A Method Based on Machine Learning Techniques for the Development of a Parametric Environmental Impact Model for Industrial Electric Vehicles

A method based on machine learning techniques for the development of a parametric environmental impact model for industrial electric vehicles

Luca Manuguerra[1][0000-0003-0832-3886], Federica Cappelletti[1][0000-0001-5592-3150], Marta Rossi[2][0000-0001-9287-8109] and Michele Germani[1][0000-0003-1988-8620]

[1] Università Politecnica delle Marche, Via Brecce Bianche 12, 60131, Ancona, Italy
[2] Università eCampus, Via Isimbardi, 10 - 22060 Novedrate
l.manuguerra@pm.univpm.it

Abstract. Designers need to be aware early in the design phase of the environmental impact of their choices over the entire product life cycle. This paper proposes an eco-design method to support designers of different categories of electric vehicles, such as self-driving vehicles, cars, shuttles and buses. The methodology developed aims to realize a model for predicting the environmental impact of industrial electric vehicles. The proposed approach exploits machine learning methods to develop models with the design features of a generic electric vehicle, such as vehicle mass and distance traveled during its entire lifetime as independent parameters, to estimate the emissions of new products. The environmental impact indicator for this study is Climate Change, the dependent parameter chosen for the impact model. Machine learning algorithms were trained on training data retrieved from an automatic environmental impact estimation software tool based on an analytical approach. 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. Finally, the model is tested by estimating the environmental impact of a tourist shuttle.

Keywords: Machine Learning, Data Science, Sustainability, Design, Life Cycle Assessment.

1 Introduction

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 [1]. This necessitates the design team to be capable of the early design process evaluation of the environmental performance of numerous potential designs [2]. Many tools aid designers in making informed judgments; many are based on the Life Cycle Assessment (LCA) technique [3]. The life cycle impact of a product or service is assessed using the analytical and methodical process known as LCA. The computation includes the phases of raw material extraction, production, distribution, usage, and final disposal of the product, returning the environmental impact values related to its life cycle. The reference standards are ISO 14040 and ISO 14044. The practical application of eco-design and circular economy approaches within the industrial sector contributes to the reduction of environmental impacts and the creation of economic value [4].

The environmental impact of electric vehicles is frequently studied using the LCA method; in particular, these analyses frequently compare electric vehicles to internal combustion vehicles or, as in the case of the study by J. Szczurowski et al. [5], examine vehicles used for passenger transportation in the context of public transportation and electro-mobility. A comprehensive and entirely consistent LCA-based assessment of a variety of light weighting strategies for compact passenger vehicles is proposed by Raugei et al. [6]. Knowing the key characteristics of the products, Marmiroli et al. [7] compare the effects of electric, compressed natural gas, and light-duty diesel cars. Held et al. [9] examined the impact of the use phase in the LCA in two commercial mobility applications. Manuguerra et al. [8] proposed a method and tool to predict the environmental impact of industrial electric vehicles in the design phase considering all life cycle phases.

Broader use of Machine Learning (ML) is challenging to implement because of the enormous amount of input factors and their uncertainties that affect the entire life cycle. ML techniques belong to a class of Artificial Intelligence (AI) techniques that can learn from data to increase their accuracy without reprogramming. By analyzing a collection of training data and generating an algorithm without the assistance of a person, ML creates a model that can identify patterns [9]. ML algorithms are typically categorized into four groups: supervised learning, unsupervised learning, semi-supervised learning, and reinforcement learning [10].

ML is one area of data science that is now used to fill up data gaps for LCA [11]. 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.

Literature offers many examples of ML applications for predicting emissions from electric vehicles. Pan et al. [12] designed a deep learning framework based on the recurrent gate to capture external factors and time dependence characteristics to predict real-time emission rates for LNG buses under naturalistic driving conditions. Antanasijevic et al. [13] 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.

