International Research Journal of Education and Technology
Peer Reviewed Journal
ISSN 2581-7795

Towards Human Ai Interface-TO SUPPORT EXPLAINABILITY AND CAUSALITY IN MEDICAL AI

NAME: MOUSIKA A

DEPARTMENT: COMPUTER SCIENCE AND ENGINEERING

COLLEGE: BANNARI AMMAN INSTITUTE OF TECHNOLOGY

ERODE SATHYAMANGALAM

mousika.cs20@bitsathy.ac.in

NAME: GOWTHAM M

DEPARTMENT: INFORMATION AND TECHNOLOGYAA

COLLEGE: BANNARI AMMAN INSTITUTE OF TECHNOLOGY

ERODE SATHYAMANGALAM

gowtham.it20@bitsathy.ac.in

NAME: SRI HARINIVAS SP

DEPARTMENT: COMPUTER SCIENCE AND BUSINESS SYSTEM

COLLEGE: BANNARI AMMAN INSTITUTE OF TECHNOLOGY

ERODE SATHYAMANGALAM

sriharinivas.cb20@bitsathy.ac.in

NAME: MITHRA A

DEPARTMENT: COMPUTER SCIENCE AND ENGINEERING

COLLEGE: BANNARI AMMAN INSTITUTE OF TECHNOLOGY

ERODE SATHYAMANGALAM

mithraa.cs20@bitsathy.ac.in

International Research Journal of Education and Technology
Peer Reviewed Journal
ISSN 2581-7795

ABSTRACT-Our definition of causality refers to the degree to which humans may comprehend a particular machine explanation. With the use of a real-world cancer research case, we demonstrate causality. We argue for the use of causality in the design and assessment of future human-AI interfaces in the field of medical artificial intelligence (AI). Since this field's inception, the goal has been to develop artificial intelligence (AI) that is on par with human intelligence. Big data and the necessary processing

capacity have now made statistical machine learning, especially deep learning, quite advanced, even in fields as complex as medicine. One such example is the Stanford machine learning group's research on dermatology1, which gained popularity in Europe under the slogan "AI is better than doctors."

Keywords—AI ,Medical ,XAI, human-AI

I. INTRODUCTION

The team simply used pixels and illness labels as inputs for the classification of skin lesions and trained a deep learning model straight from dermatological photos. They used 130,000 clinical photos with roughly 2,000 different disorders for pretraining and 1.3 million images from the 2014 ImageNet challenge. The results were comparable to or even superior to those of human dermatologists, with an average classification performance of 92%. This is a great accomplishment, and it is clear that AI will play a significant role in the future of medicine.

We must be mindful that these earlier methods rely on statistical model-free learning notwithstanding their spectacular outcomes. It can be extremely risky to rely exclusively on statistical correlations, especially in the field of medicine where correlation should not be confused with causality the existing AI utterly lacks. This is a widespread issue. A special problem is that even for domain experts, it might be challenging, if not impossible, to grasp how the findings were obtained because these approaches are so sophisticated, high dimensional, nonlinear, and nonconvex. As a result, these methods are known as black-box models.

Both issues drive us to utilize the knowledge of a human in the loop. Sometimes, but not always, a human expert can add experience, conceptual understanding, and situational awareness.

A fairly recent study on histopathology demonstrated that some morphological constructions and ontological linkages produced by humans coincide with those produced by machines.

The human experience can help to increase algorithm robustness and explain ability, which are seen as the two biggest hurdles facing contemporary AI, and these overlaps between people and AI can help solve the problems that are now plaguing society. 3 It is no accident that we mention robustness and interpretability together because both are strongly tied to the capacity for generalization. Robustness is a characteristic shared by biological systems and by humans. It ensures that certain system functions are kept up despite external and/or internal disturbances. We are all aware that the

poor data quality in the medical field makes even the greatest machine learning algorithms extremely delicate and vulnerable to even little distortions.

