The importance of interpretability and visualization in ML for medical applications

Alfredo Vellido
IDEAI-UPC Research Center, UPC BarcelonaTech, 08034 Barcelona, Spain
CIBER-BBN, Cerdanyola del Vallès, Barcelona, Spain
avellido@cs.upc.edu

***Abstract*—Many areas of science have made a sharp transition towards data-dependent methods, enabled by simultaneous advances in data acquisition and the development of networked system technologies. This is particularly clear in the life sciences, which can be seen as a perfect scenario for the use of machine learning to address problems in which more traditional data analysis approaches might struggle. But this scenario also poses some serious challenges. One of them is the lack interpretability and explainability for complex nonlinear models. In medicine and health care, not addressing such challenge might seriously limit the chances of adoption of these methods. In this summary paper, we pay specific attention to one of the ways in which interpretability and explainability can be addressed in this context: data and model visualization.**

***Index Terms*—Interpretability and explainability, Data Visualization, Machine Learning, medicine, health care, medical decision support systems.**

I. Introduction

The overabundance of data in the modern life sciences could be seen as a perfect scenario for the use of machine learning (ML), but comes accompanied by some far from trivial challenges [1]. One of them is model interpretability and explainability. In medicine and health care, where explainability is paramount and the societal impact is potentially high [2], such challenge might seriously limit the chances of adoption, in real practice, of computer-based systems that rely on opaque ML methods for data analysis. In this summary paper, we pay specific attention to one of the ways in which the challenge can be addressed: through techniques for data visualization. By doing so, we aim to stress the importance of considering the human factor when attempting to enhance model interpretability in general and the importance of integrating the medical expert in the process of developing strategies to guarantee the interpretability and explainability of medical data models.

II. Interpretable ML in medicine: a key to adoption

Data-dependence is bound to increase in medical practice, given the prominent place occupied by evidence-based medicine in the current agenda. The simultaneous creation of an information-rich medical environment and the development of techniques for knowledge extraction tailored to this domain, would seem to be a win-win situation for ML. But lack of model interpretability is a problem with obvious implications: if an ML-based Medical Decision Support System (MDSS) churns out decisions that cannot be described in comprehensible terms, an insurmountable barrier is raised between the MDSS and the human subjects. The medical expert could not trust to implement a decision that she or he cannot explain, whereas the patient might not trust experts that base their judgement on unexplainable computer outcomes. This means that formal frameworks for machine-human interaction pursuing interpretability and explainability are even more important in medicine than in other ambits of science, specially because there is a constellation of stakeholders in the health domain with possibly quite different explanation needs [3]. These frameworks should almost be considered as a pre-requisite in the development of ML-based MDSS.

III. Visualization as a problem in medicine

Visualization has been mentioned to play a central role as an interpretability tool for medicine and it is important to provide a formal framework for its use in this area. The human analyst has an active role in the interactive visualization framework proposed by Sacha and co-workers in [4], acting as a bridge between visual pattern discovery (mostly using ML tools) and knowledge validation by external experts. The importance of appraising the possible benefits of putting the "human-in-the-loop" is persuasively argued precisely as a validating actor in practical applications.

In real-world medicine, visual discovery is not always purely exploratory and, therefore, potentially interesting patterns obtained through visualization must be validated against expert knowledge from the domain. Often, this external assessment requires a committee of experts who, in turn, will provide feedback to the analyst that can help to redesign visualization experiments. This adds an extra layer of human subjectivity to the interpretation task through visualization. As a result, the framework must care not only about a cycle involving computer-based visual techniques and a human analyst, but also about a coupled cycle involving two human parts: the data analyst and the experts from the medical domain who provide the ultimate expert verification.

A detailed representation of this interpretability-through-visualization cycle can be found in Fig. 1. It involves requests from the medical experts to the data analyst, including: a) guarantees of interpretability and explainability that are adapted to the specific requirements of the medical problem; b) model compliance with clinical protocols and guidelines for

