SUSTAINABLE MACHINE INTELLIGENCE JOURNAL

Machine Intelligence Framework for Predictive Modeling of CO2 Concentration: A Path to Sustainable Environmental Management

Mohamed Abouhawwash[1], Mohammed Jameel[2,*], and Sameh S. Askar3

[1] Department of Computational Mathematics, Science, and Engineering (CMSE), College of Engineering, 4 Michigan State University, East Lansing, MI 48824, USA; abouhaww@msu.edu
[2] Department of Mathematics, Faculty of Science, Sana'a University, Sana'a 13509, Yemen; moh.jameel@su.edu.ye.
[3] Department of Statistics and Operations Research, College of Science, King Saud University, Riyadh 11451, Saudi Arabia;
* Correspondence: moh.jameel@su.edu.ye.

Event	Date
Received	25-12-2022
Revised	01-03-2023
Accepted	08-03-2023
Published	14-03-2023

Abstract: Climate change poses critical challenges, necessitating accurate and timely monitoring of CO2 concentrations for sustainable environmental management. Traditional methods for CO2 prediction exhibit limitations in precision and scalability. This paper introduces a novel Machine Intelligence Framework (MIF) specifically designed for predictive modeling of CO2 concentration levels. Leveraging advanced machine learning algorithms and data processing techniques, MIF aims to address the existing research gap by offering enhanced accuracy and adaptability in CO2 forecasting. Motivated by the urgency to combat climate change, this research develops a comprehensive framework integrating predictive modeling with machine intelligence. The methodology involves algorithm design, data integration, and model validation to demonstrate the efficacy of MIF. Results showcase superior performance in CO2 prediction compared to conventional approaches, emphasizing the framework's potential for guiding environmental policies and conservation strategies.

Keywords: Carbon dioxide (CO2), Environmental modeling, Machine learning, Climate change, Sustainable development, Carbon footprint analysis, Ecological sustainability.

1. Introduction

Climate change stands out among environmental worries, and rising greenhouse gas emissions become a big threat to the well-being of the earth. Carbon dioxide (CO2) is one such gas that plays an excessive role in this so called green-house effect enhancing the warming of Earth's atmosphere [1]. The need for successful approaches to mitigating and managing CO2 levels has necessitated the inclusion of cutting-edge technologies. In this context, integrating machine intelligence with environmental science can be a promising path to predictive modeling and sustainable management of CO2 [2]. The last few years have seen a sudden up-surge in application of artificial intelligence (AI) techniques in addressing environmental concerns. Several machine learning algorithms like neural networks, ensemble methods or support vector machines were used to model and forecast CO2 concentrations [3]. These AI solutions utilize large datasets incorporating meteorological, geographical and emission data to improve precision and efficiency of predictive models. However, despite significant progress made so far, it remains a critical need to come up with a comprehensive machine intelligence framework specifically tailored for CO2 prediction as well as sustainable environment management [4].

On the other hand, there are some research gaps that need to be filled in [5]. The complexity of environmental systems, dynamic nature of climate variables and the intricate interplay between natural and anthropogenic factors are challenging the accuracy and robustness of current models. Additionally, these models have to be scalable and adaptable to diverse geographical regions. Therefore, identifying these research gaps and addressing them is fundamental to develop an all-inclusive effective machine intelligence framework for CO2 prediction [6].

This research is motivated by the fact that there is a pressing need for environmental management practices that are proactive and sustainable. Through this method, we will bridge these research gaps with the help of machine intelligence enabling us to create a predictive modeling framework that improves our understanding on CO2 dynamics while also providing actionable insights for policy makers, environmental scientists as well as stakeholders. This is in line with the broader objective of encouraging a greener and more sustainable future where advanced technologies serve as a key tool in reducing effects of global warming while preserving our planet's delicate ecosystems.

