Technological Model for the Protection of Genetic Information using Blockchain Technology in the Private Health Sector

Julio César Arroyo-Mariños, Karla Mariella Mejia-Valle and Willy Ugarte[a]

Universidad Peruana de Ciencias Aplicadas (UPC), Lima, Peru

Keywords: Genetic Information, Blockchain, Patient, Healthcare.

Abstract: Currently, genetic information is considered a valuable element in the Health Sector, since due to more accurate diagnostic samples in medical genomics this allows to offer better treatments to patients against various types of diseases. This paper presents the development of a technological model using Blockchain as technology to ensure the protection of genetic information in the Private Health Sector because activities related to the storage or management of data of this information present many points of vulnerability. In addition, activities such as registration and control of access in the exchange of genetic information have been administered in an unauthorized manner, mainly through entities that manage the eligibility of users with genomic information and allow access to specific data sets, which shows a lack of harmonization in access policies between the owners of their genetic information and the health entities that manage it. A proof-of-concept has been performed to validate the capabilities of the model and ensure that a larger-scale deployment can be performed. The evaluation showed that experts agree with the proposal and that users would be willing to use proof of concept to ensure traceability and security of their information

1 INTRODUCTION

Today, medicine that considers the patient's genetic information as a variable to consider is already a reality. Genetic testing within clinical practice is providing results aimed at confirming or obtaining a more accurate diagnosis of genetic diseases, deciding treatment, or assessing the risk of relapse to certain types of diseases such as cancer and its derivatives, hepatitis, and others.

In Peru, the management of genetic information in health entities is not very pronounced, in view of this INEN shows a comparison of previous years in which it is highlighted that the waiting time for a consultation takes an average of 60 minutes. In addition, the number of genetic information consultations during the years 2013 to 2018 has been on average 450 consultations per year[1].

In the face of this, the great advancement of technological development has brought with it within the field of health an important and profound development of the genomic information of the human being.

In view of this, various genetic analysis methodologies have been applied, as well as the biomedical implementations necessary for the interpretation of identified alterations (Shabani, 2019a).

A clear example of this is the digitization of the genome, which is born by the need to store genetic information digitally. With this, doctors can streamline diagnostic and treatment processes in people.

However, as with personal information, it is of paramount importance that a person's genetic information is protected for privacy reasons, and they also bring with them a major challenge regarding the transport and storage of personal genetic information in a simple, secure and anonymous manner because people do not want anyone to manipulate their information[2]. Encouraging open and responsible exchange of genomic data is a central focus of many national and international programmes and initiatives for their implementation and use.

However, there are some challenges based on the adoption of centralized approaches to data storage, exchange and access due to the constraints faced by these central mechanisms and the non-automated nature of traditional data access and exchange to es-

[a] https://orcid.org/0000-0002-7510-618X

[1] Chief Resolution (In Spanish) N 490-2019-J - https://bit.ly/2RQ0k5X

[2] Emerging Tech 2015: Digital Genome - https://bit.ly/3ePkBDm

Arroyo-Mariños, J., Mejia-Valle, K. and Ugarte, W.
Technological Model for the Protection of Genetic Information using Blockchain Technology in the Private Health Sector.
DOI: 10.5220/0010422401710178
In *Proceedings of the 7th International Conference on Information and Communication Technologies for Ageing Well and e-Health (ICT4AWE 2021)*, pages 171-178
ISBN: 978-989-758-506-7

tablish central mechanisms in data access management (Shabani, 2019b).

Genomic data security risks are highlighted because data storage and management activities that contain genetic information have many points of vulnerability. Even if there are guidelines for protecting patient data in health facilities, due to law regulations, these are frequent target of computer attacks.

For example, 90% of health care organizations and partner companies that responded to a Ponemon survey in 2016 reported having experienced a data breach, and 64% reported a violation involving the leak of patient medical records (Ujibashi et al., 2016).

Recently technologies have emerged the context of genomic data with the aim of improving access to data, patient empowerment and interoperability.

