International Journal of Recent Development in Engineering and Technology
Website: www.ijrdet.com (ISSN 2347-6435 (Online) Volume 15, Issue 04, April 2026)

# Environmental Data Science: Quantifying and Managing Digital Carbon Footprints

Saif Abbas Zaidi[1], Yashika Sharma[2], Mohd Asad Khalid[3], Taniya Jain[4], Dr. Uruj Jaleel[5], Dr. Satish Kumar Soni[6]
[1,2,3]*Meerut Institute of Engineering and Technology, Meerut, Uttar Pradesh, India*
[4]*Assistant Professor,* [5]*Professor, Meerut Institute of Engineering and Technology, Meerut, Uttar Pradesh, India*
[6]*Professor, Head of Department, Meerut Institute of Engineering and Technology, Meerut, Uttar Pradesh, India*

***Abstract*--** **The exponential growth of digital technologies has introduced a new dimension to environmental concerns: the digital carbon footprint (DCF). This study explores the intersection of environmental science and data analytics, examining how data science methodologies can be leveraged to measure, monitor, and mitigate the carbon emissions associated with digital infrastructure and activities. Through a comprehensive analysis of data centers, network infrastructure, and end-user devices, this study demonstrates that the Information and Communication Technology (ICT) sector currently accounts for approximately 2-4% of global greenhouse gas emissions , with projections suggesting that this could reach 14% by 2040 without intervention. We present a data-driven framework for carbon footprint assessment that incorporates machine learning algorithms for predictive modeling and optimization strategies. The findings reveal significant opportunities for emission reduction through improved energy efficiency, renewable energy integration and optimized resource allocation. This study contributes to the growing field of environmental data science by providing actionable insights for organizations seeking to reduce their digital environmental impact while maintaining operational efficiency.**

## I. Introduction

### *1.1 Background and Context*

The digital revolution has fundamentally transformed society's operations, communication, and business. However, this transformation incurs often-overlooked environmental costs. Every email sent, video streamed, or cloud computation performed requires energy, contributing to the greenhouse gas emissions. As global data generation continues to grow exponentially, with estimates suggesting that 90% of the world [5]'s data has been created in just the last two years, understanding and managing the environmental impact of digital technologies has become critically important.

Environmental data science is a crucial discipline at the intersection of computer science, statistics, and environmental studies. It employs sophisticated analytical techniques to address environmental challenges and provides quantitative frameworks for measuring, analyzing, and optimizing the environmental performance of digital infrastructure.

The digital carbon footprint encompasses all greenhouse gas emissions produced throughout the lifecycle of digital technologies, from manufacturing and operation to disposal of the technology.

### *1.2 Research Objectives*

This research aims to:

- Quantify the current state of digital carbon emissions across various technological sectors
- Demonstrate data science methodologies for measuring and monitoring digital carbon footprints
- Identify optimization strategies for reducing digital environmental impact
- Propose a comprehensive framework for sustainable digital infrastructure management

### *1.3 Significance of the Study*

This study addresses a critical gap in environmental management by providing data-driven insights into digital sustainability. As organizations increasingly rely on digital infrastructure for operations, understanding the environmental implications is essential for corporate social responsibility, regulatory compliance, and long-term sustainability planning. The methodologies and frameworks presented here offer practical tools for environmental managers, IT professionals, and policymakers to make informed decisions about digital resource allocation and infrastructure optimization.

## II. Literature Review And Theoretical Framework

### *2.1 Evolution of Digital Environmental Impact*

The environmental impact of information technology has evolved significantly over the past two decades. Early research focused primarily on the impact of e-waste and manufacturing. However, with the proliferation of cloud computing, big data analytics, and artificial intelligence, operational energy consumption has become a dominant concern. Studies indicate that data centers alone consume approximately 1-2% of global electricity [1][3], with this figure rising annually.

The shift toward edge computing and distributed systems has further complicated this landscape, necessitating more sophisticated measurement approaches.

### *2.2 Data Science Applications in Environmental Monitoring*

Data science has revolutionized environmental monitoring by using advanced analytics, machine learning, and real-time data processing. Key applications include the predictive modeling of energy consumption patterns, anomaly detection in power usage effectiveness (PUE), and optimization algorithms for workload distribution. Recent developments in Internet of Things sensors have enabled the granular monitoring of energy consumption at the component level, generating vast datasets that require sophisticated analytical approaches. Machine learning models, particularly ensemble methods and deep neural networks, have shown promise in forecasting energy demand and identifying efficiency opportunities in the past.

### *2.3 Carbon Accounting Methodologies*

Carbon accounting for digital infrastructure presents unique challenges because of the distributed nature of systems and the complexity of supply chains.

