

## A Review on Pathology, Artificial Intelligence and the Explainability Conundrum

Mohammad Haeri<sup>1\*</sup> and Mohammad Hossein Jarrahi<sup>2</sup>

<sup>1</sup>Department of Pathology and Laboratory Medicine, University of Kansas Medical Center, Kansas, USA

<sup>2</sup>University of North Carolina, North Carolina, USA

**\*Corresponding author:** Dr. Mohammad Haeri, Department of Pathology and Laboratory Medicine, University of Kansas Medical Center, Kansas, USA, E-mail: mhaeri@kumc.edu, Mohammad Hossein Jarrahi, University of North Carolina, USA, E-mail: jarrahi@unc.edu

**Received:** 03-Feb-2023, Manuscript No. DPO-23-88536; **Accepted:** 10-Feb-2023, Manuscript No. DPO-23-88536, **Published:** 20-Feb-2023

DOI: 10.4172/2476-2024.8.1.209

**Citation:** Haeri M, Jarrahi MH (2023) Pathology, Artificial Intelligence and the Explainability Conundrum. *Diagnos Pathol Open* 8: 209.

**Copyright:** © 2023 Haeri M, et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

### Abstract

Artificial intelligence in medical diagnosis, including pathology, provides unprecedented opportunities. However, the lack of explainability of these systems raises concerns about the proper adoption, accountability and compliance. This article explores the problem of opacity in end-to-end AI systems where pathologists might only serve as trainers of the algorithm. A solution is suggested with the "pathologists-in-the-loop" approach, which involves continuous collaboration between pathologists and AI systems through the concepts of parameterization and implicitization. This human centered workflow enhances the pathologist's role in the diagnosis process to create an explainable system rather than an automated report generator.

**Keywords:** Pathology; Machine learning; Parameterization; Implicitization; Explainable Artificial Intelligence (XAI)

### Introduction

The practice of pathology is facing new hopes and challenges through the use of high-throughput glass slide scanners and, more importantly, advances in Artificial Intelligence (AI) and machine learning breakthroughs. Newly developed systems provide histopathologic diagnosis of cancers with acceptable accuracy through image analysis of standard H and E slides even without utilizing accompanying immune histochemical studies. However, these systems are adopted only sporadically in general medical practices. A key reason for the reasonable hesitation in employing AI-derived diagnoses is its unexplainable nature, commonly referred to as "black-box" or inscrutable AI. The black-box AI is the product of a system trained by thousands of similar cases to achieve an acceptable accuracy in providing a pathologic diagnosis. However, there is little explanation about the reasoning behind making a specific diagnosis. In other words, in contrast to the routine practice of pathologists to "reach a diagnosis" by analysing morphological features and other relevant data to develop a differential diagnosis, an AI-based system "jumps to a diagnosis" without providing adequate evidence for its decision.

### Literature Review

As such, AI-based diagnosis is largely unexplainable to the pathologist and even to computer scientists who developed the system, and this is the origin of hesitation or even suspicion among the community of clinicians. More importantly, the unexplainable nature of AI-based diagnoses violates the principle of medical ethics and medical accountability. Even though these opaque diagnostic processes might operate effectively, they may still fail to meet the legal requirements and regulations. This means that simply having a

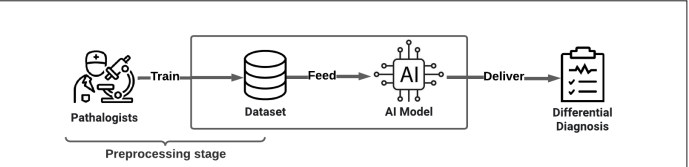
high level of efficiency in the diagnostic process is not enough to ensure compliance with the law.

This deficiency in the commonly adopted AI models by the technical community has led to the advent of explainable AI (XAI) [1]. XAI aims to make the AI-empowered process of pathologic diagnoses more transparent by providing a set of rationales for the generated diagnosis. However, this reasoning is not as detailed or transparent as the reasoning provided by a pathologist, and it may not fully capture the complexity and nuances of the diagnostic process. Current XAI methods in medical applications often give a post hoc, general explanation for AI system inferences. These can offer general insight into how the AI system reasoned, but for specific decisions, the explanations may be untrustworthy or only

Superficially informative (e.g., why the system thought specific features were important). For example, saliency maps are considered a popular AI explainability approach pinpointing the image parts central to AI predictions. However, a study's heat map for pneumonia prediction only showed a large part of one lung, leaving unclear what led to the conclusion. The lack of information on the AI's decision raises concerns about its reliance on image acquisition factors (e.g., a particular pixel value or texture), rather than disease-related factors, like an airspace opacity, heart border, or pulmonary artery shape [2].

