Machine learning for predicting environmental impacts in intensive livestock farming, overcoming data challenges

L. Foschi[1], D. Pennino[1], M. Marconi[1], M. Barbanera[1], G. Grossi[2], F. Cesarini[2], C. Rossi[2], L. Bava[3], A. Vitali[2] and N. Lacetera[2]

[1]Università degli Studi della Tuscia, Dipartimento di Economia, Ingegneria, Società e Impresa, Largo dell'Università, Viterbo, 01100, Italy
[2]Università degli Studi della Tuscia, Dipartimento di Scienze Agrarie e Forestali, Via San Camillo de Lellis, Viterbo, 01100, Italy
[3]Università degli Studi di Milano, Dipartimento di Scienze Agrarie e Ambientali, via Celoria 2, Milano, 20133, Italy
Corresponding Author: lorena.foschi@unitus.it

Abstract

Life Cycle Assessment (LCA) is a crucial methodology for evaluating environmental impacts, and its integration with machine learning (ML) regression offers promising new applications. While LCA is widely used, its combination with ML can significantly enhance predictive accuracy. However, the importance of robust datasets is often overlooked, and LCA datasets frequently suffer from issues such as missing values, which compromise the accuracy and reliability of the assessments. This study addresses these challenges by leveraging ML regression not only for final predictions but also as a tool for cleaning and preparing datasets. We propose a systematic approach to identify and select the most suitable regression algorithm for a dataset with missing values. This approach involves analyzing dataset characteristics and applying different ML regression techniques to find the best fit. Our method was validated by applying it to three distinct datasets, each with unique data quality issues. The results demonstrate that using ML regression for both prediction and data cleaning can significantly improve the robustness and reliability of LCA assessments..

Keywords: life cycle assessment (LCA), carbon footprint, machine learning regression, environmental impact prediction, intensive dairy cattle farming
Presented at the ICAR Anual Conference 2024 in Bled at the Session 11: Methane Emission-Free Communications: Genetics, Environmental, and Life Cycle Assessment Studies

Introduction

The livestock sector accounts for about 12% of total anthropogenic greenhouse gas (GHG) emissions, with dairy production contributing 30% of this total (FAO, 2022). In recent years, consumers have increasingly demanded environmental information, demonstrating a significant interest in the environmental impacts of the agri-food sector (Potter *et al*., 2022). The sustainability of animal-based products is a highly topical issue, with environmental concerns leading to conflicting perceptions among stakeholders (Leroy *et al*., 2022). Therefore, it is becoming necessary to provide adequate information to make society aware of the environmental impacts of agriculture and livestock production (Stygar *et al*., 2022). Life Cycle Assessment (LCA) is a valuable and standardized methodology (ISO 14040:2006; ISO 14044:2006) used to estimate the emissions per unit of product and identify the sources of environmental burden along the supply chain (Rossi *et al*., 2024; 2023). LCA provides a holistic approach to account for both direct and indirect environmental emissions and has been effectively

applied in numerous studies focusing on food of animal origin and the majority of which considered emissions of GHGs (Thumba *et al*., 2021).

In conducting LCA studies, one of the most significant challenges is the acquisition of a sufficiently extensive and accurate dataset of primary data. This difficulty is inherent to the methodology itself, as the accuracy of environmental impact assessments hinges greatly upon the availability and quality of primary data. However, obtaining such data can be arduous and resource-intensive, requiring meticulous data collection efforts across various stages of the product lifecycle. In instances where primary data collection proves impractical or unfeasible, it may be necessary to consider leveraging existing databases or literature data, commonly referred to as secondary data. While secondary data can offer a valuable alternative, their use introduces complexities and uncertainties, as they are inherently reliant on the assumptions and methodologies employed in previous studies. Consequently, the reliability and accuracy of the results obtained from secondary data depend heavily on the validity and applicability of these assumptions, underscoring the need for careful consideration and scrutiny. Recognizing these challenges, researchers have increasingly explored the integration of Machine Learning (ML) techniques to augment traditional LCA methodologies. The use of ML algorithms offers a potential solution to address common limitations encountered in LCA studies, particularly pertaining to inventory compilation and completeness, as well as environmental impact calculation or estimation. ML offers the potential to automate and optimize data processing tasks, enhance the accuracy of predictive models, and uncover complex patterns and relationships within large datasets. As such, the integration of ML holds promise in overcoming data-related challenges and advancing the capabilities of LCA methodologies in assessing environmental impacts comprehensively and effectively..

