Industrial pollution control based on artificial intelligence: A synergistic model using social network analysis and machine learning

Yu-Cheng Lin , Yiling Liu[*]

Business School, Quanzhou Normal University, No. 398 Donghai Street, Quanzhou City, Fujian Province 362000, China

ABSTRACT

This research examines the impact of articulated robots (ARs), the Environmental Policy Stringency Index (EPSI), and foreign direct investment (FDI) on industrial air pollution—measured by PM2.5 levels—across 12 developed countries from 1993 to 2023. Employing Social Network Analysis (SNA) for variable selection, Granger causality tests for temporal validation, and machine learning (ML) for predictive modeling, this work captures the complex, nonlinear dynamics of pollution outcomes. The results yield three key insights. First, the EPSI consistently mitigates PM2.5 emissions, lending support to the Porter Hypothesis, which posits that stringent environmental regulations can drive innovation while reducing pollution. Second, FDI demonstrates a consistent negative effect on PM2.5 emissions, primarily by facilitating the transfer of cleaner technologies and advanced management practices. This pollution-reducing impact is particularly evident in contexts with robust regulatory frameworks, indicating that foreign investment can support environmental improvement when aligned with effective institutional oversight. Third, AR (articulated robots) consistently exhibits a negative impact on PM2.5 emissions by enhancing operational precision, improving energy efficiency, and minimizing resource waste in industrial processes. The integration of robotic automation contributes to cleaner production practices, particularly when supported by clean energy adoption and effective environmental regulations. To enhance the accuracy of PM2.5 predictions, six ML models were tested: ARIMA, SARIMA, LightGBM, XGBoost, LSTM, and GRU. Among these, the integration of SNA with LSTM achieved the highest predictive accuracy, outperforming traditional models in capturing the complex, long-term dynamics of pollution. This synergistic approach not only underscores the pivotal roles of the EPSI, FDI, and AR in pollution mitigation and but also offers a practical framework for incorporating advanced technologies into industrial pollution control approaches.

Introduction

With the rapid acceleration of global economic development and industrialization, industrial pollution has become a pressing environmental challenge, posing serious threats to human health, ecosystems, and sustainable development. According to the World Health Organization (WHO), approximately 7 million individuals die prematurely annually due to air pollution, with industrial emissions identified as a major contributor. Among the diverse pollutants, PM2.5 (particulate matter with a diameter of less than 2.5 micrometers) is particularly hazardous because of its ability to penetrate deep into the respiratory system, leading to severe health issues and premature mortality (WHO, 2023). PM2.5 is also a key component of the Air Quality Index (PM2.5), making it a widely adopted metric for assessing industrial pollution levels across countries (Makhdoomi et al., 2025). Considering its health

implications, regulatory relevance, and broad data availability, PM2.5 serves as the central indicator in this study and a focal point in global pollution control efforts.

Effectively addressing industrial pollution requires a multidimensional approach that integrates technological, regulatory, and economic factors. This research centers on three critical variables representing these dimensions: articulated robots (ARs) as a proxy for industrial automation and modernization, the Environmental Policy Stringency Index (EPSI) as a measure of regulatory enforcement, and foreign direct investment (FDI) as an economic driver correlated with technology diffusion and industrial transformation. These variables were not chosen arbitrarily; rather, they reflect the complexity of industrial pollution governance. While prior research has extensively examined the roles of economic growth, energy consumption, and industrial output, the combined influence of automation, regulatory stringency, and

international capital flows has received comparatively less attention. This research seeks to address that gap.

ARs are widely deployed in advanced manufacturing sectors to enhance productivity and mitigate per-unit emissions. Countries such as Japan and Germany have made significant investments in industrial robotics, resulting in efficiency gains and reduced pollution in high-emission sectors like steel and chemicals (International Federation of Robotics, 2024; Deng et al., 2024). Nonetheless, the adoption of robotics also increases electricity demand, which can offset environmental benefits—particularly in regions reliant on fossil fuels (Song et al., 2023). Conversely, the EPSI captures the rigor of a country's environmental regulation and its ability to shape industrial behavior. Stricter environmental policies tend to encourage cleaner production methods, investment in green technologies, and efforts toward pollution control. However, their effectiveness depends on enforcement and regulatory oversight (Deng et al., 2024). FDI adds another layer of complexity. On one hand, it can mitigate pollution by promoting technology transfer and optimizing production efficiency, particularly in developing countries. On the other hand, it can exacerbate environmental degradation in regions with weak regulatory frameworks, raising concerns about the "pollution haven" effect (Brunnermeier & Levinson, 2004).

To further support the inclusion of these variables, this research employed Social Network Analysis (SNA) to extract key determinants from a large corpus of unstructured textual data correlated with industrial pollution. Through this keyword-mining process, AR, the EPSI, and FDI emerged as central nodes in the pollution discourse network, underscoring their conceptual and empirical significance.

Building on these foundations, this research pursues two primary objectives: (1) to accurately predict PM2.5 levels (PM2.5) using advanced machine learning (ML) models and (2) to identify the key factors affecting industrial pollution, thereby providing actionable insights for policymakers and industry leaders. To capture the complex, nonlinear relationships among these variables, six ML models—ARIMA, SARIMA, LightGBM, XGBoost, LSTM, and GRU—were integrated with SNA. This hybrid framework bridges the gap between traditional environmental modeling and data-driven policy analysis, offering a novel lens through which to understand and manage industrial pollution in an era shaped by automation and globalization.

The rest of this paper is organized as follows: Section 2 reviews the literature on industrial pollution control and the roles of automation, regulation, and investment; Section 3 details the research methodology, encompassing variable selection, model comparison, and data sources; Section 4 provides the empirical results, highlighting the predictive accuracy of each model and the effects of key determinants; Section 5 presents a detailed discussion of the findings and their implications; and Section 6 concludes the study, outlines its contributions, and suggests directions for future research.

Literature Review

Current status of industrial pollution control in developed nations

In the context of the current global response to climate change and industrial pollution, developed countries have adopted a range of strategies to mitigate industrial emissions, each shaped by their distinct economic, technological, and regulatory environments (Kemp, 2000; Zhang et al., 2006; Awewomom et al., 2024). These approaches not only help lower industrial emissions but also provide valuable insights for other nations aiming to enhance their environmental governance and foster sustainable industrial development (Brunet-Jailly, 2022). This section reviews the key approaches implemented by the United States, the European Union (EU), and Japan, underscoring their main features and the lessons they offer for global pollution control efforts.

The United States has developed a multilevel pollution control framework that integrates federal oversight with state-level flexibility. This system is anchored by the Clean Air Act, a foundational piece of environmental regulation that has undergone multiple revisions since the 1970s to establish strict emission limits for major pollutants such as sulfur dioxide (SO_2) and nitrogen oxides (NO_x) (Ross et al., 2012). According to the U.S. Environmental Protection Agency (EPA) reported between 2010 and 2020, national PM2.5 concentrations declined by over 40%, while sulfur dioxide and nitrogen oxide emissions fell by 60% and 50%, respectively (EPA, 2021). These reductions are largely attributable to an integration of stringent federal standards and innovative state-level initiatives. One such initiative is California's Cap-and-Trade Program, implemented through the California Air Resources Board (CARB). This program establishes a carbon credit market that incentivizes companies to adopt cleaner technologies by allowing them to trade emission allowances (California Air Resources Board, 2021). The program has proven effective in mitigating carbon emissions while supporting economic growth, illustrating the potential of market-based mechanisms to drive environmental improvements (El-Hakim & AbouZeid, 2024; Bade & Tomomewo, 2024). The U.S. model—marked by strong national regulation coupled with flexible local implementation—serves a valuable example for other countries aiming to balance economic growth with environmental protection.

Conversely, the EU has adopted a more centralized, cross-border approach to pollution control, highlighting policy consistency and strict technical standards across its member states. A cornerstone of the EU's environmental policy is the Industrial Emissions Directive, which mandates that all large industrial facilities implement best available techniques and comply with stringent emission limits (European Commission, 2019). This directive aims to minimize pollution from industrial sources by promoting cleaner production technologies and upholding high environmental standards. According to the European Environment Agency (EEA), between 2010 and 2019, sulfur dioxide emissions across the EU decreased by approximately 75%, while nitrogen oxide emissions fell by 40% (EEA, 2020). Within this framework, member states such as Germany and France have made notable progress in mitigating industrial emissions. For instance, Germany seeks to reduce industrial carbon emissions by 55% by 2030, with a particular focus on heavy industrial regions like the Ruhr region, where stringent pollution control measures have been implemented (German Federal Environment Agency, 2021). These efforts reflect the EU's commitment to integrated, cross-border environmental governance—supported by substantial financial investments in clean technologies and robust enforcement mechanisms.

Meanwhile, Japan has pursued a distinct approach by prioritizing technological innovation as a central strategy for mitigating industrial pollution. The Japanese government has advanced this agenda through the Environmental Protection Technology Research and Development Plan, which offers financial support to high-emission industries—such as steel and automotive manufacturing—to encourage the adoption of cleaner production technologies (McKean, 2023). According to the Ministry of Economy, Trade, and Industry, emissions of nitrogen oxides and sulfur dioxide from Japan's industrial sector declined by approximately 50% and 60%, respectively, between 2020 and 2024 (Ministry of Economy, Trade and Industry, 2024). Additionally, Japan has pioneered the "carbon-neutral factory" model, which integrates renewable energy, recycled water systems, and high-efficiency manufacturing technologies to minimize environmental impact. For instance, Toyota has adopted this model to significantly reduce its carbon footprint through advanced automation and clean energy solutions. Japan also leads the world in industrial robot density, with 390 industrial robots per 10,000 workers, further optimizing manufacturing efficiency while reducing emissions (Bilgen, 2021). This strong emphasis on technological innovation has allowed Japan to achieve substantial pollution reductions in industrial pollution while maintaining its position as a globally competitive manufacturing powerhouse.

Overall, these developed countries have adopted distinct yet complementary strategies for controlling industrial pollution. The EU stands out for its policy coordination and stringent technical standards,

advancing environmental protection across member states through a cohesive regulatory framework. Conversely, the United States employs a flexible governance model that integrates federal oversight with state-level adaptability, allowing for tailored pollution control measures suited to local conditions. Meanwhile, Japan has achieved notable success through its emphasis on advanced automation technologies—particularly in the manufacturing sector—where the widespread adoption of robotics and automated systems has optimized production efficiency while mitigating emissions. Together, these approaches offer valuable lessons for other nations seeking to develop effective industrial pollution control approaches. The emphasize the significance of integrating technological innovation, regulatory consistency, and market-based mechanisms (Kemp, 2000; Heller & Shukla, 2003; Awewomom et al., 2024; Brunet-Jailly, 2022).

Application of ARs in pollution prevention and control in developed countries

With the rapid advancement of automation technology, ARs have become increasingly vital in industrial pollution control, particularly in developed countries where technological innovation is more advanced (Song et al., 2023). These robots—known for their precision, operational stability, and high manufacturing accuracy—have significantly enhanced production efficiency while effectively mitigating pollution emissions. Compared with traditional manual operations, ARs are better suited for performing complex, hazardous, and labor-intensive industrial tasks, thereby reducing direct human involvement and minimizing the environmental impact associated with manual processes (Wang et al., 2023). As technology continues to evolve, ARs have been widely adopted in high-pollution industries such as chemical manufacturing, pharmaceuticals, and metal processing. Their deployment in these sectors has contributed to substantial improvements in emission control (Wang et al., 2024).

ARs offer several key advantages in in industrial pollution control. First, they enable precise control over material usage during manufacturing, thereby reducing waste and preventing the overuse of raw materials. This precision helps minimize the generation of harmful substances such as volatile organic compounds (VOCs) and particulate matter (PM2.5), which are commonly produced during manual processes like welding, spraying, and polishing (Song et al., 2022; Zhang, 2023). For instance, traditional welding methods can emit substantial amounts of PM2.5 and VOCs, contributing to air pollution and respiratory health risks. Contrarily, robots can perform these tasks with high accuracy, significantly mitigating the release of hazardous pollutants (Popescu et al., 2024; He et al., 2024). This level of precision is particularly valuable in sectors with strict environmental regulations, where even marginal reductions in emissions can cause substantial improvements in air quality and regulatory compliance.