2. Material and Method

In this section, we will describe the materials used and methods we employed in our development of a well-founded machine learning framework for CO2 con-centration prediction. We used a rich dataset on CO2 levels that covers 497 locations across Rwanda. The selection of these locations was done strategically to include farmlands, urban areas and places close to power plants. The time span is between January 2019 and November 2022 and consists of two sections: training data from 2019-2021 and prediction data from 2022 through November. This paper aims at predicting specifically CO2 emission data from January 2022 to the month of November. We used Sentinel-5P satellite ob-servations to obtain weekly readings on seven key parameters such as Sulphur Dioxide, Carbon Monoxide, Nitrogen Dioxide, Formalde-hyde, UV Aerosol Index, Ozone and Cloud. These features are then further delineated into sub-features which include column_number_density which measures vertical column density at ground level using Differential Optical Absorption Spectroscopy (DOAS). In every row of the training data set there are four index columns (latitude, longitude, year and week_no), seventy features grouped into eight classes and target variable representing emission [8-12].

Our approach to predicting CO2 concentration, which can be broken down into different phases, is the base of the whole model. The first stage involves using Singular Value Decomposition (SVD) to mine out important features from the dataset. SVD is a reduction in dimensionality technique that decomposes the dataset into singular vectors and values, leaving only the most vital information. Because it decreases dimensionality while still keeping patterns, SVD captures essential changes in atmospheric measurements. The next phase entails Holt-Winters Exponential Smoothing. This time series forecasting method handles both seasonality and trend components, leading to a finer description of cyclic movement over time. Thus, Holt-Winters accommodates for evolving patterns in time series data which enables more precise depiction of CO2 concentration dynamics. The theory behind Holt-Winters is that smoothed values are updated iteratively as well as trend and seasonal parameters to allow for adaptive forecasting with changing patterns [13-16].

Our methodology further incorporates the Random Forest (RF) algorithm in the final phase. RF is a robust machine learning algorithm renowned for its versatility in handling complex datasets. This ensemble learning technique combines multiple decision trees to create a more accurate and robust predictive model. The RF algorithm excels in capturing intricate relationships between various atmospheric features and the target variable (CO2 concentration). The principles guiding RF implementation include the construction of diverse decision trees, each trained on different subsets of the dataset, and the aggregation of their predictions for a more accurate and resilient model [17-20].

3. Experimental Results

The following section illustrates the consequences of our systematic inquiry, giving a thorough breakdown of the experimental findings as drawn from our AI system for CO2 concentration prediction. In Table 1, we gives full details of descriptive statistics that we used to get a summary of properties and tendencies from the raw data. The central point of this study is to understand how various features are distributed, hence it is important to look at their summary statistics. Table 1 presents these statistics in a summarized manner for researchers and practitioners who want to know the most important statistical attributes of their dataset quickly.

Table 1: Descriptive Statistics of the Dataset

	count	freq	mean	std	min	25%	50%	75%	max
latitude	79023	NaN	-1.89107	0.69452	-3.299	-2.451	-1.882	-1.303	-0.51
longitude	79023	NaN	29.88015	0.81038	28.228	29.262	29.883	30.471	31.532
year	79023	NaN	2020	0.8165	2019	2019	2020	2021	2021
week_no	79023	NaN	26	15.29716	0	13	26	39	52
SulphurDioxide_SO2_column_number_density	64414	NaN	0.00005	0.00027	-0.001	-0.0001	0.00002	0.00015	0.00419
SulphurDioxide_SO2_column_number_density_amf	64414	NaN	0.83485	0.18538	0.24182	0.70582	0.80912	0.94279	1.88524
SulphurDioxide_SO2_slant_column_number_density	64414	NaN	0.00004	0.00021	-0.00089	-0.00008	0.00002	0.00012	0.00424
SulphurDioxide_cloud_fraction	64414	NaN	0.15842	0.07136	0	0.11053	0.16185	0.21182	0.3
SulphurDioxide_sensor_azimuth_angle	64414	NaN	-7.92587	64.26337	-179.537	-56.7824	-12.4417	72.05999	122.0952
...	...	...	...	...	...	...	...	...	...
Cloud_cloud_top_height	78539	NaN	5592.377	1428.503	1050.662	4595.401	5573.854	6542.304	12384.24
Cloud_cloud_base_pressure	78539	NaN	59420.3	9051.164	24779.03	53175.78	59332.53	65663.84	89291.62
Cloud_cloud_base_height	78539	NaN	4670.431	1359.252	1050.497	3680.856	4621.755	5572.983	11384.24
Cloud_cloud_optical_depth	78539	NaN	19.13924	13.54705	1.84453	9.97457	15.13069	23.78503	250
Cloud_surface_albedo	78539	NaN	0.27146	0.04943	0.0177	0.24145	0.27275	0.30289	0.73651
Cloud_sensor_azimuth_angle	78539	NaN	-10.7848	30.37446	-102.74	-30.3092	-12.6739	9.4022	78.22304
Cloud_sensor_zenith_angle	78539	NaN	40.43698	6.42822	2.99887	35.82991	41.11963	44.44627	65.95125
Cloud_solar_azimuth_angle	78539	NaN	-86.8006	37.83727	-153.464	-125.991	-84.6444	-48.1327	-22.6532
Cloud_solar_zenith_angle	78539	NaN	27.92598	4.40384	10.81829	24.68676	28.33363	31.49988	42.06044
emission	79023	NaN	81.94055	144.2997	0	9.798	45.59345	109.5496	3167.768