For example, the use of Cloud Computing technology for the storage and sharing of genetic data, which allows the scaling of genetic data flexibly and ease of use for users (Yang, 2019).

However, activities such as registration and access control in the exchange of genetic information have been administered in an unauthorized manner, mainly through entities that manage the eligibility of users with genomic information and allow access to specific datasets, which shows a lack of harmonization in access policies between the owners of their genetic information and the health entities that manage it.

Adopting an "authorized" structure using Blockchain technology, which limits access to a patient's data, will allow transparency in these transactions and, in addition, the owner of that genomic information to manage and manage it.

Although the metadata that the datasets describe is available to everyone on the system, that does not mean that the data stored in Blockchain is readable to everyone and this is because Blockchain "Is based on pseudo anonymity and public key infrastructure (PKI), allowing Blockchain content to be encrypted in a prohibitive way to decrypt".

Implementing such technology could enable detection of existing datasets, while protecting people's privacy by restricting access to approved users. We propose a technological model for the management of genetic information security with Blockchain tech. Our contributions are as follows:

- We define a technological model, which allows to carry out activities such as obtaining the appropriate information, in the right way, for the right person, at the right time, in the appropriate place.
- The components used ensure transparency and data integrity, as the participants of the chain are known and there would be no chance of falsification or theft of information.
- We use the Ethereum Framework for data processing with low resource consumption.

This paper is organized as follows. Section II describes the relevant concepts and definitions related to this document. Section III presents our contribution. Section IV discusses Related Works. Section V shows the analysis of the results obtained. Finally, section VI shows the conclusions and perspectives.

2 BACKGROUND

This section provides the concepts that will be applied in the development of the model to be proposed for the protection of genetic information using Blockchain technology in the Private Health Sector.

Definition 1 (Blockchain (Tripathi et al., 2020))**.** *Blockchain is defined as a distributed and immutable digital ledger that provides data transparency and user privacy.*

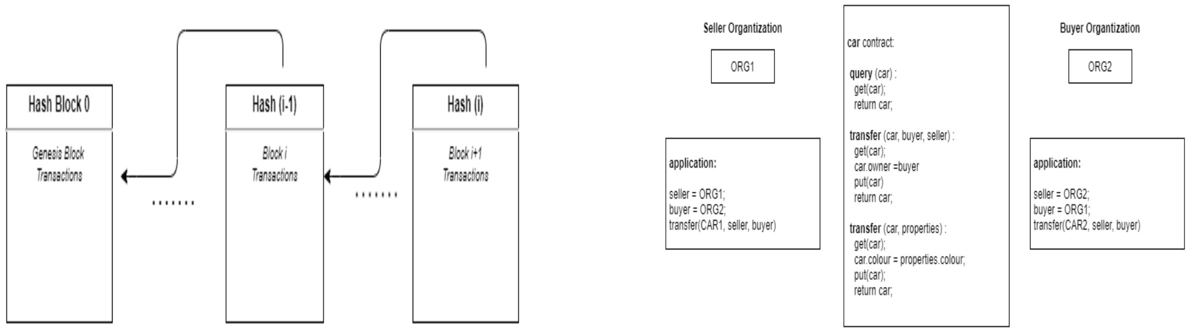
This concept is derived from bitcoin and is based on cryptography and Peer-to-Peer (P2P) networks, in which data from a given structure is organized into blocks and organized into a data chain in chronological order in a chain structure forming a blockchain (Cheng et al., 2020).

Likewise, there is no central authority and the digital ledger is shared among all peers for all to see. While it is true that this concept was initially linked to financial transactions, technological advances and its enormous benefits have made this technology part of the various areas where record security is of an important nature (Tripathi et al., 2020).

