BOHR International Journal of Engineering
2023, Vol. 2, No. 1, pp. 1–6
DOI: 10.54646/bije.2023.11
www.bohrpub.com

RESEARCH

Hospital recommender system

Shrey Sumariya*, Shreyas Rami*, Shubham Revadekar*, Vidhan Shah* and Sudhir Bagul*

Department of Computer Engineering, Dwarkadas J. Sanghvi College of Engineering, Mumbai, India

***Correspondence:**
Shrey Sumariya,
shrey.sumariya@gmail.com
Shreyas Rami,
shreyasrami24.sr@gmail.com
Shubham Revadekar,
shubham0revadekar@gmail.com
Vidhan Shah,
shahvidhan123@gmail.com
Sudhir Bagul,
Sudhir.bagul@djsce.ac.in

Received: 05 January 2023; **Accepted:** 11 January 2023; **Published:** 11 January 2023

Elderly patients require more medical effort. It is clear that early-stage disease diagnosis can support timely and appropriate treatment. But if you don't pay attention in a timely manner, it can lead to different kinds of health problems that can lead to death. Take advantage of our recommendation system to recommend hospitals. A recommender system uses algorithms to provide product or service recommendations to users. By combining blockchain technology and machine learning models, we provide users with highly accurate recommendations. This whitepaper describes how sophisticated machine learning models and blockchain can be connected to improve recommendations, providing hospitals with higher performance and more accurate recommendations. An optimized model for recommending hospitals in a better manner is the main goal behind this paper.

Keywords: machine learning, smart contracts, recommendation system, TFID Vectorizer, cosine similarity

Introduction

Typical questions that traditional recommendation systems answer is about what kind of books are worth reading, which movies are worth watching, and what products are worth buying. A recommender system helps you sift through large amounts of data to determine the options that are the most relevant tasks that users have in mind. Recommendation systems are everywhere when customers buy products from online stores, e.g., Amazon. Use an intelligent recommendation system that recommends different products using a content-based collaborative approach. Say something like, "Customers who buy this will buy that," or put it another way, "There are some products that might be useful, based on your searches here." Decisions made by customers are low risk, but decisions in other areas often have more serious consequences, and recommendations based on those decisions may yield unreliable results. Health and medicine are one such field. A recommendation system should not only avoid mistakes and support decision-making but should also understand patients, attitudes, requirements, and values in the context of disease and health care. This makes health recommendation systems more difficult to apply. Considering various diseases, hospitals or doctor recommendation has many aspects to consider before recommending someone. Cost, patient risk, location, disease type, doctor availability, and many other factors should be considered when making a decision. Mistakes or errors in considering any of the above factors could have serious consequences. Medical recommendations play an important role, but the risk factors are high, and many aspects must be considered. Hospital referrals for cancer patients are always a tedious task. There is a lot of information to consider when recommending a hospital. First of all, cancer is a high-risk disease, and in many cases, patients need to receive proper medical care, which requires proper medical care in all aspects.

In today's world, information run and privacy area units are therefore vital, and we have a tendency to tend to use blockchain for this purpose. Additionally, to provide privacy protection, we have a tendency to produce a hash layer that is troublesome to hack to forestall information loss. A blockchain may be distributed info that keeps a history of all transactions that have ever occurred during a blockchain network. All network nodes have a duplicate of this info, replicated and secured. The advantage of blockchain is that untrusted participants will communicate and exchange assets in a secure manner while not the necessity of a sure third party. As the name suggests, a blockchain may be a series of blocks joined along by a cryptological hash. Every block references the previous block by storing a cryptological hash, manufacturing a series of blocks. Additionally, to the previous block's hash, a block contains a collection of transactions that, once accepted and connected to the blockchain, cannot be updated or deleted. As a result, each consistency and double-spending problem area unit was alleviated. A wise contract may be a worm that runs on the blockchain. It is viewed as a sure third party between leery participants. It consists of contract memory, credits, and program code. By merely submitting dealing to the blockchain, it is created and utilized by any node on the network. Sensible contracts' code is fastened and cannot be updated once it is on the blockchain. During this article, I primarily wish to stipulate the associate in the nursing application I needed to make for hospital referrals for cancer patients. During the latter half, I will be able to make a case for everything we have a tendency to use, together with, however, our recommendation system considers several factors in recommending hospitals.