Moreover, ARs optimize resource efficiency and mitigate overall energy consumption, further supporting industrial pollution control efforts. Unlike traditional production equipment—which commonly consumes large amounts of energy in multi-step operations—robots can optimize energy use by precisely controlling production parameters, thereby lowering emissions of carbon dioxide and other greenhouse gases (Sękala et al., 2024; Elahi et al., 2024). For instance, Germany, a global leader in industrial automation, has widely deployed ARs in high-emission sectors such as steel and chemical manufacturing, achieving significant reductions in carbon and other harmful gas emissions (Altenburg, 2024). According to the International Federation of Robotics (IFR), global installations of industrial robots are projected to continue rising from 2019 to 2024, reflecting a broader industry shift toward automation as a strategy for mitigating emissions and optimizing production efficiency (International Federation of Robotics, 2024). This trend is especially prominent in Europe and Asia, where the integration of robotics into manufacturing has played a crucial role in controlling industrial emissions and advancing sustainability goals.

Additionally, ARs contribute to long-term cost savings by mitigating energy consumption and minimizing waste. They optimize production processes, reduce downtime, and enhance overall operational efficiency, making them a cost-effective solution for companies aiming to balance economic growth with environmental responsibility (Gadaleta et al., 2019; Javaid et al., 2022). As automation technologies continue to evolve, the capabilities of ARs are expected to expand, enabling more efficient production while further mitigating the environmental footprint of industrial operations.

In sum, the widespread adoption of ARs for industrial pollution control represents an effective approach for reducing emissions, optimizing production efficiency, and minimizing resource waste. These robots not only enhance precision and automation in manufacturing processes but also significantly reduce pollutant emissions by minimizing manual operations and optimizing resource utilization. As automation advances, ARs are poised to play an increasingly vital role in supporting the transition toward a more sustainable, low-carbon industrial economy (Gadaleta et al., 2019; Javaid et al., 2022).

Key determinants of industrial pollution and methodological innovation

Industrial pollution in developed countries is affected by several key factors, encompassing economic growth (GDP), population growth (PG), internet usage (UI), and the installation of ARs. High levels of UI, which reflects technological progress, may also result in increased energy consumption and industrial activity, thereby contributing to higher PM2.5 emissions. Both GDP and PG exhibit positive correlations with pollution, underscoring the scale effects of economic and demographic expansion. These findings are consistent with the IPAT framework (Impact = Population × Affluence × Technology), which attributes environmental degradation to the combined effects of population size, economic affluence, and technological development (Ehrlich & Holdren, 1971; Gong et al., 2023; Liu, 2023). The Environmental Kuznets Curve (EKC) further proposes that pollution rises with early economic growth but may decline as income increases and cleaner technologies are adopted (Yang et al., 2018; Huangfu & Atkinson., 2020). Nonetheless, this pattern may not hold in countries prioritizing industrialization over environmental protection (Gasimli et al., 2019).

These factors were selected according to their theoretical correlations with industrial pollution, as supported by existing literature (Yang et al., 2018; Huangfu et al., 2020; Zhao et al., 2022). Nonetheless, our empirical analysis, encompassing SNA and Granger causality testing, demonstrated that FDI, the EPSI, and ARs had the strongest and most consistent impact on pollution levels in the sampled countries. As such, these three variables constitute the primary focus of our analysis and discussion.

To better capture these interactions, this research adopts a novel three-stage analytical framework. First, SNA was used to extract key variables from large-scale textual data on industrial pollution. Through co-occurrence network mapping and centrality calculations, FDI, the EPSI, and ARs emerged as the most influential and consistently connected concepts. These three variables were subsequently tested for statistical relevance through Granger causality analysis, which confirmed their strong association with pollution levels in developed countries. Unlike previous studies that rely solely on economic output or energy use, this strategy ensures that variable selection is both data-driven and theoretically grounded. Each of these three variables represents a distinct dimension of pollution governance:

Environmental Policy Stringency Index (EPSI)
Stricter regulations play a significant role in fostering cleaner technologies and optimizing pollution abatement strategies. However, their actual effectiveness largely depends on how they are implemented and enforced. Evidence supports the idea that well-designed environmental regulations can drive firms to innovate in ways that mitigate pollution and upgrade productivity. The Porter Hypothesis posits that stringent environmental standards can incentivize businesses to invest in cleaner

technologies, resulting in the simultaneous benefits of reduced pollution and lower compliance costs (Porter & Linde, 1995; Chang & Sam, 2015).

Foreign Direct Investment (FDI)

The environmental effect of FDI is inherently dual in nature. On one hand, it facilitates the transfer of cleaner technologies and advanced management practices from multinational corporations to domestic firms, enhancing environmental performance in line with the Pollution Halo Hypothesis. On the other hand, it may encourage the relocation of pollution-intensive industries to countries with weak environmental regulations, exacerbating environmental degradation as explained by the Pollution Haven Hypothesis. This makes the regulatory context of the host country a critical factor in determining whether FDI serves as a catalyst for sustainable development or as a conduit for ecological harm (Ezeoha & Cattaneo, 2012; Gnangnon, 2022).

Articulated Robots (AR)

ARs, which account for over 50% of all annual industrial robot installations, serve as a key indicator of industrial automation. Although ARs can optimize production efficiency and mitigate emissions through precise control of manufacturing processes, their widespread adoption may also increase energy consumption—particularly in regions still dependent on fossil fuels (Song et al., 2023; Li et al., 2024). This dual effect underscores the need to balance technological advancement with energy efficiency in industrial contexts. The deployment of ARs is closely tied to the Technology Spillover Effect, whereby advanced automation reduces per-unit pollution through increased efficiency (Acemoglu & Restrepo, 2019).

To move beyond static explanations, this study introduces a methodological innovation by applying advanced ML models for PM2.5 prediction. Unlike traditional regression models, ML techniques can accommodate nonlinear relationships and dynamic feedback mechanisms, providing enhanced predictive accuracy and practical relevance (Bahmanisangesari, 2024). Specifically, we employ a hybrid approach that integrates SNA with six ML models: ARIMA, SARIMA, LightGBM, XGBoost, LSTM, and GRU. This approach enables us to identify key drivers of pollution and forecast pollution outcomes under varying scenarios, thus bridging the gap between descriptive analysis and policy-oriented forecasting.

This integrated framework—comprising data-driven variable identification (via SNA), causal verification (via Granger test), and predictive modeling (via ML)—constitutes the core innovation of this research. A summary of the key theoretical foundations and related studies is exhibited in Table 1.

Finally, we propose a hybrid model (Figure 1), which is primarily composed of two components: SNA and ML. First, key variables affecting industrial pollution are identified through big data analysis using SNA. These selected variables are then used for as inputs for six classical machine learning models in a series of comparative experiments. Industrial pollution is measured by PM2.5 levels, which serve as the output variable for the predictive models.

Methodology

The research methodology of this study begins with big data mining to obtain relevant datasets, followed by the use of SNA to visualize the

Table 1

Summary of industrial pollution study literature.

| Study	Sample	Methodology	Variables		Major findings
			Dependent variable	Major independent variables	
Boonpeng et al. (2018)	Lichen samples from the petrochemical district of Thailand	1. Lichen-air quality index (LiPM2.5) 2. Physiological measurements	Air quality levels	Distance from the industrial center	LiPM2.5 exhibited that air quality improved as the distance from the industrial site increased. Lichen effectively detected atmospheric pollutants.
Huang et al. (2019)	Chemical industrial park, China	1. Gaussian Puff Model 2. Boundary-type AQMN	Pollution detection efficiency	1. Monitor number 2. Meteorological conditions	Increasing the number of monitors improves pollution detection and source identification in industrial areas. The sparse networks restricted source accuracy, highlighting the need for higher monitor density for effective surveillance.
Kaplan et al. (2023)	Water samples from polluted industrial areas in Israel	Chemical analysis (air and water monitoring)	Water quality parameters	Industrial air pollutants	High industrial pollutants minimally impacted water quality in AWG-produced water, although nickel and dichloromethane occasionally exceeded standards.
Mărmureanu et al. (2024)	6 years of air quality data from Bucharest, Romania	1. PCA 2. Pollution data subtraction	1. PM10 2. NO$_2$ levels	1. Urban vegetation 2. Road traffic	Vegetation mitigated PM10 by 7 % and NO$_2$ by 25 % during the growing season, demonstrating vegetation's impact on pollution reduction.
Ramadan et al. (2024)	Chrome plating industry environment	1. IoT sensors 2. LSTM 3. Random Forest 4. Linear regression	Air pollutant concentration levels	IoT sensor data (NH$_3$, CO, NO$_2$, etc.)	LSTM achieved high predictive accuracy (99 % for temperature and humidity), with Random Forest effective for PM2.5 prediction. Real-time monitoring enabled proactive pollution control.
Saha et al. (2024)	South Asian cities are experiencing rapid urbanization.	1. Geographically weighted regression (GWR) 2. Remote sensing	Air quality levels (PM, CO, O$_3$)	Land use/land cover (LULC) changes	LULC changes, specifically urban expansion, increased pollutant levels. Areas with green cover exhibited a reduced pollution impact.
Liu et al. (2024)	168 cities in China (2010–2021)	1. SNA 2. Temporal exponential 3. Random Graph Model (TERGM)	Air pollution levels	1. Economic development 2. Green innovation capacity	Collaborative governance in cities can mitigate pollution through inter-regional collaboration. Pollution linkages are stronger in cities with similar levels of economic development and innovation.
Guo et al. (2024)	Rural counties in China (2011–2018)	Difference-in-differences (DID)	Air quality	Rural E-Commerce Demonstration County (REDC) policy	The REDC policy exacerbated air pollution in rural counties, particularly through secondary industry growth. Digital financial inclusion amplified these effects although environmental impacts varied by region.
Cavazzin et al. (2024)	Taranto, Italy (29 urban-industrial sites)	Biomonitoring of Mediterranean plants	Metal pollution	1. Proximity to industrial sources 2. Species type	Plants in industrial areas exhibited higher metal accumulation. Pine needles recorded the highest pollution loads, underscoring their utility for biomonitoring industrial pollutants such as Ni and Cr in urban-industrial settings.
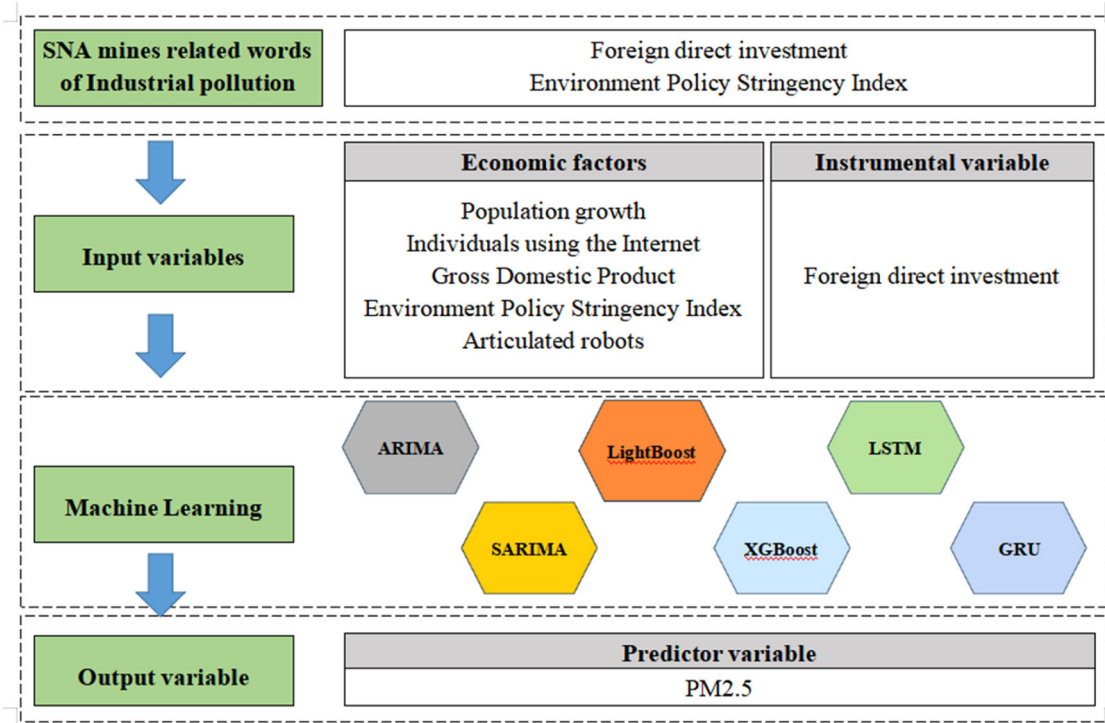
Fig. 1. Research model.

relationships between keywords and industrial pollution. The Granger causality test is then used to further verify the correlation among variables. Finally, ML models are employed to predict industrial pollution levels.