LCA critical data aspects in dairy farms

However, the application of this methodology has encountered some challenges when applied to animal-based products, primarily due to difficulties in assessing emissions arising from biological processes (Lanzoni *et al*., 2023). The rationale behind these limitations stems from the methodology's original design, which was primarily intended to evaluate environmental emissions in industrial processes before being adapted for use in various fields, including agriculture (Caffrey and Veal, 2013). The challenge of adapting the LCA framework to farming systems lies in identifying additional key aspects that must be considered to use this approach effectively for livestock production and achieve an appropriate environmental evaluation. Applying LCA to agricultural systems is challenging due to their varied factors and dynamic connections. Particularly difficult is the large amount of data required for the LCA, which may have an impact on the quality of the data and the predictability of the results. (Nirmala *et al*., 2023). Seasonal conditions and long-term processes involving field and herd management make it challenging to gather reliable information, requiring significant time to develop a comprehensive understanding of farm activities and interactions (Caffrey and Veal, 2013). Additionally, the low accuracy and limited accessibility of primary data collection can result in a lack of representativeness, insufficiency, or even a complete absence of necessary input data for LCA analysis at farm level (Pradeleix *et al*., 2022).

Machine Learning contribution to data prediction

The contemporary landscape unambiguously demonstrates the pervasive influence of data-driven paradigms, as evidenced by the fact that an enormous amount of data is continuously being generated and collected every second by almost every electronic device currently in use.

Being able to manage and use this large amount of data is quite complex, but in recent years thanks to research on artificial intelligence, the valorization and use of this data is becoming possible. This field of research is itself divided into numerous subcategories, including *Machine Learning* (ML). Machine learning is a field of research that concerns the study and implementation of mathematical algorithms capable of learning from data and interpolating or predicting unseen data. The ability to analyze huge amounts of data and extract insights has become a crucial requirement in many use cases.

Machine learning approaches can be divided into three main categories, which correspond to the ways in which the system can be trained.

- **Supervised learning:** The system is trained with tuples of inputs and the desired outputs. The objective is to learn a general rule that maps inputs to outputs.
- **Unsupervised learning:** The system is trained by providing only input data, letting it independently search for a structure in it. This type of learning is widely used when you want to discover hidden patterns within a dataset.
- **Reinforcement learning:** The system is trained using a reward strategy. As the system searches for patterns within the data, as with the previous category, it is provided with feedback on its learning. These feedbacks are a sort of reward and the system tries to maximize them.

In general, when the term “data prediction” is employed, it is understood to refer to a machine learning model that performs a regression analysis with a supervised approach. This type of analysis follows a statistical process to estimate a mathematical function between input and output data. These predictive models are used in numerous fields, from engineering sciences to social networks, agriculture, finance, security, etc.

Machine learning has become a tool capable of providing information that allows those who are using it to make choices based on statistical models rather than on intuitions, conjectures, or suppositions. Moreover, the scalability and automation enabled by machine learning algorithms facilitate process optimization, reduced operational costs, and increased productivity, thus promoting organizational growth and sustainability in an environment characterized by increasing dependency on data. To streamline the discussion, the term “Machine Learning” will be used throughout the remainder of the paper to refer to supervised regression models.

Problem formalization

The need to make the Life Cycle Assessment methodology increasingly accurate in predicting environmental impacts is leading to its conjunction with Machine Learning techniques. Current research in this field is mainly focused on improving LCA models through the application of ML algorithms, with the aim of improving the accuracy and efficiency of environmental impact assessments. A growing literature shows how machine learning is enhancing the precision of predictions derived from LCA studies. For instance, Romeiko *et al*. (2024) state that ML applications are concentrated on those stages of LCA that necessitate a substantial investment of effort, such as inventory, impact assessment, and result interpretation. With regard to the initial stage, ML was employed to estimate the overall life cycle inventory (LCI) and to predict the product properties required for the realization of the final LCI. The main application of ML to support LCA studies is the use of algorithms to predict environmental impacts. The agricultural sector represents the most extensively analyzed sector in this regard, with algorithms employed to estimate yield, energy use, and life cycle impacts. Lastly, in the interpretation stage of the results, ML was employed in a variety of ways. These included the minimization of assessed impacts, the identification of patterns and the main drivers of life cycle impacts, the understanding of uncertainty and the sensitivity

of the results and the classification and assessment of relationships between indicators and environmental impacts. In light of the aforementioned considerations, there are only a few works that exploit ML to optimize the quality of the dataset. While machine learning can significantly aid in predicting environmental impacts, the reliability of these predictions depends on the availability of high-quality data. Consequently, there is a clear need for reliable data to support more accurate ML and LCA modeling. The complexity and uncertainty of environmental data require careful consideration of the datasets used in ML models to ensure that predictions are robust, with low error rates, and therefore applicable to real-world scenarios.

