



A machine learning based method for parametric environmental impact model for electric vehicles

Luca Manuguerra^{*}, Federica Cappelletti, Michele Germani

Università Politecnica delle Marche, Faculty of Engineering, Department of Industrial Engineering and Mathematical Sciences, Via Breccia Bianche 12, 60131, Ancona, Italy

ARTICLE INFO

Handling Editor: Tomas B. Ramos

Keywords:

Machine learning
Data science
Environmental sustainability
Eco-design
Life cycle assessment

ABSTRACT

Environmental managers attempt to increasingly incorporate precautionary principles into decision making. The literature lacks Machine Learning -based approaches for forewarning lifecycle environmental impacts. This paper proposes a method to support electric vehicle design. The main innovation of the work lays in merging Life Cycle Assessment (LCA) and Machine Learning foundations to provide support and awareness to designers. The present approach overcomes the present literature because it provides a method for the design phase, is based on a well-established methodology (LCA) and provides quantitative results from little inputs. The approach exploits machine Learning Methods to develop models with the design features of a generic electric vehicle (such as vehicle mass and distance traveled) in six phases (Problem definition; Data collection; Data Preparation; Modeling; Model evaluation; Model interpretation). Differently from existing environmental analyses, all stages of the product life cycle have been considered in building the database; moreover, the model provides quantitative results. Regression models and supervised algorithms were used. The obtained model can be used by product engineers, as well as those not experts on LCA. Moreover, the model guarantees the database and hypothesis's uniqueness, ensuring the results coherence and comparability. The level of accuracy obtained in the case study (error or 17%) is comparable with studies handling full environmental analysis (that should be more accurate), and outstanding, as the present case is for the design phase. Future works will focus on additional significative indicators, similar electric vehicle design and integration with prospective LCA approaches.

1. Introduction

Leading the focus from improving the physical manufacturing processes to creating a digital representation of the physical processes to get better insights (Dalzochio et al., 2020), the fourth industrial revolution (14.0) leveraged advancements in production facilities through the inclusion of big data analysis, information technology, Internet of Things (IoT) and Additive Manufacturing (AM) (Ivanov et al., 2021). Automated and interconnected machinery and many digital technologies brought a consistent contribution to process optimization; however, their construction may require critical materials and their use phase much electricity. The design process is critical in enabling an improvement of their performances along their lifecycle and early design decisions can have a very significant impact on sustainability. Design processes are complex and conducting a Life Cycle Assessment (LCA) is also a complex task. Merging these two processes with proper theoretical knowledge will lead to the successful integration of LCA into early

product design stages (Ostad Ahmad Ghorabi et al., 2009). Machine Learning (ML) is one area of data science that can be used to fill the gaps in environmental analysis and therefore support the design process.

1.1. Design, environmental assessment, and electric vehicles

Assessing the viability of a product should extend beyond traditional design considerations like operational performance and costs to include an evaluation of its environmental effects throughout its life cycle. This requires the design team to gain the ability to evaluate the environmental performance of different concepts at the initial stages of the design process (Sousa and Wallace, 2006), going beyond geometrical and technical aspects (Mandolini et al., 2019). Various tools, including the widely acknowledged LCA, have been created to aid designers in making well-informed decisions.

This study aims to develop a predictive model, enabling the analysis of the environmental impact of Electric Vehicles (EVs) during the initial

^{*} Corresponding author.

E-mail address: l.manuguerra@pm.univpm.it (L. Manuguerra).

design phase. EVs, being a subject of significant environmental concern, warrant thorough investigation. Nordelöf et al. (2014) conducted a review of 79 works, exploring the utility of different types of LCA studies on EVs to provide comprehensive stakeholder information. They categorized studies into Well-To-Wheel (WTW), complete LCAs, and battery LCAs. Their results depict a strong interest in Greenhouse Gas Emissions (GHG) and light passenger vehicles, as confirmed by the wide availability of comparisons with combustion engine cars (Petrauskienė et al., 2020).

Despite the attention given to EV environmental impact, a clear gap exists in methodologies supporting the design phase. Existing literature primarily concentrates on the use phase or battery life, neglecting considerations of materials, maintenance, and End-of-Life (EoL) phases (Raugei et al., 2015; Burchart-Korol et al., 2018). For a tool to be effective in the early design stage, it must handle limited data and provide consistent results.

Existing studies, such as Raugei et al.'s comprehensive LCA-based comparison of lightweight options for compact passenger vehicles (Raugei et al., 2015), or Petrauskienė et al.'s reliance on market-released product data, address specific aspects but fall short in guiding the early design phase (Petrauskienė et al., 2020). The gap persists as studies related to EV environmental impact lack methods tailored for the initial design stages when data is scarce. This hinders designers from making well-informed decisions and evaluating the goodness of their choices before changes become impractical. Jakub et al. (2022) conduct a comparison of the environmental impact among electric, compressed natural gas, and diesel light-duty vehicles, considering the key features of each product. Moreover, the absence of design methods specific to EVs, irrespective of their application, further compounds the challenge. Closing this gap is crucial for providing comprehensive insights into the environmental impacts of EVs throughout their lifecycle, facilitating effective comparisons with traditional vehicles under various circumstances.

The literature lacks comprehensive works coupling the environmental assessment of EVs during the design phase and useful tools to do it. Two main groups of works can be considered: those providing approaches intended to assess the environmental impacts of EVs only and those offering design tools.

Among the first, many tools aid designers in making informed judgments; several are based on LCA (Marta et al., 2023). The reference standards are ISO 14040 (Arvanitoyannis, 2008) and ISO 14044 (Koffler et al., 2020). For example, Burchart-Korol assesses the potential environmental impacts of the life cycle of EVs, taking into account the energy used to charge EV batteries in the present and as projected into the future (with Poland and the Czech Republic as reference geographic areas) (Burchart-Korol et al., 2018). J. Szczerkowski et al. (Jakub et al., 2022) examine vehicles used for passenger transportation in the context of public transportation and electro-mobility. Some focus only on part of the whole vehicle, like the battery. Wang et al. (F. Wang et al., 2022; H. S. Wang et al., 2013) use a module that utilizes the results from a previous LCA study on SiNW- and SiNT-based Lithium-Ion Batteries (LIBs) for EVs to estimate the environmental impacts based on the specifications entered the dashboard. Results of the LCA studies are normalized per km of EV driving distance. Others face only part of the product lifecycle; for example, the study proposed by Graba M. et al. (Graba et al., 2020) is oriented to analyze only the sustainability of the use phase of an Automated Guided Vehicle (AGV) and aims at optimizing the traveled path.

In the second category, the focus is on design tools. Several eco-design approaches focus on battery design, and many others are developed for products of different natures. For example, Epp et al. (2022) belong to the first group: after inserting user-defined parameters into their High-Voltage-Battery-Optimization-Tool (HV-BOT) (e.g., energy, power, total installation space) a starting parameterization of the design variables is selected. The respective sub-models use the space to evaluate the optimal configuration for the given input parameters. The

procedure is executed for a previously defined number of iterations. Picatoste (Picatoste et al., 2022). presented a list of CE design criteria for LIBs used in BEVs and evaluated their implementation potential according to the views of industrial stakeholders. The results provided great insight into the design priorities of the industry. In addition to that, the work by Collado-Ruiz and Ostad-Ahmad-Ghorabi. (Collado-Ruiz & Ostad-Ahmad-Ghorabi, 2013) stands among those focused on the design of products that do not concern electric mobility. The work aims at introducing sustainability in design, starting with an analysis of similar products. The method can never substitute for a complete LCA; however, the results can guide in selecting priorities or key strategies. Ostad-Ahmad-Ghorabi and Collado-Ruiz (Ostad-Ahmad-Ghorabi & Collado-Ruiz, 2011) affirm that parametrization might be useful when developing such tools. Their applicability to a wider range of products may be difficult as eco-design methods and tools often rely on previous product analysis to provide quantitative results. Manuguerra et al. (2023) proposed a method and tool to predict the environmental impact of industrial EVs in the design phase considering all life cycle phases. However, the tool requires a high number of input parameters, and it is an analytical tool.

As a result, there is a requirement in the literature for a simplified environmental evaluation method and tool tailored for the initial phases of the design process, applicable to a diverse range of EVs. The current paper indeed is intended for a different range of vehicles, whose distribution of environmental impacts may not follow the same trends as passenger cars. Together with other conventional design criteria, the environmental effects of a product's life cycle should be taken into consideration while assessing the design of a product (Cappelletti et al., 2022a).

1.2. Machine learning and eco-design

In the realm of sustainability management, data science finds application across multiple tiers. The life cycle of a process involves a series of diverse processing steps, each of which can be delineated by detailing its inputs, outputs, and accompanying energy and economic metrics. The comprehensive inclusion of these varied data sets, sourced from heterogeneous origins, is essential for conducting a thorough LCA. Broader use of ML is challenging to implement because of the enormous amount of input factors and their uncertainties that affect the entire life cycle and whose sources can be highly heterogeneous (Kamm et al., 2023). ML techniques belong to a class of Artificial Intelligence (AI) techniques that can learn from data to increase their accuracy without further user action. By analyzing a collection of training data and generating an algorithm without the assistance of a human being, ML creates a model that can identify patterns (Tom M. Mitchell, 1997). Their application pervades private and industrial spheres (i.e., application in automatization of the existing requirement management workflow (Lyutov et al., 2019)) or in condition monitoring and predictive maintenance of bearings (Schwendemann et al., 2021)). ML algorithms are typically categorized into four groups: supervised learning, unsupervised learning, semi-supervised learning, and reinforcement learning (Batta Mahesh, 2018). ML is one area of data science that is now used to fill up data gaps for LCA (Ghoroghi et al., 2022). ML would provide appropriate methods and technologies for this analysis (Prioux et al., 2022). Prioux et al. (2022) suggested a merge of data science methods and environmental analysis to enhance LCA. The outlined strategy combines multidimensional scaling, the k-means algorithm, and environmental analysis. This methodology consists of five key stages: (1) defining goals and boundaries, (2) establishing data architecture, (3) conducting environmental assessment, (4) visualizing and analyzing results, and (5) making decisions. Notably, this approach is crafted to be efficient in terms of time and cost, as it eliminates the need for experiments and relies on publicly available scientific data as its database. The example of hybridization is not meant for eco-design.

They have also been utilized to create LCA ancillary tools that model

and forecast a product's environmental impact using data from the design stage. ML as a real-time algorithm enables it to evaluate production or process changes and provide prospective solutions for improved or less ecologically damaging manufacturing. With ML, it is possible to identify the most relevant attributes and concentrate on gathering them, ignoring others that might not substantially impact the model's accuracy. ML stands out in those applications where it is necessary to solve mathematical models efficiently and accurately. As a result, it can be modified to offer suggestions or techniques for an optimization process. For example, it can be used in a real-time decision-making process to identify ways to enhance a system's performance throughout its lifetime. The process can then be improved by using optimization techniques. Unlike the complete LCA, this makes it particularly helpful in the design process.