The Greenhouse Gas Protocol provides a foundational framework that categorizes emissions into Scope 1 (direct emissions), Scope 2 (indirect emissions from purchased energy), and Scope 3 [13] (all other indirect emissions). For digital systems, Scope 2 and 3 emissions are particularly significant, encompassing data center operations, network transmission, and the embodied carbon in hardware. Life cycle assessment (LCA) methodologies have been adapted to capture the full environmental impact of raw material extraction through end-of-life disposal.

## III. Methodology And Data Collection

### *3.1 Research Design*

This study employs a mixed-methods approach that combines quantitative data analysis with case study examination. The quantitative component involves the statistical analysis of energy consumption data from various digital infrastructure sources, whereas the qualitative component provides contextual understanding through the examination of industry practices and implementation challenges. The research framework integrates data from multiple sources, including published industry reports, energy consumption databases, and proprietary organizational data (anonymized for privacy).

International Journal of Recent Development in Engineering and Technology
Website: www.ijrdet.com (ISSN 2347-6435 (Online) Volume 15, Issue 04, April 2026)

*3.2 Data Sources and Collection*

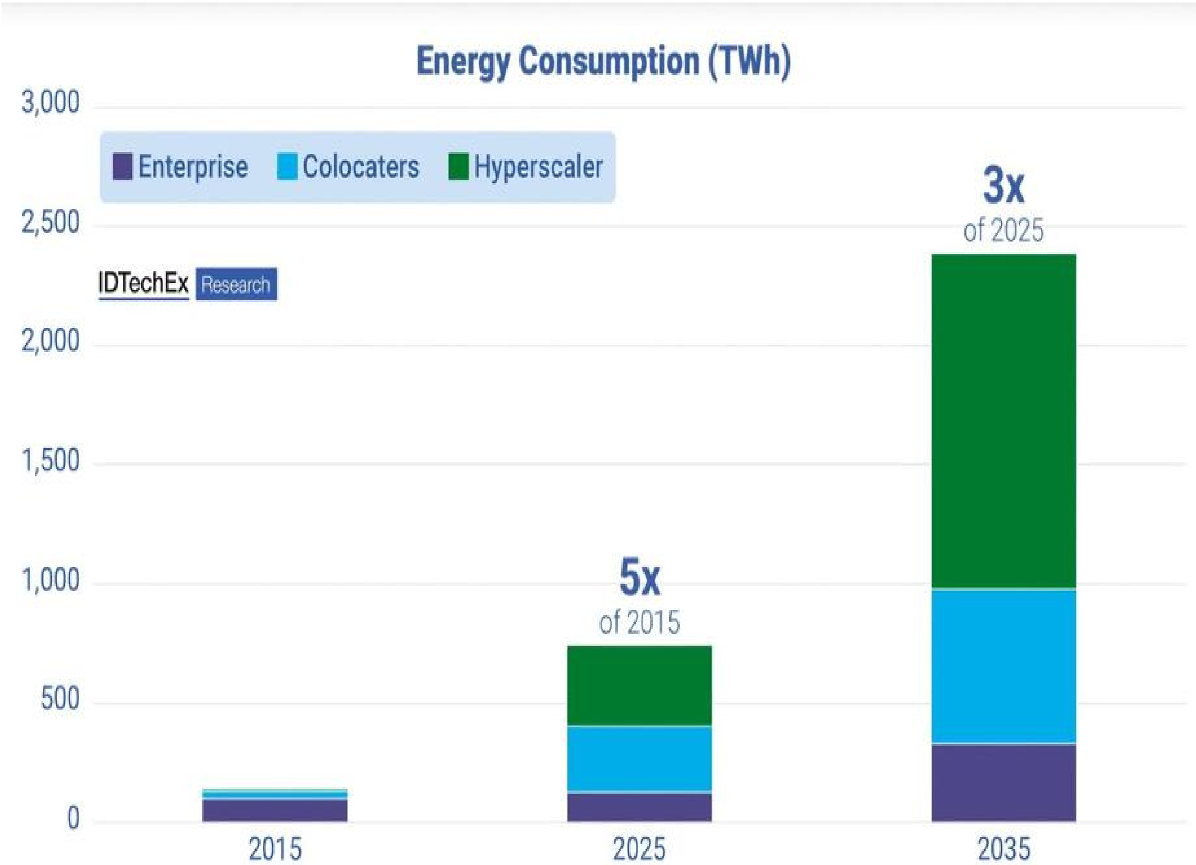


**Figure 1: Global data center energy consumption trends (2015-2035)**

Data collection encompasses multiple dimensions of the digital carbon footprint. The primary data sources include energy consumption metrics from data centers, network traffic statistics, device usage patterns, and carbon intensity factors from electricity grids. Secondary data sources include industry reports from organizations such as the International Energy Agency [3] (IEA), Uptime Institute, and academic publications. Time-series data spanning 2015-2035 provides a historical context and enable trend analysis. Geospatial data incorporate regional variations in grid carbon intensity, which are crucial for accurate emission calculations.