Step 1: Social Network Analysis (SNA) and Granger causality test

The SNA method offers an effective approach for analyzing large volumes of unstructured data by uncovering interrelated patterns and nonlinear associations. According to Lopes et al. (2023), SNA interprets key terms as nodes within a network, with the connections and interactions between these nodes forming semantic social ties. By exploring the structural components of the network, researchers can investigate the contexts where specific keywords are employed and understood in both general and specialized discussions (Piselli et al., 2022). Network theorists highlight that clusters or patterns—identified through term frequency, co-occurrence, and centrality—reveal the meanings embedded within texts (Tonta & Darvish, 2010; Puetz et al., 2021).

In this research, a Python library was used to compute the term frequency-inverse document frequency (TF-IDF) and degree centrality, specifically for English-language terms. These metrics were used to establish categories based on centrality and relevance, highlighting both term frequency and the strength of connections among nodes (representing words or texts) within semantic networks, as discussed by Widianto et al. (2024). Nodes that appear on the shortest paths connecting other node pairs are considered to hold prominent positions in this network. SNA effectively organizes words into networks, enabling the extraction of underlying meanings (Tabassum et al., 2018). After calculating the TF-IDF value, the top 50 words were selected to construct a co-occurrence matrix. This matrix was then analyzed using SNA to calculate central connectivity. Subsequently, CONCOR, a commonly employed clustering method in SNA, was employed to cluster the words based on context and relevance and visualized through a network. Then, the Granger causal interplay between the variables determined in the relevant literature and SNA and PM2.5 was analyzed. This part of the

analysis proved the availability and relevance of the selected variables, offering sufficient theoretical support for subsequent research. The Granger causality test can be formalized by estimating two regression models (Freeman, 1983; Guo et al., 2010):

1. Univariate model (without X):

$$PM2.5_t = \alpha_0 + \sum_{i=1}^{p} \beta_i PM2.5_{t-i} + \epsilon_t \tag{1}$$

where $PM2.5_t$ is regressed solely on its own lagged values up to lag p, and ϵ_t represents the error term.

2. Bivariate model (with X):

$$PM2.5_t = \alpha_0 + \sum_{i=1}^{p} \beta_i PM2.5_{t-i} + \sum_{j=1}^{q} \gamma_j X_{t-j} + \epsilon_t \tag{2}$$

In this model, $PM2.5_t$ is regressed on both its own past values (up to p lags) and the lagged values of X (up to q lags), where γ_j represents the coefficients of X.

Step 2: Machine learning (ML)

ML enables systems to extract insights from data autonomously. These adaptive systems are designed to model intricate relationships between input variables (X) and output variables (Y), which allows them to make predictions on new data sets (Jordan & Mitchel, 2015; Sharifani & Amini, 2023). ML is typically viewed as an extension of nonparametric statistics (Eyring et al., 2024). The various ML methods differ based on their underlying learning mechanisms, which determine how data is processed and interpreted. These approaches are grounded in differing assumptions about the nature of intelligence (Aliferis & Simon, 2024). Additionally, researchers can optimize ML training by refining datasets and adjusting parameters to enhance regression performance and improve generalization to new data. This is the core of this step and

a reference for conducting comparative experiments using the six models and finally selecting the optimal model.

Autoregressive Integrated Moving Average (ARIMA)

The Autoregressive Integrated Moving Average (ARIMA) model, developed by Box and Jenkins (1970) in the 1970s, is a robust statistical tool for time-series analysis, specifically in economic forecasting. Its structure encompasses three essential components: autoregression (AR), differencing (I), and moving average (MA). Each of these components plays a crucial role in capturing the temporal dependencies and patterns inherent in time-series data. The autoregressive (AR) component models the influence of past values on current observations—an especially important feature in economic contexts, where historical data often have a significant impact on future trends. By differentiating the data, ARIMA effectively removes trends and seasonality, thereby stabilizing the mean of the time series. This process is essential in economic forecasting, where trends can obscure the true relationships between variables (Shadab et al., 2019). Based on this structure, we construct the following model equation.

$$AQI_t = \delta + \sum_{i=1}^{p} \phi_i AQI_{t-i} + \sum_{j=1}^{q} \theta_j \epsilon_{t-j} + \epsilon_t \tag{3}$$

where:

AQI_t : the predicted Air Quality Index at time t
δ : constant term
ϕ_i : autoregressive (AR) coefficients
θ_j : moving average (MA) coefficients
ϵ_t : white noise term, assumed $\epsilon_t \sim N(0, \sigma^2)$
p, q : orders of the autoregressive and moving average components, respectively

Seasonal Autoregressive Integrated Moving Average (SARIMA)

The SARIMA model is a significant extension of the ARIMA framework, specifically designed to incorporate seasonality into time-series forecasting (Arumugam & Natarajan, 2023). The SARIMA model is grounded in the well-established Box–Jenkins methodology, which offers a systematic approach to model identification, parameter estimation, and diagnostic checking (Shabri, 2015). This method effectively preserves the characteristics of variables, particularly their seasonal components. Research has affirmed that SARIMA models outperform traditional ARIMA models in scenarios where seasonal effects are significant, as they can explicitly account for recurring seasonal patterns in the time series (Basnayake & Chandrasekara, 2022). Based on this, we construct the following model equation.

$$\Phi_P(B^m)(1 - B^m)^D \phi(B)(1 - B)^d AQI_t = \Theta_Q(B^m)\theta(B)\epsilon_t \tag{4}$$

where:

$\Phi_P(B^m)$: seasonal autoregressive polynomial with seasonal order P
B : backshift operator, defined as $BAQI_t = AQI_{t-1}$
m : seasonal period length
D, d : orders of seasonal and nonseasonal differencing, respectively
$\Theta_Q(B^m)$: seasonal moving average polynomial of seasonal order Q

Light Gradient Boosting Machine (LightBoost)

LightGBM, which was developed by Microsoft in 2017, represents a significant advancement in the ML field, particularly within the gradient boosting framework. The proposed framework is specifically designed to handle large datasets with high-dimensional features efficiently. Unlike other gradient boosting methods, such as XGBoost, LightGBM utilizes a histogram-based algorithm and a leaf-wise growth strategy. These features enhance its computational efficiency, reduce memory usage, and improve classification accuracy, making it especially well-suited for applications involving large-scale data (Inui et al., 2023; Wang et al., 2023). Due to its ability to effectively process vast amounts of

information, LightGBM has become a preferred choice in both academic research and industrial applications (Khan et al., 2024). Based on these strengths, we construct the following model equation.

$$A\widehat{QI}_i = F(x_i) = \sum_{m=1}^{M} \gamma_m h_m(x_i) \tag{5}$$

where:

$A\widehat{QI}_i$: predicted PM2.5 for the i-th sample
$F(x_i)$: ensemble model prediction
γ_m : weight for each decision tree in the ensemble
$h_m(x_i)$: prediction of the m-th decision tree for input x_i
M : total number of decision trees

Extreme Gradient Boosting (XGBoost)

XGBoost, or Extreme Gradient Boosting, is a powerful ML algorithm constructed by Chen and He (2015) that enhances the speed and performance of predictive models through gradient boosting techniques. The algorithm builds on the core principles of gradient boosting, which involves sequentially training weak learners (typically decision trees) to correct the errors of the ensemble model generated in previous iterations. Its ability to efficiently handle sparse data, combined with parallel processing capabilities, significantly enhances computational efficiency compared to traditional gradient boosting methods (Toghani & Allen, 2020). Therefore, we construct the following model equation.

$$L(\phi) = \sum_{i=1}^{n} l(A\widehat{QI}_i, AQI_i) + \sum_{k=1}^{K} \Omega(f_k) \tag{6}$$

where:

$L(\phi)$: overall loss function
$l(A\widehat{QI}_i, AQI_i)$: individual loss for each sample, commonly mean squared error (MSE) or cross-entropy
$\Omega(f_k) = \gamma T + \frac{1}{2}\lambda \| w \|^2$: regularization term to control model complexity
T: number of leaf nodes in each tree
λ: regularization parameter
w: leaf weights

Long Short-Term Memory (LSTM)

Long Short-Term Memory (LSTM) networks, introduced by Hochreiter and Schmidhuber (1997), mark a significant advancement in the field of recurrent neural networks (RNNs). They were specifically designed to overcome the challenges of learning long-term dependencies in sequential data. The core structure of an LSTM encompasses a memory cell that maintains its state over time, along with three primary gates: the input gate, the forget gate, and the output gate. These gates regulate the flow of information into and out of the memory cell, enabling the network to learn which information is relevant to retain or discard (Van Houdt et al., 2020). This capability is specifically beneficial in applications involving time-series data, where the relationships between inputs can span significant time intervals. Based on this structure, we construct the following model equation.

Forget gate:

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f) \tag{7}$$

Input gate:

$$i_t = \sigma(W_i \cdot |h_{t-1}, x_t| + b_i) \tag{8}$$

Candidate cell state:

$$\widetilde{C}_t = tanh(W_c \cdot [r_t h_{t-1}, x_t]) + b_c \tag{9}$$

Cell state update:

$$C_t = f_t * C_{t-1} + i_t * \widetilde{C}_t \tag{10}$$

Output gate:

$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o) \tag{11}$$

Hidden state:

$$h_t = o_t * tanh(C_t) \tag{12}$$

where:

x_t: input at time t

h_{t-1}: hidden state from the previous timestep

C_t: cell state

W_f, W_i, W_c, W_o: weight matrices

b_f, b_i, b_c, b_o: bias terms

σ: sigmoid activation function

Gated Recurrent Unit (GRU)

The GRU, which was introduced by Cho et al. (2014), functions as a simplified alternative to the LSTM network. Both models are designed to overcome the limitations of traditional RNNs, specifically the vanishing gradient problem that can impede learning in long sequences. The GRU achieves this by using a more streamlined architecture that incorporates only two gates: the update gate and the reset gate. This design enables the GRU to maintain performance comparable to that of LSTMs while requiring fewer computational resources, making it particularly advantageous for applications in which efficiency is critical (Al-Selwi, 2023). Based on this framework, we construct the following model equation.

Update gate:

$$z_t = \sigma(W_z \cdot [h_{t-1}, x_t]) \tag{13}$$

Reset gate:

$$r_t = \sigma(W_r \cdot [h_{t-1}, x_t]) \tag{14}$$

Candidate activation

$$\widetilde{h}_t = tanh(W \cdot [r_t * h_{t-1}, x_t]) \tag{15}$$

Final output:

$$h_t = (1 - z_t) * h_{t-1} + z_t * \widetilde{h}_t \tag{16}$$

where:

x_t: input at time t

h_{t-1}: hidden state from the previous timestep

W_z, W_r, W: weight matrices

σ: weight matrices

Variables and data resources

Economics Factors:

PG is included as a control variable to account for demographic changes that can significantly influence economic performance and environmental conditions. The annual PG rate is defined as the percentage change in the midyear population from one year to the next (Khursanaliev, 2023). PG can affect industrial production and energy consumption, thereby indirectly increasing pollution levels through greater demand for goods, services, and infrastructure.

