

Protecting Personal Healthcare Record Using Blockchain & Federated Learning Technologies

Satyabrata Aich*, Nday Kabulo Sinai**, Saurabh Kumar***, Mohammed Ali****, Yu Ran Choi*, Moon-IL Joo*, Hee-Cheol Kim*****

* Institute of Digital Anti-Aging Healthcare, Inje University, South Korea

** K4-security, Seoul, South Korea

*** U1 Geographic Information System, Seoul, South Korea

**** College of Medicine, Yonsei University, South Korea

***** Department of Computer Engineering/Institute of Digital Anti-Aging Healthcare/u-HARC, Inje University, South Korea

satyabrataaich@gmail.com, kabulo.nday@gmail.com, saurabh@u1gis.com, midoabdo@yonsei.ac.kr, dbfks2007@naver.com, joomi@inje.ac.kr, heeki@inje.ac.kr

Corresponding Author: heeki@inje.ac.kr

Abstract— For decades artificial intelligence (AI) has been used for various applications in the healthcare industry. Machine learning and artificial intelligence algorithms allow us to diagnose and customize medical care and follow-up plans to get better results, and during the covid19 pandemic, it was found that AI models have been using to predict the Covid-19 symptoms, understanding how it spreads, speeding up research and treatment using medical data. However, it is very challenging to make a robust AI model and use it in a real-time and real-world environment since most organizations do not want to share their data with other third parties due to privacy concerns, furthermore, it is difficult to build a generalized prediction model because of the fragmented nature of the patient data across the healthcare system. To solve the above problems, this paper presents a solution based on blockchain and AI technologies. The blockchain will securely protect the data access and AI-based federated learning for building a robust model for global and real-time usage.

Keywords— Artificial intelligence, blockchain, federated learning, privacy, pandemic

I. INTRODUCTION

The advances in AI, especially in the field of machine learning and deep learning opened into several outstanding innovations in the field of healthcare especially in the field of Radiology. Due to Covid-19 pandemic, lot of use cases using AI model has been developed and provided promising result. Radiological images such as X-ray images, Computed Tomography (CT) images were used for detection of various inflammatory lung diseases ([1],[2],[3],[4]).

The blockchain based model gives the healthcare market a new dimension by considering safety aspects for data integrity and developed standardized and formalized contracts for accessing the data. Especially when we work with the electronic health record (HER), which stores data with different workflow it is difficult to know the identity of the person who does what and when the work has been performed. Block chain-based model puts a time stamp on every workflow and also put an identity to it and the copies are distributed to each participated node in the network. So, if there is modification or update happened in any node, it is equally distributed to all the

node and can be visible to everybody who is accessing it anywhere in the world. The model ensures that the data integrity is maintained between the end points without any human intervention [5].

Federated Learning (FL) is a technique based on machine learning which trains the algorithm across the edge devices in a decentralized way without leaking the privacy. The AI model developed trained inside the device based the user input data and then the data sent to the central server for aggregation of all the data and inside the central server the AI model get updated and the improved model again sent back to the device [6]. In FL, training of the algorithm has been done collaboratively without sharing the private information. In the beginning this method was used for the use cases related to edge devices and mobile devices [7]. Recently FL has gained a lot of traction in the field of healthcare. It was found that the accuracy of the model trained with FL performed better compared to the model developed at the isolated atmospheres ([8],[9]).

Looking at the current need for controlling the pandemic, AI based robust model, and at the same time keeping the privacy intact, the approach combining federated learning and blockchain could be the best possible solution that could be useful for real time environment. Rest of the paper is organized as follows: Section 2 presents the past work related to this field. Section 3 describes the proposed framework. Section 4 describes about the discussion. Section 5 describes about the conclusion and future work.

II. RELATED WORKS

1. Application of AI techniques in medical imaging

The artificial intelligence-based technique already providing promising result in detecting various critical disease using medical images. Mainly the role of the AI based technique in diagnosis of various diseases using medical images is to act as a decision support system for the doctors so that the decision taking process will be faster and at the same time maintaining the precision and accuracy. One of the best techniques is deep

learning that has been successful in detecting various diseases using medical images. Sekeroglu and Ozsahin proposed an approach that can detect Covid-19 based on chest X-ray using Convolutional Neural Network. The proposed method performed the classification of images in three groups such as healthy, pneumonia, and Covid-19 with an accuracy of 98.50% with a small imbalanced dataset [10]. Jain et al., proposed a method, that is used to classify two groups such as healthy group and Covid-19 group using chest X-rays. They have used 6432 samples and used various deep learning techniques for classification of these groups. They have tried different deep learning techniques and found that Xception model provides high accuracy of 97.97%. This model was found to be effective for detecting covid-19 patients [11]. Ozturk et al., proposed an approach that uses binary classification for differentiating two group such as Covid-19 vs No-Findings and Multiclass classification for differentiating three groups such as Covid-19 vs No-Findings vs Pneumonia. They have tried YOLO object detection system using the darknet model. It was found that binary classification provides an accuracy of 98.08% and multiclass classification provides an accuracy of 87.02%. This model would able to help the radiologists for initial screening [12].