Elouariaghli et al. (ELOUARIAGHLI et al., 2022) link eco-design to Deep Learning (DL) and ML, while using also mathematical algorithms, to reach new perspectives for the domain of LCA. The study is in its infancy; however, aims to manually set the initial conditions of the scenario, automatically retrieve the related data and make the following calculations to have the global carbon impact of the product. The authors deal both with consequential and attributional LCA, two LCA methods; the authors apply their approach to Information and Technology (IT) products. The lack of algorithms and applications doesn't allow a proper validation of the proposed approach.

The literature offers examples of ML applications for predicting emissions from vehicles; unfortunately, they do not cover EVs or their design. Pan et al. (2021) designed a DL framework based on the gated recurrent unit to capture external factors and time dependence characteristics to predict real-time emission rates for Liquefied Natural Gas (LNG) buses under naturalistic driving conditions. The paper only examines the real-time emissions data of LNG buses and constructs a real-time prediction model for the emission rates (CO, CO₂, HC, and NO_x); no attention is paid to the emission characteristics of pure electric buses. Antanasijević et al. (2014) developed a model for GHG emissions forecasting for European countries using an Artificial Neural Network (ANN) approach with sustainability, economic and industrial indicators used as inputs. However, the study is at a very high level and considers the whole economic activities the European GHG emission can be broken down: transportation, but also energy supply and use, agriculture, industrial processes and waste.

In this context, the present work proposes an approach to get insights about the EVs' environmental impact in the first phase of the design process, when only the main design constraints are known. Specifically, the case study aims at estimating the environmental impacts through supervised learning algorithms. According to Algren et al. (2021), the method proposed belongs to the surrogate LCA field; and it is aligned with the studies according to which the main goal of the surrogate LCA is to enable product designers to predict the environmental impacts of a new product system or process. The approach aims to create a predictive model that allows it to be used in the initial design phase to analyze the environmental impact. A full LCA is the starting point; from it, a database is created, and models are obtained. The chosen impact category is Climate Change and the reference product is an electric shuttle. Ultimately, the method is tested to verify its applicability on a shuttle by estimating the environmental impact.

The work fills the gap of useful ML-based approaches, useful for the eco-design. Although some are proposed to reach new perspectives for the domain of LCA, their lack of algorithms and applications doesn't allow a proper validation or it is not intended for the design phase. Moreover, the present work provides a case study focused on electric shuttles. Electric shuttles are electric transport vehicles especially used to transport passengers in restricted areas (i.e., in terminals in airports or to provide service for students in universities between campuses or tourists among tourist attractions, in limited areas between or within cities (Anund et al., 2022)) or goods (i.e., provide transport of finished or non-finished merchandise between plants of the same company (Bibak

and Bai, 2023) closely located). Electric transport vehicles may climb rapid slope roads. Electric shuttle service vehicles are used for short distances and at low speeds (Ma et al., 2023). Examples of ML applications for predicting emissions from vehicles are available in the literature, however, those are not available for the design phase.

Compared to the existing literature, the present approach i) successfully provides a useful method for the design phase and ii) it is simultaneously based on a well-established methodology, that is LCA; iii) it fills the gap of lack of eco-design methods for EVs and iv) has wider applicability than solely passengers' cars. Existing methods do not provide a global assessment; the present work investigates the Climate Change impact category; however, it can be extended to all damage indicators.

In later stages of process design or for processes already in operation, LCA studies can provide valuable inputs to decision-making; as considerable detailed data are available, the decision making consists of a selection from a list of options. However, this approach makes LCA difficult to offer early advice and guide design decisions at the beginning of the process development steps (Karka et al., 2022). Therefore, the present study aims to contribute to the development of a streamlined, ML-based approach for early design to estimate lifecycle environmental impacts at the early stages of product development, using a limited amount of available information.

The remaining is structured as follows: Section 2 shows the approach in detail; Section 3 applies it to the electric shuttle and uses the electric shuttle model in the design process of a real vehicle; Section 4 discusses the results, before the conclusion (Section 5).

2. Methods

The methodology to develop a parametric environmental impact model for EVs using ML techniques consists of six phases (Fig. 1 (Saltz, 2021)): Problem definition; Data collection; Data Preparation; Modeling; Model evaluation; Model interpretation. The procedure structure is a standard procedure for solving a problem through data science. The introduction to the theme of LCA required customization of the Data Collection (Data Understanding and Data Preparation). The Data Understanding phase requires the creation of a simplified LCA, useful in the further choice of the most representative features of the problem; this also influences the Data Preparation phase. The approach herein proposed is intended to provide easy, handy and practical tools for the first phases of design. The creation of the model, through ML techniques, allows for repeatable analysis of the same of different EVs. The obtained quantitative results stand for the new driver of environmental sustainability to consider next to the traditional ones and are accessible and useful also for non-LCA experts, as technical roles as designers may be.

Once the model is complete, multiple choices and comparisons can be performed, as well as when designing different vehicles.

The current study is particularly directed towards the initiation of the model creation process, thereby laying the groundwork for numerous subsequent conceptual and functional EV designs. The following elucidates the six phases followed in creating the model.

2.1. Problem definition

The definition of the problem phase focuses on the conceptualization of the project requirements according to the environmental objectives. This phase aims to identify lifecycle management and product design opportunities. The present work aims to develop a parametric model with ML techniques for the environmental impact of EVs in the conceptual design phase. Thus, for this methodology, the business understanding phase defines two essential requirements: the *goal and scope definition* and the *model acceptance performance*.

As this methodology is strongly influenced by an LCA analysis, it retrieves the same findings regarding the *goal and scope definition*.

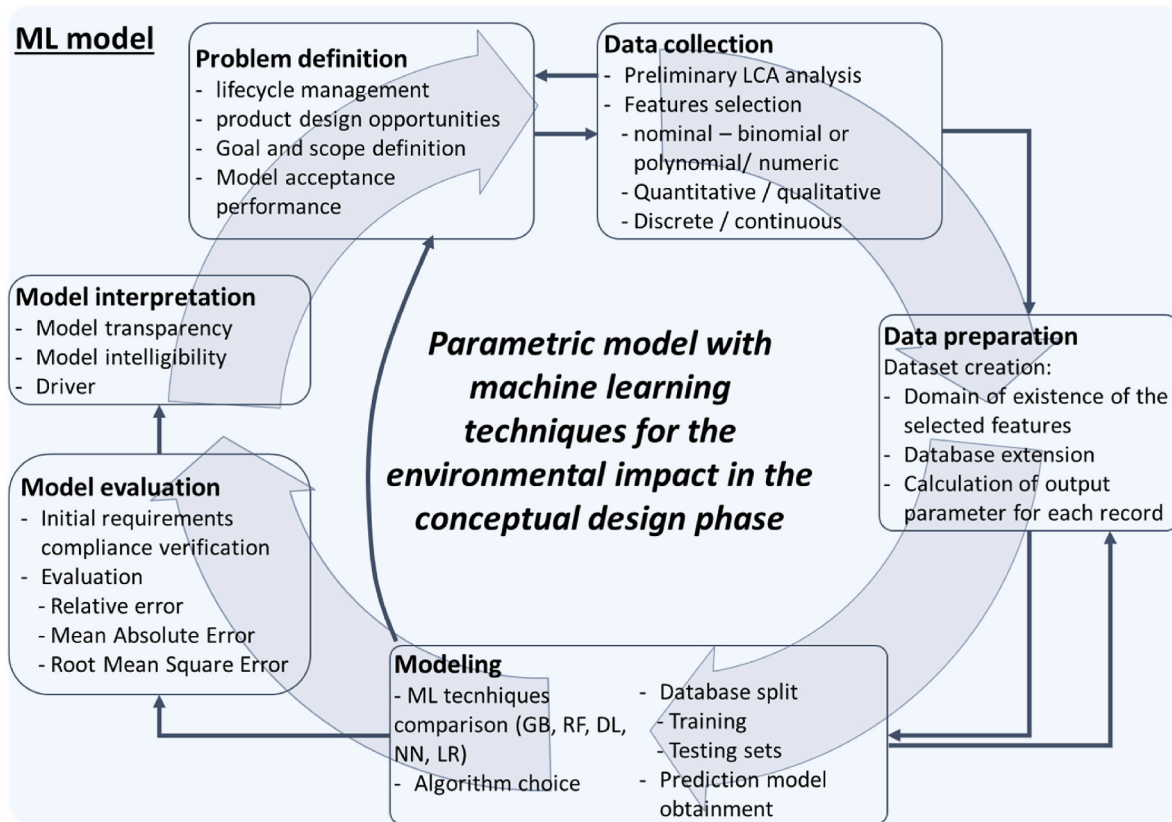


Fig. 1. Methodology to develop a parametric environmental impact model for EV design using ML techniques (adapted by (Saltz, 2021)).

Therefore, first the objective is addressed. In the present dissertation is to assess the environmental impact of an EV; the functional unit is to use the EV for a given lifetime; the system boundaries (considered stages) taken into consideration will encompass the material, the manufacturing, the use, the maintenance and EoL. Finally, LCIA (Life-cycle Impact Assessment) Method, database, and impact category must be defined.

In the *model acceptance performance*, the acceptance accuracy of the model is defined. In the case of an LCA analysis, an error of 40% can be encouraging (Algren et al., 2021).

2.2. Data collection

This phase aims to obtain the database on which the model will be trained and tested. Data collection consists of two phases: a *preliminary LCA analysis* and *feature selection*.

A *preliminary LCA analysis* is necessary to understand the impact of the product and the most impactful phases. A complete LCA analysis already previously carried out (e.g., of an existing product) can be used or a new analysis carried out. In this second case, the *preliminary LCA analysis* may consist of an analysis with little data available and many approximations, as carried out during the initial design phase and some decisions have not yet been taken and the information is reduced.

The first step of an LCA analysis is the goal and scope definition; it is the same to the one identified in the definition of the problem. The second phase, the inventory (LCI) collects information about lifecycle phases. As far as the material phase concerns, the EV is divided into two main assemblies, the electronic assembly and the frame assembly (Ambrose et al., 2020). The electronic assembly consists of the most representative components of an EV (battery pack, electric motor, battery management system, case). The chassis assembly collects all materials not belonging to the electronic assembly. For both assemblies, the total mass will be identified. In case of missing information, choosing a

commercial database can reduce the amount of data to be collected. For example, datasets in the Ecoinvent database refer to the battery pack and the glider without the battery pack. The manufacturing phase considers environmental impacts related to the consumption of resources such as heat, water, and electricity, related to the production of the different components. For the use phase the causes environmental impacts are to be called to electricity consumption and particulate emissions. The information needed to calculate these impacts is the mass of the vehicle, the average consumption of the vehicle, the distance traveled during its entire life, the electricity grid mix of the country of use and possibly whether the company has its system for producing electricity and charging vehicles. Particulate emissions are due to the consumption of wheels, asphalt and brakes. The maintenance phase represents the impact of battery pack replacement. The impact related to routine maintenance of a vehicle, i.e., replacement of wheels and other worn components, is also considered. The EoL phase considers the treatment of materials and components at the end of the life of the vehicle.

