frontiers in Sustainability

TYPE Review
PUBLISHED 13 August 2026
DOI 10.3389/frsus.2026.1854730

OPEN ACCESS

EDITED BY
Martí Rufí-Salís,
2.-0 LCA Consultants, Barcelona, Spain

REVIEWED BY
Joan Muñoz-Liesa,
Autonomous University of Barcelona, Spain
Mehran Akrami,
University of Virginia, United States
Hairong Wang,
Carnegie Mellon University Stores, United States

*CORRESPONDENCE
Nils Pauliks
✉ n.pauliks@cml.leidenuniv.nl

RECEIVED 13 April 2026
REVISED 17 July 2026
ACCEPTED 21 July 2026
PUBLISHED 13 August 2026

CITATION
Pauliks N, Donati F, Weber JM and Steubing B (2026) Machine learning in prospective LCA: insights from a systematic literature review. *Front. Sustain.* 7:1854730. doi: 10.3389/frsus.2026.1854730

Machine learning in prospective LCA: insights from a systematic literature review

Nils Pauliks[1]*, Franco Donati[1], Jana Marie Weber[2] and Bernhard Steubing[1]

[1]Institute of Environmental Sciences (CML), Leiden University, Leiden, Netherlands, [2]Department of Intelligent Systems, Delft University of Technology, Delft, Netherlands

Machine learning (ML) offers considerable opportunities for advancing life cycle assessment (LCA), yet a consolidated overview of ML models with potential for prospective LCA (pLCA) is still missing. This systematic review identifies ML models suited to pLCA's data needs and assesses the transferability of retrospective LCA models to pLCA. We apply keyword-based and automated forward/backward searches to identify studies based on algorithm type, training datasets, input/output variables, and performance metrics. We identify 50 publications and classify them into four ML application groups: (1) streamlined LCA, (2) life cycle inventory data prediction, (3) impact category prediction, and (4) characterization factor prediction. Among these, predicting life cycle inventory (LCI) data and characterization factor estimation are directly applicable to pLCA. Life cycle inventory data prediction and streamlined LCA represent the largest body of literature, offering benefits such as estimating missing unit process data, leveraging large external datasets, advanced simulation of process parameters, identifying molecular substitutes, or the employment of rapid screening tools. While characterization factor prediction shows potential, it requires further research for practical implementation. Overall, ML is rarely applied directly to pLCA. However, many ML models developed for retrospective LCA have the potential to be transferred to pLCA, either directly or with adaptations to data structures. This is most evident for predicting missing or technology-specific life cycle inventory data, where similarity-based learning, simulation-enhanced models, and large datasets offer scalable ways to estimate future inventory parameters. Across all groups, uncertainty is insufficiently assessed, with only a minority of studies addressing model, parameter, or scenario uncertainty. Overall, ML holds potential for pLCA when applied with methodological care and attention to underlying uncertainties.

KEYWORDS
artificial intelligence, emerging technologies, machine learning, predict, prospective LCA, systematic review

1 Introduction

Machine learning (ML) is widely recognized for its potential to advance analytical capabilities for researchers across diverse scientific disciplines (Horvitz and Mulligan, 2015). ML algorithms can automatically acquire knowledge, make decisions, predict outcomes, and adapt to changing conditions, allowing systems to improve their performance without explicit programming. As such, it is defined as a subdomain

of artificial intelligence, among others like automated reasoning or natural language processing. Within ML, there are three classifications: supervised, unsupervised, and reinforcement learning. Each typology represents a different way of learning from data. Supervised learning uses labeled data to predict outputs, unsupervised learning finds patterns in unlabeled data, and reinforcement learning learns through rewards and penalties (Samoili et al., 2020). Note that we use the term *prediction* to refer to the output of an ML model and to remain consistent with ML terminology; it does not refer to predictive scenarios for a future point in time.

These powerful statistical tools can become particularly valuable to life cycle assessment (LCA) practitioners, as numerous systematic reviews in the scientific literature show (Akrami et al., 2025; Baehr et al., 2024; Ghoroghi et al., 2022; Nwagwu et al., 2025; Romeiko et al., 2024; Salla et al., 2025; Wang, 2025). Here, unsupervised and supervised algorithms are identified as commonly employed in LCA (Romeiko et al., 2024), while reinforcement learning is less common. Furthermore, most of the ML models in LCA are used for regression and classification tasks (Salla et al., 2025). General applications include decision-making, impact prediction, and literature review applications (Ghoroghi et al., 2022), as well as to enable optimization, the prediction of unknown life cycle inventory (LCI) data, and the estimation of characterization factors (Romeiko et al., 2024). Other use cases employ ML for inventory modeling (Martínez-Ramón et al., 2024), use large language models to automate a manufacturer's Bill of Materials to LCIs (Castle et al., 2025), employ data mining from patents (Spreafico et al., 2023), make use of sensor data (Algren et al., 2021), and apply simulation tools (Naser et al., 2023). A full overview of existing reviews can be found in the Supplementary material 1.

All aforementioned literature reviews and examples use ML models in retrospective LCA studies, where product systems are modeled using inventory data that reflect earlier points in time. This is often because updated inventories are unavailable (Arvidsson et al., 2018, 2024). Prospective LCA (pLCA), on the other hand, aims to assess the environmental impacts of emerging technologies or long-term effects of established product systems or services (Buyle et al., 2019; Thonemann et al., 2020). Thus, pLCA introduces unique data challenges that do not exist in retrospective LCA. Examples include expert elicitation, upscaling, temporal changes in the background system, unsuitable characterization factors for new substances, and increased uncertainty due to future modeling (Cucurachi and Blanco, 2023; Van Der Giesen et al., 2020). A full overview of the unique challenges of pLCA can be found in the Supplementary material 1. Due to these different data requirements, it remains unclear which ML approaches are most suitable to be used in pLCA, as they may require a different ML model selection than retrospective LCA.

Limited attention is given in the literature to identifying suitable ML models for use in pLCA and to assessing whether ML models applied in retrospective LCA apply to pLCA. To bridge this gap, we conduct a systematic literature review exploring the use of ML in pLCA. We aim to identify which ML approaches hold potential for tackling pLCA challenges, discuss their associated limitations, and assess the transferability of ML models from retrospective LCA to pLCA. Ultimately, this study seeks to support pLCA practitioners and researchers by synthesizing insights from existing case studies and thereby facilitating the potential adoption of ML in pLCA.

2 Methods

2.1 Systematic literature review

To comprehensively evaluate ML approaches suitable for pLCA, a systematic literature review following the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) method is conducted. The PRISMA method is a guideline for conducting systematic literature reviews and meta-analyses (Page et al., 2021).

We conducted keyword searches in the Scopus database, combining the most prevalent pLCA terms: "*prospective*", "*ex-ante*", and "*anticipatory*" (Arvidsson et al., 2018), along with the phrase "*predictive life cycle assessment*". For ML coverage, we use keywords defined in the European Commission taxonomy for Artificial Intelligence in the domain of ML (Samoili et al., 2020). The full list of keywords can be found in the Supplementary material 2.

To identify further literature in this heterogeneous field, we apply automated forward and backward searches on the platform (Litmaps, n.d.), using the obtained list of literature from Scopus as an input. Litmaps applies the open-access metadata repositories (Crossref, n.d.; Semantic Scholars, n.d.) and (OpenAlex, n.d.) to identify citation-based relationships of the uploaded list of publications.

2.1.1 Selection process

Papers are included if the authors employ an ML model suitable to generate data for one of the previously mentioned pLCA challenges (Supplementary Table S2), like LCI data availability in prospective assessments or unsuitable characterization factors for future assessments. When authors apply ML models retrospectively but explicitly state that these models can also be used in pLCA, we included the study in our literature selection. If no such statements are provided, we evaluate the transferability of retrospective ML models independently. Hence, in the absence of an explicit statement, the assessment of transferability unavoidably depends on our expert judgment. We acknowledge this and other limitations of our study. A full overview of the limitations can be found in the Supplementary material 1. Reviews and theoretical discussions about ML usage in pLCA are excluded, as no algorithm can be attributed to an identified study.

