Critical Review – 2

by

Fizza Tauqeer

A research study titled "Causability and Explainability of Artificial Intelligence in Medicine" headed by Holzinger et al. employs a discriminatory review and analysis of the emphasis and need of causable and explainable artificial aptitude, even more so in the medical framework. By causability, the study refers to the ability to initiate understanding of a certain cause or concept to a peer in a definite context to an acceptably effective level, whereas by explainability, it refers to the technical attributes and their values in terms of representing an appropriate algorithmic solution of a problem predicting a certain feature to an acceptably effective level. This differentiation enables the study to evaluate and decide which ability is more liable to become a necessity in the medical field where machine learning deployment and insight is fast picking momentum, especially since explainable AI has become a growing area of interest for many technical research enthusiasts. Relevant literature is the prime documentation and evaluation metric, accompanied by an all-encompassing dive into identifying why such abilities always meet at a trade-off point where deep learning methodologies are concerned. Deep learning methods such as Recurrent Neural Networks (RNNs), due to their 'black-box' nature whilst being post-hoc systems, generally fail at providing a good measure of causability while excelling at explainability. On the other hand, ante-hoc systems such as Decision Trees are generally good in causability and poor in explainability due to existing in a 'glass-box' set-up. Based on these former criterions, this paper implies a 'fuzzy' partiality to Causability in the medicine region instead.

The study makes profound arguments over why explainable AI need not follow as strict a composition or attention, especially in the medicinal field, as traditional descriptive methods – rather they reason that the focus should be more dedicated and directed towards the cause-effect strata of any foreseeable solution's breakdown. It reasons that this is primarily because realistic datasets pertaining to the medical field are hardly ever in formats that are reliable for explanatory content, such as why certain attributes and their values could end up providing a certain result. More often than not, in a field dictated by emergencies and spontaneous situations, maintaining exhaustive knowledge bases is not the prime priority of medical or associated personnel. Moreover, treatment of any ailment is considered highly individual – explanations would not guarantee that a generic query rule, for example, would be applicable or helpful to all potential patients. Knowing the reasoning behind the ailment ultimately also carries more value, as it may help likely individuals in avoiding certain habits that could, otherwise, propel them to being diagnosed in its place. These reasons solidify why causality is far more significant in the medical realm than 'peer-to-peer' explainability, which is the main focus of this research. Additionally, treatment happens to vary in time – some diseases may require years while others may require weeks to be completely treated or cured. The study also acknowledges how visual analysis – which is not exactly considered a viable subset of explainability – can be just as effective in providing prized perceptions, that too with lesser time complexity and no knowledge graph building.

Additionally, the research study relies on various literatures to back its viewpoints due to its analytical self-segregation (it is not meant to be taken as a technical implementation-based study). The review of numerous related literature – that circle back on the emphasis of factors that heavily participate into the predictive power of any algorithmic implementation – far outweigh comparative studies that acknowledge the need for descriptive machine learning, but do not exactly indicate it as a necessity, in the medical arena. Thus, these many nuances indeed stand as strong points of argumentation in causality preference by way of such systematic literature breakdown.

Essentially, the study is able to point out a very existent and fundamental tradeoff with the research necessity: necessary prediction is amplified through strong machine learners which do not exactly correlate to sufficient explainable knowledge bases, while weak learners that co-exist with such sufficiency are no longer the need of these associated industries. This is an aspect that cannot be critiqued and disqualified due to its genuine presence - it is indeed a smart foundation upon which the authors have made a baseline for this study. This facet also presented an elaborate take on why deep learning remains excluded in the realm of exploiting causable knowledge sets, as such inherent methods treat the cohesion of data and implementation as a 'black-box' instead which thus cannot be consumed for tasks such as knowledge base generation for dedicated understanding.

To top it off, a medical histomorphological (the structure of normal or abnormal tissues) case study is presented to prove what a detailed set of explanatory behavior of the tissue breakdown is required for it to acquire a medical grade standard and inclusion via explainable AI. This was a unique (but personally effective) way of aiming to convince the paper reader of what exactly the trade-off of factor inclusion and description entails in the high-risk settings of the medical world, where lives could be on the line even if one such important attribute or its value is amiss.

However, surprisingly enough, in the former introductory halve of the study even with these notions of a stout basis presented and defended later on, the research appears confused in whether to downplay the importance of explainable AI or extrapolate its significance, as it continuously goes back and forth over its own research claims – one such example is when, after a hefty yet academic act of convincing that causality has the upper hand in the growing sector of artificial intelligence, it claims that AI descriptors eventually aid the understanding process of any medical condition profoundly. The line between necessity and enhancement thus remains blurred by the conclusion of the study, even though there are astute reasons provided in its well-placed rebuttals throughout the length of the paper. There is something left to be desired by the end of the research as the authors claim no explicit preference to either causability or explainability over the other, even though the initial implication was strongly that causability required – rightly so, in their outlined opinion – more research devotion. The study leaves the reader with the impression that both have to be in tandem extensively in order to provide coveted solutions engineered through the field of artificial intelligence, even though the article was attempting to persuade otherwise.

Overall, it was a very comprehensive and reliable research on the effects of causable artificial intelligence in comparison to explainable categories, and why it is entitled to far more consideration in the blooming collaborative field of medicine and AI. It would have been an outstanding study if the ambiguities regarding which elemental quality has the upper hand – via proof of their individual research – had been conclusively resolved in an open manner by the paper's culmination.