


Chapter 11

Exploring Explainable AI in Healthcare: Challenges and Future Directions

Adla Padma

 <https://orcid.org/0000-0002-3168-4762>
Vellore Institute of Technology, India

Vasavi Chithanuru

Vellore Institute of Technology, India

Posham Uppamma

Vellore Institute of Technology, India

R. VishnuKumar

Vellore Institute of Technology, India

ABSTRACT

Artificial intelligence (AI) has revolutionized the healthcare industry by making decisions similar to human intelligence. However, the need to illustrate AI predictions based on healthcare data is a challenging task. To address this, Explainable AI (EXAI) has emerged to provide transparent explanations for machine-generated predictions and ensure accuracy in healthcare. This review emphasizes the importance of adopting EXAI in healthcare and discusses its concept of providing reliable AI-based solutions. The authors analyze the most recent developments in EXAI-based technologies and present research findings on their extensive implementation aspects, including the challenges and limitations of existing models. The importance of EXAI in healthcare extends from early disease prediction to intelligent diagnosis. Furthermore, this survey provides insights into the future perspectives of EXAI in healthcare, offering valuable research directions. Integrating EXAI into healthcare can enhance transparency, interoperability, and trust in AI-driven healthcare solutions.

DOI: 10.4018/979-8-3693-5468-1.ch011

1. INTRODUCTION

Of late, Artificial Intelligence (AI) was established in 1950 in the field of computer science and it imitates human cognition in order to create systems capable of efficiently processing large datasets. AI is a system that reproduces human intelligence to execute real-world tasks. It trains the system based on data and learns from experiences to resolve complex challenges. AI plays a vital role in technology complication, automation, and more sophisticated, which recommends a tremendous boost in AI development (Islam et al., 2022). AI can progress individuals' overall well-being and strength by augmenting healthcare professionals' diagnostic capabilities, prioritizing preventive measures, and offering personalized treatment recommendations inside automatic health accounts (EHRs). This is important in healthcare, where AI is increasingly used to make patient health monitoring decisions. For instance, AI systems can be used to identify diseases, recommend treatments, and predict patient outcomes. However, healthcare professionals must understand how these systems work and why they make the decisions (Bharati et al., 2023).

Voice recognition, self-driving vehicles, disease prediction and classification (Kumar & Vanmathi, 2022; Uppamma & Bhattacharya, 2023) and recommendation systems are just a few examples of AI's practical achievements that have already impacted people's lives. However, despite several useful technologies, AI has yet to be widely adopted in health care. Because of this, AI, Machine Learning (ML), and Deep Learning (DL) algorithms remain secretive in several situations. Figure 1 shows the relation between AI, ML, DL, and EXAI. There are generally few explanations for decision-making processes. Moral and practical challenges are associated with AI diagnosis, as it is incredible to express whether diagnostic differences reproduce diagnostically significant differences, errors, or else overdiagnoses. AI makes user decisions based on business models but humans need to be made aware of the outcome of AI. However, it is challenging to find the output process of AI. So, Explainable Artificial Intelligence (EXAI) was introduced (Hulsen, 2023).

EXAI, coined by the Defence Advanced Research Projects Agency (DARPA), clarifies a model's internal structure, facilitating users' comprehension of the methods, procedures, and output processes. Due to its ability to elucidate a model's procedures, it is often called the white box approach. As depicted in Figure 2, the procedure begins with introducing training data. Users select the appropriate methodology for prediction based on specific requirements or application domains and employ EXAI techniques to disclose the inner workings of models through an explicative dashboard. This transparency enables consumers to comprehend the outcomes of Explainable AI, thereby increasing their confidence in AI models. Armed with insights from the output, users can improve result precision and identify model defects, allowing them to make informed decisions regarding model enhancement

33 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage: www.igi-global.com/chapter/exploring-explainable-ai-in-healthcare/349123

Related Content

The Place of Humor, Empathy, and Cultural Nuances in AI-Supported Tools

Pankaj Bhanudas Vaishnav (2026). *Pedagogical Innovations in AI-Supported Intercultural Communications* (pp. 65-100).

www.irma-international.org/chapter/the-place-of-humor-empathy-and-cultural-nuances-in-ai-supported-tools/408179

Extending Loosely Coupled Federated Information Systems Using Agent Technology

Manoj A. Thomas, Victoria Y. Yoonand Richard Redmond (2007). *International Journal of Intelligent Information Technologies* (pp. 1-20).

www.irma-international.org/article/extending-loosely-coupled-federated-information/2420

Evaluation of Logistics Development Under the Visual Field of Low-Carbon Environmental Protection Based on Hierarchical Methods

Jinjuan Wang (2024). *International Journal of Ambient Computing and Intelligence* (pp. 1-16).

www.irma-international.org/article/evaluation-of-logistics-development-under-the-visual-field-of-low-carbon-environmental-protection-based-on-hierarchical-methods/360709

Defining Smart Community in the Context of Healthcare Efficiency in the UK: Mapping the Evolution of a Concept

Tim Woolliscroft (2020). *International Journal of Intelligent Information Technologies* (pp. 1-18).

www.irma-international.org/article/defining-smart-community-in-the-context-of-healthcare-efficiency-in-the-uk/262976

How to Manage Persons Taken Malaise at the Steering Wheel Using HAaaS in a Vehicular Cloud Computing Environment

Meriem Benadda, Karim Bouamraneand Ghalem Belalem (2017). *International Journal of Ambient Computing and Intelligence* (pp. 70-87).

www.irma-international.org/article/how-to-manage-persons-taken-malaise-at-the-steering-wheel-using-haaas-in-a-vehicular-cloud-computing-environment/179290