The human experience can help to increase algorithm robustness and explain ability, which are seen as the two biggest hurdles facing contemporary AI, and these overlaps between people and AI can help solve the problems that are now plaguing society. 3 It is no accident that we mention robustness and interpretability together because both are strongly tied to the capacity for generalization. Robustness is a characteristic shared by biological systems and by humans. It ensures that certain system functions are kept up despite external and/or internal disturbances. We are all aware that the poor data quality in the medical field makes even the greatest machine learning algorithms extremely delicate and vulnerable to even little distortions. As a result, a physician in the loop4 will be crucial in medical AI, at least in the near future. In general, people are resilient, and they occasionally enhance machine learning systems through their expertise, conceptual understanding, and implicit knowledge. Despite their tendency to make mistakes, humans are adaptable, resourceful, and plastic, which helps them understand and make sense of their surroundings in the context of an application domain. AI, on the other hand, is sensitive to even minor disturbances. We stress that the current methods not only lack robustness and generality but also, and more crucially, are unable to create causal models that can support the user's deep understanding. As a result, a physician in the loop4 will be crucial in medical AI, at least in the near future. In general, people are resilient, and they occasionally enhance machine learning systems through their expertise, conceptual understanding, and implicit knowledge. Despite their tendency to make mistakes, humans are adaptable, resourceful, and plastic, which helps them understand and make sense of their surroundings in the context of an application domain. AI, on the other hand, is sensitive to even minor disturbances. We stress that the current methods not only lack robustness and generality but also, and more crucially, are unable to create causal models that can support the user's deep understanding. Such a strategy will need suitable human-AI interfaces that allow for easy interaction

with machine learning techniques. But first, let's list the accomplishments of the explainable AI (XAI) community.

II. EXPLAINABILITY AND CASUALTY

Real-world issues include two issues. 1) When making a medical diagnosis, the ground reality is not always clear cut. 2) While correlation is acknowledged as a basis for judgements, it must be seen as an intermediary stage.

Human (scientific) models are frequently built on causality as the ultimate aim for understanding the underlying explanatory mechanisms. Due to the value of validity, the need to develop human trust, and the requirement to establish "AI experience," this is extremely pertinent to the medical field. Real-world issues include two issues. 1) When making a medical diagnosis, the ground reality is not always clear-cut. 2) While correlation is acknowledged as a basis for judgements, it must be seen as an intermediary stage. Human (scientific) models are frequently built on causality as the ultimate aim for understanding the underlying explanatory mechanisms. Due to the value of validity, the need to develop human trust, and the requirement to establish "AI experience," this is extremely pertinent to the medical field. As we have mentioned, the most successful algorithms are based on

probabilistic models and provide only a rudimentary basis for establishing causal models. Consequently, when we discuss the explainability of a machine statement, we have to carefully distinguish among the following terms:

Explainability

Technically speaking, decision-relevant components of the machine representations of the algorithms that are being used and active components of the algorithmic model that either contribute to the model accuracy on the training set or to a specific prediction for one particular observation are highlighted by explainability. It doesn't specifically refer to a human model.

Usability**:**

This phrase describes the quantifiable degree to which a system satisfies a user's needs for effectiveness, efficiency, and satisfaction within a given use environment.

Causability**:**

The quantifiable degree to which a statement's explanation to a human expert achieves a given level of causal comprehension with efficacy, efficiency, and satisfaction in a given context of use is known as causality. Causality refers to a human understandable model since it is evaluated in terms of the efficacy, efficiency, and (human) satisfaction related to causal comprehension and its transparency for an expert user.

Given that the explanation is, by definition, defined in relation to a human model, this is always conceivable for an explanation of a human assertion. However, a causal model in the sense of Pearl must be used to represent causal links and allow inferences about those causal ties from data in order to test the usability of an explanation of a machine statement.

For the majority of AI algorithms, this is not the case, hence a mapping between the two must be established. Here, we must create a distinction between the explainable model (XAI) andan explanation interface that enables the expert to use and benefit from the results obtained by the explainable model.