Example 1. *The generation of each block occurs by generating the genesis block or block "0", and the corresponding transactions. Each block in this structure has a unique identifier called a hash, and this is generated each time the contents of the block are changed, which the header of the previous block becomes the hash of the next block, as shown in Figure 1a. This block interconnect provides the security feature to the blockchain, as any modification to the data in each block requires updating hash values throughout the chain, making it impossible to manipulate or hack information (Tripathi et al., 2020).*

Definition 2 (Encryption (Benil and Jasper, 2020))**.** *At Blockchain, encryption is responsible for ensuring the integrity of the information and the anonymity of the owner of that information using algorithms.*

This technology is accompanied by a hash function, which allows new data to be added to the block.



(a) Building a Block (b) Flow of a Smart Contract.

Figure 1: Blockchain Structure.

Specifically, any changes that are made to the ledger such as adding a new transaction must be recalculated (Benil and Jasper, 2020). The use of digital signatures allows confirmation that the recipient received the transactions and that the sender is not an imposter (Benil and Jasper, 2020).

Definition 3 (Smart Contract (Ozercan et al., 2018))**.** *In Blockchain, Smart Contracts are a set of instructions that have special storage for information related to the developing application.*

This storage has a record that is distributed as a copy to all nodes in the network (Zhang et al., 2018). This allows, for example, the ability to automate the use of notifications in health applications, which generates better integration between devices (Chen et al., 2019). In addition, it allows an analysis of medical data, which allows to activate alerts about certain unusual activities (Griggs et al., 2018).

Example 2. *As shown in Figure 1b, two Org1 and Org2 organizations are observed, which define a Smart Contract for consulting, transferring and updating cars*[3].

Definition 4 (Blockchain in the Health Sector (McGhin et al., 2019))**.** *The health sector has data security and privacy requirements due to the existence of legal standards in the protection of patient information established by governments.*

Malicious attacks that compromise the integrity of such information due to internet use in the exchange of patient information (Khezr et al., 2019) are now a matter of concern. Therefore, the requirements that Blockchain technology conceives in the health sector are such as the exchange and transfer of medical and genomic data, as well as authentication and interoperability (McGhin et al., 2019).

[3]Smart Contracts and Chaincode - https://bit.ly/3kvgwG2

3 MAIN CONTRIBUTION

Now we describe our proposal fo the Blockchain-based technology model for the protection of the genetic information, as shown in Figure 2.

3.1 Roles

For the design of the technological model, the roles that have been raised are:

- *Generator Agent* with the following activities:
  - **Generation:** The Generator Agent is responsible for recording the genetic information of the patient who has given him the registration permit, so that the information recorded in the blockchain ensuring traceability.
  - **Authorization/Validation:** The Generator Agent is authorized to add genetic information from the patient.
- *Consumer Agent* with the following activities:
  - **Distribution:** The Consumer Agent will be able to consult the genetic information of the patient registered in the system
  - **Authorization/Validation:** The Consumer Agent is authorized to consult and visualize the genetic information of the patient
- *Patient* with the following activities:
  - **Generation:** The patient grants registration permission to the Generator Agent so that the generator can add their genetic information.
  - **Authorization/Validation:** The patient grants the respective permissions as a Generator Agent or Consumer Agent so that they can manage their genetic data
  - **Distribution:** The patient grants the consultation permission to the Consumer Agent so that he/she can consume their genetic information