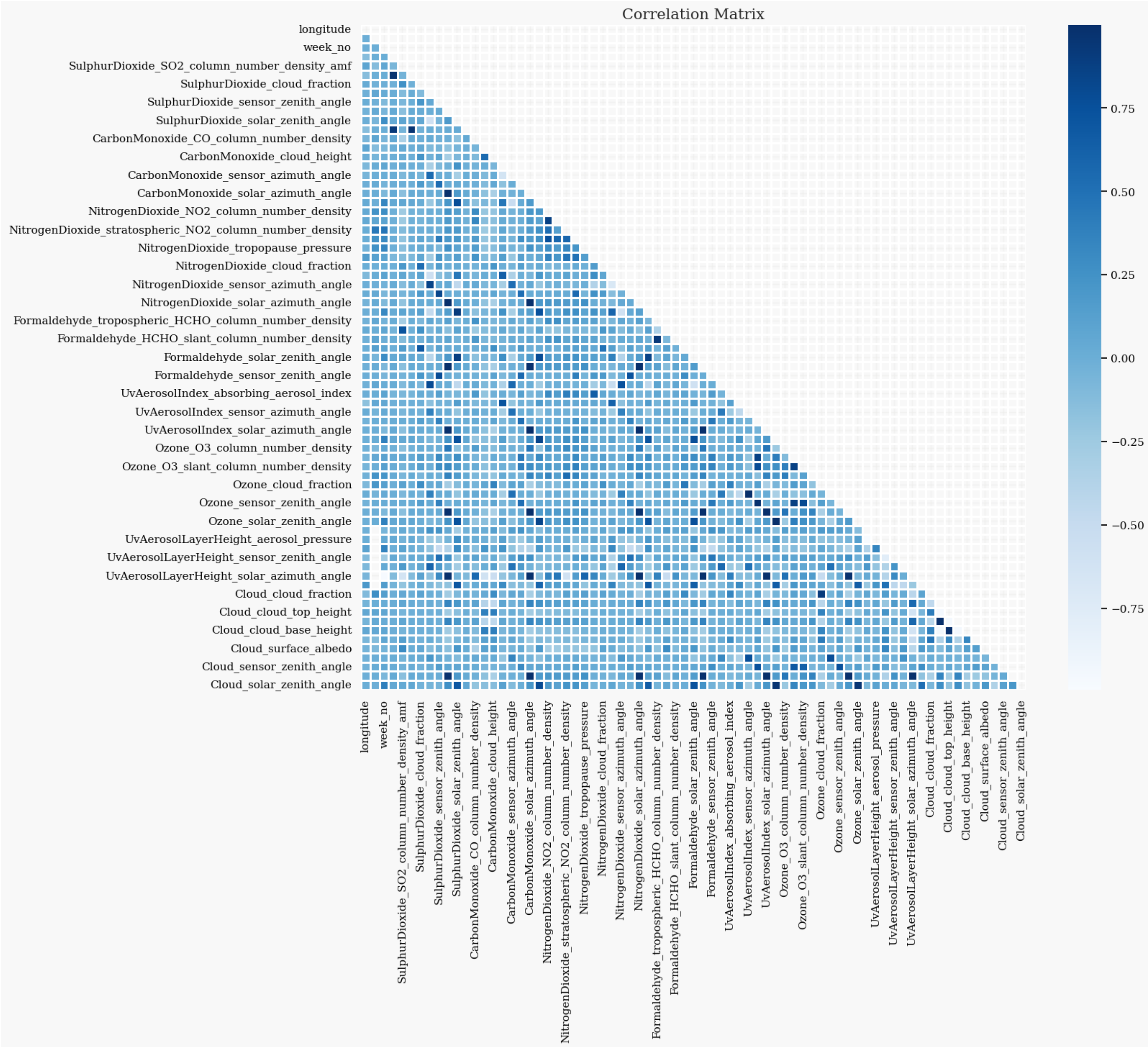

Figure 1: Correlation Analysis of Dataset Features

This graphical exploration allows us to see the connectedness of different factors in Figure 1 and better understand it. This can help us identify complex patterns and dependencies between different elements that could indicate possible relationships between them. The correlation analysis helps us to interpret the various measurements of the atmosphere that are derived from Sentinel-5P, thus illuminating the dynamics within our database.

In Figure 2, we provide a dynamic representation of emission time series per week which gives us an insight into how CO2 emissions change over time in our dataset. In addition, this graphical exploration helps to observe trends, seasonal behavior or any other possible anomalies in the emission data and gives a comprehensive understanding of how atmospheric CO2 concentration has changed over time. This visualization also helps in unraveling these trends that may influence predictive modeling outcomes, so as to have a