Literature review

In (1), the paper "Building a hospital referral expert system with a Prediction and Optimization-Based Decision Support System algorithm" proposes to design a recommendation system for hospitals using artificial intelligence. Choosing a healthcare facility with a history of providing quality healthcare can make the distinction between asked issues and incorrect issues, including death. Endured croakers learn which facilities offer stylish care. Physicians generally play a crucial part in recommending specific hospitals for their cases. Some medical situations are time-critical, and transportation time plays a veritably important part in issues. Some medical situations require immediate attention, and transportation time is critical. Distance is frequently the most important consideration when choosing a hospital for patients living in pastoral and underserved areas. Even in non-emergency circumstances, proximity is preferable. As a result, several research studies have shown that patients frequently favor local higher-risk healthcare facilities to go to lower-risk health facilities. Geographic factors may influence the effect of institutional predictors. This can be calculated by experts to assess the outcomes of their problems.

Physicians assign a case to a particular hospital focusing on the patients' physical condition and distance to travel. An experienced physician can select the hospital that poses the least risk to the patient. A patient is more likely to accept such a tailored approach. The goal of this project is to create an expert system that can help with such a personally tailored hospital selection decision. Two significant priorities are travel distance and survival likelihood. When the hospital selection decision is tailored, the tradeoff between different aims can be handled explicitly. Due to the small result space, the problem is first framed as single-objective optimization in this study and could be addressed by a systematic search (only the number of hospitals). In this optimization, a query provides the system with data about the patient, such as age, admission type, comorbidities, and the maximum suffered distance. To generate a personalized hospital selection, the optimization process will combine the given information with the captured knowledge.

The goal of the gathering and analysis tool is to find the hospital with the highest likelihood of survival considering the greatest suffered distance to a hospital. If we also want to consider other factors in the healthcare decision, such as the likelihood of difficulties, we have a multi-objective optimization problem.

From (2), our strategy requires that there will be two different kinds of independent variables in the problem. Uncontrollable (unchangeable) variables are the first category. These variables' values are fixed and cannot be changed. Patient characteristics such as demographic information, the results of diagnostic tests, the kind of admission, whether surgery was performed, the patient's comorbidity scores, and the method of payment, for instance, are uncontrollable variables in this study. Variables with controlled (adjustable) values are the opposite type. These factors can be used to base the recommendation. The values of these variables in our application describe a hospital. We suggest a new functionality for using blockchain technology to manage and clean the dataset. Based on the information provided by the blockchain, the controllable or modifiable variables in the dataset are frequently updated.

From (3), the blockchain era is redefining facts of modeling and governance deployed in lots of healthcare applications. This is in particular because of its adaptability and ability to segment, steady, and proportion clinical facts and offerings in an unheard-of way. The blockchain era is at the center of many modern-day traits withinside the healthcare enterprise. The blockchain era sits at the pinnacle of the uncooked facts layer this is taken into consideration as the middle framework in pursuit to create a secured healthcare structure this is divided into 4 components. Blockchain technology gets more and more effective and robust, as they come to be coupled with AI in numerous real-phrase healthcare solutions. Machine gaining knowledge and deep gaining knowledge is the primary using elements for the AI domain, which is likewise relentlessly enhancing the development of automation. The extra facts we feed to the device and the extra cap potential a device will benefit to categorize or expect styles accurately. The blockchain era is gaining sizable interest from individuals, in addition to groups of almost a wide variety and dimensions. It is able to rework the conventional enterprise with its features, which encompass decentralization, anonymity, persistence, and audibility. The blockchain era is anticipated to reshape the healthcare ecosystem. Not best will the method be obvious and steady, but additionally the exceptional of healthcare can be multiplied at a decreased cost. Specifically, we supplied modern-day studies on fitness fact control and the way blockchain will empower sufferers and streamline the sharing method of fitness facts. We located that there is a consensus among researchers that, with the blockchain era, affected person facts can be surely owned and managed through the rightful proprietor of the facts, i.e., the affected person. The blockchain permits fitness data to be time-stamped in order that nobody can tamper with them after turning them into a part of the distributor ledger.

The sufferers can have the proper to determine who can and cannot get the right of entry to their facts and for what purpose. However, there are nonetheless numerous open demanding situations that require additional investigation. For instance, cross-border sharing of fitness facts in which distinct and regularly conflicting jurisdictions exist may also avoid the advantage of blockchain's facts sharing. In fact, the expectancy of a character's privateness varies from one United States of America to any other primarily based totally on the authority's regulations. Therefore, destiny studies on regulation, standardization, and cross-border fitness facts retrieving rules such as retention and utilization goal are duly urgent. Furthermore, we mentioned modern-day studies of blockchain on healthcare SCM. In particular, the applicability of blockchain to cope with consider degradation and to reinforce fact transparency at the scientific trials. Several researchers advise using blockchain to enhance the medical credibility of findings from scientific trials, which can be undermined through issues including lacking facts and selective publication.