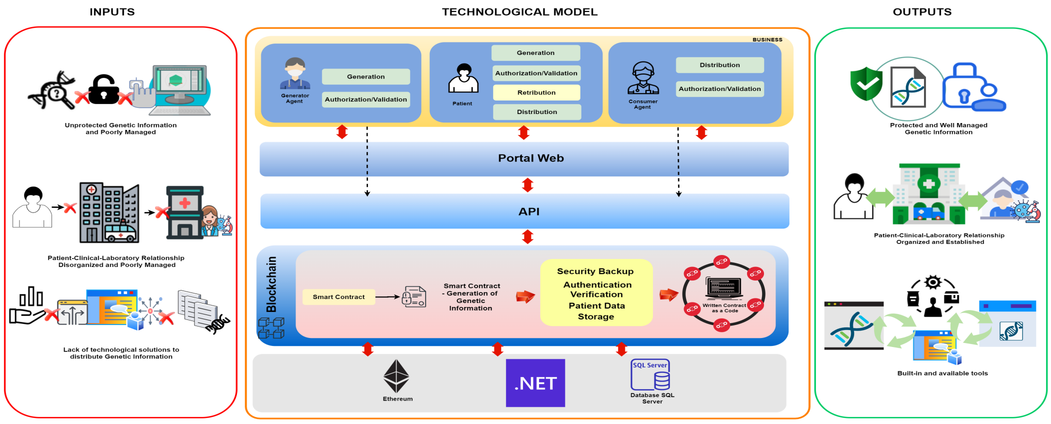


Figure 2: Proposed Technology Model.

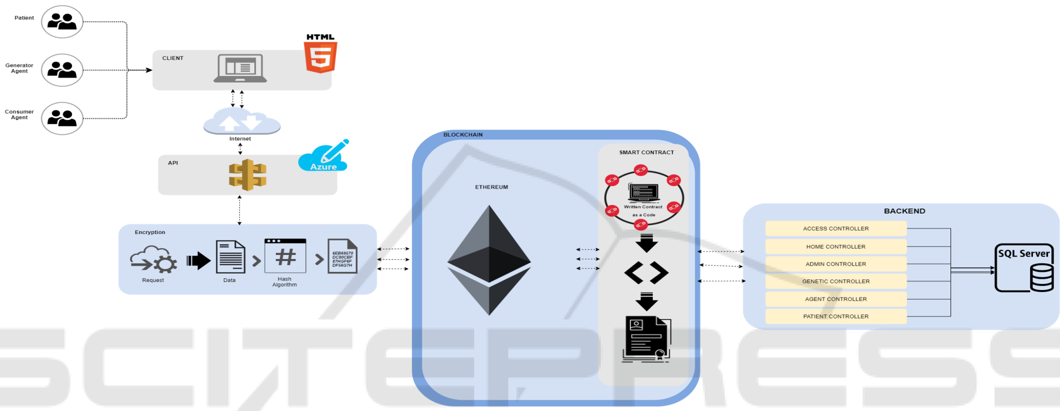


Figure 3: Web Portal Architecture.

3.2 Web Site

The model will have the representation of a web portal in order to show the registration of the genetic information of the patient and the control of accesses of this information by the same actor. The technologies involved in developing this application are HTML5 by the client view, ASP.Net for the back end, SQL Server as the database manager, and the Microsoft Azure Cloud platform.

The Blockchain part used the Ethereum Framework for the development of the Genetic Information Generation Smart Contract, which each time such information is recorded, a transaction is generated which contains the block ID, the transaction creation date, and the cryptographic hash generated. This can be seen in the architecture of the portal in Figure 3.

3.3 Blockchain Platform

For transactions made in the development of the model, the Blockchain platform acts as the technology with the ability to record transactions made in the web application and transfer it to a network of nodes where this network is worth the transaction. Moreover, each transaction has a cryptographic hash in order to validate the claims that people make against the tracked and managed assets in the blockchain, which in this case is genetic information.

4 RELATED WORKS

Now, we describe the related works of our proposal.

In (Fu et al., 2020), the authors propose a process of light mechanism for preserving the privacy of medical records, which is composed of the creation and storage of medical information. Unlike this, our proposal adds the process of distribution of information.

(Kuo et al., 2020) propose a security storage model for medical data based on registration and authentication. Our proposal contrarily gives access control by the owner of the information and the process of distribution of such information.

(Kulemin, 2017) propose a Blockchain-based architecture of the Zenome system, in which the roles

Table 1: Expert Survey Data.