The third step is the LCIA: In this phase, the environmental impact of the vehicle is calculated. Ultimately, the interpretation of the obtained results takes place.

Through the LCA analysis, it is possible to identify the most influential features (*Feature Selection* phase). Those can be different; when features are nominal, they are parameters that allow predefined names as values; in this case, it is possible to distinguish one object from another. Features can also be binomial (e.g., true or false, male or female) or polynomial (e.g., more than two options), numeric, and parameters that represent numeric values (e.g. integers, real, or dates).

Many are the identifiable features at the end of a full LCA and they can have varying impacts on the LCA outcome. The feature selection identifies those that describe the vehicle at a general level. A general description of the vehicle makes it easier to collect data at the initial design stage. The easier the model to understand, the easier to use and

interpret (Mladenić, 2006). Choosing features suitable for the problem being addressed will simplify the model, which will also improve the prediction result (Gao et al., 2023). Features can also be chosen to create specific models for certain life cycle stages or to describe specific vehicle categories. Consequently, some parameters are fixed (starting assumptions), while others constitute the variable parameters that allow different scenarios comparison. The expert who conducted the full vehicle LCA analysis and has an informed view of the case study will be responsible for choosing the most appropriate parameters to describe the product at the early design stage. Thus, it is known in detail how the parameters are related to each other and make an informed selection.

2.3. Data preparation

This step sees the creation of the database. The methodology was created for supervised learning. A generic database for supervised learning is structured by i) the Number of independent variables which is the number of parameters related to the EV used to define the predictive model, called features; ii) the Number of “records”: number of representative elements of different alternatives from which extrapolate the independent variables. The creation of the database consists of 3 stages: i) Domain of existence of the selected features; ii) Database extension; iii) Calculation of output parameter for each record.

In the *domain of existence of the selected features*, the maximum and minimum values have been defined for each feature. In this way, how the range within the parameter can vary is defined (Heath et al., 2009). The chosen variables can be qualitative or quantitative when the numerical values express their values or intensities. Quantitative variables are divided into discrete and continuous. Discretization is a fundamental technique in this phase, as it allows converting attributes from continuous into discrete, dividing the domain into a defined number of intervals, namely, determining the resolution of the parameter.

Once the features are in the *database extension* phase, all potential combinations of vehicle characteristics are identified. The impossible combinations are erased.

Calculation of output parameter for each record is the last step and it consists in calculating the dependent variable (parameter) the ML model will predict; it acts as an indicator to assess the environmental impact of an EV. This parameter is correlated with the features chosen as input to the model.

2.4. Modeling

Modeling involves the creation and evaluation of multiple models using various techniques. The method allows for obtaining labeled datasets. For each dataset, the environmental impact is calculated for a specific impact category. The chosen machine learning category for this method is Supervised Learning (SL), which indeed utilizes labeled datasets to train algorithms to predict outcomes. The primary objective is to identify the most effective ML methods for predicting the impact of environmental factors, with a dedicated model for each parameter. To achieve this, several ML techniques such as Gradient Boosting (GB), Random Forest (RF), DL, Neural Network (NN) and Linear Regression

(LR) need to be compared (Table 1) (Roberto Marmo, 2020); five algorithms from the comprehensive comparison by Gayathri et al. were depicted (Gayathri et al., 2022).

The environmental impact model building process begins by dividing the database into training and testing sets, typically 70% and 30% respectively. Specific software tools for data mining or data science are used for the modeling process. Subsequently, each algorithm produces a prediction model. For each model, the testing set is employed to calculate metrics like Relative Error, Mean Absolute Percentage Error (de Myttenaere et al., 2016) and R-Squared Error (Colin Cameron and Windmeijer, 1997). The Relative Error is the average of the absolute deviation of the prediction from the actual value divided by the maximum of the actual value and the prediction (Rapid Miner, 2023a). Based on these metrics, the optimal method is selected. The chosen algorithm is the one yielding the lowest error according to the selected metric.

2.5. Model evaluation

This phase consists of assessing whether the model performance meets the initial requirements defined in the problem statement (required accuracy). If not verified, an analysis of the procedure supports the understanding of the causes behind the mismatch. Typically, actions are made in the feature selection and the domain of existence of the selected features. In the feature selection phase, the number of features can be increased; however, this may also increase the complexity of the problem. If actions are taken in the domain of the existence of the selected features, the number of intervals for each stage can be increased, thereby enhancing the resolution for a specific parameter.

2.6. Model interpretation

The interpretation model phase enables users to handle the parametric model and assess the impact of each independent parameter on the assessment impact. This phase involves interpreting the model to make it transparent and easily understandable, rather than being treated as a mysterious entity. The study of feature importance allows design engineers to understand the meaning of each driver, which corresponds to a design variable, and how it influences the overall impact. Various algorithms can be employed to evaluate the importance of these features (Altmann et al., 2010). This information ultimately aids designers in using the model effectively during the design phase.

3. Results

The present section first goes through the six phases to develop a parametric environmental impact model for EVs using ML techniques, and then applies it to the case of an electric shuttle (3.7). The goal of the case study (summarized in Fig. 2) is to confirm that the suggested method enables the creation of an environmental model for an electric minibus with an acceptable level of accuracy. The company that sets the objective produces different types of EVs, intended for distinctly diverse purposes (i.e., tourism, industrial logistics, military and emergency

Table 1

ML algorithms' characteristics. SL = supervised; UL = Non supervised; RL: Reinforce; RE: Regression; CL = Classification; CG = Clustering.

Model	Calc speed	Learning speed	Memory required	Need to optimize	Annotations	Learning type	Solution	Ref
LR	Fast	Fast	Little	Minimal	Small problems with linear decision surfaces	SL	RE	Su et al. (2012)
RF	Fast	Fast	Little	Something	Good generalization but prone to overfitting	SL	RE	Liu et al. (2012)
GB	Fast	Fast	Little	Something	Good generalization but prone to overfitting	SL	RE	Natekin and Knoll (2013)
DL	Moderate	Slow	Medium to large	Much	Classification, understanding, recognition and predictions	SL, UL, RL	CL, CG, RE	LeCun et al. (2015)
NN	Moderate	Slow	Medium to large	Much	Classification, understanding, recognition and predictions	SL, UL, RL	CL, CG, RE	Bishop (1994)

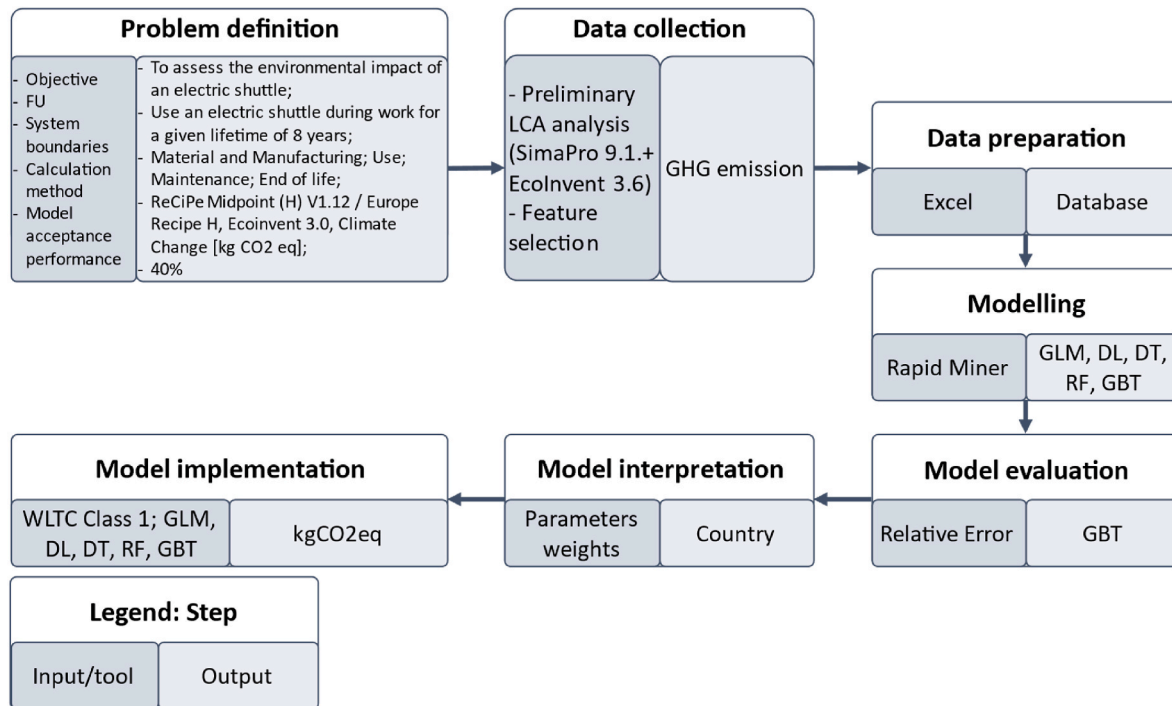


Fig. 2. Case study. Each box consists of one step of the proposed method; dark grey boxes contain tools used and requirements, and light grey boxes output the step.

cases). Based on the company requirements, the acceptance of accuracy was established during the problem definition phase. The company needs a strategy for calculating environmental impact in the early design phases. The company contributed the vehicle Bill of Material (BOM) and technical know-how required to produce the results for the preliminary LCA analysis. The case study’s starting point was established by first identifying a reference EV, from which minibus records could be extracted. Once developed, the model can be used by product engineers without the need for experience in LCA analysis.

3.1. Problem definition

The present work aims to develop a parametric model with ML techniques for the environmental impact of an electric shuttle in the conceptual design phase. The objective is to assess the environmental impact of an electric shuttle, which has a lifetime of 8 years; the system boundaries include the material and manufacturing, use, maintenance and EoL. The transportation-related impact was not considered in this study. The calculation method used is ReCiPe Midpoint (H) V1.12/ Europe Recipe H, Ecoinvent 3.0, Climate Change [kg CO2 eq] that is the most commonly used indicator (Subal et al., 2024). The model accepted accuracy is 40% Relative Error.

3.2. Data collection

Given the goal and scope definition identified in the problem definition (Table 2), the LCI is the phase that collects information about lifecycle phases. The most important information about the shuttle

Table 2
Material phase data.

Frame Assembly			Electronic Assembly		
Parts	Value	UM	Parts	Value	UM
Total Mass	4044	[kg]	Glass	28	[%]
Steel	60	[%]	Fiberglass	2	[%]
Wood	2	[%]	Rubber	2	[%]
Plastic	3	[%]	Aluminum	3	[%]
			Parts	Value	UM
			Total Mass	381	[kg]
			Electronic components	66	[kg]
			Battery Pack	260	[kg]
			Electric motor	55	[kg]

material phase is shown in Table 2.