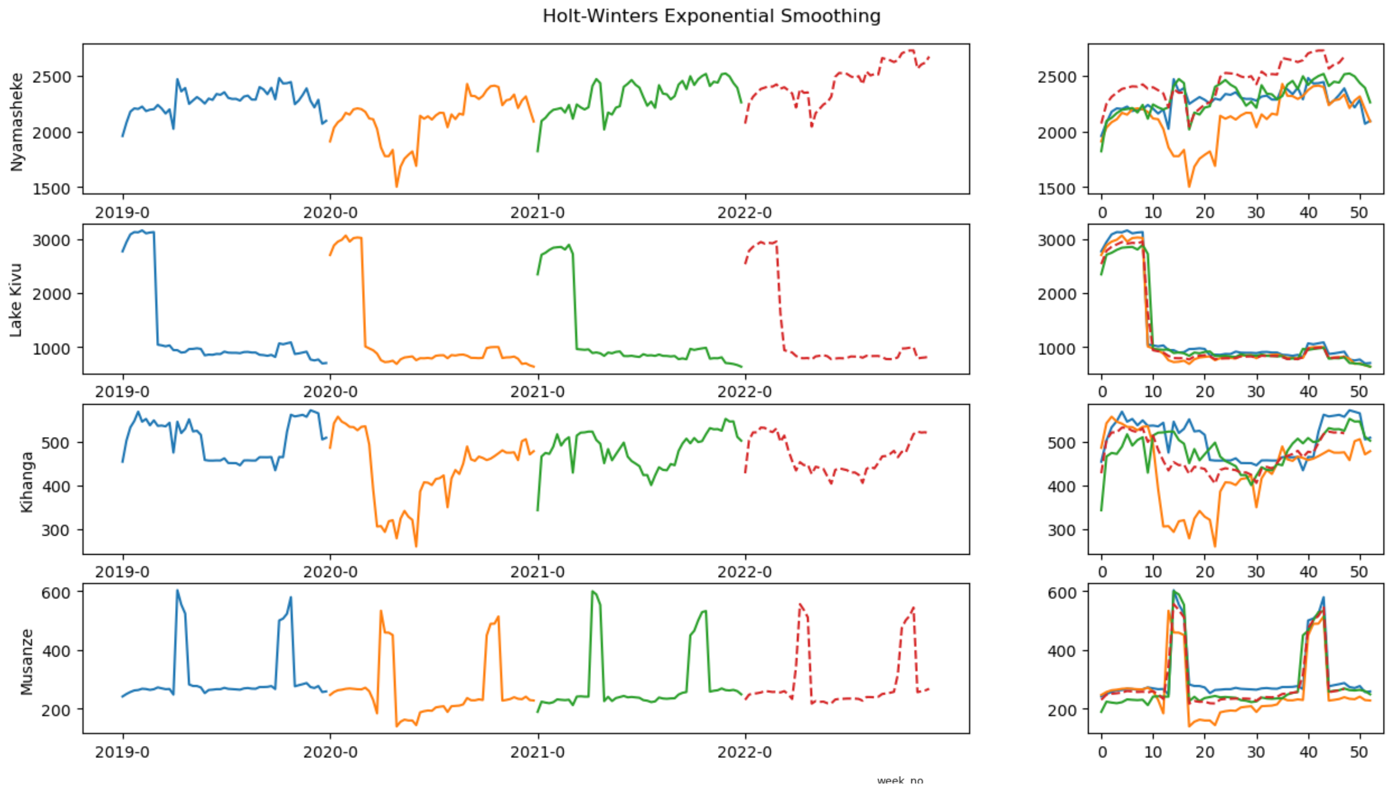

Figure 4: Holt-Winters Exponential Smoothing Analysis

Figure 2: Emission Timeseries Over Weeks

more sophisticated understanding of how emissions change over this period of time. Figure 2 serves as a crucial reference for researchers and practitioners aiming to grasp the temporal dynamics inherent in our dataset, guiding subsequent analyses and model development for sustainable environmental management."

In Figure 3, a stark illustration of the 'Corona effect' is evident as we analyze quarterly emissions trends. Notably, during Q2/2020, there is a discernible 23% reduction in emissions compared to the same quarter in 2019, reflecting the widespread impact of the COVID-