2. Blockchain Technology

Patel discusses the development of a Cross Domain Image Sharing Blockchain Network where we can share the medical and radiological images of patients to some other healthcare centres based on a Consensus Blockchain. These network helps to keep the privacy of the information. [13]. Linn and Koo identify quite simple yet very robust use of blockchain for the data storage purpose and leveraging it to the healthcare sector for storing the data of the patients. The proposed system tends to motivate the storage of the complete data starting from the data of birth to the wearable data to the lipid profile data to the MRI data everything on the patients Blockchain [14]. Liangetal mentioned that user can share data with even the healthcare providers and insurance companies for either getting the services or insurance quotes respectively. The system also employs to be a user centric system where the user has got complete rights regarding the data to be shared or not. This kind of systems can be very helpful in terms of medical research and also preservation of personal medical records which are a real matter of privacy can be very much achieved with Blockchain as an architecture [15].

3. Federated Learning

FL based technique would able to help the healthcare providers because it can improve the accuracy and robustness of the AI model and that helps to make the model more generalizable so that it could be used in real time environment. At the same time this approach would able to save time and cost. The most important part is that we would able to get all the benefits from FL without leakage of information.

The software and hardware manufactures would get lot of benefits from the FL because the collaborative learning is possible between hardware devices and software applications without information leakage and the other benefit will come in terms of validation of AI based model since it is combined with the hardware devices that helps in continuous improvement of AI based models.

Researchers and developers get lot of benefit from the FL based services because a huge amount of real data would be available to them. This will help the AI researchers and developers to think about different algorithm strategies that would help them to make robust models [16].

III. PROPOSED FRAMEWORK

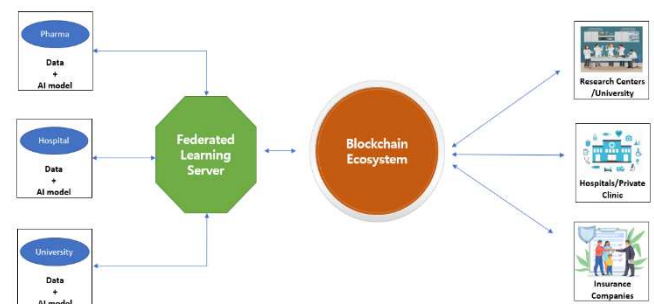


Figure 1. Blockchain- Federated Learning Architecture

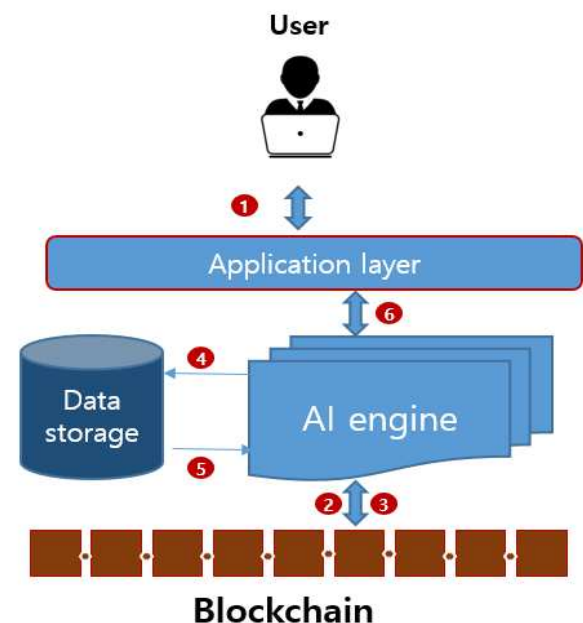


Figure 2. Token Usage Flowchart

In this section the overall architecture of the proposed framework shown in figure 1 is discussed in detail. In the beginning hospitals, universities, pharma work in the isolated place and run the model isolated to each other but later all the trained model are shared to the federated server where the

models learn among each other without leaking the private information. Usually all the devices connected to the centralized federated server used to send the data to the server and inside the server all the model averaged and made to a single model. This process keeps on repeating until a high-quality model has been generated. This kind of architecture provides lot of benefits as follows: 1) Since the learning process is collaborative in nature the models become smarter and smarter which helps in producing good result at the later stage with unknown data. 2) Since the prediction is happening locally in the new model, it lowered down the latency 3) In the present approach it is not necessary to be available near the data to get more insight about the data, insight could be obtained from anywhere. 4) The most important point is that the privacy remained through the process because all the data are isolated to that particular device, only the trained models are shared to the server.