For the manufacturing phase, the impact is calculated using the flow that calculates energy resource consumption as a function of frame mass. The flow was customized according to the reference production site, which is Italy in this case. Table 3 shows the main data for the use phase analysis. The mass of the vehicle also considers the average mass of the driver, equal to 75 kg. The average consumption of the vehicle was assessed considering a customized driving cycle, considering the maximum load of 100% equal to 1840 kg. It is assumed that the company’s photovoltaic system meets the demand for 50% of the electricity needed to charge the vehicle.

The maintenance phase considers the replacement of the battery pack every 106000 km. The impact of routine maintenance is related to the distance traveled and the weight of the vehicle. At the EoL phase 85% of the metal is recycled and 50% of the remaining materials; in both cases, the remaining untreated is sent to landfills. For electronic components, 50% is sent to a specific treatment, while the remaining is landfilled. For the cells of the battery pack, 100% of the batteries are

Table 3
Use phase data.

Vehicle Data	Value	UM
Mass	4500	[kg]
Average consumption	0.281	[kWh/km]
Distance travelled	106000	[km]
Country electricity grid mix	Italy	
Photovoltaic system	50	[%]

sent to an EoL treatment that includes 50% hydrometallurgical treatment and 50% pyrometallurgical treatment. The vehicle was modeled in the SimaPro software. The overall results presented in Fig. 3 are for the Italian use case.

The Material phase turns out to be the most impactful. In particular, the steel material in the frame assembly for the Climate Change category; it has the greatest mass. The Use phase is also very impactful. The change in the electricity mix, in favor of renewable sources, results in a lowering of the impact. Recycling and proper treatment of EoL components and materials result in a benefit to the total impact.

The potential features were high in number (more than 100); however, a limited quantity was chosen for the model (Fig. 4). The selection followed the main approach guidelines. Features considered fixed for the type of vehicle considered were discarded. Features that describe the vehicle life cycle in a general way were chosen. Finally, a choice was made to avoid having features that were too correlated. The LCI of the full LCA analysis made it possible to understand how the features are related to each other's. Exclusion of dependent features (i.e., speed vs consumption, guide profile vs consumption, average passenger mass and mass), listed below and shown. Grey boxes in Fig. 4 show fixed features, which are features always valid for the type of product considered, electric shuttle in the current case. White features can vary, according to the specific product and design process; some are dependent on each other's. They are design constraints (i.e., maximum load, empty mass, maximum speed), or deduced from components datasheets (i.e.,

chemical battery, number of cells). The chosen parameters for creating the database are.

- Country Use Phase: indicates the electric mix of the nation where the vehicle operates.
- Country Manufacturing Phase: indicates the electrical mix of the country in which the vehicle was produced.
- Distance [km]: indicates the total distance traveled by the vehicle throughout its entire life.
- Consumption [kWh/km]: indicates the average electrical consumption of the vehicle per unit of distance traveled.
- Mass [kg]: indicates the average mass of the vehicle.
- Autoproduced Energy [%]: indicates the percentage of electric energy consumed by the vehicle that is self-produced through a private photovoltaic system.
- N. Battery Pack: indicates the number of battery packs installed on the vehicle.
- Change Battery Pack: indicates whether the battery pack replacement is anticipated during its useful life.

These parameters were chosen based on the results of the preliminary LCA analysis. Not all available parameters were included as they were deemed redundant, which is why the main ones were selected. The material phase is influenced by the vehicle's mass and the number of battery packs. Parameters related to material percentages were not

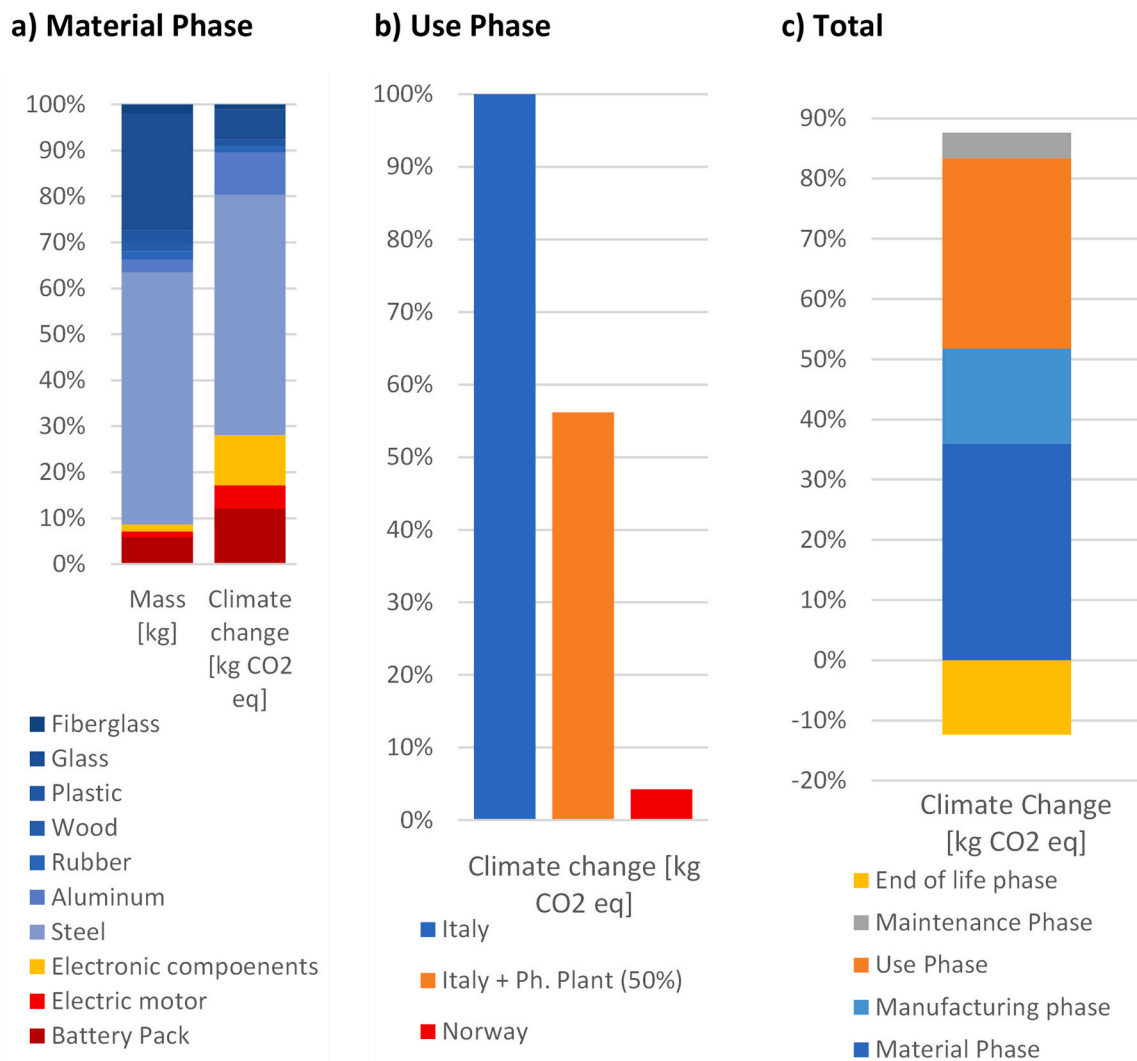


Fig. 3. LCA results modeled from SimaPro software for material (a), use phase (b) and overall lifecycle (c).

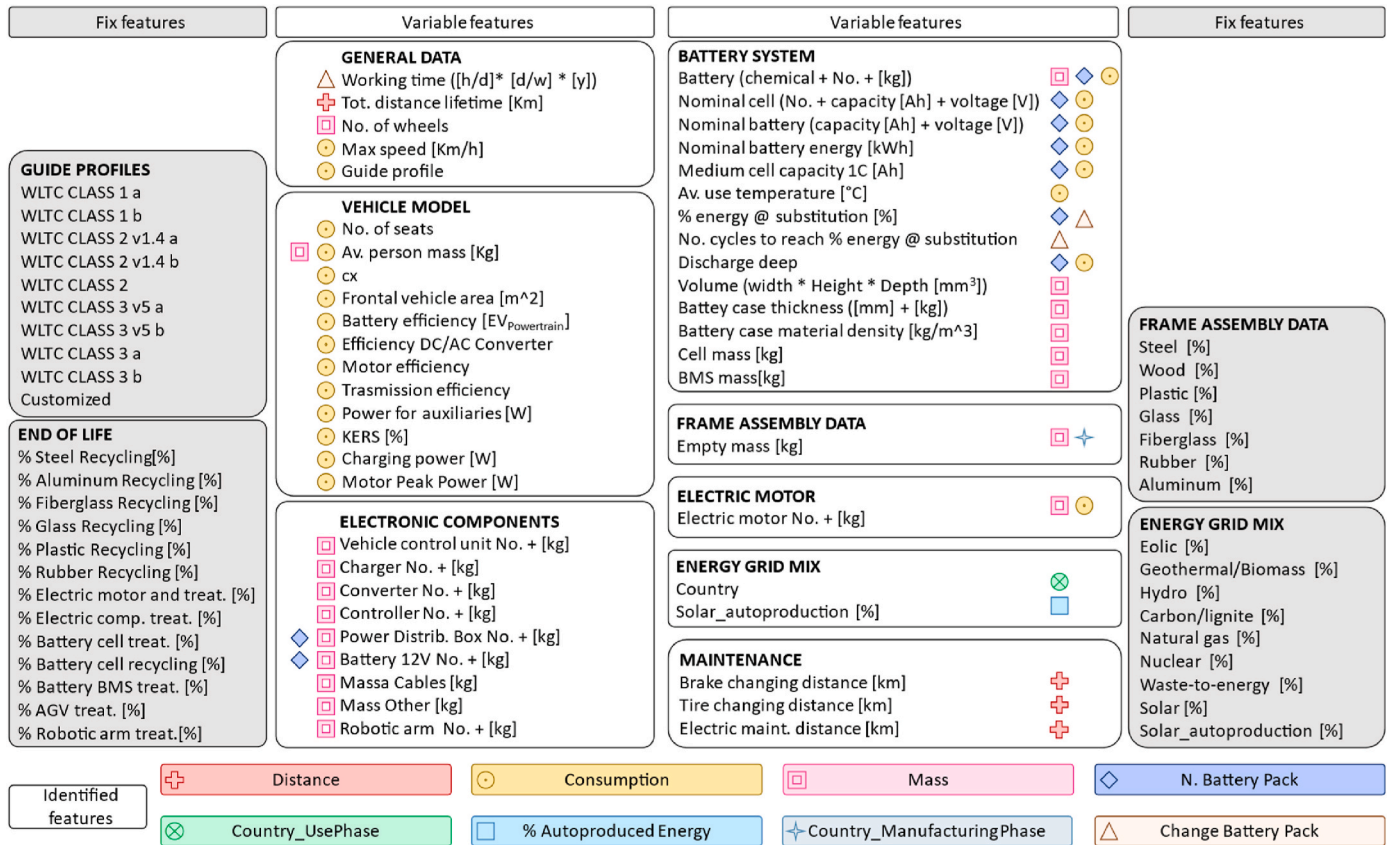


Fig. 4. Features Selection for the electric shuttle.

chosen because the intended model pertains to shuttles with similar percentages, and these parameters were considered redundant compared to the representing vehicle mass. The only design parameter introduced in the material phase is the number of battery packs, which serves as a reference design parameter. The manufacturing phase is influenced by the vehicle mass and the production country. The use phase is influenced by the distance traveled, average consumption, autoproduced energy, and country. Parameters related to the carried load mass and selected driving cycle were discarded as redundant about the vehicle’s average consumption. The maintenance phase is influenced by battery pack replacement, as well as EoL treatment of the consumed battery pack, with the impact of constructing the replacement battery pack added in this phase. The impact of routine maintenance is related to the distance traveled and the weight of the vehicle. The EoL phase’s impact is influenced by the vehicle’s mass and the number of battery packs.