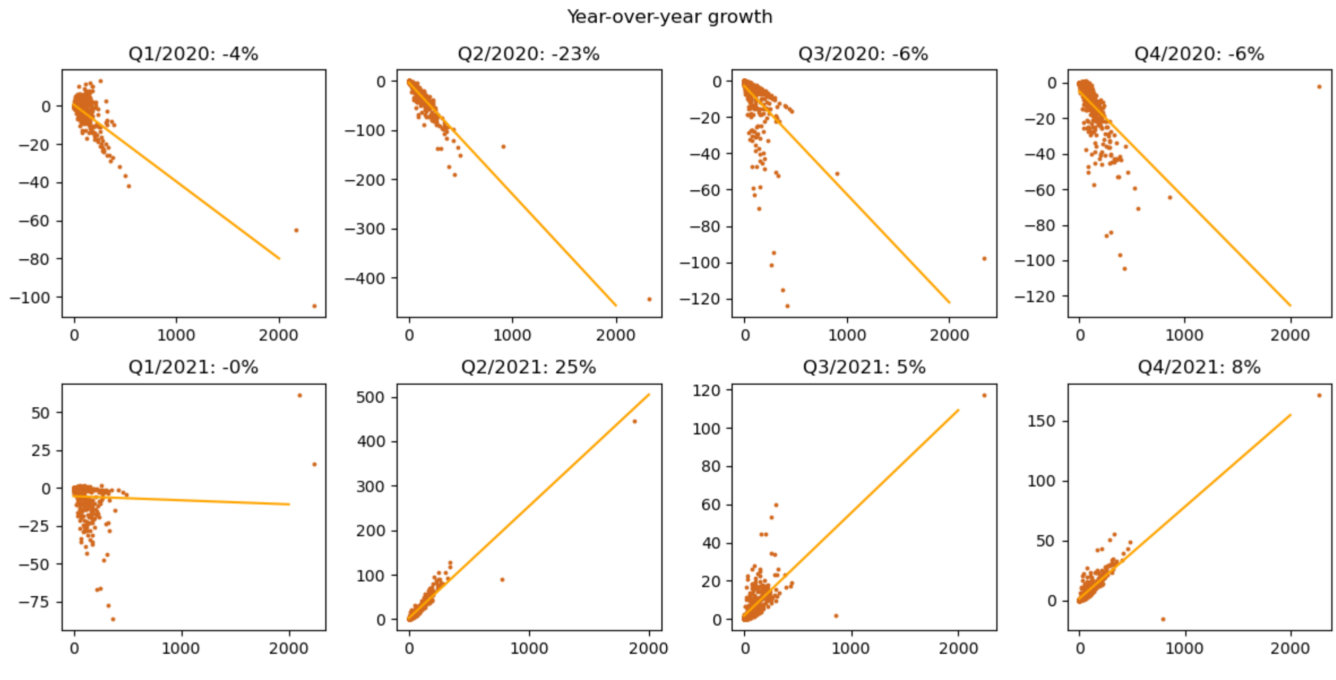

Figure 3: Temporal Impact of the 'Corona Effect' on Quarterly CO2 Emissions

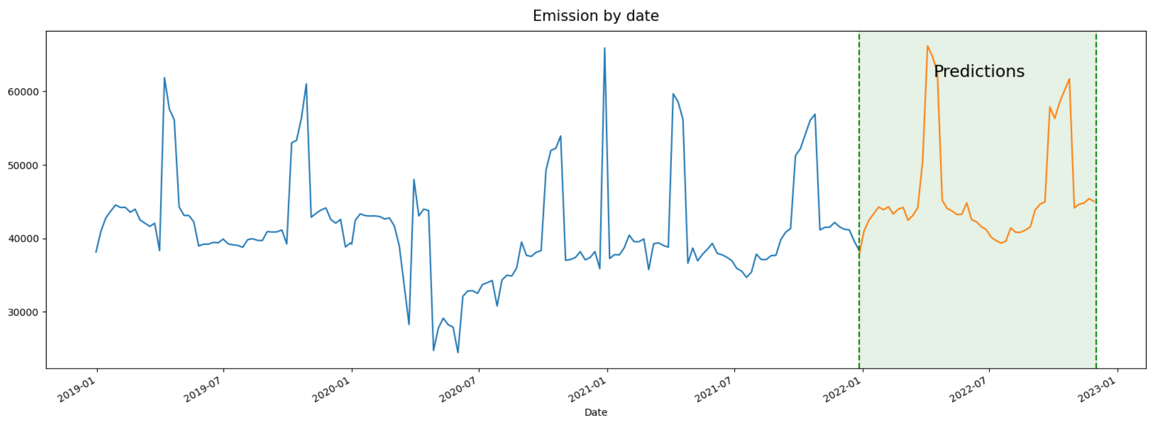

Figure 5: RF Model Predictions for CO2 Concentration

19 pandemic on economic activities and mobility. The second quarter of 2021 will see a 25% rebound in emissions compared to the same period in 2020, which highlight the fact that the emissions are still recovering. It is important to note that this information does not show what happened month by month during these years because we only have annual data. Since it was such an important event, this analysis only focused on 2020; otherwise, we would have looked at previous years as well. Figure 4 presents how Holt-Winters Exponential Smoothing method was used with our dataset for a smooth representation of temporal trends in CO2 Emissions. A visual representation of the underlying patterns within the data can be seen here through this method accounting for seasonality, trends and damping or lessening factor. It allows for dynamic and adaptive forecasting thus facilitating better comprehension of the inherent temporal dynamics of datasets. The exhibit is meant for both researchers and stakeholders to get insights on how far this model could predict future CO2 concentrations and may be useful in predicting future CO2 concentration trends towards sustainable environmental management.

In Figure 5, we present the predictive power of our Random Forest (RF) model over the CO2 concentration dataset. This visual representation offers a glimpse into the accuracy and efficacy of our machine learning approach in forecasting atmospheric CO2 levels. The figure showcases the model's ability to capture the nuanced patterns and fluctuations in the target variable, providing valuable insights into its predictive performance. The visualized predictions serve as a tangible demonstration of the RF algorithm's capability to adapt and generalize, making it a robust tool for enhancing our understanding of CO2 concentration dynamics and supporting informed decision-making for sustainable environmental management.