We include peer-reviewed journal articles and conference papers with an implied peer-review process, published in English up to September 30, 2024. The combined search results from Scopus and Litmaps are deduplicated based on digital object identifiers and title- and year-matching, using an initial automated procedure followed by a manual verification step. We then screen each record by reviewing the title, abstract, keywords, and, when necessary, the methodology section, to determine whether a study meets the inclusion criteria. Any additional references identified in the retrieved literature that meet the inclusion criteria are added to

the dataset. Publications that do not meet the criteria are excluded. All screening decisions, including inclusion and exclusion, are made by the first author. The full track record of paper selection, whether pLCA or retrospective LCA is employed, and which study is identified through which selection mechanism, is provided in the Supplementary material 2.

2.1.2 Classification and data collection

Studies are categorized by mapping their ML applications to the corresponding LCA phases defined in ISO 14040 (International Organization for Standardization, 2006). Although this framework describes standard LCA methodology, which is typically retrospective, pLCA is still considered an LCA methodology (Arvidsson et al., 2018). For this reason, we apply the same four-phase framework to pLCA.

To summarize the identified studies meaningfully, we create groups of studies representing the purpose of the employed ML algorithm. This process yields four overarching groups: *streamlined LCA*, *predicting LCI data*, *characterization factor prediction*, and *impact category prediction*. Algorithms in *streamlined LCA* couple technology-specific information with environmental impacts, aiming to directly predict the environmental impacts of a prospective technology based on available information at the design stage. Models within the *predicting LCI data* group are designed to predict LCI data, either directly or indirectly, with subsequent calculation steps, using information available at the early stages of product design. ML approaches in *characterization factor prediction* link chemical descriptors to characterization factors, to include uncharacterized substances into newly calculated characterization

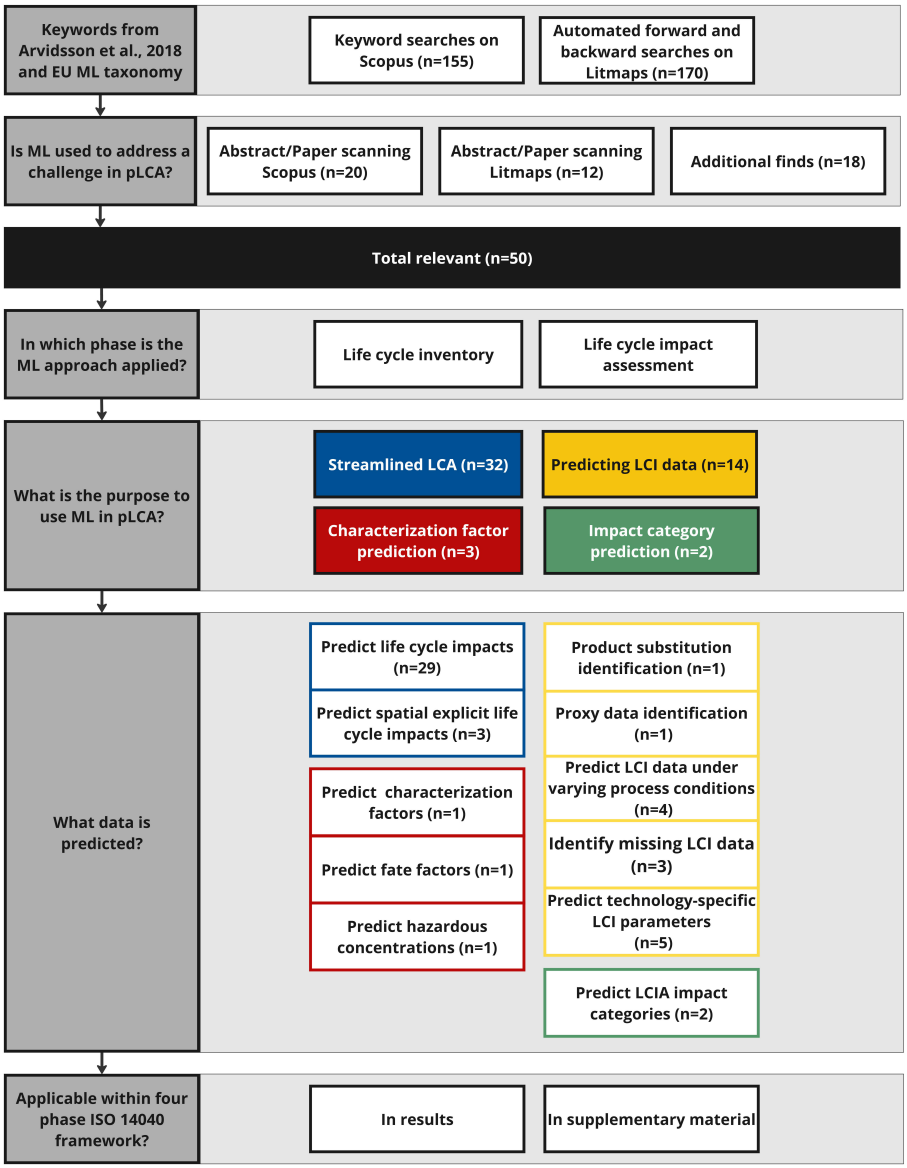

FIGURE 1
Flow chart of the systematic review, the automated back and forward searches, the additional identified papers, and the grouping process.

factors. The group *impact category prediction* aims to streamline the life cycle impact assessment (LCIA) phase and use statistical relationships from one calculated impact category to predict other impact categories. Because the ML models in each group target different types of data for the same overall objective, we further divide each group into subgroups. However, not always can multiple examples be found per subgroup; in this case, we refer to such subgroups as proof-of-concept cases.

The *streamlined LCA* group contains two subgroups: *predict life cycle impacts* and *predict spatially explicit life cycle impacts*. These models are used across the LCI and LCIA phase. The *predicting LCI data* group comprises three subgroups only in the LCI phase: *predicting LCI data under varying process conditions*, *identifying missing LCI data*, and *predicting technology-specific LCI parameters,* as well as two proof-of-concept cases, *product substitution identification* and *proxy data identification.* The group *characterization factor prediction* includes three proof-of-concept cases placed in the LCIA phase: *predict characterization factors*, *predict fate factors*, and *predict hazardous concentrations*. Finally, the *impact category prediction* group placed in the LCIA phase contains a single subgroup: *predict LCIA impact categories*.

For each selected study, the training set, the input variables, the target variables, the model performance indicator, and the type of algorithm are collected. Additionally, the list includes a description of the ML algorithm in pLCA, the functional unit, the system boundaries, the target LCA matrices, whether uncertainty was addressed, and the field of application. Data is presented in part in section 3.4 and in full in the Supplementary material 2.

Finally, because this paper aims to support pLCA practitioners specifically, we highlight only those ML models in the results that offer direct applicability to pLCA practitioners. Studies that do not directly support the practitioners' workflow, for example, those replacing the four-phase ISO 14040 framework with an ML-model, are moved to the Supplementary material 1, where they are presented in full. This does not imply that these studies lack value for pLCA practitioners or product designers; for example, they may support rapid screening tasks. Our emphasis is rather on ML models that have more potential to be integrated in the pLCA practitioners' workflow, while still ensuring that all groups are covered with an equivalent level of detail. The complete systematic review process, including study selection and categorization, is shown in Figure 1.

3 Results

Our search yields 155 publications from Scopus, 170 from Litmaps, and 18 identified through manual citation tracking. Of these, 50 papers satisfy the selection criteria. Publication volumes have grown over time, and the majority of relevant publications are published between 2015 and 2024. The most common fields of application across all studies are organic chemicals ($n = 16$), followed by the built environment ($n = 10$), agriculture ($n = 7$), and chemical engineering ($n = 7$). The underlying data in the supporting information shows that both main authors and co-authors, as well as their affiliated institutions, span a broad and diverse set of research groups, which strengthens the robustness of the findings. An overview of publications over time and the respective fields of applications can be seen in Figure 2.

3.1 Rationale for applying the ML models

We identify four distinct categories that capture the different rationales for employing ML models, as described in Chapter 2.1.2. The groups and subgroups are shown in Figure 3, with regard to their specific LCA phase(s). Furthermore, the identified approaches of ML in relation to pLCA are shown. All identified models are found in the LCI and the LCIA phase; no ML model is found to generate data to support the goal and scope or the interpretation phase. *Streamlined LCA* and *characterization factor estimation* span across two phases, while *predicting LCI data* and *impact category prediction* are only applied within one LCA phase. Some subgroups use ML to directly generate data required for pLCA, while others predict parameters that need additional processing steps before they can be used in a pLCA.