3.3. Data preparation

Table 4 shows the maximum, intermediate and minimum values

Table 4
Vehicle parameters domain of existence and steps.

Parameter	Variables				
Distance [km]	50000	100000	150000	200000	250000
Consumption [kWh/km]	0.2	0.3	0.4	0.5	0.6
Mass [kg]	2000	3000	5000		
N. Battery Pack	1	2	4		
% Auto produced Energy	0	25	50	75	
Country Use Phase	Italy	Norway			
Country Manufacturing Phase	Italy	China			
Change Battery Pack	Yes	No			

chosen for each parameter. The values were chosen based on data provided by the literature case study (Nordelöf et al., 2014; Raugei et al., 2015). The maximum percentage of energy that is covered by the self-generated through a photovoltaic system is set to 75% because it is often coupled with additional source(s) of power (Peffley and Pearce, 2020). Country Norway was chosen as it represents an electricity mix based on renewable sources. The Norwegian electricity grid mix is low in carbon as over 90% of electricity production comes from hydroelectric sources (Godzimirski, 2014).

All possible combinations of the given parameters have been identified. Through SimaPro 9.1, it was possible to analyze the impact in terms of kg CO2 eq for each record. This resulted in a database of 7200 records. An excerpt from the database is shown in Table 5.

3.4. Modeling

The modeling step in the process used RapidMiner (Auto Model plugin for RapidMiner Studio), a data science platform developed by Altair Engineering. RapidMiner offers a wide range of ML techniques and features. While the software validation for data science may not be as exhaustive as full cross-validation, this approach strikes a balance between runtime efficiency and model validation quality (Rapid Miner, 2023b). RapidMiner enables the testing and comparison of five algorithms: Generalized Linear Model (GLM), DL, Decision Tree (DT), RF, and Gradient Boosted Trees (GBT).

3.5. Model evaluation

The Relative Error, which is easily understandable by engineers and directly relevant to the initial design phase, is adopted by the authors (Table 6). The findings indicate that both GLM and DL fail to achieve the required level of accuracy. In general, decision trees demonstrate

Table 5

Excerpt from the database. Full database is obtained from combinations of parameters.

Distance [km]	Consumption [kWh/km]	Country_UsePhase	%Autoproduced Energy	Country_ManufacturingPhase	Mass [kg]	N. Battery Pack	N. Change Battery Pack	Climate Change [kg CO2 eq]
150000	0.2	Norway	0.5	Italy	5000	2	No	27280.738
200000	0.6	Italy	0	Italy	5000	1	No	98438.125
200000	0.6	Norway	0	China	2000	2	Yes	18680.872

Table 6

Models performance results.

Model	Relative Error [%]
Generalized Linear Model	20.74
Deep Learning	2.18
Decision Tree	4.65
Random Forest	11.22
Gradient Boosted Trees	0.90

compliance with constraints and outperform deep learning and linear regression. Specifically, GBT performs better than RF and DT.

The use of commercial software for the life cycle evaluation enables the development of a reliable database for training ML algorithms. Additionally, LCA estimation software facilitates the gathering of impacts under various conditions. Consequently, the proposed method is not reliant on any specific software and can be implemented using data science tools and other analytical disciplines related to environmental impact estimation. All the parameters selected for the predictive models are known during the early stages of designing a shuttle. However, it is important to note that this research only provides a comprehensive estimate of the environmental impact of a shuttle. The detailed analysis obtained from a comprehensive LCA analysis at the end of the design process cannot be replaced by the results obtained through this approach.

3.6. Model interpretation

Feature importance analysis allows engineers to highlight the impact of parameters on the model. In the case of a shuttle model, the algorithm has identified the country parameter (Fig. 5) as the most relevant (i.e., the electricity mix of the reference nation).

3.7. Model application to an electric shuttle

In this paragraph, the analyzed models are used to forecast the environmental impacts of an electric shuttle. How the created model paves the way for numerous subsequent conceptual and functional EV designs is shown (Fig. 6). The considered attributes are the technical features expressed by the Italian Ministry of Infrastructure and Transport (Ministero delle infrastrutture e dei trasporti, 2015).

The electric shuttle is characterized by.

- Maximum speed lower than 25 km/h;
- Maximum admitted mass in the fully loaded range: 2500 kg;
- Maximum nominal continuous power: 15 kW;
- Maximum power and total mass (fully loaded) ratio: 7.5 kW/t;
- Maximum capacity: 8 passengers and driver.

The WLTC class 1 with a maximum speed limited to 25 km/h (Fig. 7) is suitable for quantifying the energy consumption of the vehicle.

Reasonably, an overall distance of 100000 km is assumed traveled without battery replacement. The total mass is 2180 kg, considering the frame assembly mass is 1380 kg, the average load and the electronic assembly mass (400 kg). The estimated consumption when the shuttle is half-loaded is 0.229 kWh/km (Table 7).

Table 8 shows the values associated with the features of the predictive problem.

The environmental impact of the lifecycle of the electric shuttle is 17754.181 kg CO2 eq with the GBT model.

4. Discussion

The present work uses product design parameters to estimate LCA results for new products; it enables product designers to predict the environmental impacts of a new product system or process. The method contributes to building knowledge in the context of eco-design (Rossi et al., 2022) as it allows the addressees to make environmentally conscious choices, the knowledge is easily consultable, and it allows an immediate understanding and interpretation of results.

The need for the present work stands in the lack of tools for the design process that can retrieve quantitative results, accessible and useful also for non-LCA experts, as technical roles as designers may be. Moreover, the work overcomes the obstacles of using LCA-bases at various levels of design as the manual data inputs and the lack of fully automated processes that undermine their advantages (Hansen et al., 2023).

The case study and the construction of the database come from the need for a company that produces a wide variety of electric vehicles for multiple purposes and applications, such as electric shuttles for tourism, for inter-quarters logistics. The highly different applications of the EV determined the ranges of the parameters and the scenarios.

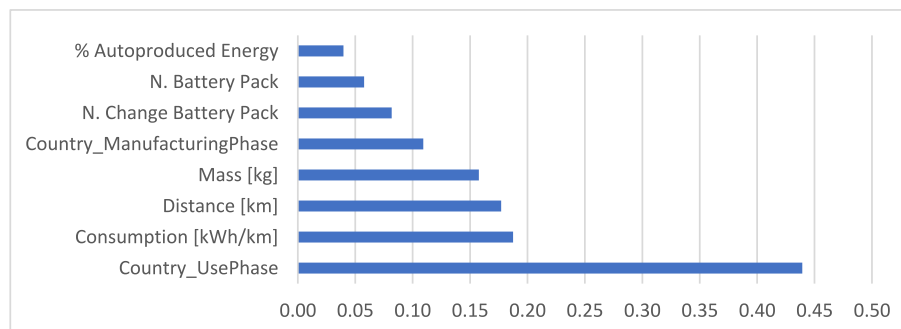


Fig. 5. GBT model, weights of parameters.



Fig. 6. ML model workflow in EVs concept design.

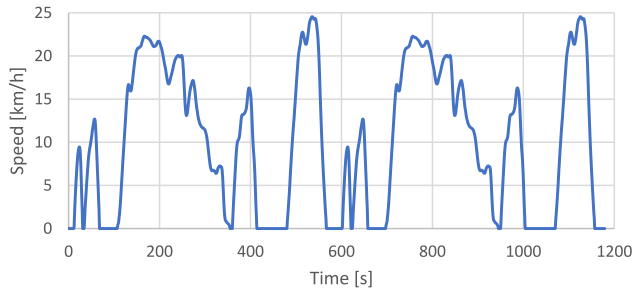


Fig. 7. Customized driving cycle, adapted from WLTC class 1.

Table 7
Vehicle consumption [kWh/km] related to the transported load [kg].

Weight [kg]	Load	Consumption [kWh/km]
1860	0%	0.219
2020	25%	0.224
2180	50%	0.229
2340	75%	0.235
2500	100%	0.240

Table 8
Features values for the shuttle.

Parameters	Value	UM
Distance	100000	km
Consumption	0.229	kWh/km
Mass	2180	kg
N. Battery Pack	2	n.
% Auto produced Energy	58	%
Country_Use Phase	Italy	
Country_Manufacturing Phase	Italy	
Change Battery Pack	No	

4.1. Method strengths and limitations

The present work extensively goes beyond the state of the art, still exploiting well known methodologies. The database relies on the strong objectivity of LCA, which is the method employed in building each line. The database contains the combinations of the parameters that vary within the ranges displayed in Table 4.

The work provides researchers and designers with a new approach to evaluate the environmental impact of EVs during the design phase, with technologies never used before in the specific context of the design of EVs. Unlike the existing literature, which often focuses on certain phases for the sake of comparison with traditional combustion engines as Raugai et al. (Raugai and Winfield, 2019), all the product lifecycle stages are considered. The output is a surrogated LCA model. According to Algren et al. (2021), the term surrogate LCA is used to describe ML algorithms trained on process LCA data to predict LCA results. Differently from Castaneda et al. (2021), who propose a model that returns a qualitative result and makes a comparison with other more complicated products, the proposed model returns a quantitative result, allowing a quick comparison with the impact caused by other vehicles. The need for such approaches arose for three main reasons; first, the environmental

driver is slowly making its way into the design process. Consequently, designers are asked to assess the consequences of their choices (Cappelletti et al., 2022b), in terms of environmental sustainability. Secondly, a full LCA requires time, resources and detailed data about production processes and product structure; a surrogated LCA can retrieve quantitative results with an acceptable error, without high volumes of input information. Ultimately, LCA made through commercial tools are not flexible, nor customizable. Parametrization is possible, but rigid.

With the surrogated approach presented hereby, it was possible to estimate the overall environmental impacts of the lifecycle of an electric shuttle, for the Climate Change impact category, with seven parameters.