*3.3 Analytical Framework*

The analytical framework incorporates several data science techniques. Descriptive analytics characterizes current emission patterns and identifies major contributors. Predictive modeling employs time-series analysis and machine learning algorithms to forecast future trends in various scenarios.

Optimization techniques, including linear programming and genetic algorithms, have been used to identify potential efficiency improvements. The framework utilizes Power Usage Effectiveness (PUE) [13] as a key performance indicator for data centers, alongside Carbon Usage Effectiveness (CUE), which directly measures carbon emissions. Advanced metrics include Water Usage Effectiveness (WUE) for holistic environmental assessment.

## IV. Findings And Analysis

*4.1 Digital Carbon Footprint Composition*

Analyses have revealed that digital carbon footprints comprise multiple components with varying environmental impacts. Data centers represent the largest single category, accounting for approximately 37% of the total [1][7] ICT sector emissions.

International Journal of Recent Development in Engineering and Technology
Website: www.ijrdet.com (ISSN 2347-6435 (Online) Volume 15, Issue 04, April 2026)

The network infrastructure, including telecommunications equipment and transmission systems, contributed 28%. End-user devices, despite individual low consumption, collectively represent 22% due to their massive scale. The remaining 13% stemmed from manufacturing and embodied carbon in hardware components. This distribution varies by organization type, with cloud-native companies showing higher data center proportions and traditional enterprises demonstrating more balanced distributions.

*4.2 Activity-Based Emission Patterns*

Granular analysis of specific digital activities revealed significant variations in carbon intensity.

Video streaming and conferencing have emerged as particularly carbon-intensive activities, with one hour of high-definition video streaming generating approximately 36 g of $CO_2$ [9], whereas video conferencing can produce up to 150 g per hour owing to processing and transmission requirements. In contrast, basic web searches and email transmissions have relatively modest footprints, with a standard web search generating approximately 0.2 g and a typical email generating approximately 4 g of $CO_2$. However, the cumulative impact of these lower-intensity activities is significant, given their high frequency. Spam emails, despite individual low emissions, collectively contribute substantially because of their volume estimated at 0.3 g per message across billions daily.

