International Journal of Research in Management & Social Science
Volume 14, Issue 2: April - June 2026

ISSN 2322 - 0899

PREDICTIVE AI MODELS FOR CARBON EMISSION FORECASTING IN DIGITALLY TRANSFORMED ENTERPRISES

[1]Mrs. Anjali Dandekar and [2]Dr. Suvarna Dhaamdhere
[1]Assistant Professor, MES's D.G. Ruparel College of Arts, Science and Commerce
[2]MKSSS's Smt. Hiraben Nanavati Institute of Management and Research for Women (HNIMR)

ABSTRACT
This paper studies how prediction models based on AI can help companies to forecast their carbon emissions using data that is already publicly available. This study leverages sustainability disclosures, government emission records, and prior research to examine how organisations manage their carbon footprints instead of creating fresh datasets. It evaluates methods spanning simple regression to complex time-series models, and assesses how effectively they work in practice. The outcome indicate that there is significant influence of AI tools is on corporate environmental decision-making, regulatory compliance, and sustainability planning. The paper integrates evidence across various disciplines and proposes a practical framework to help transforming companies digitally to forecast emissions with greater reliability.

Keywords: *Carbon Emissions, Predictive AI, Secondary Data, Sustainability Analytics, Digital Transformation*

1. INTRODUCTION
Climate change is not just a scientific debate or a distant political problem but it is something that businesses are being held responsible for right now, and their accountability is only growing. Governments, investors, customers, and communities are all asking organizations to be clearer about how much carbon they produce and what they are doing to reduce it. One of the most useful things an organization can do in this space is to get good at predicting its own emissions, because without reliable forecasts, it is nearly impossible to set meaningful targets or make smart decisions about reducing environmental impact.

Over a long time, forecasting emissions was genuinely hard. Organizations did not always have the data they required, and even when they did, making sense of it was very difficult. But now the situation has changed considerably. The tremendous transformation of digital technology that has swept through most industries over the past few decades has produced enormous amounts of data as a by-product. Sensors on equipment, cloud-based management platforms, real-time energy tracking systems, and integrated supply chain software all instantiate continuous streams of information. A lot of this information is directly useful in understanding and predicting carbon emissions.

A practical approach is taken in this study by working with data that is already out there, rather than trying to collect new data. There is already a lot of valuable material available in government databases, annual sustainability reports, and academic studies that has not been fully put to use for this purpose. The aim is to review what forecasting tools based on AI are available, compare their performance, and propose a framework that businesses can actually use. It also helps to connect the dots between sustainability, digital technology, and data science, which are usually discussed separately but belongs together.

The remaining of the paper is laid out as follows. The key background concepts around carbon emissions, digital transformation, and AI models are covered in Section 2. Section 3 contains what researchers have already found. The research methodology is explained in Section 4. Section 5 has comparison between different AI approaches. Section 6 discusses the important takeaways, and a practical framework for enterprises is proposed in Section 7. Sections 8 and 9 address the challenges faced in real-world and also the directions for future work, and Section 10 concludes things off with a summary of conclusions.

2. CONCEPTUAL BACKGROUND

2.1 Carbon Emissions in Enterprises
All carbon emissions are not the same, and knowing the difference really matters a lot for companies trying to get control on their environmental impact. They can be grouped into three scopes. Scope 1 covers the direct emissions from sources the company owns or controls, such as fuel burned in company vehicles or machinery and so on. Scope 2 is about the indirect emissions. They are linked to the energy a company buys, mainly electricity or heating. Scope 3 is broader and covers everything else along the value chain such as emissions generated by suppliers all the way through to those produced when customers use or dispose of the company's products.

 ISSN 2322 - 0899

Companies are consistently required to measure and report all three of these scopes. There are well-established frameworks to help them do it to a greater extent. The most widely used are ESG reporting architecture and the Global Reporting Initiative, or GRI. These standards are most important as the companies can be compared against each other. A consistent basis for evaluation is provided to investors, regulators, and the public. For companies, getting this right is most important than compliance. It leads to real implications for companies' reputation, access to capital, and ability to attract and retain customers who care about sustainability.

