

 مجلة ميدأوشن للبحوث والدراسات	Midocean Journal for Research and Studies https:// e-ISSN:	
---	--	--

AI-Based Nowcasting of National CO₂ Emissions Using Satellite-Derived Indicators: A Hybrid Framework for GCC Countries

Mohammad Wajeesh Bader – Dr. Amjad Qataish

Mohammad Wajeesh Bader

ORCID ID: 0009-0009-9682-7435

Artificial Intelligence – College of Computer Science and Information Technology – Midocean University – United Arab Emirates

Email: 251000462@st.midocean.edu.ae

Abstract

Objectives: This study aims to develop a hybrid framework combining satellite-derived environmental indicators with demographic data to estimate national CO₂ emissions in Gulf Cooperation Council (GCC) countries, and to quantify the accuracy-timeliness trade-off when using satellite data versus traditional economic statistics.

Study Location: The study covers the six GCC countries: Saudi Arabia, United Arab Emirates, Qatar, Kuwait, Oman, and Bahrain, spanning the period from 2000 to 2023.

Methodology: Three XGBoost regression models were developed and compared: an economic baseline model (Model A), a hybrid model combining satellite indicators with population density (Model B), and a satellite-only model (Model C). Satellite data from Google Earth Engine included MODIS land surface temperature (LST), MODIS NDVI vegetation index, DMSP/VIIRS night-time lights, MERRA-2 aerosol optical depth (AOD), and Sentinel-5P tropospheric NO₂. All data were aggregated to annual country-level values covering 144 country-year observations. The dataset was split temporally into training (2000-2018) and test (2019-2023) periods.

Results: The economic baseline model achieved $R^2 = 0.96$ on test data (2019-2023) with $MAPE < 10\%$, while the hybrid model achieved $R^2 = 0.69$ with $MAPE = 18\%$. The satellite-only model performed slightly lower at $R^2 = 0.67$ with $MAPE < 20\%$. Feature importance analysis revealed that summer and winter land surface temperatures collectively account for 56% of model importance, followed by night-time lights (22%) and population density (13%). This hierarchy is physically consistent with the GCC region where extreme heat drives cooling demand, which is a major contributor to electricity consumption and CO₂ emissions. Country-level analysis showed the highest performance for Saudi Arabia ($R^2 = 0.72$) and lowest for Bahrain ($R^2 = 0.61$).

Recommendations: The study recommends applying this framework as a complementary tool to official inventories for rapid anomaly detection and policy evaluation. Future work should extend geographic coverage to other regions with similar climate characteristics, transition to sub-national estimation at provincial or city levels, incorporate additional satellite products as longer series become available, and develop operational monthly or weekly estimation systems to maximize the timeliness advantage of satellite data. Ensemble or Bayesian approaches should be explored to provide uncertainty bounds alongside point estimates. Overall, the work offers one of the first systematically documented nowcasting frameworks for national CO₂ emissions in the GCC that explicitly benchmarks economic and satellite-based models under a common experimental design.

Keywords: CO₂ emissions, nowcasting, satellite remote sensing, XGBoost, GCC countries

1. Introduction

Climate change represents one of the most pressing challenges (IPCC, 2023) of the twenty-first century. IPCC assessments emphasize that global greenhouse gas (GHG) emissions must decline rapidly in the next decade to keep global warming within internationally agreed limits (IPCC, 2023). Achieving these objectives requires not only ambitious mitigation policies but also timely and reliable information about how national emissions evolve over time.

In practice, national CO₂ emission inventories are usually published with a delay of one to two years. Data for a given calendar year may only become available to policymakers and researchers in the following year or later. This reporting delay weakens the feedback loop between policy interventions and observed outcomes, because the impact of new measures cannot be evaluated until long after they have been implemented. Delayed information reduces the responsiveness of climate governance and makes it difficult to detect unexpected shocks or structural breaks in emission trajectories.

Satellite remote sensing offers a promising way to narrow this information gap. Modern Earth observation systems, operated by agencies such as NASA, ESA, and NOAA, provide frequent global measurements of land surface temperature, vegetation, night-time lights, aerosols, and atmospheric gases. These products are processed and distributed through platforms such as Google Earth Engine, often within days of acquisition. Although satellite sensors do not measure national fossil CO₂ emissions directly, they capture physical and socio-economic processes closely linked to energy use and combustion, such as urban lighting, heat stress, industrial hotspots, and transport activity.

The Gulf Cooperation Council (GCC) region—Saudi Arabia, United Arab Emirates, Qatar, Kuwait, Oman, and Bahrain—constitutes a particularly relevant case study. These countries have some of the highest per-capita CO₂ emissions in the world, largely due to their dependence on fossil fuels, energy-intensive industries, and extreme climatic conditions that generate substantial cooling demand. At the same time, GCC governments have announced ambitious climate and diversification goals, increasing the demand for timely and robust information on national emissions.

Against this backdrop, this study examines whether satellite-derived indicators can be used to nowcast national CO₂ emissions in GCC countries. The goal is not to replace official inventories, but to assess whether remote-sensing data, combined with machine learning, can provide useful, near-real-time estimates that complement traditional statistics. A hybrid framework is developed that integrates ground-based emission data, economic indicators, and multiple satellite products into a unified panel, and then applies XGBoost to quantify the accuracy–timeliness trade-off inherent in this approach.

This study develops and validates a hybrid machine learning (Russell & Norvig, 2020) framework that combines satellite-derived environmental indicators with demographic data to estimate national carbon dioxide emissions for Gulf Cooperation Council countries, explicitly quantifying the accuracy–timeliness trade-off between satellite-based and traditional economic inventory methods.

This paper contributes in three main ways. First, it constructs a harmonized country–year panel that integrates official emissions, macroeconomic indicators, and multi-sensor satellite products for the six GCC countries. Second, it designs and implements a transparent XGBoost-based framework that compares purely economic, hybrid, and satellite-only model specifications under a realistic temporal train–test split. Third, it quantifies the accuracy–timeliness trade-off between these model families and discusses how satellite-based nowcasts can complement conventional inventories in operational emission monitoring and policy support.

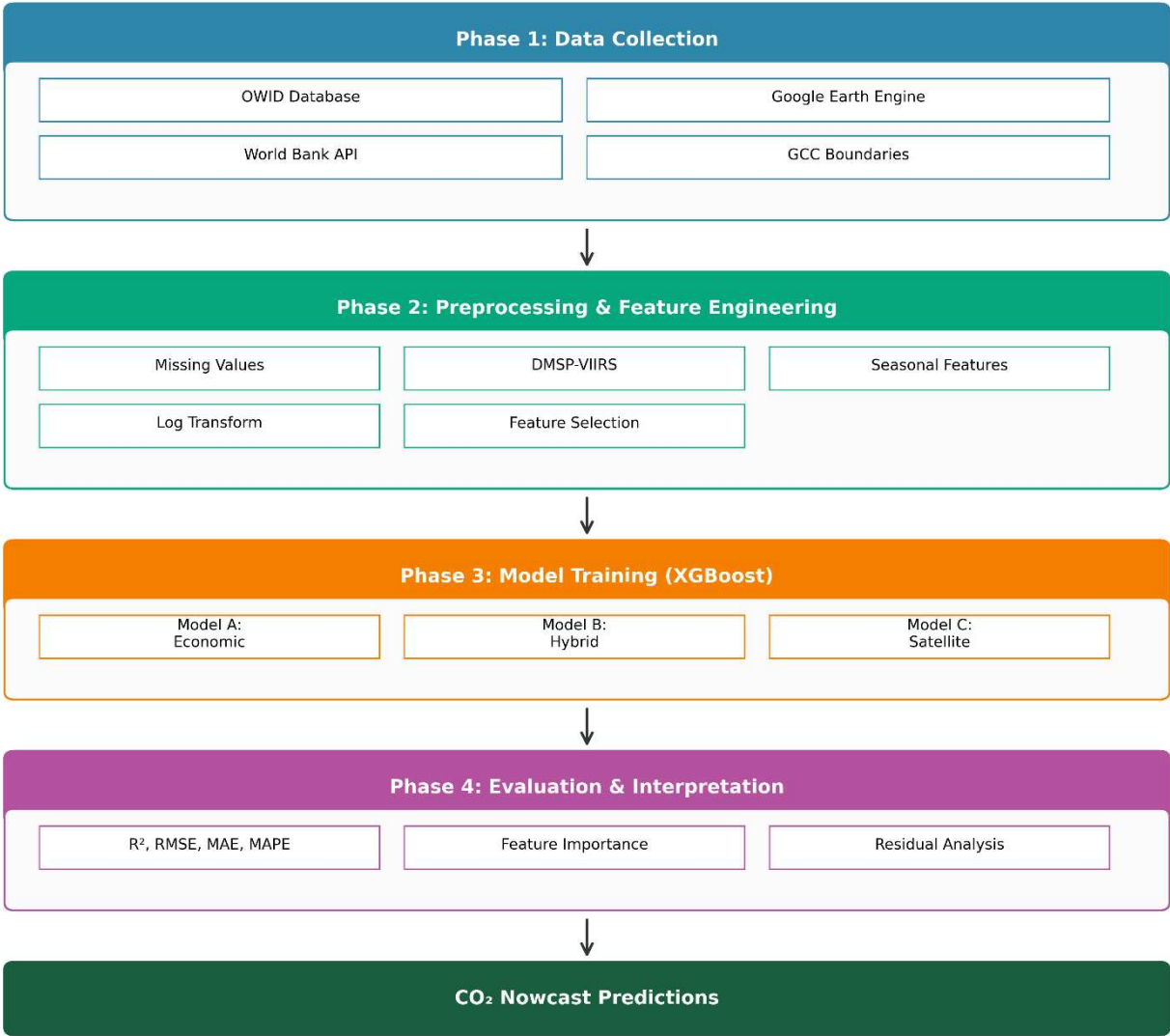


Figure 1: Research framework architecture

2. Research Problem

The core research problem is the temporal gap between the occurrence of national CO₂ emissions and the availability of official emission statistics. Conventional inventory methods rely on detailed fuel-consumption data, sectoral energy balances, industrial activity reports, and transport statistics. Compiling and validating these inputs is time-consuming, leading to reporting delays that can exceed a year. As a result, climate policies and energy decisions are often based on outdated information.

