ISSN: 2454-9940

INTERNATIONAL JOURNAL OF APPLIED SCIENCE ENGINEERING AND MANAGEMENT

E-Mail :
editor.ijasem@gmail.com
editor@ijasem.org

www.ijasem.org

INTERNATIONAL JOURNAL OF APPLIED SCIENCE ENGINEERING AND MANAGEMENT
ISSN 2454-9940
www.ijasem.org
Vol 15, Issue 4, 2021
https://zenodo.org/records/15835370

EXPLAINABLE AI FOR ENHANCED INTERPRETABILITY IN MEDICAL DIAGNOSIS

[1]Shekhar Katukoori, [2]Dr.Sandeep Chahal
[1]Research Scholar, shekhar.ktk@gmail.com
Department of Computer Science and Engineering
[2]NIILM UNIVERSITY, Kaithal, Haryana, India. Associate Professor
Computer Science and Engineering, NIILM UNIVERSITY

ABSTRACT

This research study outlines the investigation into the transformative role of Explainable Artificial Intelligence (XAI) in the realm of medical diagnosis, with a focus on enhancing interpretability. As AI algorithms increasingly contribute to diagnostic processes, the opacity of these black-box models poses significant challenges in understanding, trusting, and optimizing their outcomes. This study aims to explore the technical foundations of XAI, unraveling how interpretability can be integrated into machine learning models specifically tailored for medical diagnostic applications. Evaluating the performance of interpretable AI against conventional black-box models, the research addresses the impact of transparency on the decision-making processes of healthcare professionals. Additionally, ethical considerations, patient perspectives, and regulatory implications related to the deployment of XAI in medical diagnosis will be scrutinized. By comprehensively examining these dimensions, this research seeks to provide insights that contribute to the development of guidelines and best practices, fostering responsible and transparent integration of XAI for enhanced interpretability in medical settings.

KEYWORDS : Explainable Artificial Intelligence (XAI), Medical diagnosis, Interpretability, AI algorithms, Black-box models, Diagnostic processes, Transparency, Machine learning models

I. INTRODUCTION

The rapid advances in AI development over the past few decades have resulted in increasingly available AI applications to support human experts in their work, including decision-making. In this paper, we examine how dermatologists use or envisage using AI in their diagnostic work.

We chose medicine as it is one of the most-developed AI application areas, there is already substantial experience in using AI, and the high quality of this use is critical – i.e. lives are at stake. We decided to choose one single area in medicine in order to achieve high consistency. Dermatological diagnosis is a particularly suitable area of study as it makes use of image processing aspect of AI, which is particularly well-developed. Specifically, we focus on the process of diagnosing melanoma; this provides a useful basis for comparing the participants' accounts. In addition, the lead author has access to the participants, which provides the benefit of an "insider view". We have designed an exploratory, qualitative empirical study, aimed at understanding how dermatologists think and feel about AI and using AI, as well as how the use of AI would alter their established diagnostic processes.

This positions our work in the broad area of 'future of work', where the emergence of automation and AI transform human work [[1], [2], [3]]. The human-computer interaction (HCI), which is an important aspect of future work, raises the question of automation or augmentation [4]. Both automation and augmentation are about identifying and resolving employees' weaknesses and limitations, but the approaches and possible consequences for the future of work are quite different (see [5] for a review based on three recent books). Employees rather dislike automation, but they have a much more positive attitude towards augmentation: while automation threatens their jobs and salaries, augmentation increases their work quality and productivity by complementing their weaknesses and enhancing their capabilities [5,6]. Although many different terms are used – including human-AI symbiosis, teams, and collaborations – currently AI-based technologies complement and augment human capabilities rather than replace them, and this can be expected for the foreseeable future [[7], [8], [9], [10], [11]]. The biggest challenge is consistently getting right the integration of AI into the existing organizational processes [12,13].

