PLOS CLIMATE

REVIEW

Integrating machine learning into life cycle assessment: Review and future outlook

Hairong Wang*

Department of Civil and Environmental Engineering, Carnegie Mellon University, Pittsburgh, Pennsylvania, United States of America

* hairongw@andrew.cmu.edu

Citation: Wang H (2025) Integrating machine learning into life cycle assessment: Review and future outlook. PLoS Clim 4(10): e0000732. https://doi.org/10.1371/journal.pclm.0000732

Editor: Taoyuan Wei, Center for International Climate and Environmental Research: CICERO, NORWAY

Received: May 26, 2025

Accepted: September 29, 2025

Published: October 16, 2025

Funding: The author(s) received no specific funding for this work.

Competing interests: The authors have declared that no competing interests exist.

Abstract

Life Cycle Assessment (LCA) is widely used to quantify environmental impacts but often faces data gaps, heterogeneous practices, and limited timeliness. This review examines how machine learning (ML) can strengthen LCA across all four phases—goal & scope, life cycle inventory (LCI), life cycle impact assessment (LCIA), and interpretation—while providing a reproducible bibliometric map of recent research. We performed a bibliometric search and keyword co-occurrence visualization (VOSviewer) and organized the literature by LCA phases. We highlight actionable opportunities: NLP-assisted scope definition, probabilistic imputation and uncertainty quantification for LCI, surrogate and hybrid models for LCIA, and calibrated, decision-oriented interpretation. Compared with prior reviews, we (i) deliver phase-specific guidance instead of generic lists, (ii) extend coverage to recent work with reproducible bibliometrics, and (iii) foreground early-phase opportunities that remain under-explored. These insights—together with open materials for reuse—aim to make LCA more data-robust, transparent, and actionable in research and practice.

Author summary

Life Cycle Assessment (LCA) is a vital tool for understanding the environmental footprint of products and services, but it often struggles with incomplete data and an inability to adapt to changing conditions. We review how Machine Learning (ML), a type of artificial intelligence, is revolutionizing LCA. ML can automatically gather and fill in missing data, make LCA models more dynamic, and help us make better environmental decisions. For instance, ML can process text to define the scope of an LCA or predict environmental impacts when direct data are scarce. While ML offers powerful new capabilities, we also discuss challenges like ensuring ML models are understandable and can be widely applied. Our work highlights the need for standardized approaches and greater collaboration between LCA experts and ML specialists to unlock the full potential of ML in making sustainability assessments more robust and actionable.

Introduction

Life Cycle Assessment (LCA) is a widely used methodology for systematically evaluating the environmental impacts of products, processes, and services throughout their life cycle. Standardized under ISO 14040 and 14044 [1,2], LCA provides a structured approach to assess resource consumption, emissions, and overall sustainability from raw material extraction to end-of-life disposal. It serves as a fundamental decision-making tool in environmental management, eco-design, and policy development. However, despite its widespread adoption, LCA faces several well-documented challenges, including data scarcity, high uncertainty, and a static nature that may not fully reflect dynamic real-world processes [3,4].

Traditional LCA methodologies are heavily based on extensive life cycle inventory (LCI) datasets, which are often incomplete or inconsistent between different industries and regions. This limitation introduces significant uncertainty into impact assessments, which requires assumptions and simplifications that may reduce reliability. In addition, conventional LCA models provide static 'snapshot' analyses that struggle to incorporate temporal, geographical, and technological variations. As environmental regulations and sustainability goals evolve, there is a growing demand for more adaptive and predictive LCA approaches that can integrate diverse data sources and update dynamically in response to real-time changes [5,6].

Recent advances in big data analytics and machine learning (ML) offer promising solutions to overcome these challenges. ML techniques excel at handling complex, high-dimensional, and non-linear datasets, making them highly suitable for automating data acquisition, harmonization, and predictive modeling within LCA. Using artificial intelligence, ML can improve the accuracy of LCA by filling data gaps, estimating missing values, and integrating real-time environmental and operational parameters. For example, research comparing traditional LCA simulations with ML-based predictive models for ship emissions has shown that ML can incorporate factors such as engine performance, operational conditions, and external environmental variables, resulting in more precise and reliable impact estimations [7,8]. Such studies highlight the potential of ML to refine LCA methodologies, making them more adaptable to real-world complexities.

Beyond improving data quality and uncertainty quantification, ML also streamlines computational complexity in LCA. Regression models, generative techniques, and optimization algorithms have been used successfully to reduce the number of indicators required while maintaining predictive accuracy. For example, combining multilinear regression with mixed-integer linear programming has enabled the development of simplified LCA models that accurately estimate environmental impacts using a reduced set of proxy metrics [9]. This approach not only improves computational efficiency, but also facilitates the broader adoption of LCA among industries and policymakers by making the analysis more accessible and interpretable.

Despite its transformative potential, the integration of ML into LCA presents its own set of challenges. Issues such as model generalizability, explainability, and integration with existing LCA frameworks need to be addressed to ensure robustness and reliability. Furthermore, as ML-driven LCA models become more sophisticated, their transparency and interpretability will be critical to gaining stakeholder trust and regulatory acceptance.

Several recent studies have reviewed the intersection of machine learning (ML) and life cycle assessment (LCA), notably the work of Ghoroghi et al. (2022), Romeiko et al. (2024), and Salla et al. (2025) [4,5,10]. Building upon these valuable surveys, our review provides a more structured and methodologically-grounded analysis. Our work is distinguished by its organization around the four standard LCA phases, an approach that allows us to emphasize emerging opportunities in early stages (such as goal and scope definition) that were not the primary focus of previous work. To support this analysis, we incorporate a bibliometric review of literature through 2025 and identify future research directions that build upon and extend beyond prior findings.