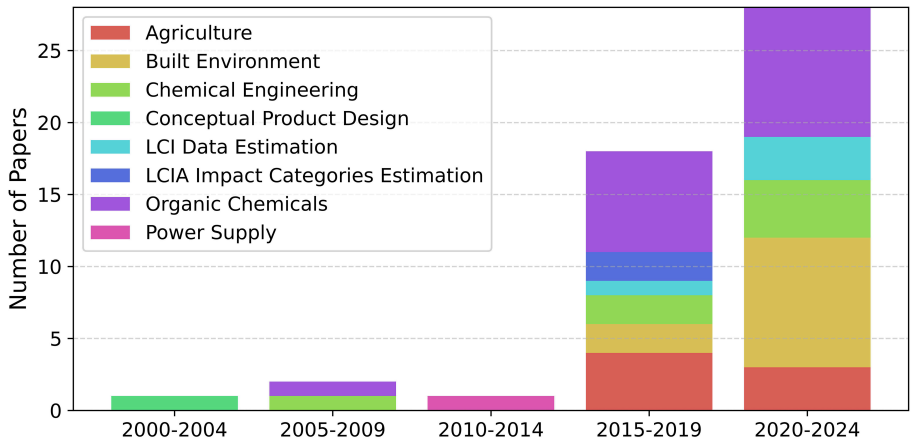

FIGURE 2
ML applications identified as relevant for pLCA data generation over time and including their field of application.

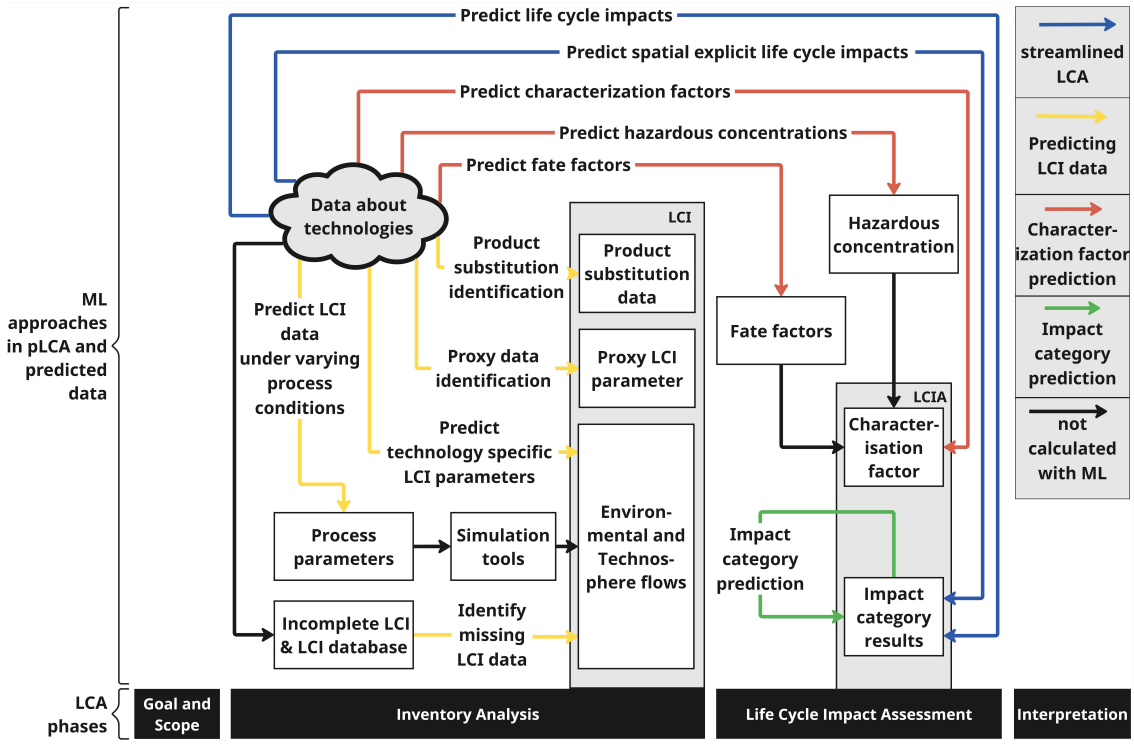

FIGURE 3
Overview of LCA applications from the literature on how ML is applied in pLCA and retrospective LCA with relevance for pLCA. Each arrow represents one of the identified subgroups, for streamlined LCA and Impact category prediction only the groups are presented.

Streamlined LCA and *Impact category prediction* replace the LCA model with an ML model, i.e., they skip the conventional LCA model entirely (LCI and LCIA calculations) and estimate instead LCIA results from the input variables. This creates the possibility to rapidly screen potential environmental impacts or to gain an understanding of how impact categories are related to each other. However, both groups partly replace the four-phase framework. Hence, these two groups are moved to the Supplementary material 1 for a detailed description. However, it is noteworthy to mention that *streamlined LCA* represents the largest body of identified literature.

3.2 Predicting life cycle inventory data

A major challenge in pLCA is the collection of data needed to generate LCIs for prospective technologies, which are often available only at laboratory or pilot scale. Identifying suitable substitutes for processes or products is difficult, particularly when the prospective technology does not fulfill the same function as existing alternatives. In addition, such technologies may operate under varying process conditions, and manufacturing technologies for them may not exist yet, further limiting the availability of reliable data. We identified the following five ML models, which aim to improve data collection for incomplete LCIs of emerging technologies.

3.2.1 Product substitution identification

The identification of product substitutes is often challenging. Substitutes are expected to act as a replacement for the existing product; hence, they must fulfill the same or a very similar function as the previous one, but with substantially reduced environmental impacts. We identified a proof-of-concept where (Zhu et al., 2020) conduct a screening assessment to identify a substitute for the molecule sitagliptin. The substitute should fulfill the same function as the sitagliptin, but with a less environmentally polluting supply chain. Substitution in this case may be possible because molecules with similar structures often exhibit similar compound activities. The authors conduct a screening assessment of the chemical substance database PubChem, identifying substitutes with known chemical descriptors using the Python ML package Rdkit. In a second step, a principal component analysis is used to reduce the number of correlated variables. The latter, therefore, helps to reduce the number of parameters required to identify suitable alternatives for sitagliptin.

After identifying potential substitutes to the molecule sitagliptin, the authors introduce a streamlined LCA by using a deep neural network, which is why we grouped this application initially under *streamlined LCA*. The authors speculate that this approach could be used to reduce the environmental impacts of similar chemical production processes and to support future sustainable development.

3.2.2 Proxy data identification

The authors of (Ayagapin et al., 2021) conduct an LCA for 34 different single-family houses, a representative sample of the built environment of the geographic scope of the study. ML is used in the form of a principal component analysis to standardize the calculated impact category results. The input variables consist of seven midpoint impact categories for the construction, the use, and the end-of-life phases, resulting in 21 variables in total. They then apply *k*-means clustering algorithms to identify the houses whose environmental impacts are most representative of the existing building stock. In this way, the authors show how clustering algorithms can be used to identify a data proxy, in this case, three single-family house types as proxies for the existing building stock.

3.2.3 Predict LCI data under varying process conditions

In this subgroup, artificial neural networks and multi-linear regression are applied to predict process parameters under varying process conditions. Inputs are technical product parameter like water-cement ratio or specified biomass input compositions. ML is used to predict process parameters like yields, environmental emissions, and material composition shares, only from the inputs. The predicted parameters are, in the following step, fed into simulation tools to generate corresponding mass and energy requirements of the modeled process (Liao et al., 2020; Meng et al., 2019; Tanhadoust et al., 2023; Zainul et al., 2024). This modeled data can serve as flow or process-level information of the LCI data of any technology where only limited LCI data is available.

3.2.4 Identifying missing LCI data

Decision trees, random forest, gradient boosting, adaptive boosting, similarity learning, and eXtreme gradient boosting are used to predict missing economic and environmental flows in unit processes by identifying similarities to other analogous processes in LCI databases. Using collected inventory data, the algorithms estimate economic and environmental flow quantities based on statistically similar unit processes within the LCI database (Hou, 2019; Saad et al., 2024; Zhao et al., 2021). With that, missing data can be approximated based on similar LCIs from LCI databases.