As we see it, the objective is not to replace the human decision-makers with AI; it is to produce accurate algorithmic predictions, which are then supplemented with the (value) judgments by human experts. The algorithmic predictive capability of AI is an input into the decision-making process, and the human expert's final decision (judgment) remains critical. Thus, our standpoint is what can be legitimately called a "decision support" [14], and what is referred to more recently as "decision augmentation" [15]. We found that the process of melanoma diagnosis consists of two components: a prediction and a judgment. In a human-only scenario, i.e. when no AI is used,

the predictions are created and the judgments are made by the doctor: therefore the two components of the process are intertwined and the process is nonlinear. When AI is used, the two components are disentangled – contributed by separate entities – and thus they are arguably less intertwined; rather, they build on each other in a fashion that resembles a linear process. This does not suggest a lower complexity however, as what constitutes the complexity changes as well – namely, the human-AI interaction becomes an additional source of complexity.

In order to depict how dermatologists think and feel about using AI in their diagnosis, in what follows, we first provide a brief overview of the background knowledge on using AI in medicine. Then we outline our methodological considerations, explain our choices, and describe the scope of the study. Next, we present our findings, organized around four themes (aggregate dimensions): the role of AI, responsibility, explainability, and the mindset needed to work with AI. Subsequently, we discuss the findings in the light of the extant literature, highlighting what is significant about our improved understanding, and exploring the implications of the findings. We finish with a final commentary in which we account for the lessons learned for designing AI to support physicians in a manner that suits them.

Explainable Artificial Intelligence (XAI) plays a pivotal role in revolutionizing medical diagnosis by enhancing interpretability and transparency in the decision-making process. In the realm of healthcare, where accurate and trustworthy diagnoses are paramount, XAI provides a pathway for understanding and validating the complex algorithms employed in medical decision support systems. By incorporating transparent models and interpretable features, Explainable AI ensures that clinicians and healthcare professionals can comprehend the reasoning behind AI-driven diagnoses, fostering a symbiotic relationship between human expertise and machine intelligence. This not only instills confidence in the reliability of AI-assisted medical diagnoses but also enables practitioners to make informed and accountable decisions, ultimately contributing to improved patient outcomes and the overall advancement of medical practices.

II. LITERATURE REVIEW

Amina Adadi et al (2018)

We are witnessing a rapid and pervasive integration of artificial intelligence (AI) into our daily lives at the onset of the fourth industrial revolution, which contributes to the acceleration of the transition to a more algorithmic society. Nevertheless, despite these unprecedented

developments, a significant barrier to the implementation of AI-based systems is that they frequently lack transparency. Undoubtedly, these systems' black-box characteristics enable formidable prognostications; nevertheless, they defy direct explanation. This matter has instigated a fresh discourse regarding explainable artificial intelligence (XAI). A domain of study exhibits considerable potential in enhancing the confidence and openness of AI-powered systems. It is acknowledged as an absolute necessity for AI to maintain a consistent and uninterrupted progression. This survey offers interested researchers and practitioners an initial opportunity to gain knowledge on fundamental elements of the nascent and swiftly expanding field of XAI research. In this literature review, we critically examine the current methodologies employed in the field, analyze prevalent trends that pertain to it, and outline significant avenues for future research.

Andreas Holzinger et al (2019)

In the medical field, explainable artificial intelligence (AI) is generating considerable interest. Explainability has been a technical challenge since the inception of artificial intelligence (AI), with classic AI utilizing comprehensible and traceable methods. Nevertheless, their deficiency lay in managing the uncertainties inherent in the actual world. The implementation of probabilistic learning led to a rise in the success of applications, albeit at the expense of greater transparency. Explainable AI pertains to the execution of traceability and transparency measures for statistical black-box machine learning techniques, with a specific focus on deep learning (DL). Our position is that something must be done beyond explainable AI. Causability is required to attain an air of explanation in medicine. Similar to how causability incorporates metrics to assess the quality of explanations, usability encompasses metrics to evaluate the quality of use. This article presents essential definitions for differentiating between explainability and causability, along with an illustration of the application of both DL interpretation and human explanation in the field of histopathology. The primary contribution of this article is the definition of causability, which distinguishes itself from explainability by stating that causability pertains to an individual, whereas explainability pertains to a system.

