Healthcare Trust Evolution with Explainable Artificial Intelligence

Subjects: Computer Science, Artificial Intelligence

Contributor: Pummy Dhiman , Anupam Bonkra , Amandeep Kaur , Yonis Gulzar , Yasir Hamid , Mohammad Shuaib Mir , Arjumand Bano Soomro , Osman Elwasila

The developments in IoT, big data, fog and edge networks, and AI technologies have had a profound impact on a number of industries, including medical. The use of artificial intelligence (AI) for therapeutic purposes has been hampered by its inexplicability. Explainable Artificial Intelligence (XAI), a revolutionary movement, has arisen to solve this constraint. By using decision-making and prediction outputs, XAI seeks to improve the explicability of standard AI models.

Explainable Artificial Intelligence XAI healthcare neural network

1. Introduction

New applications for artificial intelligence (AI) have been generated by recent developments in machine learning (ML), the Internet of Things (IoT) [1], big data, and assisted fog and edge networks, which offer several benefits to many different sectors. However, many of these systems struggle to justify their own decisions and actions to those who are not computers. The emphasis on explanation, according to some AI researchers, is incorrect, unrealistic, and perhaps unnecessary for all applications of AI [2]. The authors of [3] proposed the phrase "explainable AI" to highlight a training system developed for the US Army's capacity to justify its automation choices. The Explainable Artificial Intelligence (XAI) program was started in 2017 via the Defense Advanced Research Projects Agency (DARPA) [3] to construct methods for comprehending intelligent systems. DARPA refers to a collection of methods as XAI to describe how they develop explainable models that, when combined with successful explanation procedures, allow end-users to grasp, correctly trust, and efficiently manage the next generation of AI systems.

In keeping with the perception of keeping humans in the loop, XAI aims to make it simpler for people to comprehend opaque AI systems so they may use these tools to help with their work more successfully. Recent applications of XAI include those in the military, healthcare, law, and transportation. In addition to software engineering, socially sensitive industries, including edification, law enforcement and forensics, healthcare, and agriculture, are also seeing an increase in the usage of ML and deep learning feature extraction and segmentation techniques [4][5]. This makes using them considerably more difficult, especially given that many people who are dubious about the future of these technologies just do not know how they operate.

AI has the potential to help with a number of critical issues in the medical industry. The fields of computerized diagnosis, prospects, drug development, and testing have made significant strides in recent years.

Within this particular framework, the importance of medical intervention and the extensive pool of information obtained from diverse origins, including electronic health records, biosensors, molecular data, and medical imaging, assume crucial functions in propelling healthcare forward and tackling pressing concerns within the medical sector. Establishing treatments, decisions, and medical procedures specifically for individual patients is one of the objectives of AI in medicine. The current status of artificial intelligence in medicine, however, has been described as heavy on promise and fairly light on evidence and proof. Multiple AI-based methods have succeeded in real-world contexts for the diagnosis of forearm sprains, histopathological prostate cancer lesions [4], very small gastrointestinal abnormalities, and neonatal cataracts. But in actual clinical situations, a variety of the systems that encompass them have been demonstrated to be on par with or even better than those used by specialists in experimental studies and have large false-positive rates. By improving the transparency and interpretability of AI-driven medical applications, Explainable Artificial Intelligence has the potential to completely transform the healthcare system. Healthcare practitioners must comprehend how AI models make judgments in key areas, including diagnosis, therapy suggestions, and patient care.

Clinical decision making is more informed and confident thanks to XAI, which gives physicians insights into the thinking underlying AI forecasts. Doctors may ensure patient safety by identifying potential biases, confirming the model’s correctness, and offering interpretable explanations. Additionally, XAI promotes the acceptance of AI technology in the healthcare industry, allaying worries about the “black box” nature of AI models. By clearly communicating diagnoses and treatment plans, transparent AI systems can improve regulatory compliance, resolve ethical concerns, and increase patient participation.

Healthcare professionals may fully utilize AI with XAI while still maintaining human supervision and responsibility. In the end, this collaboration between AI and human knowledge promises to provide more individualized and accurate healthcare services, enhance patient outcomes, and influence the course of medical research.