	Validation Date	Expert	Recording Link	Survey Link	Answers
Blockchain	06/10/2020	Expert 1	https://bit.ly/3kVQFYr	https://bit.ly/31uVArq	https://bit.ly/3mqyoCV
	07/10/2020	Expert 2	https://bit.ly/34sRTEL	https://bit.ly/31uVArq	https://bit.ly/3mqyoCV
Laboratory	06/10/2020	Expert 1	https://bit.ly/3dYeYCs	https://bit.ly/3dOrwMv	https://bit.ly/3kTXeuB
	09/10/2020	Expert 2	https://bit.ly/37CukLH	https://bit.ly/3dOrwMv	https://bit.ly/3kTXeuB

defined in their system are 3: Person, Data Consumer and Service Provider. However, our proposal involves other access to the person's information as the literature mentions levels of privacy of such information, and in our case is granted directly to the other roles.

(Murugan et al., 2020) propose a medical data exchange solution using Blockchain technology, this solution encompasses processes such as healthcare and data exchange. However, our proposal covers both the process of recording information and distributing

Most of these works cover the process of data logging and storage, but they do not cover distribution or access control, or in reverse, if they attack processes such as distribution or access control do not cover processes such as data logging and storage.

Therefore, with our proposal we seek to contemplate the processes globally both the generation and distribution of genetic information.

5 EXPERIMENTS

Now, we test scenarios to validate the feasibility of our proposal.

5.1 Experimental Protocol

The tools used for the experiments are as follows:

- Visual Studio 2017 with C#
- Microsoft SQL Server Management Studio 18
- Microsoft Azure with "App Services"
- Ethereum

Also, the entire project both the code developed for proof-of-concept that validates the functionalities of the model and the data registered in the database is publicly available at the following link: https://bit.ly/37HAcDk

In addition, virtual presentations have been made for the validation of the proof of concept in which the project proposal was explained to the experts selected through the Zoom platform. After the presentations made, surveys were submitted which aimed to seek feedback from the experts and the user.

For this we have segmented surveys into different groups that make references to expert groups, Blockchain Expert and Laboratory Expert, and to the user.

1. Blockchain Expert Survey: The Blockchain Expert survey was divided into 13 questions aimed at receiving feedback based on your Blockchain experience and feedback on our proposal.

   Two Blockchain experts have also been selected.

2. Laboratory Expert Survey: The survey to be conducted by the Laboratory Expert was divided into 14 questions whose objective is to receive feedback at the functional level of the Model and comments towards our proposal.

   Two laboratory experts are also selected.

3. User Survey: A survey was conducted on the user, who represents the patient in our model.

   For the purposes of our proposal we have a sample of 40 people, which belong to an age range of 20 to 30 years.

   All these sample users are resident in the city of Lima Metropolitana and genetic testing has been done at least once.

   Some of the questions to be answered by them are: "What level of importance do you think should be given to the security of your genetic information?" and "At what level would you be willing to share your genetic information with others?" This survey is divided into 9 questions and the user survey link is as follows: https://bit.ly/2TRl5iA

Table 1 gives the links of the Survey given to the Blockchain and Laboratory experts, the link of the recordings made, the date of the sessions and the platform on which the presentation was made.

Similarly, statistical data collected by INEN were obtained, which can be seen in Table 2, which writes the average genetic consultation time, which is estimated at 60 minutes ranging from patient identification, genetic testing registration and diagnosis.

In addition, Figure 5 shows a statistic collected by INEN4 describing the increase in genetic information