Diptesh Das et al (2019)

Sparse high-order interaction model with rejection option (SHIMR) is an interpretable machine learning model that we propose for medical diagnosis. A long rule (i.e., conjunction of many intervals) is utilized by a decision tree to convey the diagnosis to a patient, whereas SHIMR

employs a weighted sum of short rules. By employing proteomics data from the Alzheimer's Disease Neuroimaging Initiative (ADNI) dataset encompassing 151 subjects, it is demonstrated that SHIMR exhibits comparable accuracy to alternative non-interpretable methods (SN = 0.84 ± 0.1; SP = 0.69 ± 0.15; AUC = 0.86 ± 0.09). SHIMR is equipped with a clinical function that prevents it from rendering erroneous diagnoses when it lacks sufficient confidence. This enables physicians to select more precise pathologies, albeit those that are more invasive and/or expensive. Incorporating a rejection option into SHIMR facilitates the development of a framework for cost-effective multistage diagnosis. By employing a baseline concentration of cerebrospinal fluid (CSF) and plasma proteins obtained from a standardized cohort of 141 subjects, SHIMR has demonstrated efficacy in the development of a cost-effective Alzheimer's disease (AD) pathology that is tailored to individual patients. Therefore, due to its interpretability, dependability, and capacity to develop a patient-specific multistage cost-effective diagnosis framework, SHIMR has the potential to become an essential instrument in the era of precision medicine. This would satisfy the needs of both healthcare professionals and patients, while also alleviating the substantial financial strain associated with medical diagnosis.

III. HUMAN-CENTRIC IN MEDICAL AI SYSTEMS

In the rapidly advancing landscape of healthcare, the integration of Artificial Intelligence (AI) introduces unprecedented possibilities, reshaping the way medical systems operate. At the heart of this transformative journey lies the principle of human-centric design, a philosophy that transcends the traditional bounds of technology by prioritizing the human experience in the development and implementation of AI-driven solutions. In the context of medical AI systems, the notion of human-centric design takes on profound significance, as it seeks not only to enhance the efficiency and precision of diagnostics and treatments but also to ensure that these advancements are imbued with compassion, empathy, and a deep understanding of the nuanced human condition. This introduction unravels the essence of human-centric design in the realm of medical AI, exploring how it revolutionizes healthcare by aligning technological progress with the fundamental aspects of patient care, thus ushering in an era where innovation and compassion coalesce to redefine the future of medical practices.

3.1. Patient-Centered Approach in Medical AI

In the rapidly advancing landscape of healthcare, the adoption of Artificial Intelligence (AI) has emerged as a transformative force, promising to enhance diagnostics, treatment, and overall

patient care. Central to this technological evolution is the imperative of adopting a patient-centered approach, a philosophy that places the individual at the core of medical decision-making and leverages AI to tailor healthcare solutions to meet the unique needs, preferences, and circumstances of each patient. As we navigate the intricate intersection of medicine and AI, the patient-centered approach becomes not just a guiding principle but a revolutionary paradigm that redefines the dynamics between healthcare providers, technology, and the individuals seeking care.

3.2. Empathy as a Foundation for AI in Healthcare

In the rapidly advancing landscape of healthcare, the integration of Artificial Intelligence (AI) stands poised to redefine the way we approach diagnostics, treatment, and patient care. At the heart of this technological revolution lies a profound shift—one where the concept of empathy emerges as a foundational pillar for the development and deployment of AI in healthcare. Far beyond being a mere buzzword, empathy becomes the compass guiding the ethical, compassionate, and patient-centric evolution of AI applications in the medical realm.

The application of AI in healthcare promises unprecedented capabilities, from predictive analytics for early disease detection to personalized treatment plans based on intricate patient data. However, the transformative potential of AI can only be fully realized when it is rooted in empathy—a quality traditionally associated with human interactions but increasingly recognized as indispensable in the realm of artificial intelligence. In the context of healthcare, empathy becomes the conduit through which technology not only delivers accurate and efficient outcomes but also fosters a profound understanding of the human experience, making the patient's well-being its ultimate priority.

