Order number:
Serial number:

**PEOPLE'S DEMOCRATIC REPUBLIC OF ALGERIA**
**Ministry of Higher Education and Scientific Research**

**UNIVERSITY OF ECHAHID HAMMA LAKHDAR - EL OUED**
**FACULTYOFEXACT SCIENCES**
**Computer Science Department**

**Thesis**
**Submitted in partial fulfillment of the requirements for the degree of**

# ACADEMIC MASTER

Field: **Mathematics and Computer Science**
Option: **Computer Science**
Specialty: **Distributed Systems and Artificial Intelligence**


Presented by:

- **LEHMIDI Oum elhana**
- **BOULIFA Manel**


# Title

# A new direction of secure data representation: medical domain as case of study

Defended on : 07\06\2023

**Examination Committee:**

| | | | |
|---|---|---|---|
| M. | Abdelnacer KHALAIFA | MCA | President |
| M. | Abdelkader LAOUID | MCA | Supervisor |
| M. | M$^{ed}$ lamine YAGOUB | MCB | Examiner |

Academic year: 2022-2023

# THANKS

First and foremost, we thank God Almighty who guided us to his book and by which we were able to complete this humble work.

We also extend our sincere thanks and gratitude from the bottom of our hearts to

LAOUID Abdelkader

who helped us, directed us, and treated us like a dear brothers, we wish to her happiness and success in various fields.

We sincerely thank all the members of the jury who accepted the verdict on our humble work, and all who helped us from near or far : our teachers, our families, our friends, our colleagues all named.

# Dedication

The journey was not short, and it should not have been. The dream was not close, and the road was not easily paved. But I did it.

I dedicate my graduation to the one whose name I proudly carry, to the one who cleared the thorns from my path and paved the way for knowledge, to my beloved father.

After the grace of Allah, everything I have achieved is attributed to my father, the man who didn't receive even a small part of what we have accomplished, and the man who strived throughout his life for us to become better than him.

To the invisible hand that removed the thorns from my way, and to the one who endured every moment of pain I went through and supported me in my weakness and failures, to my dear mother.

To my siblings and companions of the years, and to everyone who has been a support and pillar on this journey, I am grateful to all of you. I wouldn't have reached this point without your grace and assistance, after the grace of Allah Almighty.

# Dedication

The journey was not short, and it should not have been. The dream was not close, and the road was not easily paved. But I did it.

I would like to dedicate this to all my family and friends, classmates, and most especially for my family someday and I pray to cod I hope to pass together all my classmates for I Assessment this coming school year. And I would like to think to all my instructors and most especially Supervisor and think again for sacrifice teaching me and understand to all my classmates and my try my best Assessment and even I'm schooling this is because I want to learn and have knowledge and skills for my future someday.

Allah Bless and Good luck to all my classmates.

I wouldn't't have reached this point without your grace and assistance, after the grace of Allah Almighty.

Study Hard Before Your Love!!

**Abstract**

The generation of large quantities of data poses significant challenges, particularly in the realm of modern information technology. One area where technology is rapidly being adopted to enhance patient care is the medical field.

In the present day, digital data is produced by numerous inexpensive devices and stored in various formats. This necessitates careful consideration of data processing, transmission, and storage. By employing an appropriate format for representing data, several advantages can be gained at both the node (handling and storage) and arch (transport) levels. While blockchain has emerged as a prominent storage technology, it still encounters issues, particularly with regards to its rapidly expanding volume.To address this concern, a new approach to data representation is proposed, offering multiple advantages. The concept involves converting a binary matrix I, sized m × n bits, into two vectors: ho and ve, with sizes m′ and n′ respectively. The compression rate, expressed as (m × n)/(m′ + n′), exhibits exponential growth, highlighting the remarkable efficiency of the proposed technique in reducing data size. The original data can be restored using ho, ve, and Hsh(I). The conversion from I to (ho, ve) is simple, resulting in improved power consumption for low-power devices.

The challenge of recovering the original data can be leveraged through two operations, such as determining mining compatibility based on which node retrieves a predefined set of vectors. Additionally, the verification of data integrity will be limited to nodes with significant computational power.

**keywords:** Data representation; Blockchain; Resource computation variety; Distributed computing.

**Résumé**

La génération de grandes quantités de données pose des défis importants, en particulier dans le domaine des technologies de l'information modernes. Un domaine où la technologie est rapidement adoptée pour améliorer les soins aux patients est le domaine médical.

De nos jours, les données numériques sont produites par de nombreux appareils peu coûteux et stockés dans différents formats. Cela nécessite une réflexion minutieuse sur le traitement, la transmission et le stockage des données. En utilisant un format approprié pour représenter les données, plusieurs avantages peuvent être obtenus à la fois au niveau des nœuds (gestion et stockage) et au niveau de l'architecture (transport). Bien que la technologie de la blockchain soit devenue une technologie de stockage importante, elle rencontre encore des problèmes, notamment en ce qui concerne son volume en expansion rapide. Pour remédier à cette préoccupation, une nouvelle approche de représentation des données est proposée, offrant plusieurs avantages. Le concept consiste à convertir une matrice binaire I, de taille m × n bits, en deux vecteurs : ho et ve, de tailles respectives m′ et n′. Le taux de compression, exprimé comme (m × n)/(m′ + n′), présente une croissance exponentielle, mettant en évidence l'efficacité remarquable de la technique proposée dans la réduction de la taille des données. Les données d'origine peuvent être restaurées à l'aide de ho, ve et Hsh(I). La conversion de I en (ho, ve) est simple, ce qui permet une consommation d'énergie améliorée pour les appareils à faible puissance.