There are some important taxonomies of XAI that exist to show the antithesis of some AI, ML, and particularly DL models’ black-box characteristics. The following terms are distinguished in **Figure 1**.

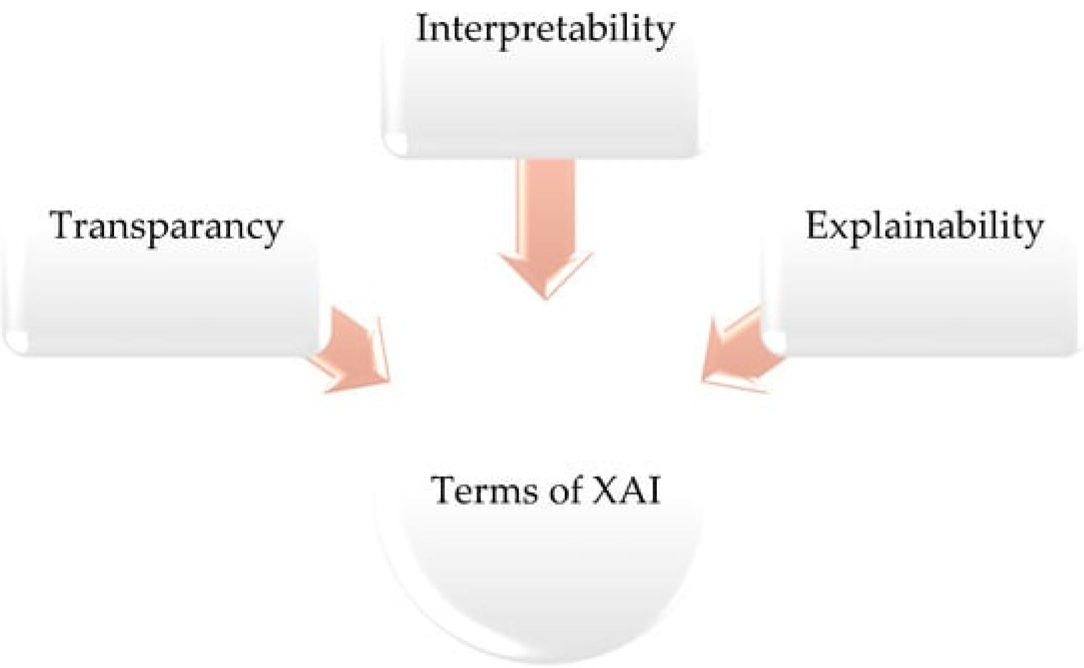

Figure 1. Terms of XAI.

- Transparency: A sculpture is said to be translucent if it has the capacity to make sense on its own. Thus, lucidity is the contradiction of a black box [5].

- Interpretability: The term “interpretability” describes the capacity to comprehend and articulate how a complicated system, such as a machine learning model or an algorithm, makes decisions. It entails obtaining an understanding of the variables that affect the system’s outputs and how it generates its conclusions [6]. Explainability is an area within the realm of interpretability, and it is closely linked to the notion that explanations serve as a means of connecting human users with artificial intelligence systems. The process encompasses the categorization of artificial intelligence that is both accurate and comprehensible to human beings [6].

According to the authors of [7], XAI is required within any of the following scenarios:

- Where in the interest of fairness and to help customers make an informed decision, an explanation is necessary.

- Where the consequences of a wrong AI decision can be very far-reaching (such as recommending surgery that is unnecessary).

- In cases where a mistake results in unnecessary financial costs, health risks, and trauma, such as malignant tumor misclassification.

- Where domain experts or subject matter experts must validate a novel hypothesis generated by the AI.

- The EU’s General Data Protection Regulation (GDPR) [8] gives consumers the right to explanations when data are accessed through an automated mechanism.

2. Taxonomy of XAI

Translucent Model

The authors of [5] provide a list of a few well-known transparent models, including Fuzzy systems, decision trees, principal learning, and K-nearest neighbors (KNN). Typically, these models yield decisions that are unambiguous; however, it should be noted that mere transparency does not guarantee that a given concept will be easily comprehensible, as illustrated in **Figure 2**.