From (4), another function we have got protected is data-to-textual content technology systems, which are more and more used withinside the fitness domain. They can, for instance, be used for automation of fitness reports, scientific selection support, inspiring behavioral change, making sure affected person engagement, or helping sufferers with making fitness decisions. Many sufferers are not able to advantage of these established fitness records, due to the fact that those files fail to speak vital records that impact the affected person's expertise of those materials (e.g., the affected person's character risks, issues, and values). Tailoring fitness verbal exchange to character sufferers appears a fruitful solution. In addition, despite the fact that such customized fitness records may be effective, it is not without demanding situations. Personalizing fitness records manually is time-consuming and highly priced, and the outputs are regularly inconsistent. Natural language technology (NLG) strategies can address those issues and are consequently more and more used withinside the fitness domain. The main instance of tailor-made fitness records with the use of NLG is the BabyTalk device evolved through Gatt et al. (2009).

From (5–9), Ethereum is a public, open-source, blockchain-based distributed computing platform and operating system with smart contract functionality. Using blockchain, we are developing a decentralized and self-tallying Internet voting protocol with maximum voter privacy. The healthcare industry is one of the largest in the world, accounting for more than 10% of the gross domestic product (GDP) of the most developed countries. Blockchain and its smart contract capabilities, in particular, have the potential to address healthcare interoperability issues such as enabling effective interactions between users and medical applications, securely delivering patient data to a wide range of organizations and devices, and improving the overall efficiency of medical practice workflow. Accessing and managing a large amount of medical data is also part of the job. The associated cost for this system has been estimated in terms of a feasibility study as part of the implementation of the workflows of the medical smart contract system for healthcare management. Not all customized NLG information is effective. Reiter et al. (2003) developed "STOP," a system for creating personalized smoking cessation letters. Nontailored letters, in contrast, were just as effective as tailored letters. We do not need to change health behavior in our system, but rather inform patients in the best way possible to help them make decisions (4).

Design

Recommender system

First, we take source datasets from the websites that host data about doctors. This dataset contains important information such as consultant fees, locations where the doctors work, ratings, hospitals, etc. The dataset will be generated using Web Scraping technology. The dataset is then cleaned by removing unwanted outliers and missing values. Incorrect tuples are either removed or corrected by scrapping another source.

Some part of the cleaned data is then passed to the ML model where the model learns how to give suggestions based on certain parameters. The system will be using sophisticated algorithms and will return with a ranked list of doctors. The remaining part of the cleaned dataset is used to test the accuracy and performance of the ML model. After testing, the ML model is then implemented on the application.

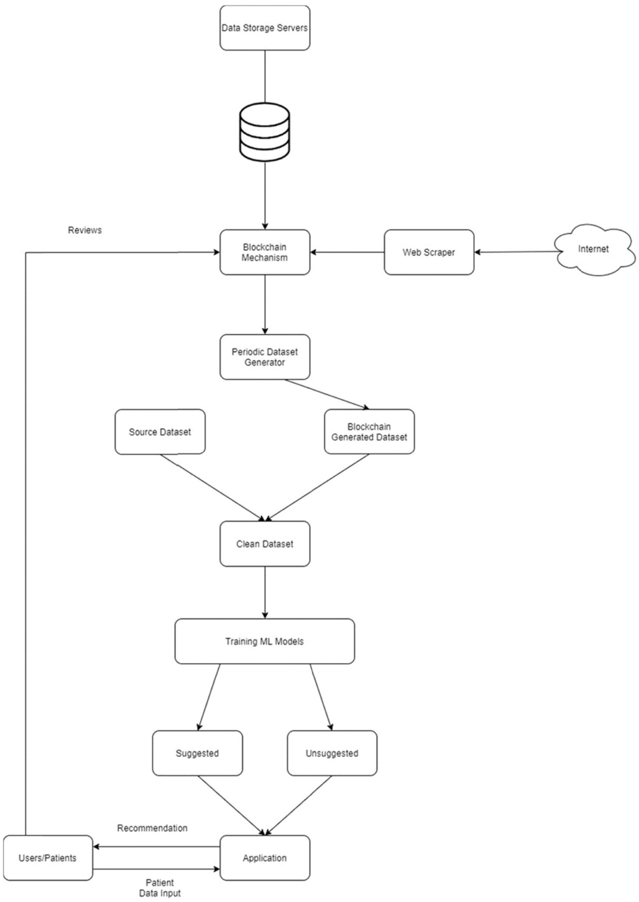


FIGURE 1 | Proposed design for the system.