Le défi de récupération des données d'origine peut être relevé grâce à deux opérations, telles que la détermination de la compatibilité minière en fonction du nœud qui récupère un ensemble prédéfini de vecteurs. De plus, la vérification de l'intégrité des données sera limitée aux nœuds disposant d'une puissance de calcul significative.

**Mots-clés:** Représentation des données, Blockchain, Diversification des ressources, Informatique distribuée.

**الملخص**

تواجه توليد كميات كبيرة من البيانات تحديات هامة، خاصة في مجال التكنولوجيا المعلوماتية الحديثة. واحدة من المجالات التي يتم اعتماد التكنولوجيا فيها بسرعة لتحسين رعاية المرضى هو مجال الرعاية الطبية.

في الوقت الحالي، يتم إنتاج البيانات الرقمية من قبل العديد من الأجهزة الرخيصة ويتم تخزينها في تنسيقات مختلفة. وهذا يتطلب تفكيرًا دقيقًا في معالجة ونقل وتخزين البيانات. باستخدام تنسيق مناسب لتمثيل البيانات، يمكن الحصول على العديد من المزايا على مستوى العقد (الإدارة والتخزين) وعلى مستوى البنية (النقل). على الرغم من أن تكنولوجيا البلوكشين أصبحت تكنولوجيا تخزين هامة، إلا أنها لا تزال تواجه مشاكل، بما في ذلك زيادة حجمها بسرعة. للتغلب على هذه المخاوف، يتم اقتراح نهج جديد لتمثيل البيانات يقدم العديد من المزايا. يتمثل الفكرة في تحويل مصفوفة ثنائية I بحجم $m \times n$ بت إلى نوعين من البيانات: ho و ve بأحجام $m'$ و $n'$ على التوالي. نسبة الضغط، المعبر عنها بواسطة $(m + n) / (m' \times n')$، تظهر نموًا متسارعًا، مما يبرز الكفاءة الملحوظة للتقنية المقترحة في تقليل حجم البيانات. يمكن استعادة البيانات الأصلية باستخدام ho و ve و Hsh(I). عملية تحويل I إلى (ho, ve) بسيطة، مما يتيح استهلاك طاقة محسن للأجهزة ذات الطاقة المنخفضة.

يمكن التغلب على تحدي استعادة البيانات الأصلية من خلال عمليتين، مثل تحديد التوافق المعدني بناءً على العقد الذي يسترد مجموعة محددة مسبقًا من النواتج. بالإضافة إلى ذلك، ستكون عملية التحقق من سلامة البيانات محدودة بالعقد التي تحتوي على قدرة حسابية كبيرة.

**الكلمات الرئيسية:** تمثيل البيانات، البلوكشين، تنويع الموارد، الحوسبة الموزعة.

# Contents


**1 State of Art** **12**
- 1.1 Information technology . . . 12
- 1.2 Information security . . . 12
- 1.3 Medical information security . . . 13
- 1.4 Features medical information security . . . 14
- 1.5 Challenges of medical information security . . . 14
- 1.6 Limitation of medical information security . . . 15

**2 Background** **16**
- 2.1 Introduction . . . 16
- 2.2 Related work . . . 16
  - 2.2.1 Motivation . . . 18
- 2.3 Medical data . . . 19
  - 2.3.1 Types of medical images . . . 19
  - 2.3.2 Body temperature measurements . . . 21
  - 2.3.3 Blood glucose levels for diabetic patients . . . 22
  - 2.3.4 Cholesterol levels . . . 22
  - 2.3.5 Allergies and adverse reactions to medications . . . 23
  - 2.3.6 Medication dosage and frequency information . . . 23
  - 2.3.7 Family medical history . . . 23

**3 The proposed technique** **25**
- 3.1 The idea behind the suggested method for representing data . . . 25
  - 3.1.1 Description of the model that is being suggested . . . 25
  - 3.1.2 Proposed model: recuperation . . . 26
  - 3.1.3 Current magnitude of the blockchain . . . 27

**4 Discussion and analysis** **29**
- 4.1 Introduction . . . 29
- 4.2 Development environment : . . . 29
  - 4.2.1 Hardware environment : . . . 29
  - 4.2.2 Software environment : . . . 30
  - 4.2.3 Used data . . . 32
- 4.3 Implementation . . . 32
  - 4.3.1 Implementation of the proposed method . . . 32
  - 4.3.2 Blockchain Size Comparaison . . . 36
- 4.4 Results analysis . . . 36

**General Conclusion** **42**

# List of Figures


1 Internet of Things application . . . . . 10

2.1 X-ray image [19] . . . . . 20
2.2 Ultrasound image[21] . . . . . 20
2.3 CT scan image . . . . . 21
2.4 Magnetic Resonance Imaging . . . . . 21
2.5 Continuous Glucose Monitor (CGM) . . . . . 22

3.1 Simplified block representations: basic and ours, where $ho_i$ denotes the number of 1 in row $i$ $(tx : i)$ and $ve_i$ denotes the number of 1 in column $i$ . . . . . 27

4.1 Python Logo . . . . . 30
4.2 visual studio Logo[32] . . . . . 31
4.3 blockchain Logo[34] . . . . . 31
4.4 Medical pictures used . . . . . 32
4.5 Blockchain . . . . . 36
4.6 The main interface of the application . . . . . 37
4.7 Stages of code execution . . . . . 38
4.8 The size of the blockchain after storing the data . . . . . 39
4.9 Blockchain containing Matrix and Vector data . . . . . 39
4.10 Content Size of Image Compared to Binary Matrix and Vector . . . . . 40
4.11 The total number of possible matrices . . . . . 41
4.12 Plot of Complexity:O(nlog(n)) . . . . . 41

