ACTA MÉDICA PORTUGUESA
A Revista Científica da Ordem dos Médicos

Application of Artificial Intelligence in Healthcare: The Need for More Interpretable Artificial Intelligence

Aplicação de Inteligência Artificial em Cuidados de Saúde: A Necessidade de Mais Inteligência Artificial que Seja Interpretável

Jorge TAVARES✉[1]
Acta Med Port 2024 Jun;37(6):411-414 ▪ https://doi.org/10.20344/amp.20469

Keywords: Artificial Intelligence; Delivery of Health Care; Machine Learning
Palavras-chave: Aprendizagem Automática; Inteligência Artificial; Prestação de Cuidados de Saúde

INTRODUCTION

Understanding artificial intelligence (AI) and its different types is of the utmost importance for the application of this technology in healthcare.[1,2] Artificial intelligence is a field of knowledge which combines computer science and advanced statistics to support problem-solving.[3] It is divided in two sub-fields: machine learning (ML) and deep learning.[1] The ML concept resides in the ability of using computer algorithms that have the capability to recognize patterns and efficiently learn to train the model to predict, make recommendations or find data patterns.[1,3] After a sufficient number of repetitions and algorithm adjustments, the machine becomes capable to accurately predict an output.[1,3] Deep learning is a newer and more complex approach of AI that uses deep neural networks. The neural network starts with an input layer that then progresses to a variable number of hidden layers.[1] Since the algorithm uses multiple layers with deep neural networks, it can successively refine itself, without explicitly programmed directions.[1] It is a fact that, by using deep learning, the models usually achieve higher accuracy compared with ML. Still, when using ML, it is frequently possible to better understand which are the input variables that have more influence on the output variables.[4]

In both medical and clinical practices, it is often particularly relevant to understand why an AI technique is suggesting a certain classification or direction for a certain action.[1] Not only in healthcare but also in other fields of knowledge, explainable AI (also called XAI) is growing its influence.[4] The current European legal regulation, specifically the General Data Protection Regulation (GDPR), requires that automated models provide meaningful information about the rationale on how the algorithm operates.[4]

The goal of this article is not to provide an exhaustive view about all existing AI models and explainable AI, but instead to provide a summarized and easy to understand view of what should be considered when implementing AI in healthcare and in clinical practice.

Definition of explainable artificial intelligence

Most likely, the best way to start describing the goal of explainable AI is to use an example from the literature. The case that is described here is about the classification of patients with pneumonia.[5] When a patient was first diagnosed with pneumonia, the hospital (located in the USA) needed to make one critical decision early on: whether to treat the patient as an inpatient or an outpatient.[5] An AI/ML group of experts was tasked with building models to predict patient survival rates and identify which patients were at greatest risk, which could help the hospital triage new patients.[5] The result was a head-to-head of traditional ML models (logistic regression, rule-learning model, decision tree) and a neural network.[5] Among all the models tested, the neural network achieved the best accuracy at identifying and classifying the patients with the lowest survival rates.[5] The most obvious decision would be to use the neural network, but in the end it was not. Another researcher had been training a rule-based model on the same dataset. Rule based models are among the most easily interpreted ML models. They typically take the form of a list of 'if x, then y' rules, that are easier to be interpreted by humans.[5] During the verification of the rules a strange rule was identified. The rule read that if a patient had a history of asthma, then they had a lower risk of death and should be treated as an outpatient.[5]

Based on this strange and contradictory rule, the researchers decided to approach the physicians. The physicians said that the fact that the asthma patients had better survival rates was most likely because they immediately received high standards of care, and not only stayed immediately in the hospital but were also transferred to the intensive care unit.[5] Another issue was that the neural network model was also classifying the asthma patients as outpatients.[5] A major classification issue with serious consequences was therefore avoided because it was possible to comprehend the rule-based model. The same ability to

1. NOVA Information Management School (NOVA IMS). Universidade NOVA de Lisboa. Lisbon. Portugal.
✉ **Autor correspondente:** Jorge Tavares. d2012072@novaims.unl.pt
Recebido/*Received*: 30/07/2023 - **Aceite/*Accepted*:** 27/12/2023 - **Publicado *Online*/*Published Online*:** 05/04/2024 - **Publicado/*Published*:** 03/06/2024

comprehend how the neural network was classifying the patients was not available.[5] Explainable AI/ML should be accurate and robust and the models need to be transparent and comprehensible.[4] This means that it has to be possible to explain how the algorithm works, starting from the inputs, how the data is processed and what is the rationale on how the outputs are generated.[4]

Types of artificial intelligence and machine learning models concerning explainability.