3.2.5 Predict technology-specific LCI parameters

Authors in this subgroup predict building lifetimes, agricultural machinery fuel quantities, machinery usage hours, fertilizer quantities and emission intensities of coal power plants at different locations from available statistics or other larger datasets. The authors use autoregressive integrated moving average, neural networks, multiple regression and boosting algorithms trained on input parameters like building and demolishing statistics, historic fuel and usage machinery data, plant specific fertilizer use and emission intensities of power plants (Dai et al., 2022; Ji et al., 2023, 2021; Ma and Kim, 2015; Steinmann et al., 2014). Hence, this subgroup summarizes the prediction of technology-specific life-cycle parameters from related available large datasets.

3.3 Characterization factor prediction

All models in this group focus on predicting characterization factors. Three proof-of-concept cases are identified, each with a distinct prediction scope. Across all studies, molecular-specific descriptors are used as inputs to the respective ML algorithms. All identified studies apply retrospective approaches, and none addresses the adjustment of characterization factors for future scenarios.

3.3.1 Predict characterization factors

A challenge in pLCA is to obtain characterization factors of novel substances or adjustments of already existing ones. Servien et al. (2022) link molecular descriptors of molecules directly to characterization factors. They use a partial least squares algorithm, a random forest algorithm, and a support vector machine. The authors train the models on molecular descriptors in the UseTox database and existing characterization factors for freshwater ecotoxicity and human toxicity. The models learn the relationship between molecular descriptors to characterization factors to estimate characterization factors of non-characterized substances. However, the authors emphasize the proof-of-concept stage of this research.

3.3.2 Predict fate factors

In this proof-of-concept case, the authors of (Marvuglia et al., 2015) link molecular descriptors to fate factors, describing the chemical's behavior in the environment, including physical solubility and the breakdown over time. The authors use an artificial neural network to link nine molecular descriptors of substances in the UseTox database to fate factors for human toxicity and freshwater ecotoxicity research. Fate factors can be used in exposure models to calculate characterization factors. The authors, however, mention the further need for research in this area before applicability to other fields is given.

3.3.3 Predict hazardous concentrations

In this proof-of-concept case study, (Hou et al., 2020) use molecular descriptors of chemical substances to predict hazardous concentrations (HC_{50}) and thereby fill gaps in missing experimental data. This is done with the following models: k-nearest neighbor, support vector machines, artificial neural network, random forest, AdaBoost, gradient boosting machine.

From the predicted concentrations, they calculate characterization factors and compare them to existing values from USEtox. The resulting model demonstrates that ML can support with simulation of missing data.

3.4 Overview of employed algorithms, training data and input & target variables

Overall, the authors of the groups *predicting LCI data* and *characterization factor estimation* utilized a wide variety of different ML models. We find both supervised and unsupervised algorithms, even though supervised algorithms predominate. Supervised algorithms learn from labeled data. The labels are learned by the algorithms based on input-output pairs and then predicted for unknown instances. An example from the subgroup *identify missing LCI data*, the authors used decision-tree supervised ML algorithms that link elementary and intermediate flows to purposely removed flows from the same dataset. Another example from the subgroup *predict characterization factors* illustrates supervised learning. The authors link molecular descriptors to existing characterization factors for freshwater ecotoxicity using algorithms such as partial least squares, random forests, and support vector machines.

Unsupervised algorithms learn from unlabeled data. Here, clustering can be applied to identify groups of similar unlabeled data, which then allows the user to make assumptions based on the identified groups. For example, in the subgroup *proxy data identification*, the authors apply k-means clustering to group existing buildings into proxy categories, with the intention of selecting samples that best represent the remaining dataset. However, it remains uncertain whether the identified clusters truly capture the full variability of the underlying data.

TABLE 1 Summary of the collected data in the Supplementary material, consisting of training dataset, input and target variables, as well as the applied algorithms per group and subgroup.

Group	Subgroup	Training dataset	Input variables	Target variables	Applied algorithms	# of studies
Predict LCI data	Product substitution identification	224 datasets each containing 125 molecular and 25 LCIA indicators	125 molecular descriptors	25 LCIA mid- and endpoint indicators	Artificial neural network	1
	Proxy data identification	21 LCIA indicators	not applicable	3 proxy datapoints out of 34	principal component analysis, *K*-means clustering	1
	Predict LCI data under varying process conditions	72–245 datasets containing of 9–168 datapoints	2–17 technical product specific parameters	process parameters	artificial neural networks, multi linear regression	4
	Identify missing LCI data	315—all Ecoinvent unit processes	collected LCI data for the product or biosphere flows	missing technological or biosphere exchanges in unit process	eXtreme Gradient Boosting, random forest, similarity learning, decision tree, adaptive boosting	3
	Predict technology-specific LCI parameters	52-971,514 datasets containing 1–22 datapoints	technology specific parameters like location, design features	technology specific life cycle parameters	exponential smoothing, autoregressive integrated moving average, artificial neural networks, linear regression, eXtreme Gradient Boosting, Light gradient boosting machine, multi linear regression, gaussian regression	5
Character-ization factor prediction	Predict characterization factors	274 datasets	273 molecular descriptors	USEtox characterization factor	partial least squares, random forest, support vector machines	1
	Predict fate factors	2307 datasets	14 parameters from CompTox	HC 50 values of chemical	*k*-nearest neighbor, support vector machines, artificial neural network, random forest, AdaBoost, gradient boosting machine	1
	Predict hazardous concentrations	3073 datasets	13 USEtox output factors	fate factor of chemical	artificial neural network	1

As the pLCA applications are diverse, we see both classification and regression tasks in the reviewed literature. Specific algorithm types are amongst other k-means-based methods, artificial neural networks, tree-based methods, and ensemble methods. Notably, we did not encounter recent architectures such as foundational models, multi-agent systems, geometric deep learning, or generative algorithms; nor did we see any approaches using reinforcement learning or studies with "humans-in-the-loop". This is not unexpected in a field in which quality data in sufficient quantity is cumbersome to collect.

Training datasets vary greatly in size across groups and subgroups. We observe large datasets, especially in subgroups that focus on the prediction of technology-specific variables. Examples are the models in the subgroups *predict LCI data under varying process conditions* and *predict technology-specific LCI parameters*. Furthermore, the predicted target variables show that the ML models are employed mostly as highly specialist models, designed and trained to predict data for a clear use case. Larger, more generalist ML models are not employed, with the exception of ML models in the subgroup *identify missing LCI data*. The model is trained on the Ecoinvent database and represents a model that can easily be applied in other pLCA studies. The model only requires the collected LCI data of the assessed technology to predict the missing LCI data. The data in Table 1 represents a summary of the algorithm-specific data; other collected parameters on model performance, system boundaries, and the fields of application can be found in the Supplementary material 2 under 'details of reviewed studies'.

4 Discussion

Overall, we observe that a distinct focus on pLCA is rare in the reviewed literature, with only two studies explicitly stating that they conduct a pLCA (Ayagapin et al., 2021; Karka et al., 2022). Other studies do not clearly distinguish between retrospective LCA and pLCA as defined by (Arvidsson et al., 2018). Consequently, they do not explicitly use the term pLCA, even though they apply ML for prospective purposes such as early-stage design interventions (Feng et al., 2019; Fozer et al., 2023; Karka et al., 2019; Kleinekorte et al., 2023, 2019; Park et al., 2001; Płoszaj-Mazurek et al., 2020; Zhu et al., 2020) or the incorporation of future climate scenarios (Lee et al., 2020). The remaining studies do not adopt a pLCA perspective, indicating that ML, for the time period considered, is not widely integrated into pLCA research.

However, we identify significant potential for applying ML models from retrospective LCA to pLCA. Transferability is particularly strong when pLCA and retrospective LCA challenges overlap, as existing models can be used in pLCA without adjustment. This is evident in the group *LCI parameter prediction*, especially in the subgroups *identify missing LCI data* and *predict technology-specific LCI parameters*.

Similarly, *streamlined LCA* models trained on retrospective data can, in times, directly serve as rapid environmental impact screening tools for new molecules or technology parameters. This function is particularly helpful at the design stage of a product to quickly assess the potential environmental impacts of a design or a new molecule. Examples include the studies of (Kleinekorte et al., 2019; Song et al., 2017; Zhang et al., 2024). However, it is important to mention that reliable results depend on the similarity of new molecules or materials to the training data (Kretschmer et al., 2025). The model is therefore only supportive when sufficient proximity to already assessed molecules or technologies exists.