# Introduction General

The IoT is a network of physical objects that includes home appliances, vehicles, and devices that are designed with sensors, software, and network connectivity. This enables them to communicate and interact with each other without the involvement of humans. The growth rate of IoT is rapid, and it is expected that there will be more than 22 billion IoT devices in operation by 2025.[1]

The key advantage of IoT is its potential to have a significant impact on different aspects of our daily lives and environment, such as smart spaces. This can result in enhancements in various domains, such as healthcare, energy, security, safety, industry, environment, and entertainment. As shown in Figure 1.1, these domains can be positively impacted by IoT technology.[2]


Figure 1: Internet of Things application

With the proliferation of IoT devices and their collection of massive data, ensuring information security becomes an essential concern. The vulnerability of IoT devices to cyberattacks poses significant risks, encompassing not only data breaches but also potential physical harm.

The diversity and heterogeneity of IoT devices present a challenge in terms of securing them. These devices come from various manufacturers, adopt different protocols, and serve different purposes. Consequently, securing the IoT ecosystem necessitates a multi-layered approach that encompasses device security, communication security, and data security.

To tackle these challenges, several security mechanisms and protocols have been suggested. These include device authentication, data encryption, access control, and intrusion detection. Moreover, leveraging machine learning and artificial intelligence techniques can aid in identifying abnormal behavior and detecting potential security threats.

In essence, information security is of paramount importance for the Internet of Things, requiring a multi-faceted strategy to account for the diverse and heterogeneous nature of IoT devices. By implementing a combination of security mechanisms, protocols, and advanced technologies such as machine learning and artificial intelligence, effective solutions can be devised to safeguard the IoT

ecosystem[3]. Among these solutions is the contemporary blockchain technology, which serves as a decentralized and tamper-proof digital ledger, ensuring data integrity and continuity by securely recording transactions or information across multiple nodes or computers through cryptographic methods[4].

In this thesis a

To sum up, the main contributions in this work are summarized as follows:

- In **the first Chapter** In the initial chapter, we provided a definition of medical information security along with an exploration of its benefits and the key challenges it faces.
- In **the second Chapter**We have provided an overview of previous works in the extensive field of information security, giving a glimpse into some of the research and advancements made in this area.
- In **the third Chapter** In this section, we will discuss crucial data in the field of medicine and delve into the concept underlying the suggested approach for representing this data.
- And in **the final chapter** we will discuss the practical application of the proposed approach and present the outcomes of the conducted experiment.

# Chapter 1

# State of Art

## 1.1 Information technology

Informatics plays a crucial role in the fields of Information Technology (IT) and Computer Science (CS), providing significant benefits to professionals in these domains. These benefits include[5]:

- Improved Efficiency: Informatics can help automate processes, reduce manual errors, and streamline workflows in IT and CS, leading to increased productivity and efficiency.
- Better Decision-making: Informatics provides professionals with real-time data and analytic that can help them make informed decisions about software design, functionality, and user experience.
- Enhanced Quality: Informatics can help improve the quality of IT and CS products and services, ensuring that they are reliable and secure
- Increased Accessibility: Informatics can help make IT and CS products and services more accessible to users with disabilities, enabling them to use software and digital services with ease.
- Collaboration: Informatics facilitates collaboration among various stakeholders in IT and CS, enabling teams to work together effectively and efficiently to achieve common goals.
- Innovation: Informatics drives innovation in IT and CS by enabling the development of new products, services, and technologies that can enhance productivity, improve user experience, and solve complex problems[5].

In today's highly competitive and rapidly changing technological landscape, professionals in IT and CS must leverage informatics to stay ahead of the curve and deliver high-quality products and services to their clients and users. By embracing informatics, professionals in these fields can achieve greater efficiency, better decision-making, enhanced quality, increased accessibility, collaboration, and innovation, leading to improved outcomes and greater success.

## 1.2 Information security

Informatics, which is also known as computer science, is a vast and constantly evolving field that is essential in today's world. It involves the study of computer hardware, software, and their applications in various industries[6]. Some of the areas of use of informatics are:

- Business and finance: Many businesses and financial institutions rely heavily on computer systems to process and analyze data. Informatics is used to design, develop, and maintain software applications that help these organizations manage their finances and operations efficiently. Education: Informatics is used in the development of educational software and applications, as well as in the management of educational resources and data. It has also made distance learning possible, allowing students to learn remotely.
- Entertainment: The entertainment industry relies heavily on computer technology to create and distribute media content. Informatics is used in the development of video games, animation, and other digital media.
- Manufacturing: Informatics is used in the design and development of computer-aided manufacturing systems, which automate the production process and increase efficiency.
- Transportation: The transportation industry uses informatics to develop and manage systems for traffic control, logistics, and navigation. It has also made possible the development of autonomous vehicles.
- Communications: Informatics is used in the development of communication systems, such as email, instant messaging, and social media. It has also made possible the development of video conferencing and other remote communication technologies.
- Medical: The healthcare industry is rapidly embracing technology to improve patient care and outcomes. Informatics is used in the development of electronic medical records, health information exchange systems, and medical imaging technologies, among others[6].

All the information should be confident

## 1.3 Medical information security

Medical information security refers to the protection of sensitive patient information, such as medical records, treatment plans, and personal identifying information, from unauthorized access, use, disclosure, or modification. In the healthcare industry, protecting patient information is critical for maintaining patient privacy, ensuring the integrity of medical records, and complying with legal and regulatory requirements[7].