Broadly, AI/ML models can be defined as being transparent or opaque/black box models.[4] For a model to be considered transparent it should follow into one or more of the three categories. The first category is simulatability, and it refers to the ability to be simulated by a human.[4] A good example of this type of models is the rule-based model explained in the previous section.[4] The second category is decomposability, and it denotes the ability to break down a model into parts.[4] Decision trees fall into this category.[3,4] The last category is the algorithmic transparency, and it is the ability to understand the way the model generates its output.[4] It is often only possible to inspect it through a mathematical analysis, which is still sufficient to validate it as transparent.[4] Some examples of the models that fall into this category are linear/logistic regression, and k-means clustering.[3,4] Opaque models lack these categories of transparency, and newer techniques are now being studied to provide explainability to them, which are still in their early stages of development.[4,5] Deep learning models are the most well-known example of this type of models.[3,4] Figure 1 shows the different types of models and their graphical outputs.

How to decide which models to implement

Healthcare deals with very specific and sensitive types of data and approaches topics of high complexity. An AI/ML model should not be implemented without the participation of a group of experts composed of healthcare

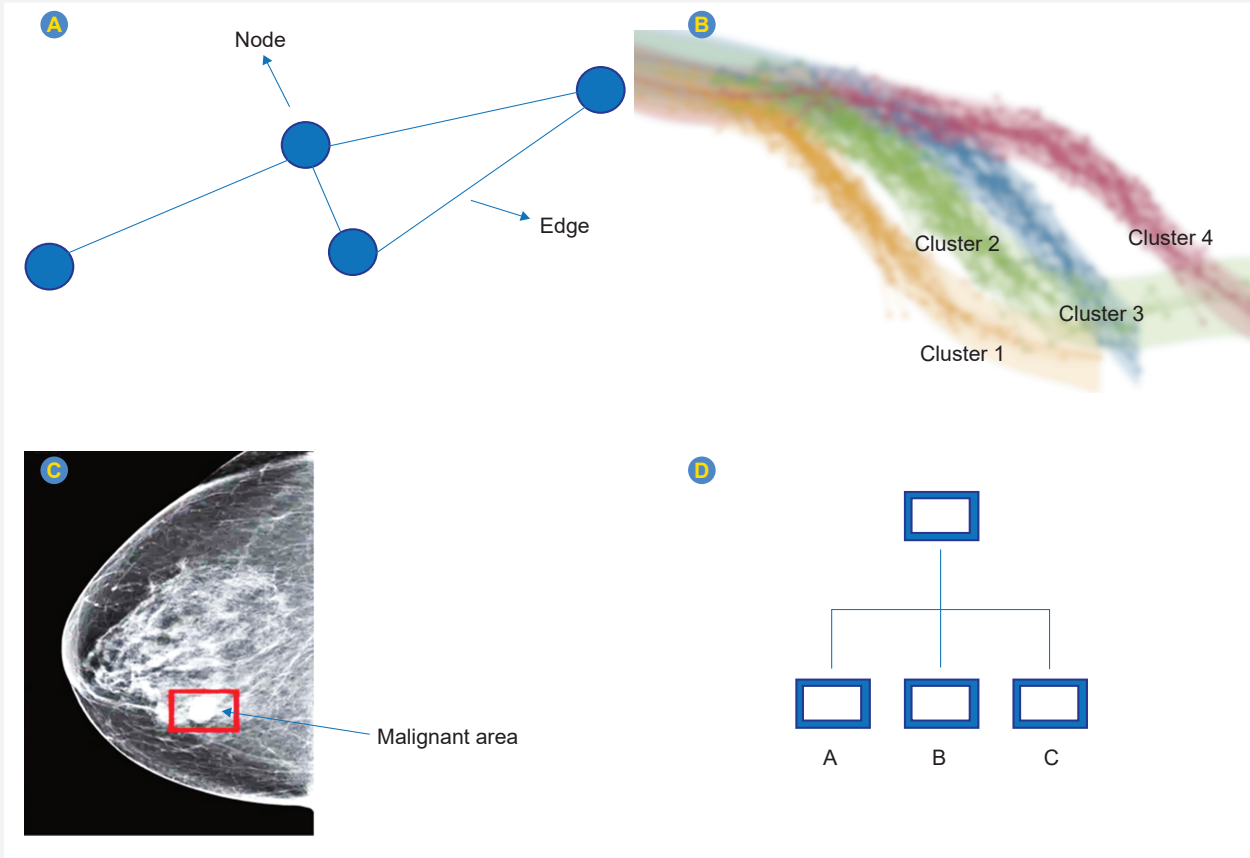

Figure 1 – AI/ML models with graphical capabilities that allow interpretation of results. ML graph, the model learns to make predictions based on the graph's structure and the attributes of nodes and edges (A). Mixture of Gaussian processes, that in this case aggregates patient disease progression trajectories into clusters using a non-parametric approach (B). AI computer vision models supported by attention plots that identify the relevant areas for disease diagnosis (C). Simple decision tree that following certain rules, can visually help to interpret the relevant parameters for class classification (D).

professionals, biostatisticians, data scientists, regulators and/or members of an ethics committee.[6-8] The European Union (EU) is working on specific legislation for AI: the EU AI Act.[4] It emphasizes that AI systems used in the EU should be transparent, safe, protect confidentiality, traceable, non-discriminatory and should be overseen by humans, rather than by automation, to prevent harmful outcomes.[4] Recent studies showed cases of risk of biased AI connected to specific ethnic groups, particularly the ones with lower socioeconomic strata.