2.2 Sustainability and Digital Transformation

Digital transformation basically means that businesses use technology to operate more intelligently and efficiently. This practically means things like connecting factory machines and infrastructure through the Internet of Things, using integrated software platforms that help manage everything from procurement to finance, moving computing resources to the cloud, and using data analytics to make better and more effective decisions faster. All of these technologies generate data as a side effect. Most of that data is directly relevant to understand a company's energy use and environmental impact.

Studies have found a consistent relationship between how digitally mature a company is and how well it performs on sustainability metrics. This link makes intuitive sense. A company with good systems for tracking energy consumption, logistics movements, and production processes already has most of the data it needs to monitor its carbon footprint. It is also better placed than a less digitally developed organisation to take the next step and deploy AI tools for forecasting, because it already has the infrastructure, the data pipelines, and the technical talent to make those tools work.

2.3 Predictive AI Models

A wide spectrum is covered by AI models used for forecasting carbon emissions. These tools are from various categories - relatively simple to highly complex and sophisticated. These models can be categorised in three sections: traditional statistical models, models based on machine learning approaches and deep learning models that are more advanced. The traditional statistical models are simpler tools. The historical data is analysed to find patterns and to make predictions by these tools. The machine learning approaches like Random Forests and Gradient Boosting, can recognize the complex relationships in large datasets which simpler tools may not detect. And the last category is deep learning models such as Long Short-Term Memory networks and transformer-based architectures. Extended sequences of data and pick up on subtle, long-running patterns can be processed by them.

Choosing the right model does not mean choosing the most accurate one that is available. The selection of a model also depends on various other elements such as what type of data the company possesses, the company's operations complexity, and the technical resources that can actually be used. A smaller company with limited historical data and no data science team needs a very different approach from a large multinational with years of detailed operational records and a fully staffed analytics function. Choosing the suitable tool according to the situation is very significant.

3. LITERATURE REVIEW (SECONDARY DATA ANALYSIS)

For past few decades, the study on AI and machine learning for carbon emission forecasting has increased considerably, and there is now a useful evidence is available. Earlier work in this domain was based on mainly traditional statistical methods like ARIMA models. When emission trends were relatively steady, they worked reasonably well. But they could not be effective when conditions changed quickly or when there are complex and non-linear patterns in the data. Researchers started incorporating machine learning tools as they were easier to use, widely available. Also they were giving better results, specially when emission data was combined with other variables like economic activity, energy prices, or temperature.

In comparative studies, the Gradient boosting methods, especially XGBoost and LightGBM, have repeatedly come out as strong performers. They are valued partly because they manage to be both accurate and computationally efficient, which matters when working with large datasets. For tasks involving long sequences of historical data, deep learning models, particularly LSTM networks, have produced some of the strongest results in the literature. They are especially effective at capturing patterns that only emerge when you look at data over extended time periods.

These studies use the data sources that span quite a range. Some research scholars had worked with national-level data from agencies such as the US Environmental Protection Agency or the European Environment Agency. The datasets from international organisations such as the International Energy Agency or the World Bank is used by others. The Carbon Disclosure Project has become an important resource at the company level.

 ISSN 2322 - 0899

Also, sustainability reports published by large corporations and aggregated databases that compile environmental performance data across multiple organisations has been a significant resource.

Inspite all of this progress, there are some significant gaps that still stand out. Often the largest portion of a company's total footprint is Scope 3 emissions. Surprisingly, it has received remarkably little attention from the perspective of forecasting. Most existing models have been developed and tested in specific industries or regions. Their findings can not be applied elsewhere as they are industry specific or region specific. There is also very little work on incorporating real-time data from IoT systems into emission forecasting pipelines, which seems like an obvious and valuable direction. And across the board, the issue of how to make these models interpretable and explainable to non-specialists has not been adequately resolved.

Below is a **clear, academic-ready Global vs India comparison table** focused on **Predictive AI Models for Carbon Emission Forecasting in Digitally Transformed Enterprises**.