This review explores the integration of ML into LCA across four key stages: goal and scope definition, life cycle inventory (LCI) analysis, life cycle impact assessment (LCIA), and interpretation. By systematically analyzing existing studies, we identify the strengths and limitations of ML-enhanced LCA, highlight current research gaps, and propose future directions for advancing the field. The objective is to provide a comprehensive understanding of how ML can transform LCA into a more efficient, adaptive, and decision-oriented tool for sustainability assessments.

Theoretical background

Life cycle assessment (LCA)

Life Cycle Assessment (LCA) is a globally recognized standardized methodology used to comprehensively evaluate the environmental burdens associated with a product, process, or service. It takes a "cradle-to-grave" (or "cradle-to-cradle") perspective, encompassing all stages of a system's life, from raw material extraction and processing, through manufacturing, distribution, use, and end-of-life management (e.g., recycling, disposal). This holistic approach is essential for identifying potential trade-offs and preventing problem-shifting, where solving one environmental issue might inadvertently create another elsewhere in the life cycle. The core principle of LCA is to provide a complete picture of environmental interactions, allowing informed decision-making in design, policy, and business strategy.

The LCA framework, as formalized by the International Organization for Standardization (ISO) in Standards 14040 and 14044, is structured around four interconnected phases: goal and scope definition, life cycle inventory (LCI) analysis, life cycle impact assessment (LCIA) and interpretation. These phases are iterative, meaning that findings in later stages may lead to refinements in earlier ones. The initial phase, goal and scope definition, is paramount, establishing the purpose, target audience, system boundaries, and the functional unit (a quantified measure of the service provided) to ensure fair comparisons between alternatives.

The LCI phase involves detailed accounting of all material and energy inputs and outputs (emissions and waste) associated with each process within the defined system boundaries. This data collection can be very demanding, often relying on a combination of primary data (directly from specific processes) and secondary data (from databases, literature, or industry averages). The LCIA phase then uses known characterization factors to turn the LCI data into a set of environmental impact indicators, including global warming potential, acidification, eutrophication, and resource depletion. Different LCIA methodologies exist, offering various perspectives and levels of detail (e.g., midpoint vs. endpoint indicators).

The final interpretation phase is where the results of the LCI and LCIA are analyzed to identify significant environmental hotspots, assess the robustness of the findings through sensitivity and uncertainty analyses, and draw conclusions. This phase is crucial to translating the quantitative results into actionable recommendations to improve environmental performance. LCA can be broadly categorized into attributional LCA, which provides a static snapshot of environmental flows, and consequential LCA, which models the dynamic consequences of decisions, including market and economic effects [11,12].

Machine learning (ML)

Machine Learning (ML), a core branch of artificial intelligence (AI), allows computer systems to learn from data without explicit programming. Instead of relying on predefined rules, ML algorithms identify patterns, make predictions, and adapt their performance as they are exposed to new data. This ability to learn and adapt makes machine learning particularly valuable for tackling complex, data-rich problems in various fields, including environmental science and sustainability assessment [13].

ML encompasses a spectrum of techniques, broadly categorized as supervised, unsupervised, reinforcement, and deep learning. Labeled data sets (with known inputs and outputs) are used in supervised learning to train models for regression (predicting continuous values) or classification (predicting categorical labels). Unsupervised learning, on the other hand, looks at data that hasn't been labeled to find hidden structures, like groups of data points that are similar or simplified representations of large datasets. Reinforcement learning takes a different approach: training agents to make optimal decisions through interactions with an environment and learning from rewards and penalties [14].

Deep learning, a powerful subset of ML, utilizes artificial neural networks with multiple layers (deep neural networks) to learn intricate patterns from vast amounts of data [15]. These deep networks have achieved breakthroughs in areas such as image recognition and natural language processing and are increasingly being applied to sustainability challenges. Different kinds of neural networks have different strengths: for example, convolutional neural networks (CNNs) are well suited for image data, recurrent neural networks (RNNs) for sequential data, and autoencoders for reducing dimensionality.

The integration of ML with other analytical methods is also growing, with approaches such as physics-informed machine learning (PIML). Furthermore, ensemble methods, which combine predictions from multiple models, and techniques like Gaussian Process Regression provide enhanced robustness and uncertainty quantification. These advancements are continually developing a powerful toolkit for applications in various fields [16].

Synergies between LCA and ML

Integrating ML into LCA can effectively address traditional methodological limitations by automating data acquisition and harmonization through Natural Language Processing (NLP), filling data gaps and improving data quality using regression models and generative techniques, dynamically analyzing temporal changes, and simplifying complex multi-criteria decision-making processes through clustering, optimization, and explainable AI (XAI). Table 1 summarizes the synergies between ML and LCA along with relevant references from the literature.

Bibliometric analysis

We conducted a reproducible bibliometric search (time window: 2015–2025; last search: Aug 22, 2025) in Web of Science Core Collection. The exact query string was:

```
TS = (
  ("Life Cycle Assessment" OR LCA OR "life cycle inventory" OR
      LCI OR
   "life cycle impact assessment" OR LCIA OR
   "Environmental Life Cycle Assessment" OR "Sustainable Life
      Cycle Assessment" OR
   "Life Cycle Analysis")
```

Table 1. Synergies between ML and LCA stages.