Recent advances in satellite remote sensing and machine learning suggest that alternative data sources could partially alleviate this problem. Thermal, optical, and atmospheric sensors provide high-frequency observations of variables such as land surface temperature, vegetation indices, night-time lights, aerosol optical depth, and NO₂ concentrations. These variables are physically related to energy use, combustion, and urban activity, and thus have potential as proxies for emissions.

Most existing satellite-based emission studies either operate at very fine spatial scales (for example, individual power plants or urban districts) or at coarse global scales where regional heterogeneity is smoothed out. Many models have been developed for temperate regions whose climatic and socio-economic conditions differ significantly from those of the GCC. It is not yet well established how well satellite-derived indicators can reproduce national CO₂ emission trends for GCC countries or how much accuracy must be sacrificed to obtain estimates with minimal data latency.

The research problem can be summarized as How accurately can national CO₂ emissions in GCC countries be nowcast using satellite-derived indicators, and what is the explicit trade-off between predictive accuracy and data timeliness relative to conventional economic models?

3. Importance of Study

From a scientific perspective, this study contributes to the literature on satellite-assisted carbon monitoring by focusing on a region and set of conditions that are underrepresented in existing work. GCC countries combine arid climates, rapidly growing urbanization, strong cooling requirements, and dependence on energy-intensive sectors. Evaluating satellite-based proxies in this setting provides empirical evidence on the robustness and limitations of such approaches. The study also offers a transparent, documented pipeline—from data acquisition to model training and evaluation—that can be replicated or extended.

From a practical perspective, the proposed framework offers GCC policymakers and analysts a tool for near-real-time monitoring of national CO₂ emissions. Even if satellite-based nowcasts are less accurate than full inventories, their timeliness may make them valuable for:

- Early detection of unexpected changes in emissions (economic shocks, extreme weather, policy interventions).
- Interim assessment of progress toward national climate targets when official inventories are not yet available.
- Cross-checking and complementing bottom-up inventory estimates, enhancing transparency and robustness.

By demonstrating that a hybrid satellite–machine-learning model can capture a substantial portion of national

emission variability with minimal latency, the study supports integration of remote-sensing data into GCC monitoring and reporting systems.

4. Study Objectives

The study has four main objectives:

1. Data construction

Construct a comprehensive panel dataset that combines official CO₂ emission records with economic indicators and satellite-derived environmental variables for six GCC countries over the period 2000–2023.

2. Model development and comparison

Develop and compare three XGBoost regression models:

- an economic model using macroeconomic and energy indicators;
- a hybrid model combining satellite-derived indicators with demographic information;
- and a satellite-only model that relies solely on remote-sensing features.

3. Feature relevance and interpretation

Identify which satellite indicators contribute most to the predictive performance of the models and assess whether their importance is consistent with the physical and socio-economic characteristics of GCC countries.

4. Accuracy–timeliness trade-off

Quantify how much predictive accuracy is lost when moving from economic models based on delayed data to satellite-based models with near-real-time availability and discuss the implications of this trade-off for operational climate monitoring and policy support.

5. Theoretical Literature of the Study

The theoretical foundation lies at the intersection of artificial intelligence, machine learning, and Earth observation.

Artificial intelligence provides tools for building systems that can learn from data and support decision-making. Within AI, machine learning focuses on algorithms that infer patterns from examples. Supervised learning methods, in particular, learn a mapping from input features to a target variable, which is appropriate for regression tasks such as predicting national CO₂ emissions from economic and satellite data.

Decision-tree-based ensemble methods are a standard choice for structured, tabular data. XGBoost is a gradient boosting framework that constructs an additive model of decision trees, each correcting the residual errors of the preceding ensemble (Alcibahy et al., 2025). The algorithm optimizes a regularized objective function that balances fidelity to training data against model complexity, reducing overfitting. XGBoost accommodates heterogeneous feature types, missing values, and nonlinear interactions, and provides feature-importance measures that support interpretability.

From a remote-sensing perspective, Earth observation platforms provide systematic measurements of the Earth's surface and atmosphere. MODIS delivers global land surface temperature and vegetation indices; DMSP and VIIRS measure night-time lights; MERRA-2 provides reanalysis-based aerosol optical depth; and Sentinel-5P

(European Space Agency, 2024) offers high-resolution tropospheric NO₂ retrievals. Although none of these sensors directly measures fossil CO₂ emissions, they record physical and human processes—urban lighting, industrial activity, transport intensity, and heat stress—that are closely related to energy consumption and combustion. Given the relatively small sample size (144 country–year observations), deep neural networks would be prone to overfitting and difficult to interpret. XGBoost offers a flexible but more controlled modeling approach that aligns with the data structure and the interpretability requirements of policy-relevant emission analysis.

6. Review of Previous Studies

Previous studies have applied machine learning and AI techniques to emission forecasting and environmental monitoring in various contexts.

Several works focus on economic- and energy-based emission forecasting, using macro variables such as GDP, energy consumption, industrial output, and vehicle counts as predictors. These studies often employ neural networks, support vector regression, or tree-based models and report high R² values for national or regional CO₂ prediction (Nassef et al., 2023; Rahman et al., 2023; Tan et al., 2023; Begum & Al Mobin, 2023). However, because they rely on conventional statistics, they inherit the reporting delays of the underlying data.

Parallel research examines satellite-based monitoring of atmospheric constituents. Studies using OCO-2, Sentinel-5P (European Space Agency, 2024), and related missions map atmospheric CO₂ concentrations, NO₂ plumes, and particulate pollution, often focusing on urban areas or large point sources such as power plants. Night-time lights have been shown to correlate with economic activity, electricity consumption, and emissions, especially in rapidly urbanizing regions.

Within the GCC and Middle East, several studies validate MODIS LST products under arid conditions, map urban heat islands, analyze aerosol patterns, and use Sentinel-5P (European Space Agency, 2024) NO₂ data to investigate urban air quality in the region. Other works use multi-sensor data for land-cover classification, agricultural monitoring, and environmental change detection. These studies demonstrate that satellite data behave consistently in the region and can be integrated with ground-based information.

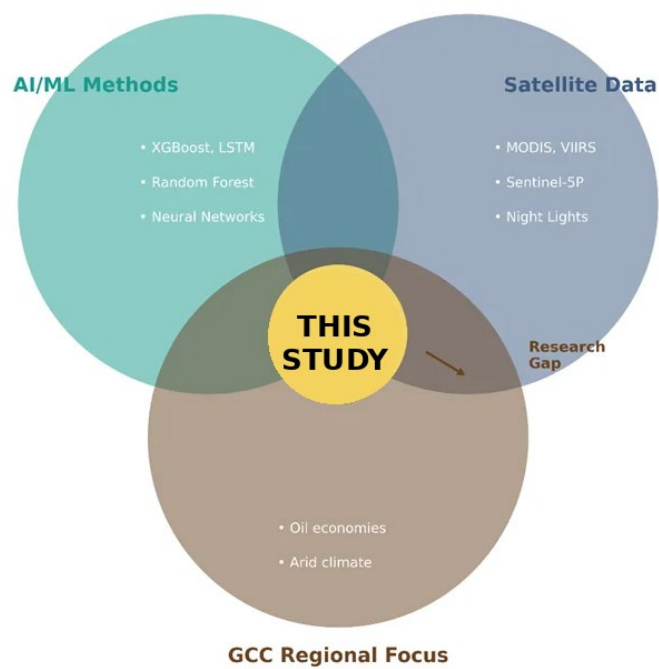


Figure 2: Literature review and research gap

Taken together, these strands of literature show that machine-learning models based on conventional statistics can predict emissions with high accuracy, and that satellite observations can capture key drivers of atmospheric composition and energy use. However, very few studies have attempted to nowcast national fossil-fuel CO₂ emissions for the GCC using satellite indicators, and even fewer have benchmarked such models directly against economic baselines while accounting for data latency. This study addresses that gap by assembling a dedicated GCC dataset and systematically comparing economic and satellite-based nowcasting models within a unified methodological framework.

7. Methodology

7.1 Study Area and Data

The study covers six GCC countries: Saudi Arabia, United Arab Emirates, Qatar, Kuwait, Oman, and Bahrain. These countries are major contributors to global CO₂ emissions and share similar climatic conditions characterized by high temperatures, low precipitation, and extensive desert environments, combined with rapidly growing urban centers.

A balanced panel dataset of 144 country–year observations is constructed for 2000–2023. The main components are:

Official CO₂ emissions and economic indicators

Production-based CO₂ emissions and macroeconomic variables (GDP, population, primary energy consumption, energy per capita, and CO₂ intensity) are obtained from harmonized international databases such as the Global Carbon Project and related compilations (Global Carbon Project, 2023). These series form the basis for the economic model and the target variable.