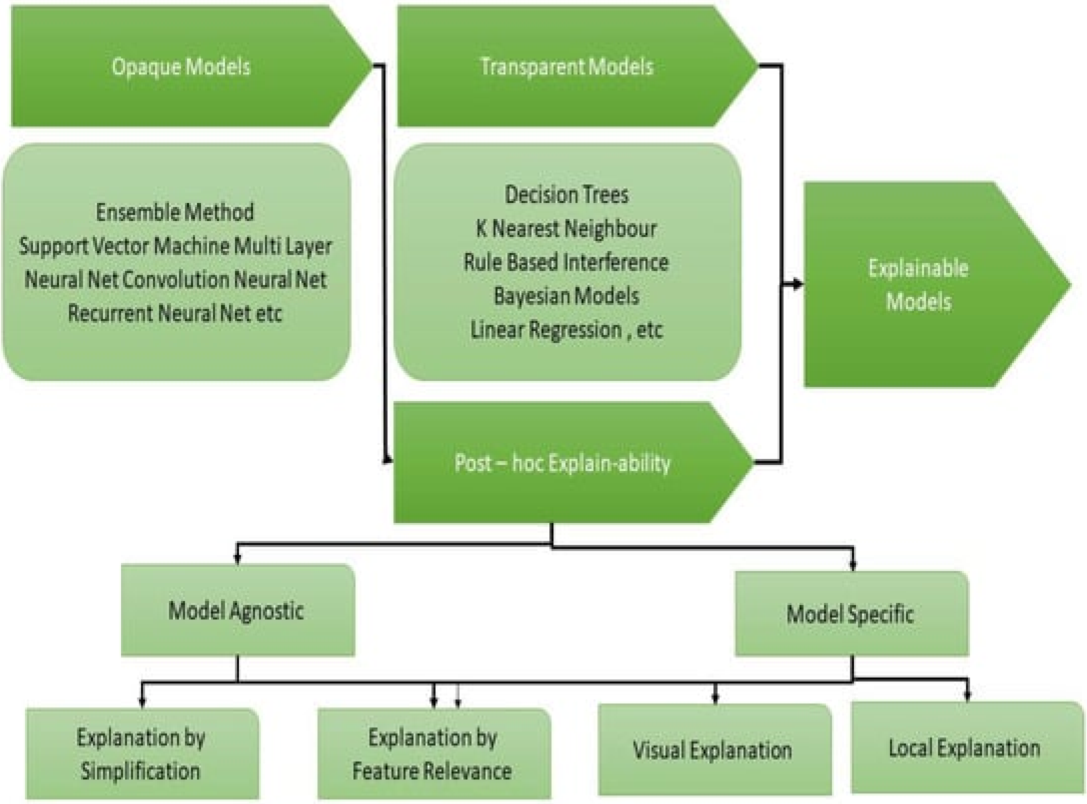

Figure 2. Taxonomy of XAI.

Opaque Models

Black-box or opaque models are those used in machine learning or artificial intelligence that lack transparency and are challenging for humans to understand. These models are difficult to comprehend or describe in a way that is intelligible to humans since they base their choices on intricate relationships between the input characteristics. Transparent models encompass various types, such as rule-based models, decision trees, and linear regression. The comprehension of predictions made by opaque models can be achieved through the utilization of transparent models or procedures, such as post hoc explanations like LIME (Local Interpretable Model-Agnostic Explanations) or SHAP (Shapley Additive Explanations). These approaches provide a balance between accuracy and interpretability.

Model-Agnostic Techniques

Explainable Artificial Intelligence techniques that are model-agnostic are meant to make machine learning model decisions more understandable and interpretable without relying on the specifics of the model’s internal architecture. These methods attempt to be applicable to many models and may be used in a variety of contexts, making them extremely adaptive and versatile. Establishing a clear and understandable link between the input characteristics (data) and the model’s output (predictions) is the main goal of model-agnostic XAI. These methods do not need access to the model’s internal parameters, intermediate representations, or procedures. They are “agnostic” to the model’s underlying complexity since they only focus on the input–output connection [9].

