Aprendizado Federado e Inteligência Artificial Explicável em Saúde Digital
Federated Learning and Explainable Artificial Intelligence in Smart Healthcare

Nauber Milicio Prestes[1], Augusto Foronda[2], Rui Tadashi Yoshino[3]

RESUMO

O objetivo principal deste artigo é conduzir uma examinação compreensível sobre as tecnologias atualmente estudadas na área da saúde, por meio da análise de artigos feitos por pesquisadores e bancas respeitáveis, com foco no uso de sistemas de Aprendizado Federado (FL) e Inteligência Artificial Explicável (XAI), que estão sendo estudadas e aplicadas por hospitais e especialistas. Este artigo discute algumas das aplicações, como gêmeos digitais (DT), registros eletrônicos de saúde (EHR) e imagiologia médica, que estão sendo usadas em pesquisas por meio de FL e XAI, assim como as vantagens e desvantagens da utilização de FL e XAI em tais aplicações. Este artigo encontrou como resultado que Fl e XAI são tecnologias modernas com a capacidade de revolucionar a experiência de pacientes e profissionais de saúde nas áreas de segurança, confiança, conforto e eficiência do sistema. Os problemas encontrados incluem incertezas em partes específicas da segurança do sistema, considerações éticas dos estudos e principalmente as limitações tecnológicas existentes.

PALAVRAS-CHAVE: Aprendizado de Máquina; Inteligência Artificial; Saúde Digital.

ABSTRACT

The primary objective of this paper is to conduct a comprehensive examination of healthcare technologies currently being researched, through the examination of respected papers and authors, focusing on the use of Federated Learning (FL) and Explainable AI (XAI) systems, which are currently being researched and adopted by some healthcare systems and experts. This paper discusses a few applications, such as digital twins (DT), Electronic Health Records (EHR) and Medical Imaging, which are currently being applied in studies and hospitals with the use of FL and XAI, as well as the pros and cons of using FL and XAI to these applications. The findings of this review conclude that Fl and XAI are new technologies that can revolutionize the experience of patients and healthcare professionals in terms of security, trust, comfort and system efficiency, the issues standing in the way include problems regarding specific security aspects of these applications, ethical considerations of studies and mainly the technological limitations existing today.

KEYWORDS: Machine Learning; Artificial Intelligence; Smart Healthcare.

INTRODUCTION

Technologies such as artificial intelligence (AI) have advanced at a large rate in recent years, allowing for complex and effective applications of machine learning (ML) models, blockchain, cloud systems and more. Those applications greatly benefited healthcare systems, especially Explainable Artificial Intelligences (XAI) and Federated Learning (FL), being two of the main technologies currently studied in the area.

XAI fundamentally refers to the illustration and explanation of the output given by an AI algorithm, done by an external model or the same model, with the objective of helping

[1] Discente no curso de Engenharia Elétrica. Universidade Tecnológica Federal do Paraná - Campus Ponta Grossa, Ponta Grossa, Paraná, Brasil. E-mail: nauber@alunos.utfpr.edu.br. ID Lattes: 6330809255888605.

[2] Docente no Departamento Acadêmico de Informática. Universidade Tecnológica Federal do Paraná - Campus Ponta Grossa, Ponta Grossa, Paraná, Brasil. E-mail: foronda@utfpr.edu.br. ID Lattes: 7103296555987124.

[3] Docente no Departamento Acadêmico de Engenharia de Produção. Universidade Tecnológica Federal do Paraná - Campus Ponta Grossa, Ponta Grossa, Paraná, Brasil. E-mail: ruiyoshino@utfpr.edu.br. ID Lattes: 1374012206166960.

external users, such as patients or healthcare professionals, to understand what the model's output is and why it was given, through the use of text, images and audio that justify that output.

ML is an AI algorithm that is capable of learning and evolving based on the feedback of its output and the changes in its inputs that led to such output. In traditional ML models, the collection of data is done by gathering external users' data and transferring them to a centralized database, meaning that all of the collected data is stored in the same place, afterwards a ML model is trained with the collected data. In this scenario, the system has a hold of all of its users' data and the ML model.