One of the most significant challenges in LCA is the presence of missing data within datasets. This can result from data incompleteness, measurement errors, or data collection limitations. Traditional approaches for handling missing data, such as mean- or median-based inferences, are often limited in their effectiveness and may lead to biased or unreliable results. In contrast, machine learning techniques present a compelling solution to this challenge, offering sophisticated algorithms for data completion and imputation. ML models can infer missing values with remarkable accuracy by leveraging the inherent structure and patterns within LCA datasets. This enhances the efficiency and reliability of LCA analyses. Furthermore, ML systems can capture complex relationships and dependencies within the data, enabling more robust assessments of environmental impacts.

In conclusion, although the integration of ML into LCA is a promising development for predicting environmental impacts, the current research tends to overlook the importance of good datasets. Ensuring the quality and relevance of the data used is crucial for the accuracy and reliability of the predictions made by ML-LCA models. This work wants to reduce this gap by focusing on the creation of high-quality datasets that can support the predictive capabilities of ML algorithms within the LCA framework.

Material and methods

In this section, we present our methodology, which consists of steps leading to the choice of a regression machine learning model for predicting missing data from a given dataset.

As mentioned in the previous sections, there are various regression models, each with advantages and disadvantages that make it more or less suitable for a specific use case.

The following is a list of the main regression algorithms involved in our approach, with a description of their modell.

1. **Linear regression** is one of the most frequently used regression algorithms in Machine Learning. It finds the linear relationship between the input variables and output variables using a best-fit straight line. Basically, the linear regression algorithms assume that there is a linear relationship between the inputs and the outputs. Formula 1 shows a simplification and generalization of the mathematical equation on which the computation of the linear regression is performed.

$$y = \alpha_0 + \alpha_1 x_1 + \alpha_2 x_1 + \ldots\ldots + \alpha_n x_n \qquad (1)$$

In the above formula,
y is the output variable of the model
x_i is the i-th input variable
α_i is the i-th coefficient that is estimated by the model

2. **Ridge regression** is a variant of the linear regression that use penalty terms to minimize the error between the predicted and actual values.

   Referring to the previous equation, the penalty term is the sum of the squares of the coefficients. This regularization is called L2 regularization and it shrinks the coefficients but doesn't set any of them to zero.

   It is particularly useful to mitigate the problem of multicollinearity in linear regression, which commonly occurs in models with large numbers of parameters. Multicollinearity occurs when an input variable is highly correlated with one or more of the other input variables in a regression. Multicollinearity is a problem because it undermines the statistical significance of an input variable.

3. **Lasso regression** is also a variant of linear regression that uses penalty terms to minimize the error between the predicted and actual values. Unlike the previous technique, it uses a regularization called L1 regularization. In this technique, the penalty term is the sum of the absolute value of the coefficients. It can reduce some coefficients to zero, effectively performing input selection. It is particularly useful to reduce the overfitting of the model.

4. **Polynomial regression** extends linear regression by fitting a polynomial equation to the data, allowing for non-linear relationships between input and output variables. Formula 2 shows a simplified and generalized example of a polynomial function of n-th degree.

$$y = \alpha_0 + \alpha_1 x_1^1 + \alpha_2 x_1^2 + \ldots\ldots + \alpha_n x_n^n \qquad (2)$$

5. **Decision tree regression** is a non-linear regression model. The main function of this technique is to split the dataset into smaller sets. The splitting of the data results in a tree-like structure. A decision tree is like a flowchart where each prediction starts from the root node and, based on some criteria, moves along a path of internal nodes called decision nodes until reaching the leaves that contain the result.

6. **Random forest regression** is a collection of multiple decision trees. Each tree is independently trained so that each tree turns out to have a slightly different structure from each other. When we run a prediction using this model, we are essentially asked for the prediction at each individual tree, then their predictions are aggregated to identify the most suitable result.