Model-Specific Techniques

The purpose of model-specific XAI techniques is to enable interpretability and transparency for a particular model or a subset of models by leveraging the internal architecture, knowledge, and features of a given model. Model-specific XAI focuses on maximizing interpretability for a specified model, as opposed to model-agnostic techniques, which are flexible across multiple models. These methods make use of the model’s built-in structure, such as neural network attention processes or decision tree decision rules, to produce explanations that are consistent with the model’s knowledge. Model-specific XAI strives to give more precise and educational insights for that particular model or a specific group of related models by customizing explanations to the model’s complexities [10].

Simplification of Enlightenment

By using a rough framework, it is possible to find different ways in which computer versions could be made that help explain the prophecy being looked into. To show a more complicated structure, a regression analysis or decision tree can be used as it is based on the model’s predictions [11].

Relevance of Explanation by Feature

This concept is comparable to generalization. After considering all potential combinations, this kind of XAI approach evaluates an attribute according to its predicted marginal contribution to the model’s decision [10][11].

Graphic Explanation

This particular XAI strategy is built around visualization. In light of this, it is possible to use the family of data visualization techniques to construe the prophecy or decision taken in light of the input data [12].

Narrow Explanation

Narrow explanations shed light on the behavior of the model. They carry out the model’s decision-making process in a constrained space centered on an interesting case [13]

Healthcare has advanced significantly, with more innovative research and a shift towards healthcare 5.0. In order to transition into the new era of smart disease control and detection, virtual care, smart health management, smart monitoring, and decision making and decision explanation, the healthcare industry is in the midst of a paradigm shift. The purpose of XAI is to offer machine learning and deep learning algorithms that perform better while being understandable, making it easier for users to trust, comprehend, accept, and use the system [8]. Several studies give insight into how XAI is utilized in healthcare [14].

Exploring XAI in the healthcare industry is essential for ensuring ethical, transparent, and responsible AI use, which will improve patient care and boost confidence in AI-driven medical decisions. Furthermore, it aids in reducing the risk of diagnostic errors and ensures compliance with healthcare standards. With this in mind, the objective of this study is to evaluate a prior research-based analysis in order to obtain insight into the work performed and opportunities given by AI advancement, and the explainability feature of AI in the healthcare sector.

3. Healthcare Trust Evolution with Explainable Artificial Intelligence

Explainable AI (XAI) provides users with an explanation of why a method produces a particular result. The outcome can then be understood in a particular context. A crucial application of XAI is in clinical decision support systems (CDSSs) [15]. These methods aid doctors in making judgments in the clinic but, because of their complexity, may lead to issues with under- or overreliance. Practitioners will be better able to make decisions that, in some circumstances, could save lives as the practitioners are given explanations for the processes used to arrive at recommendations. The demand for XAI in CDSS and therapeutic industries in general has arisen as a result of the necessity for principled and equitable decision making as well as the actuality that AI skilled in chronological data might perpetuate pre-existing behaviors and prejudices that should be exposed [15].

- Medical Imaging and Diagnosis

Medical imaging and diagnosis often benefit from the use of these techniques. XAI can provide valuable insights into the decision-making process and model behavior, setting it apart from other artificial intelligence methods, such as deep learning. Current advancements have placed significant emphasis on the utilization of Explainable Artificial Intelligence (XAI) in the domains of surgical procedures and medical diagnoses. For instance, Explainable Artificial Intelligence (XAI) has the potential to improve the comprehensibility and transparency of medical image analysis [16], specifically in the context of breast cancer screening. This statement pertains to the issue presented by the lack of transparency in AI systems [17][18].

- Chronic Disease Detection

Chronic disease management poses a continuous healthcare burden, particularly in places such as India, where diseases like diabetes and asthma prevail. Artificial intelligence (AI), including explainable AI (XAI), assumes a crucial role in facilitating the coordination of therapies for chronic illnesses. It offers valuable insights into the physical and mental well-being of individuals, thus assisting patients in efficiently managing their health [19][20].