The blockchain service allows the users (researchers, clinicians, institutions, etc) to access the data by providing a small transaction fee. The non-repudiation and patient's private data are guaranteed. By using the blockchain system, we can know who is accessing the data, when the data was requested, for what reasons. The data access mechanism is described as follow: 1) The user creates a user account on the platform, 2) The user requests access to a certain resource/asset, 3) The smart contract checks if the resource is available in the ledger. If the requested data is available, the smart contract reserves it and notifies the user. The user checks the requirements to access the data and signs the contract. Once the validation is done the user will be charged according to the smart contract's rules. Accordingly, a usage token is issued to the user.

Token usage flow is shown in the figure 2. The steps of the token usage flow are as follows: 1) The buyer or user logs in the platform using their credentials. 2) The buyer requests the data using their provided token. 3) The ML engine requests the usage rights to the ledger before proceeding. 4) If the buyer acknowledged by the system, the blockchain replies with the hash of the requested data. 5) The engine queries the data using the hash and other information returned by the blockchain. 6) The buyer can download the data or use the ML cloud engine to train it.

IV. DISCUSSION

In this study, the authors focus on the protection of personal healthcare records using blockchain-federated learning. Privacy is most important, especially in the case of medical imaging data. As the AI based model is getting traction these days especially in the field of medical imaging, protecting the private information is necessary before used in the real time environment. Particularly in the time of pandemic, AI based model has been used for detecting Covid-19 using medical imaging, however these models have not been used because of the leakage of the personal information as well as the models are not robust. With the availability of Federated learning it is ensure that robust model could be made without leaking the private information and at the same time combination with the

blockchain helps to transact data securely with different service providers. Recently NVIDIA Clara FL – a FL based framework has been using for testing purposes in medical devices. If it is successful then the toolkits developed using this framework will help the radiologist to develop the algorithm in their own institutions using their own patient data. The start-up Owkin works on machine learning developed a federated learning platform that gives the owner the power to track of their data usage [17]. The way the federated learning is gaining the traction and giving hope for building robust AI model and make is generalized. It can be assured that FL based learning will be used in other healthcare areas such as pharmaceutical industries and also life science industries and FL in combination with blockchain technology will provides hope that the medical data could be shared to all service providers without leaking the private information.

V. CONCLUSION

In this paper, a solution based on blockchain and AI Federated learning has been proposed to protect and share safely the patient's healthcare data. After analysing the potentials of combining blockchain and AI, we realized that the proposed solution would help to build robust models and share them without being compromised. However, the application wasn't developed and tested, yet this solution is indeed promising. In the future, our proposed framework will be implemented and tested in a real-world environment and its effectiveness will be studied more in detail.

ACKNOWLEDGMENT

This research was supported by the Ministry of Trade, Industry, and Energy (MOTIE), KOREA, through the Education Program for Creative and Industrial Convergence (Grant Number N0000717)

REFERENCES

- [1] Choe, J., Lee, S.M., Do, K.H., Lee, G., Lee, J.G., Lee, S.M. and Seo, J.B., 2019. Deep learning-based image conversion of CT reconstruction kernels improves radiomics reproducibility for pulmonary nodules or masses. *Radiology*, 292(2), pp.365-373.
- [2] Kermay, D.S., Goldbaum, M., Cai, W., Valentim, C.C., Liang, H., Baxter, S.L., McKeown, A., Yang, G., Wu, X., Yan, F. and Dong, J., 2018. Identifying medical diagnoses and treatable diseases by image-based deep learning. *Cell*, 172(5), pp.1122-1131.
- [3] Negassi, M., Suarez-Ibarrola, R., Hein, S., Miernik, A. and Reiterer, A., 2020. Application of artificial neural networks for automated analysis of cystoscopic images: a review of the current status and future prospects. *World Journal of Urology*, pp.1-10.
- [4] Gulshan, V., Peng, L., Coram, M., Stumpe, M.C., Wu, D., Narayanaswamy, A., Venugopalan, S., Widner, K., Madams, T., Cuadros, J. and Kim, R., 2016. Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. *Jama*, 316(22), pp.2402-2410.
- [5] Tandon, A., Dhir, A., Islam, N. and Mäntymäki, M., 2020. Blockchain in healthcare: A systematic literature review, synthesizing framework and future research agenda. *Computers in Industry*, 122, p.103290.
- [6] Yang, Q., Liu, Y., Chen, T. and Tong, Y., 2019. Federated machine learning: Concept and applications. *ACM Transactions on Intelligent Systems and Technology (TIST)*, 10(2), pp.1-19.
- [7] Kairouz, P., McMahan, H.B., Avent, B., Bellet, A., Bennis, M., Bhagoji, A.N., Bonawitz, K., Charles, Z., Cormode, G., Cummings, R. and d'Oliveira, R.G., 2019. Advances and open problems in federated learning. *arXiv preprint arXiv:1912.04977*.