3.3. Understanding User Perspectives in Medical AI

In the dynamic landscape of healthcare, the integration of Artificial Intelligence (AI) introduces a transformative wave, promising unparalleled advancements in diagnostics, treatment, and patient care. However, at the heart of this technological revolution lies a critical consideration—how do users, including healthcare professionals, patients, and other stakeholders, perceive and interact with medical AI systems? The success of these cutting-edge technologies hinges not only on their technical prowess but also on their ability to align seamlessly with the diverse perspectives, needs, and expectations of those who engage with them.

At the forefront of this intersection between technology and healthcare is the imperative to understand user perspectives in the realm of Medical AI. Healthcare professionals, who are pivotal decision-makers in patient care, navigate a landscape where AI augments their expertise. Understanding how these professionals perceive and integrate AI insights into their clinical workflow is essential. Are they viewing AI as a supportive tool, a second opinion, or a revolutionary force challenging traditional practices? The answers to these questions shape the acceptance and efficacy of AI in the medical community.

3.4. Inclusive Design for Diverse Healthcare Needs

In the intricate tapestry of healthcare, the concept of inclusive design has emerged as a beacon, illuminating a path towards equitable, accessible, and person-centered medical solutions. The essence of inclusive design lies in recognizing the rich diversity of human experiences, needs, and capabilities, and weaving this understanding into the very fabric of healthcare practices, services, and technologies. It transcends the conventional boundaries of healthcare design, embracing a holistic approach that caters to individuals of all abilities, backgrounds, and conditions. Inclusive design not only seeks to break down physical barriers but also addresses the nuanced complexities of healthcare delivery, ensuring that every person, irrespective of their unique circumstances, has access to empathetic and effective healthcare solutions. In this exploration, we navigate the realm of inclusive design within healthcare, unveiling its transformative potential to nurture holistic well-being and create a healthcare landscape that is truly universal in its embrace of diversity.

4.5. User Feedback Integration in AI Medical Systems

The integration of Artificial Intelligence (AI) into medical systems represents a revolutionary leap forward in the quest for more accurate diagnostics and personalized treatment plans. However, the success of these AI applications is not solely dependent on their algorithmic prowess but equally on their seamless integration with the end-users - the healthcare professionals and patients. In this dynamic interplay, user feedback emerges as a linchpin, providing a crucial feedback loop that refines, validates, and humanizes AI-driven medical systems. This iterative process not only ensures the continuous improvement of algorithms but also fosters a sense of trust, transparency, and patient-centric care in the rapidly evolving landscape of AI in healthcare.

INTERNATIONAL JOURNAL OF APPLIED SCIENCE ENGINEERING AND MANAGEMENT
ISSN 2454-9940
www.ijasem.org
Vol 15, Issue 4, 2021
https://zenodo.org/records/15835370

At the core of user feedback integration is the recognition that healthcare professionals are not merely recipients of AI-generated insights; they are active collaborators in the decision-making process. AI algorithms, despite their sophistication, may not fully capture the intricacies of clinical nuances or the diverse array of patient presentations. Thus, healthcare providers play a pivotal role in fine-tuning these systems by offering invaluable insights derived from their practical experiences. The iterative nature of user feedback allows AI algorithms to evolve in tandem with real-world scenarios, adapting to the ever-changing dynamics of medical practice.

Moreover, user feedback serves as a bridge between the language of algorithms and the intuition of healthcare professionals. Explainable AI becomes a cornerstone in this context, providing a transparent framework that enables users to comprehend the rationale behind AI-generated recommendations. As healthcare professionals provide feedback on the accuracy, relevance, and usability of AI outputs, the system becomes more adept at aligning with the decision-making processes of its human counterparts. This alignment is not just a technical requirement but a strategic move towards building trust and acceptance of AI in the medical community.