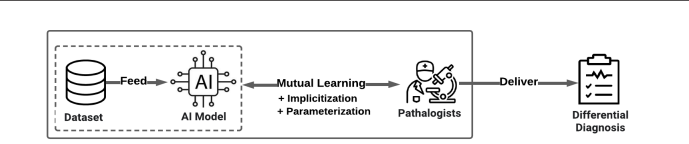
Conclusively, while the XAI provides a degree of explanation for its decision, it is not practically useful for a clinician and may not be even equally meaningful for different members of the technical community. A focus on the end-to-end nature of AI systems is necessary to comprehend the underlying reasons for their opaqueness. While these systems may prove valuable in contexts where transparency is not a paramount concern (e.g., Netflix's recommendations algorithm), in high-stakes decision-making situations such as medical diagnosis, their adoption could be

counterproductive. In this workflow, pathologists are typically involved during the pre-processing phase, where they help train the system (Figure 1). The role of pathologists in this context is therefore implicitly relegated to augmenting the capabilities of the AI system, rather than the other way around.



**Figure 1:** End-to-end (black-box) AI workflow.

In this piece, we propose an explainable AI workflow that we refer to as “pathologists-in-the-loop”. The workflow involves the active participation of pathologists as domain experts throughout the process, beyond just the data pre-processing and training stages (Figure 2). The partnership between pathologists and the AI system is ongoing and is a source of mutual learning for both. In this scenario, the AI system is utilized to augment the work of the pathologist, rather than to automate their work processes (Table 1).



**Figure 2:** Explainable, pathologists-in-the-loop workflow.

Approach	The role of pathologists	The role of AI systems	The nature of interaction	Explainability
End-to-end AI-based workflow	Pathologists only involved in pre-processing stages	Automating pathology workflow	AI being trained by pathologists during the pre-processing stages	Unexplainable to pathologists and other stakeholders
“Pathologists in the loop” workflow	Pathologists continuously involved throughout the workflow as domain experts and final decision makers	Augmenting pathologist’s workflow	Mutual and continuous learning between pathologists and AI	Explainable to AI systems, pathologists and other stakeholders

**Table 1:** The comparison between the end-to-end AI-based and “pathologists in the loop” workflows.

Our vision for pathologists in the loop is based on two concepts; parameterization and implicitization [3]. This approach helps gain the trust of pathologists, clinicians, and regulators. In simple terms, parameterization is about breaking down the problem into smaller components (parameters), while implicitization combines these components back into a final diagnosis by taking a holistic perspective and placing it within a broader context. Parameterization involves identifying and expressing implicit equations (e.g., a tumor

using a variety of parameters (e.g., histopathological features). This “divides and conquers” approach breaks down the entire pathologic state into histopathological features that are simple enough to be observed and understood directly. Implicitization is the inverse process, converting parameters and their associations back to a single implicit disease state (e.g., the tumor type).

For instance, we train the AI to detect mitosis events with high accuracy. Highly accurate mitosis detection is acceptable even if it is not fully explainable for pathologists. This is a task defined by pathologists, carried out by AI, and will be utilized to reach a final integrated diagnosis by the pathologist to close the loop. Once again, the pathologist can accept the highly accurate mitosis detection rate even if a clear-cut explanation is not available. In this setting, the system is tailored to meet the pathologist’s need while the final word is made by the expert. To provide a more sensible example, we recently utilized the same parameterization and implicitization approach to determine the tumor grade in meningioma [4]. We parameterized (broke down) features associated with different tumor grades and trained the AI system to find those features to assist the pathologist to make the final integrated diagnosis (implicitization). This system assists the pathologist to make the final diagnosis utilizing an explainable AI that takes a transparent technological path.