[4]Chief Resolution (in spanish) N 490-2019-J - https://bit.ly/2RQ0k5X

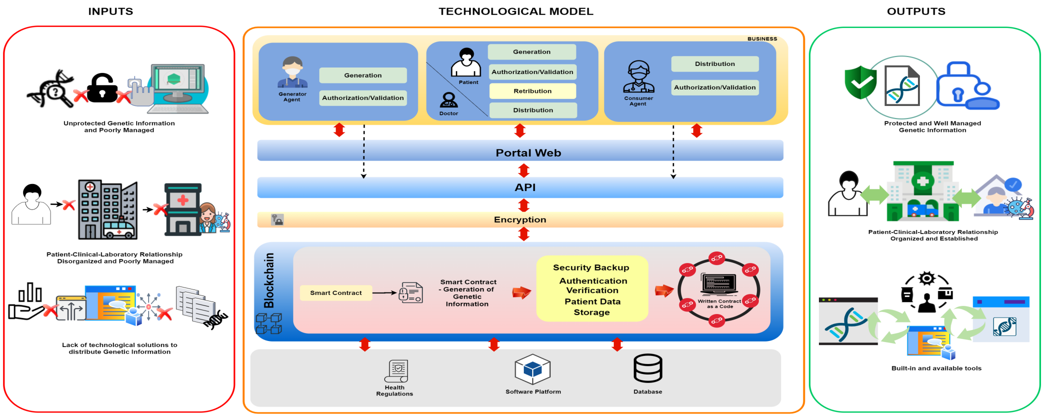


Figure 4: Final Proposal for the Technology Model.

Table 2: Average genetic consultation time report[4].

Medical Genetics Consultation	Average Time (minutes)
Patient Admission	5
Medical Consultation	5
Exams	40
Genetic Pretest Counseling	10
Total Average time	60

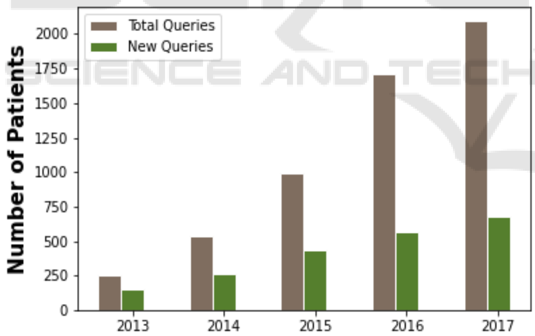


Figure 5: Relationship between total consultations per year and new consultations of the Medical Genetics Office[4].

consultation transactions from 2013 to 2018 achieving an average of 450 consultations per year and identifying a gap between supply and demand.

5.2 Experimental Results

On the one hand, Blockchain experts agree that protecting genetic information transparently and using Blockchain as a technology to ensure the management and assurance of this asset is a correct choice. They also agree with the idea that the patient owns and can manage their genetic information. However, they indicated that additional encryption processes should be established to the model in order to achieve better assurance of genetic information, to define the components of the model, since, being a model, these can be reused for later architectures and solutions.

On the other hand, laboratory experts indicated that they agree with the project presented as it facilitates the registration and access of genetic information and safeguards it with Blockchain technology.

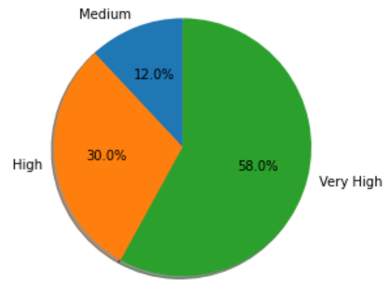
They also found the need identified by the vulnerabilities that occur frequently within health systems correct, and the design of the model is simple and easy to understand, the roles managed are sufficient and prefer a web environment in terms of proof of concept for management issues in health facilities.

Whereas they emphasized that not always the patient can manage their information since depending on the clinical condition that he presents would need the support of a doctor for his management and that a classification of genetic information can be established as paternity studies, autoimmune diseases, degenerative diseases, among others.

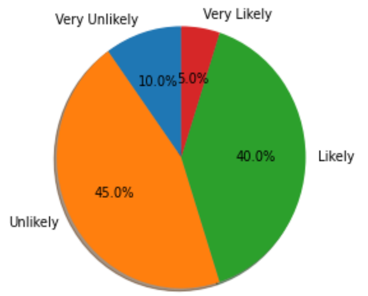
Based on the comments and recommendations provided by the experts, the following components were added in Figure 4:

- The encryption layer, which adds additional encryption processes to provide greater security.
- A “Health Regulations” component, which aims to detail that the Legal Rules regarding the use of patient information are considered.
- A “Doctor” role, which assume the patient’s activities according to the patient’s clinical picture.