During the runtime of the application, the users can input their parameters such as location and price range, and the model will be returning the preferred doctors' list. Feedback from the users will be recorded, which will be used to determine the rank of the recommended doctors. The positive feedback will increase the rank of the doctors on the list. The negative ones will decrease the rank. The blockchain will be storing the ranks of the doctors.

The ranks of doctors are also affected by the latest data that web crawlers will be mining. This new data will be cleansed and will be used to retrain the ML model with the latest trends in the data. A periodic dataset generator will be used that will handle all the latest dataset corrections. The ML model will then be trained periodically to update its recommendations (**Figure 1**).

Implementation

There are numerous tools and libraries available for DApp development. The tools and libraries used in developing this Hospital Recommender System (HRS) are, however, briefly discussed in **Table 1**.

TABLE 1 | Tools and libraries used in hospital recommender system (HRS).

Tools and libraries	Descriptions
Truffle framework	Allows the development of DApps on the Ethereum network. Provides a set of tools for creating solidity-based smart contracts. Provides a framework for contract testing. Provides tools for deploying Smart contracts on the blockchain. Inside truffle, this is used to create client-side applications.
Ganache	Runs a local Ethereum instance. Provides sample accounts for use in the development and testing processes.
Solidity	Knowledge of object-oriented programming is required to write and implement smart contracts on the Ethereum blockchain.
Web3js	JavaScript open-source library. API for interacting with local Ethereum nodes.
Django	Server for the backend of our system.
MetaMask	Chrome browser extension acting as a Web 3 wallet. Inserts Web3js libraries into the browser to allow reading and writing on the Ethereum blockchain.
Web Front End	ReactJs, which heavily relies on Web3js to connect to blockchain nodes.

We use the local Ethereum framework, MetaMask, Web3js, and Django to create the prototype of the proposed framework.

We are using smart contracts for reading and writing the ranks of doctors. Currently, we are using a rank-based system, where we assign a rank to each of our recommended doctors, based on whether the user likes our recommendation or not. If the user likes the recommendation, the rank will increase.

Smart contract stores data in their own private part of the blockchain state, and every non-light node has a copy of this state, and, while usually not directly exchanging blockchain states, nodes use hashes to make sure their copies of the blockchain state are in sync.

When a smart contract accesses its state variable, the node that executes the corresponding transaction reads the value of this variable from the local copy of the blockchain state.

We searched for available datasets in research papers, Kaggle, and other resources on the Internet. But, a dataset that satisfied our requirements was not available. We found a website which showed all the doctors and other details that we require. We used web scraping to scrape the data as per our requirements. The data were then preprocessed to remove any null values and get the specific fields such as doctor experience, fees, ratings, and location.

To convert the city to numeric form so that it can be used to train our regression model, one hot encoder was used. Ranks for each doctor were derived using the ratings and fee values. A doctor with higher ratings will have a higher rank and for two doctors with the same ratings, the one with less fees will have a higher rank than the other.