UI measures the percentage of the population that has accessed the internet within the past three months. This variable reflects the level of digital connectivity and technological engagement in a society, which can indirectly shape industrial activity by fostering digital transformation and the growth of information-intensive industries (Perez et al., 2023). Greater internet penetration can also shift consumer behavior and business practices, potentially mitigating traditional industrial emissions through increased digitalization and the expansion of remote work.

Gross Domestic Product (GDP) is represented by the annual growth rate of GDP at market prices, serving as a key economic indicator that reflects a country's overall economic performance (Norkus et al., 2024). It captures the total market value of all goods and services produced within a country and functions as a broad measure of economic health and industrial output. Higher GDP growth is typically correlated with increased industrial activity and energy consumption, which can elevate pollution levels if not accompanied by effective environmental regulations.

The EPSI measures the strictness of a country's environmental regulations, ranging from 0 (least stringent) to 6 (most stringent) (Oate & Portney, 2003). It reflects the intensity of governmental efforts to mitigate industrial emissions and promote sustainable practices, indicating the regulatory pressures industries face to minimize their environmental impact.

AR is represented by the number of articulated robots installed in industrial settings and serves as a proxy for automation and technological advancement (Cagigas-Muñiz, 2023). These robots are known for their high versatility, precision, and efficiency, making them essential for automating complex industrial tasks and enhancing production efficiency. While their widespread adoption is associated with increased productivity, it can also contribute to higher energy consumption and carbon emissions if not supported by clean energy integration.

Instrumental Variable:

FDI is included as an instrumental variable and is calculated as the net inflows from new foreign investments minus divestments. Measured in U.S. dollars, FDI reflects the level of international economic integration and may indicate the extent of industrial activity affected by foreign investment (Udemba et al., 2024).

Predictor Variable:

PM2.5 refers to the concentration of fine particulate matter with a diameter of 2.5 micrometers or less, measured in micrograms per cubic meter ($\mu g/m^3$). Widely recognized as a critical indicator of ambient air quality, PM2.5 levels are strongly associated with emissions from industrial processes and fossil fuel combustion. As such, they are frequently used in empirical research as a proxy variable for industrial pollution (Zhou et al., 2024).

Data Source:

This study uses monthly data from 1993 to 2023 for 12 developed countries: the United States, Canada, Chile, Denmark, England, Finland, France, Germany, Japan, South Korea, Spain, and Switzerland. The dataset comprises a total of 2,604 observations, differentiated by the OECD classification. Data were sourced from the World Bank, the OECD, the IFR, and the National Bureau of Statistics. Additionally, the quadratic match sum method in Eviews 13 was applied to convert specific data into monthly frequency where needed

Results

SNA

Text Mining: TF-IDF and Degree Centrality

This paper takes industrial pollution as the central word, and finally extracts 2660 words through big data mining and data cleaning. The importance of a term increases with its frequency of occurrence in the document, but this is balanced with its occurrence rate in the entire corpus (Salton & Buckley, 1988). Therefore, we select the top 50 terms in the text by calculating TF-IDF, that is, the top 50 terms in TF-IDF value. We also calculate the centrality of each node, which represents the position of the term in the network and the degree of dependence of other terms on the term. Table 2 provides a summary of the text mining results for keywords linked to industrial pollution.

Table 3 lists 50 keywords and their TF-IDF and centrality, representing the most common and influential words related to industrial

Table 2
Result of text mining on industrial pollution of developed nations.

	Main Words (TF-IDF/Degree Centrality)
	Industrial pollution
Number of clusters	4
Average clustering coefficient	0.751
Major hub nodes	Air (441/0.05), industry (337/0.028), water (317/0.018), emission (261/0.013), investment (237/0.008), policy (202/0.008), health (188/0.01), month (146/0.006), impact (124/0.006), government (91/0.002)

Table 3
TF-IDF and Degree Centrality for industrial pollution in developed nations.

Words	TF-IDF (Rank)	Degree Centrality (Rank)	Words	TF-IDF (Rank)	Degree Centrality (Rank)
Air	441 (1)	0.05 (2)	climate	139 (26)	0.005 (29)
Industry	337 (2)	0.028 (3)	source	134 (27)	0.006 (25)
Water	317 (3)	0.018 (4)	person	133 (28)	0.006 (26)
Emission	261 (4)	0.013 (5)	India	125 (29)	0.005 (30)
Investment	237 (5)	0.008 (9)	impact	124 (30)	0.006 (27)
Control	205 (6)	0.011 (6)	EPA	124 (31)	0.004 (32)
Policy	202 (7)	0.008 (10)	resident	113 (32)	0.003 (39)
Health	188 (8)	0.01 (7)	carbon	111 (33)	0.003 (40)
Waste	181 (9)	0.006 (17)	factory	110 (34)	0.003 (41)
Pollution	178 (10)	0.103 (1)	day	109 (35)	0.005 (31)
Study	175 (11)	0.008 (11)	production	107 (36)	0.004 (33)
Plant	174 (12)	0.007 (12)	Delhi	102 (37)	0.004 (34)
Area	171 (13)	0.006 (18)	oil	101 (38)	0.004 (35)
Community	167 (14)	0.006 (19)	company	100 (39)	0.003 (42)
Pollutant	166 (15)	0.007 (13)	chemical	100 (40)	0.003 (43)
Quality	165 (16)	0.009 (8)	development	97 (41)	0.003 (44)
Plastic	159 (17)	0.006 (20)	soil	97 (42)	0.003 (45)
City	153 (18)	0.006 (21)	report	95 (43)	0.003 (46)
Environment	151 (19)	0.007 (14)	analysis	95 (44)	0.003 (47)
State	150 (20)	0.005 (28)	change	94 (45)	0.004 (36)
Risk	149 (21)	0.007 (15)	fashion	94 (46)	0.004 (37)
Level	147 (22)	0.007 (16)	problem	93 (47)	0.004 (38)
Month	146 (23)	0.006 (22)	facility	92 (48)	0.003 (48)
World	143 (24)	0.006 (23)	government	91 (49)	0.002 (50)
Gas	141 (25)	0.006 (24)	effect	90 (50)	0.003 (49)

pollution, including "air", "industry", "water", "emission", "investment", "policy", "health", "month", "impact", "government" and "exchange rate". Among them, "investment" (ranked 5th) and "policy" (ranked 7th) occupy a high position in terms of centrality, indicating that investment (FDI) and policy (EPSI) play an important role in guiding and strengthening industrial pollution control in developed countries.

CONCOR Analysis

In this analysis, we used the CONCOR clustering method in SNA. Employing UCINET software version 6.8, we constructed a co-occurrence matrix based on English terminology, specifically targeting unstructured data associated with two main challenges-investment (FDI) and policy (EPSI) (Analytic Technologies, 2023). The subsequent step involved comparing the row vectors of each node with the CONCOR matrix, using either correlation or eigenvalue analysis to interpret patterns in these relationships (Lu & Park, 2022). The CONCOR method identifies semantic clusters in a network by identifying hidden subgroups and exploring the dynamics between clusters. In this study, lexical correlations form a matrix and SNA groups terms through CONCOR analysis. This clustering method can reveal the existence and characteristics of important clusters in the network. As shown in Table 2.

CONCOR clustered 50 terms related to industrial pollution into four clusters. The average clustering coefficient was 0.751, which indicates that the nodes are closely connected. By analyzing the SNA network diagram, as shown in Figure 2, we identified 11 central hub nodes. Among them, the key nodes directly related to industrial pollution include investment (FDI) and policy (EPSI). In addition, the four terms of pollutants, carbon, waste and emissions show a strong correlation with industrial pollution, especially in developed countries. The analysis results show that "investment" (FDI) and "policy" (EPSI) have a strong correlation with industrial pollution and are important factors affecting the level of industrial pollution. Therefore, introducing FDI and EPSI into the analysis of industrial pollution can more accurately obtain the development trend of pollution.

Granger Causality Test

In this research, we applied the Granger causality test to explore the directional influence between PM2.5 and selected economic and environmental variables (Granger, 1969). Table 4 shows the results of Granger causality test. The analysis results show that AR, EPSI, FDI,
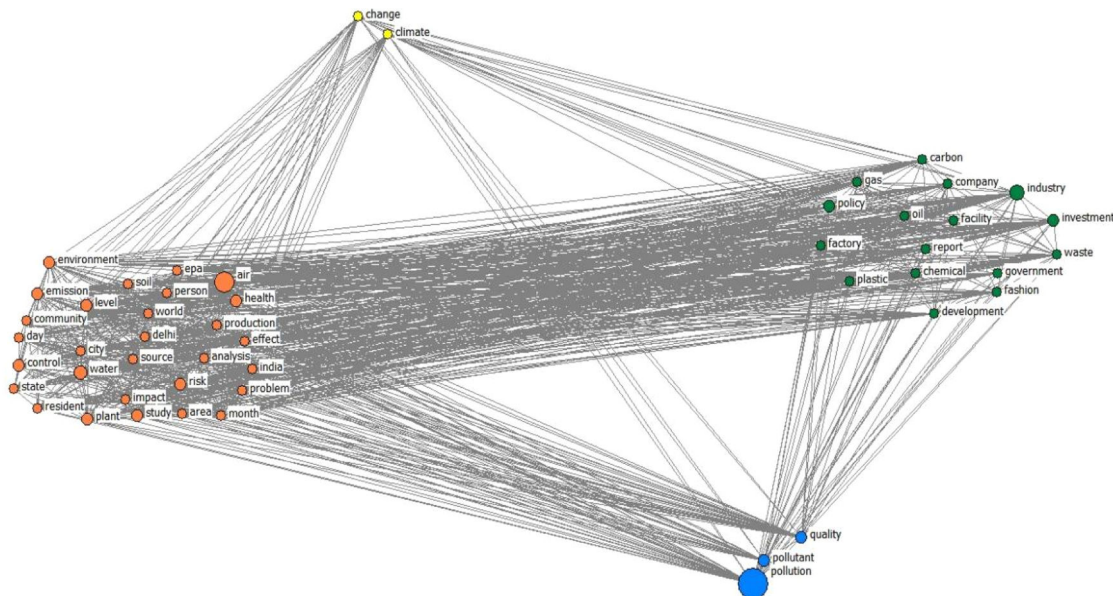
Fig. 2. SNA for industrial pollution.

Table 4
Granger causality test.

H0 Hypothesis	F-Statistic	p-Value	Decision
AR does not Granger Cause PM2.5	3.133**	0.045	Reject H0
PM2.5 does not Granger Cause AR	0.231	0.794	Cannot reject H0
EPSI does not Granger Cause PM2.5	3.909**	0.021	Reject H0
PM2.5 does not Granger Cause EPSI	2.256	0.106	Cannot reject H0
FDI does not Granger Cause PM2.5	8.055***	0.000	Reject H0
PM2.5 does not Granger Cause FDI	0.656	0.520	Cannot reject H0
GDP does not Granger Cause PM2.5	10.486***	0.000	Reject H0
PM2.5 does not Granger Cause GDP	0.562	0.571	Cannot reject H0
PG does not Granger Cause PM2.5	4.177**	0.016	Reject H0
PM2.5 does not Granger Cause PG	1.408	0.246	Cannot reject H0
UI does not Granger Cause PM2.5	3.116**	0.046	Reject H0
PM2.5 does not Granger Cause UI	1.470	0.231	Cannot reject H0

Note: The symbols ***, **, and * indicate significance at the 1 %, 5 %, and 10 % levels, respectively.

GDP, PG, and UI all have a Granger relationship with PM2.5. For example, at the 5% significance level, AR and EPSI have a Granger relationship with PM2.5, with F-statistics of 3.133 and 3.909, respectively, and corresponding p-values of 0.045 and 0.021, respectively. At the 1% significance level, FDI and GDP exhibit stronger causal effects on PM2.5, with F-statistics of 8.055 and 10.486, respectively, and p-values of 0.000, emphasizing their large impact. Furthermore, PG and UI are PM2.5 significant at the 5% threshold with F-statistics of 4.177 and 3.116, respectively, and p-values of 0.016 and 0.046, respectively.