The chosen design parameters were selected, considering those that have a greater influence on the impact through a preliminary LCA analysis. This minimized the required design parameters. The present work started from more than 100 parameters. Gao et al. (2023) also handled the design parameters of electric vehicles before. They focused on the Chinese market from 2017 to 2022 and identified more than 50. However, those parameters only refer to the structure and functionalities of the vehicle and do not examine any aspect of the remaining lifecycle phases, as the present work does.

The need for a few design parameters is among the main strengths of the present approach. Four are the main reasons that limit the influence of LCA within decision-making: lower priority of environmental aspects compared to other decision criteria, complex LCA methods, complex results, and a lack of data to perform a reliable LCA (Subal et al., 2024). The present work, unlike many analytical models not based on ML (Manuguerra et al., 2023; Marmioli et al., 2020), does not require much input data; therefore, the effort required to build kinematic models to estimate consumption does not justify the large uncertainty of the input data.

Together with its complexity and time consumption, the implementation of LCA at early design stages has so far remained a challenge; in addition to that, the requirement of many data, that are not available since then (Jusselme et al., 2018). The construction of the model allows for quantitative results, despite the use of a limited number of parameters.

The drawback of small inventory effort is the cumulate results, despite the analysis considering all lifecycle phases. The proposed model is trained on a database, whose total impact is obtained by summing the impacts of four life cycle phases and not focusing only on the use (Onat et al., 2021).

The traditional LCA is a very rigorous and rigid method. It is one of the most diffused, appreciated and reliable methodologies to provide quantitative evaluation of the sustainability of technologies and support decision-making in design. The traditional LCA provides detailed quantitative output, that describes the current state. The main limitations are the high requirements and quality of data and the incapability of providing a quantitative perspective evaluation of future developments, as prospective and surrogate LCA can do (Spreafico et al., 2023). The current approach overcomes the linearity of traditional LCA because: i) it returns quantitative results, given little inputs; ii) it requires a much shorter time for collecting data compared to existing approaches in the literature; iii) the results are acceptable at the design phase when little production information is available and yet to change along the process.

Through the exploitation of ML approaches, the present work successfully exploits what was suggested by Spreafico et al. (2023), who

look for studies focused on diversified parameters (such as operational scenarios) to enrich the setting of the forecasts they work on their prospective LCA. Eco-design decision-making process could benefit from the integration of the two approaches, as the prospective LCA would enrich the product and processes' parameters.

The design process is characterized by high variability of hypothesis and is the time when project constraints are set. The resulting model makes the designer aware of the consequences of choices through comparisons, thanks to the quantitative results returned by the model. At first, the design process is limited by project constraints but has much freedom for what has not been set yet. Below the main parameters are reported.

Mossali et al. (2020) use the House of Quality tool to assess the most important design interventions, once the technical specifications to produce an EV battery pack and the redesign requirements are derived to enable CE strategies of reuse, remanufacturing and recycling. Similarly, through the output of the proposed approach under environmental terms, designers can decide whether to expect the changing of the battery pack. For example, the planner is interested in comparing the result of the vehicle described in Chapter 4 with the same vehicle but traveling 200,000 km and replacing the battery pack; the impact in the latter case will be 27891.833 kg CO₂ eq (increase by 57%).

The emission reduction benefits of EVs are dependent on the time and location of charging and CO₂eq emissions are highly dependent on the percentage of fossil fuels in the grid mix (McLaren et al., 2016). This is among the main reasons why the auto production and energy grid mix are among the selected features. The company could evaluate at an early stage, quantitatively, the benefit of using a photovoltaic system to recharge the EV and then decide whether to increase its efficiency or build one specifically for the vehicle. For example, in the case study of Chapter 4, if the company does not use photovoltaic panels for vehicle charging, the impact will be 22929.080 kg CO₂ eq, which is equal to an increase of 29%. The country use phase parameter makes it possible to compare the difference in the impact of vehicle use, in different geographical contexts. This can be useful both for the company, to check in advance the impact in different plants and for an EV vendor who can simulate multiple stages of use for different countries. For example, if the vehicle in the case study were used in Norway, the impact would increase by 34% and reach 11722.087 kg CO₂ eq.

The country manufacturing phase parameters allow the comparison of vehicle production impact, in different geographical contexts. The electricity mix can have a very noticeable impact on carbon emissions during the manufacturing phase. As affirmed by Lai et al. (2023), as the percentage of clean energy in the electricity mix increases, the carbon emissions during the battery manufacturing process are steadily declining. In the current model, if the vehicle were built in China, the impact would be 19769.350 kg CO₂ eq, that is an 11% increase. When extending to the production stage of the entire life cycle, there emerge disparities in research conclusions; therefore, more detailed studies may contribute to reducing uncertainties for the manufacturing phase results (Shang et al., 2024).

Parameters such as mass, consumption, and battery packs can be used to compare the impact between different versions of EVs to choose the vehicle to purchase with the least impact. The development of efficient EVs is related to the management of different parts of the powertrain (Maiorino et al., 2024). For example, considering a vehicle that has a consumption of 0.4 kWh/km, 4 battery packs and a mass of 3000 kg; leaving the remaining parameters unchanged, its impact is 30279.739 kg CO₂ eq, that is, compared to the vehicle in Chapter 4, an increase of 71%.

The database on which the model is trained must be built considering all possible scenarios. Extension to additional impact categories may be considered, as Shang et al. (2024) claim a lack of research based on production status, especially to identify the main emission types in the manufacturing process of lithium batteries, components, and complete vehicles. This would be the baseline for further emission reduction

strategies for EV production strategy development.

4.2. Results comparison: ML-based ex-ante and LCA post design

This section proposes a comparison of results obtained by applying the method herein presented (chapter 2) and the environmental analysis, carried out at the end of the design process. The adapted input data are summarized in Table 9. The environmental impact calculated at the end of the design differs by 17% from the ML model results. The input data main data changes are due to the different composition of the frame of the vehicle. The model considers a generic mixture of metals, non-metals, plastics, etc. Moreover, the product's material composition is subjected to high variability along the design process, before reaching the final design. In their work, De Wolf et al. (De Wolf et al., 2023) outline works with high variations of environmental impacts in the characterization of software tools and databases for LCA.

For example, Dalla Mora et al. (Dalla Mora et al., 2020) noticed variations of up to 22% of the environmental impact of the materials for the buildings they analyzed. Emami et al. (2019), in their comparison of the embodied environmental impact assessment of a concrete-element, multi-story residential building and a detached wooden house in Finland, registered significant discrepancies: the whole building assessments yielded a 15% difference and most individual categories showed a difference of 40% or more. Applying their optimization model for selecting optimal prefabricated column design considering environmental impacts and costs using a genetic algorithm, Kumi and Jeong (Kumi and Jeong, 2023) got an average error for the environmental impact of 25%, with peaks of 37%. Therefore, the error (17%) obtained in the present work, is acceptable because it is in line with similar errors observed in the literature (Algren et al., 2021). The quantification of other impact categories can lead to adverse results (Moro and Helmers, 2017). Moreover, the works above mentioned refer to LCAs of final products, which should be more accurate than surrogate LCA for the design phase. The current approach refers to errors in forecasts of environmental impacts made at the design phase and obtained to incorporate precautionary principles into the decision making and get an overview of the lifecycle emissions, rather than obtaining punctual values.

The present approach and the case study herein provided go through the manual extension of the database; in addition to that, the comparison of the maximum expected error and the obtained error are user-based. Proper automation of the two phases would allow for time reduction and resource efficiency.

5. Conclusions

The need for the present work has deep connection with the establishment of environmental impact among the main design drivers. Designers need to be aware of the environmental impact of their choices over the entire product life cycle. However, they often have technical backgrounds and are not LCA experts. In addition, such well-known and accepted methodologies (for example, LCA) although are accurate, need much detailed data. Contrarily to traditional full LCA, not all lifecycle information is available yet in the design phase and commercial tools are

Table 9
Shuttle data, as defined in the design process.

Parameters	Value	UM
Distance	106.000	km
Consumption	0.281	kWh/km
Mass	4500	kg
N. Battery Pack	2	n.
% Auto produced Energy	50	%
Country_UsePhase	Italy	
Country_ManufacturingPhase	Italy	
Change Battery Pack	Yes	

very rigid, against the high flexibility and variability of the design phase. Therefore, lean approaches are needed for the design phase, together with simpler, surrogated LCA, that this paper accomplishes. The present work proposes a method to support EV design. The methodology allows the development of a parametric environmental impact model for electric shuttles exploiting ML techniques and consists of six phases: problem definition; data collection; data preparation; modeling; model evaluation; model interpretation. All stages of the product life cycle have been considered in the construction of the database, and the model provides quantitative results that consider the consumption of material and energy resources. Regression models and supervised algorithms were used. The resulting model is among the main strengths of the proposed approach: the model: i) allows for engineers to make informed decisions during the design process, even when designers are not experts in environmental sciences; ii) ensures comparison of different analyses, as the model relies on the same database and initial hypothesis, iii) consequently, results are coherent and comparable. Therefore, the decision-making process encompasses the environmental driver, without requiring much effort, or resources. It urges to underline that the current approach is intended for the concept design phase, without meaning to replace a complete LCA analysis, which is feasible at the end of the design process, when detailed information is available. However, the results are within an acceptable range or error.

The case study confirms that the suggested method enables the creation of an environmental model for an electric shuttle with an acceptable level of accuracy: the environmental impact calculated when the design phase is over, and all product information is known differs by 17% from the ML model results. This error would be acceptable even for full LCA, where results variations are registered at about 20%. The model works with varying parameters (i.e. country energy mix); the full LCA considers specific material composition for the considered electric shuttle. The error is even more satisfactory if the purpose of the study is considered: provide a surrogate LCA for the design phase.

The emissions (Climate Change indicator) over the whole lifecycle for the considered electric shuttle are 17754.181 kg CO₂; the estimated distance traveled is 100000 km, without battery replacement; the total mass is 2180 kg (frame assembly mass is 1380 kg, average load and the electronic assembly mass is 400 kg) and the estimated consumption when the shuttle is half loaded is equal to 0.229 kWh/km. The chosen model is the GBT model and the relative error is 0.90%. The highest relative error is registered by the Generalized Linear Model (20.74%). The method contributes to building knowledge in the context of eco-design as it allows the addressees to make environmentally conscious choices, the knowledge is easily consultable, and it allows an immediate understanding and interpretation of results. Future works will focus on additional significant indicators for EVs' lifecycle and on similar EV designs (i.e., both freight and passengers). The present work would significantly gain benefits from the integration with prospective LCA approaches, which consider future trends (i.e., getting data from patents) to forewarn future impacts. Ultimately, future studies could also focus on the automation of initial assumptions editing, so that the extension of the database and the error comparison would speed up and require fewer resources.