FL is a decentralized ML model, where users' data is not transferred and gathered by the system to a central database. Instead, the system sends a copy of the main ML model to selected users that are of interest, the users afterwards train the model locally with their own data, and send only the trained model back to the system, the system gathers all of those different and already trained models to create a main model, the main model is given to new users continuing the learning process. In this way the system is never aware of nor collects the data of users.

FEDERATED LEARNING APPLICATIONS

Multiple studies were made where FL was directly applied to healthcare systems, in this section, a selection of those studies will be presented along with their conclusion based on their application of the technology.

DIGITAL TWINS

One of the studied applications of FL are Digital Twins (DT) systems. A DT is a series of collected data of a patient, made with the use of sensors and IoMT (Internet of Medical Things) devices that can be directly attached to the patient or be external. The collected data effectively creates a digital copy of the patient, allowing for accurate and real-time monitoring of a patient by healthcare professionals, and ML algorithms. The data handling and gathering propose great risks to patients' data, but can be mitigated with the use of FL models.

Gupta, et al. in 2021 created an anomaly detection system based on FL. They created a case use of a remote patient monitoring system (a DT) with their system and concluded that in the LSTM anomaly detection system, the use of FL prevented confidential patient data from being gathered.

ELECTRONIC HEALTH RECORD

An electronic health record (EHR) is a digital history of a patient's health record, used to automate the access of a patient's information and help healthcare professionals. ML algorithms are widely used in EHR systems, but contain severe security risks in the usage of patients' data, the use of FL can greatly increase the security and trust of such systems.

Hao, et al. in 2020 used a FL model capable of EHR analysis of patients, and concluded that its use helped in keeping patients' data safe, along with the use of noise in the learning data to prevent memorization attacks.

IMAGE PROCESSING

Image processing, sometimes called Medical Image Processing, refers to the use of image technologies in the automated diagnostics of patients. The use of FL is still experimental in this application, but studies are being made that predict better patient stay and increased security in hospitals.

Darzidehkalani, et al. in 2022 argued in detail the considerations of using FL in image processing in two articles. They introduce FL and how it handles the collection and communication of patients' data, and the challenges and benefits of such systems in medical image processing. The authors compared multiple methods of FL, weighting the pros and cons of each one, the methods were FedAvg, SWT, CWT, Ensemble Methods and CDS, with their findings concluding an increase in the identification accuracy of patients and their data security. Their findings contradict the findings of the creators of the FedAvg method, with the authors' performance being worse than the alleged by the creators when handling heterogenous data.

The authors found many problems with the methods studied when applying them to medical imaging systems, citing issues with bias, privacy, architecture and data heterogeneity, concluding that current technologies must advance further before FL models can be used by hospitals in medical image processing.

EXPLAINABLE ARTIFICIAL INTELLIGENCE APPLICATIONS

There are studies in which the use of XAI in clinics and hospitals were argued, this section describes the methods and results obtained by selected studies.

ELECTRONIC HEALTH RECORD

Chen, et al. in 2021 used their AI model called PHASE (PHysiologicAl Signal Embedding) in an EHR based dataset to predict the outcome of surgeries using psychological data, and measured the effectiveness of using the model.

Their findings were favorable to the use of PHASE, outperforming state of the art techniques being currently used.

IMAGE PROCESSING

Chetoui, et al. in 2021 published a study explaining how they made an x-ray image processing system of patients' chests to detect cases of COVID-19 or pneumonia, using GradCAM explainable models to analyze and explain the resulting image. The authors' findings concluded that the model had great efficiency and can successfully generate gradient images indicative of areas of interest from the image where the initial ML model drew its conclusion from.

Limitations were found in the explainable model developed, the main one being false positives and negatives resulting from poor quality x-ray images, since common image artifacts can be interpreted as the lung's transparency by the model. Human error such as wires, cables and medical equipment being present in the image also generate problems with the model's output.

The authors recommend the exploration of the use of XAI in different diagnostics systems, and to build upon existing systems such as their own.

CONCLUSION

This review presented technologies that are innovative to the field of smart healthcare, based on the application of FL and XAI in specific processes such as DT, EHR, and image processing. The findings of the selected studies indicates that current advancements in technologies such as AI, ML, sensors and IoT can revolutionize the experience of patients and the everyday of the healthcare professional, in terms of security, comfort and trust, the advancements are also capable of increasing process efficiency, reducing costs and improving public perception of many companies.