The method proposed in this paper aims to create a predictive model that allows it to be used in the initial design phase to analyze the environmental impact; the category chosen is Climate Change which allows verifying the kg CO2 eq produced by an industrial electric vehicle. Finally, the method is tested to verify its applicability on a tourist shuttle by estimating the environmental impact.

2 Method

The method is a standard procedure for solving a problem through data science. It takes inspiration from the CRISP-DM method. The novelty of the method lays on its application in the environmental sustainability and eco-design fields. It consists of two main steps (Fig. 1). The first step involves creating the database. For the design phase, all product life cycle stages have been considered in the construction of the database. The second step of the method is the choice of the best algorithm for predicting results. The method can be applied using any type of software, the important thing is that they perform the required function. The method can also be used with other machine learning techniques and other metrics to evaluate the error committed.

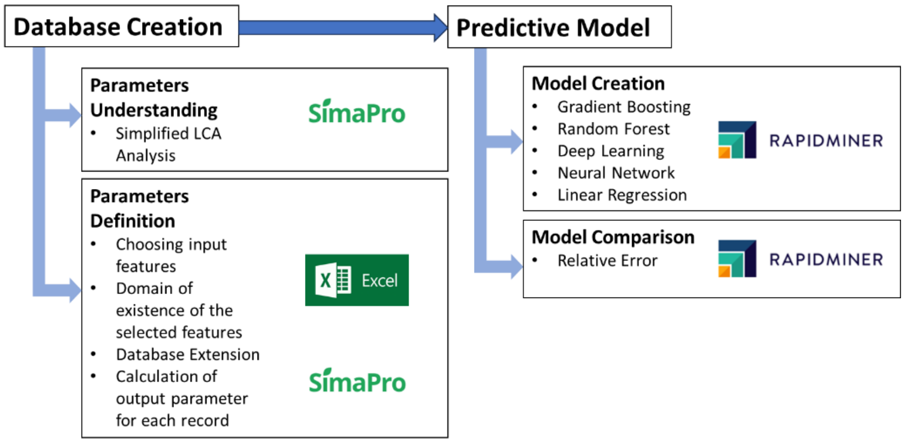

Fig. 1. Method diagram

2.1 Database Creation

The goal of this step is to obtain the database on which to train and test the model. First, it is necessary to understand the vehicle benchmarks and define their impact on the phases that make up the life cycle.

Parameters Understanding

A simplified LCA analysis is performed to understand the benchmarks. It is defined as simplified because very little data are available, and the lack of information is made up of data taken from literature or by choosing products already modeled within the software used for LCA analysis. The phases examined in this study are the material phase, the use phase, the maintenance phase and the end-of-life phase. In this way, it is possible to understand the most impactful phases and start identifying the parameters that influence the phases.

Parameters Definition

In this sub-step the database is created. A generic database is structured by I) Number of independent variables: the number of parameters related to the electric vehicle used to define the predictive model, called features; II) Number of "records": the number of elements representative of different vehicle alternatives from which to extrapolate the independent variables. The creation of the database consists of 4 stages: I) Choice of input features; II) Domain of existence of the selected features; III) Database extension; IV) Calculation of output parameter for each record.

Choosing input features

At this stage, those parameters are identified for each phase that is to be analyzed that are the most significant and that by their variability go a long way in influencing the results. The features can be different: nominal, which is a parameter that allows predefined names as values; in this way, it is possible to distinguish one object from another. They can be binomial (true or false, male or female) or polynomial (more than two options), numeric, parameters that represent numeric values (integers, real, or dates).

The domain of existence of the selected features

To use the database, it is necessary to define the range in which each feature varies, consistent with the objectives of the analysis and the natural values associated with the attribute [14]. The variables chosen can be qualitative or quantitative when numerical values express their values or intensities. Quantitative variables are divided into discrete and continuous. A fundamental technique in this phase that allows the transformation of attributes is discretization, through which it is possible to convert a continuous attribute into a discrete one, dividing the domain into a defined number of intervals. The use of clustering techniques to identify optimal ranges can also be used.