- [8] Li, W., Milletari, F., Xu, D., Rieke, N., Hancox, J., Zhu, W., Baust, M., Cheng, Y., Ourselin, S., Cardoso, M.J. and Feng, A., 2019, October. Privacy-preserving federated brain tumour segmentation. In International Workshop on Machine Learning in Medical Imaging (pp. 133-141). Springer, Cham.
- [9] Sheller, M.J., Reina, G.A., Edwards, B., Martin, J. and Bakas, S., 2018, September. Multi-institutional deep learning modeling without sharing patient data: A feasibility study on brain tumor segmentation. In International MICCAI Brainlesion Workshop (pp. 92-104). Springer, Cham.
- [10] Sekeroglu, B. and Ozsahin, I., 2020. <? covid19?> Detection of COVID-19 from Chest X-Ray Images Using Convolutional Neural Networks. SLAS TECHNOLOGY: Translating Life Sciences Innovation, p.2472630320958376.
- [11] Jain, R., Gupta, M., Taneja, S. and Hemanth, D.J., 2020. Deep learning based detection and analysis of COVID-19 on chest X-ray images. Applied Intelligence, pp.1-11.
- [12] Ozturk, T., Talo, M., Yildirim, E.A., Baloglu, U.B., Yildirim, O. and Acharya, U.R., 2020. Automated detection of COVID-19 cases using deep neural networks with X-ray images. Computers in Biology and Medicine, p.103792.
- [13] Patel, V., 2019. A framework for secure and decentralized sharing of medical imaging data via blockchain consensus. Health informatics journal, 25(4), pp.1398-1411.
- [14] Linn, L.A. and Koo, M.B., 2016. Blockchain for health data and its potential use in health it and health care related research. In ONC/NIST Use of Blockchain for Healthcare and Research Workshop. Gaithersburg, Maryland, United States: ONC/NIST (pp. 1-10).
- [15] Ozturk, T., Talo, M., Yildirim, E.A., Baloglu, U.B., Yildirim, O. and Acharya, U.R., 2020. Automated detection of COVID-19 cases using deep neural networks with X-ray images. Computers in Biology and Medicine, p.103792.
- [16] Rieke, N., Hancox, J., Li, W., Milletari, F., Roth, H., Albarqouni, S., Bakas, S., Galtier, M.N., Landman, B., Maier-Hein, K. and Ourselin, S., 2020. The future of digital health with federated learning. arXiv preprint arXiv:2003.08119.
- [17] <https://www.zuehlke.com/en/insights/federated-learning-healthcare-ai-that-doesnt-risk-patient-data-privacy>, extracted on 17/11/2020.



Satyabrata Aich, is working as a Lecture at the Institute of Digital Anti-Aging Healthcare, Inje University. He has published many research papers in journals and conferences in the realms of Machine learning, Text mining, and Artificial intelligence. His research interests are natural language processing, Machine learning, supply chain management, Text mining, Bio informatics, and Blockchain.



Nday kabulo Sinai is a PhD student majoring in digital healthcare at INJE University. He has received his bachelor's degree in software engineering at the "Université Protestante de Lubumbashi", He also attended at KOOKMIN University from where He got his master's degree in Computer Engineering. His research interests include Blockchain, artificial intelligence, micro-services, software quality and IoT.



Saurabh Kumar received his PhD. degree from Chungbuk National University, South Korea in Feb 2020. He has wide industry experience in companies like Samsung Electronics and HCL Technologies. He has published various journal and conference papers in IEEE and Springer. His research interests include wireless communication, computer vision, artificial intelligence, robotics and block chain.



Mohammed Ali, is a PhD student majoring in Medicine at Yonsei University. He has received his MSc degree in medical sciences from Yonsei University. He has received Korean government scholarship for graduate studies, and currently being member of American Society of Microbiology. His research interests include artificial intelligence, Drug discovery, Bio informatics



Yu Ran Choi is a master student at the Institute of Digital Anti-Aging Healthcare, Inje University. She has worked in many real-life projects related to artificial intelligence. Her research interests are artificial intelligence, Machine learning, clinical research, healthcare big data, and data analysis



Moon-Il Joo received the PhD degree in computer engineering from Inje University in 2018. He is currently working as a research professor at the Institute of digital anti-aging healthcare, Inje University, Korea. His research interests are in software engineering, human computer computer interaction, smart phone programming, and component-based development.



Hee-Cheol Kim, received his BSc at Department of Mathematics, MSc at Department of Computer Science in SoGang University in Korea, and PhD at Numerical Analysis and Computing Science, Stockholm University in Sweden in 2001. He is Professor at Department of Computer Engineering and Head of the Institute of. Digital Anti-aging Healthcare, Inje University in Korea. His research interests include Machine learning, Text mining, Bio Informatics.