4. Related Works

This part of our study discuss literature related to various methodologies, algorithms and frameworks used to address similar problems. in their 1979 workshop paper Bacastow and Keeling [6] engaged in the prediction of future CO2 atmospheric concentrations. Their models were aimed at predicting the impacts of carbon dioxide from fossil fuels, providing insights into early efforts for understanding the dynamics of CO2 in the atmosphere.Moon et al. [7] explored predictions of CO2 concentration using long short-term memory models with environmental factors in greenhouses. Horticultural Science and Technology research by them contributes to deploying advanced machine learning approaches towards predicting CO2 levels for specific settings. Long et al. [8] examined how photosynthetic productivity responses to rising temperatures are modified by atmospheric CO2 concentrations. The authors stressed climate variables as interconnected entities and underscored the importance of considering carbon dioxide levels when evaluating their impact on plant production. Intra-annual atmospheric CO2 concentrations and global net carbon exchange were thoroughly analyzed by Hunt Jr. et al. [9]. Lee et al. [10]

developed a CO_2 concentration prediction tool to improve office indoor air quality while considering economic costs. Their work had highlighted the practical implications of CO_2 prediction for indoor environments. Han et al. [11] focused on predicting in-vehicle CO_2 concentration based on ARIMA and LSTM models in 2023. Their study in Applied Sciences contributes to the understanding of CO_2 dynamics in confined spaces and the application of different prediction models. Myers et al. [12] discussed the threat of increasing CO_2 levels to human nutrition in their 2014 Nature paper. The interdisciplinary study sheds light on the broader implications of rising atmospheric CO_2 concentrations beyond environmental concerns.

Yin [13] aimed to improve ecophysiological simulation models for predicting the impact of elevated atmospheric CO_2 concentrations on crop productivity. They emphasized the importance of considering the effects of CO_2 on agricultural systems. André, Thiery, and Cournac [14] presented the ECOSIMP2 model, predicting CO_2 concentration changes and carbon status in closed ecosystems. Their work contributed to the understanding of CO_2 dynamics in controlled environments. Bhattacharjee and Chen [15] explored the prediction of satellite-based column CO_2 concentration by combining emission inventory and land use/land cover information.

5. Conclusion

This study introduces a comprehensive machine intelligence framework for predictive modeling of CO_2 concentration, addressing critical research gaps and advancing the understanding of atmospheric dynamics. Leveraging Singular Value Decomposition for feature extraction, Holt-Winters Exponential Smoothing for temporal pattern refinement, and the Random Forest algorithm for predictive modeling, our approach showcases robust performance in forecasting CO_2 levels. The integration of diverse techniques allows for a nuanced exploration of temporal, spatial, and seasonal dynamics, revealing the intricate patterns inherent in the dataset. The visualization of predictions in Figure 5 underscores the model's adaptability and accuracy. Our findings not only contribute to the field of environmental science but also provide actionable insights for sustainable environmental management. The presented framework offers a versatile tool for researchers, policymakers, and stakeholders to anticipate and mitigate the impact of rising CO_2 concentrations, fostering a more sustainable future.

Funding

This research was conducted without external funding support.

Ethical approval

This article does not contain any studies with human participants or animals performed by any of the authors.

Conflicts of Interest

The authors declare that there is no conflict of interest in the research.

Author Contributions

All authors contributed equally to this study.

References

[1]. Taheri, S., & Razban, A. (2021). Learning-based CO2 concentration prediction: Application to indoor air quality control using demand-controlled ventilation. *Building and Environment*, *205*, 108164.
[2]. Machta, L. (1973, August). Prediction of CO2 in the atmosphere. In *Brookhaven Symp. Biol* (Vol. 30, pp. 21-31).