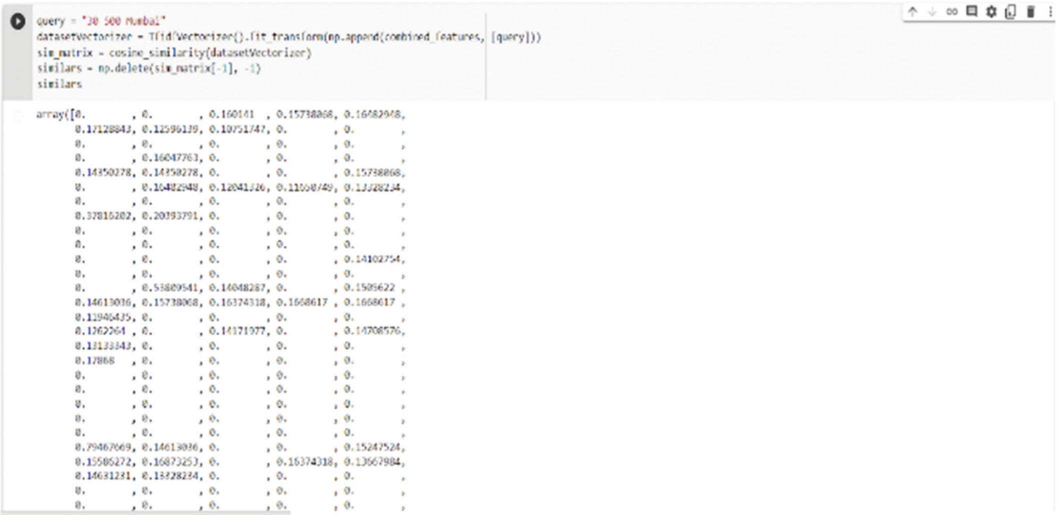
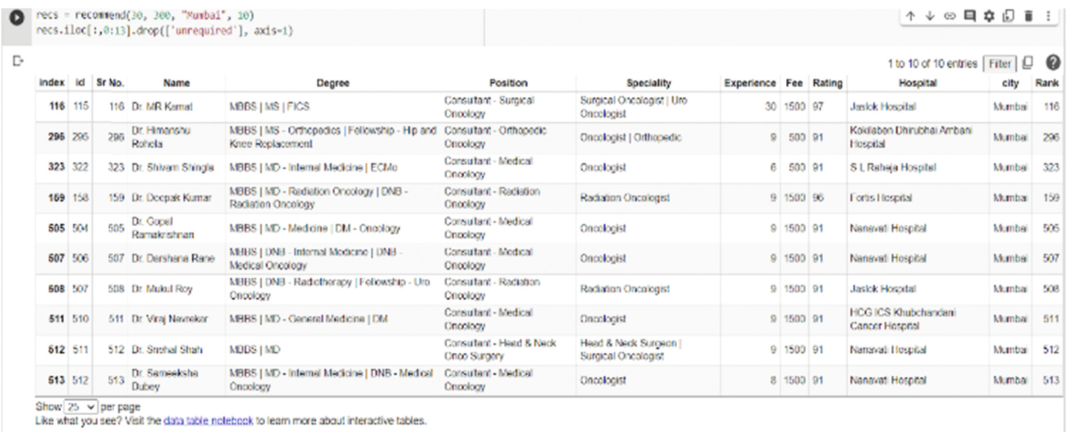


FIGURE 2 | Result of TFID Vectorizer for some sample parameters.



```
recs = recommend(30, 300, "Mumbai", 10)
recs.iloc[:,0:13].drop(['unrequired'], axis=1)
```

1 to 10 of 10 entries

index	id	Sr No.	Name	Degree	Position	Speciality	Experience	Fee	Rating	Hospital	city	Rank
116	115	116	Dr. MR Kamat	MBBS | MS | FICS	Consultant - Surgical Oncology	Surgical Oncologist | Uro Oncologist	30	1500	97	Jaslok Hospital	Mumbai	116
296	295	296	Dr. Himanshu Rohela	MBBS | MS - Orthopedics | Fellowship - Hip and Knee Replacement	Consultant - Orthopedic Oncology	Oncologist | Orthopedic	9	500	91	Kokilaben Dhirubhai Ambani Hospital	Mumbai	296
323	322	323	Dr. Shivam Shingla	MBBS | MD - Internal Medicine | ECMo	Consultant - Medical Oncology	Oncologist	6	500	91	S L Raheja Hospital	Mumbai	323
159	158	159	Dr. Deepak Kumar	MBBS | MD - Radiation Oncology | DNB - Radiation Oncology	Consultant - Radiation Oncology	Radiation Oncologist	9	1500	96	Fortis Hospital	Mumbai	159
505	504	505	Dr. Gopal Ramakrishnan	MBBS | MD - Medicine | DM - Oncology	Consultant - Medical Oncology	Oncologist	9	1500	91	Nanavati Hospital	Mumbai	505
507	506	507	Dr. Darshana Rane	MBBS | DNB - Internal Medicine | DNB - Medical Oncology	Consultant - Medical Oncology	Oncologist	9	1500	91	Nanavati Hospital	Mumbai	507
508	507	508	Dr. Mukul Roy	MBBS | DNB - Radiotherapy | Fellowship - Uro Oncology	Consultant - Radiation Oncology	Radiation Oncologist	9	1500	91	Jaslok Hospital	Mumbai	508
511	510	511	Dr. Viraj Navnekar	MBBS | MD - General Medicine | DM	Consultant - Medical Oncology	Oncologist	9	1500	91	HCG ICS Khubchandani Cancer Hospital	Mumbai	511
512	511	512	Dr. Snehal Shah	MBBS | MD	Consultant - Head & Neck Onco Surgery	Head & Neck Surgeon | Surgical Oncologist	9	1500	91	Nanavati Hospital	Mumbai	512
513	512	513	Dr. Sameeksha Dubey	MBBS | MD - Internal Medicine | DNB - Medical Oncology	Consultant - Medical Oncology	Oncologist	8	1500	91	Nanavati Hospital	Mumbai	513