Explainability

The XAI community is actively working on techniques to make black-box approaches understandable. There are several methods that can be used to arrive at such "mechanical explanations." In the simplest approach, the deep neural network is viewed as a function, and the explanation dependson the function's gradient, which is available via the backpropagation algorithm.

Gradients are used as a multivariable generalization of the derivative in this approach. The XAI community is actively working on techniques to make black-box approaches understandable. The effective strategies human-understandable model mh, which is judged in terms of emphasize identifying which input factors contribute more to a efficacy, efficiency, satisfaction connected to causal

particular classification result by using heat mapping, for comprehension, and its transparency for a user. Since an example, to highlight the elements that have contributed to each explanation of a human statement is inherently defined in conclusion. There are several methods that can be used to relation to mh, this is always a possibility. In conclusion, arrive at such "mechanical explanations." In the simplest either mh must be based on a causal model (which is not the approach, the deep neural network is viewed as a function, and case for most machine learning algorithms) or a mapping

the explanation depends on the function's gradient, which is between mm and mh must be defined in order to measure the available via the backpropagation algorithm. Gradients are used usability of an explanation of a machine statement sm.

as a multivariable generalization of the derivative in this

approach.

A. HUMAN-AI INTERFACE : EFFECTIVE MAPPING OFEXPLAINABILITY WITH CAUSABILITY

An effective and consistent mapping of explainability with causability is essential for successful human-AI interaction and, by extension, future human-AI interactions. This "mapping" (or "map metaphor") is not about creating a new map; rather, it is about creating links and interconnections between already existing places. Instead, it involves locating the same, or at least comparable, regions in two entirely separate "maps" of AI explainability and human causability. Because of this, "mapping" is a very useful term. Effective and efficient mapping is required, but it won't suffice to understand an explanation, of course.

We can see that an explanatory statement s, where s is a function s = f (r, k, c) with the following parameters: can either be made by a human sh or a machine sm. k, pre existing knowledge, which is incorporated into an algorithm for a machine or created for a human by explicit, implicit, and/or tacit knowledge; r, representations of an unknown (or unobserved) fact ue related to an entity; c, context, which for a machine is the technical runtime environment and for a human is the physical environment in which the decision was made (the pragmatic dimension). A ground truth that we attempt to simulate with a machine mm or as a human mh is represented by an unknown (or undiscovered) fact.

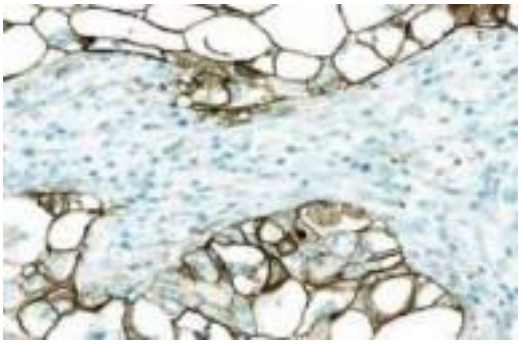
The primary objectives are that the offered explanation of the statement s fits this ground truth and that the statement s is identical to the given ground truth gt. In an ideal circumstance, the ground truth, which is specified for machines and people inside the same framework, is congruent (mh mm) and identical to both the human and the machine statement. The issue with machine learning in the medical field is that the most effective models are based on correlation or closely related ideas of similarity and distance. All of this must be viewed as an intermediary stage that can only serve as a foundation for the development of additional causal models because it is probabilistic in nature.

The decision-relevant components of the machine representations rm and machine models mm, or the components that contributed to model accuracy during training or to a particular prediction, are highlighted by explainability in a technical sense. The fact that explainability does not refer to a human model mh must be emphasized. In order to accomplish a specific level of causal comprehension with effectiveness, efficiency, and satisfaction in a specified context of usage, an explanation of a statement to a user must be as causally plausible as possible. Causability refers to a Here it is important to classify, filter, and make use ofaffective computing methods to

1. measure the effectiveness of explainability (does the user really understand a given explanation), which can be measured via sensors by the success
   rate to which a given explanation has been understood by humans; see the examples in the section “A Clinical Case”
2. adapt the visual communication according to themental model (a priori knowledge) of the
   user.