Some of the key aspects of medical information security include:

- Access control: Limiting access to patient information to only authorized individuals and implementing measures such as passwords, encryption, and two-factor authentication to prevent unauthorized access.
- Data encryption: Encrypting sensitive patient information during transmission and storage to prevent it from being intercepted or stolen.
- Security policies and procedures: Establishing policies and procedures for managing patient information and ensuring that all employees are trained on these policies and procedures.
- Network security: Implementing firewalls, intrusion detection systems, and other security measures to protect against unauthorized access to the healthcare organization's network.
- Risk management: Conducting regular risk assessments to identify vulnerabilities and implementing measures to mitigate those risks.

- Incident response: Developing a plan for responding to security incidents, including reporting and investigating incidents and implementing measures to prevent future incidents[7]

Medical information security is crucial for protecting patient privacy, maintaining trust in the healthcare system, and complying with legal and regulatory requirements. Healthcare organizations must take appropriate measures to safeguard patient information and ensure that it is only accessible to authorized individuals.

## 1.4 Features medical information security

Medical information security involves several features to ensure that sensitive patient information is protected from unauthorized access, use, disclosure, or modification. Here are some of the key features of medical information security[8]:

Confidentiality: Medical information security requires that patient information is kept confidential and only accessed by authorized individuals. This can be achieved through access controls, such as passwords, encryption, and two-factor authentication.

Integrity: Medical information security ensures that patient information is accurate and has not been modified or altered. This can be achieved through data validation, audit trails, and version control.

Availability: Medical information security ensures that patient information is available when needed. This can be achieved through redundancy and backup systems.

Risk management: Medical information security involves identifying and managing risks to patient information. This can include conducting regular risk assessments and implementing measures to mitigate those risks.

Compliance: Medical information security must comply with legal and regulatory requirements, such as HIPAA (Health Insurance Portability and Accountability Act) in the United States. Compliance requires implementing appropriate safeguards, policies, and procedures to protect patient information.

Incident response: Medical information security requires a plan for responding to security incidents, such as data breaches or unauthorized access. This includes reporting incidents, investigating them, and taking appropriate actions to prevent future incidents.

These features are essential to ensuring that sensitive patient information is protected and that healthcare organizations maintain the trust of their patients. By implementing appropriate measures, healthcare organizations can ensure that patient information remains confidential, accurate, and available when needed.

## 1.5 Challenges of medical information security

Medical information security faces several challenges, which can make it difficult to protect sensitive patient information effectively. Here are some of the key challenges of medical information security[9]:

Complexity: Medical information systems are complex, involving multiple stakeholders, interconnected systems, and diverse technologies. This complexity can make it difficult to implement and maintain effective security measures.

Human error: Human error, such as unintentional disclosure of information or failure to follow security protocols, can compromise the security of patient information.

Insider threats: Insider threats, such as employees or contractors with access to patient information who misuse or disclose it, can be difficult to detect and prevent.

Cybersecurity threats: Medical information systems are vulnerable to cyber threats, such as hacking, malware, and ransomware attacks, which can compromise patient information and disrupt healthcare operations.

Third-party risks: Medical information is often shared with third-party vendors, such as electronic health record providers, medical device manufacturers, and telemedicine platforms. These vendors may have their security vulnerabilities, making them vulnerable to data breaches.

Compliance: Compliance with legal and regulatory requirements, such as HIPAA in the United States, can be challenging and complex, requiring healthcare organizations to implement specific security measures, policies, and procedures.

Addressing these challenges requires a comprehensive and holistic approach to medical information security, including implementing technical, administrative, and physical safeguards, conducting regular risk assessments, providing ongoing training and education to staff, and regularly reviewing and updating security policies and procedures. Healthcare organizations must also stay informed of emerging threats and trends in medical information security to adapt their security measures accordingly[9].

## 1.6 Limitation of medical information security

Despite efforts to implement effective medical information security measures, there are several limitations and challenges that can make it difficult to protect patient information adequately. Here are some of the key limitations of medical information security[10]:

Limited resources: Healthcare organizations may have limited resources to invest in medical information security, such as staffing, funding, and technology, which can compromise the effectiveness of security measures.

Technological limitations: Medical information systems may have technological limitations, such as outdated or unsupported software, which can make it difficult to implement effective security measures.

User behavior: User behavior, such as the failure to follow security protocols or poor password hygiene, can compromise the security of patient information.

Insider threats: Insider threats, such as employees or contractors with access to patient information who misuse or disclose it, can be challenging to detect and prevent.

Evolving threats: Cybersecurity threats are continually evolving, and healthcare organizations may struggle to keep up with the latest threats and trends in medical information security.

Privacy concerns: Patients may be reluctant to share their medical information due to privacy concerns, which can limit the effectiveness of medical information security measures.

Legal and regulatory limitations: Legal and regulatory limitations, such as the lack of standardized security requirements across different jurisdictions, can make it challenging to implement effective medical information security measures.

Addressing these limitations requires a multifaceted approach that involves investing in resources, technology, and staff training to ensure that healthcare organizations can implement effective security measures. Healthcare organizations must also stay informed of emerging threats and trends in medical information security and be willing to adapt their security measures accordingly. Finally, healthcare organizations must work to address patient concerns about privacy and ensure that patients understand the importance of medical information security.