7. **Support vector regression** is a technique that aims to find a hyperplane that best fits the data points in a continuous space. This is achieved by mapping the input variables to a high-dimensional feature space and finding the hyperplane that maximizes the distance (margin) between the hyperplane and the closest data points, while also minimizing the prediction error.

8. **Gradient boosting regression** is a technique that builds a sequence of weak learners, typically decision trees, with each new learner correcting errors made by the previous ones. In each iteration, the algorithm computes the residual error, which is equal to the actual value minus the predicted value. This error is then used to train a new weak model, with the objective of minimizing it. The process is repeated until a stopping criterion is met.

Our approach

The approach presented below is designed to assist in the selection of an optimal machine learning algorithm for prediction purposes, while also facilitating the incorporation of missing values from a given dataset.

For simplicity and usability, our approach is presented as a decision tree.

As represented in Figure 1 the methodology guides the selection of a model based on key considerations derived from a dataset. The initial inquiry concerns the presence of missing values within the dataset, directing to either proceed with further analysis or conclude that no regression model is required. The subsequent step is to interrogate the linearity of the input-output relationship, in case a linear relationship exists, the Linear Regression approach should be selected. Conversely, an additional evaluation is necessary to determine whether the relationship exhibits non-linearity. If this is the case, then complex interactions of high-order relationships between inputs should be revised. This process can be accomplished by either Polynomial Regression if there are such interactions, or Decision tree Regression if not there are not. In the event that the relationship is found to be non-linear, further investigation should be conducted to ascertain the necessity for effective handling of outliers or noise. If these are present, an evaluation regarding the size and dimension of the dataset is necessary. With a small to medium-sized and high-dimensional dataset, the Support Vector regression model should be used. While with another type of dataset, the Random Forest regression would be the ideal solution. In the event that there is no necessity to address outliers or noise, it would be beneficial to determine whether the input selection is a priority or not. If so, it would be advisable to identify any instances of multicollinearity and select Ridge Regression if the latter is present or Lasso Regression if not. In the case where the input selection is not a primary concern, an assessment of the importance of achieving a high degree of predictive accuracy should be conducted.

Results and discussion

In this Section, we present and discuss the results of our approach applied to a real dataset.

Our dataset consisted of information from approximately 100 dairy cattle farms in Italy, primarily from studies conducted over the past decade. The majority of these farms utilize intensive farming practices. The data collection focused on key inputs for Carbon Footprint analysis, such as production metrics like Fat and Protein Corrected Milk (FPCM) produced, herd composition, cultivated land area, and diet details, including the amount and origin of forage and concentrates, energy and fuel used, as well as the amount of CO2 eq. as output. By using our approach, we were able to increase the number of rows used for training by an order of magnitude, resulting in a significant reduction in prediction error.

Due to the complexity of the study, the dataset was decomposed into subdatasets to make it easier to manage and apply our approach. Specifically, the three subsets of data on soy meal, heard size, and total feed are used in the following to show the accuracy of our approach.

The characteristics of each dataset are listed below:

- **soy meal dataset:** Since the value of soy meal is a component that is influenced by use case components such as herd size, lactating cows, concentrate values, etc., this dataset results in a large number of heterogeneous inputs. It is also important to consider that some of the inputs have noise, due to the inherent limitations of the data collection methods employed.