Other examples identified as suitable for use in pLCA require adaptations in their data structure or the combination with additional tools. Hence, their applicability largely depends on whether pLCA practitioners find the use of ML promising enough to chose ML aided modeling approaches. The group *LCI parameter prediction* stands out as particularly promising, as these models can be seamlessly integrated into the existing workflows of pLCA scholars. Additionally, it is the most well-represented group, with numerous subgroups and proof-of-concept cases. *Characterization factor estimation* through ML prediction also represents a suitable application; however, the limited number of case studies calls for more explorative research before it can be employed in pLCA. The following section explores the transferability of retrospective LCA cases to pLCA in detail.

4.1 Methodological implementations and limitations for ML usage in pLCA

4.1.1 Models to predict similar unit process data

Predicting economic and environmental flows, based on collected LCIs of the target technology and analogous background processes, offers a scalable method to fill gaps in inventory data, particularly when new technologies evolve from existing ones. The case studies (Hou et al., 2018; Saad et al., 2024; Zhao et al., 2021) offer a directly applicable approximation method for pLCA, if the missing inventory data is closely related to existing technologies.

We speculate that a similarity-based learning approach could be combined with predicted LCI background data from integrated assessment tools, for example, through the tool *premise* (Sacchi et al., 2022). This would allow the pLCA practitioner to estimate missing unit process data of predicted data. A comparable approach is adopted by (Lee et al., 2020): They combine climate projections derived from an integrated assessment model, specifically following the IPCC's Representative Concentration Pathways, with detailed time-series data on agricultural practices. Using an eXtreme Gradient Boosting algorithm, they then predicted the life cycle impacts of farming at the county level across the United States.

A combined use of premise and ML is particularly interesting, as the structure of the projected inventory data does not change (Sacchi et al., 2022), which is beneficial for the ML algorithm. However, similarity-based flow predictions proved most effective in cases where approximately 95% of the inventory data had already been collected (Hou et al., 2018), which limits the identified approach to cases where most inventory data is already available.

4.1.2 Enhanced simulation tools

ML also enables the prediction of technology-specific parameters under varying conditions. The ML predictions makes use of physical properties of the technology to predict hard-to-obtain technology parameters, like the yield of a production process. This modeling is comparable with the use of simulation tools like Aspen Plus. At times, ML models are also used in combination with simulation software (Zainul et al., 2024).

Additionally, to the direct applicability of predicting technology-specific parameters of a (novel) technology, we see potential to use such ML models to support upscaling challenges. Models could be trained to estimate inventory data depending on the specific feedstock, raw material composition, or energy requirements, at different production scaling steps. Examples are shown in the literature by (Martínez-Ramón et al., 2024). Such models could reduce the need for expensive prototypes or assist in pre-screen upscale experiments.

4.1.3 Use of large datasets

The ML models in the subgroups *predict technology-specific LCI parameters* and *product substitution identification* show how ML can be used to make use of large datasets. We see particular use cases in the chemical and pharmaceutical sectors, where ML and large databases can help identify structurally similar compound molecules that could replace environmentally harmful ones (Zhu et al., 2020). PLCA practitioners may also use the initial screening of substances independently from the presented streamlined approach to identify shortlisted chemical substitutes.

Additionally, the use of large datasets to predict individual LCI parameters, with examples in the literature of building lifetimes (Ji et al., 2023, 2021) or agricultural machinery (Ma and Kim, 2015) parameters, represents the use of additional data sources outside commonly applied LCI databases. While large datasets are unlikely to exist for emerging technologies, pLCAs focusing on the future development of an existing technology could benefit from such possibilities. An additional large and beneficial dataset for pLCA with the support of ML could be the use of data mining from patent data (Spreafico et al., 2025).

4.1.4 Characterization factor prediction

All identified studies in the group *prediction of characterization factors* present ML models for incorporating currently uncharacterized substances into toxicity impact assessment models. Here, the models learn relationships between molecular descriptors and characterization factors (Servien et al., 2022), fate factors (Marvuglia et al., 2015), or hazardous concentrations of substances in the environment (Hou et al., 2020). For novel molecules, such models could help to estimate their characterization factors; however, only if the new molecules are similar enough to the ones in the training set (Kretschmer et al., 2025).

Artificial neural networks, k-nearest neighbor, and random forest algorithms, among others shows a high prediction performance in this domain; however, the lack of explicit focus on prospective molecule characterization and only three identified proof-of-concept cases implementation into pLCA is not yet given. Furthermore, the described studies do not account for temporal changes in environmental conditions. The latter is important because the environmental impacts of substances depend on the specific conditions of the environment in which they occur (Vázquez-Vázquez et al., 2025). When these conditions change, the resulting environmental impacts change as well, which then in turn require adopted characterization factors.

The unknown future environmental conditions and the reliance on previously established toxicity characterization make the prediction difficult to interpret. Limitations arise as many key input parameters simply lack enough measured data to train reliable models (Von Borries et al., 2023). We therefore see limited direct applicability of such ML predictions to support pLCA in the short term.

4.1.5 Streamlined LCA

For product developers wanting to rapidly screen environmental impacts of new technologies early on in the development phase, *streamlined LCA* offers a quick way to estimate potential impacts, saving time and resources compared to full pLCA studies. A chemist could, for example, quickly estimate the potential environmental impacts of a new molecule. Such rapid screening models are particularly relevant to the chemical industry (Blanco et al., 2024) and are in recent studies, increasingly trained on large industrial datasets (Gao et al., 2026; Zhang et al., 2024). Some of the listed approaches are directly applicable for pLCA scholars, for example, to support early-stage assessments of different process-route alternatives (Kleinekorte et al., 2023, 2019) or alternative manufacturing set-ups (Karka et al., 2022, 2019). While models are trained with retrospective impact category results and built on identified statistical relationships between descriptors and target variables, more recent approaches also investigate end-to-end learning—a technique where descriptors are also learned by the model (Gao et al., 2026).

A limiting factor for the use of some of the models is that it can be very difficult to understand why the model predicts a target variable. Thus, especially neural-network and deep-learning approaches are often referred to as "black-box models" (Barbierato and Gatti, 2024). Consequently, the resulting predictions of environmental emissions of such models do not permit contribution analyses, assessments of data uncertainty, and introduce their own model uncertainties. Streamlined LCA predictions also show a lack of accuracy in the predicted environmental impacts compared to the use of process calculations, process simulations, or stoichiometry-based approaches (Parvatker and Eckelman, 2019). We therefore consider *streamlined LCA* models useful as rapid early-stage screening tools, but not as substitutes for a full pLCA.

4.1.6 Impact category prediction

Impact category prediction aims to replace the LCIA phase with ML approaches, which adds to a long discussion of converging

results of several impact categories, mostly because of the use of fossil fuels in the supply chain (Huijbregts et al., 2010, 2006). We argue that ML approaches, which build on statistical relationships between different impact categories, can, in the best case, deliver identical results to the existing characterization models. The two studies in this group, however, cannot, due to their architecture, improve or exceed the accuracy of existing characterization factors in pLCA, which is why we have not considered them further.

4.2 Uncertainty and study appraisal

Performance indicators, such as the coefficient of determination (R^2) or root mean square error (RMSE), evaluate an ML model's prediction accuracy by comparing the predicted target variables to a test dataset. We report all available model performance indicators of the identified studies in the Supplementary material 2. Yet, these metrics are only valid for comparing models trained on the same dataset (Gosiewska et al., 2022). Thus, model performance indicators are sometimes used to indirectly account for model uncertainty by comparing models trained on the same dataset (Romeiko et al., 2020). Model performance indicators, however, cannot assess parameter or scenario uncertainty and should therefore be distinguished from uncertainty assessments (Baehr et al., 2024).

Model uncertainty originates from model structure, implicit assumptions, or limitations of the model itself, while parameter uncertainty comes from variability, measurement and reporting errors, or data gaps. Scenario uncertainty originates from normative choices or assumptions for future predictions or extrapolations (Huijbregts et al., 2003). While model and parameter uncertainty are frequently mentioned in the literature, only 14 of the 50 identified studies explicitly assess or mitigate any type of uncertainty. Model uncertainty is often addressed through the employment of multiple ML models, out of which the best-performing is selected. But models are not systematically chosen to distinguish between different model uncertainties, and thus, the inherent model uncertainty remains unquantified. For parameter uncertainty, a common approach is to calculate confidence intervals of the input data and use these metrics to include uncertainty measures in the ML models, e.g., through bootstrapping (Fleitmann et al., 2021; Karka et al., 2022). Parameter uncertainty originating from assumptions, modeling choices, or LCIA quantification is rarely discussed or addressed within the reviewed literature.