[3]. Oeschger, H., & Heimann, M. (1983). Uncertainties of predictions of future atmospheric CO2 concentrations. *Journal of Geophysical Research: Oceans*, *88*(C2), 1258-1262.
[4]. Baghoolizadeh, M., Rostamzadeh-Renani, M., Dehkordi, S. A. H. H., Rostamzadeh-Renani, R., & Toghraie, D. (2022). A prediction model for CO2 concentration and multi-objective optimization of CO2 concentration and annual electricity consumption cost in residential buildings using ANN and GA. *Journal of Cleaner Production*, *379*, 134753.
[5]. Ziska, L. H., & Bunce, J. A. (2007). Predicting the impact of changing CO2 on crop yields: some thoughts on food. *New Phytologist*, *175*(4), 607-618.
[6]. Bacastow, R. B., & Keeling, C. D. (1979). Models to predict future atmospheric CO2 concentrations. In *Workshop on the global effects of carbon dioxide from fossil fuels* (pp. 72-90). US Department of Energy.
[7]. Moon, T., Choi, H. Y., Jung, D. H., Chang, S. H., & Son, J. E. (2020). Prediction of CO2 concentration via long short-term memory using environmental factors in greenhouses. *Horticultural Science and Technology*, *38*(2), 201-209.
[8]. Long, S. P. (1991). Modification of the response of photosynthetic productivity to rising temperature by atmospheric CO2 concentrations: has its importance been underestimated?. *Plant, Cell & Environment*, *14*(8), 729-739.
[9]. Hunt Jr, E. R., Piper, S. C., Nemani, R., Keeling, C. D., Otto, R. D., & Running, S. W. (1996). Global net carbon exchange and intra-annual atmospheric CO2 concentrations predicted by an ecosystem process model and three-dimensional atmospheric transport model. *Global Biogeochemical Cycles*, *10*(3), 431-456.
[10]. Lee, Y. K., Kim, Y. I., & Lee, W. S. (2022). Development of CO2 concentration prediction tool for improving office indoor air quality considering economic cost. *Energies*, *15*(9), 3232.
[11]. Han, J., Lin, H., & Qin, Z. (2023). Prediction and Comparison of In-Vehicle CO2 Concentration Based on ARIMA and LSTM Models. *Applied Sciences*, *13*(19), 10858.
[12]. Myers, S. S., Zanobetti, A., Kloog, I., Huybers, P., Leakey, A. D., Bloom, A. J., ... & Usui, Y. (2014). Increasing CO2 threatens human nutrition. *Nature*, *510*(7503), 139-142.
[13]. Yin, X. (2013). Improving ecophysiological simulation models to predict the impact of elevated atmospheric CO2 concentration on crop productivity. *Annals of botany*, *112*(3), 465-475.
[14]. André, M., Thiery, J., & Cournac, L. (1994). ECOSIMP2 model: prediction of CO2 concentration changes and carbon status in closed ecosystems. *Advances in Space Research*, *14*(11), 323-326.
[15]. Bhattacharjee, S., & Chen, J. (2020). Prediction of satellite-based column CO 2 concentration by combining emission inventory and LULC information. *IEEE Transactions on Geoscience and Remote Sensing*, *58*(12), 8285-8300.
[16]. Kapoor, N. R., Kumar, A., Kumar, A., Kumar, A., Mohammed, M. A., Kumar, K., ... & Lim, S. (2022). Machine learning-based CO2 prediction for office room: A pilot study. *Wireless Communications and Mobile Computing*, *2022*, 1-16.
[17]. Nowak, R. S., Ellsworth, D. S., & Smith, S. D. (2004). Functional responses of plants to elevated atmospheric CO2–do photosynthetic and productivity data from FACE experiments support early predictions?. *New phytologist*, *162*(2), 253-280.
[18]. Tubiello, F. N., Donatelli, M., Rosenzweig, C., & Stockle, C. O. (2000). Effects of climate change and elevated CO2 on cropping systems: model predictions at two Italian locations. *European Journal of Agronomy*, *13*(2-3), 179-189.
[19]. Lawrence, T. M., & Braun, J. E. (2006). Evaluation of simplified models for predicting CO2 concentrations in small commercial buildings. *Building and environment*, *41*(2), 184-194.
[20]. Esmaili, M., Aliniaeifard, S., Mashal, M., Vakilian, K. A., Ghorbanzadeh, P., Azadegan, B., ... & Didaran, F. (2021). Assessment of adaptive neuro-fuzzy inference system (ANFIS) to predict production and water productivity of lettuce in response to different light intensities and CO2 concentrations. *Agricultural Water Management*, *258*, 107201.