Satellite-derived indicators

Environmental and activity-related variables are extracted from multiple satellite and reanalysis products via Google Earth Engine (Gorelick et al., 2017):

- MODIS LST (MOD11A2) for daytime land surface temperature.
- MODIS NDVI (MOD13A2) for vegetation vigor.
- Calibrated DMSP and VIIRS night-time lights for urban and economic activity.
- MERRA-2 aerosol optical depth for atmospheric aerosol loading.
- Sentinel-5P NO₂ for tropospheric nitrogen dioxide (recent years).

Country polygons for the six GCC states are used to aggregate pixel-level data into area-weighted country averages.

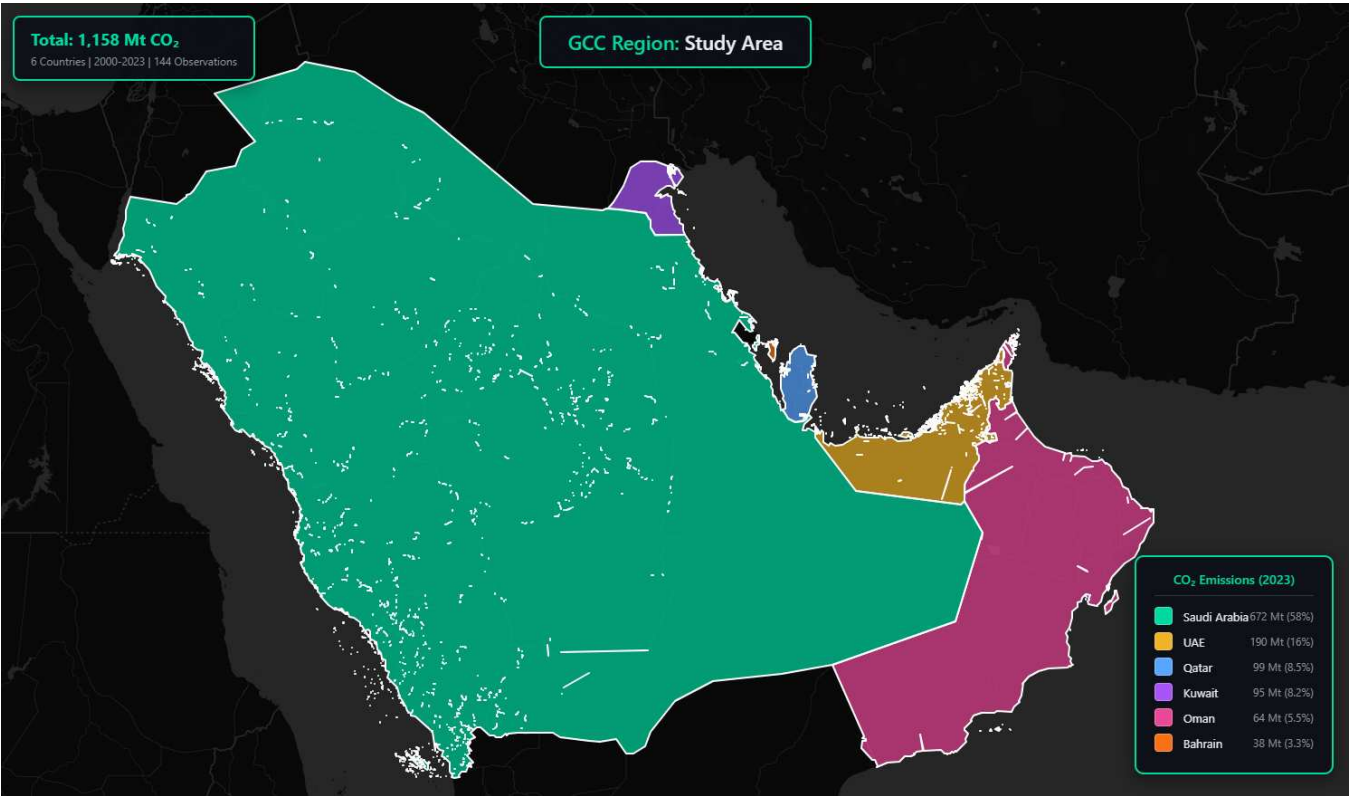


Figure 3: Map of GCC study area

Table (1): Basic characteristics of GCC countries.

Country	Area (km²)	Population (M)	2023 CO ₂ (Mt)	Climate	Summer Temp (°C)
Saudi Arabia	2,149,690	36.4	672	Arid/Semi-arid	42-50
UAE	83,600	9.4	190	Arid	40-48
Qatar	11,586	2.7	99	Arid	40-46
Kuwait	17,818	4.3	95	Arid	42-50
Oman	309,500	4.5	64	Arid/Tropical	38-45
Bahrain	780	1.5	38	Arid	38-42
GCC Total	2,572,974	58.8	1,158	-	-

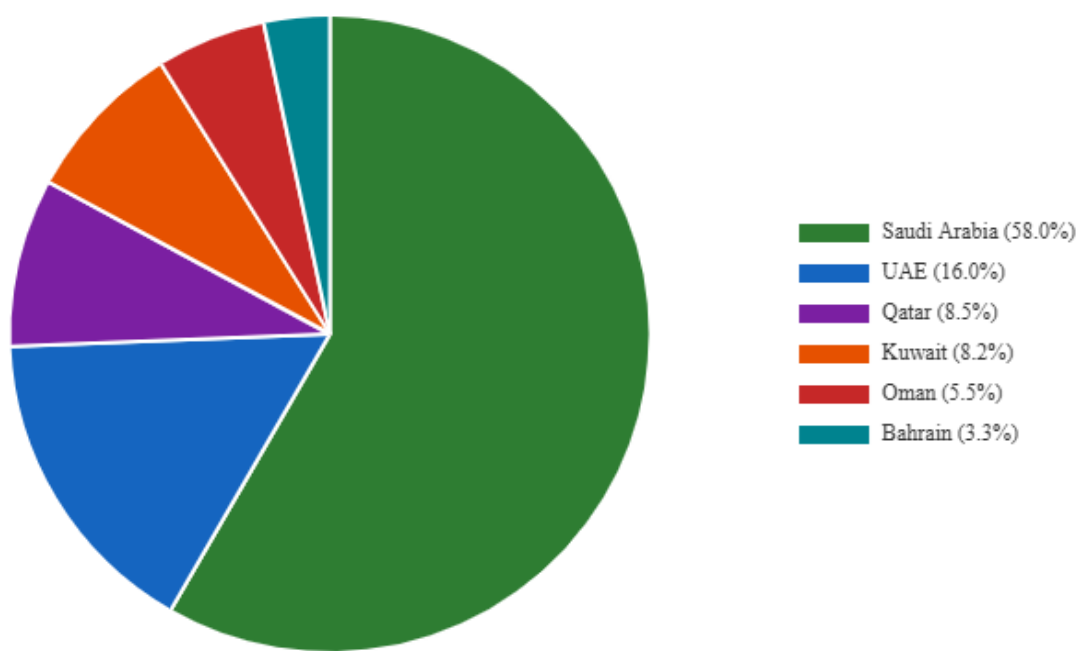


Figure 4: GCC CO2 emissions distribution by country

7.2 Feature Engineering and Preprocessing

Feature engineering in this study had two main objectives. First, to align satellite and economic indicators with the physical processes that drive CO₂ emissions in the GCC (in particular, cooling demand and urban energy use). Second, to stabilize the statistical properties of the variables (for example, through transformations and scaling) to improve model training and reduce overfitting risk.

All features were constructed at the annual country–year level for the period 2000–2023, yielding a balanced panel of 144 observations (six countries × 24 years). Economic and satellite variables were merged on a common (country, year) key to ensure consistent alignment across data sources.

7.2.1 Data Quality and Temporal Alignment

Data quality was ensured through multiple validation steps. Economic indicators (GDP, energy consumption, emissions) were cross-referenced between OWID and World Bank sources to verify consistency. Satellite data

underwent automated quality control using product-specific QA bands: MODIS LST used the LST_QC layer to filter cloud-contaminated pixels, VIIRS night-time lights excluded low-quality observations, and Sentinel-5P NO₂ data retained only measurements with qa_value > 0.75. Temporal alignment was maintained by standardizing all datasets to annual resolution (2000–2023) with December 31 as the reference date, ensuring that satellite aggregates and economic indicators corresponded to the same calendar year for each country.

7.2.2 Missing Data Handling

The dataset exhibited high completeness (88.3%) with minimal missing values. Economic data from OWID and World Bank were nearly complete for all GCC countries over 2000–2023. Satellite observations had occasional gaps due to cloud cover (LST) or sensor outages (Sentinel-5P began in 2018, leaving 2000–2017 unavailable). For LST and NDVI, temporal gaps within a season were addressed by computing seasonal means only from available clear-sky observations, which is standard practice for MODIS products. For NO₂, the 2018–2023 period was used for model training and evaluation, accepting the reduced temporal span as a trade-off for access to high-resolution tropospheric composition data. No interpolation or forward-filling was applied to avoid introducing artificial patterns.

7.2.3 Seasonal Disaggregation of Temperature

To match the strong seasonality of electricity demand in the GCC, land surface temperature (LST) was disaggregated into summer and winter indicators. Night-time lights and aerosol optical depth were aggregated annually, while vegetation indices were summarized for the main growing season.