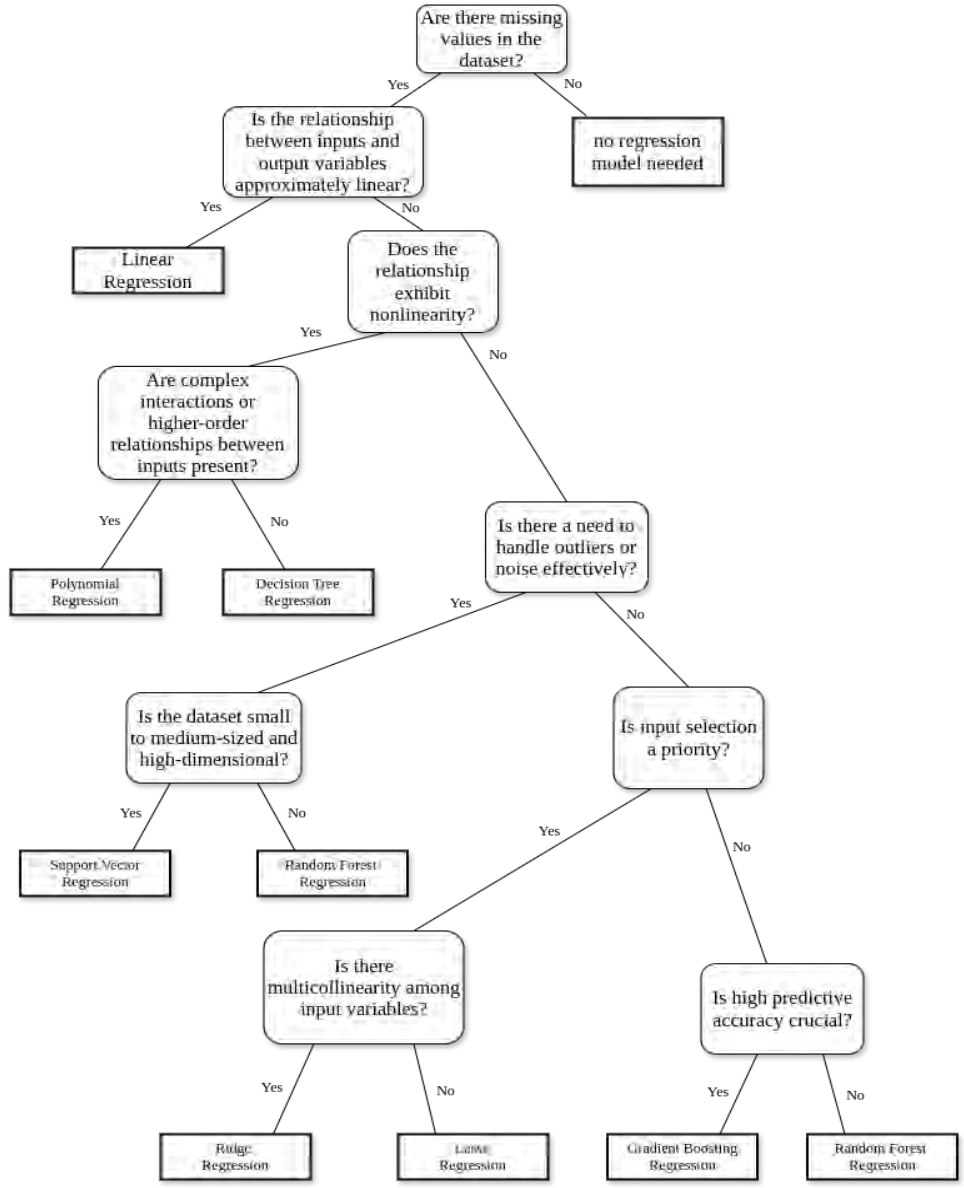

Figure 1. Decision tree representing our approach.

- **heard size dataset:** This value does not depend on an overly complex mathematical calculation and depends very much on the herd farm in question, so we use a dataset with many parameters that can help characterize the type of farm and thus estimate the possible herd size. These parameters are numerous as in the previous case, but they are easier to collect. In this dataset, elements such as the size of the farm, the cultivated area, the type of the farm, etc. are used as inputs.
- **total feed dataset:** This dataset is very similar to the previous one, there is not a complex mathematical relationship between the inputs and the output, it depends

a lot on the way the farm is managed. However, unlike the previous dataset, many inputs are correlated with each other such as soy meal, concentrates, cultivated area, etc.

To evaluate a regression model, various techniques can be used.

In this work, we decided to use the Root Mean Square Error (RMSE). The RMSE represents the standard deviation of the prediction errors. These errors are a measure of the distance of the actual data from the regression line; RMSE is a measure of the spread of these residual values. In other words, it indicates the concentration of actual value around the line of best approximation.

Table 1 shows the RMSE calculated on the three datasets. In order to facilitate a more comprehensive understanding and comparison of performance across the different datasets, the RMSE was normalized between 0 and 1.

In the following, we apply our approach to each dataset and can see that the resulting regression algorithm is equal to the best-performing one as shown in the table.