[8] This should be managed by including diverse groups in clinical studies and controlling the model outputs via interpretability.[8] The new EU AI legislation emphasizes the need of having explainable AI/ML models and using them as a first choice.[7] Still, it is important to understand how to choose which type of models to implement. In applications where explainability is relevant, it is of the utmost importance to use a transparent model (e.g.: treatment decision algorithm, algorithm to decide patient inclusion in a clinical trial).[2,4] But it is not always possible to solve problems using the simpler and/or more transparent approaches.[5,7,8] A second layer of options is to use models that may be more complex but still retain some ability to be understood. These models are called semi-opaque models, because they can provide feature importance (which variables are more important in our model to explain the problem we want to understand) and/or allow the extraction of rules or decision paths that explain how the model arrived at a particular prediction.[4,5,8] Examples of these valuable approaches are random forests, ML graphs with visual interpretation, gradient-boosted trees and Mixture of Gaussian processes in combination with a clustering approach.[4,5,8,9] A recently published article using the Mixture of Gaussian process with a clustering approach, which is a method sustained by a robust statistical approach, showed potential superiority to analyze disease progression in patients with amyotrophic lateral sclerosis compared with more traditional parametrical approaches like Kaplan-Meier curves.[9] In cases where high accuracy is required and other alternatives do not show good results, the use of an opaque model may be justified (e.g.: tumor detection in MRI using deep learning networks).[4] The European Medicines Agency (EMA) published on July 10, 2023, a reflection paper on the use of AI/ML in drug development.[7] Fig. 2 provides an overview on how the EMA approaches the application of AI in healthcare and clinical practice.

CONCLUSION

The increase in the usage of AI in healthcare poses questions about how these algorithms work and how transparent they are. Therefore, it is of the utmost importance to develop explainable AI/ML in healthcare. The EU is developing new AI legislation that emphasizes the need of having transparent models. The decision to implement a certain type of AI/ML model should take into consideration not only the need of being able to explain what the model does, but it should also consider specific legislation and ethical concerns. Explainable AI/ML frequently relies on statistical models, and there is an opportunity to bridge it