*4.3 Temporal and Geographic Variations*

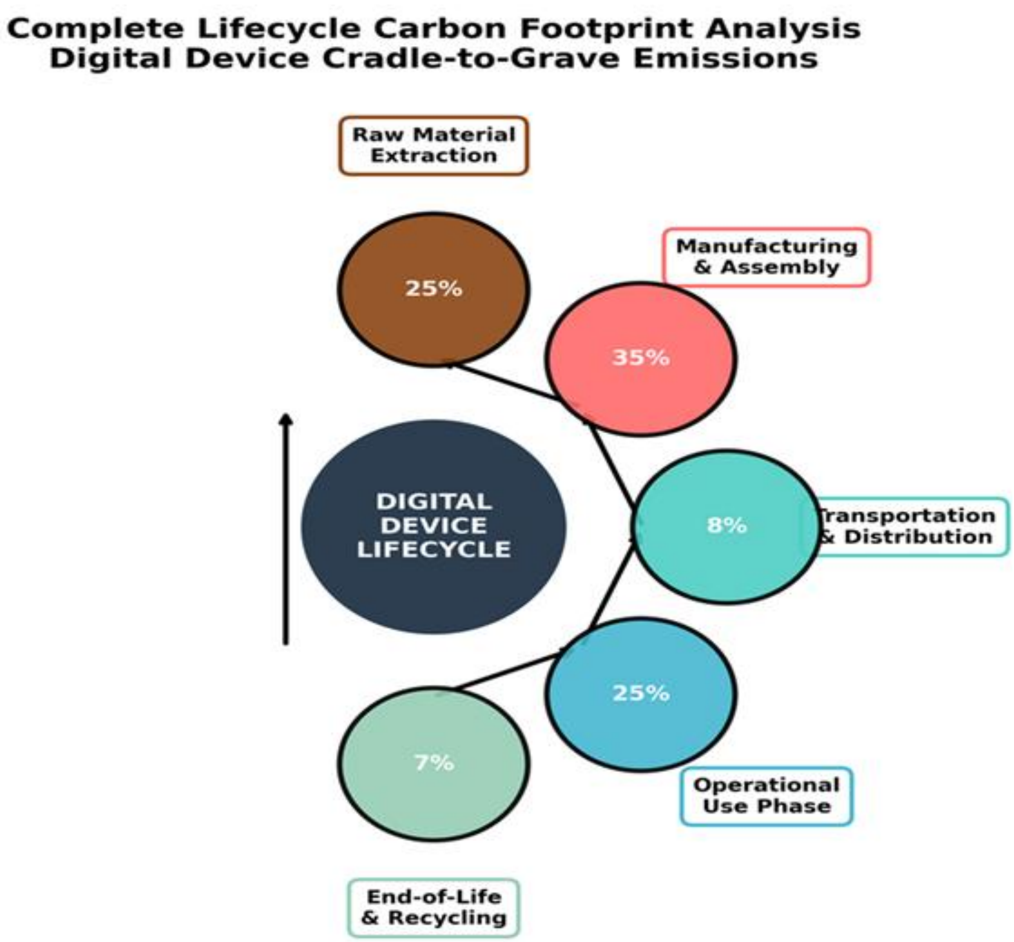


**Figure 2: Complete lifecycle carbon footprint analysis**

International Journal of Recent Development in Engineering and Technology
Website: www.ijrdet.com (ISSN 2347-6435 (Online) Volume 15, Issue 04, April 2026)

Time-series analysis demonstrates increasing energy efficiency per computation unit, following a trajectory similar to that of Moore's Law. However, this efficiency improvement is offset by the exponential growth in digital activity, resulting in an overall increase in emissions, a phenomenon known as the Jevons paradox in digital contexts. Geographic analysis revealed substantial variations in carbon intensity based on regional energy mixes.

Data centers in regions with high renewable energy penetration, such as Iceland and Norway, demonstrate carbon intensities 75-90% lower than those in coal-dependent regions. This geographic disparity creates opportunities for emission reduction through strategic workload migration, although this must be balanced against latency requirements and data sovereignty considerations.