The limited attention to uncertainty assessment in the reviewed studies aligns with findings from other reviews on ML applications in retrospective LCA (Baehr et al., 2024; Romeiko et al., 2024; Wang, 2025). When uncertainty is assessed, predicted results can change substantially (Akrami et al., 2025). Model predictions are typically validated against available data using train-test splits, *k*-fold, or nested cross-validation; however, validation is rarely conducted from a scenario-narrative perspective. Moreover, few studies publish their training data or model code, which limits reproducibility and hinders a fully transparent understanding for the reader.

To improve the assessment of parameter uncertainty in ML models for pLCA, time-based data splits can be employed for algorithmic benchmarking (Yao et al., 2022). This approach could be particularly valuable for technology-specific parameters, such as evolving energy demands of a technology over time. Yet, they provide only limited insight into an algorithm's ability to capture future, potentially disruptive developments. Bayesian models could be used to quantify and propagate parameter uncertainty (Akrami et al., 2025), but only if reliable uncertainty parameters can be generated for the input variables. As these input variables represent parameter estimates, often in a first-of-a-kind study, reliable distribution ranges remain a challenge (Jouannais et al., 2024). To reduce the uncertainty arising from using individual models, ensemble models can be employed (Wang, 2025).

Generally, assumptions in ML predictions must be communicated transparently, and readers should be able to scrutinize them. Moreover, the use of ML and any potential mispredictions should be carefully documented, just as in any well-defined scenario- or narrative-based future exploration assessment. ML in pLCA should therefore be applied cautiously, primarily for parameters with a strong, evidence-based track record and validation through proven case studies to predict future parameters.

4.3 Future research outlook

Future research should place explicit emphasis on ML use in pLCA to investigate whether challenges arise that differ from those described in this article. This is particularly evident for the group of *characterization factor estimation*. The work should additionally extend beyond the current focus on freshwater toxicity impacts and include additional impact categories.

Although the group *predicting LCI data* represents the largest share of the identified literature, more research is needed to apply a genuinely prospective focus across all subgroups, especially for the various proof-of-concept cases. The subgroup *predict technology-specific LCI parameters* provide promising tools, yet these approaches remain highly tailored to specific technologies. Future work should therefore explore how more generalizable ML models can be developed that extend beyond a single technological context. One avenue to rapidly estimate model parameters could be to train ML models specifically on simulation-tool outputs across different production stages.

For the subgroup *predict similar unit process data*, simulation-based learning could be combined with projected future datasets, for example, those generated through *premise*. This would allow the investigation of long-term trends embedded in modified LCI databases and reveal patterns that are difficult to detect using conventional statistical analysis.

5 Conclusion

This study addressed the underexplored intersection of ML and pLCA through a systematic literature review. We identify

50 publications and categorize them into four distinct ML application groups: *predicting LCI data, characterization factor estimation, streamlined LCA*, and *LCIA estimation*. We identify that *predicting LCI data* and *characterization factor estimation* are the most directly applicable to pLCA workflows. Within these groups, ML is leveraged to simulate technology-specific parameters, estimate missing LCI data, and predict parameters from large datasets. Additionally, we identified individual proof-of-concept cases that support chemical compound substitution and estimate proxy data. Within the characterization factor estimation group, ML models support the direct prediction of characterization factors, fate factors, or hazardous concentrations of chemicals in the environment.

Despite its potential, ML remains rarely directly applied to pLCA, revealing a significant gap in its adoption. Moreover, our findings suggest that ML models from retrospective LCA can often be transferred to pLCA, either directly or with adaptations to data structures, offering a promising pathway for future applications.

More studies with an explicit prospective focus are needed. Future research should particularly focus on characterization factor estimation beyond freshwater toxicity and across all impact categories. For technology-specific LCI parameters, generalizable models trained on simulation outputs could enable broader applicability, e.g., to address upscaling challenges in a specific sector like the chemical industry. Additionally, combining simulation-based learning with projected datasets, e.g., from integrated assessment models, could uncover long-term trends in LCI databases, offering insights beyond traditional statistical methods. Overall we see large application potential of ML in pLCA, if the models are employed with care and awareness of existing uncertainties in the data and the models.

Author contributions

NP: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Software, Validation, Visualization, Writing – original draft, Writing – review & editing. FD: Conceptualization, Project administration, Supervision, Validation, Visualization, Writing – review & editing. JW: Project administration, Supervision, Validation, Visualization, Writing – review & editing. BS: Project administration, Supervision, Validation, Visualization, Writing – review & editing.

Funding

The author(s) declared that financial support was received for this work and/or its publication. This work was partly supported by a fellowship of the German Academic Exchange Service (DAAD); and the Circular Circuits project (P20-13) funded by the Dutch Research Council (www.nwo.nl).

Acknowledgments

We sincerely thank Carlos Felipe Blanco Rocha for his thorough insights and constructive feedback. We also extend our gratitude to Christina Drotenko and Niklas Engberg for their valuable comments and suggestions.

Conflict of interest

The author(s) declared that this work was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Generative AI statement

The author(s) declared that Generative AI was used in the creation of this manuscript. During the preparation of this work, the author(s) used Mistral AI - Le Chat in order to improve language and readability. After using this tool/service, the author(s) reviewed and edited the content as needed and take full responsibility for the content of the publication.

Any alternative text (alt text) provided alongside figures in this article has been generated by Frontiers with the support of artificial intelligence and reasonable efforts have been made to ensure accuracy, including review by the authors wherever possible. If you identify any issues, please contact us.

Publisher's note

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/frsus.2026.1854730/full#supplementary-material

References

Akrami, M., Porter, M. D., and Colosi, L. M. (2025). Addressing uncertainty in machine learning-integrated life cycle assessment (ML+LCA). *J. Environ. Manage.* 389:126225. doi: 10.1016/j.jenvman.2025.126225

Algren, M., Fisher, W., and Landis, A. E. (2021). "Machine learning in life cycle assessment," in *Data Science Applied to Sustainability Analysis* (Amsterdam: Elsevier), 167–190. doi: 10.1016/B978-0-12-817976-5.00009-7

Arvidsson, R., Svanström, M., Sandén, B. A., Thonemann, N., Steubing, B., and Cucurachi, S. (2024). Terminology for future-oriented life cycle assessment: review and recommendations. *Int. J. Life Cycle Assess.* 29, 607–613. doi: 10.1007/s11367-023-02265-8

Arvidsson, R., Tillman, A. M., Sandén, B. A., Janssen, M., Nordelöf, A., Kushnir, D., et al. (2018). Environmental assessment of emerging technologies: recommendations for prospective LCA. *J. Ind. Ecol.* 22, 1286–1294. doi: 10.1111/jiec.12690

Ayagapin, L., Praene, J. P., Jaggeshar, D., and Surroop, D. (2021). Prospective life cycle assessment: effect of electricity decarbonization in building sector. *Energies*. 14:17. doi: 10.3390/en14113184

Baehr, J., Koyamparambath, A., Dos Reis, E., Weyand, S., Binnig, C., Schebek, L., et al. (2024). Predicting product life cycle environmental impacts with machine learning: uncertainties and implications for future reporting requirements. *Sustain. Prod. Consum.* 52, 511–526. doi: 10.1016/j.spc.2024.11.005

Barbierato, E., and Gatti, A. (2024). The challenges of machine learning: a critical review. *Electronics* 13:416. doi: 10.3390/electronics13020416

Blanco, C. F., Pauliks, N., Donati, F., Engberg, N., and Weber, J. (2024). Machine learning to support prospective life cycle assessment of emerging chemical technologies. *Curr. Opin. Green Sustain. Chem.* 50:100979. doi: 10.1016/j.cogsc.2024.100979

Buyle, M., Audenaert, A., Billen, P., Boonen, K., and Passel, S. V. (2019). The future of ex-ante LCA? Lessons learned and practical recommendations. *Sustainability* 11:5456. doi: 10.3390/su11195456

Castle, S., Schneider, J. M., Hennig, L., and Rehm, G. (2025). Entity linking using LLMs for automated product carbon footprint estimation. doi: 10.48550/ARXIV.2502.07418

Crossref (n.d.). Crossref [WWW Document]. Crossref. Available online at: https://www.crossref.org/ (accessed 13 May 2024).