Database Extension

Once the features have been defined, it is possible to identify all possible combinations of vehicle parameters. However, not all combinations are realistic; some will be discarded.

Calculation of output parameter for each record

The last step in creating the database is to choose the dependent variable that the ML model will predict, usually called the dependent parameter, which acts as an indicator to assess the environmental impact of an electric vehicle. This parameter is correlated with the features chosen as input to the model. For this study, the impact category chosen is Climate Change; any other category can be chosen.

2.2 Predictive Model

The step to obtain the predictive model is further divided into two sub-steps: Model Creation and Interpretation Model.

Model Creation

Modeling is the stage in which numerous models are created and assessed using various methods. The key objective is finding the best ML methods to predict environmental impact items (one for each parametric impact model). To do that, several ML techniques, such as Gradient Boosting (GB) [15], Random Forest (RF) [16], Deep Learning (DL) [17], Neural Network (NN) [18] and Linear Regression (LR) [19], must be contrasted. The creation of a model for environmental impact begins by dividing the database into training (for example, 70%) and testing sets (for example, 30%). Software tools for data mining or data sciences must be used for the modeling. Finally, each algorithm generates a prediction model.

Model Comparison

For each model, the testing set is utilized to calculate metrics like Relative Error [20]. Then, the ideal method is chosen based on these factors. The algorithm chosen is the one that allows obtaining the lowest error, based on the metric of the choice.

3 Case study

The case study was used to verify that the proposed method was applicable to obtain a parametric prediction model for the Climate Change impact category [kg CO2 eq] for a tourist shuttle. First, a simplified LCA analysis was conducted on SimaPro that allowed identifying the parameters shown in Fig. 2 and its results are shown in Table 1.

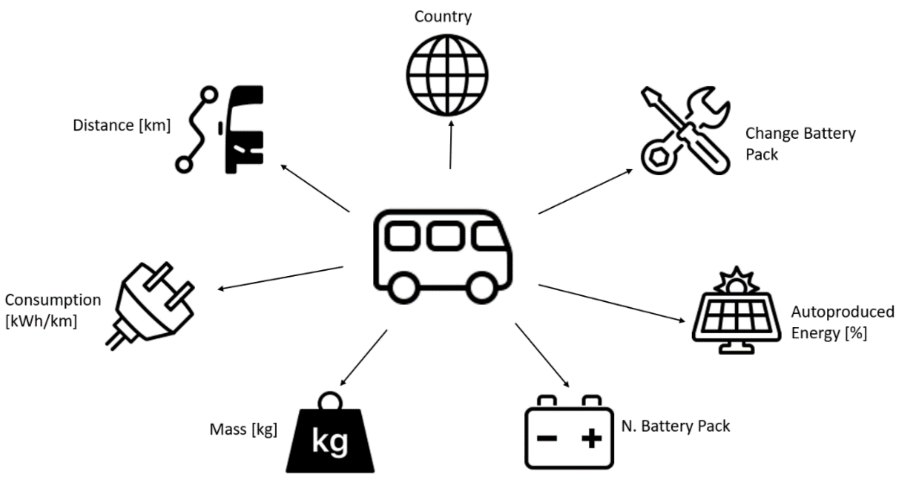

Fig. 2. Shuttle parameters

The phases examined in this study are the material phase, the use phase, the maintenance phase and the end-of-life phase. The material phase of the electric vehicle was modeled by secondary data from SimaPro, provided with the EcoInvent database. The software analyzes the battery pack separately from the rest of the electric vehicle. Components modeled in the material phase influence the impact of end-of-life. The use phase is linked to different aspects, such as the distance traveled, the consumption per kilometer traveled, the reference country's electrical mix, and the percentage of self-produced energy through a private photovoltaic system. Finally, the maintenance phase is closely linked to the replacement of the battery pack.