Obviously, the current black-box AI will also be able to classify meningioma into three different grades by acceptable accuracy when trained by thousands of subjects; however, no pathologist would agree to sign out a case by solely relying on a tumor grade generated by an end-to-end AI system that is largely unexplainable. In contrast, a tumor grade suggested by a smart system with provided evidence defined by standard criteria will be utilized by a pathologist with confidence. In the latter, the system performs the lengthy task of detecting mitotic figures, small foci of necrosis among other features, and the pathologist makes the final tumor grade and diagnosis by looking at the evidence. This system is designed for pathologists’ needs and practices and is explainable for all clinicians, regulators, and other stakeholders. There is no doubt that the development of an evidence-based and explainable AI is more complex compared to creating a black-box system. However, pathologists and those who are regulating the medical devices and laboratory tests will not accept a black-box system when an alternative explainable AI is available. The future will show how AI is implemented in medical practices and how perhaps a combined system with comprehensive explainability and high accuracy paves the road for future pathologists.

A less appreciated advantage of the “pathologist-in-the-loop” strategy is the valuable mutual learning with an emphasis on less-explained pathologist’s gain of new knowledge. The computational power and AI-assisted image analysis provide new morphological features unseen by the pathologist’s eye. Human eyes are not tuned to detect minor but significant variations in size, shape, and contour to mention only a limited list of possible features. In contrast, computational power assisted by AI can provide significant and previously unexplained histological features of specific pathologies. For instance, in order to differentiate renal oncocytoma from chromophobe renal cell carcinoma we noticed significant nuclear density between the two tumors which was not explained previously [5,6]. We believe this is just the beginning of this body of knowledge that can be discovered through mutual learning gained by utilizing the “pathologist-in-the-loop” strategy.

## Discussion and Conclusion

The utilization of AI in automating the work of pathologists has proven to be both ineffective and infeasible due to the opacity of AI systems, empowered by deep learning. A viable alternative is the adoption of a "pathologist-in-the-loop" strategy that emphasizes continuous collaboration between pathologists and AI systems through means of parameterization and implicitization. This collaboration helps address the previously opaque AI process and makes it transparent to all relevant stakeholders. In this alternative workflow, the pathologist remains a key player and decision-maker, not merely an annotator. For example, pathologists interact with data and actively participate in feature extraction (parameterization) while collaborating with the AI system. They also produce the final decision by putting the AI-enabled parameterization into a holistic pathologic perspective (implicitization).

The pathologists' ability to contextualize is a central aspect of the pathologist-in-the-loop approach. While AI systems may identify contextual features, it is the pathologist who integrates all relevant information (e.g. patients' history) to make the final integrated diagnosis. Such a transparent collaboration can improve trust and accountability as the pathologist can better understand how the AI system arrives at its decisions. This approach leverages the strengths of both partners and addresses the disconnect between AI's knowledge and expert's experience commonly seen in medical AI applications.

## Acknowledgments

Authors thank Ms. Cynthia Shaddy-Gouvion and Dr. Robert Hoyt from the Department of Pathology and Laboratory Medicine of the University of Kansas Medical Center for their valuable comments.

## References

1. Adadi A, Berrada M (2018) Peeking Inside the Black-Box: A Survey on Explainable Artificial Intelligence (XAI). *IEEE Access* 6:52138-60.
2. Ghassemi M, Oakden-Rayner L, Beam AL (2021) The false hope of current approaches to explainable artificial intelligence in health care. *Lancet Digit Health*.3:e745-e750.
3. Jarrahi MH, Davoudi V, Haeri M (2022) The key to an effective AI-powered digital pathology: Establishing a symbiotic workflow between pathologists and machine. *J Pathol Inform* 13:100156.
4. Gu H, Liang Y, Xu Y, Williams CK, Magaki S, et al. (2020) Improving Workflow Integration with xPath: Design and Evaluation of a Human-AI Diagnosis System in Pathology. *ACM Trans. Comput.-Hum. Interact* 29:1-9.
5. Haeri M, Zarrin-Khameh N, Citron D, Finch CJ, Wheeler T (2018) Computer-assisted image analysis offers accurate diagnostic aid; differentiating chromophobe renal cell carcinoma from renal oncocytoma. In: *Laboratory investigation*. Nature Publishing Group, New York, USA, pp: 10013-11917.