Cucurachi, S., and Blanco, C. F. (2023). *Practical Solutions for Ex-Ante LCA Illustrated by Emerging PV Technologies*. Lausanne: EPFL.

Dai, T., Jordaan, S. M., and Wemhoff, A. P. (2022). Gaussian process regression as a replicable, streamlined approach to inventory and uncertainty analysis in life cycle assessment. *Environ. Sci. Technol.* 56, 3821–3829. doi: 10.1021/acs.est.1c04252

Feng, K., Lu, W., and Wang, Y. (2019). Assessing environmental performance in early building design stage: an integrated parametric design and machine learning method. *Sustain. Cities Soc.* 50:101596. doi: 10.1016/j.scs.2019.101596

Fleitmann, L., Kleinekorte, J., Leonhard, K., and Bardow, A. (2021). COSMO-susCAMPD: sustainable solvents from combining computer-aided molecular and process design with predictive life cycle assessment. *Chem. Eng. Sci.* 245:116863. doi: 10.1016/j.ces.2021.116863

Fozer, D., Nimmegeers, P., Toth, A. J., Varbanov, P. S., Klemeš, J. J., Mizsey, P., et al. (2023). Hybrid prediction-driven high-throughput sustainability screening for advancing waste-to-dimethyl ether valorization. *Environ. Sci. Technol.* 57, 13449–13462. doi: 10.1021/acs.est.3c01892

Gao, Q., Schulze Balhorn, L., Laera, A., Meys, R., Goßen, J., Weber, J. M., et al. (2026). Environmental impacts prediction using graph neural networks on molecular graphs. *Comput. Chem. Eng.* 204:109362. doi: 10.1016/j.compchemeng.2025.109362

Ghoroghi, A., Rezgui, Y., Petri, I., and Beach, T. (2022). Advances in application of machine learning to life cycle assessment: a literature review. *Int. J. Life Cycle Assess.* 27, 433–456. doi: 10.1007/s11367-022-02030-3

Gosiewska, A., Woznica, K., and Biecek, P. (2022). Interpretable meta-score for model performance. *Nat. Mach. Intell.* 4, 792–800. doi: 10.1038/s42256-022-00531-2

Horvitz, E., and Mulligan, D. (2015). Data, privacy, and the greater good. *Science.* 349, 253–255. doi: 10.1126/science.aac4520

Hou, P. (2019). *Data-Driven Environmental System Analysis: Addressing Data Gaps in Life Cycle Assessment* (Dissertation). The University of Michigan.

Hou, P., Cai, J., Qu, S., and Xu, M. (2018). Estimating missing unit process data in life cycle assessment using a similarity-based approach. *Environ. Sci. Technol.* 52, 5259–5267. doi: 10.1021/acs.est.7b05366

Hou, P., Jolliet, O., Zhu, J., and Xu, M. (2020). Estimate ecotoxicity characterization factors for chemicals in life cycle assessment using machine learning models. *Environ. Int.* 135:105393. doi: 10.1016/j.envint.2019.105393

Huijbregts, M. A. J., Gilijamse, W., Ragas, A. M. J., and Reijnders, L. (2003). Evaluating uncertainty in environmental life-cycle assessment. a case study comparing two insulation options for a dutch one-family dwelling. *Environ. Sci. Technol.* 37, 2600–2608. doi: 10.1021/es020971+

Huijbregts, M. A. J., Hellweg, S., Frischknecht, R., Hendriks, H. W. M., Hungerbühler, K., and Hendriks, A. J. (2010). Cumulative energy demand as predictor for the environmental burden of commodity production. *Environ. Sci. Technol.* 44, 2189–2196. doi: 10.1021/es902870s

Huijbregts, M. A. J., Rombouts, L. J. A., Hellweg, S., Frischknecht, R., Hendriks, A. J., Van De Meent, D., et al. (2006). Is cumulative fossil energy demand a useful indicator for the environmental performance of products? *Environ. Sci. Technol.* 40, 641–648. doi: 10.1021/es051689g

International Organization for Standardization (2006). *Environmental management—Life cycle assessment—Principles and framework (ISO 14040:2006).* Available online at: https://www.iso.org/standard/37456.html

Ji, S., Lee, B., Cho, Y., and Yi, M. Y. (2023). Effect of realistically estimated building lifespan on life cycle assessment: a case study in Korea. *J. Build. Eng.* 75:107028. doi: 10.1016/j.jobe.2023.107028

Ji, S., Lee, B., and Yi, M. Y. (2021). Building life-span prediction for life cycle assessment and life cycle cost using machine learning: a big data approach. *Build. Environ.* 205:108267. doi: 10.1016/j.buildenv.2021.108267

Jouannais, P., Blanco, C. F., and Pizzol, M. (2024).ENvironmental Success under Uncertainty and Risk (ENSURe): a procedure for probability evaluation in ex-ante LCA. *Technol. Forecast. Soc. Change* 201:123265. doi: 10.1016/j.techfore.2024.123265

Karka, P., Papadokonstantakis, S., and Kokossis, A. (2019). Environmental impact assessment of biomass process chains at early design stages using decision trees. *Int. J. Life Cycle Assess.* 24, 1675–1700. doi: 10.1007/s11367-019-01591-0

Karka, P., Papadokonstantakis, S., and Kokossis, A. (2022). Digitizing sustainable process development: from ex-post to ex-ante LCA using machine-learning to evaluate bio-based process technologies ahead of detailed design. *Chem. Eng. Sci.* 250:117339. doi: 10.1016/j.ces.2021.117339

Kleinekorte, J., Kleppich, J., Fleitmann, L., Beckert, V., Blodau, L., and Bardow, A. (2023). Appropriate life cycle assessment: a PROcess-specific, predictive impact assessment method for emerging chemical processes. *ACS Sustain. Chem. Amp Eng.* 11, 9303–9319. doi: 10.1021/acssuschemeng.2c07682

Kleinekorte, J., Kröger, L., Leonhard, K., and Bardow, A. (2019). A neural network-based framework to predict process-specific environmental impacts. *Comput. Aided Chem. Eng.* 46, 1447–1452. doi: 10.1016/B978-0-12-818634-3.50242-3

Kretschmer, F., Seipp, J., Ludwig, M., Klau, G. W., and Böcker, S. (2025). Coverage bias in small molecule machine learning. *Nat. Commun.* 16:554. doi: 10.1038/s41467-024-55462-w

Lee, E. K., Zhang, W.-J., Zhang, X., Adler, P. R., Lin, S., Feingold, B. J., et al. (2020). Projecting life-cycle environmental impacts of corn production in the U.S. Midwest under future climate scenarios using a machine learning approach. *Sci. Total Environ.* 714:136697. doi: 10.1016/j.scitotenv.2020.136697

Liao, M., Kelley, S., and Yao, Y. (2020). Generating energy and greenhouse gas inventory data of activated carbon production using machine learning and kinetic based process simulation. *ACS Sustain. Chem. Eng.* 8, 1252–1261. doi: 10.1021/acssuschemeng.9b06522

Litmaps (n.d.). *Litmaps | Your Literature Review Assistant.* Available online at: https://www.litmaps.com/ (Accessed October 21, 2024).

Ma, J., and Kim, H. M. (2015). Predictive usage mining for life cycle assessment. *Transp. Res. Part Transp. Environ.* 38, 125–143. doi: 10.1016/j.trd.2015.04.022

Martínez-Ramón, N., Calvo-Rodríguez, F., Iribarren, D., and Dufour, J. (2024). Frameworks for the application of machine learning in life cycle assessment for process modeling. *Clean. Environ. Syst.* 14:100221. doi: 10.1016/j.cesys.2024.100221

Marvuglia, A., Kanevski, M., and Benetto, E. (2015). Machine learning for toxicity characterization of organic chemical emissions using USEtox database: learning the structure of the input space. *Environ. Int.* 83, 72–85. doi: 10.1016/j.envint.2015.05.011

Meng, F., LaFleur, C., Wijesinghe, A., and Colvin, J. (2019). Data-driven approach to fill in data gaps for life cycle inventory of dual fuel technology. *Fuel* 246, 187–195. doi: 10.1016/j.fuel.2019.02.124

Naser, A. Z., Defersha, F., Pei, E., Zhao, Y. F., and Yang, S. (2023). Toward automated life cycle assessment for additive manufacturing: a systematic review of influential parameters and framework design. *Sustain. Prod. Consum.* 41, 253–274. doi: 10.1016/j.spc.2023.08.009

Nwagwu, C. C., Ogorodnyk, O., Sølvsberg, E., Eleftheriadis, R. J., and Meskers, C. (2025). Integrating artificial intelligence into life cycle assessment: a framework for balancing automation and human expertise. *J. Sustain. Metall.* 11, 3590–3605. doi: 10.1007/s40831-025-01305-x

OpenAlex (n.d.). *OpenAlex: The Open Catalog to the Global Research System* [WWW Document]. Available online at: https://openalex.org/ (accessed May 13, 2024).