- COVID-19 Diagnosis

During the COVID-19 pandemic, AI, including XAI, has significantly improved diagnostic accuracy. For instance, chest radiography, a critical screening tool, is employed to identify COVID-19 cases, particularly when traditional methods like polymerase chain reaction fall short. XAI contributes by elucidating the factors influencing COVID-19 detection, thereby enhancing the screening process [21].

- Global Health Goals

Global health objectives can be effectively pursued through the utilization of digital health technologies, which encompass artificial intelligence (AI). These technologies are in line with several initiatives proposed by the United Nations and play a significant role in advancing global health goals. These technologies utilize patient data, environmental information, and connectivity to enhance healthcare delivery, demonstrating significant value during times of emergencies and disease epidemics [22].

- Pain Assessment

The assessment of pain in patients has been significantly enhanced by advancements in Artificial Intelligence (AI), particularly with a focus on Explainable AI (XAI). Artificial intelligence (AI) and machine learning (ML) models are capable of analyzing facial expressions, which can serve as reliable indicators of pain and suffering. This technical application exhibits potential in the field of healthcare, specifically in the assessment of pain levels among patients [23].

- Biometric Signal Analysis

The study presented in [24] employed a modified bidirectional LSTM network with Bayesian optimization for the automated detection and classification of ECG signals. This system has the capability to improve accuracy and has practical implications for effectively addressing issues associated with categorizing biometric data. A Bayesian optimization-modified bidirectional LSTM (BiLSTM) network is employed in this inquiry to provide an automated ECG detection and classification approach. The two hyperparameters of the BiLSTM network that are optimized using Bayesian techniques are the initial learning rate and the total number of hidden layers. When categorizing five ECG signals in the MIT-BIH arrhythmia database, the improved network’s accuracy rises to 99.00%, an increase of 0.86% from its pre-optimization level. The potential practical relevance of the presented approach to further quasi-periodic biometric signal-based categorization problems might be considered in future research [24].

- Stroke Recognition

Stroke, a deadly medical ailment that occurs when the brain's blood supply is cut off, is the subject of much investigation. Brain cells die if blood flow is abruptly interrupted. In [25], the author's study of numerous research methods begins with artificial intelligence's interpretability and explainability. Two explainable AI-based cardiac disease prediction experiments are compared. This comparison can help AI beginners choose the best techniques [26]. In another study, deep learning models in electronic health records (EHRs) are examined, along with interpretability in medical AI systems [27].

Overall, medical professionals employ AI to speed up and improve several healthcare processes. These include forecasting, risk assessment, diagnosis, and decision making. They finish by carefully studying medical images to find hidden anomalies and patterns that humans cannot see. Many healthcare professionals have already integrated AI into their workflows, but doctors and patients often become frustrated with its operations, especially when making important judgments. This industry's demand for explainable AI (XAI) drives its adoption. Complex AI suggestions like surgical procedures or hospital hospitalizations need explanations for patients and physicians [28]. Examining the revised contributions revealed substantial implications for academics and professionals, prompting an exploration of methodological aspects to promote medical AI applications. The philosophical foundations and contemporary uses of 17 Explainable Artificial Intelligence (XAI) techniques in healthcare were examined. Finally, legislators were given goals and directions for building electronic healthcare systems that emphasize authenticity, ethics, and resilience. The study examined healthcare information fusion methodologies, including data fusion, feature aggregation, image analysis, decision coordination, multimodal synthesis, hybrid methods, and temporal considerations [29]. AI has been used to manage healthcare services, forecast medical outcomes, enhance professional judgment, and analyze patient data and diseases. Despite their success, AI models are still ignored since they are seen as "black boxes". Lack of trust is the largest barrier to their widespread usage, especially in healthcare. To ease this concern, Explainable Artificial Intelligence (XAI) has evolved. XAI improves AI model predictions by revealing their logic. XAI helps healthcare providers embrace and integrate AI systems by explaining the model's inner workings and prediction approach [30]. They thoroughly analyze healthcare scenarios involving explanation interfaces.

