Applications of Design of Experiments and machine learning on product innovation: a literature review

Rosa Arboretti Giancristofaro[1], Riccardo Ceccato[2], **Luca Pegoraro**[2] and Luigi Salmaso[2]

[1] Department of Civil, Environmental and Architectural Engineering [2] Department of Management and Engineering, University of Padova, Padua, Italy.

INTRODUCTION

In recent years there has been a spike in the application of machine learning (ML) models in industrial contexts [1]. A practice that is gaining traction consists in the application of ML algorithms for the analysis of data collected through experimental designs (DOE) [2]. This is particularly relevant in contexts affected by data scarcity, such as the ones of product improvement and innovation (PI) [2,3].

A Systematic Literature Review (SLR) has been conducted to delineate the state of the art and identify the main trends in this topic [4].

METHODOLOGY

Literature research questions:

LRQ1: *Which are the advantages and challenges of using ML methods with respect to traditional parametric statistics approaches?*

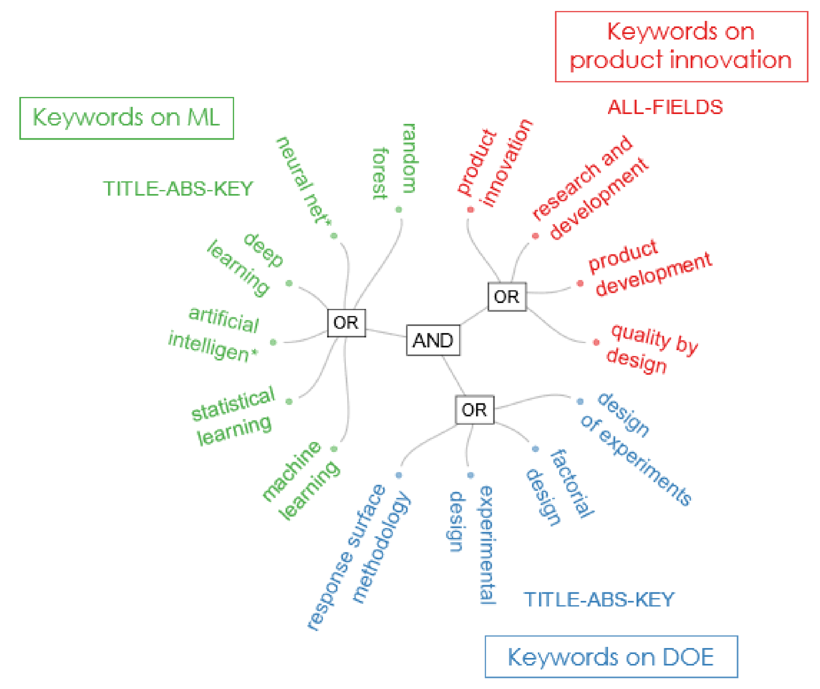
LRQ2: *Considering a DOE+ML framework, which are the DOE strategies and DOE types adopted?*

LRQ3: *Considering a DOE+ML framework, which are the ML algorithms adopted?*

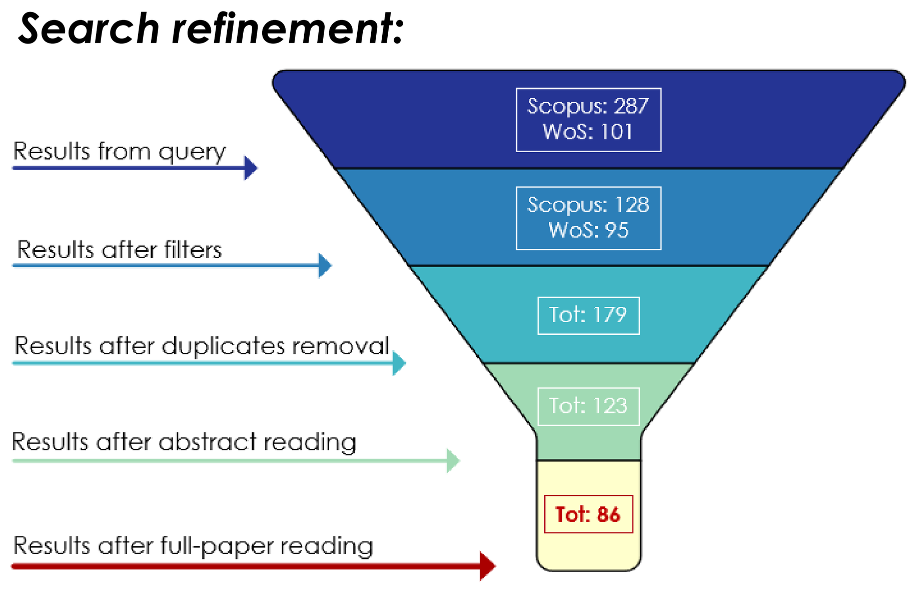
LRQ4: *Which are the implications of adopting a DOE+ML framework?*