LCA Stage	ML Applications	Relevant Literature
Goal and Scope Definition	NLP for automated extraction of goals, scope, and functional units; ontology-based consistent boundary definition; clustering/classification for comparative analysis.	Zargar et al. [17]; Koeck et al. [3]; Ibn-Mohammed et al. [18]
Inventory Analysis (LCI)	Regression models for data imputation; generative models (GANs, VAEs); ANNs and Gaussian processes for complex predictions; decision trees, random forests; clustering and classification for data mapping; transfer and federated learning.	Khadem et al. [19]; Dai et al. [20]; Saad et al. [21]; Zargar et al. [17]; Pascual-González et al. [9]; Kane et al. [22]; Tomic et al. [23]; Elhami et al. [24]; Wright et al. [25]
Impact Assessment (LCIA)	Gaussian processes for uncertainty modeling; ML regression for accurate characterization factors; sensitivity analysis methods (Sobol, Morris); Bayesian networks for probabilistic impact predictions.	Marvuglia et al. [26]; Dai et al. [20]; Di Lullo et al. [27]; Rivera and Sutherland [28]; Khang et al. [29]; Stemmle et al. [30]; Bala et al. [7]; Salla et al. [10]
Interpretation and Decision-Making	Explainable AI (SHAP, LIME); clustering for result simplification; MCDA integrated with ML; optimization algorithms (genetic algorithms); visualization (Pareto fronts, Sankey diagrams).	Hassan et al. [31]; Elhami et al. [24]; Tushar et al. [32]; Zhang et al. [8]; Belucio et al. [33]; Mabey et al. [34]; Eddy et al. [35]; Manuguerra et al. [36]; Zhang et al. [8]

https://doi.org/10.1371/journal.pclm.0000732.t001

```
AND
  ("Machine Learning" OR "Artificial Intelligence" OR AI OR "
      Deep Learning" OR
   "Neural Network*" OR "Support Vector Machine*" OR SVM OR
   "Random Forest*" OR "Gradient Boost*" OR "Regression Model
      *" OR "Clustering" OR
   "Data Mining" OR "Predictive Model*" OR "Supervised
      Learning" OR
   "Unsupervised Learning" OR "Reinforcement Learning")
AND
  (("Data Gap*" OR "Data Scarcity" OR "Missing Data" OR "Data
      Imputation" OR
    "Data Quality" OR "Data Uncertainty" OR "Data Integration"
       OR
    "Data Harmonization" OR "Data Deficiency" OR "Data
       Incompleteness")
   OR
   ("Parameter Estimation" OR "Sensitivity Analysis" OR "
      Uncertainty Quantification" OR
    "Dynamic LCA" OR "Scenario Analysis" OR "Design
       Optimization" OR
    "Eco-design" OR "Sustainable Design")
   OR
   ("Pattern Recognition" OR "Decision Support" OR "Predictive
      LCA" OR
    "Real-time LCA" OR "LCA Modeling" OR "LCA Simulation"))
)
AND PY = (2015-2025)
```

Document types included articles and reviews; language was English. Scripts and the .bib file are deposited in the repository referenced in the Data Availability statement. VOSviewer processed the dataset, generating a co-occurrence network that revealed three major clusters reflecting distinct research themes (Fig 1).

One cluster (colored red) predominantly featured terms such as "artificial neural networks," "carbon footprint," and "sensitivity analysis," indicating efforts to enhance environmental impact quantification through ML techniques. A second cluster (colored green) focused on "data quality," "predictive modeling," and "decision support," reflecting how ML can help fill in missing data and make LCA more reliable. The third cluster (colored blue) centered on applications in the built environment and dynamic analysis, using keywords such as "circular economy," "building," "construction," and "scenario analysis." These distinct clusters suggest a partitioning of the research: one segment emphasizes ML methods for improving impact calculation, another focuses on predictive modeling and data quality, and the last involves sector-specific dynamic sustainability assessments.