# Chapter 2

# Background

## 2.1 Introduction

Medical information security pertains to shielding confidential patient data within healthcare systems. This encompasses protecting electronic health records, preserving the anonymity of patients and their medical histories, and ensuring the accuracy and accessibility of medical information. The safety of medical data is fundamental to ensure the privacy of patients, prevent identity theft, and sustain the confidentiality and confidence between healthcare providers and patients. Physical and technological safety measures, including access control, encryption, backup, and recovery systems, are implemented. Furthermore, adhering to regulations like HIPAA[11] is crucial in ensuring the security of medical information. In the medical sector, there exists an extensive literature on privacy-preserving technologies for IoT and AI systems. Scholars have introduced various technologies and models that aim to address privacy concerns, including:cryptosystem[12] , Blockchain[13] [14] [15], privacy techniques to a blueprint architecture and efficient implementation[16] .

## 2.2 Related work

- **A Decentralized Privacy-Preserving Healthcare Blockchain for IoT:**

  The proposed blockchain model for IoT devices is specifically designed to do away with the concept of proof-of-work. Speakers relied on the distributed nature of the network and additional security features to ensure its suitability. In their blockchain, transactions are broadcast publicly and contain additional information about both the sender and recipient. To address concerns about user anonymity and authenticity, a lightweight privacy-preserving ring signature scheme was introduced for anonymous transactions by authenticated users. The scheme used a lightweight digital signature to ensure data integrity and loop signature to allow anonymous signing.

  To provide additional security, double encryption of data with lightweight encryption algorithms (ARX ciphers) and generic ciphers has been used. The data is first encrypted with a symmetric encryption key and then the symmetric key itself is encrypted with a public key. ARX ciphers have been found to be particularly suitable for IoT devices because they use simple arithmetic operations such as modular addition, bit-level rotation, and exclusive OR. In addition, the Diffie-Hellman key exchange technology was used for the secure exchange of cryptographic keys over a public channel.

  The proposed architecture provided a solution [13] to the major security and privacy threats

faced by IoT devices. A hybrid approach that combines private and public key cryptography with blockchain technology and other lightweight cryptographic core elements to develop patient-centered access control for electronic medical records has been introduced. Although open questions have been raised regarding potential attacks such as DoS and modding attacks, IoT resource limitations have remained a major challenge in addressing these issues.

- **A privacy-preserving cryptosystem for IoT E-healthcare** Their paper[12] proposes a new image encryption cryptosystem that provides fast and secure protection for patients' privacy. The system ensures that medical data is kept secure during dissemination to healthcare centers. A new PRNG algorithm is also introduced, which uses two chaotic maps to produce the encryption keys. To enable real-time processing of wireless capsule endoscopy videos, they developed a summarization method that utilizes integral-image based features. The proposed cryptosystem employs a block cipher encryption method with one round of confusion and diffusion operations. To enhance the security of the system, their encryption algorithm is probabilistic, producing different encrypted data each time even with the same secret key. Their proposed system has been tested through simulations and security analysis, demonstrating its robustness against various attacks. The system has a large key space, making exhaustive attacks infeasible. Overall, the system reduces energy consumption, communication bandwidth, specialist analysis time, and searching efforts while maintaining the security and privacy of patient data. Their proposed PRNG algorithm is one of the first to be based on the Zaslavsky system, to the best of their knowledge.

- **Privacy preserving Internet of Things: From privacy techniques to a blueprint architecture and efficient implementation**

  The focus of the paper [16] is to introduce innovative techniques that can ensure end-to-end privacy and security in the next generation of IoT systems. The approach involves randomizing the collected data from IoT devices and expressing it as a sum of multiple numbers, where each addend is stored in a separate data store that only keeps one component number. This requires the use of several data stores, which can be located in a server or in the cloud. Additionally, the paper [16] proposes the implementation of a homomorphic encryption scheme that enables secure access to the IoT data by aggregating the addends, without exposing sensitive information to hackers or data store administrators. Even if all but one of the data stores are compromised, the IoT data remains protected. Overall, these techniques can be an effective solution for safeguarding sensitive IoT sensor data from unauthorized disclosure and hacking.

- **Block Summarization and Compression in Bitcoin Blockchain** This research paper[14] proposes a solution to address the issue of blockchain size, specifically in transferable entity transactions. The solution involves summarizing and compressing the blockchain's blocks, which reduces the block size and the number of blocks that needs to be downloaded. This simplifies payment verification for a node and improves blockchain scalability. However, there is a challenge to implementing this solution, such as creating an algorithm for an adaptive value of that can handle different numbers of blocks, establishing a standard for block summarization and compression that optimizes space saving, and developing a standard summary block form for use in various blockchain platforms.

  Despite this challenge, the proposed solution holds promise, and further research and development is necessary to overcome this challenge and make it more widely applicable to different blockchain platforms.

- **Lightweight blockchain consensus mechanism and storage optimization for resource constrained IoT devices**

  The article [15] describes the experimental findings of two proposed algorithms for blockchain systems: a consensus algorithm and a storage algorithm. These algorithms have been optimized to reduce latency and overhead. The experiments conducted investigate the impact of various factors, such as block size and block partition number, on storage space consumption, block recovery delay, and encoding speed. The proposed consensus algorithm is found to outperform other algorithms in terms of throughput, delay, and communication times, meaning that it can process more transactions in less time with less overhead. The proposed storage algorithm is also shown to reduce the average storage footprint of the blockchain by up to 48.14compared to traditional storage methods. The experiments demonstrate that larger blocks and more block divisions result in faster encoding of data, but longer recovery times in case of block failure.

  The article [15] introduces two proposed algorithms: the SPBFT consensus mechanism and the RS-BSOS storage optimization strategy, which are designed for lightweight blockchain systems. The SPBFT consensus mechanism evaluates the status of each node and selects only the consensus nodes to participate in the consensus process, thereby reducing the computing and communication burden on resource-constrained devices. The RS-BSOS storage optimization strategy encodes each block into multiple segments using RS erasure code technology, which is distributed across appropriate blockchain nodes. This reduces the storage overhead and threshold for resource-constrained devices. Overall, these proposed algorithms show promising results for improving the performance and efficiency of blockchain systems.