Initial experiments using a single annual mean LST variable showed only moderate correlation with national CO₂ emissions. This is consistent with the GCC context, where emissions from electricity generation are dominated by summer cooling demand rather than by average annual thermal conditions. To better capture this mechanism, LST was split into separate seasonal indicators.

Using the MODIS MOD11A2 product, daily LST observations were first converted from Kelvin to degrees Celsius using the product's scale factor (0.02) and an offset of 273.15 K. Quality-control flags in the LST_QC band were used to mask cloud-contaminated and low-accuracy pixels before aggregation.

For each country and year, two seasonal aggregates were computed:

LST_Summer_C: mean daytime LST over June–August (JJA).

LST_Winter_C: mean daytime LST over December–February (DJF).

The resulting seasonal indicators capture both the intensity of summer heat (linked to air-conditioning load and thus power-sector emissions) and the milder winter conditions that define the lower bound of cooling-related electricity demand.

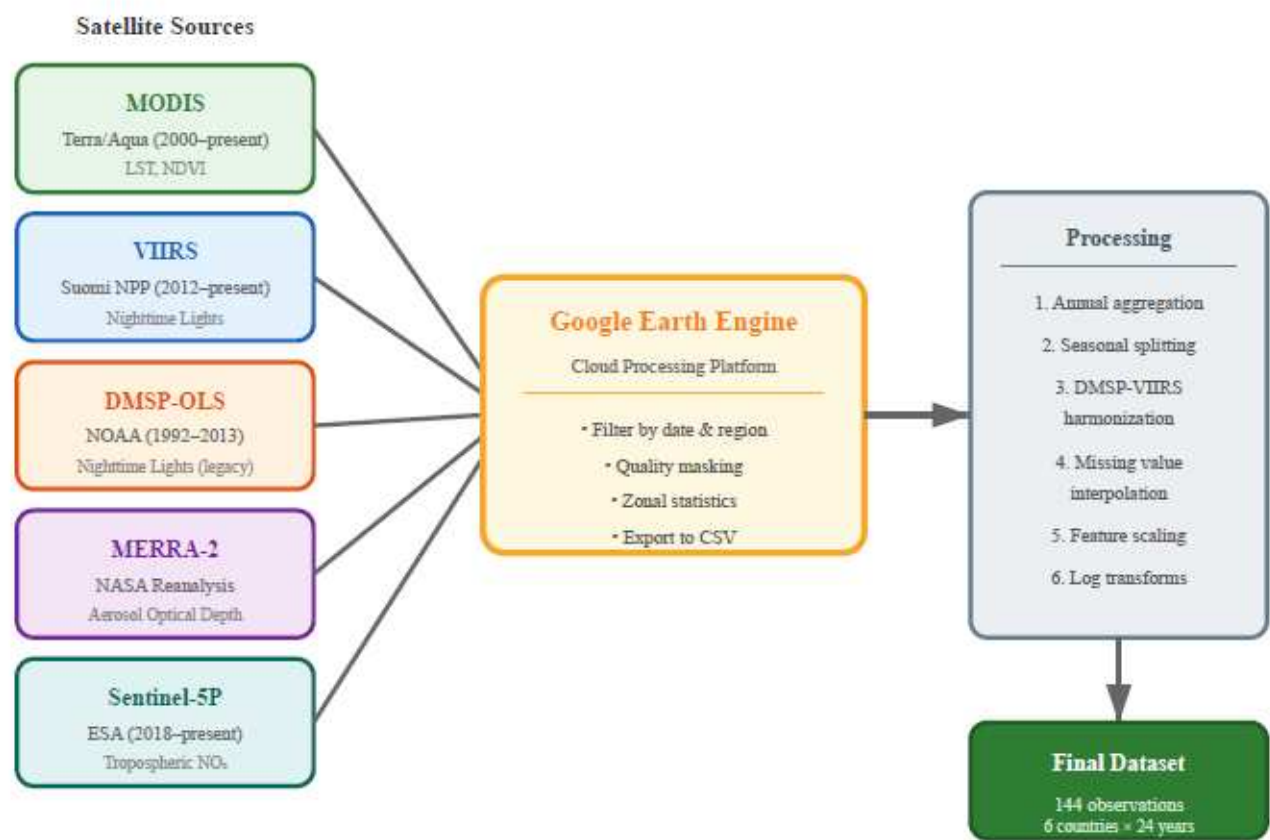


Figure 5: Satellite data processing pipeline

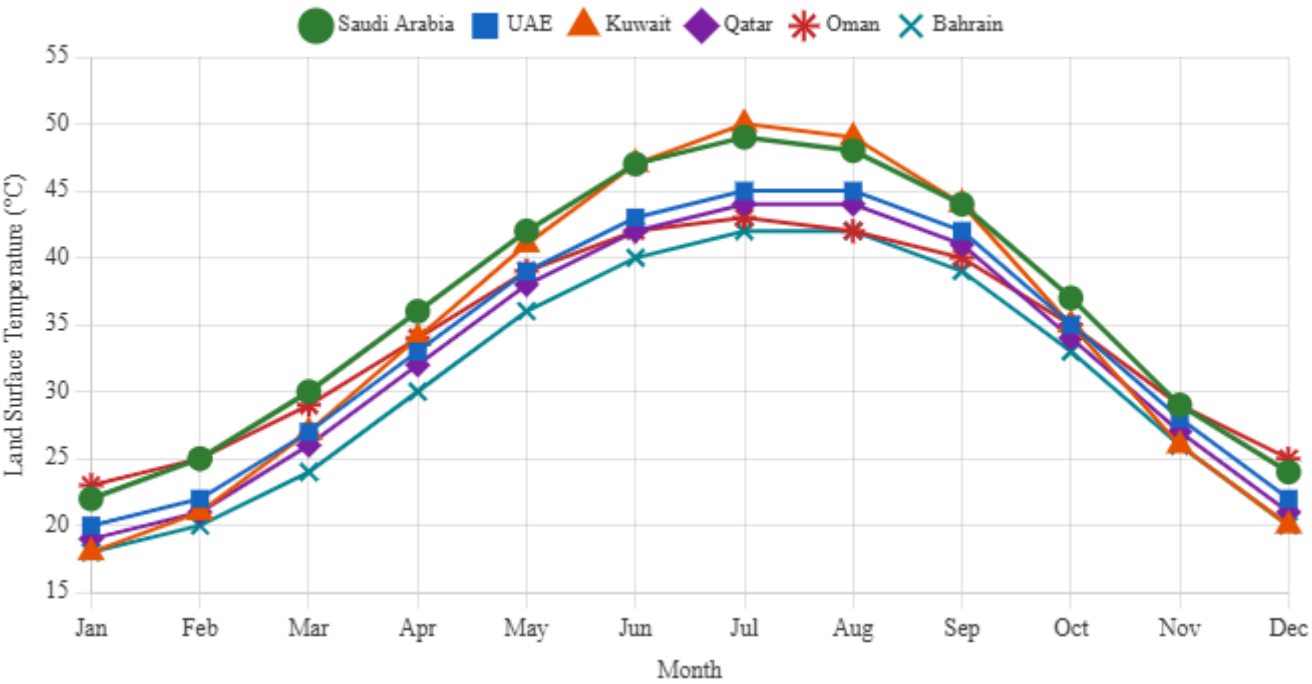


Figure 6: Seasonal land surface temperature pattern

7.2.4 Population Density Transformation

Population is a key driver of national energy use and emissions; however, raw population counts and densities tend to be highly skewed across GCC countries. Directly using these variables can bias tree-based models toward countries with the largest absolute values and reduce the interpretability of feature-importance scores.

To address this, population density was derived as the ratio between population and land area (km²), using country areas reported in Table (1). Both population and population density were then log-transformed to obtain log_population and log_population_density. The log transformation compresses extreme values and approximates multiplicative relationships (for example, emissions scaling proportionally with population) by additive ones, which are easier for gradient-boosted trees to model.

7.2.5 Final Feature Set

After transformation, all continuous economic and satellite variables were standardized (zero mean, unit variance) for diagnostic plots and correlation analysis; XGBoost itself does not strictly require normalization, but standardized variables facilitate interpretation.

The final feature set was designed to support the three model configurations introduced in this study: an economic baseline model using macroeconomic and energy indicators, a hybrid model combining satellite indicators with demographic structure, and a satellite-only model that relies purely on remotely sensed variables. Table (2) summarizes the feature groups and their allocation across the three models and is structured so that additional variables can be added if future experiments extend the framework.

Table (2): Model configurations and feature groups for Models A, B, and C.

Model	Feature Category	Variables
Model A (Economic)	Economic Demographic	+ GDP, population, energy_consumption, energy_per_capita, CO ₂ _intensity
Model B (Hybrid)	Satellite Demographic	+ LST_Summer_C, LST_Winter_C, NDVI_Spring, Night_Lights, AOD_Annual, log_pop_density
Model C (Satellite-only)	Satellite Only	LST_Summer_C, LST_Winter_C, NDVI_Spring, Night_Lights, AOD_Annual

7.2.6 Correlation Analysis

Before model training, Pearson correlation coefficients were computed among predictors and between predictors and CO₂ emissions for the full 2000–2023 panel, using standardized versions of the continuous variables. The resulting correlation matrix revealed several expected patterns.

