"Review of Causability and Explainability of Artificial Intelligence in Medicine"

mkaramat.bee17seecs

April 2021

This paper presents the growth of AI in a story-like fashion. The main point emphasized in this paper is that as AI algorithms continue to be developed and cutting-edge research is being conducted all over the world to improve AI, there is a need to introduce the concept of causability in AI algorithms. The paper uses the terms "white-box" and "black-box" to describe explainability of AI algorithms and stress upon the fact that to enhance the explainability of an AI model for medical purposes, we must introduce causability in them. The paper raises the point that current state-of-the-art Deep Learning (DL) algorithms are based on statistical models and lack the notion of "context". To design efficient Deep Learning algorithms for medical purposes and to reach a high level of explainability, we must incorporate context in the development of our models. Moving on, the paper differentiates between the terms; "explanation" and "interpretation". Furthermore, the two types of explainable AI; post-hoc systems and ante-hoc systems are described along with examples. The paper then discusses the various methods to gauge an AI model for uncertainty, attribution, and prototypes. Following up after explaining and describing various terms used in this area of research along with a historical perspective, the paper then presents two examples of analysis done by a professional on liver pathology, one each for post-hoc and ante-hoc explainable AI. Lastly, the paper gives three recommendations; introduce weakly supervised learning in medical AI since completely supervised learning is too laborious of a task, develop structural causal models that are better in terms of explainability as compared to the current statistical models, and establish causability as a new scientific field.

This paper is an excellent story of the development of AI models from their birth in the 1950s to the current era where AI is ubiquitous. The paper focuses on the tradeoff between the performance and explainability of AI models. However, this point could be better elaborated with some graphical representation. The paper, overall, lacks the use of statistics and graphical explanations.To present their case effectively to the reader, the authors have provided an extensive literature review in the paper. The literature review excellently presents the case of the authors from the start of the AI era to the present need for causability in AI models. The definitions of terms are exemplified to provide a

better understanding to the reader.

To provide concrete basis for their arguments, the paper brings in the perspective of a professional pathologist to elaborate on the differences between the two types of explainable AI discussed in this paper: post-hoc and ante-hoc. The labelled diagram of a liver tissue augments the explanation provided for the two types of explainable AI.

The paper is presented in a clear and understandable format. The arguments are well-placed and connected to the previous ones. However, in the section "Causability as a new scientific field", there seems to be a typo. The question presented in this section is "what does a feature in a histology slide the pathologist about a disease?" Instead, it should be "what does a deature in a histology slide tell the pathologist about a disease?"

This paper provides a solid basis for development in explainable AI and to incorporate causability in AI models. They propose using structural causal models instead of the statistical models used currently. In addition to this, they signify the need to make causability a full-fledged research area. These propositions could be an important contribution in making improvements in explainable AI.

The paper talks about the need for development of causability in AI models. However, such models require great effort and require the contribution of medical professionals as presented in the examples for post-hoc and ante-hoc explainable AI. This may turn out to be a tedious task.

Overall, this paper presents viable solutions to make causability in AI models a reality. From a practical point of view, the arguments are sound and the propositions presented in the conclusion of this paper hold great potential to make this a reality in the future.