References

1. Bonkra, A.; Dhiman, P. IoT Security Challenges in Cloud Environment. In Proceedings of the 2021 2nd International Conference on Computational Methods in Science & Technology, Mohali, India, 17–18 December 2021; pp. 30–34.
2. Barredo Arrieta, A.; Díaz-Rodríguez, N.; Del Ser, J.; Bennetot, A.; Tabik, S.; Barbado, A.; Garcia, S.; Gil-Lopez, S.; Molina, D.; Benjamins, R.; et al. Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. Inf. Fusion 2020, 58, 82–115.

3. Van Lent, M.; Fisher, W.; Mancuso, M. An explainable artificial intelligence system for small-unit tactical behavior. In Proceedings of the Nineteenth National Conference on Artificial Intelligence, Sixteenth Conference on Innovative Applications of Artificial Intelligence, San Jose, CA, USA, 25–29 July 2004; pp. 900–907.
4. Mukhtar, M.; Bilal, M.; Rahdar, A.; Barani, M.; Arshad, R.; Behl, T.; Bungau, S. Nanomaterials for diagnosis and treatment of brain cancer: Recent updates. Chemosensors 2020, 8, 117.
5. Adadi, A.; Berrada, M. Peeking Inside the Black-Box: A Survey on Explainable Artificial Intelligence (XAI). IEEE Access 2018, 6, 52138–52160.
6. Gilpin, L.H.; Bau, D.; Yuan, B.Z.; Bajwa, A.; Specter, M.; Kagal, L. Explaining explanations: An overview of interpretability of machine learning. In Proceedings of the 5th IEEE International Conference on Data Science and Advanced Analytics, DSAA 2018, Turin, Italy, 1–3 October 2018; pp. 80–89.
7. Ahmad, M.A.; Eckert, C.; Teredesai, A. Explainable AI in Healthcare. SSRN Electron. J. 2019.
8. Sheu, R.K.; Pardeshi, M.S. A Survey on Medical Explainable AI (XAI): Recent Progress, Explainability Approach, Human Interaction and Scoring System. Sensors 2022, 22, 68.
9. Dieber, J.; Kirrane, S. Why model why? Assessing the strengths and limitations of LIME. arXiv 2012, arXiv:2012.00093.
10. Bach, S.; Binder, A.; Montavon, G.; Klauschen, F.; Müller, K.R.; Samek, W. On pixel-wise explanations for non-linear classifier decisions by layer-wise relevance propagation. PLoS ONE 2015, 10, e0130140.
11. Tritscher, J.; Ring, M.; Schlr, D.; Hettinger, L.; Hotho, A. Evaluation of Post-hoc XAI Approaches through Synthetic Tabular Data. In Foundations of Intelligent Systems, 25th International Symposium, ISMIS 2020, Graz, Austria, 23–25 September 2020; 2020; pp. 422–430.
12. Chattopadhay, A.; Sarkar, A.; Howlader, P.; Balasubramanian, V.N. Grad-CAM++: Generalized gradient-based visual explanations for deep convolutional networks. In Proceedings of the 2018 IEEE Winter Conference on Applications of Computer Vision, WACV 2018, Lake Tahoe, NV, USA, 12–15 March 2018; pp. 839–847.
13. Selvaraju, R.R.; Cogswell, M.; Das, A.; Vedantam, R.; Parikh, D.; Batra, D. Grad-CAM: Visual Explanations from Deep Networks via Gradient-Based Localization. Int. J. Comput. Vis. 2020, 128, 36–359.
14. Alsharif, A.H.; Md Salleh, N.Z.; Baharun, R.; A. Rami Hashem, E. Neuromarketing research in the last five years: A bibliometric analysis. Cogent Bus. Manag. 2021, 8, 1978620.
15. Antoniadi, A.M.; Du, Y.; Guendouz, Y.; Wei, L.; Mazo, C.; Becker, B.A.; Mooney, C. Current challenges and future opportunities for xai in machine learning-based clinical decision support

systems: A systematic review. Appl. Sci. 2021, 11, 5088.