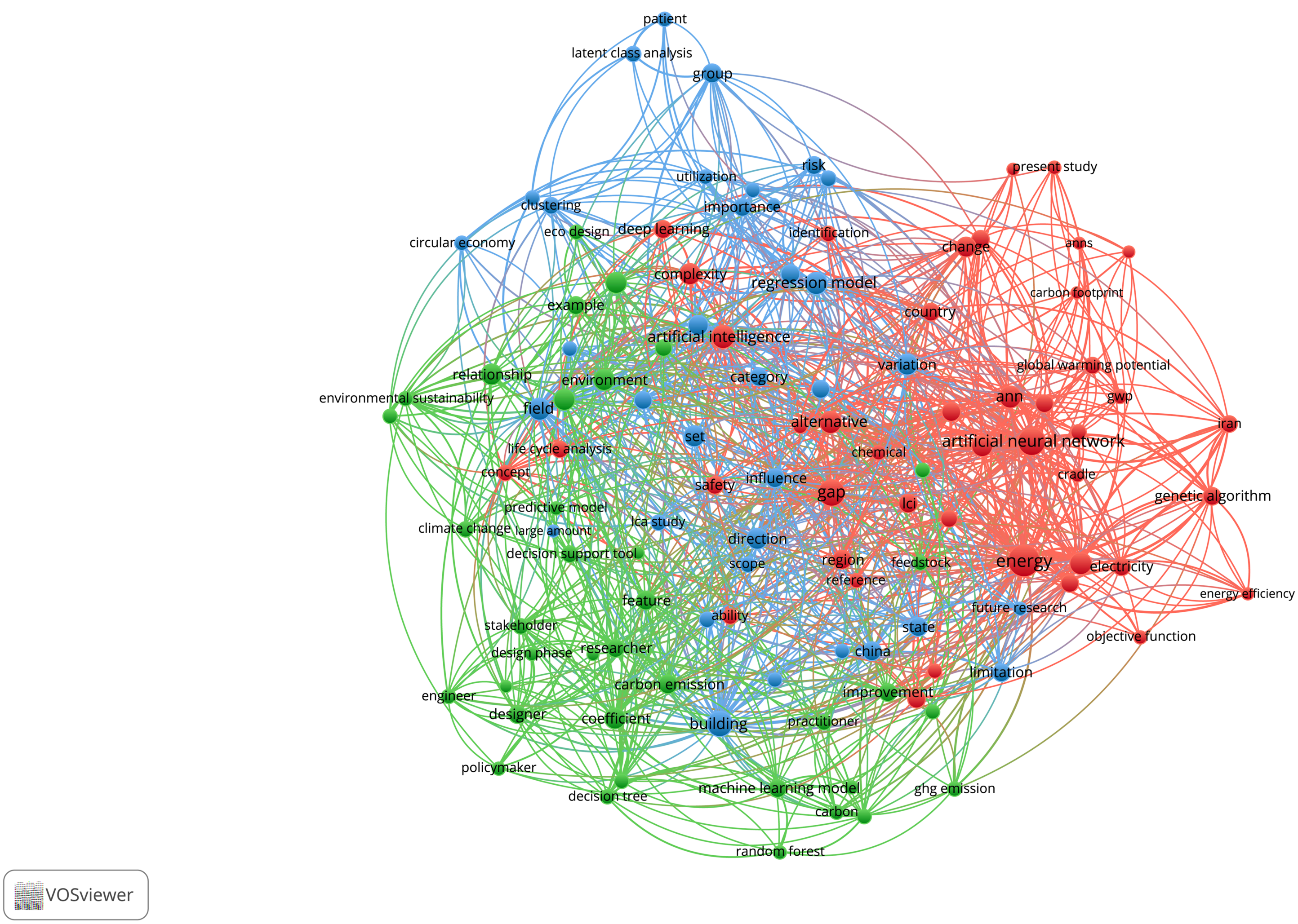

Fig 1. Network visualization of keyword co-occurrence in the literature. This figure illustrates the co-occurrence network of keywords from the bibliometric analysis, highlighting major research clusters. Red Cluster = ML Methods & Environmental Impact Metrics; Green Cluster = Data Quality & Predictive Decision Support; Blue Cluster = Sectoral & Dynamic Sustainability Applications.

https://doi.org/10.1371/journal.pclm.0000732.g001

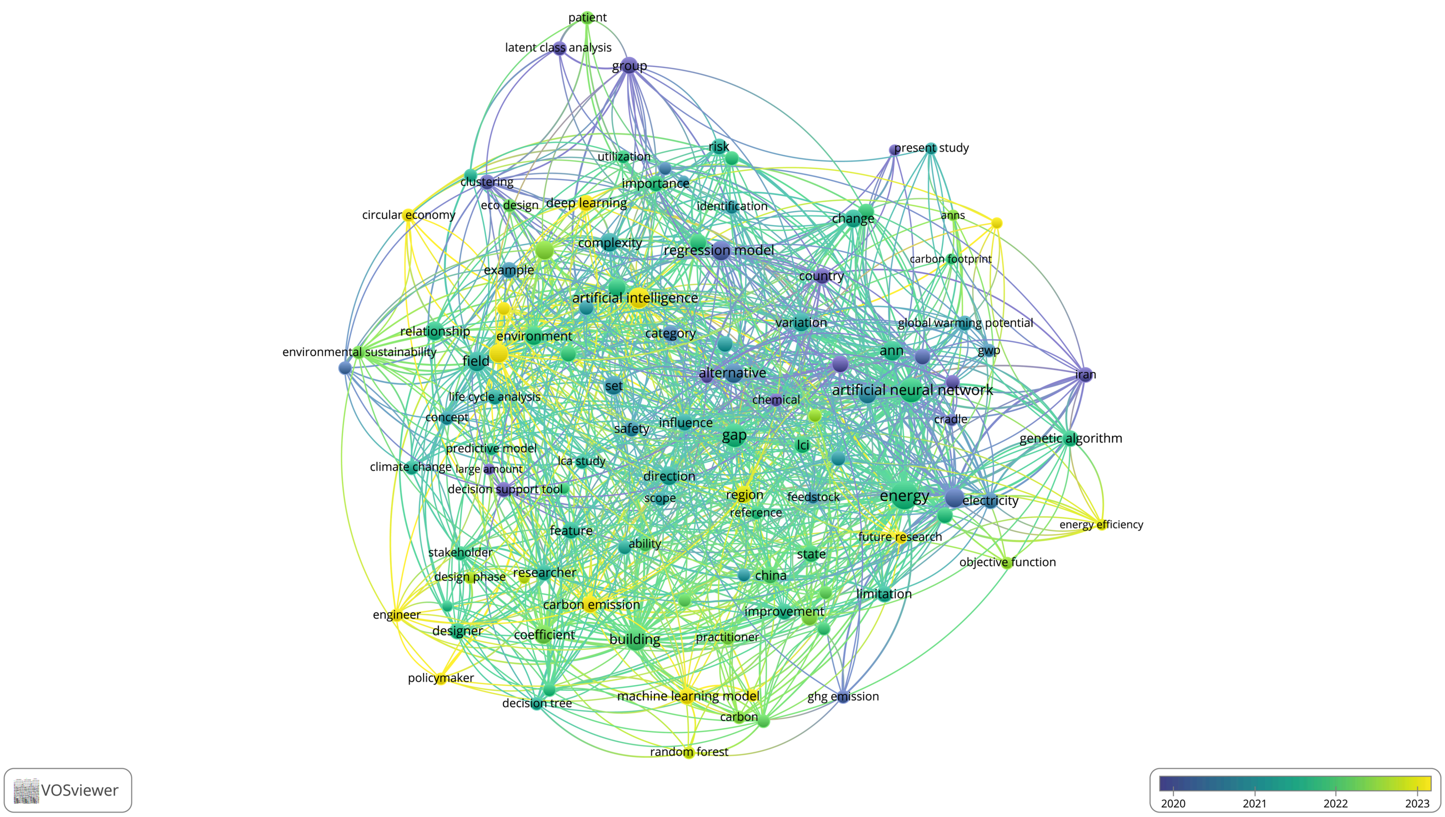

Fig 2. Overlay visualization of keyword trends over time. This figure shows the temporal trends of keywords, indicating the evolution of research focus over time.

https://doi.org/10.1371/journal.pclm.0000732.g002

In the co-occurrence map, carbon emissions and carbon footprint fall into different clusters because they co-occur with distinct neighborhoods: the former is tied to methods/-metrics (e.g., factor modeling, sensitivity, uncertainty), while the latter is embedded in application/sector contexts (e.g., product-level assessments and consumption-based studies). This split reflects topic usage contexts rather than semantic disagreement.