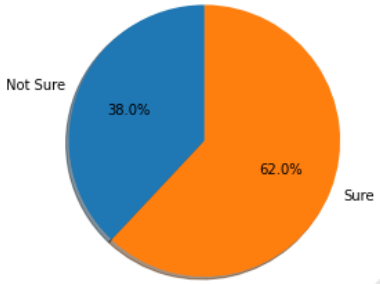
From the users surveyed, Figure 6a shows that 58% rate as “Very High” the level of importance to be considered with the security of their genetic information. Figure 6b also shows that 45% of all respondents define “Unlikely” in sharing their genetic information



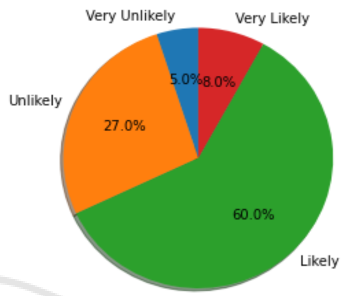
(a) Security importance level of genetic information



(b) Level of sharing of genetic information with other people



(c) Perception of having genetic information over the Internet



(d) Level of willingness to share genetic information with our proof of concept

Figure 6: Comparison of users' perceptions.

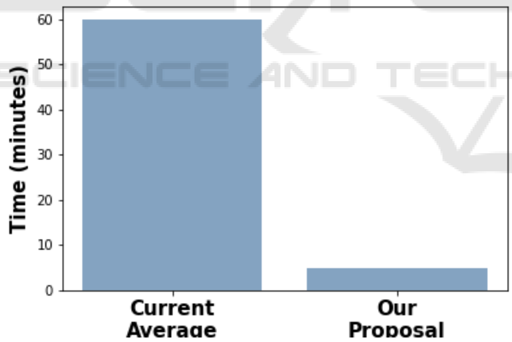


Figure 7: Comparison of average genetic consultation time.

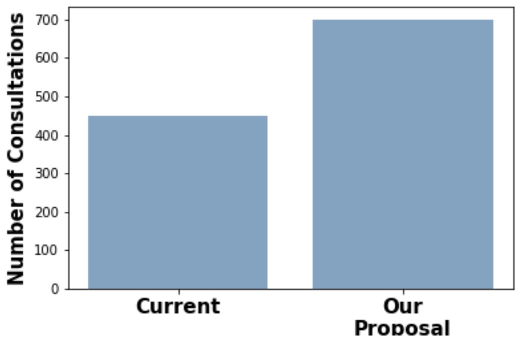


Figure 8: Comparison of average genetic consultation time.

with others. And this is supported by Figure 6c, as the percentage of respondents who do not feel confident that their genetic information is on the internet reaches 63% of the total. However, by using our proof of concept, 60% would be willing to share your information, this can be viewed in Figure 6d.

Similarly, we have the metric of the average time of genetic consultations which with our proposal we seek to reduce the time of this metric, for which a comparison was made with the current time and the average time offered by the proposal.

As seen in Figure 7, the current average time is 60 minutes, while our proposal's average time is 5 minutes, reducing the current average time by 92%.

Nevertheless, another metric that is handled is that of the number of genetic queries. As shown in Figure 8, the current number of consultations is 450, while our proposal is 700, which is increased by 56%.

6 CONCLUSIONS AND PERSPECTIVES

In conclusion, we have shown that Blockchain technology is effective and efficient, because, having an average genetic consultation time of 60 minutes, with this technology allows us to have it in 5 minutes and save the time of genetic consultations by up to 92%. In addition, it allows us to increase the number of genetic information queries, as the number of current average queries is 450, while with the use of this technology a total of 700 is estimated, which improves by 56% and users are much more willing to share their genetic information if they have a platform that guarantees the traceability and protection of their data. However, it would be interesting to be able to work with geolocalisation (Cueva-Sánchez et al., 2020), which would allow to classify the genetic information of the patient in order to provide a better diagnosis according to a region or using a chatbot to detect the interact with the users (Solis-Quispe et al., 2021).

