MEDINFO 2023 — The Future Is Accessible
J. Bichel-Findlay et al. (Eds.)

doi:10.3233/SHTI231182

A Blockchain-Based Approach for Patient Data Alignment Across Institutions

Tianshu ZHOU[a], Yao LU[a], Yin ZHANG[a], Jiaqi WANG[a], Yu TIAN[b], Jingsong LI[a,b,1]
[a] *Research Center for Healthcare Data Science, Zhejiang Laboratory, China.*
[b] Zhejiang University, China

Abstract. Clinical studies need multi-center, long-term patient data, which are difficult to align. We present a blockchain-based approach that uses cryptographic matching and attribute-based encryption for secure data alignment, aggregation, and access. It improves efficiency, lowers data synchronization, and facilitates cross-institutional patient data association and visualization.

Keywords. Blockchain, patient data alignment, cryptographic data matching

1. Introduction

Complete clinical data are crucial for high quality machine learning models for clinical analysis [1-3]. However, it is challenging to align and summarize clinical data across different institutions [4-6]. We present a lightweight, distributed system based on blockchain for cross-institutional patient information alignment.

2. Methods

The system uses hyper ledger Fabric and PBFT consensus algorithm. The process is: A researcher imports patients' identity information at one node, and requests alignment.

3. Results

The system is evaluated on a distributed platform of 9 hospitals. A reference blockchain system with full data synchronization is also developed for comparison. 5000 kidney disease patients' data are used, distributed as follows: hospital 1 has 5000 patients' demography, examination, medication and surgery records.

Table 1 compares the running time of the proposed and reference system for block lengths of 500, 1000 and 1500. The proposed system achieves an operating efficiency improvement of 124% - 349%, and this gap will widen as the block length and patient information increase.

[1] Corresponding Author: Li Jingsong, email: ljs@zju.edu.cn.

Table 1. Comparison of runtime with data volume between different systems

Block length	Running time (ms)		
	Reference system	Proposed system	Efficiency improve (%)
500	2681.465	1194.5112	124.4821982
1000	4307.4243	1261.2795	241.5122738
1500	6668.1317	1487.1842	349.0255435

Under full data synchronization mode, the data volume for synchronization grows linearly with the patient information. Table 2 shows the percentage of data reduction compared with full data synchronization mode for different patient numbers. The larger the overall data, the more significant the data reduction.

Table 2. Comparison of synchronization data volume between different system

Patient number	Synchronous data size of reference system	Synchronous data size of proposed system	Synchronous data size reduction (%)
1000	126.61 KB	17.88 KB	85.87
3000	758.63 KB	102.5 KB	86.48
5000	1230 KB	116.68 KB	90.51

4. Conclusions

The system aligns and analyzes clinical data securely and easily across institutions. It supports data collection and research, and can be deployed on a multicenter platform. It is also an ISO draft standard core content.

Acknowledgements

This work was supported by the National Key Research and Development Program of China (No. 2022YFC2504605), and the Major Scientific Project of Zhejiang Lab (No. 2022ND0AC01).

References

[1] Srinivas TR, Taber DJ, Su Z, Zhang J, Mour G, Northrup D, Tripathi A, Marsden JE, Moran WP, Mauldin PD. Big data, predictive analytics, and quality improvement in kidney transplantation: a proof of concept. Am J Transplant. 2017 Mar;17(3):671-81, doi: 10.1111/ajt.14099.

[2] Xiao C, Choi E, Sun J. Opportunities and challenges in developing deep learning models using electronic health records data: a systematic review. J Am Med Inform Assoc. 2018 Oct;25(10):1419-28, doi: 10.1093/jamia/ocy068.

[3] Panahiazar M, Taslimitehrani V, Pereira N, Pathak J. Using EHRs and Machine Learning for Heart Failure Survival Analysis. Stud Health Technol Inform. 2015;216:40-4,

[4] Van Panhuis WG, Paul P, Emerson C, Grefenstette J, Wilder R, Herbst AJ, Heymann D, Burke DS. A systematic review of barriers to data sharing in public health. BMC Public Health. 2014 Dec;14:1144, doi: 10.1186/1471-2458-14-1144.

[5] Just BH, Marc D, Munns M, Sandefer R. Why patient matching is a challenge: research on master patient index (MPI) data discrepancies in key identifying fields. Perspect health inf manag. 2016 Apr; 13(Spring): 1e.

[6] Windle JR. 2016 ACC/ASE/ASNC/HRS/SCAI health policy statement on integrating the healthcare enterprise. J Nucl Cardiol. 2016 Sep;23(6):1527-43.