An examination of keyword usage over time reveals that foundational terms such as “carbon footprint,” “energy consumption,” and “global warming potential” maintained consistent prominence (which shown in Fig 2). However, after 2020, there has been a clear increase in the use of ML-specific keywords such as “random forest,” “clustering,” and “predictive model.” This temporal trend indicates that advanced computing techniques are being used increasingly in LCA research as the field shifts focus toward managing uncertain data and creating reliable dynamic sustainability assessments.

Findings

ML in goal and scope definition

The initial goal and scope definition phase of LCA involves defining the functional unit, the system boundaries, and the data requirements. Natural Language Processing (NLP), a prominent subfield of ML, offers significant potential to automate the extraction of information from unstructured data such as scientific publications, technical reports, and product specifications. For example, regression-based modeling combined with optimization can streamline LCA processes by automatically identifying critical indicators, potentially aiding early-stage definitions by efficiently identifying key variables [9,37]. Furthermore, recent studies underscore the utility of automation and NLP for managing unstructured data, improving the initial stages of LCA [32]. However, NLP-based approaches face challenges regarding the transparency of data sources and context interpretation, as large language models often leverage web-scraped information without clear provenance. Despite this potential, explicit applications of ML in the definition of goals and scope remain limited, presenting opportunities for future research to develop robust NLP models.

ML in life cycle inventory (LCI) analysis

The LCI stage, known for its high data intensity, has seen extensive use of ML to address data scarcity, improve data quality, and enable dynamic evaluations.

Data gap filling and imputation. Zhang et al. (2024) applied linear predictive models to estimate the intensity of embodied carbon in Chinese residential buildings. Through clustering techniques, they identified critical materials, significantly reducing data collection burdens [8]. Khadem et al. (2022) used optimized Artificial Neural Networks (ANNs) to predict greenhouse gas emissions in the Canadian fuel LCI database, greatly accelerating calculation times [19]. Similarly, Stemmle et al. (2021) developed parametric regression models with Monte Carlo simulations to dynamically assess the environmental impacts of Aquifer Thermal Energy Storage (ATES) [30]. Although these approaches reduce data collection efforts and improve computational efficiency, their accuracy is highly dependent on the representativeness of training data, limiting generalizability.

Data quality assessment. Ibn-Mohammed et al. (2023) used anomaly detection to efficiently identify and correct outliers and errors within LCI datasets, significantly enhancing the reliability of the results. However, careful threshold setting—which can be subjective—remains a critical challenge [18].

Data integration and harmonization. Zargar et al. (2022) reviewed NLP techniques for automating the harmonization of diverse textual data into standardized formats. Although this approach improves data comprehensiveness, robust NLP models are required to accurately manage various textual sources [17].

Parameter estimation and uncertainty quantification. Dai et al. (2022) demonstrated Gaussian process regression to predict nitrogen fertilizer usage, effectively addressing missing data and quantifying uncertainty. Bayesian methods, combined with probabilistic programming, offer rigorous uncertainty quantification but demand considerable computational resources and expertise [20].

Dynamic LCA. Kane and Miller (2024) developed predictive models for biomass pyrolysis, capturing the temporal dynamics of environmental impacts [22]. Ji et al. (2022) used regression models to estimate how much energy an electric vehicle would use during operation, which made dynamic LCA analyses easier. Such approaches effectively model temporal dynamics but rely heavily on detailed time-series data [38].

ML in life cycle impact assessment (LCIA)

Impact category prediction. Dai et al. (2022) employed Gaussian process regression for direct prediction of environmental impacts [20]. Although this approach improves prediction accuracy, its effectiveness depends on the quality of training data, and interpretability remains limited.

Characterization factor prediction. Marvuglia et al. (2015) demonstrated that ML can be used to predict toxicity characterization factors using the USEtox model [26]. This illustrates the potential to derive characterization factors through data-driven models; however, such applications remain rare and require further investigation to validate their accuracy.

ML in interpretation and decision support

Pattern recognition and result simplification. Elhami et al. (2022) applied ANN and Adaptive Neuro-Fuzzy Inference Systems (ANFIS) to identify influential agricultural inputs, facilitate targeted environmental interventions, and simplify decision-making [24]. However, these models require substantial domain-specific expertise for accurate interpretation [24,31].

Design optimization. Hassan et al. (2024) integrated ML with optimization techniques for sustainable polymer 3D printing design [31]. Liu et al. (2018) developed ML models to evaluate CO_2 emissions for highway construction, informing materials and maintenance strategies [39]. Zhang and Zhang (2021) applied a multi-objective genetic algorithm to a reinforced concrete building design, finding that a modest increase in construction cost (approximately 6%) could reduce life-cycle carbon emissions by nearly 15%, illustrating how ML-enabled optimization can explore cost–carbon trade-offs [40]. Manuguerra et al. (2024) proposed AI-assisted evaluation tools for electric vehicle designs, providing rapid environmental impact assessments [23,36]. Although effective, the interpretability of these models remains challenging.

Scenario analysis. Zhang et al. (2024) applied clustering and regression models to evaluate building carbon emissions under various scenarios, providing comprehensive decision support [8]. However, extensive scenario analyses entail substantial computational costs [33].

Explainable AI (XAI). Eddy et al. (2015) utilized surrogate modeling with Latin Hypercube sampling designs to enhance transparency in sustainability assessments [34,35]. Although Explainable AI significantly improves interpretability and credibility, interpreting complex ML outputs still requires specialized expertise.

Discussions

In this subsection, we explore the key challenges (and corresponding opportunities) of integrating ML into LCA, expanding upon those identified in earlier sections.