16. Kaur, H.; Koundal, D.; Kadyan, V. Image fusion techniques: A survey. Arch. Comput. Methods Eng. 2021, 28, 4425–4447.
17. Kaushal, C.; Kaushal, K.; Singla, A. Firefly optimization-based segmentation technique to analyse medical images of breast cancer. Int. J. Comput. Math. 2021, 98, 1293–1308.
18. Naik, H.; Goradia, P.; Desai, V.; Desai, Y.; Iyyanki, M. Explainable Artificial Intelligence (XAI) for Population Health Management—An Appraisal. Eur. J. Electr. Eng. Comput. Sci. 2021, 5, 64–76.
19. Dash, S.C.; Agarwal, S.K. Incidence of chronic kidney disease in India. Nephrol. Dial. Transplant. 2006, 21, 232–233.
20. Refat, M.A.R.; Al Amin, M.; Kaushal, C.; Yeasmin, M.N.; Islam, M.K. A Comparative Analysis of Early Stage Diabetes Prediction using Machine Learning and Deep Learning Approach. In Proceedings of the 2021 6th International Conference on Signal Processing, Computing and Control (ISPCC), Solan, India, 7–9 October 2021; pp. 654–659.
21. Tiwari, S.; Kumar, S.; Guleria, K. Outbreak Trends of Coronavirus Disease-2019 in India: A Prediction. Disaster Med. Public Health Prep. 2020, 14, e33–e38.
22. Pai, R.R.; Alathur, S. Bibliometric Analysis and Methodological Review of Mobile Health Services and Applications in India. Int. J. Med. Inform. 2021, 145, 104330.
23. Madanu, R.; Abbod, M.F.; Hsiao, F.-J.; Chen, W.-T.; Shieh, J.-S. Explainable AI (XAI) Applied in Machine Learning for Pain Modeling: A Review. Technologies 2022, 10, 74.
24. Li, H.; Lin, Z.; An, Z.; Zuo, S.; Zhu, W.; Zhang, Z.; Mu, Y.; Cao, L.; Garcia, J.D.P. Automatic electrocardiogram detection and classification using bidirectional long short-term memory network improved by Bayesian optimization. Biomed. Signal Process. Control 2022, 73, 103424.
25. Merna Said, A.S.; Omaer, Y.; Safwat, S. Explainable Artificial Intelligence Powered Model for Explainable Detection of Stroke Disease. In Proceedings of the 8th International Conference on Advanced Intelligent Systems and Informatics, Cairo, Egypt, 20–22 November 2022; pp. 211–223.
26. Tasleem Nizam, S.Z. Explainable Artificial Intelligence (XAI): Conception, Visualization and Assessment Approaches towards Amenable XAI. In Explainable Edge AI: A Futuristic Computing Perspective; Springer International Publishing: Cham, Switzerland, 2023; pp. 35–52.
27. Rajkomar, A.; Oren, E.; Chen, K.; Dai, A.M.; Hajaj, N.; Hardt, M.; Liu, P.J.; Dean, J. Scalable and accurate deep learning with electronic health records. NPJ Dig. Med. 2018, 1, 18.
28. Praveen, S.; Joshi, K. Explainable Artificial Intelligence in Health Care: How XAI Improves User Trust in High-Risk Decisions. In Explainable Edge AI: A Futuristic Computing Perspective; Springer International Publishing: Cham, Switzerland, 2022; pp. 89–99.

29. Albahri, A.S.; Duhaim, A.M.; Fadhel, M.A.; Alnoor, A.; Baqer, N.S.; Alzubaidi, L.; Deveci, M. A systematic review of trustworthy and explainable artificial intelligence in healthcare: Assessment of quality, bias risk, and data fusion. Inf. Fusion 2023, 96, 156–191.

30. Loh, H.W.; Ooi, C.P.; Seoni, S.; Barua, P.D.; Molinari, F.; Acharya, U.R. Application of explainable artificial intelligence for healthcare: A systematic review of the last decade (2011–2022). Comput. Methods Programs Biomed. 2022, 226, 107161.

Retrieved from https://encyclopedia.pub/entry/history/show/116207