Machine Learning (ML)

Heatmap of Pearson Correlation Coefficients

This correlation analysis reveals significant associations between multiple variables, providing robust support for the study (Gogtay & Thatte, 2017). As shown in Figure 8, the SHAP values of UI are mainly in the negative range, indicating that higher UI values tend to reduce the predicted values. The SHAP values of EPSI are mostly positive, with the maximum effect close to 0.15, indicating that stricter environmental policies have a positive impact on the model's predictions. The SHAP values of GDP and PG range from about -0.1 to 0.1, showing a mixed impact of positive and negative effects, indicating that their relationship with the predicted results is complex. AR and FDI have relatively small effects and high SHAP values. Meanwhile, the low correlations of GDP and FDI with PM2.5 (0.035 and 0.29, respectively) indicate that short-term economic growth and investment may have limited direct effects on air quality, prompting further exploration of other factors with more immediate impacts on PM2.5 (Figure 3).

Evaluate the Performance of ARIMA, SARIMA, LightBoost, XGBoost, LSTM and GRU (10-fold Time-Derived Split)

The culmination of our research lies in deriving final model estimations, which provide a quantitative basis for our economic hypotheses while shedding light on the economic phenomena's underlying dynamics (Sullivan, 2022). Table 5 and Figure 4 presents the indices used to evaluate the results of the SARIMA, LightGBM, XGBoost, LSTM, and GRU models, offering a comparative perspective on their predictive
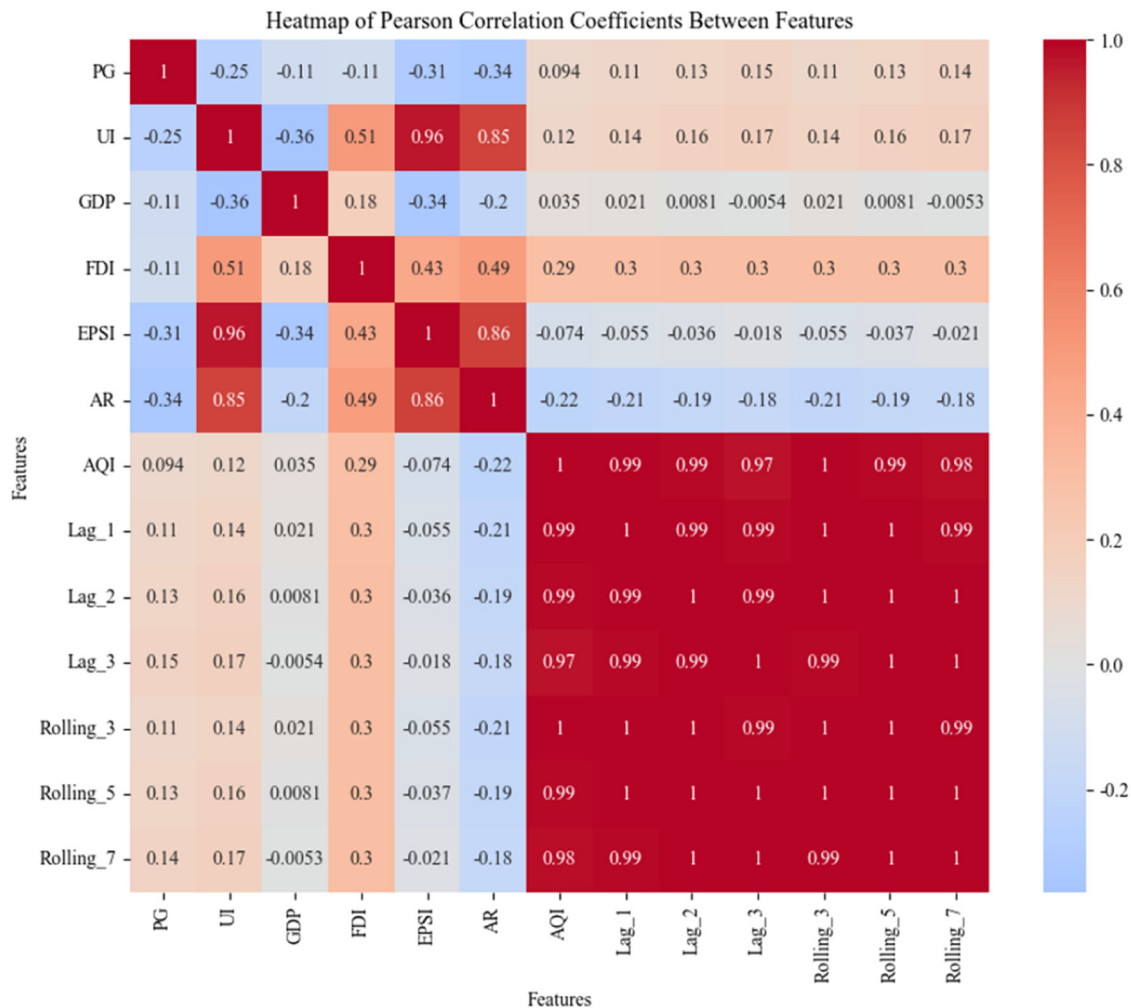
Fig. 3. Pearson Correlation Coefficients.

Table 5
Indices of model.

	ARIMA	SARIMA	LightBoost	XGBoost	LSTM	GRU
Train R^2	0.8452	0.8505	0.9901	0.9999	0.9998	0.9998
Test R^2	0.9819	0.9651	0.5166	0.6257	0.9808	0.9369
MSE	0.0128	0.0248	0.3117	0.2413	0.0123	0.0406
RMSE	0.1132	0.1576	0.5583	0.4912	0.111	0.2016
MAE	0.0801	0.1018	0.3609	0.3363	0.0818	0.1614
SMAPE	0.8182	1.0228	3.8106	3.5445	0.4332	0.8477

accuracy and performance. This table presents the performance of ARIMA, SARIMA, LightGBM, XGBoost, LSTM, and GRU models across various evaluation metrics, including R^2, Mean Squared Error (MSE), Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), and Symmetric Mean Absolute Percentage Error (SMAPE), on both training and testing sets. The results indicate that the LSTM model performs the best across multiple metrics, achieving R^2 values of 0.9998 on the training set and 0.9808 on the testing set, along with the lowest MSE (0.0123), RMSE (0.1110), MAE (0.0818), and SMAPE (0.4333), reflecting its minimal prediction error and strong generalization ability. In contrast, XGBoost and LightGBM exhibit a notable drop in R^2 on the testing set (0.6257 and 0.5166), coupled with higher MSE and SMAPE values, suggesting potential overfitting issues. The GRU model, while not outperforming LSTM, still demonstrates competitive performance across several metrics. Overall, LSTM shows superior accuracy and stability in time series forecasting for this study, providing empirical support for model selection.

The comparison plot of actual versus predicted PM2.5 values on the test dataset highlights the model's strong ability to accurately capture both general trends and finer variations. With PM2.5 values ranging between approximately 8.0 and 11.5, the predicted outcomes closely match the observed data throughout the test period. In the early phase (samples 0–10), the model successfully tracks a minor decrease in PM2.5 to around 9.0 and its subsequent recovery. Between samples 10 and 30, the PM2.5 steadily rises, peaking at approximately 11.5, and the model reflects this trend with minimal deviations. At the sharp apex near sample 30, the model shows a slight delay in response but maintains overall alignment with the actual values. During the rapid decline from 11.5 to 8.0 (samples 30–40), the model continues to perform effectively,

with errors consistently remaining below 0.2. These results demonstrate the model's robustness in handling both gradual PM2.5 changes and abrupt transitions, reinforcing its reliability and precision in forecasting PM2.5 fluctuations (Figure 5).

Figure 6 presents the residuals of predictions for the LSTM model on the test set, representing the difference between the predicted and actual values. The vertical axis shows the residual values, while the horizontal axis represents the index. Observing the figure, most residuals are distributed around zero and fluctuate within the range of -0.15 to 0.25. Overall, the residual values exhibit no apparent systematic pattern or trend, indicating that the prediction errors are relatively random and that there is no significant systematic bias. Most residuals are concentrated between -0.05 and 0.05, demonstrating the LSTM model's high accuracy in predicting the majority of samples.

SHapley Additive exPlanations (SHAP)

SHAP analysis is a tool used to explain machine learning model predictions, particularly suited for understanding the impact of various features on the target variable. From an economist's perspective, SHAP analysis provides transparency when interpreting complex models,
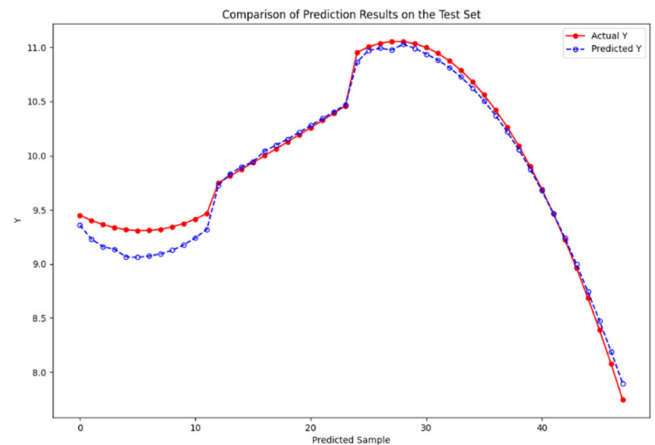
Fig. 5. Actual vs. predicted PM2.5-LSTM.
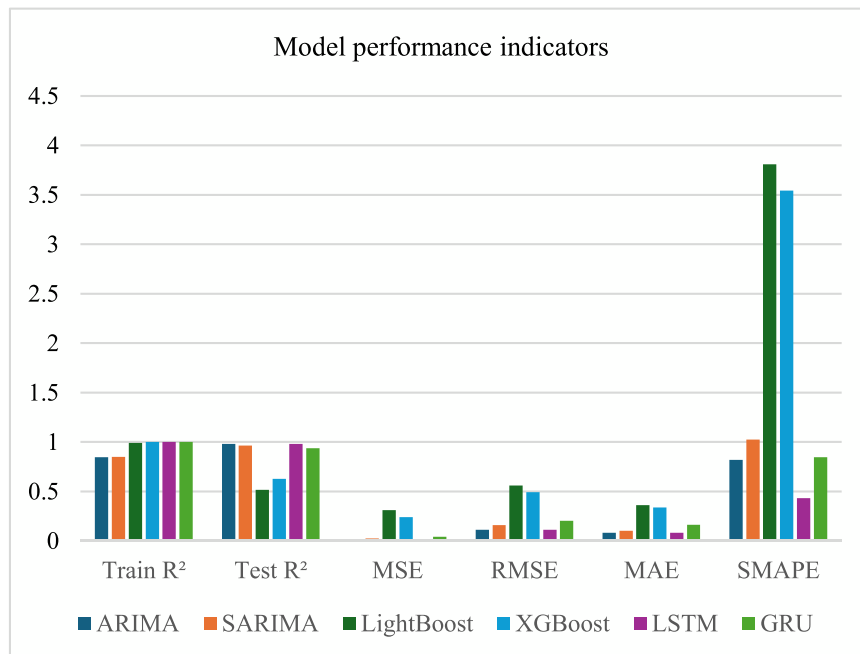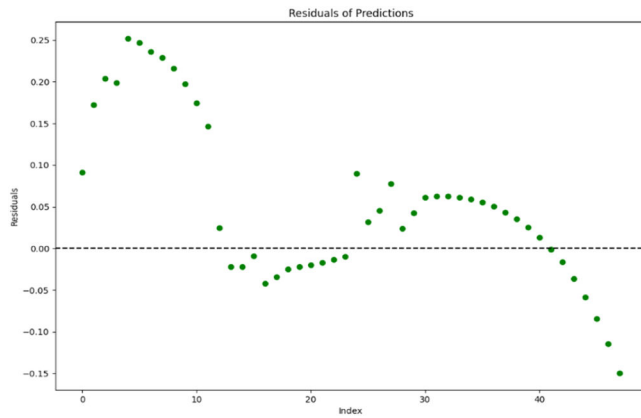
Fig. 4. Model performance indicators.

Fig. 6. Residuals of predictions-LSTM.

helping to uncover the decision-making process of the model and enhancing insights into economic phenomena (Parsa et al., 2020; Cakiroglu et al., 2024).