REFERENCES

Benil, T. and Jasper, J. (2020). Cloud based security on outsourcing using blockchain in e-health systems. *Comput. Networks*, 178:107344.

Chen, Y., Ding, S., Xu, Z., Zheng, H., and Yang, S. (2019). Blockchain-based medical records secure storage and medical service framework. *J. Medical Syst.*, 43(1):5:1–5:9.

Cheng, X., Chen, F., Xie, D., Sun, H., and Huang, C. (2020). Design of a secure medical data sharing scheme based on blockchain. *J. Medical Syst.*, 44(2):52.

Cueva-Sánchez, J. J., Coyco-Ordemar, A. J., and Ugarte, W. (2020). A blockchain-based technological solution to ensure data transparency of the wood supply chain. In *2020 IEEE ANDESCON*, pages 1–6.

Fu, J., Wang, N., and Cai, Y. (2020). Privacy-preserving in healthcare blockchain systems based on lightweight message sharing. *Sensors*, 20(7):1898.

Griggs, K. N., Ossipova, O., Kohlios, C. P., Baccarini, A. N., Howson, E. A., and Hayajneh, T. (2018). Healthcare blockchain system using smart contracts for secure automated remote patient monitoring. *J. Medical Syst.*, 42(7):130:1–130:7.

Khezr, S., Moniruzzaman, M., Yassine, A., and Benlamri, R. (2019). Blockchain technology in healthcare: A comprehensive review and directions for future research. *Applied Sciences*, 9(9).

Kulemin, N. (2017). The zenome project : Whitepaper blockchain-based genomic ecosystem.

Kuo, T., Kim, J., and Gabriel, R. A. (2020). Privacy-preserving model learning on a blockchain network-of-networks. *J. Am. Medical Informatics Assoc.*, 27(3):343–354.

McGhin, T., Choo, K. R., Liu, C. Z., and He, D. (2019). Blockchain in healthcare applications: Research challenges and opportunities. *J. Netw. Comput. Appl.*, 135:62–75.

Murugan, A., Chechare, T., Muruganantham, B., and Kumar, S. (2020). Healthcare information exchange using blockchain technology. *International Journal of Electrical and Computer Engineering (IJECE)*, 10:421.

Ozercan, H. I., Ileri, A. M., Ayday, E., and Alkan, C. (2018). Realizing the potential of blockchain technologies in genomics. *Genome research*, 28 9:1255–1263.

Shabani, M. (2019a). Blockchain-based platforms for genomic data sharing: a de-centralized approach in response to the governance problems? *J. Am. Medical Informatics Assoc.*, 26(1):76–80.

Shabani, M. (2019b). Blockchain-based platforms for genomic data sharing: a de-centralized approach in response to the governance problems? *J. Am. Medical Informatics Assoc.*, 26(1):76–80.

Solis-Quispe, J. M., Quico-Cauti, K. M., and Ugarte, W. (2021). Chatbot to simplify customer interaction in e-commerce channels of retail companies. In *International Conference on Information Technology & Systems*, pages 1–6.

Tripathi, G., Ahad, M. A., and Paiva, S. (2020). S2hs- a blockchain based approach for smart healthcare system. *Healthcare*, 8(1):100391.

Ujibashi, Y., Kawaba, M., and Harada, L. (2016). Proposal of a database type and aggregation function for accelerating medical genomics study on RDBMS. In *EDBT*, pages 672–673. OpenProceedings.org.

Yang, J. (2019). Cloud computing for storing and analyzing petabytes of genomic data. *Journal of Industrial Information Integration*, 15:50 – 57.

Zhang, P., White, J., Schmidt, D. C., Lenz, G., and Rosenbloom, S. T. (2018). Fhirchain: Applying blockchain to securely and scalably share clinical data. *Computational and Structural Biotechnology Journal*, 16:267 – 278.