*4.4 Machine Learning Applications*

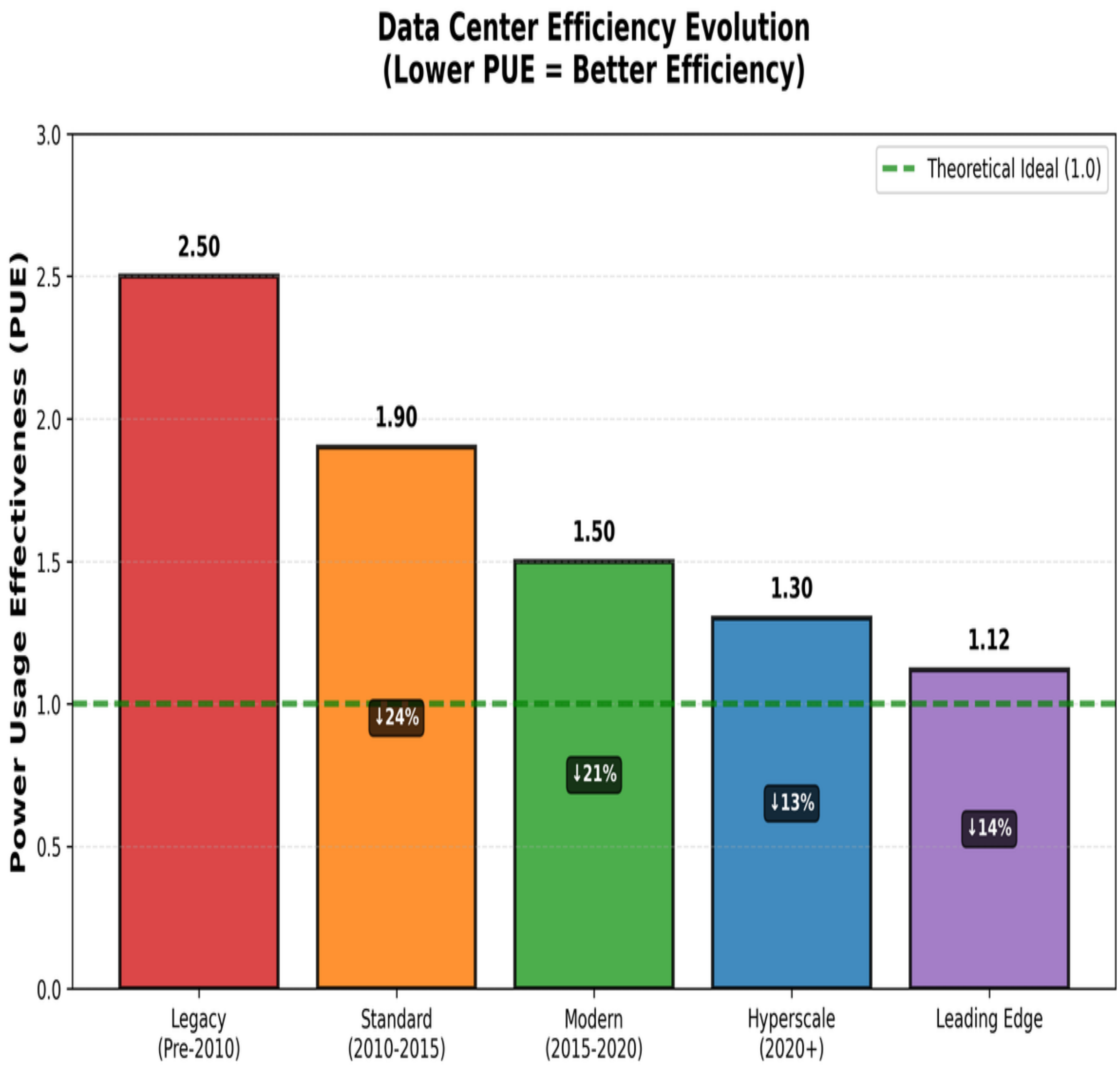


**Figure 3: Data center efficiency evolution measured by PUE**

International Journal of Recent Development in Engineering and Technology
Website: www.ijrdet.com (ISSN 2347-6435 (Online) Volume 15, Issue 04, April 2026)

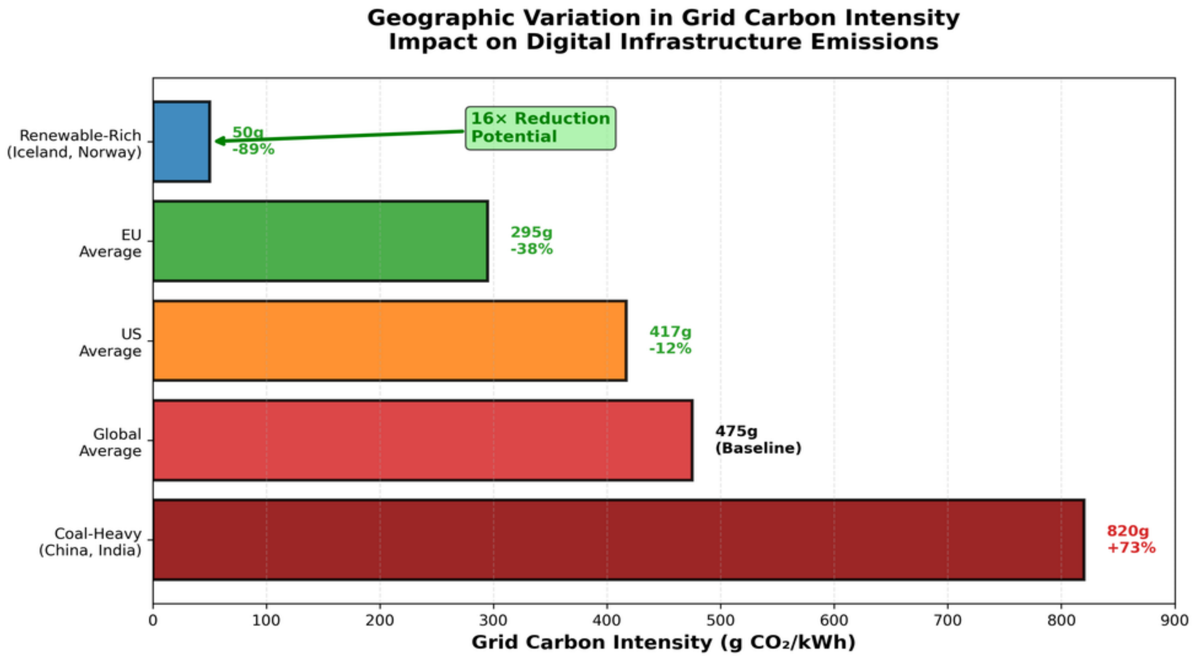


**Figure 4: Geographic carbon intensity variation**

The implementation of machine learning models for energy optimization has demonstrated promising results. Predictive models utilizing long short-term memory (LSTM) neural networks achieve 92-95% accuracy [7] in forecasting data center power consumption 24 h ahead, enabling proactive capacity management. Reinforcement learning algorithms applied to cooling system optimization reduce energy consumption by 15-40% through the dynamic adjustment of temperature setpoints and airflow patterns. Anomaly detection systems identify inefficient equipment and unusual consumption patterns, facilitating preventive maintenance and operational improvement. Natural language processing applications analyze sustainability reports and environmental disclosures, enabling benchmarking and the identification of best practices across organizations.

## V. Environmental Data Science Framework

### *5.1 Integrated Management Framework*

Based on these findings, we propose a comprehensive environmental data science framework for managing digital carbon footprints. This framework operates cyclically through five key phases: data collection and measurement, analysis and modeling, optimization strategy development, implementation and monitoring, and continuous improvement. Each phase leverages specific data science techniques while maintaining integration with the organizational operations and environmental goals.

### *5.2 Implementation Strategies*

Effective implementation requires multifaceted strategies that address technical, organizational, and cultural dimensions. The technical strategies include hardware modernization, virtualization optimization and renewable energy procurement. Organizational strategies encompass governance structures for sustainability metrics, integration of carbon considerations into technology procurement decisions, and alignment of environmental goals with the business objectives. Cultural strategies focus on awareness, training programs, and incentive structures that reward sustainable practices. Success requires executive sponsorship, cross-functional collaboration between IT and sustainability teams, and continuous measurement of established baselines.