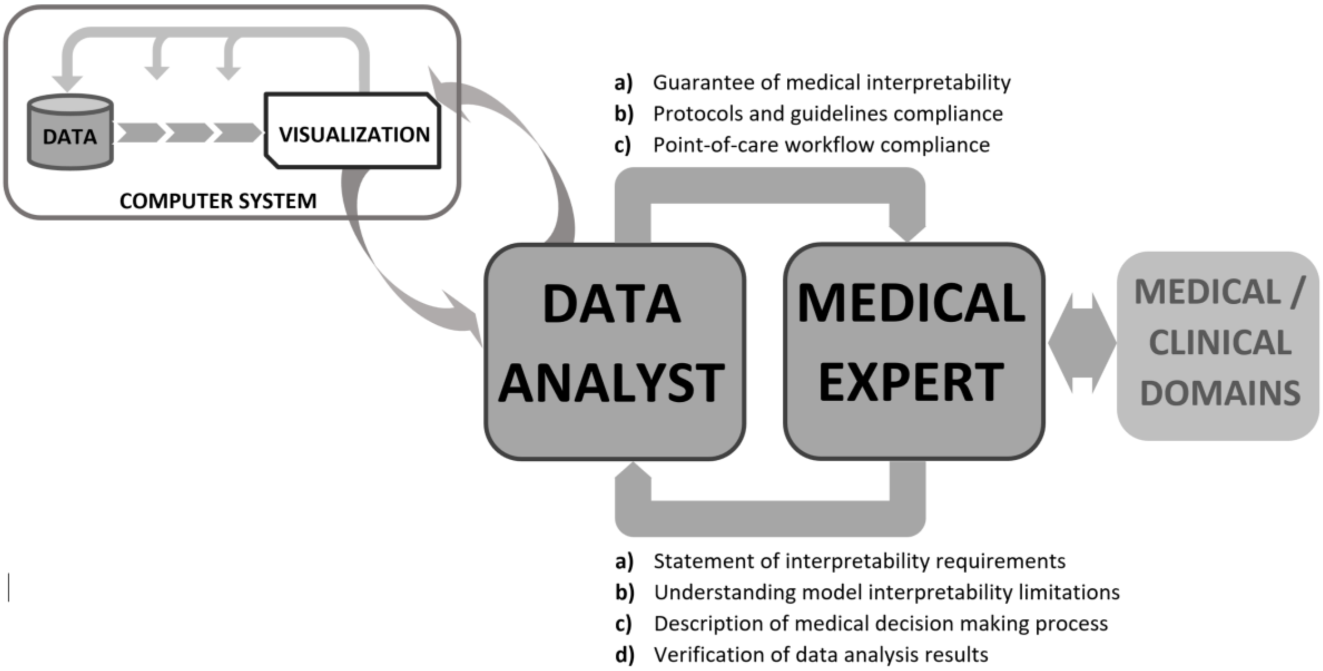
DATA
VISUALIZATION
COMPUTER SYSTEM
a) Guarantee of medical interpretability
b) Protocols and guidelines compliance
c) Point-of-care workflow compliance
DATA ANALYST
MEDICAL EXPERT
MEDICAL / CLINICAL DOMAINS
a) Statement of interpretability requirements
b) Understanding model interpretability limitations
c) Description of medical decision making process
d) Verification of data analysis results

Fig. 1. Extension of the human analyst-computer ML interpretability cycle through interactive visualization proposed in [4] to account for a new sub-cycle of importance to the medical and health care domains. This new sub-cycle covers the necessary interaction between the human analyst, who must deliver data models that are interpretable and/or explainable from a medical viewpoint, and the medical expert, who must ensure that the data analyst is informed of the requirements that make interpretability valid from a medical standpoint. Arrows in the graphical depiction of the interaction between these two agents point from the agent that can deliver the interpretability item to the agent that requires it.

a given problem; c) model compliance with system-human interaction workflows at the point of care. It also involves requests from the data analyst to the medical experts, such as: a) a clear statement of the medical requirements concerning interpretability and explainability; b) a realistic understanding of the interpretability limitations and possibilities of the analytical models; c) a clear description of the real medical decision making process in place at the point of care; and d) a guarantee of verification of the data analysis results.

IV. CONCLUSIONS

The life sciences are at the avant-garde of an irreversible trend that is placing data at the heart of scientific discovery. Medicine and health care, at their own pace, are following suit. This is an unprecedented opportunity for ML, CI and related techniques for knowledge extraction from data. In this paper summary, we have argued that there are still many barriers to overcome before these techniques become mainstream. One of them is model interpretability and explainability, which must be guaranteed before ML-based MDSS are trusted by medical practitioners. Model interpretability has become a central issue for ML in recent times, due the success of deep learnng models, paradigmatic examples of lack of interpretability. We have tried to convey the message that medical data analysts must widen their scope to ensure the interpretability of the complete analytical process by involving medical experts in it, with special attention paid to interpretability achieved by interactive visualization. We should ensure that the interaction between the data analysts and the medical experts adheres to a formal protocol in which the specific requirements of each of these parties are clearly and unambiguously laid out. This form of interactive ML makes methodically correct analytical processes more difficult to implement, evaluate and replicate. These difficulties, though, should be offset by the advantages of adhering to such protocol, which would maximize the chances of ML-based MDSS being integrated in the routine of clinical practice.

ACKNOWLEDGMENT

This is an abbreviated version of [5] Vellido, A. (2020). doi.org/10.1007/s00521-019-04051-w. The study is funded by Spanish research project PID2019-104551RB-I00.

REFERENCES

[1] Cabitza, F., Rasoini, R. and Gensini, G.F. “Unintended consequences of machine learning in medicine.” JAMA-J. Am. Med. Assoc. Vol. 318(6), pp. 517–518 (2017)

[2] Vellido, A., “Societal issues concerning the application of artificial intelligence in medicine.” Kidney Diseases Vol. 5, pp. 23–27 (2019)

[3] Gerlings, J., Jensen, M.S. and Shollo, A., “Explainable AI, but explainable to whom?.” ArXiv preprint arXiv:2106.05568 (2021)

[4] Sacha, D., *et al*. “What you see is what you can change: Human-centered machine learning by interactive visualization.” Neurocomputing Vol. 268, pp. 164–175 (2017)

[5] Vellido, A., “The importance of interpretability and visualization in machine learning for applications in medicine and health care.” Neural. Comput. & Applic. Vol. 32, pp. 18069–18083 (2020)