Through a literature study [6–8], it was possible to identify the domain of existence of each parameter to discretize and identify all the possible realistic combinations (Table 1). The Maximum and Minimum columns do not refer to Climate Change impact values, but to the boundaries of existence ranges. The third column shows the result of the simplified LCA analysis Through SimaPro, it was possible to analyze the impact in terms of kg CO2 eq for each record. This resulted in a database of 6000 records.

Table 1. Shuttle parameters domain of existence

Parameters	**Minimum**	**Maximum**	**Simplified LCA**
Distance [km]	50000	250000	150000
Consumption [kWh/km]	0,2	0,6	0,4
Mass [kg]	1500	5000	2500
N. Battery Pack	1	4	2
Auto-produced energy [%]	0%	75%	50%
Country	Norway	Italy	Italy
Change Battery Pack	No	Yes	Yes
Material phase [kg CO2 eq]	6675	23072	11536
Use phase [kg CO2 eq]	256	30961	20310
Maintenance phase [kg CO2 eq]	376	9924	3710
End of Life phase [kg CO2 eq]	17629	72098	36049
Total [kg CO2 eq]	24936	136054	71605

For the Prediction Model step RapidMiner (by Altair Engineering) was used. It is a data science platform with a wide variety of ML techniques. The Auto Model plugin for RapidMiner Studio streamlines the creation and validation of models. Although data science software validation is not as thorough as full cross-validation, this approach balances runtime and model validation quality [20]. Five algorithms: Generalized Linear Model (GLM), DL, Decision Tree (DT), RF, and Gradient Boosted Trees (GBT), can be tested and compared using RapidMiner.

4 Results and discussion

The Relative Error is simple for engineers to understand and directly relates to the initial design phase, so the authors adopt it. The results show that neither GLM nor DL can ensure the necessary accuracy. Generally, decision trees respect the constraints and outperform deep learning and linear regression. Mainly, GBT performs better than RF and DT (Table 2).

Table 2. Relative Error for Climate Change [kg CO2 eq.] shuttle predictive model

Model	**Relative Error [%]**
Generalized Linear Model	10,14
Deep Learning	2,14
Decision Tree	2,64
Random Forest	8,74
Gradient Boosted Trees	0,63

Software for life cycle evaluation allowed for the creation of a trustworthy database for training ML algorithms. Moreover, LCA estimation software enabled the collection of impacts under various conditions. Therefore, the suggested method is not dependent on any specific piece of software. The concept can be used using software tools from data science and other analytical environmental impact estimation disciplines.

Every parameter selected for the predictive models is known at the early stages of a tourist shuttle's design. The critical drawback of this research is that it only provides a comprehensive estimate of the environmental impact of a tourist shuttle. The detailed analysis received at the design's conclusion from a comprehensive LCA analysis is not replaced by the results of this approach.

5 Conclusion

The method for creating parametric models for electric vehicles at the conceptual or early design phases was provided in the study. The method is based on ML algorithms that were taught using databases created by an autonomous software tool for estimating the impact of LCAs.

One model for a shuttle for tourists was created using the methodology. The case study allowed for the real-world testing of the method's suggested stages. The strategy also made it possible to choose the most effective ML algorithm for this challenge.

Further research should evaluate the approach's applicability for additional electric vehicles by testing it on different goods and industrial contexts.

References

1. Cappelletti, F., Rossi, M., Germani, M.: How de-manufacturing supports circular economy linking design and EoL - a literature review. J Manuf Syst. 63, 118–133 (2022). https://doi.org/10.1016/j.jmsy.2022.03.007.
2. Sousa, I., Wallace, D.: Product classification to support approximate life-cycle assessment of design concepts. Technol Forecast Soc Change. 73, 228–249 (2006). https://doi.org/10.1016/j.techfore.2004.03.007.