6 Glossary

Additive Manufacturing	AM	High-Voltage-Battery-Optimization-Tool	HV-BOT
Artificial Neural Network	ANN	Industrial Revolution	14.0
Artificial Intelligence	AI	Internet of Things	IoT
Attributional LCA	LCA-A	Life Cycle Assessment	LCA
Automated Guided Vehicle	AGV	Lifecycle Impact Assessment	LCIA
Bills of Materials	BOM	Life Cycle Inventory	LCI

(continued on next column)

(continued)

Additive Manufacturing	AM	High-Voltage-Battery-Optimization-Tool	HV-BOT
Classification	CL	Linear Regression	LR
Clustering	CG	Liquefied Natural Gas	LNG
Decision Tree	DT	Lithium-Ion Batteries	LIB
Deep Learning	DL	Machine Learning	ML
Electric Vehicles	EV	Neural Network	NN
End-of-Life	EoL	Regression	RL
Generalized Linear Model	GML	Random Forest	RF
Gradient Boosted Trees	GBT	Reinforce	RL
Gradient Boosting	GB	Supervised algorithm	SL
Greenhouse Gas Emissions	GHG	Non supervised	UL
		Weel-To-Wheel	WTW

CRediT authorship contribution statement

Luca Manuguerra: Writing – original draft, Validation, Software, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Federica Cappelletti:** Writing – review & editing. **Michele Germani:** Project administration.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgment

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

References

Algren, M., Fisher, W., Landis, A.E., 2021. Machine learning in life cycle assessment. In: Data Science Applied to Sustainability Analysis. Elsevier, pp. 167–190. <https://doi.org/10.1016/B978-0-12-817976-5.00009-7>.

Altmann, A., Tološi, L., Sander, O., Lengauer, T., 2010. Permutation importance: a corrected feature importance measure. *Bioinformatics* 26 (10), 1340–1347. <https://doi.org/10.1093/bioinformatics/btq134>.

Ambrose, H., Kendall, A., Lozano, M., Wachke, S., Fulton, L., 2020. Trends in life cycle greenhouse gas emissions of future light duty electric vehicles. *Transport. Res. Transport Environ.* 81, 102287 <https://doi.org/10.1016/j.trd.2020.102287>.

Antanasijević, D.Z., Ristić, M.D., Perić-Grujić, A.A., Pocajt, V.V., 2014. Forecasting GHG emissions using an optimized artificial neural network model based on correlation and principal component analysis. *Int. J. Greenh. Gas Control* 20, 244–253. <https://doi.org/10.1016/j.ijggc.2013.11.011>.

Anund, A., Ludovic, R., Caroleo, B., Hardestam, H., Dahlman, A., Skogsmo, I., Nicaise, M., Arnone, M., 2022. Lessons learned from setting up a demonstration site with autonomous shuttle operation – based on experience from three cities in Europe. *Journal of Urban Mobility* 2, 100021. <https://doi.org/10.1016/j.urbmob.2022.100021>.

Arvanitoyannis, I.S., 2008. ISO 14040: life cycle assessment (LCA) – principles and guidelines. In: Waste Management for the Food Industries. Elsevier, pp. 97–132. <https://doi.org/10.1016/B978-012373654-3.50006-7>.

Batta Mahesh, 2018. Machine learning algorithms - a review. *International Journal of Science and Research (IJSR)* 9 (1). <https://doi.org/10.21275/ART20203995>. ISSN: 2319-7064.

Bibak, B., Bai, L., 2023. An optimization approach for managing electric vehicle and reused battery charging in a vehicle to grid system under an electricity rate with demand charge. *Sustainable Energy, Grids and Networks* 36, 101145. <https://doi.org/10.1016/j.segan.2023.101145>.

Bishop, C.M., 1994. Neural networks and their applications. *Rev. Sci. Instrum.* 65 (6), 1803–1832. <https://doi.org/10.1063/1.1144830>.

Burchart-Korol, D., Jursova, S., Folęga, P., Korol, J., Pustejovska, P., Blaut, A., 2018. Environmental life cycle assessment of electric vehicles in Poland and the Czech

- Republic. *J. Clean. Prod.* 202, 476–487. <https://doi.org/10.1016/j.jclepro.2018.08.145>.
- Cappelletti, F., Rossi, M., Germani, M., 2022a. How de-manufacturing supports circular economy linking design and EoL - a literature review. *J. Manuf. Syst.* 63, 118–133. <https://doi.org/10.1016/j.jmsy.2022.03.007>.
- Cappelletti, F., Rossi, M., Manuguerra, L., Germani, M., 2022b. Design for X Tool to introduce sustainability in the design process. *Computer-Aided Design and Applications* 234–247. <https://doi.org/10.14733/cadaps.2023.234-247>.
- Castaneda, J., Cardona, J.F., Martins, L. do C., Juan, A.A., 2021. Supervised machine learning algorithms for measuring and promoting sustainable transportation and green logistics. *Transport. Res. Procedia* 58, 455–462. <https://doi.org/10.1016/j.trpro.2021.11.061>.
- Colin Cameron, A., Windmeijer, F.A.G., 1997. An R-squared measure of goodness of fit for some common nonlinear regression models. *J. Econom.* 77 (2), 329–342. [https://doi.org/10.1016/S0304-4076\(96\)01818-0](https://doi.org/10.1016/S0304-4076(96)01818-0).
- Collado-Ruiz, D., Ostad-Ahmad-Ghorabi, H., 2013. Estimating environmental behavior without performing a life cycle assessment. *J. Ind. Ecol.* 17 (1), 31–42. <https://doi.org/10.1111/j.1530-9290.2012.00510.x>.
- Dalla Mora, T., Bolzonello, E., Cavalliere, C., Peron, F., 2020. Key parameters featuring BIM-LCA integration in buildings: a practical review of the current trends. *Sustainability* 12 (17), 7182. <https://doi.org/10.3390/su12177182>.
- Dalzocho, J., Kunst, R., Pignaton, E., Binotto, A., Sanyal, S., Favilla, J., Barbosa, J., 2020. Machine learning and reasoning for predictive maintenance in Industry 4.0: current status and challenges. *Comput. Ind.* 123, 103298 <https://doi.org/10.1016/j.compind.2020.103298>.
- de Myttenaere, A., Golden, B., Le Grand, B., Rossi, F., 2016. Mean absolute percentage error for regression models. *Neurocomputing* 192, 38–48. <https://doi.org/10.1016/j.neucom.2015.12.114>.
- De Wolf, C., Cordella, M., Dodd, N., Byers, B., Donatello, S., 2023. Whole life cycle environmental impact assessment of buildings: developing software tool and database support for the EU framework Level(s). *Resour. Conserv. Recycl.* 188, 106642 <https://doi.org/10.1016/j.resconrec.2022.106642>.
- Elouariaghli, F.N., Kozderka, S.M., Quaranta, T.G., Pena, F.D., Rose, F.B., Hoarau, S.Y., 2022. Eco-design and life cycle management: consequential life cycle assessment, artificial intelligence and green IT. *IFAC-PapersOnLine* 55 (5), 49–53. <https://doi.org/10.1016/j.ifacol.2022.07.638>.
- Emami, N., Heinonen, J., Marteinsson, B., Sänäjoki, A., Junnonen, J.-M., Laine, J., Junnila, S., 2019. A life cycle assessment of two residential buildings using two different LCA database-software combinations: recognizing uniformities and inconsistencies. *Buildings* 9 (1), 20. <https://doi.org/10.3390/buildings9010020>.
- Epp, A., Wendland, R., Behrendt, J., Gerlach, R., Sauer, D.U., 2022. Holistic battery system design optimization for electric vehicles using a multiphysically coupled lithium-ion battery design tool. *J. Energy Storage* 52, 104854. <https://doi.org/10.1016/j.est.2022.104854>.
- Gao, X., Simeone, A., Zhang, J., 2023. Smart decision-making for design adaptation of electric vehicles using big sales data. *Procedia CIRP* 119, 710–715. <https://doi.org/10.1016/j.procir.2023.02.161>.
- Gayathri, R., Rani, S.U., Čepová, L., Rajesh, M., Kalita, K., 2022. A comparative analysis of machine learning models in prediction of mortar compressive strength. *Processes* 10 (7), 1387. <https://doi.org/10.3390/pr10071387>.
- Ghoroghi, A., Rezgui, Y., Petri, I., Beach, T., 2022. Advances in application of machine learning to life cycle assessment: a literature review. *Int. J. Life Cycle Assess.* 27 (3), 433–456. <https://doi.org/10.1007/s11367-022-02030-3>.
- Godzimirski, J.M., 2014. The Norwegian energy security debate: domestic and international dimensions. In: *The Political Economy of Renewable Energy and Energy Security*. Palgrave Macmillan UK, pp. 116–136. https://doi.org/10.1057/9781137338877_6.
- Graba, M., Kelouwani, S., Zeghmi, L., Amamou, A., Agbossou, K., Mohammadpour, M., 2020. Investigating the impact of energy source level on the self-guided vehicle system performances, in the industry 4.0 context. *Sustainability* 12 (20), 8541. <https://doi.org/10.3390/su12208541>.
- Hansen, R.N., Hoxha, E., Rasmussen, F.N., Ryberg, M.W., Andersen, C.E., Birgisdóttir, H., 2023. Enabling rapid prediction of quantities to accelerate LCA for decision support in the early building design. *J. Build. Eng.* 76, 106974 <https://doi.org/10.1016/j.jobe.2023.106974>.
- Heath, G.A., Hsu, D.D., Inman, D., Aden, A., Mann, M.K., 2009. Life cycle assessment of the energy independence and security act of 2007: ethanol—global warming potential and environmental emissions. *ASME 2009 3rd International Conference on Energy Sustainability* 1, 857–861. <https://doi.org/10.1115/ES2009-90037>.
- Ivanov, D., Tang, C.S., Dolgui, A., Battini, D., Das, A., 2021. Researchers' perspectives on Industry 4.0: multi-disciplinary analysis and opportunities for operations management. *Int. J. Prod. Res.* 59 (7), 2055–2078. <https://doi.org/10.1080/00207543.2020.1798035>.
- Jakub, S., Adrian, L., Mieczysław, B., Ewelina, B., Katarzyna, Z., 2022. Life cycle assessment study on the public transport bus fleet electrification in the context of sustainable urban development strategy. *Sci. Total Environ.* 824, 153872 <https://doi.org/10.1016/j.scitotenv.2022.153872>.
- Jusselme, T., Rey, E., Andersen, M., 2018. An integrative approach for embodied energy: towards an LCA -based data-driven design method. *Renew. Sustain. Energy Rev.* 88, 123–132. <https://doi.org/10.1016/j.rser.2018.02.036>.
- Kamm, S., Veekati, S.S., Müller, T., Jazdi, N., Weyrich, M., 2023. A survey on machine learning based analysis of heterogeneous data in industrial automation. *Comput. Ind.* 149, 103930 <https://doi.org/10.1016/j.compind.2023.103930>.
- Karka, P., Papadokonstantakis, S., 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 <https://doi.org/10.1016/j.ces.2021.117339>.
- Koffler, C., Amor, B., Carbajales-Dale, M., Cascio, J., Conroy, A., Fava, J.A., Gaudreault, C., Gloria, T., Hensler, C., Horvath, A., Humbert, S., Manzardo, A., Margni, M., Osset, P., Sinistore, J., Vigon, B., Wallace, M.L., Wang, M., Prox, M., 2020. On the reporting and review requirements of ISO 14044. *Int. J. Life Cycle Assess.* 25 (3), 478–482. <https://doi.org/10.1007/s11367-019-01706-7>.
- Kumi, L., Jeong, J., 2023. Optimization model for selecting optimal prefabricated column design considering environmental impacts and costs using genetic algorithm. *J. Clean. Prod.* 417, 137995 <https://doi.org/10.1016/j.jclepro.2023.137995>.
- Lai, X., Chen, J., Chen, Q., Han, X., Lu, L., Dai, H., Zheng, Y., 2023. Comprehensive assessment of carbon emissions and environmental impacts of sodium-ion batteries and lithium-ion batteries at the manufacturing stage. *J. Clean. Prod.* 423, 138674 <https://doi.org/10.1016/j.jclepro.2023.138674>.
- LeCun, Y., Bengio, Y., Hinton, G., 2015. Deep learning. *Nature* 521 (7553), 436–444. <https://doi.org/10.1038/nature14539>.
- Liu, Y., Wang, Y., Zhang, J., 2012. New machine learning algorithm: random forest, 246–252. https://doi.org/10.1007/978-3-642-34062-8_32.
- Lyutov, A., Uygun, Y., Hütt, M.-T., 2019. Managing workflow of customer requirements using machine learning. *Comput. Ind.* 109, 215–225. <https://doi.org/10.1016/j.compind.2019.04.010>.
- Ma, J., Chen, X., Xing, Z., Zhang, Y., Yu, L., 2023. Improving the performance of airport shuttle through demand-responsive service with dynamic fare strategy considering mixed demand. *J. Air Transport. Manag.* 112, 102459 <https://doi.org/10.1016/j.jairtraman.2023.102459>.
- Maiorino, A., Cilenti, C., Petruzzello, F., Aprea, C., 2024. A review on thermal management of battery packs for electric vehicles. *Appl. Therm. Eng.* 238, 122035 <https://doi.org/10.1016/j.applthermeng.2023.122035>.
- Mandolini, M., Marconi, M., Rossi, M., Favi, C., Germani, M., 2019. A standard data model for life cycle analysis of industrial products: a support for eco-design initiatives. *Comput. Ind.* 109, 31–44. <https://doi.org/10.1016/j.compind.2019.04.008>.
- Manuguerra, L., Cappelletti, F., Manes, F., Germani, M., 2023. A predictive eco-design method and tool for electric vehicles of industry 4.0. *Procedia Comput. Sci.* 217, 248–257. <https://doi.org/10.1016/j.procs.2022.12.220>.
- Marmiroli, B., Venditti, M., Dotelli, G., Spessa, E., 2020. The transport of goods in the urban environment: a comparative life cycle assessment of electric, compressed natural gas and diesel light-duty vehicles. *Appl. Energy* 260, 114236. <https://doi.org/10.1016/j.apenergy.2019.114236>.
- Marta, R., Federica, C., Michele, G., 2023. A step forward life cycle assessment to optimize products and increase company eco-design competencies, 62–74. https://doi.org/10.1007/978-3-031-15928-2_6.
- McLaren, J., Miller, J., O'Shaughnessy, E., Wood, E., Shapiro, E., 2016. CO2 emissions associated with electric vehicle charging: the impact of electricity generation mix, charging infrastructure availability and vehicle type. *Electr. J.* 29 (5), 72–88. <https://doi.org/10.1016/j.tej.2016.06.005>.
- Ministero delle infrastrutture e dei trasporti, 2015. *Regolamento recante norme relative all'individuazione dei criteri di assimilazione ai fini della guida e della circolazione ed all'accertamento dei requisiti tecnici di idoneità della «navetta turistica»*, 15G00206.
- Mitchell, Tom M., 1997. *Machine Learning*. McGraw-Hill Science/Engineering/Math (March 1, 1997). ISBN: 0070428077.
- Mladenović, D., 2006. Feature selection for dimensionality reduction, 84–102. https://doi.org/10.1007/11752790_5.
- Moro, A., Helmers, E., 2017. A new hybrid method for reducing the gap between WTW and LCA in the carbon footprint assessment of electric vehicles. *Int. J. Life Cycle Assess.* 22 (1), 4–14. <https://doi.org/10.1007/s11367-015-0954-z>.
- Mossali, E., Gentilini, L., Merati, G., Colledani, M., 2020. Methodology and application of electric vehicles battery packs redesign for circular economy. *Procedia CIRP* 91, 747–751. <https://doi.org/10.1016/j.procir.2020.01.139>.
- Natekin, A., Knoll, A., 2013. Gradient boosting machines, a tutorial. *Front. Neurobot.* 7 <https://doi.org/10.3389/fnbot.2013.00021>.
- Nordelöf, A., Messagie, M., Tillman, A.-M., Ljunggren Söderman, M., Van Mierlo, J., 2014. Environmental impacts of hybrid, plug-in hybrid, and battery electric vehicles—what can we learn from life cycle assessment? *Int. J. Life Cycle Assess.* 19 (11), 1866–1890. <https://doi.org/10.1007/s11367-014-0788-0>.
- Onat, N.C., Abdella, G.M., Kucukvar, M., Kutty, A.A., Al-Nuaimi, M., Kumbaroğlu, G., Bulu, M., 2021. How eco-efficient are electric vehicles across Europe? A regionalized life cycle assessment-based eco-efficiency analysis. *Sustain. Dev.* 29 (5), 941–956. <https://doi.org/10.1002/sd.2186>.
- Ostad Ahmad Ghorabi, H., Collado-Ruiz, D., Wimmer, W., 2009. *Towards integrating LCA into CAD*. In: Norell, M., Grimheden, M., Skogstad, P. (Eds.), *Proceedings of the International Conference on Engineering Design*.
- Ostad-Ahmad-Ghorabi, H., Collado-Ruiz, D., 2011. Tool for the environmental assessment of cranes based on parameterization. *Int. J. Life Cycle Assess.* 16 (5), 392–400. <https://doi.org/10.1007/s11367-011-0280-z>.
- Pan, Y., Zhang, W., Niu, S., 2021. Emission modeling for new-energy buses in real-world driving with a deep learning-based approach. *Atmos. Pollut. Res.* 12 (10), 101195. <https://doi.org/10.1016/j.apr.2021.101195>.
- Peffley, T.B., Pearce, J.M., 2020. The potential for grid deflection of small and medium sized enterprises using solar photovoltaic, battery and generator hybrid systems. *Renew. Energy* 148, 193–204. <https://doi.org/10.1016/j.renene.2019.12.039>.
- Petrauskienė, K., Skvarnavičiūtė, M., Dvarionienė, J., 2020. Comparative environmental life cycle assessment of electric and conventional vehicles in Lithuania. *J. Clean. Prod.* 246, 119042 <https://doi.org/10.1016/j.jclepro.2019.119042>.