Data availability and quality

A fundamental challenge in ML applications within LCA is obtaining extensive, high-quality, and representative datasets for model training. Data scarcity, heterogeneity, and uncertainty significantly impact model accuracy and reliability. LCA datasets frequently suffer from data gaps and inconsistencies due to variations in data collection methods, regional differences, and incomplete documentation of processes [17,18,36]. Furthermore, uncertainty associated with measurement errors and subjective expert judgments can introduce significant bias into training datasets, affecting the accuracy and credibility of ML predictions [19,27,41]. Additionally, LCA data come in a variety of formats (e.g., EcoSpold, ILCD, JSON-LD), depending on the database or software used, which adds another layer of complexity to data harmonization [10].

Model generalizability and transferability

Developing ML models that generalize across different products, processes, or geographical regions remains a significant obstacle. Models trained on specific datasets, such as the Canadian fuel LCI database [19], often struggle to maintain predictive accuracy when applied outside of their initial scope. Transfer learning, which involves leveraging knowledge gained from one context to enhance model performance in another, presents a promising approach to improve generalizability. However, its effectiveness in LCA contexts remains underexplored and requires further research to evaluate its viability and potential limitations [8,38]. Additionally, validating ML models in LCA is challenging—supervised learning approaches require ground-truth values for training and testing, but many environmental impact measures lack a single accepted true value. In practice, model accuracy must often be inferred from proxy metrics or expert judgment, introducing uncertainty and limiting confidence in claims of improved predictive performance.

Model interpretability and transparency

The interpretability and transparency of ML models are critical to ensuring their adoption and trustworthiness in decision-making processes [25,42–44]. Complex models like neural networks and ensemble methods frequently function as “black boxes,” obscuring the underlying decision logic from users and stakeholders [35]. Consequently, there is a growing emphasis on the development and implementation of Explainable AI (XAI) techniques—such as surrogate models, feature importance analyses, and visualization tools—to elucidate ML model results. Enhancing transparency through XAI is essential to gain the trust of stakeholders and to enable informed decisions [35].

Integration with existing LCA frameworks and software

Integrating ML methodologies into existing LCA frameworks and software systems presents significant technical challenges. Current LCA software tools typically lack the ability to directly incorporate sophisticated ML algorithms, requiring additional software interfaces or dedicated platforms. This integration gap hinders the widespread adoption and practical implementation of ML-enhanced LCAs, highlighting the need for more robust and flexible integration solutions, standardized data formats, and improved compatibility between ML tools and traditional LCA software [21,30].

Computational cost

Training and deploying complex ML models require substantial computational resources, including high-performance computing infrastructure and extended processing times. The computational burden of sophisticated algorithms (such as deep learning and Bayesian methods) limits their applicability to routine or real-time evaluations, particularly in resource-constrained settings [20]. Moreover, the energy consumption and associated carbon emissions of training large ML models can partially offset their sustainability benefits if not managed with efficient hardware or renewable energy sources. Future research must focus on developing efficient algorithms and computational strategies that reduce resource demands while preserving model accuracy and predictive capabilities.

Lack of standardized methodologies

Currently, there are no widely accepted standardized methodologies or guidelines for integrating ML techniques within LCA processes. The absence of standards complicates comparisons between studies and reduces the reliability and reproducibility of ML-based LCA results. Establishing clear and universally applicable methodological standards and guidelines would significantly facilitate broader adoption, improve the comparability of results, and enhance methodological transparency and rigor in the LCA community [9,17,29,45].

Domain expertise

The effective application of ML within LCA requires close interdisciplinary collaboration between experts in the environmental assessment domain and ML specialists. The development and deployment of ML models demand both a deep understanding of environmental processes and proficiency in advanced computational methodologies. Consequently, fostering interdisciplinary collaborations is essential to successfully bridge knowledge gaps, improve model relevance, and ensure accurate interpretation and implementation of ML results within the practical context of LCA [24,31,46].

Conclusion

ML is already reshaping how LCA is planned, executed, and interpreted. First, a phase-aware reading of the literature reveals complementary strengths: NLP and knowledge graphs can clarify goal & scope; probabilistic imputation and learned harmonization strengthen LCI; surrogate and hybrid models accelerate LCIA without discarding interpretability; and calibrated, uncertainty-aware analytics improve interpretation for decisions. Second, progress depends on transparent data provenance, shareable pipelines, and consistent reporting of assumptions. Third, validation must go beyond point accuracy: proxy ground truth, out-of-domain splits, and calibration metrics (e.g., prediction-interval coverage, calibration curves, CRPS) make claims auditable. Fourth, computational cost and carbon should be reported (hardware, runtime, FLOPs) and managed (efficient training and low-carbon energy) so that modeling does not undermine sustainability goals. Fifth, we release open materials to support reuse and replication. Looking ahead, we see three priorities: (i) standard benchmarks and metadata schemas for LCA-ML tasks; (ii) robust transfer across regions/sectors and over time; and (iii) human-in-the-loop workflows that align model predictions with domain expertise and value choices. With these elements, ML-enabled LCA can become more efficient, reliable, and decision-oriented.

Author contributions

Conceptualization: Hairong Wang.

Formal analysis: Hairong Wang.

Methodology: Hairong Wang.

Validation: Hairong Wang.

Visualization: Hairong Wang.

Writing – original draft: Hairong Wang.

Writing – review & editing: Hairong Wang.