LRQ5: *Which are the most important literature gaps and research opportunities?*

Query formulation:

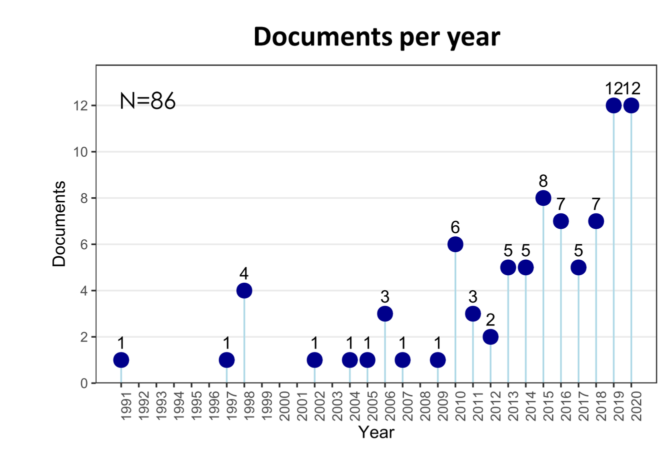


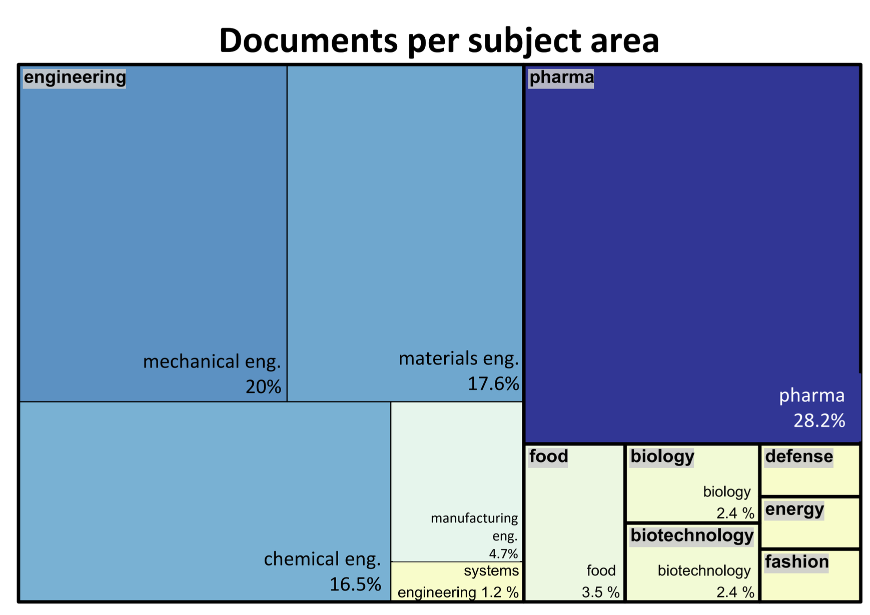
Search refinement:



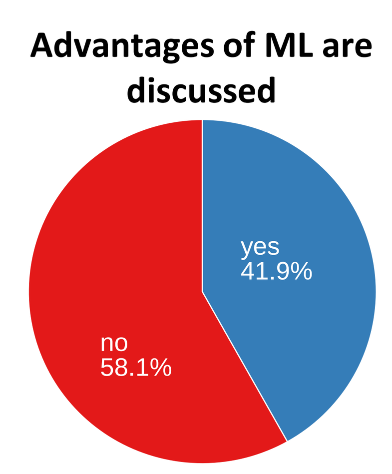
RESULTS AND DISCUSSION

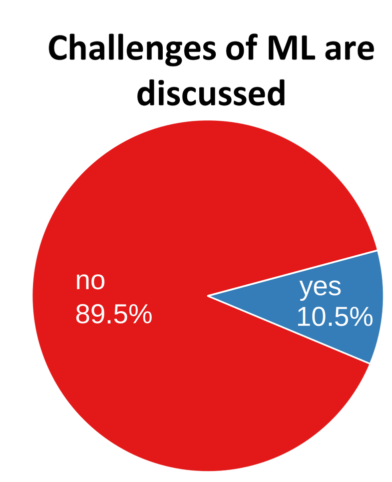
Descriptive results:



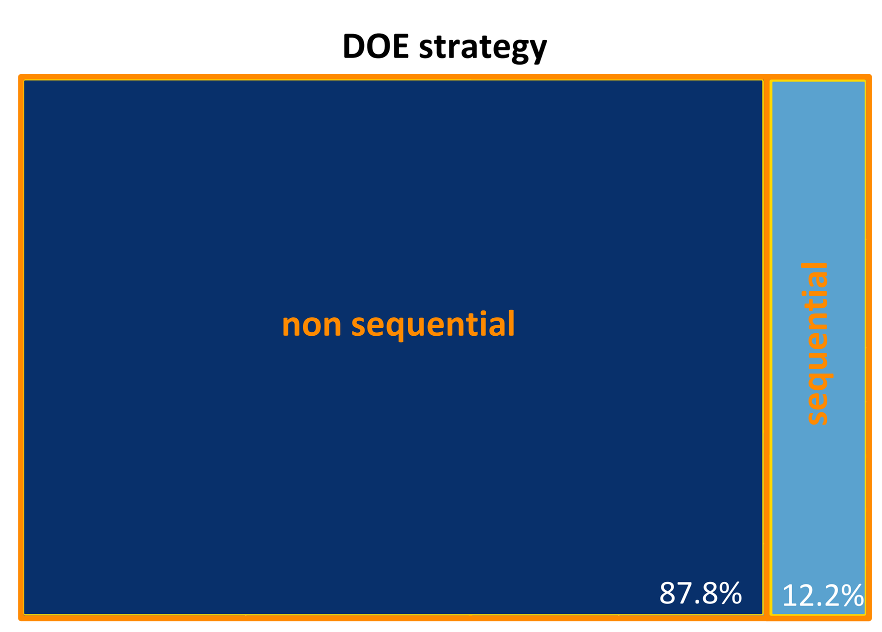


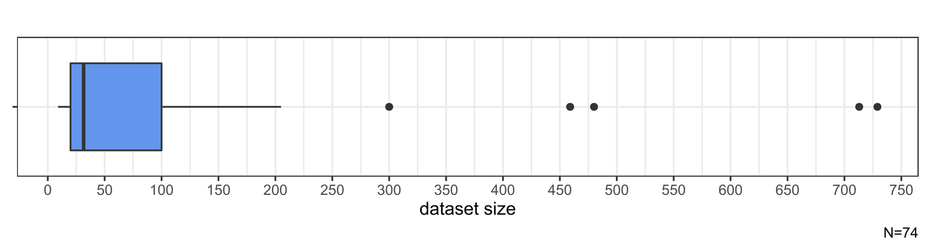
Content analysis results:
LRQ1:

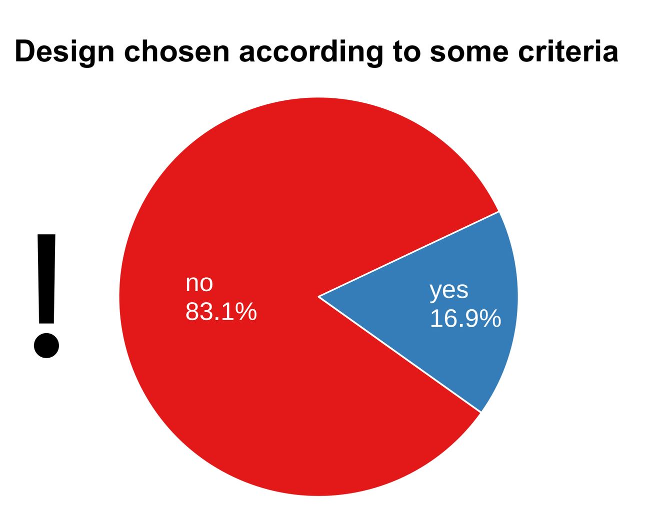




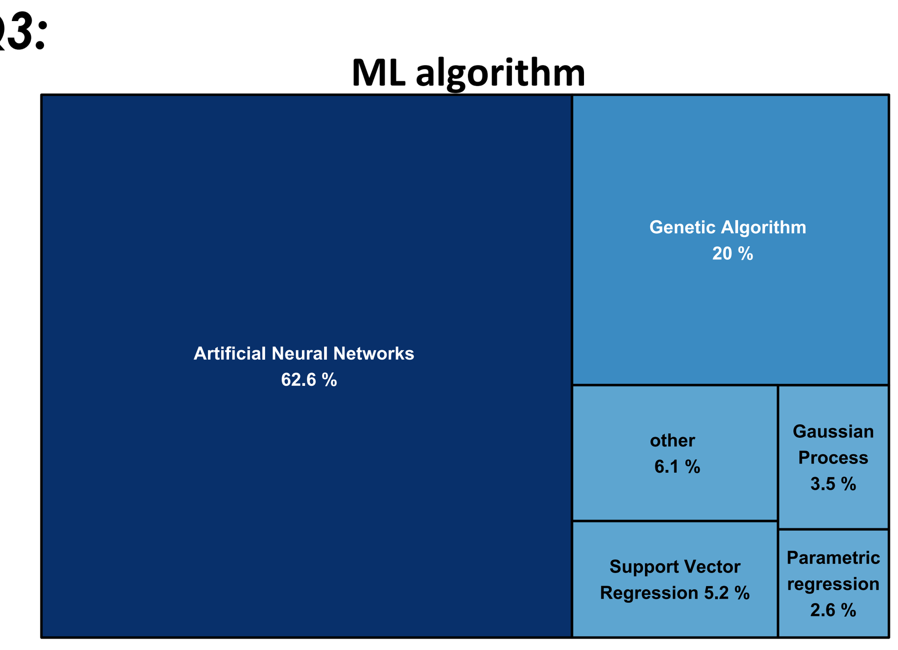
LRQ2:

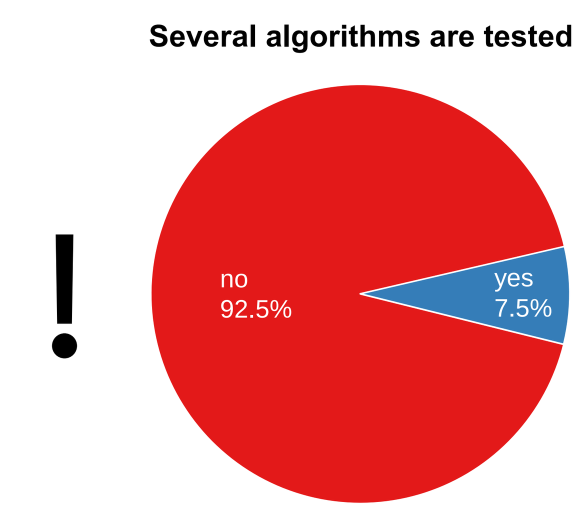






LRQ3:





LRQ4:

Description	%	N
In DOE+ML, a minimization of the number of experiments is achieved	35.5	11
In DOE+ML, ML can suggest optimal configurations for DOE trials	29	9
DOE+ML can lead to the exploration of experimental regions omitted by DOE alone	25.8	8
DOE+ML provides better final product quality	25.8	8
In DOE+ML, DOE analysis can be used to explain the relationships which drive the ML algorithms and select variables	25.8	8
DOE+ML can lead to full automation in experimentation	22.6	7
DOE+ML provides a systematic, non-subjective method for PI	16.1	5
In DOE+ML, DOE can optimize ML hyperparameters	6.5	2
In DOE+ML, DOE provides reasonable datasets in small data settings	6.5	2
In DOE+ML, DOE controlled experiments provide a support for causal claims	3.2	1

Table 1: Implications of the joint adoption of DOE+ML in PI

CONCLUSIONS AND FUTURE WORK

LRQ5: Literature Gaps and Open Questions

RQ1: *What are the most appropriate DOE strategies (seq. vs non-seq.) and DOE types for a DOE + ML framework?*

RQ2: *What are the most appropriate ML algorithms in a DOE + ML framework? How to choose them?*

It is the authors' intention to answer to the RQs through application on real data and simulation studies.

REFERENCES

[1] Wuest, Thorsten, et al. "Machine learning in manufacturing: advantages, challenges, and applications." Production & Manufacturing Research 4.1 (2016): 23-45.

[2] R. Arboretti, R. Ceccato, L. Pegoraro, L. Salmaso, C. Housmekerides, L. Spadoni, E. Pierangelo, S. Quaggia, C. Tveit, S. Vianello, Machine learning and design of experiments with an application to product innovation in the chemical industry, Journal of Applied Statistics (2021).

[3] Y. Tian, R. Yuan, D. Xue, Y. Zhou, X. Ding, J. Sun, T. Lookman, Role of uncertainty estimation in accelerating materials development via active learning, Journal of Applied Physics 128 (2020) .

[4] R. Arboretti, R. Ceccato, L. Pegoraro, L. Salmaso, Design of Experiments and machine learning for product innovation: a systematic literature review, Quality and Reliability Engineering International (submitted).