Application of our approach to the SoyMeal dataset:

1. Are there missing values in the dataset? Yes.
2. Is the relationship between inputs and outputs approximately linear? No, the inputs of the dataset are many and there is no linearity between them and the output.
3. Is there a need to handle outliers or noise effectively? Yes. Since many inputs to the dataset come from measurements that by their nature may contain errors, this dataset definitely has presence of noise within it.
4. In the dataset small to medium-size and high-dimensional? No. In general, a dataset is considered “high-dimensional” when it has a large number of variables relative to the number of observations.
5. Our approach suggests to use a Random Forest Regression model.

Application of our approach to the Herd Size dataset:

1. Are there missing values in the dataset? Yes.
2. Is the relationship between inputs and outputs approximately linear? No, the inputs of the dataset are many and there is no linearity between them and the output.
3. Is there a need to handle outliers or noise effectively? No. Since the inputs come from simpler and therefore more accurate data collection and gathering methods.
4. Is input selection a priority? Yes. It is not easy to choose appropriate inputs to predict this value, so it is preferred to have the regression algorithm do the input selection.
5 Is there multicollinearity among input variables? No. The inputs are different from each other, and although there is a slight correlation among some of them, we can consider the dataset unaffected by multicollinearity
6. Our approach suggests to use a Lasso Regression model.

Application of our approach to the Total Feed dataset:

1. Are there missing values in the dataset? Yes.
2. Is the relationship between inputs and outputs approximately linear? No, the inputs of the dataset are many and there is no linearity between them and the output.
3. Is there a need to handle outliers or noise effectively? No. Since the inputs come from simpler and therefore more accurate data collection and gathering methods.
4. Is input selection a priority? Yes. It is not easy to choose appropriate inputs to predict this value, so it is preferred to have the regression algorithm do the input selection.
5. Is there multicollinearity among input variables? Yes. Many inputs are correlated with each other such as soy meal, concentrates, and cultivated area. We can consider this dataset affected by multicollinearity,
6. Our approach suggests to use a Ridge Regression model.

Table 1. Normalized RMSE of three different datasets.

Dataset/ ML model	Linear	Ridge	Lasso	Polynomial	Decision tree	Random forest	Support vector	Gradient boosting
Soy Meal	0.118	0.116	0.116	0.114	0.114	**0.105**	0.121	0.115
Heard Size	0.023	0.019	**0.020**	0.033	0.088	0.091	0.023	0.081
Total Feed	0.066	**0.058**	0.061	0.086	0.09	0.08	0.059	0.076

Conclusion

In this paper, we introduced a novel approach to address the challenge of predicting missing values within datasets, particularly within the context of environmental impact assessment in the livestock sector. Our approach was designed to select the most appropriate algorithm for predicting missing values in a dataset, taking into account its specific characteristics and requirements. By leveraging Machine Learning (ML) techniques, we sought to develop a systematic framework that could effectively handle missing data and enhance the completeness of the dataset.

To assess the effectiveness of our approach, we conducted testing using three distinct datasets, each representing different variables that influence the environmental impacts of cattle milk farming. Through rigorous testing and validation procedures, we assessed the performance of our approach across these datasets, measuring its ability to accurately predict missing values and improve dataset completeness. The results of the testing demonstrated the versatility and effectiveness of our approach, showcasing its capability to address missing data challenges across diverse datasets. Furthermore, to demonstrate the practical applicability of our approach in real-world scenarios, the three individual datasets were consolidated into a single comprehensive dataset tailored for environmental impact prediction, which enables us to explore and analyze environmental impacts across various domains. The application of the approach

to the unified dataset resulted in a significant increase in the size of the dataset by an order of magnitude, thereby enhancing its utility for environmental impact prediction.

The integration of ML techniques with LCA methodology represents a promising avenue for addressing data-related challenges in environmental impact assessment. The use of ML algorithms to predict missing values and enhance dataset completeness represents a promising avenue for overcoming common limitations encountered in LCA studies, such as incomplete or unreliable data. This synergistic approach enables more robust and comprehensive environmental impact assessments, thereby facilitating informed decision-making and sustainable practices across diverse sectors and industries.

In summary, this study underscores the potential of combining ML techniques with LCA methodologies to address data-related challenges and enhance the completeness and accuracy of environmental datasets.

Acknowledgment

This study was carried out within the Agritech National Research Center and received funding from the European Union Next-GenerationEU (PIANO NAZIONALE DI RIPRESA E RESILIENZA (PNRR) – MISSIONE 4 COMPONENTE 2, INVESTIMENTO 1.4 – D.D. 1032 17/06/2022, CN00000022). This manuscript reflects only the authors' views and opinions, neither the European Union nor the European Commission can be considered responsible for them.