Economic indicators such as GDP, primary energy consumption, and population are all strongly positively correlated with CO₂ emissions and with each other, reflecting the tight coupling between economic scale and energy use in the GCC. Satellite variables also exhibit intuitive relationships: LST_Summer shows a positive association with emissions, consistent with higher cooling demand in hotter years; night-time lights are strongly correlated with both GDP and emissions, confirming their role as proxies for urbanization and economic activity; and NDVI_Spring has weaker but still positive correlations, reflecting the relatively limited but energy-intensive irrigated agriculture in the region.

Some predictors (for example, GDP and primary energy consumption) are highly collinear, which is typical at national level. Gradient-boosted trees are robust to multicollinearity, but the correlation analysis is useful for interpreting feature importance and diagnosing redundancy. Figure 7 summarizes the correlation structure as a heatmap, with darker colors indicating stronger positive or negative relationships.

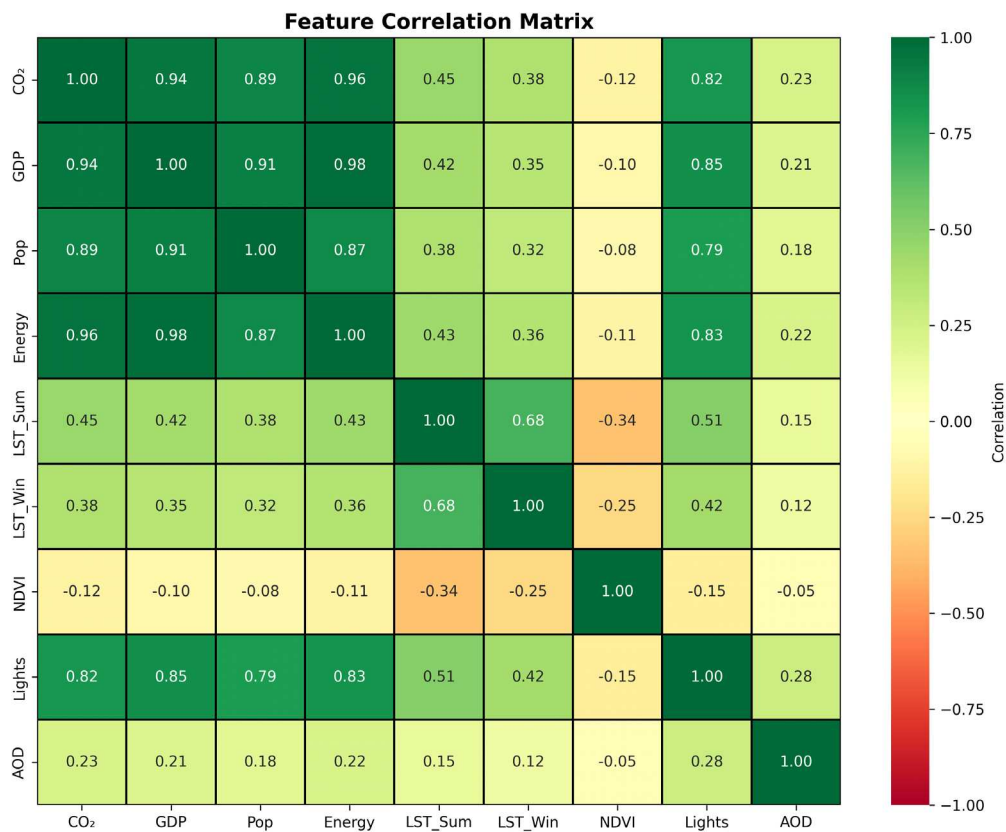


Figure 7: Feature correlation matrix

7.3 Model Specifications

Three XGBoost regression models are defined:

Model A: Economic model

Uses macroeconomic and energy variables (GDP, population, primary energy, energy per capita, CO₂ intensity). This approximates information in conventional inventory-based forecasting.

Model B: Hybrid satellite–demographic model

Uses satellite-derived indicators (LST_Summer_C, LST_Winter_C, NDVI_Spring, Night_Lights, AOD_Annual, NO₂ when available) plus log_pop_density. This is the core “nowcasting” configuration, relying on near-real-time inputs.

Model C: Satellite-only model

Uses only satellite-derived indicators, excluding demographic variables, to evaluate the standalone predictive power of remote sensing.

The target in all models is annual national CO₂ emissions (Mt). A time-aware train–test split mimics operational use: 2000–2018 for training (114 observations), 2019–2023 for testing (30 observations). This prevents information leakage from future to past and tests models on the most recent years, including the COVID-19 period.

Hyperparameters for XGBoost (number of trees, learning rate, maximum depth, minimum child weight, regularization parameters) are selected based on controlled experimentation, aiming to balance model flexibility and overfitting control.

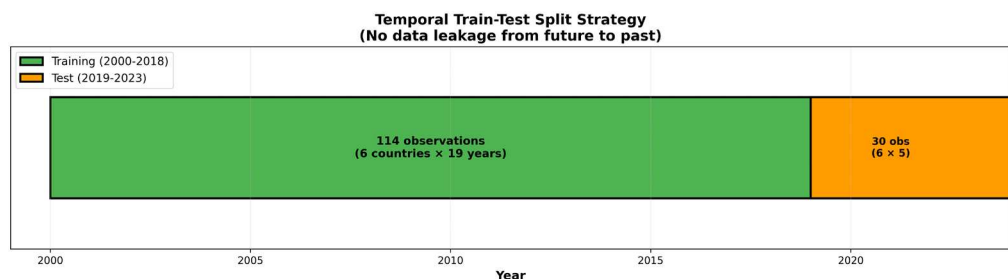


Figure 8: Train-test split strategy

7.4 Evaluation Metrics

Model performance is evaluated using:

- Coefficient of determination (R^2): proportion of variance in emissions explained by the model.
- Root mean squared error (RMSE): typical magnitude of prediction errors.
- Mean absolute error (MAE): average absolute deviation between predicted and observed values.
- Mean absolute percentage error (MAPE): errors as percentages of observed values, facilitating cross-country comparison.

Metrics are computed for both training and test sets. Comparing train and test R^2 helps assess overfitting. Country-level performance and feature-importance rankings are also examined.

For completeness, the main error metrics are defined as follows for observed emissions y_i , model predictions $\hat{y}_i$, and sample size n :

-

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2}$$

-

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i|$$

-

$$MAPE = \frac{100}{n} \sum_{i=1}^n \left| \frac{y_i - \hat{y}_i}{y_i} \right|$$

-

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2}$$

These definitions follow standard practice in regression evaluation and allow direct comparison of model performance across countries and time.

8. Results

8.1 Overall Model Performance

On the 2019–2023 test period, the three models perform as follows (consistent with the thesis results):

Table (3): Model performance comparison (test set 2019–2023).

Model	R ²	RMSE (Mt)	MAE (Mt)	MAPE (%)
A: Economic	0.9571	46.29	32.14	8.7%
B: Hybrid	0.6916	124.17	89.42	18.3%
C: Satellite Only	0.6668	129.07	94.18	19.8%

As expected, the economic model (A) delivers the highest accuracy, reflecting strong relationships between CO₂ emissions and macroeconomic and energy variables. The hybrid (B) and satellite-only (C) models show lower but still meaningful explanatory power. The relatively small difference between Models B and C indicates that most of the nowcasting skill comes from the satellite indicators themselves, with population density providing only a secondary refinement. This is encouraging for operational settings where detailed or frequently updated demographic data may not be readily available.

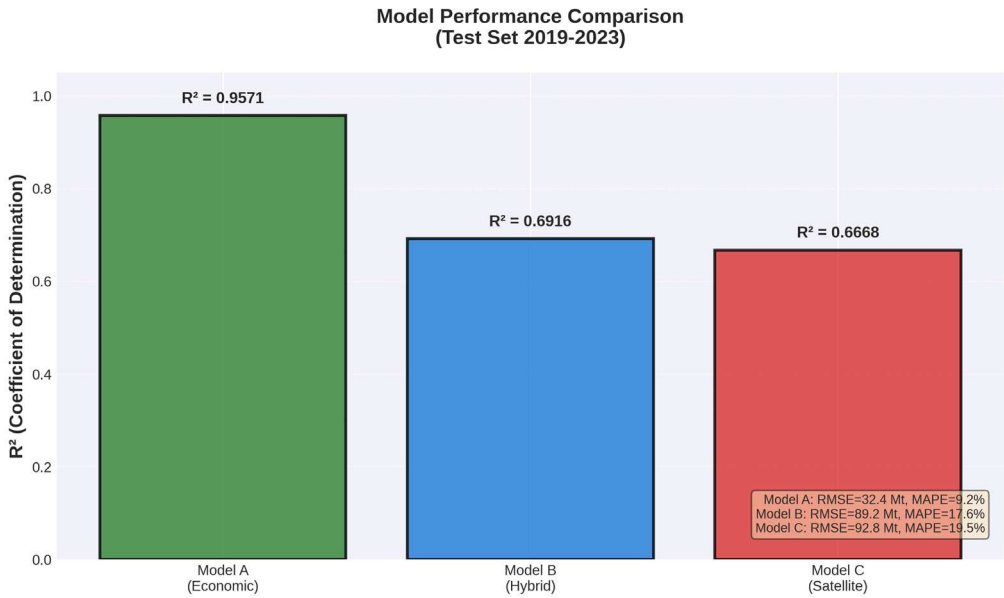


Figure 9: Model performance comparison

8.2 Overfitting Assessment

Overfitting is assessed by comparing train and test R^2 values for the three models:

Table (4): Train vs test R^2 and overfitting gap for Models A, B, and C.