Dimension	Global Scenario	India Scenario
Total GHG emissions	**~51–53 $GtCO_2e$ annually** (global emissions continue to rise, 2023–2024 levels) [indiaai.gov.in]	**~2.96 $GtCO_2e$** (excluding LULUCF, latest official baseline) [globaleffi…yintel.com]
Global ranking	Largest aggregate emissions worldwide	**3rd largest emitter globally**, but low per-capita emissions [globaleffi…yintel.com]
Net-zero target	Varies by country (EU 2050, USA 2050, China 2060)	**2070 net-zero commitment** (India) [globaleffi…yintel.com]
Enterprises using AI (any function)	**~88–90%** of enterprises globally use AI in at least one function [documents1…ldbank.org]	**~64%** of Indian enterprises actively use AI for sustainability use cases [tmcnet.com]
AI maturity level	Widespread pilots; ~65% still not fully scaled across enterprises [documents1…ldbank.org]	Rapid adoption, but largely **pilot-to-early scale** phase [tmcnet.com]
Share of Scope 3 emissions	**70–90%** of total corporate emissions globally [beatcarbon.com], [link.springer.com]	**70–98%** in several Indian sectors (IT, FMCG, auto, services) [energydigital.com]
Scope 3 disclosure rate	~**29%** overall; **48%** for large-cap firms globally [anthropic.com]	Mandatory only via **BRSR**; quality and coverage still uneven [tmcnet.com]
Primary need for predictive AI	Scenario planning, regulatory compliance, supply-chain forecasting	Supply-chain estimation, data gap-filling, rapid scaling forecasts
Accuracy of AI forecasting models	**~95–97% prediction accuracy** in ML-based carbon forecasts [indiaai.gov.in]	Error rates reduced to **~1.7–3% MAPE** using hybrid AI models [youtube.com]
Dominant AI model types	ML, Deep Learning, Hybrid ML-Physics, Federated Learning [github.com]	Time-series + ML hybrids, Physics-Informed Neural Networks (PINNs) [youtube.com]
Digital infrastructure emissions	Data centers consume **~460 TWh (2025)**; projected ~1,000 TWh by 2030 [indiacsrsummit.in], [indiaai.gov.in]	IT & digital sector contributes **~35–50 $MtCO_2e$ (~1–1.5%)** [iea.org]
Regulatory drivers	CSRD, ISSB, California Climate Acts, SBTi	BRSR, BRSR-Core assurance, CBAM exposure
Climate-tech ecosystem size	US$40B+ global climate-tech investment annually	**800+ climate-tech startups**, US$3.6B+ invested (2014–2024) [isg-one.com]
Role of predictive AI	Strategic risk forecasting and compliance automation	Operational forecasting and ESG capability building
Key challenge	Translating AI pilots into enterprise-wide carbon decision systems	Data quality, supplier emissions visibility, SME integration

4. RESEARCH METHODOLOGY

4.1 Research Design

This research study takes a review-based approach. It blends qualitative analysis of conceptual ideas with a more structured comparison of the performance of different models in existing research. The objective is not to produce original empirical results but to take a careful look at what has already been done, figure out what the evidence actually shows, and translate that into something useful for both researchers and practitioners. This kind of synthesis is particularly valuable when research on a topic is spread across many different fields and hard to navigate as a whole.

The decision to rely on secondary data rather than primary data collection was deliberate. Drawing on a much wider pool of evidence is possible due to working with the already existing study rather than any single original study could provide. It also makes it possible to inspect the findings that have varied across different industries, countries, and time periods. The quality of the conclusions depends partly on the quality of the underlying sources, which are not always consistent, is the real challenge.

4.2 Data Sources

The main inputs for this study came from peer-reviewed journal articles and conference papers found through databases including Scopus, Web of Science, and Google Scholar. Corporate sustainability reports from large companies were also examined, with particular attention to those that described using data and technology in their environmental management. Public emission datasets from government bodies and international organisations were used to provide context for the model comparisons. The introspection of industry reports from technology firms and consulting organisations was also done for practical perspectives on usage of these tools in the real world.

4.3 Data Screening and Selection Criteria

To keep the review manageable and relevant, the focus was on studies published between 2010 and 2024 that used AI or machine learning methods to forecast carbon or greenhouse gas emissions, and that drew on data from companies, sectors, or national economies. Studies were excluded if they dealt only with household or individual-level emissions, did not explain their methods clearly enough to allow meaningful comparison, or were published without independent review. Priority was given to work that had clear relevance for how businesses actually operate, rather than purely theoretical or laboratory-based research.