IV. ETHICAL CONSIDERATIONS IN EXPLAINABLE AI FOR MEDICAL DIAGNOSIS

In the rapidly advancing landscape of artificial intelligence (AI), particularly in the realm of medical diagnosis, the integration of Explainable AI (XAI) has become paramount. As the capabilities of AI systems to analyze complex medical data and provide diagnostic insights continue to evolve, so does the importance of ethical considerations in the development and deployment of these technologies. The ethical dimension gains particular significance when considering the potential impact on patients' lives and the responsibility of healthcare professionals. Ensuring transparency, accountability, and the ability to interpret AI-generated recommendations are crucial aspects of ethical AI in medical diagnosis. This introduction sets the stage for a deeper exploration of the ethical considerations surrounding Explainable AI, delving into its implications for patient trust, decision-making processes, and the broader societal impact of these groundbreaking technologies in healthcare.

4.1 Informed Consent and Patient Autonomy in Explainable AI for Medical Diagnosis

In the rapidly advancing field of medical diagnosis, the integration of Explainable Artificial Intelligence (XAI) has become paramount, raising crucial ethical considerations such as informed consent and patient autonomy. Informed consent serves as the foundation for ethical

medical practices, ensuring that patients are adequately informed about the potential risks, benefits, and alternatives associated with a particular medical procedure or intervention. As this concept extends to the realm of XAI in medical diagnosis, it becomes imperative to establish transparent communication channels between healthcare providers, AI developers, and patients. This transparency is essential in elucidating the intricacies of AI algorithms, promoting a comprehensive understanding of how these technologies influence diagnostic decision-making.

Patient autonomy, a cornerstone of modern medical ethics, emphasizes an individual's right to make informed decisions about their healthcare. In the context of XAI for medical diagnosis, patient autonomy takes on a new dimension, as individuals need to navigate the complexities of AI-driven decision support systems. Ensuring that patients are not only aware of the AI's role in diagnosis but also comprehend the implications of its recommendations is critical. Healthcare professionals must engage in open and collaborative discussions with patients, fostering an environment where individuals feel empowered to express their preferences, concerns, and values regarding the use of AI in their medical care.

4.2 Transparency in Decision-Making Processes in Explainable AI for Medical Diagnosis

Transparency in decision-making plays a pivotal role in the domain of Explainable AI (XAI) for medical diagnosis processes, offering a profound impact on the integration and acceptance of artificial intelligence in healthcare. As we navigate the complex landscape of healthcare, where critical decisions can be a matter of life and death, understanding the rationale behind AI-driven diagnoses becomes paramount. Transparent AI models provide a clear window into the decision-making process, enabling healthcare professionals and patients to comprehend the underlying mechanisms and trust the recommendations provided. This transparency fosters a symbiotic relationship between AI and its users, instilling confidence in the technology and facilitating more informed decision-making.

4.3 Minimizing Bias and Fairness Concerns in Explainable AI for Medical Diagnosis

In the realm of medical diagnosis, the integration of Explainable Artificial Intelligence (XAI) is both a promising and challenging endeavor. As we navigate the intricate landscape of healthcare, it becomes imperative to not only maximize the accuracy and efficiency of AI-driven diagnostic tools but also to address the pressing issues of bias and fairness. Minimizing bias in XAI for medical diagnosis is a multifaceted task that requires meticulous attention to data sources, model development, and interpretability.

INTERNATIONAL JOURNAL OF APPLIED SCIENCE ENGINEERING AND MANAGEMENT
ISSN 2454-9940
www.ijasem.org
Vol 15, Issue 4, 2021
https://zenodo.org/records/15835370

5.4 Data Privacy and Security in Medical AI

In the ever-evolving landscape of healthcare, the intersection of artificial intelligence (AI) and medical data presents unprecedented opportunities and challenges, with data privacy and security at the forefront of concerns. As healthcare providers increasingly leverage AI technologies to enhance diagnostics, treatment planning, and patient care, the sensitive nature of medical data intensifies the need for robust privacy measures. The vast amounts of personal health information generated and processed by AI systems pose significant risks if not adequately protected.