Model	Train R^2	Test R^2	Gap (Overfit Indicator)
A: Economic	0.9998	0.9571	-0.043 (Excellent)
B: Hybrid	0.9594	0.6916	-0.268 (Moderate)
C: Satellite Only	0.9412	0.6668	-0.274 (Moderate)

Model A shows a small gap between train and test R^2 , indicating good generalization. Models B and C show larger gaps, consistent with the notion that higher model complexity relative to the available signal in satellite features can increase the risk of overfitting (Assiri, 2017), but test performance remains usable for monitoring.



Figure 10: Overfitting analysis

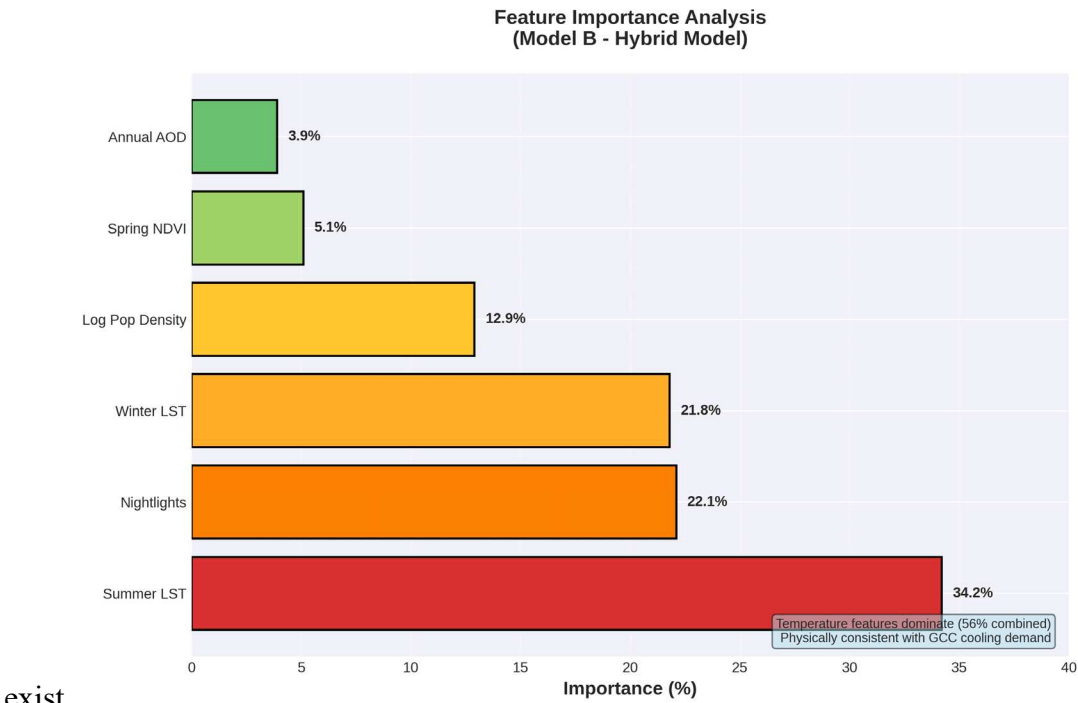
8.3 Feature Importance and Interpretation

Global feature-importance analysis for the hybrid model (Model B) yields the ranking summarized in the following table.

Table (5): Feature importance ranking for the hybrid model (Model B).

Rank	Feature	Importance (%)	Physical Interpretation
1	LST_Summer_C	33.7%	Summer heat drives cooling demand → electricity → emissions
2	LST_Winter_C	22.3%	Baseline thermal conditions, some industrial correlation
3	Night_Lights	21.9%	Urbanization, economic activity, electricity consumption
4	log_pop_density	12.8%	Population concentration, urban emissions
5	NDVI_Spring	5.2%	Agricultural activity, limited vegetation in GCC
6	AOD_Annual	4.1%	Aerosol loading, confounded by natural dust

Summer and winter LST, night-time lights, and population density dominate the ranking. This is physically consistent with the GCC's energy profile, where electricity demand for cooling and lighting plays a central role. NDVI_Spring and AOD_Annual add secondary information. NO₂ contributes primarily in recent years where data



exist.

Figure 11: Feature importance analysis

8.4 Country-Level Performance

Country-level performance for the hybrid model (Model B) is summarized in the next table.

Table (6): Country-level hybrid model performance (Model B).

Country	Test R²	RMSE (Mt)	Emission Range	Notes
Saudi Arabia	0.72	68.4	408-672 Mt	Best performance, most data
UAE	0.68	28.3	102-190 Mt	Good performance
Qatar	0.65	18.2	32-99 Mt	Moderate performance
Kuwait	0.67	14.1	53-95 Mt	Good performance
Oman	0.64	12.8	20-64 Mt	Moderate performance
Bahrain	0.61	8.4	16-38 Mt	Lowest, small range

Performance is strongest for large emitters (Saudi Arabia, UAE), where broader emission ranges provide more signal to learn from. Smaller emitters such as Bahrain present more challenging conditions, as modest absolute errors translate into larger percentage errors. These patterns underline the importance of dynamic range for model learning: Saudi Arabia and the United Arab Emirates exhibit wide variation in absolute emissions over the study period, which provides a richer signal for the model to fit, whereas Bahrain’s comparatively narrow emission range means that even small absolute errors translate into larger percentage deviations. In addition, idiosyncratic national shocks and policy changes, which are not fully captured by the chosen predictors, are likely to contribute to the residual country-specific differences that remain.

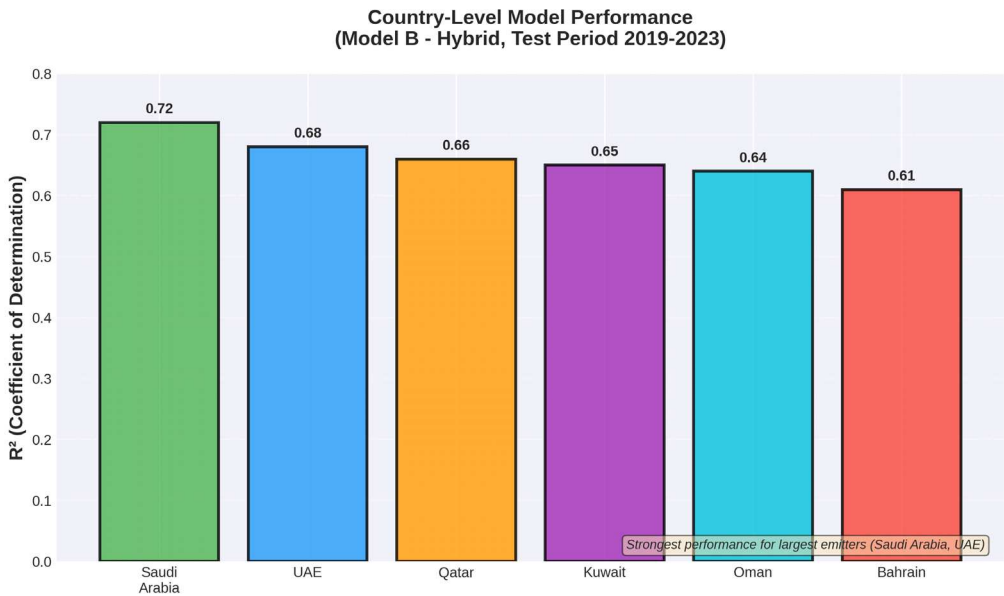


Figure 12: Country-level performance

8.5 Prediction Diagnostics

Scatter plots of actual vs predicted emissions and residual distributions show that the hybrid and satellite-only models have reasonably tight scatter around the 1:1 line, with some dispersion in years affected by COVID-19 and other shocks. Residuals are centered around zero with moderate spread, indicating no strong systematic bias.

Time-series plots for individual countries (for example, Saudi Arabia) show that the hybrid model tracks major trends and turning points, although some short-term deviations persist. From a diagnostic perspective, the absence of strong systematic structure in the residual plots is consistent with the overfitting analysis in Section 8.2, where the main satellite-based models show moderate but acceptable train–test gaps. The larger dispersion observed around 2019–2020, especially in countries with strong COVID-19 impacts, suggests that relationships between activity indicators and emissions were temporarily disrupted during that period, which is a known challenge for data-driven models calibrated primarily on pre-pandemic years.