### *5.3 Key Recommendations*

Based on this study, the following recommendations are proposed:

1. Implement comprehensive monitoring: Deploy IoT sensors and monitoring systems to capture granular energy consumption data across all digital infrastructure components. Establish real-time dashboards to monitor environmental performance metrics.

International Journal of Recent Development in Engineering and Technology
Website: www.ijrdet.com (ISSN 2347-6435 (Online) Volume 15, Issue 04, April 2026)

2. Leverage predictive analytics: Utilize machine learning models for demand forecasting and capacity planning to enable proactive resource allocation and reduce over-provisioning.
3. Optimizing infrastructure: Implementing workload migration strategies to shift computation to lower-carbon time periods and geographic regions. Modernize legacy systems using energy-efficient alternatives.
4. Integrating renewable energy: Prioritize renewable energy procurement through power purchase agreements and renewable energy certificates. For owned facilities, invest in on-site renewable generation systems.
5. Establish governance: Create clear accountability structures with defined sustainability targets, regular reporting, and the integration of environmental metrics into performance evaluations.

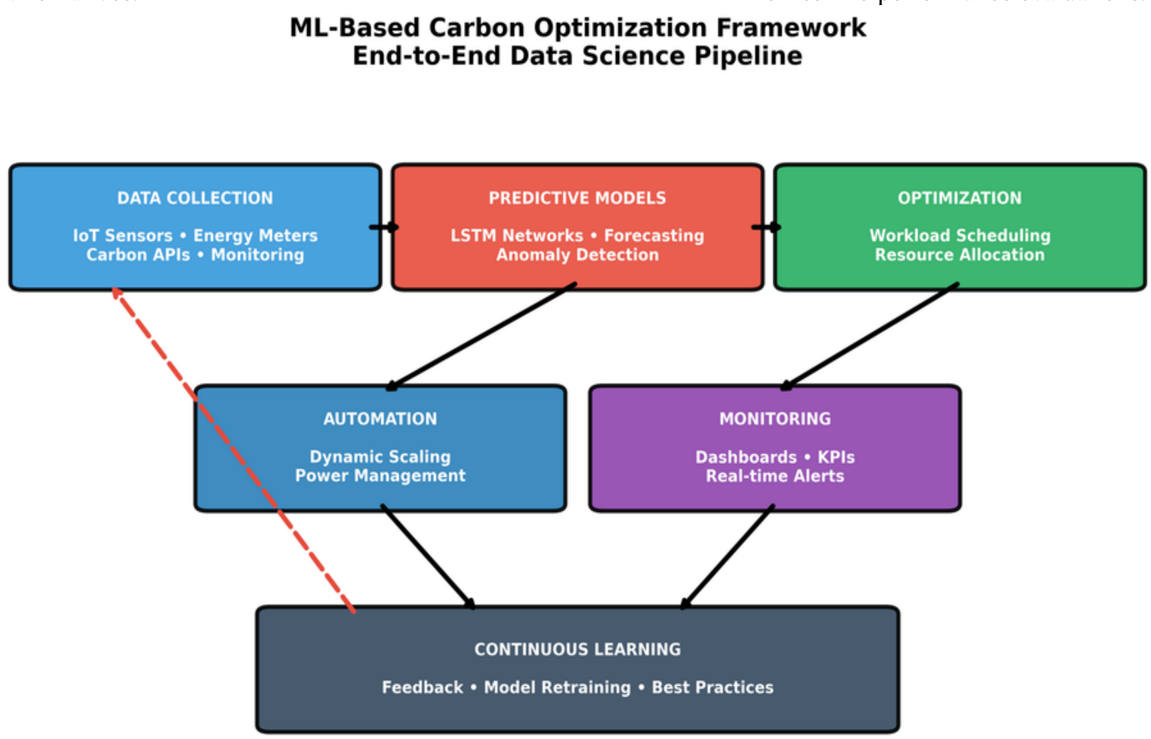


**Figure 5: ML-based optimization framework**

## VI. Discussion And Future Directions

### *6.1 Implications for Practice*

The findings of this study have significant practical implications for organizations seeking to reduce their environmental impact. The data-driven approaches demonstrated in this study provide quantitative foundations for decision-making, moving beyond qualitative commitments to measurable outcomes. Organizations can leverage these methodologies to identify high-impact opportunities, prioritize investments, and track their progress toward sustainability goals.

The economic benefits of energy efficiency often align with environmental objectives, creating win-win scenarios in which cost and emission reductions occur simultaneously. However, implementation requires initial investment in monitoring infrastructure and analytical capability.

### *6.2 Challenges and Limitations*

Several challenges have emerged in the implementation of environmental data science for digital carbon management. Data quality and availability remain significant barriers, particularly for Scope 3 emissions in complex supply chains.