Show 25 per page
Like what you see? Visit the data table notebook to learn more about interactive tables.

FIGURE 3 | Top 10 doctors based on the sample.

After preparing the final dataset, we trained our machine learning models using the data. Experience, fees, and city were the features that we used to predict the target rank for each target. We used TF-IDF Vectorizer and Cosine Similarity (Unsupervised Learning) to find the most similar tuples from the dataset. The top 10 most similar tuples from the dataset matching the city are picked and sorted on the basis of user feedback.

Results and discussion

To supply input to our models, we created a function that will take the basic input and convert it to a form that can be given to our model. To start with, we supplied user input to the TF-IDF Vectorizer, which converted the inputs into vectors. We then provided these vectors to the Cosine Similarity Algorithm, which showed the value of each record between 0 and 1 based on the similarity. **Figure 2** shows the result of one prediction when the user entered particular parameters, and **Figure 3** shows the top 10 doctors based on the sample parameters. After getting the top 10 doctors list from the model, we passed those ids to the blockchain, which then gives likes and dislikes of the top 10 doctors, accordingly on basis of likes, we showed the user in descending order of likes with the doctor with higher like at the top and the one with lowest at the bottom.

Conclusion

We have designed a solution for the difficulty in finding doctors and hospitals for cancer patients. In our solution, the user will have to sign in first followed by which the user will have to provide parameters such as cost, location, and experience. Our solution will then use the TF-DIF Vectorizer to convert the given parameters into vectors, and these vectors are then given to cosine similarity to get similar records. It then fetches data from the blockchain and sorts similar records based on the number of likes, and the top

10 doctors from the list are then recommended to the user. Thus, we have used blockchain with machine learning to improve the efficiency and authenticity of the recommender system, which will make finding hospitals and doctors super easy and reliable.

Furthermore, this can be extended to support all diseases. Features such as voice navigation, interpretation of medical reports, and medicine prediction can also act as add-ons to this system.

References

1. Chi CL, Street WN, Ward MM. Building a hospital referral expert system with a prediction and optimization-based decision support system algorithm. *J Biomed Inform.* (2008) 41:371–86.
2. Miller A, Juels A, Shi E, Parno B, Katz J. Permacoin: Repurposing bitcoin work for data preservation. *Proceedings of the 2014 IEEE Symposium on Security and Privacy*. Berkeley, CA: IEEE (2014). p. 475–90.
3. Khezr S, Moniruzzaman M, Yassine A, Benlamri R. Blockchain technology in healthcare: a comprehensive review and directions for future research. *Appl Sci.* (2019) 9:1736.
4. Hommes S, van der Lee C, Clouth F, Vermunt J, Verbeek X, Krahmer E. A personalized data-to-text support tool for cancer patients. *Proceedings of the 12th International Conference on Natural Language Generation*. Tokyo: Association for Computational Linguistics (2019). p. 443–52.
5. Zhang Q, Xu B, Jing H, Zheng Z. Ques-chain: an Ethereum based e-voting system. *arXiv [preprint]* (2019). doi: 10.48550/arXiv.1905.05041
6. McCorry P, Shahandashti SF, Hao F. A smart contract for boardroom voting with maximum voter privacy. *Proceedings of the International Conference on Financial Cryptography and Data Security*. Cham: Springer (2017). p. 357–75.
7. Katuwal GJ, Pandey S, Hennessey M, Lamichhane B. Applications of blockchain in healthcare: current landscape & challenges. *arXiv [preprint]* (2018). doi: 10.48550/arXiv.1812.02776
8. Zhang P, White J, Schmidt DC, Lenz G. Applying software patterns to address interoperability in blockchain-based healthcare apps. *arXiv [preprint]* (2017). doi: 10.48550/arXiv.1706.03700
9. Khatoon A. A blockchain-based smart contract system for healthcare management. *Electronics.* (2020) 9:94.