5. COMPARATIVE ANALYSIS OF PREDICTIVE AI MODELS

Within the studies that were analysed, the models based on AI used for emission forecasting can be classified into three groups: statistical models, machine learning models, and deep learning models. All the three groups have their significant role to play. On the basis of context, the best choice depends heavily.

The most established tools are statistical models like ARIMA in this space. They are relatively simple to set up, do not require huge amounts of historical data, and produce results that are easy to explain to people without a technical background. These qualities make them a practical first option for many organisations. Their main weakness is that they tend to struggle when emission patterns are driven by many overlapping factors or when conditions shift unexpectedly.

Machine learning models represent a meaningful step up in terms of capability. Random Forest, Gradient Boosting, and Support Vector Regression are the approaches that handle complex, multi-dimensional data. Also, they cope better with missing values and unusual observations. But the problem with these models is that they can be harder to interpret. To understand why a particular prediction was made is not always straightforward. In the context where companies need to explain and justify their forecasts to external parties, this is where it matters. Methods like SHAP values help, but the transparency challenge persists in such case.

Some of the best results in recent research have been achieved by Deep learning models, particularly LSTM networks and their variants. They excel at extracting patterns from long time-series datasets, making them well-suited to situations where historical context matters a lot. However, they are demanding in terms of data, require significant computing resources, and take real expertise to train and validate properly. They are a sensible choice for large, technically capable organisations, but probably not the right starting point for most companies.

The practical conclusion is that no single model fits all situations. Organisations that are just getting started with AI for sustainability would be better served by simpler, more interpretable approaches that can be implemented quickly and understood by a broad audience. As capabilities grow and data accumulates, it becomes more worthwhile to explore more powerful tools. A step-by-step approach, moving up in complexity as readiness allows, is likely to serve most companies better than trying to leap straight to the most sophisticated option available.

6. DISCUSSION

While looking at all the evidence, a few themes can be seen clearly. To begin with, models based on AI consistently demonstrate reliable performance as compared to traditional statistical tools when they are judged on accuracy, provided there is substantial data. The result is as expected. Looking at various studies and contexts, it validates the consistent support. Efficient resource allocation, more informed investment decisions, and more trustworthy commitments towards sustainability are relatively modest improvements in forecasting accuracy.

How much digital transformation matters as a foundation is the second theme. Companies that have invested in modern technology systems are at a better position to capitalize on forecasting tools that are AI-based. It is not just about having the right model but it is also about having the right data, the right way of collecting it, flowing through reliable systems. Most sophisticated AI approaches will struggle to deliver useful results without that infrastructure.

Third, the practical value of accurate emission forecasting is real and growing. Companies that know what their emissions are likely to be in the coming months or years can make much smarter decisions about energy purchasing, production planning, logistics, and capital investment. They are also in a much stronger position when it comes to meeting disclosure requirements, which are becoming more demanding and more detailed as sustainability reporting frameworks continue to evolve.

For managers, the main message here is that investing in data infrastructure and AI capability is a strategic move, not just a technical one. For those responsible for policy, the evidence supports the case for clearer, more consistent emission reporting standards and for practical support to help smaller companies that want to adopt these tools but lack the resources to do so independently.

7. CONCEPTUAL FRAMEWORK PROPOSAL

Drawing on the insights from this review, it is possible to sketch out a practical framework that enterprises can use to build their own AI-driven emission forecasting capability. The framework is organised around three connected layers, each of which depends on the one before it.

Getting the right data is the job of the first layer. This means collection of information from various sources. The examples of sources can be emission figures from sustainability reports and regulatory submissions, real-time operational data from sensors and software systems, external variables like energy prices and weather patterns, and supply chain data. At this stage, making sure all of this information is clean, consistent, and comparable is the key challenge. Data that has been collected in different ways or reported against different standards needs to be standardised before it can be used reliably. Skipping this step is one of the most common reasons AI forecasting projects fail in practice.