- **Recycling Smart Contracts: Compression of the Ethereum Blockchain** The authors of this work [17] have demonstrated that significant space savings can be achieved on Ethereum smart contracts by reusing bytecode stored on the Ethereum blockchain. They found that smart contracts often reuse similar bytecode structures, enabling them to achieve space savings of 70or more. They also propose a method that can be used in conjunction with other methods, such as zlib, to achieve space savings of over 90Although their compression method has a slow compression time, they were able to perform it on a laptop within the 14-second average block time of Ethereum. By introducing a new pseudo opcode to the Ethereum virtual machine, the authors can find a previous smart contract with the same sequence of operations and replace them with a pointer to the previous smart contract. This new pseudo opcode will indicate the presence of a pointer to the EVM and where to retrieve the data. The authors ensure that their pointers are pointing to a bytecode sequence longer than their pseudocode, which reduces the size of their smart contract. Therefore, their compression method is described as the recycling of already used bytecode, which results in significant space savings.

### 2.2.1 Motivation

In previous works aimed at improving information security, shortcomings or negative aspects have been observed that hinder the process.Such as when Ali Yavari and all[16] presented a paper titled "Preserving Privacy for the Internet of Things: From Privacy Technologies to Scheme Architecture," which discusses an IoT system that safeguards privacy by using data generated from IoT devices. According to the preliminary evaluations, the suggested technologies have little impact on the overall performance of the system. Nonetheless, the process is slow, and there seems to be a disregard for storage capacity.

Khan Muhammad and all In their work [12] , this is what we have come to notice Experimental results may not be generalizable to other types of medical data or IoT systems.

Dr. Gautam Srivastava and colleagues [13] proposed the adoption of a double encryption system. . Rather, it means that the encryption technique employed could potentially add complexity and prolong the time required for encryption and decryption operations. Furthermore, if the recipient's public key is compromised, the entire encryption process could be rendered useless. While Diffie-Hellman key exchange is considered a secure means of transferring public keys, it may still be vulnerable to attacks by sophisticated attackers.

Some techniques involve creating summary blocks and compressing them to increase space savings and make it easier for nodes to verify transactions. The proposed method in [14] is applied to the Bitcoin blockchain, and the space-saving is calculated by comparing the original block size with the summarized and compressed block sizes. The experiment results show that the space-saving for summary blocks is 22.318 while the space-saving for compressed summary blocks is 78.104

There is also a research on compression ethereum by modify the smart contract by using using pointing the new smart contract to the previous one which same [15]. Previous research has been done on block summarization for the customized Bitcoin blockchain on the light node, but this research still faces challenges, namely the summary block size is still same as the original block size.

In [17] , a compression technique for Ethereum smart contracts is proposed by the authors that can achieve high compression ratios without a slow decompression time. However, there are some limitations and potential drawbacks associated with it. Firstly, the use of low-level functions like connect, delegate, and call token may compromise Solidity type security and hinder network performance, making it less commonly used. Secondly, the compression algorithm takes longer to stabilize and only achieves stable space savings of around 75in the middle of the data set.

Therefore, we will try to suggest a way to solve some of the problems faced by previous works[16][12][13][14][15][17], which is a way to compress huge data in the blockchain, which helps in its transfer, use, and ease of storage, especially for devices with low power.

## 2.3 Medical data

In healthcare, data encompasses a range of patient-related information such as medical history, lab results, vital signs, imaging studies, medication records, allergies, and more. In the following sections, we will provide a simplified explanation of some of these categorie.

### 2.3.1 Types of medical images

Medical images play a crucial role in contemporary healthcare, offering a non-invasive means for medical professionals to examine the internal body parts. The imaging techniques utilized vary depending on the specific purpose of the examination, whether it is to detect tumors, visualize organs, monitor disease progression or identify broken bones. There are different medical imaging methods available, each designed to cater to specific applications, including X-ray imaging, computed tomography (CT), magnetic resonance imaging (MRI), ultrasound, functional nuclear medicine imaging, tactile imaging, magnetic brain mapping, and thermography, among others.

✓ **X-rays :** are a form of high-energy electromagnetic radiation that can penetrate through objects, including the human body. They are commonly used in medical imaging to visualize internal structures and capture tissue details. As the X-rays pass through the body, they create shadows that represent the structures within[18].

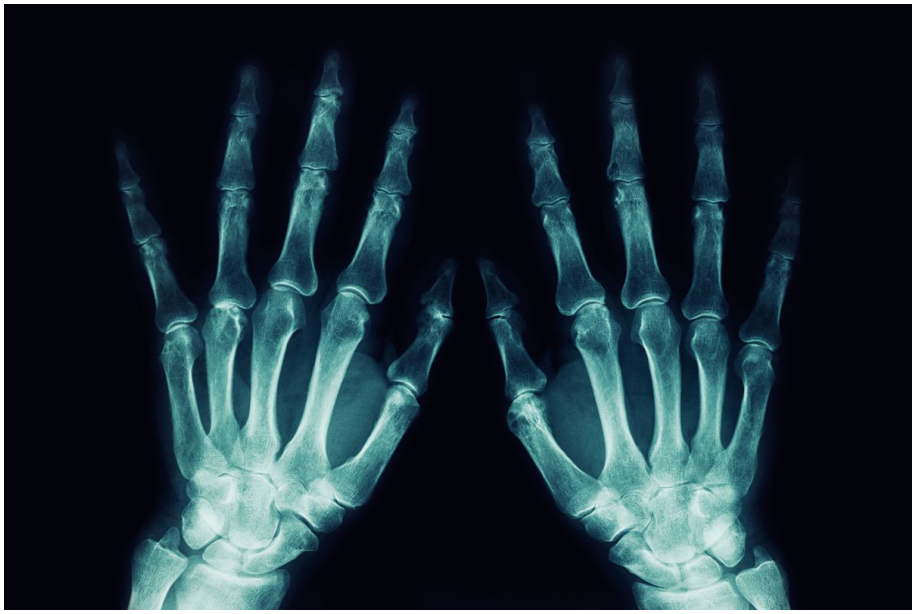

Figure 2.1: X-ray image [19]