Clinical-Case

Lung cancer will be used as a real-world illustration. The use of immunotherapies against programmed cell death ligand 1 (PD-L1) and its receptor PD-1 can increase the survival of patients with lung cancer. A surface protein called PD-L1 is involved in the suppression of the immune response. The expression of the PD-L1 protein has become a valuable diagnostic for identifying people who will respond better to immunotherapy. 19 Whether or not the immune system targets the tumor-causing degenerate cells as a possible threat determines whether a tumor can spread throughout the body. To distinguish between the body's own cells and foreign cells, the immune system scans every cell. T cells play a major role in this process as part of the immune system. T cells can identify cancer cells, but the immune system does not fight them because they can mask their appearance by releasing the protein PD-L1. PD-L1 functions as a mask that enables cancer cells to hide and go undetected, and immunohistochemistry is one method for detecting PD-L1. Figure 3 depicts a typical portion of a complete slide image used in an ongoing validation project that focuses on both clinical performance and the human-AI interface (see the ethics statement in the "Acknowledgments" section for more information).



To the human pathologist, the tumor is visible above and below in the image.The stroma is visible in the center (tumor identification is 100% positive).

The most effective AI techniques, however, are what are known as "black boxes," which are so complex that a human expert would find it difficult or impossible to grasp how a conclusion was reached. Due to expanding legal requirements, future outcomes must be rendered retractable, comprehendible, and clear enough that a human expert would find it difficult or ACKNOWLEDGMENTS:

We certify that real-world facts were utilized for this study, but that just the fundamentals of the human-AI interaction were looked into and utilized for the section under "A Clinical Case." We appreciate Luca Brcic and Markus Plass' comments on the Roche Ventana Study's favorable ethical verdict. The European Union's Horizon 2020 research and innovation programme provided financing for some of this work under project agreements 857122 (CY-Biobank), 824087 (European Open Science Cloud-Life), 874662 (Human Exposome Assessment Platform), and 826078 (Feature Cloud).

Reference:

1.A. Esteva et al., “Dermatologist-level classification of skin cancer with deep neural networks,” *Nature*, vol. 542, no. 7639, pp. 115–118, 2017. doi: 10.1038/nature21056.

2.K. Faust et al., “Intelligent feature engineering and ontological mapping of brain tumor histomorphologic by deep learning,” *Nature Mach. Intell.*, vol. 1, no. 7, pp. 316–321, 2019. doi: 10.1038/s42256-019-0068-6.

3.R. Hamon, H. Junklewitz, and I. Sanchez, *Robustness and Explainability of Artificial Intelligence—From Technical to Policy Solutions*. Luxembourg: Publications Office of the European Union, 2020.

A. Figures and Tables

a) Positioning Figures and Tables: Place figures and tables at the top and bottom of columns. Avoid placing them in the middle of columns. Large figures and tables may span across both columns. Figure captions should be below the figures; table heads should appear above the tables. Insert figures and tables after they are cited in the text. Use the abbreviation “Fig. 1”, even at the beginning of a sentence.

TABLE I. TABLE TYPE STYLES

Table Head	**Table Column Head**		
	Table column subhead	***Subhead***	***Subhead***
copy	More table copy[a]		

[a.] Sample of a Table footnote. (*Table footnote*)

Fig. 1. Example of a figure caption. (*figure caption*)

impossible to grasp how a conclusion was reached. Due to expanding legal requirements, future outcomes must be rendered retractable, comprehendible, and clear to a human expert. The rapid expansion The XAI research group has already created a number of very effective explainability techniques. The XAI community defines explainability in terms of highlighting decision-relevant elements of a result. The fact that these analyses do not explicitly use a human model is another critical element. This inspired us to presentour theory of causality.