From Figure 7, UI shows a significant influence on the model output, with SHAP values primarily in the negative range, indicating that higher UI values tend to decrease the prediction, with SHAP values extending to around -0.15. EPSI, on the other hand, exhibits mostly positive SHAP values, with a maximum impact close to 0.15. GDP and PG have SHAP values ranging from approximately -0.1 to 0.1, showing a mixed influence with both positive and negative effects, indicating the complexity of their relationship with the predicted outcomes. AR and FDI have relatively minor impacts, with SHAP values.

Figure 8 shows the normalized SHAP values for each feature in the LSTM model, providing a clear view of the relative importance of each variable in influencing the model's output. The horizontal axis represents the normalized SHAP values, where higher absolute values indicate a greater impact on the model predictions. From the chart, UI stands out as the most influential feature, with a normalized SHAP value of 1.642, indicating that this variable has the strongest effect on the model's output. EPSI follows with a negative normalized SHAP value of -0.926, suggesting a substantial but inverse effect, where higher policy stringency tends to reduce the model's output. GDP and PG have moderate impacts, with SHAP values of 0.337 and 0.263, respectively, showing a positive relationship with the output. In contrast, FDI and AR have minimal influence, with SHAP values of -0.138 and -0.178.

Discussion

This research explores the interplay between ARs, the EPSI, FDI, and industrial pollution (PM2.5) in 12 developed countries from 1993 to 2023. These variables were identified through SNA as the most central and influential in the discourse on industrial pollution. Their integration into the ML model and interpretation through Granger causality testing and SHAP values enhances the understanding of PM2.5 dynamics by uncovering complex, nonlinear, and context-dependent effects often obscured in traditional models.

The SHAP summary plot clearly indicates that articulated robots (ARs) exert a stable and negative influence on PM2.5 emissions. At all levels of deployment, ARs are associated with negative SHAP values, implying that the use of robotic automation consistently contributes to pollution reduction. This effect is largely attributed to improved process accuracy, enhanced energy efficiency, and reduced resource waste in industrial operations. The findings affirm the environmental benefits of automation technologies, suggesting that increased AR adoption, when aligned with supportive regulatory and energy systems, can serve as a reliable means of mitigating industrial air pollution in developed economies.

The EPSI consistently shows a negative association with PM2.5 emissions across all feature values. Notably, higher EPSI values produce stronger negative SHAP impacts, highlighting the effectiveness of stringent and well-enforced environmental regulations.

FDI also demonstrates a predominantly negative SHAP contribution, particularly at higher feature values, implying that increased foreign capital inflows tend to mitigate PM2.5 emissions through mechanisms such as technology transfer, modern production techniques, and enhanced environmental management.

Among the remaining economic variables, UI exhibits the most consistently positive SHAP values, specifically at higher levels. This suggests that digital expansion—while associated with knowledge dissemination and operational efficiency— may also correspond with increased energy demand and emissions. This outcome may reveal the environmental cost of data centers, digital infrastructure, and increased industrial productivity driven by digitalization. GDP and PG also display positive contributions to PM2.5, revealing the scale effect where economic expansion and demographic increases naturally raise environmental pressure. These results conceptualize environmental degradation as the outcome of interacting demographic and economic forces (Gong et al., 2023; Liu, 2023).
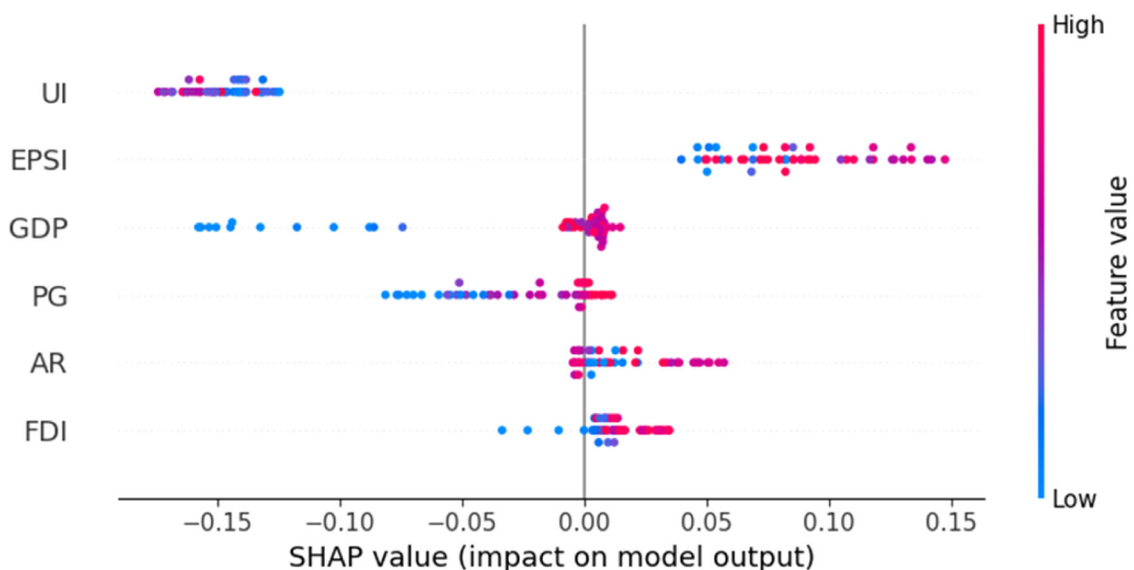

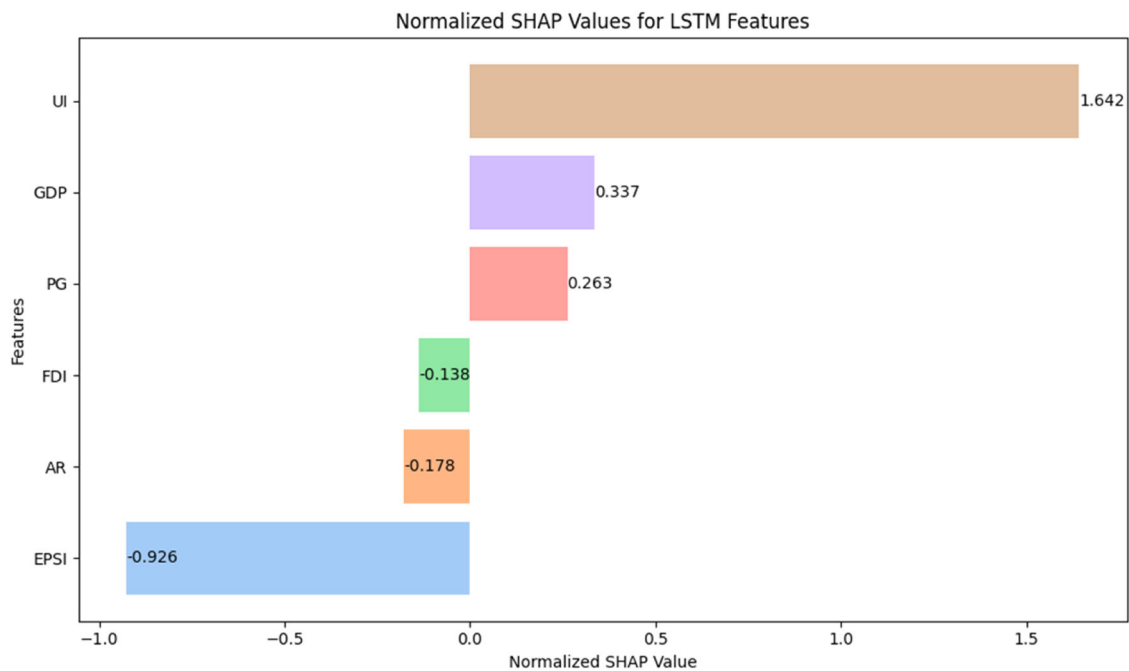**Fig. 7.** SHAP scatter plot.

Fig. 8. Histogram of Normalized SHAP values.

Together, these results verify the methodological value of integrating SNA for variable identification, Granger testing for temporal causality, and SHAP interpretation for explaining ML predictions. This multi-method strategy enhances the accuracy of industrial pollution forecasting and reveals the underlying structural roles that economic variables play in influencing environmental outcomes. Among the six models employed, the LSTM model demonstrated the highest predictive performance, reaching an R^2 of 0.9808 on the testing set along with the lowest MSE, RMSE, and MAE values. The LSTM model's outperformance not only validates its applicability in environmental forecasting and demonstrates the policy relevance of advanced deep learning strategies in industrial pollution management. These findings highlight the significance of integrating technological, regulatory, and economic heterogeneity into analytical frameworks for designing effective pollution control strategies in advanced economies.

Conclusion

This research advances the theoretical understanding of industrial pollution by integrating multiple environmental–economic frameworks within a unified empirical model. The consistent negative impact of the EPSI on PM2.5 emissions supports the Porter Hypothesis, which argues that well-designed environmental regulations stimulate innovation and pollution reduction. The consistent pollution-reducing effect of FDI supports the Pollution Halo Hypothesis, which posits that foreign investment can lead to environmental improvement through the transfer of cleaner technologies and sustainable management practices. This relationship underscores the importance of institutional quality in amplifying the positive environmental outcomes associated with FDI. The consistently positive impact of AR on environmental quality reflects the Composition and Technique effects outlined in the Scale-Composition-Technique framework. It also refines the IPAT model by demonstrating how automation contributes to emission reductions through improved efficiency, cleaner industrial processes, and the shift toward low-emission production systems. Moreover, the EKC framework proposes that pollution tends to rise in the early stages of industrialization, peaking before declining as economies mature and invest in cleaner technologies (Yang et al., 2018; Huangfu et al., 2020). Nonetheless, in countries where industrial output is heavily prioritized, the

EKC turning point may be delayed or absent, leading to sustained pollution over time (Gasimli et al., 2019). By introducing these theories into conversation with data-driven approaches, this work presents a comprehensive framework for exploring and addressing the drivers of industrial pollution in advanced economies.

The empirical findings validate that the EPSI, FDI, and ARs all play significant roles in influencing industrial pollution outcomes, providing valuable implications for policymakers and industry stakeholders.

Governments in developed countries should strengthen environmental regulations through robust policies such as carbon pricing, emission trading systems, and mandatory adoption of pollution control technologies. These measures not only mitigate emissions but also foster innovation, which is consistent with the Porter Hypothesis. To further support cleaner production, financial incentives—such as tax credits, green subsidies, and preferential loans—should be directed toward firms that implement energy-efficient and low-emission technologies. Moreover, public investment in green R&D is essential to lower the cost of clean energy solutions and accelerate their widespread adoption across industries.

Given the consistently positive environmental impact of articulated robots (ARs) through enhanced production efficiency and emission reduction, coordinated action between policymakers and industry is essential. Governments should support the integration of renewable energy into automated systems by implementing measures such as renewable energy mandates, green energy certificates, and strategic infrastructure investment. In parallel, firms adopting automation technologies are encouraged to conduct lifecycle assessments of robotic systems and actively transition to clean energy sources, thereby maximizing the environmental benefits of industrial automation.

Moreover, foreign direct investment (FDI) consistently contributes to environmental improvement in host countries by facilitating the transfer of advanced technologies, cleaner production methods, and sustainable management practices. This cross-border dynamic enables the diffusion of higher environmental standards and promotes industrial upgrading, particularly in economies seeking to modernize their manufacturing base. These outcomes support the Pollution Halo Hypothesis, highlighting FDI's role in enhancing environmental performance and fostering long-term sustainable development.

From an industry perspective, companies should adopt a holistic

strategy that aligns technological advancements with sustainability objectives. Integrating ARs and digital tools—such as IoT, big data analytics, and real-time monitoring—can mitigate waste and improve energy use. Nonetheless, companies must complement these gains with circular economy practices—such as recycling, material reuse, and resource efficiency—to ensure long-term environmental benefits. This approach aligns with the Dematerialization Hypothesis, which affirms that digitalization can help decouple economic growth from environmental degradation (Weber & Matthews, 2008).