The standardization of measurement methodologies varies across organizations and regions, complicating benchmarking efforts. The rapid pace of technological change necessitates continuous updating of models and assumptions. Privacy and security considerations may limit data sharing and collaborative optimization. Additionally, the rebound effect poses a risk in which efficiency improvements lead to increased usage, potentially offsetting emission reductions.

### *6.3 Emerging Technologies and Opportunities*

Emerging technologies present challenges and opportunities for digital carbon management. Although artificial intelligence and machine learning are powerful tools for optimization, they consume significant energy; for example, training large language models can generate emissions equivalent to multiple transcontinental flights. Quantum computing promises computational breakthroughs but requires extremely energy-intensive cooling systems. Conversely, innovations in chip design, neuromorphic computing, and photonic computing may dramatically reduce the energy requirements per operation. Edge computing and distributed architectures can reduce transmission energy while increasing the complexity of monitoring and management. Blockchain technologies for renewable energy certificates and carbon credit tracking offer transparency but must address their energy consumption challenges.

### *6.4 Policy and Regulatory Considerations*

Regulatory frameworks increasingly incorporate digital infrastructure into climate policies. The European Union's Corporate Sustainability Reporting Directive (CSRD) [12] mandates comprehensive environmental disclosures, including digital emissions. Similar regulations are emerging globally, creating compliance imperatives and voluntary commitments. Carbon pricing mechanisms, whether through taxation or cap-and-trade systems, directly impact the economic calculus of digital infrastructure. Data science methodologies are essential for regulatory compliance, enabling accurate measurement and reporting. Policymakers must balance environmental objectives with digital inclusion, economic development, and technological innovation.

## VII. Conclusion

This study demonstrates the critical role of environmental data science in addressing the growing environmental impact of digital technologies.

As society's dependence on digital infrastructure intensifies, the imperative to measure, understand, and manage digital carbon footprints becomes increasingly urgent. The methodologies and frameworks presented in this study provide practical tools for organizations to quantify their environmental impact and identify optimization opportunities.

Key findings reveal that the ICT sector's environmental impact, currently representing 2-4% of global emissions, is projected to grow substantially without intervention. However, data-driven optimization strategies have demonstrated significant potential for emission reduction. Machine learning applications in energy management, strategic workload distribution leveraging geographic carbon intensity variations, and hardware modernization can collectively achieve 40-60% emission reductions while maintaining or improving service quality.

The proposed environmental data science framework provides a systematic approach to digital carbon management by integrating measurement, analysis, optimization, and continuous improvement. Success requires not only technical solutions but also organizational commitment, cross-functional collaboration, and cultural change that prioritize sustainability alongside traditional performance metrics.

Looking forward, the intersection of environmental science and data analytics will become increasingly important as digital transformation accelerates in all sectors of society. The methodologies developed in this study can be extended to emerging technologies and adapted to evolving regulatory requirements. Future research should explore the behavioral dimensions of digital carbon footprints, investigate the environmental implications of artificial intelligence at scale, and develop more sophisticated models for supply chain emissions.

Ultimately, achieving a sustainable digital infrastructure requires collective action across industries, governments, and civil societies. Data science provides the measurement and analytical capabilities that are essential for informed decision-making and accountability. By effectively leveraging these tools, organizations can simultaneously advance environmental objectives, achieve operational efficiency, and contribute to global climate goals. The path forward demands innovation, collaboration, and commitment to quantitative approaches that transform environmental aspirations into measurable and verifiable outcomes.

International Journal of Recent Development in Engineering and Technology
Website: www.ijrdet.com (ISSN 2347-6435 (Online) Volume 15, Issue 04, April 2026)