Ensuring data privacy in medical AI involves safeguarding patient information from unauthorized access, use, or disclosure. Stringent protocols must be in place to authenticate users and restrict access to only those with legitimate reasons to interact with the data. Encryption plays a crucial role in securing medical information during transmission and storage, preventing unauthorized interception or compromise. Moreover, adopting a privacy-by-design approach entails integrating privacy features into AI systems from their inception, promoting a proactive rather than reactive stance in safeguarding patient data.

One critical consideration in the realm of medical AI is the anonymization and de-identification of patient data. While AI relies on large datasets for training and validation, these datasets must be carefully curated to remove personally identifiable information. Striking a delicate balance between preserving data utility for AI algorithms and protecting patient privacy is a persistent challenge. Techniques like differential privacy, which adds controlled noise to the data, can help mitigate the risk of re-identification while maintaining the overall integrity of the dataset.

In addition to privacy concerns, the security of medical AI systems is paramount in preventing cyber threats and ensuring the integrity of patient data. The interconnected nature of healthcare networks and the proliferation of internet-connected medical devices make them potential targets for malicious actors. Robust cybersecurity measures, including regular software updates, network monitoring, and intrusion detection systems, are imperative to fortify the defense against cyber threats.

VI. CONLCUSION

This research underscores the vital role of Explainable Artificial Intelligence (XAI) in advancing medical diagnostics, highlighting the transformative impact of transparency and interpretability

on the effectiveness and trustworthiness of AI models. As AI continues to play a pivotal role in healthcare, the study emphasizes that the inherent opacity of black-box models can hinder their adoption and limit their potential. By investigating the integration of interpretability into machine learning models, this research demonstrates that XAI can bridge the gap between complex AI systems and healthcare professionals, fostering a more informed and confident decision-making process.

Furthermore, the research sheds light on the ethical, regulatory, and patient-centered challenges associated with implementing XAI in medical diagnostics, calling for a balance between technical performance and responsible AI deployment. The insights gained from evaluating XAI's performance compared to conventional models offer valuable guidance for the creation of best practices and standards, promoting the responsible use of transparent AI systems in healthcare.

In conclusion, this study contributes to the development of frameworks that support the ethical and transparent integration of XAI, ensuring that AI technologies in medical diagnosis not only enhance clinical outcomes but also prioritize trust, accountability, and patient well-being.

REFERENCES

1. G. Valdes, J. M. Luna, E. Eaton, C. B. Simone, L. H. Ungar, and T. D. Solberg, "MediBoost: a patient stratification tool for interpretable decision making in the era of precision medicine," Scientific reports, vol. 6, no. 1, pp. 1-8, 2016.
2. H. Lakkaraju, E. Kamar, R. Caruana, and J. Leskovec, "Interpretable & explorable approximations of black box models," arXiv preprint arXiv:1707.01154, 2017.
3. A.B.Arrieta et al., "Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI," Information fusion, vol. 58, pp. 82-115, 2020.
4. A.B.Arrieta et al., "Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI," Information fusion, vol. 58, pp. 82-115, 2020.
5. A.B.Tosun, F. Pullara, M. J. Becich, D. Taylor, S. C. Chennubhotla, and J. L. Fine, "Histomapr™: An explainable ai (xai) platform for computational pathology solutions," in Artificial Intelligence and Machine Learning for Digital Pathology: Springer, 2020, pp. 204-227.
6. A.Das and P. Rad, "Opportunities and challenges in explainable artificial intelligence (xai): A survey," arXiv preprint arXiv:2006.11371, 2020.

7. A.Gomolin, E. Netchiporouk, R. Gniadecki, and I. V. Litvinov, "Artificial intelligence applications in dermatology: where do we stand?," Frontiers in medicine, vol. 7, p. 100, 2020.

8. A.Gupta, A. Anpalagan, L. Guan, and A. S. Khwaja, "Deep learning for object detection and scene perception in self-driving cars: Survey, challenges, and open issues," Array, vol. 10, p. 100057, 2021.

9. A.Gupta, A. Anpalagan, L. Guan, and A. S. Khwaja, "Deep learning for object detection and scene perception in self-driving cars: Survey, challenges, and open issues," Array, vol. 10, p. 100057, 2021.