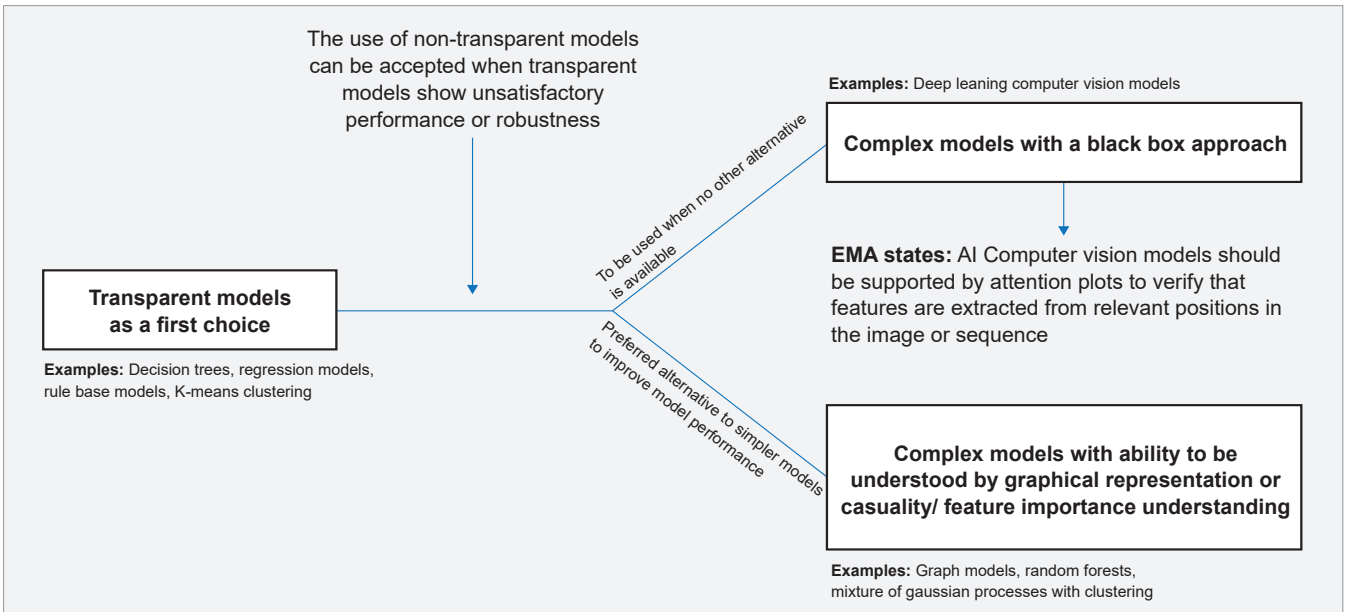

Figure 2 – Decision flow about selecting AI/ML models following the latest guidance from EMA that favors interpretable models as the first approach. SHAP and/or LIME analyses are also recommended by EMA.

with Biostatistics in order to increase the knowledge that we can obtain from research studies. The new EMA reflection paper requires that AI should be implemented considering the principles of Biostatistics guidelines.[7] Due to the high complexity of AI/ML in healthcare, multidisciplinary teams should include healthcare professionals during the development stage of the AI model algorithm, to ensure that the model meets the clinical and ethical requirements.

COMPETING INTERESTS

The author has declared that no competing interests exist.

FUNDING SOURCES

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

REFERENCES

1. Matthew Helm J, Swiergosz MA, Haeberle HM, Karnuta JL, Schaffer JE, Krebs V, et al. Machine learning and artificial intelligence: definitions, applications, and future directions. Curr Rev Musculoskelet Med. 2020;13:69-76.
2. Rasheed K, Qayyum A, Ghaly, M, Al-Fuqaha A, Razi, A, Qadir J. Explainable, trustworthy, and ethical machine learning for healthcare: a survey. Comput Biol Med. 2022;149:106043.
3. Müller AC, Guido S. Introduction to machine learning with Python. 4th ed. Paris: O'Reilly Editions; 2018.
4. Belle V, Papantonis I. Principles and practice of explainable machine learning. Front Big Data. 2021;4:688969.
5. Caruana R, Lou Y, Gehrke J, Koch P, Sturm M, Elhadad N. Intelligible models for healthcare: predicting pneumonia risk and hospital 30-day readmission. Proceedings of the 21th ACM SIGKDD Sydney: International Conference on Knowledge Discovery and Data Mining; 2015.
6. European Comission: EUR-Lex 2021. Proposal for a regulation of the European Parliament and of the Council laying down harmonized rules on artificial intelligence (Artificial Intelligence Act) and amending certain Union Legislative Acts. [cited 2023 Jul 10]. Available from: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex%3A52021PC0206.
7. European Medicines Agency. Reflection paper on the use of artificial intelligence in the lifecycle of medicines. [cited 2023 Jul 22]. Available from: https://www.ema.europa.eu/en/news/reflection-paper-use-artificial-intelligence-lifecycle-medicine.
8. Hunter DJ, Holmes C. Where medical statistics meets artificial intelligence. N Engl J Med. 2023;389:1211-9.
9. Ramamoorthy D, Severson K, Ghosh S, Sachs K, Als A, Glass JD, et al. Identifying patterns in amyotrophic lateral sclerosis progression from sparse longitudinal data. Nat Comput Sci. 2022;2:605-16.