Figure Labels: Use 8 point Times New Roman for Figure labels. Use words rather than symbols or abbreviations when writing Figure axis labels to avoid confusing the reader. As an example, write the quantity "Magnetization", or "Magnetization, M", not just "M". If including units in the label, present them within parentheses. Do not label axes only with units. In the example, write "Magnetization (A/m)" or "Magnetization {A[m(1)]}", not just "A/m". Do not label axes with a ratio of quantities and units. For example, write "Temperature (K)", not "Temperature/K".

ACKNOWLEDGMENT *(Heading 5)*

The preferred spelling of the word "acknowledgment" in America is without an "e" after the "g". Avoid the stilted expression "one of us (R. B. G.) thanks ...". Instead, try "R.
B. G. thanks...". Put sponsor acknowledgments in theunnumbered footnote on the first page.

REFERENCES

The template will number citations consecutively within brackets [1]. The sentence punctuation follows the bracket [2]. Refer simply to the reference number, as in [3]—do not use "Ref. [3]" or "reference [3]" except at the beginning of a sentence: "Reference [3] was the first ..."

Number footnotes separately in superscripts. Place the actual footnote at the bottom of the column in which it was cited. Do not put footnotes in the abstract or reference list. Use letters for table footnotes.

Unless there are six authors or more give all authors' names; do not use "et al.". Papers that have not been published, even if they have been submitted for publication, should be cited as "unpublished" [4]. Papers that have been accepted for publication should be cited as "in press" [5]. Capitalize only the first word in a paper title, except for proper nouns and element symbols.

For papers published in translation journals, please give the English citation first, followed by the original foreign language citation [6].

[1] G. Eason, B. Noble, and I. N. Sneddon, "On certain integrals of Lipschitz-Hankel type involving products of Bessel functions," Phil. Trans. Roy. Soc. London, vol. A247, pp. 529–551, April 1955. *(references)*

[2] J. Clerk Maxwell, A Treatise on Electricity and Magnetism, 3rd ed.,vol. 2. Oxford: Clarendon, 1892, pp.68–73.

[3] I. S. Jacobs and C. P. Bean, "Fine particles, thin films and exchange anisotropy," in Magnetism, vol. III, G. T. Rado and H. Suhl, Eds. New York: Academic, 1963, pp. 271–350.

[4] K. Elissa, "Title of paper if known," unpublished.

[5] R. Nicole, "Title of paper with only first word capitalized," J. NameStand. Abbrev., in press.

[6] Y. Yorozu, M. Hirano, K. Oka, and Y. Tagawa, "Electron spectroscopy We suggest that you use a text box to insert a graphic (which is ideally a 300 dpi TIFF or EPS file, with all fonts embedded) because, in an MSW document, this method is somewhat more stable than directly inserting a picture.

To have non-visible rules on your frame, use the MSWord "Format" pull-down menu, select Text Box > Colors and Lines to choose No Fill and No Line.

studies on magneto-optical media and plastic substrate interface," IEEE Transl. J. Magn. Japan, vol. 2, pp. 740–741, August 1987 [Digests 9th Annual Conf. Magnetics Japan, p. 301, 1982].

[7] M. Young, The Technical Writer's Handbook. Mill Valley, CA:University Science, 1989.

IEEE conference templates contain guidance text for composing and formatting conference papers. Please ensure that all template text is removed from your conference paper prior to submission to the conference. Failure to remove template text from your paper may result in your paper notbeing published.

The team simply used pixels and illness labels as inputs for the classification of skin lesions and trained a deep learning model straight from dermatological photos. They used 130,000 clinical photos with roughly 2,000 different disorders for pretraining and 1.3 million images from the 2014 ImageNet challenge. The results were comparable to or even superior than those of human dermatologists, with an average classification performance of 92%. This is a great accomplishment, and it is clear that AI will play a significant role in the future of medicine.

We must be mindful that these earlier methods rely on statistical model-free learning notwithstanding their spectacular outcomes. It can be extremely risky to rely exclusively on statistical correlations, especially in the field of medicine where correlation should not be confused with causality the existing AI utterly lacks.