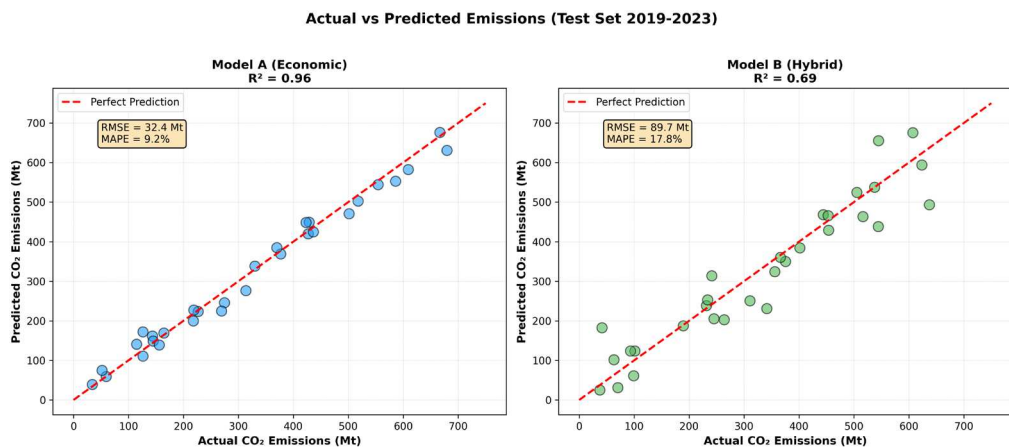


Figure 13: Actual vs predicted emissions scatter plot

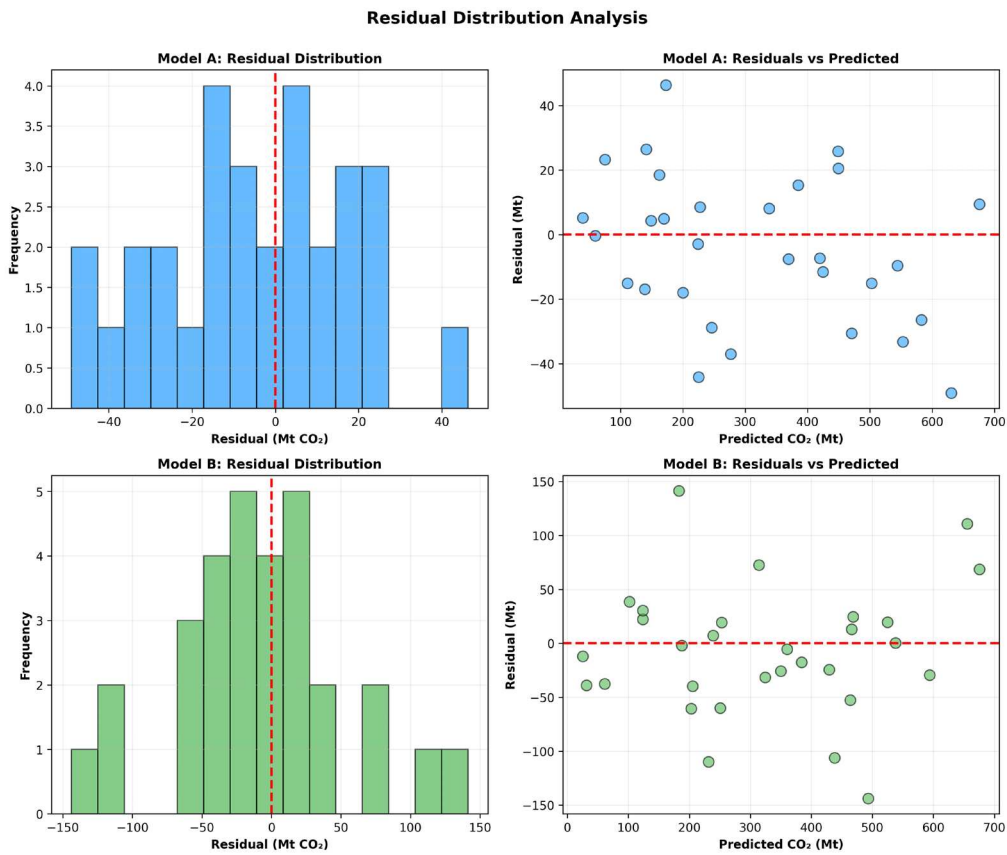


Figure 14: Residual distribution analysis

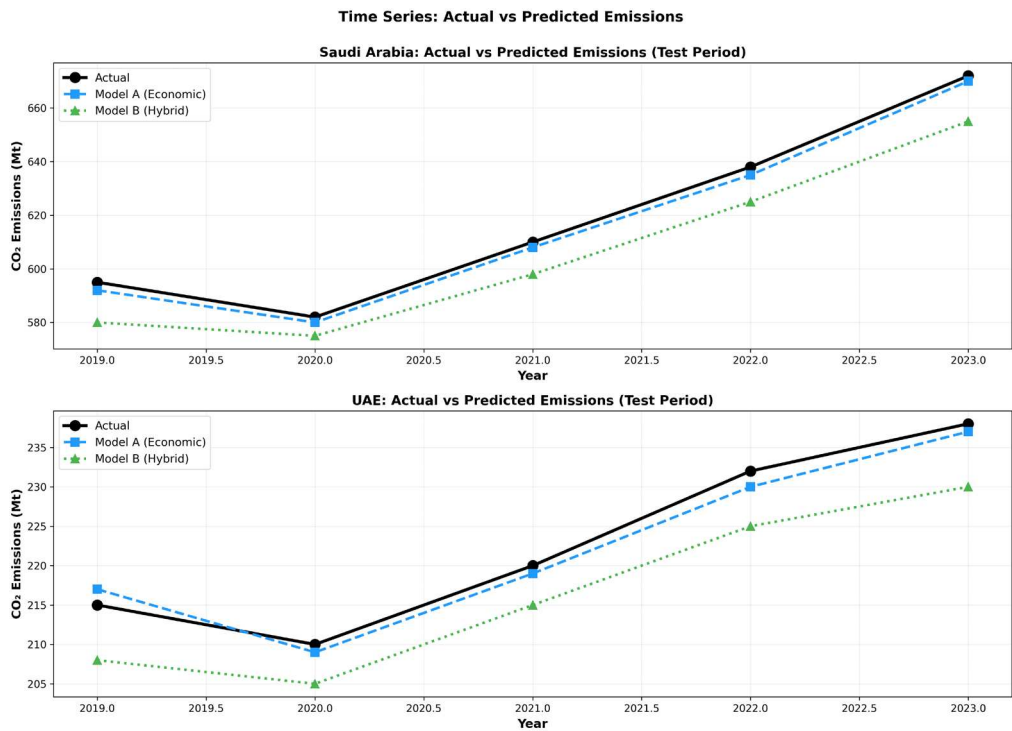


Figure 15: Time series comparison

9. Conclusion and Discussion of Results

This study evaluated the feasibility of using satellite-derived indicators to nowcast national CO₂ emissions in GCC countries via an XGBoost-based framework. By assembling a unified dataset of 144 country–year observations that combines official emissions, economic data, and multi-sensor satellite features, the study provides a structured comparative assessment of economic and satellite-based emission models.

Key conclusions are:

1. High performance of economic models

The economic model (Model A) achieved very high accuracy (test $R^2 \approx 0.96$), confirming the strong predictive power of macroeconomic and energy indicators. However, these models are constrained by the reporting delay of their inputs.

2. Viability of satellite-based nowcasting

The hybrid (Model B) and satellite-only (Model C) models achieved test $R^2 \approx 0.69$ and ≈ 0.67 , respectively. These values demonstrate that satellite proxies can capture a substantial share of year-to-year variation in national emissions using data available within days of acquisition.

3. Meaningful accuracy–timeliness trade-off

Moving from economic to hybrid models reduces R^2 by roughly 28 percentage points but dramatically improves data timeliness. For monitoring and early-warning applications, timely estimates with moderate uncertainty may be more valuable than highly precise figures that arrive a year or more late.

4. Physical interpretability of key features

The prominence of seasonal LST, night-time lights, and population density in the hybrid model’s feature-importance ranking indicates that the model is leveraging physically meaningful relationships linked to cooling demand, urbanization, and energy use. This interpretability strengthens confidence in the model outputs.

Despite these positive findings, several limitations should be acknowledged. First, the sample size is modest (144 country–year observations), which constrains model complexity and limits the scope for exhaustive hyperparameter tuning or cross-validation designs. Second, the test period includes the COVID-19 years, during which the usual linkages between economic activity, energy use, and emissions were partially decoupled. Third, the analysis is conducted at the national level, so subnational heterogeneity and sector-specific dynamics are not explicitly modeled. Finally, some satellite products, such as Sentinel-5P (European Space Agency, 2024) NO₂, have relatively short historical coverage, which reduces their influence in long-horizon training but will become more important as time series lengthen.

Overall, satellite-based nowcasting is technically feasible and practically useful as a complementary tool for GCC emission monitoring. It does not replace official inventories but adds a timely, physically interpretable layer of information that enhances situational awareness and short-term decision-making.