References

1. ISO (International Organization for Standardization). Environmental management—life cycle assessment—principles and framework (ISO 14040:2006); 2006. Note: ISO 14040:2006. 2006.
2. ISO (International Organization for Standardization). Environmental management—life cycle assessment—requirements and guidelines (ISO 14044:2006); 2006. Note: ISO 14044:2006. 2006.
3. Köck B, Friedl A, Serna Loaiza S, Wukovits W, Mihalyi-Schneider B. Automation of life cycle assessment—a critical review of developments in the field of life cycle inventory analysis. Sustainability. 2023;15(6):5531. https://doi.org/10.3390/su15065531
4. Ghoroghi A, Rezgui Y, Petri I, Beach T. Advances in application of machine learning to life cycle assessment: a literature review. Int J Life Cycle Assess. 2022;27(3):433–56. https://doi.org/10.1007/s11367-022-02030-3
5. Romeiko XX, Zhang X, Pang Y, Gao F, Xu M, Lin S, et al. A review of machine learning applications in life cycle assessment studies. Sci Total Environ. 2024;912:168969. https://doi.org/10.1016/j.scitotenv.2023.168969 PMID: 38036122
6. Martínez-Ramón N, Calvo-Rodríguez F, Iribarren D, Dufour J. Frameworks for the application of machine learning in life cycle assessment for process modeling. Cleaner Environmental Systems. 2024;14:100221. https://doi.org/10.1016/j.cesys.2024.100221
7. Bala A, Raugei M, Teixeira C, Fernández A, Pan-Montojo F, Fullana-i-Palmer P. Assessing the environmental performance of municipal solid waste collection: a new predictive LCA model. Sustainability. 2021;13(11):5810. https://doi.org/10.3390/su13115810
8. Zhang X, Sun J, Zhang X, Fenglai W. Statistical characteristics and scenario analysis of embodied carbon emissions of multi-story residential buildings in China. Sustainable Production and Consumption. 2024;46:629–40. https://doi.org/10.1016/j.spc.2024.03.019
9. Pascual-González J, Pozo C, Guillén-Gosálbez G, Jiménez-Esteller L. Combined use of MILP and multi-linear regression to simplify LCA studies. Computers & Chemical Engineering. 2015;82:34–43. https://doi.org/10.1016/j.compchemeng.2015.06.002
10. Salla JVE, de Almeida TA, Silva DAL. Integrating machine learning with life cycle assessment: a comprehensive review and guide for predicting environmental impacts. Int J Life Cycle Assess. 2025. https://doi.org/10.1007/s11367-025-02437-8
11. Hauschild MZ, Rosenbaum RK, Olsen SI. Life cycle assessment: theory and practice. Cham: Springer; 2018.
12. de Oliveira JA, da Luz Silva DAL, Pugliery FN, Saavedra YMB. Life cycle engineering and management of products: theory and practice. Cham: Springer; 2021.
13. Bhandari R, Kumar B, Mayer F. Life cycle greenhouse gas emission from wind farms in reference to turbine sizes and capacity factors. Journal of Cleaner Production. 2020;277:123385. https://doi.org/10.1016/j.jclepro.2020.123385
14. Das K, Behera RN. A survey on machine learning: concept, algorithms and applications. International Journal of Innovative Research in Computer and Communication Engineering. 2017;5:1301–9.
15. Brownlee J. How much training data is required for machine learning? 2019. https://machinelearningmastery.com/much-training-data-required-machine-learning/
16. Faceli K, Lorena AC, Gama J, Almeida TA, Carvalho ACPLF. Inteligência artificial: uma abordagem de aprendizado de máquina. 2nd ed. LTC; 2022.
17. Zargar S, Yao Y, Tu Q. A review of inventory modeling methods for missing data in life cycle assessment. J of Industrial Ecology. 2022;26(5):1676–89. https://doi.org/10.1111/jiec.13305
18. Ibn-Mohammed T, Mustapha KB, Abdulkareem M, Fuensanta AU, Pecunia V, Dancer CEJ. Toward artificial intelligence and machine learning-enabled frameworks for improved predictions of lifecycle environmental impacts of functional materials and devices. MRS Communications. 2023;13(5):795–811.
19. Khadem SA, Bensebaa F, Pelletier N. Optimized feed-forward neural networks to address CO2-equivalent emissions data gaps – Application to emissions prediction for unit processes of fuel life cycle inventories for Canadian provinces. Journal of Cleaner Production. 2022;332:130053. https://doi.org/10.1016/j.jclepro.2021.130053
20. Dai T, Jordaan SM, Wemhoff AP. Gaussian process regression as a replicable, streamlined approach to inventory and uncertainty analysis in life cycle assessment. Environ Sci Technol. 2022;56(6):3821–9. https://doi.org/10.1021/acs.est.1c04252 PMID: 35201765
21. Saad M, Zhang Y, Jia J, Tian J. Decision tree-based approach to extrapolate life cycle inventory data of manufacturing processes. J Environ Manage. 2024;360:121152. https://doi.org/10.1016/j.jenvman.2024.121152 PMID: 38759550