The second layer is the forecasting engine itself. The framework recommends a tiered approach, starting with simpler models and introducing more complexity as the organisation's data and capabilities develop. Whatever models are used, it is important to build in some form of explainability so that the outputs make sense to managers, auditors, and regulators, not just data scientists. Predictions that are lacking interpretability or can’t be explained are of limited use for effective decision-making or compliance purposes.

Generating the useful outputs is the task at third layer. Generating forecasts across various time horizons, running scenario analyses to check how emissions might change under different business conditions, and producing reports that can feed directly into sustainability disclosure architecture are some of the examples. Ideally, the company's performance dashboards include all of this, so that the forecasts are visible in real time and can drive effective and accurate decisions instead of just writing down in the reports.

8. CHALLENGES AND LIMITATIONS

Like any research study, this one comes with limitations that are worth being upfront about. The most fundamental is data quality. AI models can only ever be as good as the information they are fed, and the publicly available emission data that forms the backbone of secondary research is not always consistent or reliable. Companies report differently, cover different activities, and use different measurement approaches, which can create inconsistencies that are hard to correct for after the fact.

Getting access to detailed company-level data is a persistent challenge in this space. While there is plenty of sector-level and national data around, the granular, facility-level information that would allow the most precise forecasting is rarely made public. Companies understandably share summary figures rather than detailed operational data, which limits what is possible with secondary sources alone.

Reporting standards also vary considerably from one country or sector to another. Even organisations nominally following the same framework can interpret requirements quite differently, which creates noise in the data and makes it harder to build models that generalise well. Progress is being made on international harmonisation of reporting standards, but there is still a long way to go.

Finally, because this study does not use primary data, it is not possible to actually test whether the proposed framework works in practice. The conclusions here are grounded in existing research rather than real-world deployment, and what works on paper does not always translate cleanly to a specific company's operations. That gap between conceptual framework and practical validation is the most important limitation to acknowledge.

9. FUTURE RESEARCH DIRECTIONS

There is a lot of productive ground still to cover in this area. The most important next step is to test the ideas put forward in this paper against real company data, either by collecting it directly from organisations or through research partnerships that provide access to operational systems. This would allow the framework to be validated and refined in a way that is simply not possible with secondary data alone. It would also open the door to developing more targeted models for specific industries, since the emission drivers and data characteristics of a chemical manufacturer are quite different from those of a logistics company or a bank.

Another valuable direction is to explore how AI forecasting tools can be more tightly integrated with formal sustainability reporting processes. There is genuine demand from companies for tools that do not just predict emissions but also help produce the reports they need for regulators and investors. Closing that gap between technical forecasting and regulatory reporting would be practically very useful, and it is not something the research community has addressed in much depth yet.

The explainability question also deserves dedicated attention. As AI models play a bigger role in supporting sustainability disclosures, the ability to explain and justify those models to non-technical audiences will become increasingly important. Research into which explainability techniques work best in this context, and how to present their outputs clearly to managers and auditors, would be a valuable practical contribution.

Looking further ahead, real-time data from IoT networks and satellite monitoring systems offers exciting possibilities. The idea of models that update themselves continuously as new data flows in, rather than being retrained from scratch at intervals, could offer real improvements in accuracy and timeliness. Working out how to build and deploy these kinds of adaptive systems within existing enterprise technology environments is a promising and largely open research challenge.

10. CONCLUSION

This paper has reviewed a broad range of research on AI-based carbon emission forecasting with the aim of drawing out what is known, what works, and where the opportunities lie. The overall message is a positive one. Whether organisations use straightforward machine learning models or more sophisticated deep learning approaches, AI tools consistently offer meaningful improvements over traditional forecasting methods. But the technology on its own is not enough. It needs to sit on a foundation of good data, which in turn depends on the kind of digital infrastructure that only comes with sustained investment in modern technology systems.

The framework proposed here is intended to give companies a practical starting point rather than an abstract blueprint. Start with what you have, make sure the data is solid, keep the outputs interpretable, and build from there is the simply main idea. Getting those fundamentals right matters far more than jumping straight to the most sophisticated model available.

