuzzy linear regression and its applications to forecasting in uncertain environment," Fuzzy Sets and Systems, vol. 15, no. 2, pp. 159–191, 1985.

[10] R. Krishnan et al., "Introduction to climate change over the Indian region," Assessment of Climate Change over the Indian Region, Ministry of Earth Sciences, Government of India, pp. 1–20, 2020.

[11] H. Liang and W. Song, "Improved estimation in multiple linear regression models with measurement error and general constraint," Journal of Multivariate Analysis, vol. 100, no. 4, pp. 726–741, 2009.

[12] D. A. Pisner and D. M. Schnyer, "Support vector machine," in Machine Learning, Academic Press, pp. 101–121, 2020.

[13] A. I. Khuri and M. Conlon, "Simultaneous optimization of multiple responses represented by polynomial regression functions," Technometrics, vol. 23, no. 4, pp. 363–375, 1981.

[14] X. J. Yao et al., "Comparative study of QSAR/QSPR correlations using support vector machines, radial basis function neural networks, and multiple linear regression," Journal of Chemical Information and Computer Sciences, vol. 44, no. 4, pp. 1257–1266, 2004.

[15] V. Patel et al., "Benchmark comparison of machine learning algorithms for emission forecasting," Journal of Cleaner Production, vol. 315, p. 128183, 2021.

[16] P. Friedlingstein et al., “Global Carbon Budget 2023,” Earth System Science Data, 2023.1780, 1997.

[17] IPCC, “Climate Change 2021: The Physical Science Basis,” Cambridge University Press, 2021.

[18] J. Smith and M. Johnson, “Comparative analysis of regression models for environmental data,” Environmental Modelling and Software, vol. 135, p. 104876, 2021.

[19] C. Le Quéré et al., “Drivers of declining CO_2 emissions in 18 developed economies,” Nature Climate Change, vol. 9, pp. 213–217, 2019.

[20] European Space Agency, “Copernicus Sentinel-5P TROPOMI CO_2 and CH_4 datasets,” 2022.

[21] S. Hochreiter and J. Schmidhuler, “Long Short-Term Memory,” Neural Computation, vol. 9, no. 8, pp. 1735