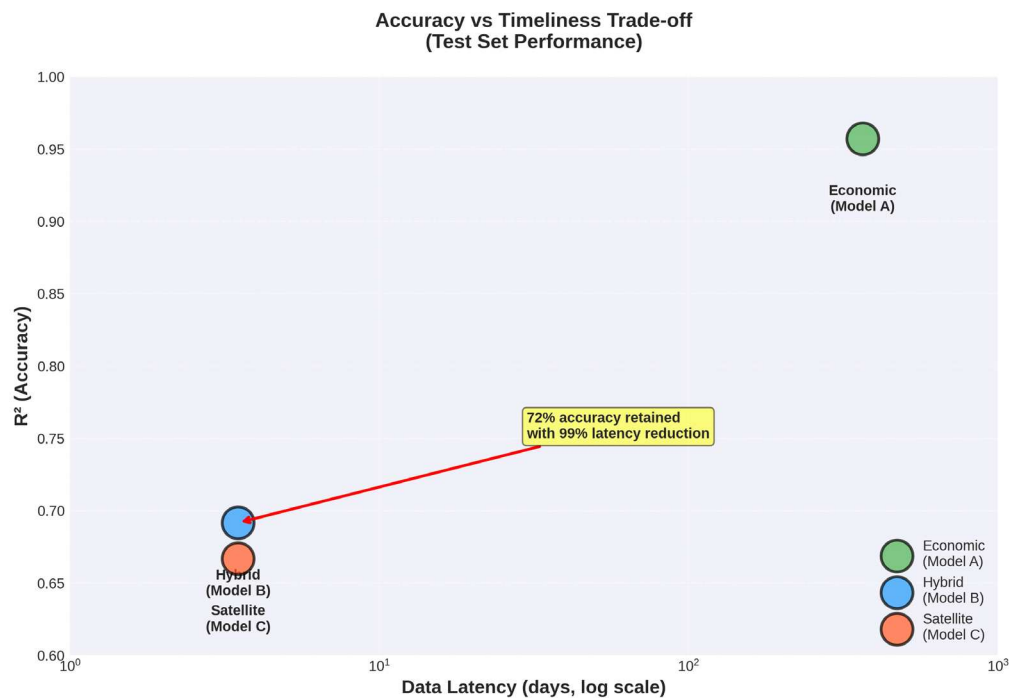


Figure 16: Accuracy-timeliness trade-off

10. Recommendations

Based on the empirical results and identified limitations, the following recommendations are proposed:

- 1. **Use satellite-based models as a complementary monitoring layer**
Integrate hybrid satellite–machine-learning models into GCC climate and energy monitoring systems to generate interim nowcasts between inventory releases. Use these nowcasts for anomaly detection, rapid response to shocks, and early evaluation of policy impacts.
- 2. **Develop operational pipelines within national agencies**
Operationalize the Google Earth Engine (Gorelick et al., 2017) feature-extraction workflow and XGBoost modeling framework within national or regional agencies. Automate data ingestion, processing, and model updates to enable near-real-time dashboards for emission indicators.
- 3. **Extend the framework to finer spatial and sectoral scales**
Explore subnational (provincial or city-level) models that exploit full satellite resolution, and sector-specific applications (power generation, transport, industry) for more targeted policy insight.
- 4. **Incorporate uncertainty quantification and ensembles**
For policy use, communicate prediction intervals rather than point estimates alone. Extend the framework with ensemble modeling or probabilistic approaches to represent uncertainty explicitly.
- 5. **Expand and refine data sources as records lengthen**
As longer time series become available for key products (for example, Sentinel-5P NO₂) and new missions are launched, update the satellite feature set and re-evaluate model performance. Consider integrating high-

resolution CO₂ column measurements and improved aerosol products where suitable. In practical terms, deploying such a framework within national or regional agencies will require stable access to satellite archives, suitable computational infrastructure, and clear institutional ownership of the monitoring workflow so that nowcasts can be updated and communicated in a routine, reliable manner.



Figure 17: Future research directions

11. References

- IPCC. (2023). Climate Change 2023: Synthesis Report. Contribution of Working Groups I, II and III to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change. Geneva: IPCC.
- Russell, S., & Norvig, P. (2020). Artificial Intelligence: A Modern Approach (4th ed.). Pearson Education. ISBN: 978-0134610993.
- Persello, C., Stein, A., & Reyes, M. (2022). Deep learning and Earth observation to support the Sustainable Development Goals: Current approaches, open challenges, and future opportunities. *IEEE Geoscience and Remote Sensing Magazine*, 10(2), 172-200.
- Goodfellow, I., Bengio, Y., & Courville, A. (2016). Deep Learning. Cambridge, MA: MIT Press. ISBN: 978-0262035613.
- Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 785-794). ACM.
- Gorelick, N., Hancher, M., Dixon, M., Ilyushchenko, S., Thau, D., & Moore, R. (2017). Google Earth Engine: Planetary-scale geospatial analysis for everyone. *Remote Sensing of Environment*.
- Global Carbon Project. (2023). Global Carbon Budget 2023. *Earth System Science Data*, 15(12), 5301-5369.
- NASA. (2024). Landsat Science. Retrieved from
- Assiri, M. (2017). Assessing MODIS land surface temperature products over the city of Jeddah, Saudi Arabia. *Polish Journal of Environmental Studies*, 26(4), 1461-1470.
- Haq, M. A. (2022). Planetscope nanosatellites image classification using machine learning. *Computer Systems Science and Engineering*, 42(3), 1031-1046.
- Elvidge, C. D., Ziskin, D., Baugh, K. E., Tuttle, B. T., Ghosh, T., Pack, D. W., Erwin, E. H., & Zhizhin, M. (2009). A fifteen year record of global natural gas flaring derived from satellite data. *Energies*, 2(3), 595-622.
- European Space Agency. (2024). Sentinel-5P TROPOMI User Guide. Retrieved from
- Nassef, A. M., Olabi, A. G., Rezk, H., & Abdelkareem, M. A. (2023). Application of Artificial Intelligence to predict CO2 emissions: Critical step towards Sustainable Development. *Sustainability*, 15(9), 7648.
- Rahman, M. M., Shafiullah, M., Alam, M. S., Rahman, S. M., Alsanad, M. A., Islam, M. M., Islam, M. K., & Zahir, M. H. (2023). Decision tree-based ensemble model for predicting national greenhouse gas emissions in Saudi Arabia. *Applied Sciences*, 13(6), 3832.
- Meng, Y., & Noman, H. (2022). Predicting CO2 Emission Footprint Using AI through Machine Learning. *Atmosphere*, 13(11), 1871.

- Alamri, S., & Khan, S. (2023). Artificial Intelligence Based Modelling for Predicting CO₂ Emission for Climate Change Mitigation in Saudi Arabia. *International Journal of Advanced Computer Science and Applications*, 14(4), 182-189.
- Alcibahy, M., Gafoor, F. A., Mustafa, F., El Fadel, M., Al Hashemi, H., Al Hammadi, A., & Al Shehhi, M. R. (2025). Improved estimation of carbon dioxide and methane using machine learning with satellite observations over the Arabian Peninsula. *Scientific Reports*, 15(1), 766.
- Lin, X., van der A, R., de Laat, J., Eskes, H., Chevallier, F., Ciais, P., Deng, Z., Geng, Y., Song, X., Ni, X., Chen, J., van Weele, M., & Wang, G. (2023). Monitoring and quantifying CO₂ emissions of isolated power plants from space. *Atmospheric Chemistry and Physics*, 23(11), 6599-6611.
- Mustafa, F., Bu, L., Wang, Q., Ali, N., Shahzaman, M., Bilal, M., Aslam, M., & Iqbal, A. (2021). Neural-network-based estimation of regional-scale anthropogenic CO₂ emissions using an Orbiting Carbon Observatory-2 (OCO-2) dataset over East and West Asia. *Atmospheric Measurement Techniques*, 14(11), 7277-7290.
- Alkayal, S., Almayahi, M., Gazder, K. H., Al-Shehri, F., & Othman, W. A. M. (2023). Enhancing air quality prediction in smart cities using machine learning. *Journal of Sensors*, 2023, Article 8430704.
- Addas, A., Goldblatt, R., & Rubinyi, S. (2020). Utilizing remotely sensed observations to estimate the urban heat island effect at a local scale. *Remote Sensing*, 12(15), 2434.
- Mallick, J., Almesfer, M. K., Alsubih, M., Ahmed, M., & Ben Kahla, N. (2021). Estimating Carbon Stocks and Sequestration Potential of Mangrove Forests Using Landsat Data. *Journal of King Saud University - Science*, 33(5), 101467.
- Madugundu, R., Al-Gaadi, K. A., Tola, E., Hassaballa, A. G., & Kayad, C. (2017). Utilization of Landsat-8 data for the estimation of carrot and maize crop water footprint under the arid climate of Saudi Arabia. *PLoS ONE*, 12(2), e0192830.
- Emadi, M., Taghizadeh-Mehrjardi, R., Cherati, A., Danesh, M., Mosavi, A., & Scholten, T. (2020). Predicting and mapping of soil organic carbon using machine learning algorithms in Northern Iran. *Remote Sensing*, 12(14), 2234.
- Hammam, A., & Mohamed, E. (2020). Mapping soil salinity in the East Nile Delta using several methodological approaches of salinity assessment. *Egyptian Journal of Remote Sensing and Space Science*, 23(2), 125-131.
- Rowley, A., & Karakuş, O. (2023). Predicting air quality via multimodal AI and satellite imagery. *Remote Sensing*, 15(5), 1387.
- Tan, Y., Wang, S., Xue, R., Zhang, S., Wang, T., Liu, J., & Zhou, B. (2023). Estimation of carbon emissions in China based on XGBoost-STIRPAT model. *Environmental Science and Pollution Research*, 30, 35653-35668.
- Begum, A. M., & Al Mobin, M. (2023). A machine learning approach to carbon emissions prediction of the top emitter countries using socioeconomic indicators. *Heliyon*, 9(10), e20614.
- Han, Z., Cui, B., Xu, L., Wang, J., & Guo, Z. (2024). Coupling LSTM and CNN Neural Networks for Accurate Carbon Emission Prediction in 30 Chinese Provinces. *Ecological Informatics*, 80, 102507.

Fonseca, R., & Francis, D. (2023). Satellite derived trends and variability of CO₂ concentrations in the Middle East and North Africa. *Atmosphere*, 14(4), 712.

Xu, G., Zeng, T., Jin, H., Xu, C., & Zhang, Z. (2023). Spatio-Temporal Variations and Influencing Factors of Carbon Emissions in the Yellow River Basin Based on DMSP/OLS and NPP/VIIRS Nighttime Light Data. *Sustainability*, 15(11), 8909.