For the research community, this paper adds value by drawing together evidence from different disciplines and making it accessible in one place. For practitioners, the key takeaway is that investing in data and AI for sustainability is increasingly a strategic necessity, not an optional extra. The companies that build this capability now will be much better placed to manage what are likely to be growing demands for transparency, accountability, and performance in the years ahead.

REFERENCES

[1] J. Luo, W. Zhuo, S. Liu, and B. Xu, "The optimization of carbon emission prediction in low carbon energy economy under big data," IEEE Access, vol. 12, pp. 14690–14702, 2024, doi: 10.1109/ACCESS.2024.3356418.

[2] X. Wang, R. Li, and C. Zhou, "A survey of machine learning methods for carbon emission prediction," IEEE Access, vol. 11, pp. 23610–23628, 2023.

[3] S. M. J. Jalali, S. Ahmadian, and A. Khosravi, "A novel evolutionary-based deep convolutional neural network model for intelligent load forecasting," IEEE Trans. Ind. Inform., vol. 17, no. 12, pp. 8243–8253, Dec. 2021, doi: 10.1109/TII.2021.3065718.

[4] Y. Chen, Q. Liu, and H. Zhang, "PSO-CNN-LSTM hybrid model for carbon emission forecasting in smart energy cities," in Proc. IEEE Int. Conf. Smart Energy Systems and Technologies (SEST), 2022, pp. 1–6, doi: 10.1109/SEST53650.2022.9898461.

[5] Y. Zhang, G. Chen, and J. Li, "TCN-LSTM with attention mechanism for time-series carbon emission prediction," Journal of Inner Mongolia University (Natural Science Edition), vol. 53, no. 5, pp. 516–521, 2022.

[6] M. Hassan, A. Al-Dulaimi, and W. Xu, "Recurrent neural network models for national CO2 emission forecasting: A comparative study," IEEE Access, vol. 9, pp. 101914–101926, 2021, doi: 10.1109/ACCESS.2021.3097645.

[7] L. Shen, C. Zhou, W. Liu, Q. Li, Y. Hao, and D. He, "Autoformer-LSTM hybrid model for carbon emission intensity prediction in power markets," in Proc. 5th Int. Academic Exchange Conf. Science and Technology Innovation (IAECST), Guangzhou, China, 2023, pp. 1589–1592.

[8] Z. Huang, P. Wang, and Y. Feng, "CNN-LSTM method for industrial carbon emission prediction under dual carbon targets using extended Kaya identity," IEEE Trans. Ind. Electron., vol. 71, no. 3, pp. 2905–2914, Mar. 2024.

[9] F. Jiang, Y. Dong, and K. Liang, "Big data-driven deep LSTM approach for carbon emission forecasting in dynamic energy environments," IEEE Access, vol. 10, pp. 88421–88433, 2022, doi: 10.1109/ACCESS.2022.3199410.

[10] IEEE Young Professionals Climate and Sustainability Task Force, "Digitalization and sustainability: Leveraging Industry 4.0 for carbon emission reduction," IEEE Technology and Society Magazine, vol. 42, no. 2, pp. 45–54, Jun. 2023.

[11] D. R. Santos, L. F. Silva, and C. M. G. Matos, "AI and IoT-driven energy optimization in smart manufacturing: A systematic review," IEEE Internet of Things Journal, vol. 10, no. 6, pp. 5419–5427, Mar. 2023, doi: 10.1109/JIOT.2022.3218671.

[12] N. S. Jadhav, P. R. Patel, and A. S. Gupta, "Digital transformation and environmental sustainability: Challenges and opportunities for low-carbon enterprise management," IEEE Trans. Green Commun. Netw., vol. 7, no. 1, pp. 410–421, Mar. 2023.

[13] K. Zhang, H. Liu, and M. Wang, "AI-powered real-time monitoring and carbon emission reduction in the construction sector," IEEE Trans. Ind. Inform., vol. 19, no. 8, pp. 9011–9022, Aug. 2023, doi: 10.1109/TII.2022.3230198.

[14] M. Hans, E. Hikmawati, and K. Surendro, "Predictive analytics model for optimizing carbon footprint from learning activities in technology-intensive environments," IEEE Access, vol. 11, pp. 114976–114991, 2023, doi: 10.1109/ACCESS.2023.3324701.