3. Marta, R., Federica, C., Michele, G.: A Step Forward Life Cycle Assessment to Optimize Products and Increase Company Eco-design Competencies. Presented at the (2023). https://doi.org/10.1007/978-3-031-15928-2_6.
4. Lamba, H.K., Kumar, N.S., Dhir, S.: Circular economy and sustainable development: a review and research agenda. International Journal of Productivity and Performance Management. (2023). https://doi.org/10.1108/IJPPM-06-2022-0314.
5. Jakub, S., Adrian, L., Mieczysław, B., Ewelina, B., Katarzyna, Z.: Life cycle assessment study on the public transport bus fleet electrification in the context of sustainable urban development strategy. Science of The Total Environment. 824, 153872 (2022). https://doi.org/10.1016/j.scitotenv.2022.153872.
6. Raugei, M., Morrey, D., Hutchinson, A., Winfield, P.: A coherent life cycle assessment of a range of lightweighting strategies for compact vehicles. J Clean Prod. 108, 1168–1176 (2015). https://doi.org/10.1016/j.jclepro.2015.05.100.
7. Marmiroli, B., Venditti, M., Dotelli, G., Spessa, E.: 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 (2020). https://doi.org/10.1016/j.apenergy.2019.114236.
8. Manuguerra, L., Cappelletti, F., Manes, F., Germani, M.: A predictive eco-design method and tool for electric vehicles of Industry 4.0. Procedia Comput Sci. 217, 248–257 (2023). https://doi.org/10.1016/j.procs.2022.12.220.
9. Tom M. Mitchell: Machine Learning. (1997).
10. Batta Mahesh: Machine Learning Algorithms - A Review. International Journal of Science and Research (IJSR). (2018).
11. Ghoroghi, A., Rezgui, Y., Petri, I., Beach, T.: Advances in application of machine learning to life cycle assessment: a literature review. Int J Life Cycle Assess. 27, 433–456 (2022). https://doi.org/10.1007/s11367-022-02030-3.
12. Pan, Y., Zhang, W., Niu, S.: Emission modeling for new-energy buses in real-world driving with a deep learning-based approach. Atmos Pollut Res. 12, 101195 (2021). https://doi.org/10.1016/j.apr.2021.101195.
13. Antanasijević, D.Z., Ristić, M.Đ., Perić-Grujić, A.A., Pocajt, V. v.: Forecasting GHG emissions using an optimized artificial neural network model based on correlation and principal component analysis. International Journal of Greenhouse Gas Control. 20, 244–253 (2014). https://doi.org/10.1016/j.ijggc.2013.11.011.
14. Algren, M., Fisher, W., Landis, A.E.: Machine learning in life cycle assessment. In: Data Science Applied to Sustainability Analysis. pp. 167–190. Elsevier (2021). https://doi.org/10.1016/B978-0-12-817976-5.00009-7.
15. Natekin, A., Knoll, A.: Gradient boosting machines, a tutorial. Front Neurorobot. 7, (2013). https://doi.org/10.3389/fnbot.2013.00021.
16. Liu, Y., Wang, Y., Zhang, J.: New Machine Learning Algorithm: Random Forest. Presented at the (2012). https://doi.org/10.1007/978-3-642-34062-8_32.
17. LeCun, Y., Bengio, Y., Hinton, G.: Deep learning. Nature. 521, 436–444 (2015). https://doi.org/10.1038/nature14539.
18. Bishop, C.M.: Neural networks and their applications. Review of Scientific Instruments. 65, 1803–1832 (1994). https://doi.org/10.1063/1.1144830.
19. Su, X., Yan, X., Tsai, C.-L.: Linear regression. Wiley Interdiscip Rev Comput Stat. 4, 275–294 (2012). https://doi.org/10.1002/wics.1198.
20. RapidMiner Documentation: https://docs.rapidminer.com/9.3/studio/auto-model/.