10. A.Holzinger, G. Langs, H. Denk, K. Zatloukal, and H. Müller, "Causability and explainability of artificial intelligence in medicine," Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, vol. 9, no. 4, p. e1312, 2019.

11. A.Khamparia et al., "Diagnosis of Breast Cancer Based on Modern Mammography using Hybrid Transfer Learning," Multidimensional Systems and Signal Processing, 2020.

12. A.Khamparia et al., "Diagnosis of Breast Cancer Based on Modern Mammography using Hybrid Transfer Learning," Multidimensional Systems and Signal Processing, 2020.

13. A.Polino, R. Pascanu, and D. Alistarh, "Model compression via distillation and quantization," arXiv preprint arXiv:1802.05668, 2018.

14. A.Rajkomar et al., "Scalable and accurate deep learning with electronic health records," NPJ Digital Medicine, vol. 1, no. 1, pp. 1-10, 2018.

15. A.Rajkomar et al., "Scalable and accurate deep learning with electronic health records," NPJ Digital Medicine, vol. 1, no. 1, pp. 1-10, 2018.

16. A.Rajkomar, J. Dean, and I. Kohane, "Machine learning in medicine," New England Journal of Medicine, vol. 380, no. 14, pp. 1347-1358, 2019.

17. Adadi, A., & Berrada, M. (2018). Peeking inside the black-box: a survey on explainable artificial intelligence (XAI). IEEE access, 6, 52138-52160.

18. Alicioglu, G., & Sun, B. (2022). A survey of visual analytics for Explainable Artificial Intelligence methods. Computers & Graphics, 102, 502-520.

19. Angelov, P. P., Soares, E. A., Jiang, R., Arnold, N. I., & Atkinson, P. M. (2021). Explainable artificial intelligence: an analytical review. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 11(5), e1424.

20. Ayano, Y. M., Schwenker, F., Dufera, B. D., & Debelee, T. G. (2022). Interpretable machine learning techniques in ECG-based heart disease classification: a systematic review. Diagnostics, 13(1), 111.

21. B.Letham, C. Rudin, T. H. McCormick, and D. Madigan, "Interpretable classifiers using rules and bayesian analysis: Building a better stroke prediction model," The Annals of Applied Statistics, vol. 9, no. 3, pp. 1350-1371, 2015.

22. B.Mahbooba, M. Timilsina, R. Sahal, and M. Serrano, "Explainable artificial intelligence (xai) to enhance trust management in intrusion detection systems using decision tree model," Complexity, vol. 2021, 2021.

23. B.Mahbooba, M. Timilsina, R. Sahal, and M. Serrano, "Explainable artificial intelligence (xai) to enhance trust management in intrusion detection systems using decision tree model," Complexity, vol. 2021, 2021.

24. B.N.Patro, M. Lunayach, S. Patel, and V. P. Namboodiri, "U-cam: Visual explanation using uncertainty based class activation maps," presented at the Proceedings of the IEEE/CVF International Conference on Computer Vision, 2019, 2019.

25. Banegas-Luna, A. J., Peña-García, J., Iftene, A., Guadagni, F., Ferroni, P., Scarpato, N., ...& Pérez-Sánchez, H. (2021). Towards the interpretability of machine learning predictions for medical applications targeting personalised therapies: A cancer case survey. International Journal of Molecular Sciences, 22(9), 4394.

26. C.Molnar, G. Casalicchio, and B. Bischl, "Quantifying interpretability of arbitrary machine learning models through functional decomposition," 2019.

27. C.Roggeman, W. Fias, and T. Verguts, "Salience maps in parietal cortex: imaging and computational modeling," Neuroimage, vol. 52, no. 3, pp. 1005-1014, 2010.

28. C.Sudlow et al., "UK biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age," PLoS medicine, vol. 12, no. 3, p. e1001779, 2015.

29. C.Xiao, T. Ma, A. B. Dieng, D. M. Blei, and F. Wang, "Readmission prediction via deep contextual embedding of clinical concepts," PloS one, vol. 13, no. 4, p. e0195024, 2018.