✓ **Medical ultrasound :** has two main types: diagnostic and therapeutic. Diagnostic ultrasound is used to create images of the inside of the body. On the other hand, therapeutic ultrasound uses high-frequency sound waves that are beyond the range of human hearing. Unlike diagnostic ultrasound, therapeutic ultrasound does not produce images. Instead, its purpose is to interact with tissues in the body to modify or destroy them[20].


Figure 2.2: Ultrasound image[21]

✓ **Computerized tomography**(CT)A scan involves capturing multiple X-ray images of the body and using computer processing to generate cross-sectional images of bones, blood vessels, and other structures[22]. This method provides more precise and detailed information compared to traditional X-rays.

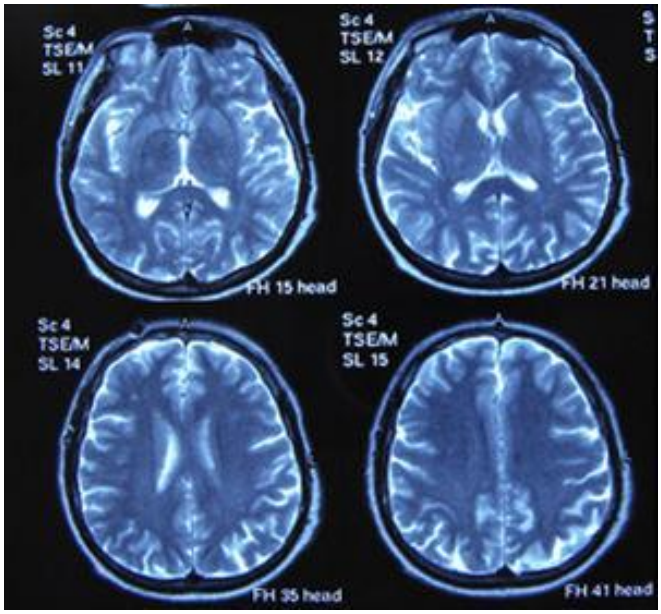


Figure 2.3: CT scan image

✓ **Magnetic Resonance Imaging**(MRI) is a type of non-invasive medical imaging technology that is commonly utilized for the purposes of disease diagnosis, treatment monitoring, and the creation of highly detailed three-dimensional anatomical images[23].

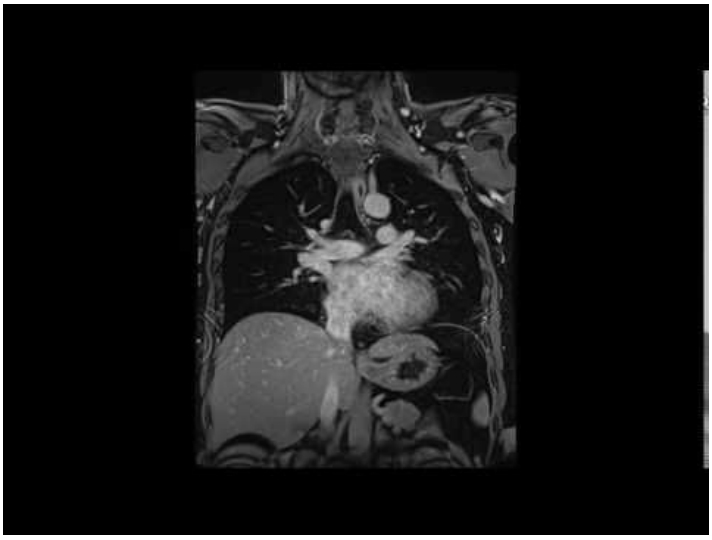

Figure 2.4: Magnetic Resonance Imaging

### 2.3.2 Body temperature measurements

Body temperature is a significant indicator of a person's health status, and abnormal readings can suggest an underlying illness or infection. The normal range of body temperature is 97.7°F to 99.5°F (36.5°C to 37.5°C), which may vary depending on various factors such as age, time of day, and physical activity levels. Body temperature can be measured using different methods, including oral, rectal, ear, axillary, and temporal artery measurements. In healthcare settings, measuring body temperature is an essential part of routine check-ups, disease diagnosis, and monitoring patients' conditions.

It is also crucial to record body temperature in non-healthcare settings to identify fever during outbreaks or epidemics. Therefore, including body temperature measurements in patient information is crucial[24].

### 2.3.3 Blood glucose levels for diabetic patients

Medical data related to blood glucose levels is crucial for patients with diabetes, a chronic disease that affects how the body processes blood sugar. Diabetic patients require frequent monitoring of their blood glucose levels to manage their condition effectively. The normal range for blood glucose levels varies, but in general, it should be maintained between 70-130 mg/dL before meals and less than 180 mg/dL after meals[25].