This research offers a data-driven framework linking industrial automation, environmental policy, and foreign investment to pollution control, but it encounters limitations due to incomplete reporting of variables such as the EPSI and ARs across countries. Future studies should expand the dataset to include developing economies, enabling more comprehensive cross-national comparisons. Incorporating dynamic system modeling and advanced SNA approaches can further enhance policy insights by capturing complex interdependencies. Additionally, deeper exploration of sector-specific impacts and the roles of renewable energy and waste management will strengthen our understanding and support the development of more effective sustainability strategies.

Funding statement

Ethic statement

Not applicable.

CRediT authorship contribution statement

Yu-Cheng Lin: Writing – original draft, Data curation, Conceptualization. **Yiling Liu:** Writing – review & editing, Software, Resources.

Declaration of interest statement

The authors declare that no competing interests exist.

Data availability

All relevant data are available from the authors on request.

References

Acemoglu, D., & Restrepo, P. (2019). Automation and new tasks: How technology displaces and reinstates labor. *Journal of economic perspectives, 33*(2), 3–30. https://doi.org/10.1257/jep.33.2.3

Aliferis, C., & Simon, G. (2024). Foundations and Properties of AI/ML Systems. *Artificial intelligence and machine learning in health care and medical sciences: best practices and pitfalls* (pp. 33–94). Cham: Springer International Publishing. https://doi.org/10.1007/978-3-031-39355-6_2

Altenburg, T. (2024). Industrial policy trends in Germany (No. 80836). Naciones Unidas Comisión Económica para América Latina y el Caribe (CEPAL). https://repositorio.cepal.org/server/api/core/bitstreams/75d226ef-36a3-40fc-8f19-c2f417cbecc5/content.

Arumugam, V., & Natarajan, V. (2023). Time series modeling and forecasting using autoregressive integrated moving average and seasonal autoregressive integrated moving average models. *Instrumentation Mesure Métrologie, 22*(4), 161–168. https://doi.org/10.18280/i2m.220404

Awewomom, J., Dzeble, F., Takyi, Y. D., Ashie, W. B., Ettey, E. N. Y. O., & Afua, P. E. (2024). Addressing global environmental pollution using environmental control techniques: a focus on environmental policy and preventive environmental management. *Discover Environment, 2*(1), 8. https://doi.org/10.1007/s44274-024-00033-5

Bade, S. O., & Tomomewo, O. S. (2024). A review of governance strategies, policy measures, and regulatory framework for hydrogen energy in the United States. *International Journal of Hydrogen Energy, 78*, 1363–1381. https://doi.org/10.1016/j.ijhydene.2024.06.338

Bahmanisangesari, S. (2024). A novel transformer-based multi-step approach for predicting common vulnerability severity score. https://unbscholar.lib.unb.ca/handle/1882/38052.

Basnayake, B. R. P. M., & Chandrasekara, N. V. (2022). Forecasting exchange rates in sri lanka: a comparison of the double seasonal autoregressive integrated moving average models (dsarima) and sarima models. *Journal of Science of the University of Kelaniya, 15*(2), 192–209. https://doi.org/10.4038/josuk.v15i2.8067

Bilgen, H. (2021). A global comparison methodology to determine critical requirements for achieving industry 4.0. Technological Forecasting and Social Change, 172, 121036. https://doi.org/10.1016/j.techfore.2021.121036.

Boonpeng, C., Srivivoon, C., Polyiam, W., Sangiamdee, D., Watthana, S., & Boonpragob, K. (2018). Assessing atmospheric pollution in a petrochemical industrial district using a lichen-air quality index (LiPM2.5). *Ecological Indicators, 95*(1), 589–594. https://doi.org/10.1016/j.ecolind.2018.08.012

Box, G., & Jenkins, G. (1970). *Time series analysis-forecasting and control, 553* p. 1970). San Francisco: Holden Day.

Brunet-Jailly, E. (2022). Cross-border cooperation: a global overview. *Alternatives, 47*(1), 3–17. https://doi.org/10.1177/03043754211073463

Brunnermeier, S. B., & Levinson, A. (2004). Examining the evidence on environmental regulations and industry location. *The Journal of Environment & Development, 13*(1), 6–41. https://doi.org/10.1177/1070496503256500

Cagigas-Muñiz, D. (2023). Artificial Neural Networks for inverse kinematics problem in articulated robots. *Engineering Applications of Artificial Intelligence, 126*, Article 107175. https://doi.org/10.1016/j.engappai.2023.107175

Cakiroglu, C., Demir, S., Ozdemir, M. H., Aylak, B. L., Sariisik, G., & Abualigah, L. (2024). Data-driven interpretable ensemble learning methods for the prediction of wind turbine power incorporating SHAP analysis. *Expert Systems with Applications, 237*, Article 121464. https://doi.org/10.1016/j.eswa.2023.121464

California Air Resources Board. (2021). File an Air Pollution Complaint. Retrieved from https://ww2.arb.ca.gov/.

Cavazzin, B., MacDonell, C., Green, N., & Rothwell, J. J. (2024). Air pollution biomonitoring in an urban-industrial setting (Taranto, Italy) using Mediterranean plant species. *Atmospheric Pollution Research, 15*(6), Article 102105. https://doi.org/10.1016/j.apr.2024.102105

Chang, C. H., & Sam, A. G. (2015). Corporate environmentalism and environmental innovation. *Journal of environmental management, 153*, 84–92. https://doi.org/10.1016/j.jenvman.2015.01.010

Chen, T., & He, T. (2015). Higgs boson discovery with boosted trees. *NIPS 2014 workshop on high-energy physics and machine learning* (pp. 69–80). PMLR. https://proceedings.mlr.press/v42/chen14.html.

Cho, K., Van Merriënboer, B., Gulcehre, C., Bahdanau, D., Bougares, F., Schwenk, H., & Bengio, Y. (2014). Learning phrase representations using RNN encoder-decoder for statistical machine translation. arXiv preprint https://doi.org/10.48550/arXiv.1406.1078.

Deng, L., Plümpe, V., & Stegmaier, J. (2024). Robot adoption at German plants. *Jahrbücher für Nationalökonomie und Statistik, 244*(3), 201–235. https://doi.org/10.1515/jbnst-2022-0073

Ehrlich, P. R., & Holdren, J. P. (1971). Impact of Population Growth: Complacency concerning this component of man's predicament is unjustified and counterproductive. *Science, 171*(3977), 1212–1217. https://doi.org/10.1126/science.171.3977.121

Elahi, E., Li, G., Han, X., Zhu, W., Liu, Y., Cheng, A., & Yang, Y. (2024). Decoupling livestock and poultry pollution emissions from industrial development: A step towards reducing environmental emissions. *Journal of Environmental Management, 350*, Article 119654. https://doi.org/10.1016/j.jenvman.2023.119654

El-Hakim, Y., & AbouZeid, M. N. (2024). Towards Mitigating Climate Change Negative Impact: The Role of Regulations and Governance in the Construction Industry. *Sustainability, 16*(16), 6822. https://doi.org/10.3390/su16166822

European Environment Agency. (2020). Analysis and data. Retrieved from https://www.eea.europa.eu/en/analysis.

Eyring, V., Collins, W. D., Gentine, P., Barnes, E. A., Barreiro, M., Beucler, T., … Zanna, L. (2024). Pushing the frontiers in climate modelling and analysis with machine learning. *Nature Climate Change, 14*, 916–928. https://doi.org/10.1038/s41558-024-02095-y

Ezeoha, A. E., & Cattaneo, N. (2012). FDI flows to sub-Saharan Africa: The impact of finance, institutions, and natural resource endowment. *Comparative Economic Studies, 54*, 597–632.

Freeman, J. R. (1983). Granger causality and the times series analysis of political relationships. *American Journal of Political Science, 27*(2), 327–358. https://doi.org/10.2307/2111021

Gadaleta, M., Pellicciari, M., & Berselli, G. (2019). Optimization of the energy consumption of industrial robots for automatic code generation. *Robotics and Computer-Integrated Manufacturing, 57*, 452–464. https://doi.org/10.1016/j.rcim.2018.12.020

Gasimli, O., Haq, I.u., Gamage, S. K. N., Shihadeh, F., Rajapakshe, P. S. K., & Shafiq, M. Z. (2019). Energy, trade, urbanization and environmental degradation nexus in Sri Lanka: bounds testing approach. *Energies, 12*(9), 1655. https://doi.org/10.3390/en12091655

German Federal Environment Agency. (2021). Data on the Environment. Retrieved from https://www.umweltbundesamt.de/en/data.

Gnangnon, S. K. (2022). Does poverty deter foreign direct investment flows to developing countries? *International Journal of Economic Policy Studies, 16*(1), 297–330.

Gogtay, N. J., & Thatte, U. M. (2017). Principles of correlation analysis. *Journal of the Association of Physicians of India, 65*(3), 78–81. https://www.kem.edu/wp-content/uploads/2012/06/9-Principles_of_correlation-1.pdf.

Gong, C., Yang, X., Tan, H., & Lu, X. (2023). Industrial robots, economic growth, and sustainable development in an aging society. *Sustainability, 15*(5), 4590. https://doi.org/10.3390/su15054590

Granger, C. W. (1969). Investigating causal relations by econometric models and cross-spectral methods. *The Econometric Society, 37*(3), 424–438. https://doi.org/10.2307/1912791

Guo, A., Wu, G., Miao, Z., & Chen, X. (2024). E-commerce and air pollution: Evidence from China. *Journal of Cleaner Production, 480*, Article 144030. https://doi.org/10.1016/j.jclepro.2024.144030

Guo, S., Ladroue, C., & Feng, J. (2010). Granger Causality: Theory and Applications. In J. Feng, W. Fu, & F. Sun (Eds.), *Frontiers in computational and systems biology. computational biology* (pp. 83–111). London: Springer. https://doi.org/10.1007/978-1-84996-196-7_5.

He, X., Teng, R., Feng, D., & Gai, J. (2024). Industrial robots and pollution: Evidence from Chinese enterprises. *Economic Analysis and Policy, 82*, 629–650. https://doi.org/10.1016/j.eap.2024.03.001

Heller, T. C., & Shukla, P. R. (2003). Development and climate: Engaging developing countries. *Beyond Kyoto: Advancing the international effort against climate change*, 111–140. https://citeseerx.ist.psu.edu/document?repid=rep1&type=pdf&doi=992b4e8c4538bb974868ab12100d112cc7d28425.

Hochreiter, S., & Schmidhuber, J. (1997). Flat minima. *Neural Computation, 9*(1), 1–42. https://doi.org/10.1162/neco.1997.9.1.1

Huang, Z., Yu, Q., Ma, W., & Chen, L. (2019). Surveillance efficiency evaluation of air quality monitoring networks for air pollution episodes in industrial parks: Pollution detection and source identification. *Atmospheric Environment, 215*, Article 116874. https://doi.org/10.1016/j.atmosenv.2019.116874

Huangfu, P., & Atkinson, R. (2020). Long-term exposure to NO2 and O3 and all-cause and respiratory mortality: A systematic review and meta-analysis. *Environment international, 144*, Article 105998. https://doi.org/10.1016/j.envint.2020.105998

Huangfu, Z., Hu, H., Xie, N., Zhu, Y., Chen, H., & Wang, Y. (2020). The heterogeneous influence of economic growth on environmental pollution: evidence from municipal data of china. *Petroleum Science, 17*(4), 1180–1193. https://doi.org/10.1007/s12182-020-00459-5

International Federation of Robotics. (2024). World Robotics 2023 Report. Retrieved from https://ifr.org/.

Inui, A., Nishimoto, H., Mifune, Y., Yoshikawa, T., Shinohara, I., Furukawa, T., … Kuroda, R. (2023). Screening for osteoporosis from blood test data in elderly women using a machine learning approach. *Bioengineering, 10*(3), 277. https://doi.org/10.3390/bioengineering10030277

Javaid, M., Haleem, A., Singh, R. P., Suman, R., & Gonzalez, E. S. (2022). Understanding the adoption of Industry 4.0 technologies in improving environmental sustainability. *Sustainable Operations and Computers, 3*, 203–217. https://doi.org/10.1016/j.susoc.2022.01.008

Jordan, M. I., & Mitchell, T. M. (2015). Machine learning: Trends, perspectives, and prospects. *Science, 349*(6245), 255–260. https://doi.org/10.1126/science.aaa8415

Kaplan, A., Ronen-Eliraz, G., Ratner, S., Aviv, Y., Wolanov, Y., & Avisar, D. (2023). Impact of industrial air pollution on the quality of atmospheric water production. *Environmental Pollution, 325*, Article 121447. https://doi.org/10.1016/j.envpol.2023.121447

Kemp, R. (2000). Technology and Environmental Policy-Innovation effects of past policies and suggestions for improvement. *Innovation and the Environment, 1*, 35–61.