List of references

Caffrey, K.R., and Veal, M.W., 2013. Conducting an Agricultural Life Cycle Assessment: Challenges and Perspectives. The Scientific World Journal. 2013, 13. https://doi.org/10.1155/2013/472431

FAO, 2022. GLEAM 3 Dashboard. In: Shiny Apps. https://foodandagricultureorganization.shinyapps.io/GLEAMV3_Public/

ISO 14040, 2006. Environmental Management. Life Cycle Assessment. Principles and Framework. International Organisation for Standardization. Geneve, Switzerland.

ISO 14044, 2006. Environmental Management. Life Cycle Assessment. Requirements and Guidelines. International Organization for Standardization. Geneve, Switzerland.

Lanzoni, L., Whatford, L., Atzori, A. S., Chincarini, M., Giammarco, M., Fusaro, I., and Vignola, G., 2023. The challenge to integrate animal welfare indicators into the Life Cycle Assessment. Animal. 17(5), 100794. https://doi.org/10.1016/j.animal.2023.100794

Leroy, F., Abraini, F., Beal, T., Dominguez-Salas, P., Gregorini, P., Manzano, P., Rowntree, J., and van Vliet, S., 2022. Animal board invited review: Animal source foods in healthy, sustainable, and ethical diets – An argument against drastic limitation of livestock in the food system. Animal. 16(3). https://doi.org/10.1016/j.animal.2022.100457

Nirmala, T. V., Harikumar, S., Anu, G., and Reddy Devivaraprasad, A., 2023. Green House Gas Emissions and Mitigation Strategies for Sustainable Dairy Farming. Journal of Experimental Agriculture International. 45 (11), 92-103. https://doi.org/10.9734/jeai/2023/v45i112239

Potter, C., Pechey, R., Clark, M., Frie, K., Bateman, P.A., Cook, B., Stewart, C., Piernas, C., Lynch, J., Rayner, M., Poore, J., and Jebb, S. A., 2022. Effects of environmental impact labels on the sustainability of food purchases: Two randomised controlled trials in an experimental online supermarket. PLoS ONE. 17(11), 1–17. https://doi.org/10.1371/journal.pone.0272800

Pradeleix, L., Roux, P., Bouarfa, S., and Bellon-Maurel, V., 2022. Multilevel environmental assessment of regional farming activities with Life Cycle Assessment: Tackling data scarcity and farm diversity with Life Cycle Inventories based on Agrarian System Diagnosis. Agricultural Systems. 196, 103328. https://doi.org/10.1016/j.agsy.2021.103328

Romeiko, X. X., Zhang, X., Pang, Y., Gao, F., Xu, M., Lin, S., and Babbitt, C., 2024. A review of machine learning applications in life cycle assessment studies. Science of The Total Environment, 912, 168969. https://doi.org/10.1016/J.SCITOTENV.2023.168969

Rossi, C., Bernabucci, U., Grossi, G., Cesarini, F., Lacetera, N., Evangelista, C., Turriziani, G., and Vitali, A., 2023. Cradle-to-grave life cycle assessment of buffalo mozzarella cheese supply chain in central Italy. Journal of Agriculture and Food Research, 14, 100871. https://doi.org/10.1016/j.jafr.2023.100871

Rossi, C., Grossi, G., Lacetera, N., and Vitali, A., 2024. Carbon Footprint and Carbon Sink of a Local Italian Dairy Supply Chain. Dairy, 5(1), 201-216. https://doi.org/10.3390/dairy5010017

Stygar, A. H., Krampe, C., Llonch, P., and Niemi, J. K., 2022. How Far Are We From Data-Driven and Animal-Based Welfare Assessment? A Critical Analysis of European Quality Schemes. Frontiers in Animal Science. 3, 1–12. https://doi.org/10.3389/fanim.2022.874260

Thumba, D. A., Lazarova-Molnar, S., and Niloofar, P., 2021. Estimating Livestock Greenhouse Gas Emissions: Existing Models, Emerging Technologies and Associated Challenges. 6th International Conference on Smart and Sustainable Technologies (SpliTech). Bol and Split, Croatia. http://doi.org/10.23919/SpliTech52315.2021.9566450