The applications of systems using FL and XAI remain extremely complex, partially due to the lack of standards and metrics for specific applications, as well as current technological limitations.

The studies of the applications of these technologies are mainly still in experimental phases, which limits the scope of this review. As concluded by Jung, et al. in 2023, few studies follow rigorous and defined metrics for their findings and, in the case of XAI, close to none take security concerns into consideration. There are many and diverse healthcare processes in which FL and XAI can be applied and are currently being studied, this review includes relatively few of them.

The study of the use of FL in this review indicates that FL is a useful technology already in use in some hospitals, better system optimization and security are study areas capable of improving the use of FL applications. The use of XAI can benefit from robust studies dedicated to security concerns of different models in different applications.

Acknowledgments

To UTFPR-PG and my advisors.

Conflict of Interest

There is no conflict of interest.

REFERENCES

AHMAD ABDULLAH ALJABR; KUMAR, K. Design and implementation of Internet of Medical Things (IoMT) using artificial intelligent for mobile-healthcare. Measurement: Sensors, v. 24, dez. 2022.

ALI, M. et al. Federated learning for privacy preservation in smart healthcare systems: A comprehensive survey. IEEE Journal of Biomedical and Health Informatics, v. 27, n. 2, p. 778–789, fev. 2023.

CHEN, H. et al. Forecasting adverse surgical events using self-supervised transfer learning for physiological signals. npj Digital Medicine, v. 4, 2021.

CHETOUI, M. et al. Explainable COVID-19 detection on chest X-rays using an end-to-end deep convolutional neural network architecture. Big Data and Cognitive Computing, v. 5, 2021.

DUCKWORTH, C. et al. Using explainable machine learning to characterise data drift and detect emergent health risks for emergency department admissions during COVID-19. Scientific Reports, v. 11, 2021.

ERFAN DARZIDEHKALANI; GHASEMI-RAD, M.; P.M.A. VAN OOIJEN. Federated learning in medical imaging: Part II: Methods, challenges, and considerations. Journal of the American College of Radiology, v. 19, p. 975–982, 2022a.

ERFAN DARZIDEHKALANI; GHASEMI-RAD, M.; P.M.A. VAN OOIJEN. Federated learning in medical imaging: Part I: Toward multicentral health care ecosystems. Journal of the American College of Radiology, v. 19, p. 969–974, 2022b.

JAMEELA AL-JAROODI et al. Healthcare 4.0 - managing a holistic transformation. SysCon 2022 - 16th Annual IEEE International Systems Conference, Proceedings, 2022.

JUNG, J. et al. Essential properties and explanation effectiveness of explainable artificial intelligence in healthcare: A systematic review. Heliyon, v. 9, p. e16110, 2023.

KHAN, L. U. et al. Federated learning for edge networks: Resource optimization and incentive mechanism. IEEE Communications Magazine, v. 58, p. 88–93, out. 2020.

LI, H. et al. Review on security of federated learning and its application in healthcare. Future Generation Computer Systems, v. 144, p. 271–290, 2023.

MOHAMMED, B. G.; HASAN, D. S. Smart healthcare monitoring system using IoT. International Journal of Interactive Mobile Technologies (iJIM), v. 17, p. pp. 141–152, 2023.

RAHMAN, A. et al. Federated learning-based AI approaches in smart healthcare: concepts, taxonomies, challenges and open issues. Cluster Computing, 2022.

RANI, S. et al. Federated learning for secure IoMT-applications in smart healthcare systems: A comprehensive review. Knowledge-Based Systems, v. 274, ago. 2023.

REHMAN, A. et al. A secure healthcare 5.0 system based on blockchain technology entangled with federated learning technique. Computers in Biology and Medicine, v. 150, nov. 2022.

ROY, S.; MEENA, T.; SE JUNG LIM. Demystifying supervised learning in healthcare 4.0: A new reality of transforming diagnostic medicine. Diagnostics, v. 12, n. 10, out. 2022.

SAEED HAMOOD ALSAMHI et al. Survey on Federated Learning enabling indoor navigation for industry 4.0 in B5G. Future Generation Computer Systems, v. 148, p. 250–265, nov. 2023.

SARANYA A; SUBHASHINI R. A systematic review of Explainable Artificial Intelligence models and applications: Recent developments and future trends. Decision Analytics Journal, v. 7, p. 100230, 2023.