Khan, M. A., Asad, B., Vaimann, T., & Kallaste, A. (2024). Improved diagnostic approach for brb detection and classification in inverter-driven induction motors employing sparse stacked autoencoder (ssae) and lightgbm. *Electronics, 13*(7), 1292. https://doi.org/10.3390/electronics13071292

Khursanaliev, B. (2023). The impact of population growth on the country's economic development. *Qo 'qon Universiteti Xabarnomasi, 1*, 8–11. https://doi.org/10.54613/ku.v6i6.236

Li, Q., Zhang, J., Feng, Y., Sun, R., & Hu, J. (2024). Towards a high-energy efficiency world: Assessing the impact of artificial intelligence on urban energy efficiency. *Journal of Cleaner Production, 461*, Article 142593. https://doi.org/10.1016/j.jclepro.2024.142593

Liu, J., Wang, R., Tian, Y., & Zhang, M. (2024). The driving mechanisms of industrial air pollution spatial correlation networks: A case study of 168 Chinese cities. *Journal of Cleaner Production, 470*, Article 143255. https://doi.org/10.1016/j.jclepro.2024.143255

Liu, Y. (2023). Impact of industrial robots on environmental pollution: evidence from China. *Scientific Reports, 13*(1), Article 20769. https://doi.org/10.1038/s41598-023-47380-6

Lopes, T., Ströele, V., Braga, R., David, J. M. N., & Bauer, M. (2023). A broad approach to expert detection using syntactic and semantic social networks analysis in the context of Global Software Development. *Journal of Computational Science, 66*, Article 101928. https://doi.org/10.1016/j.jocs.2022.101928

Lu, Y., & Park, S. D. (2022). Time series analysis of policy discourse on green consumption in China: Text mining and network analysis. *Sustainability, 14*(22), Article 14704. https://doi.org/10.3390/su142214704

Makhdoomi, A., Sarkhosh, M., & Ziaei, S. (2025). PM2. 5 concentration prediction using machine learning algorithms: an approach to virtual monitoring stations. *Scientific Reports, 15*(1), 8076. https://doi.org/10.1038/s41598-025-92019-3

Mărmureanu, L., Leca, Ș., Pitar, D., Pascu, I., De Marco, A., Sicard, P., … Badea, O. (2024). Estimation of plant pollution removal capacity based on intensive air quality measurements. *Environmental Research, 261*, Article 119703. https://doi.org/10.1016/j.envres.2024.119703

McKean, M. (2023). *Environmental protest and citizen politics in japan*. Univ of California Press. https://books.google.nl/books?id=Z2LhEAAAQBAJ&dq=60.McKean,+M.+(2023).+Environmental+protest+and+citizen+politics+in+Japan.+Univ+of+California+Press.&lr=&hl=zh-CN&source=gbs_navlinks_s.

Ministry of Economy, Trade and Industry. (2024). Technical Manual on Nitrogen Oxide Emission Reduction Countermeasures. Retrieved from https://www.meti.go.jp/.

Norkus, Z., Markevičiūtė, J., Grytten, O., Šiliņš, J., & Klimantas, A. (2024). Benchmarking Latvia's economy: a new estimate of gross domestic product in the 1930s. *Cliometrica, 18*(1), 251–325. https://doi.org/10.1007/s11698-022-00260-x

Oates, W. E., & Portney, P. R. (2003). The political economy of environmental policy. *Handbook of Environmental Economics, 1*, 325–354. https://doi.org/10.1016/S1574-0099(03)01013-1

Parsa, A. B., Movahedi, A., Taghipour, H., Derrible, S., & Mohammadian, A. K. (2020). Toward safer highways, application of XGBoost and SHAP for real-time accident detection and feature analysis. *Accident Analysis & Prevention, 136*, Article 105405. https://doi.org/10.1016/j.aap.2019.105405

Perez, A. J., Siddiqui, F., Zeadally, S., & Lane, D. (2023). A review of IoT systems to enable independence for the elderly and disabled individuals. *Internet of Things, 21*, Article 100653. https://doi.org/10.1016/j.iot.2022.100653

Piselli, C., Colladon, A. F., Segneri, L., & Pisello, A. L. (2022). Evaluating and improving social awareness of energy communities through semantic network analysis of online news. *Renewable and Sustainable Energy Reviews, 167*, Article 112792. https://doi.org/10.1016/j.rser.2022.112792

Popescu, C. A., Olteanu, S. C., Ifrim, A. M., Petcu, C., Silvestru, C. I., & Ilie, D. M. (2024). The Influence of Energy Consumption and the Environmental Impact of Electronic Components on the Structures of Mobile Robots Used in Logistics. *Sustainability, 16*(19), 8396. https://doi.org/10.3390/su16198396

Porter, M. E., & Linde, C. V. D. (1995). Toward a new conception of the environment-competitiveness relationship. *Journal of economic perspectives, 9*(4), 97–118. https://doi.org/10.1257/jep.9.4.97

Puetz, K., Davis, A. P., & Kinney, A. B. (2021). Meaning structures in the world polity: A semantic network analysis of human rights terminology in the world's peace agreements. *Poetics, 88*, Article 101598. https://doi.org/10.1016/j.poetic.2021.101598

Ramadan, M. N., Ali, M. A., Khoo, S. Y., Alkhedher, M., & Alherbawi, M. (2024). Real-time IoT-powered AI system for monitoring and forecasting of air pollution in industrial environment. *Ecotoxicology and Environmental Safety, 283*, Article 116856. https://doi.org/10.1016/j.ecoenv.2024.116856

Ross, K., Chmiel, J. F., & Ferkol, T. (2012). The impact of the clean air act. *The Journal of pediatrics, 161*(5), 781–786. https://doi.org/10.1016/j.jpeds.2012.06.064

Saha, M., Al Kafy, A., Bakshi, A., Nath, H., Alsulamy, S., Rahaman, Z. A., & Saroar, M. (2024). The urban air quality nexus: Assessing the interplay of land cover change and air pollution in emerging South Asian cities. *Environmental Pollution, 361*, Article 124877. https://doi.org/10.1016/j.envpol.2024.124877

Salton, G., & Buckley, C. (1988). Term-weighting approaches in automatic text retrieval. *Information processing & management, 24*(5), 513–523. https://doi.org/10.1016/0306-4573(88)90021-0

Sękala, A., Blaszczyk, T., Foit, K., & Kost, G. (2024). Selected Issues, Methods, and Trends in the Energy Consumption of Industrial Robots. *Energies, 17*(3), 641. https://doi.org/10.3390/en17030641

Shabri, A. (2015). Least square support vector machines as an alternative method in seasonal time series forecasting. *Applied Mathematical Sciences, 9*, 6207–6216. https://doi.org/10.12988/ams.2015.58525

Shadab, A., Said, S., & Ahmad, S. (2019). Box–Jenkins multiplicative ARIMA modeling for prediction of solar radiation: a case study. *International Journal of Energy and Water Resources, 3*, 305–318. https://doi.org/10.1007/s42108-019-00037-5

Sharifani, K., & Amini, M. (2023). Machine learning and deep learning: A review of methods and applications. *World Information Technology and Engineering Journal, 10*(07), 3897–3904. https://ssrn.com/abstract=4458723.

Song, J., Chen, Y., & Luan, F. (2023). Air pollution, water pollution, and robots: Is technology the panacea. *Journal of Environmental Management, 330*, Article 117170. https://doi.org/10.1016/j.jenvman.2022.117170

Song, J., Wang, Y., & Wang, J. (2022). The impact of so2 emissions trading scheme on firm's environmental performance: a channel from robot application. *International Journal of Environmental Research and Public Health, 19*(24), Article 16471. https://doi.org/10.3390/ijerph192416471

Sullivan, E. (2022). Understanding from machine learning models. *The British Journal for the Philosophy of Science, 73*(1). https://doi.org/10.1093/bjps/axz035

Tabassum, S., Pereira, F. S., Fernandes, S., & Gama, J. (2018). Social network analysis: An overview. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 8*(5), e1256. https://doi.org/10.1002/widm.1256

Toghani, M. T., & Allen, G. I. (2020). MP-Boost: Minipatch Boosting via Adaptive Feature and Observation Sampling. In *, 2021. 2021 IEEE International Conference on Big Data and Smart Computing (BigComp)* (pp. 75–78). https://doi.org/10.1109/BigComp51126.2021.00023. Jeju Island, Korea (South),.

Tonta, Y., & Darvish, H. R. (2010). Diffusion of latent semantic analysis as a research tool: A social network analysis approach. *Journal of Informetrics, 4*(2), 166–174. https://doi.org/10.1016/j.joi.2009.11.003

Udemba, E. N., Dagar, V., Peng, X., & Dagher, L. (2024). Attaining environmental sustainability amidst the interacting forces of natural resource rent and foreign direct investment: Is Norway any different. *OPEC Energy Review, 48*(1), 19–35. https://doi.org/10.1111/opec.12292

Van Houdt, G., Mosquera, C., & Nápoles, G. (2020). A review on the long short-term memory model. *Artificial Intelligence Review, 53*(8), 5929–5955. https://doi.org/10.1007/s10462-020-09838-1

Wang, W., Wang, J., Ye, H., & Wu, H. (2024). Polluted air, smarter factories? China's robot imports shed light on a potential link. *Energy Economics, 134*, Article 107621. https://doi.org/10.1016/j.eneco.2024.107621

Wang, X., Liu, M., Liu, C., Ling, L., & Zhang, X. (2023). Data-driven and Knowledge-based predictive maintenance method for industrial robots for the production stability of intelligent manufacturing. *Expert Systems with Applications, 234*, Article 121136. https://doi.org/10.1016/j.eswa.2023.121136

Weber, C. L., & Matthews, H. S. (2008). Quantifying the global and distributional aspects of American household carbon footprint. *Ecological economics, 66*(2-3), 379–391. https://doi.org/10.1016/j.ecolecon.2007.09.021

Widianto, A., Pebriyanto, E., Fitriyanti, F., & Marna, M. (2024). Document Similarity Using Term Frequency-Inverse Document Frequency Representation and Cosine Similarity. *Journal of Dinda: Data Science, Information Technology, and Data Analytics, 4*(2), 149–153. https://doi.org/10.20895/dinda.v4i2.1589

Yang, N., Zhang, Z., Xue, B., Ma, J., Chen, X., & Lu, C. (2018). Economic growth and pollution emission in China: structural path analysis. *Sustainability, 10*(7), 2569. https://doi.org/10.3390/su10072569

Zhang, C., Shukla, P. R., Victor, D. G., Heller, T. C., Biswas, D., & Nag, T. (2006). Baselines for carbon emissions in the Indian and Chinese power sectors: Implications for international carbon trading. *Energy Policy, 34*(14), 1900–1917. https://doi.org/10.1016/j.enpol.2005.01.009

Zhang, L. (2023). Carbon emission performance of robot application: influencing mechanisms and heterogeneity characteristics. *Discrete Dynamics in Nature and Society, 2023*, 1–18. https://doi.org/10.1155/2023/4380575

Zhao, H., Cheng, Y., & Zheng, R. (2022). Impact of the digital economy on pm2.5: experience from the middle and lower reaches of the yellow river basin. *International Journal of Environmental Research and Public Health, 19*(24), Article 17094. https://doi.org/10.3390/ijerph192417094

Zhou, S., Wang, W., Zhu, L., Qiao, Q., & Kang, Y. (2024). Deep-learning architecture for PM2. 5 concentration prediction: A review. *Environmental Science and Ecotechnology, 21*, Article 100400. https://doi.org/10.1016/j.ese.2024.100400