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., et al. (2021). The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *BMJ* 372:n71. doi: 10.1136/bmj.n71

Park, J.-H., Seo, K.-K., and Wallace, D. (2001). "Approximate life cycle assessment of classified products using artificial neural network and statistical analysis in conceptual product design," in *Proceedings - 2nd International Symposium on Environmentally Conscious Design and Inverse Manufacturing* (Tokyo), 321–326. doi: 10.1109/ECODIM.2001.992373

Parvatker, A., and Eckelman, M. (2019). Comparative evaluation of chemical life cycle inventory generation methods and implications for life cycle assessment results. *ACS Sustain. Chem. Eng.* 7, 350–367. doi: 10.1021/acssuschemeng.8b03656

Płoszaj-Mazurek, M., Ryńska, E., and Grochulska-Salak, M. (2020). Methods to optimize carbon footprint of buildings in regenerative architectural design with the use of machine learning, convolutional neural network, and parametric design. *Energies* 13:5289. doi: 10.3390/en13205289

Romeiko, X. X., Guo, Z., Pang, Y., Lee, E. K., and Zhang, X. (2020). Comparing machine learning approaches for predicting spatially explicit life cycle global warming and eutrophication impacts from corn production. *Sustainability* 12:1481. doi: 10.3390/su12041481

Romeiko, X. X., Zhang, X., Pang, Y., Gao, F., Xu, M., Lin, S., et al. (2024). A review of machine learning applications in life cycle assessment studies. *Sci. Total Environ.* 912:168969. doi: 10.1016/j.scitotenv.2023.168969

Saad, M., Zhang, Y., Jia, J., and Tian, J. (2024). Decision tree-based approach to extrapolate life cycle inventory data of manufacturing processes. *J. Environ. Manage.* 360:121152. doi: 10.1016/j.jenvman.2024.121152

Sacchi, R., Terlouw, T., Siala, K., Dirnaichner, A., Bauer, C., Cox, B., et al. (2022). Prospective environmental impact assement (premise): a streamlined approach to producing databases for prospective life cycle assessment using integrated assessment models. *Renew. Sustain. Energy Rev.* 160:112311. doi: 10.1016/j.rser.2022.112311

Salla, J. V. E., De Almeida, T. A., and Silva, D. A. L. (2025). Integrating machine learning with life cycle assessment: a comprehensive review and guide for predicting environmental impacts. *Int. J. Life Cycle Assess.* 30, 2423–2445. doi: 10.1007/s11367-025-02437-8

Samoili, S., López Cobo, M., Gómez, E., De Prato, G., Martínez-Plumed, F., and Delipetrev, B. (2020). *AI Watch. Defining Artificial Intelligence. Towards an Operational Definition and Taxonomy of Artificial Intelligence*, EUR 30117 EN. Luxembourg: Publications Office of the European Union. doi: 10.2760/382730,JRC118163

Semantic Scholars (n.d.). Semantic Scholar | Publishers [WWW Document]. Available online at: https://www.semanticscholar.org/about/publishers (Accessed May 13, 2024).

Servien, R., Latrille, E., Patureau, D., and Hélias, A. (2022). Machine learning models based on molecular descriptors to predict human and environmental toxicological factors in continental freshwater. *Peer Community J.* 2. doi: 10.24072/pcjournal.90

Song, R., Keller, A. A., and Suh, S. (2017). Rapid life-cycle impact screening using artificial neural networks. *Environ. Sci. Technol.* 51, 10777–10785. doi: 10.1021/acs.est.7b02862

Spreafico, C., Landi, D., and Russo, D. (2023). A new method of patent analysis to support prospective life cycle assessment of eco-design solutions. *Sustain. Prod. Consum.* 38, 241–251. doi: 10.1016/j.spc.2023.04.006

Spreafico, C., Thonemann, N., Pizzol, M., Arvidsson, R., Steubing, B., Cucurachi, S., et al. (2025). Using patents to support prospective life cycle assessment: opportunities and limitations. *Int. J. Life Cycle Assess.* 30, 201–220. doi: 10.1007/s11367-024-02404-9

Steinmann, Z. J. N., Venkatesh, A., Hauck, M., Schipper, A. M., Karuppiah, R., Laurenzi, I. J., et al. (2014). How to address data gaps in life cycle inventories: a case study on estimating CO_2 emissions from coal-fired electricity plants on a global scale. *Environ. Sci. Technol.* 48, 5282–5289. doi: 10.1021/es500757p

Tanhadoust, A., Emadi, S. A. A., Nasrollahpour, S., Dabbaghi, F., and Nehdi, M. L. (2023). Optimal design of sustainable recycled rubber-filled concrete using life cycle assessment and multi-objective optimization. *Constr. Build. Mater.* 402:132878. doi: 10.1016/j.conbuildmat.2023.132878

Thonemann, N., Schulte, A., and Maga, D. (2020). How to conduct prospective life cycle assessment for emerging technologies? A systematic review and methodological guidance. *Sustainability* 12:1192. doi: 10.3390/su12031192

Van Der Giesen, C., Cucurachi, S., Guinée, J., Kramer, G. J., and Tukker, A. (2020). A critical view on the current application of LCA for new technologies and recommendations for improved practice. *J. Clean. Prod.* 259:120904. doi: 10.1016/j.jclepro.2020.120904

Vázquez-Vázquez, B., Lazzari, M., and Hospido, A. (2025). Terrestrial characterization factors for bio- and fossil-based plastics: microplastics ingestion and additives release. *Waste Manag.* 196, 106–114. doi: 10.1016/j.wasman.2025.02.008

Von Borries, K., Holmquist, H., Kosnik, M., Beckwith, K. V., Jolliet, O., Goodman, J. M., et al. (2023). Potential for machine learning to address data gaps in human toxicity and ecotoxicity characterization. *Environ. Sci. Technol.* 57, 18259–18270. doi: 10.1021/acs.est.3c05300

Wang, H. (2025). Integrating machine learning into life cycle assessment: review and future outlook. *PLoS Clim.* 4:e0000732. doi: 10.1371/journal.pclm.0000732

Yao, H., Choi, C., Cao, B., Lee, Y., Koh, P. W., and Finn, C. (2022). Wild-time: a benchmark of in-the-wild distribution shift over time. *Adv. Neural Inf. Process. Syst.* 35, 10309–10324. doi: 10.48550/arXiv.2211.14238

Zainul, R., Basem, A., Jasim, D. J., Yadav, A., Hasson, A. R., Logroño, J. P. A., et al. (2024). Predictive analysis of methane-enhanced conversion of organic waste into sustainable fuel: a machine learning approach. *Process Saf. Environ. Prot.* 189, 1428–1449. doi: 10.1016/j.psep.2024.06.129

Zhang, D., Wang, Z., Oberschelp, C., Bradford, E., and Hellweg, S. (2024). Enhanced deep-learning model for carbon footprints of chemicals. *ACS Sustain. Chem. Eng.* 12, 2700–2708. doi: 10.1021/acssuschemeng.3c07038

Zhao, B., Shuai, C., Hou, P., Qu, S., and Xu, M. (2021). Estimation of unit process data for life cycle assessment using a decision tree-based approach. *Environ. Sci. Technol.* 55, 8439–8446. doi: 10.1021/acs.est.0c07484

Zhu, X., Ho, C. H., and Wang, X. (2020). Application of life cycle assessment and machine learning for high-throughput screening of green chemical substitutes. *ACS Sustain. Chem. Eng.* 8, 11141–11151. doi: 10.1021/acssuschemeng.0c02211