22. Kane S, Miller SA. Predicting biochar properties and pyrolysis life-cycle inventories with compositional modeling. Bioresour Technol. 2024;399:130551. https://doi.org/10.1016/j.biortech.2024.130551 PMID: 38458265
23. Tomic M, Micic R, Kiss F, Dedovic N, Simikic M. Economic and environmental performance of oil transesterification in supercritical methanol at different reaction conditions: experimental study with a batch reactor. Energy Conversion and Management. 2015;99:8–19. https://doi.org/10.1016/j.enconman.2015.04.010
24. Elhami B, Ghasemi Nejad Raeini M, Taki M, Marzban A, Heidarisoltanabadi M. Application of classic and soft computing for modeling yield and environmental final impact in vegetable production (a case study: transplanting onion in Isfahan province, Iran). Environ Sci Pollut Res Int. 2022;29(23):35314–37. https://doi.org/10.1007/s11356-022-18700-6 PMID: 35048351
25. Wright MM, Tan ECD, Tu Q, Martins A, Parvatker AG, Yao Y, et al. Life cycle inventory availability: status and prospects for leveraging new technologies. ACS Sustain Chem Eng. 2024;12(34):12708–18. https://doi.org/10.1021/acssuschemeng.4c02519 PMID: 40017908
26. Kobryn HT, Lantzke R, Bell R, Admiraal R. Remote sensing for assessing the zone of benefit where deep drains improve productivity of land affected by shallow saline groundwater. J Environ Manage. 2015;150:138–48. https://doi.org/10.1016/j.jenvman.2014.11.017 PMID: 25481499
27. Di Lullo G, Gemechu E, Oni AO, Kumar A. Extending sensitivity analysis using regression to effectively disseminate life cycle assessment results. Int J Life Cycle Assess. 2019;25(2):222–39. https://doi.org/10.1007/s11367-019-01674-y
28. Rivera JL, Sutherland JW. A design of experiments (DOE) approach to data uncertainty in LCA: application to nanotechnology evaluation. Clean Techn Environ Policy. 2014;17(6):1585–95. https://doi.org/10.1007/s10098-014-0890-9
29. Khang DS, Tan RR, Uy OM, Promentilla MAB, Tuan PD, Abe N, et al. Design of experiments for global sensitivity analysis in life cycle assessment: the case of biodiesel in Vietnam. Resources, Conservation and Recycling. 2017;119:12–23. https://doi.org/10.1016/j.resconrec.2016.08.016
30. Stemmle R, Blum P, Schüppler S, Fleuchaus P, Limoges M, Bayer P, et al. Environmental impacts of aquifer thermal energy storage (ATES). Renewable and Sustainable Energy Reviews. 2021;151:111560. https://doi.org/10.1016/j.rser.2021.111560
31. Hassan M, Misra M, Taylor GW, Mohanty AK. A review of AI for optimization of 3D printing of sustainable polymers and composites. Composites Part C: Open Access. 2024;15:100513. https://doi.org/10.1016/j.jcomc.2024.100513
32. Tushar Q, Zhang G, Bhuiyan MA, Navaratnam S, Giustozzi F, Hou L. Retrofit of building façade using precast sandwich panel: an integrated thermal and environmental assessment on BIM-based LCA. Buildings. 2022;12(12):2098. https://doi.org/10.3390/buildings12122098
33. Belucio M, Rodrigues C, Antunes CH, Freire F, Dias LC. Eco-efficiency in early design decisions: a multimethodology approach. Journal of Cleaner Production. 2021;283:124630. https://doi.org/10.1016/j.jclepro.2020.124630
34. Mabey CS, Dickerson TJ, Salmon JL, Mattson CA. An approach for predicting social, environmental, and economic product impacts and navigating the associated impact trade-space in engineering design. In: Proceedings of the ASME 2023 International Design Engineering Technical Conferences and Computers and Information in Engineering Conference (IDETC-CIE 2023). 2023.
35. Eddy DC, Krishnamurty S, Grosse IR, Wileden JC, Lewis KE. A predictive modelling-based material selection method for sustainable product design. Journal of Engineering Design. 2015;26(10–12):365–90. https://doi.org/10.1080/09544828.2015.1070258
36. Manuguerra L, Cappelletti F, Germani M. A machine learning based method for parametric environmental impact model for electric vehicles. Journal of Cleaner Production. 2024;454:142308. https://doi.org/10.1016/j.jclepro.2024.142308
37. Zhou Y, Ma M, Tam VWY, Le KN. Design variables affecting the environmental impacts of buildings: a critical review. Journal of Cleaner Production. 2023;387:135921. https://doi.org/10.1016/j.jclepro.2023.135921
38. Nguyen Q, Diaz-Rainey I, Kitto A, McNeil BI, Pittman NA, Zhang R. Scope 3 emissions: data quality and machine learning prediction accuracy. PLOS Clim. 2023;2(11):e0000208. https://doi.org/10.1371/journal.pclm.0000208
39. Liu Y, Wang Y, An D. Life-cycle CO2 emissions and influential factors for asphalt highway construction and maintenance activities in China. International Journal of Sustainable Transportation. 2018;12(7):497–509. https://doi.org/10.1080/15568318.2017.1402108
40. Zhang Y, Zhang X. Low-carbon design optimization for reinforced concrete buildings using multi-objective machine learning. Journal of Cleaner Production. 2021;279:123735.

41. Marzban A, Elhami B, Bougari E. Integration of life cycle assessment (LCA) and modeling methods in investigating the yield and environmental emissions final score (EEFS) of carp fish (Cyprinus carpio) farms. Environ Sci Pollut Res Int. 2021;28(15):19234–46. https://doi.org/10.1007/s11356-020-12116-w PMID: 33394451

42. MacLeod MJ, Vellinga T, Opio C, Falcucci A, Tempio G, Henderson B, et al. Invited review: a position on the Global Livestock Environmental Assessment Model (GLEAM). Animal. 2018;12(2):383–97. https://doi.org/10.1017/S1751731117001847 PMID: 28789724

43. Li P, Chen B, Cui Q. A probabilistic life-cycle assessment of carbon emission from magnesium phosphate cementitious material with uncertainty analysis. Journal of Cleaner Production. 2023;427:139164. https://doi.org/10.1016/j.jclepro.2023.139164

44. Manan A, Pu Z, Weiyi C, Ahmad J, Alattyih W, Umar M, et al. Machine learning prediction of recycled concrete powder with experimental validation and life cycle assessment study. Case Studies in Construction Materials. 2024;21:e04053. https://doi.org/10.1016/j.cscm.2024.e04053

45. Arias A, Feijoo G, Moreira MT. How could Artificial Intelligence be used to increase the potential of biorefineries in the near future? A review. Environmental Technology & Innovation. 2023;32:103277. https://doi.org/10.1016/j.eti.2023.103277

46. Ramachandran R, McShea MJ, Howson SN, Burkom HS, Chang H-Y, Weiner JP, et al. Assessing the value of unsupervised clustering in predicting persistent high health care utilizers: retrospective analysis of insurance claims data. JMIR Med Inform. 2021;9(11):e31442. https://doi.org/10.2196/31442 PMID: 34592712