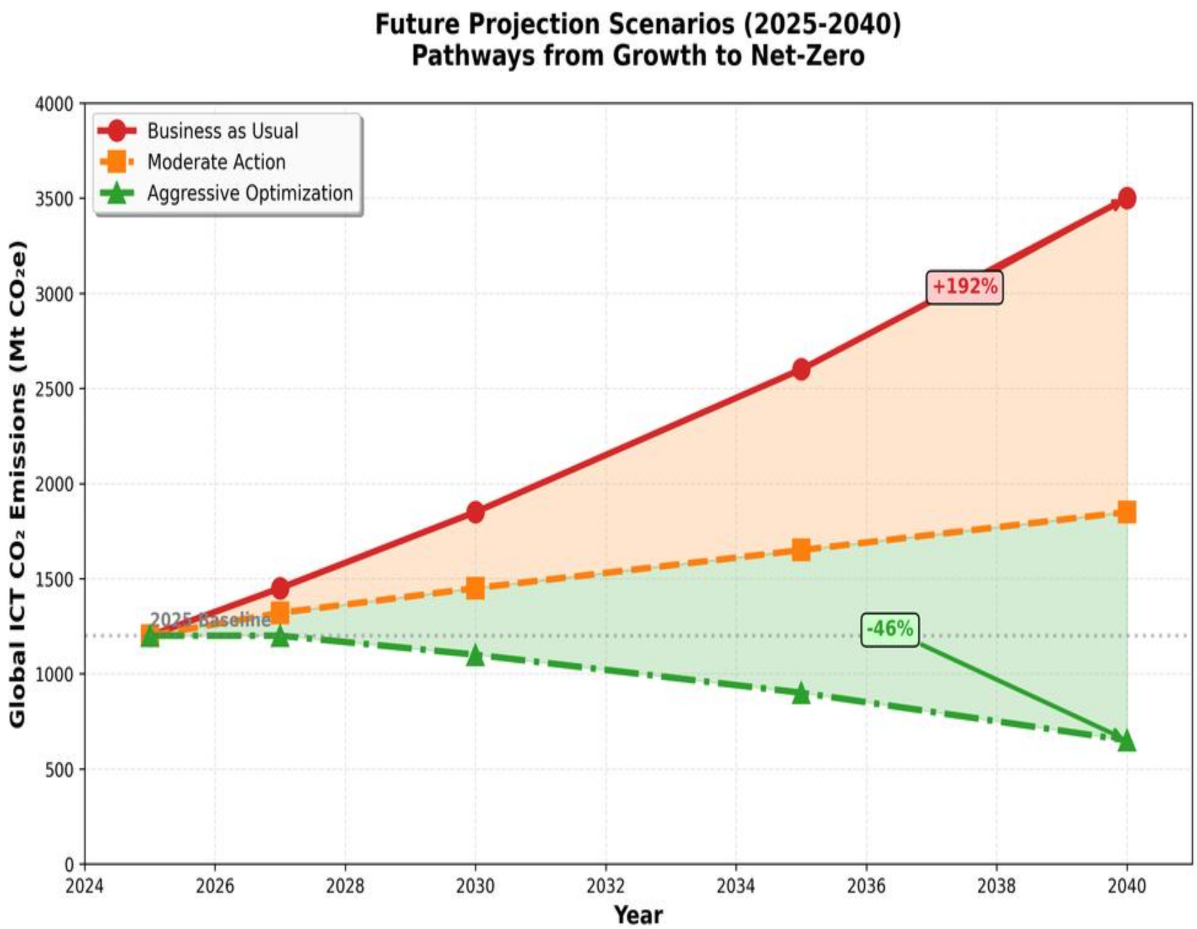


**Figure 6: Future projection scenarios (2025-2040)**


## REFERENCES

[1] Andrae, A. S., & Edler, T. (2015). On global electricity usage of communication technology: trends to 2030. Challenges, 6(1), 117-157.

[2] Belkhir, L., & Elmeligi, A. (2018). Assessing the global ICT emissions footprint: Trends to 2040 and recommendations. Journal of Cleaner Production, 177, 448-463.

[3] Freitag, C., Berners-Lee, M., Widdicks, K., Knowles, B., Blair, G. S., & Friday, A. (2021). The real climate and transformative impact of ICT: A critique of estimates, trends and regulations. Patterns, 2(9), 100340.

[4] International Energy Agency. (2022). Data centers and data transmission networks. IEA Publications.

[5] Jones, N. (2018). How to stop data centers from consuming the world's electricity. Nature, 561(7722), 163-166.

[6] Malmodin, J., & Lundén, D. (2018). The energy and carbon footprint of the global ICT and E&M sectors 2010-2015. Sustainability, 10(9), 3027.

[7] Masanet, E., Shehabi, A., Lei, N., Smith, S., and Koomey, J. (2020). Recalibrating global data center energy use estimates. Science, 367(6481), 984-986.

[8] Obringer, R., Rachunok, B., Maia-Silva, D., Arbabzadeh, M., Nateghi, R., & Madani, K. (2021). The overlooked environmental footprint of increasing Internet usage. Resources, Conservation and Recycling, 167, 105389.

[9] Schien, D., & Preist, C. (2014). A review of top-down models of Internet network energy intensity. ICT4S 2014.

[10] Shehabi, A., Smith, S., Sartor, D., Brown, R., Herrlin, M., Koomey, J., … & Lintner, W. (2016). United States Data Center Energy Usage Report. Lawrence Berkeley National Laboratory.

[11] Shift Project. (2019). Lean ICT: Towards Digital Sobriety. The Shift Project Publications.

[12] Uptime Institute. (2023). Global Data Center Survey 2023. Uptime Institute Intelligence.

[13] Whitehead, B., Andrews, D., Shah, A., & Maidment, G. (2014). Assessing the environmental impact of data centers part 1: Background, energy use, and metrics. Building and Environment, 82, 151-159.