- Picatoste, A., Justel, D., Mendoza, J.M.F., 2022. Exploring the applicability of circular design criteria for electric vehicle batteries. *Procedia CIRP* 109, 107–112. <https://doi.org/10.1016/j.procir.2022.05.222>.
- Prioux, N., Ouaret, R., Hêtreux, G., Belaud, J.-P., 2022. Environmental assessment coupled with machine learning for circular economy. *Clean Technol. Environ. Policy*. <https://doi.org/10.1007/s10098-022-02275-4>.
- Rapid miner. https://docs.rapidminer.com/latest/studio/operators/validation/performance/predictive/performance_regression.html, 2023.
- Rapid miner. <https://docs.rapidminer.com/9.3/studio/auto-model/>, 2023.
- Raugei, M., Winfield, P., 2019. Prospective LCA of the production and EoL recycling of a novel type of Li-ion battery for electric vehicles. *J. Clean. Prod.* 213, 926–932. <https://doi.org/10.1016/j.jclepro.2018.12.237>.
- Raugei, M., Morrey, D., Hutchinson, A., Winfield, P., 2015. A coherent life cycle assessment of a range of lightweighting strategies for compact vehicles. *J. Clean. Prod.* 108, 1168–1176. <https://doi.org/10.1016/j.jclepro.2015.05.100>.
- Roberto Marmo, 2020. Algoritmi per l'intelligenza artificiale. *Progettazione dell'algoritmo, dati e machine learning, neural network, deep learning*. Hoepli.
- Rossi, M., Cappelletti, F., Germani, M., 2022. Design for environmental sustainability: collect and use company information to design green products. *Procedia CIRP* 105, 823–828. <https://doi.org/10.1016/j.procir.2022.02.136>.
- Saltz, J.S., 2021. CRISP-DM for data science: strengths, weaknesses and potential next steps. In: 2021 IEEE International Conference on Big Data (Big Data), pp. 2337–2344. <https://doi.org/10.1109/BigData52589.2021.9671634>.
- Schwendemann, S., Amjad, Z., Sikora, A., 2021. A survey of machine-learning techniques for condition monitoring and predictive maintenance of bearings in grinding machines. *Comput. Ind.* 125, 103380 <https://doi.org/10.1016/j.compind.2020.103380>.
- Shang, H., Sun, Y., Huang, D., Meng, F., 2024. Life cycle assessment of atmospheric environmental impact on the large-scale promotion of electric vehicles in China. *Resources, Environment and Sustainability*, 100148. <https://doi.org/10.1016/j.resenv.2024.100148>.
- Sousa, I., Wallace, D., 2006. Product classification to support approximate life-cycle assessment of design concepts. *Technol. Forecast. Soc. Change* 73 (3), 228–249. <https://doi.org/10.1016/j.techfore.2004.03.007>.
- Spreafico, C., Landi, D., 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. <https://doi.org/10.1016/j.spc.2023.04.006>.
- Su, X., Yan, X., Tsai, C.-L., 2012. Linear regression. *Wiley Interdisciplinary Reviews: Comput. Stat.* 4 (3), 275–294. <https://doi.org/10.1002/wics.1198>.
- Subal, L., Braunschweig, A., Hellweg, S., 2024. The relevance of life cycle assessment to decision-making in companies and public authorities. *J. Clean. Prod.* 435, 140520 <https://doi.org/10.1016/j.jclepro.2023.140520>.
- Wang, H.S., Wang, Y.N., Wang, Y.C., 2013. Cost estimation of plastic injection molding parts through integration of PSO and BP neural network. *Expert Syst. Appl.* 40 (2), 418–428. <https://doi.org/10.1016/j.eswa.2012.01.166>.
- Wang, F., Intrator, C., Salopek, N., Yuan, C., 2022. An environmental sustainability analysis tool for next generation lithium ion batteries of electric vehicles. *Procedia CIRP* 105, 489–494. <https://doi.org/10.1016/j.procir.2022.02.081>.