There are different ways to measure blood glucose levels, including using a glucose meter to test a drop of blood from a finger prick, or through continuous glucose monitoring systems that use a small sensor placed under the skin Figure 2.5. Blood glucose levels can be affected by factors such as food intake, physical activity, stress, medications, and illness. Healthcare providers use blood glucose level measurements to monitor diabetic patients' control over their blood sugar levels, assess the effectiveness of diabetes medications, and make necessary adjustments to treatment plans. Accurate and timely blood glucose level monitoring and recording is essential for proper diabetes management and prevention of long-term complications.

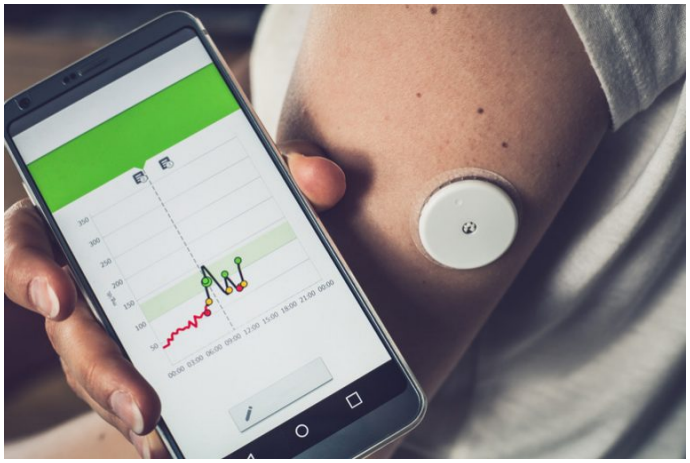

Figure 2.5: Continuous Glucose Monitor (CGM)

### 2.3.4 Cholesterol levels

Cholesterol measurement is an important component of medical data that can provide insights into a patient's risk of developing heart disease. Cholesterol is a fatty substance that is essential for the body's function, but high levels of cholesterol in the blood can increase the risk of developing heart disease, stroke, and other cardiovascular problems.

Cholesterol levels can be measured through a blood test that measures the levels of different types of cholesterol in the blood, including low-density lipoprotein (LDL) cholesterol, high-density lipoprotein (HDL) cholesterol, and triglycerides. A healthy total cholesterol level is less than 200 milligrams per deciliter (mg/dL), while LDL cholesterol levels should be below 100 mg/dL, and HDL cholesterol levels should be above 60 mg/dL[26].

Medical professionals use cholesterol measurement data to assess a patient's risk of developing heart disease and to develop treatment plans to reduce the risk. Treatments may include lifestyle changes such as diet and exercise, medication to lower cholesterol levels, and regular monitoring of

cholesterol levels. It is important to include cholesterol measurement data in a patient's medical record to ensure comprehensive and effective management of cardiovascular health.

### 2.3.5 Allergies and adverse reactions to medications

Allergies and adverse reactions to medications are important pieces of medical information that should be included in a patient's medical record. Adverse reactions can range from mild symptoms such as rashes and headaches to severe, life-threatening conditions such as anaphylaxis. Anaphylaxis is a severe allergic reaction that can cause difficulty breathing, a drop in blood pressure, and can be fatal if not treated immediately.

Some of the most dangerous allergies and adverse reactions to medications include anaphylaxis, Stevens-Johnson syndrome, and toxic epidermal necrolysis. Stevens-Johnson syndrome and toxic epidermal necrolysis are severe skin reactions that can cause extensive damage to the skin and mucous membranes. They are often caused by medications and can be life-threatening.

It is important for healthcare providers to be aware of a patient's allergies and adverse reactions to medications to prevent adverse drug reactions and ensure patient safety. Patients should inform their healthcare providers of any known allergies or adverse reactions to medications they have experienced in the past, and should also report any new symptoms or reactions they experience while taking medication[27].

### 2.3.6 Medication dosage and frequency information

Medication dosage and frequency information are crucial pieces of medical data that help healthcare providers ensure that patients receive the appropriate treatment for their conditions. Medications must be prescribed at the correct dosage and frequency to achieve the desired therapeutic effect while minimizing the risk of adverse reactions or toxicity.

Medical data related to medication dosage and frequency can be collected through various methods, such as patient self-reporting, medication reconciliation, or electronic health records (EHRs). It is important for patients to accurately report their medication use, including over-the-counter and herbal remedies, to their healthcare provider to avoid potential drug interactions or adverse events[28].

Healthcare providers use this information to adjust medication dosages based on patient factors such as age, weight, and kidney or liver function. They may also modify medication frequency to achieve the desired therapeutic effect or reduce the risk of adverse reactions.

Incorrect medication dosage or frequency can have serious consequences, such as treatment failure, adverse drug reactions, or even death. It is important for patients to follow their prescribed medication regimen and to inform their healthcare provider if they experience any adverse effects.

Overall, medication dosage and frequency information is essential to providing safe and effective medical care to patients.

### 2.3.7 Family medical history

Family medical history is an important type of medical data that can provide valuable information about a patient's risk for certain health conditions. It includes information about diseases or conditions that have affected members of a patient's family, such as parents, siblings, grandparents, and aunts or uncles. This information can help healthcare providers assess a patient's risk for certain conditions and develop an appropriate care plan.

Family medical history can be collected through various methods, such as patient interviews, medical records, and family pedigree charts. Common conditions that may be included in a

family medical history include heart disease, stroke, diabetes, high blood pressure, certain types of cancer, and genetic disorders. This information can be particularly useful in identifying inherited conditions or diseases that may run in families[29].

Healthcare providers can use family medical history to recommend appropriate screenings or tests, make lifestyle recommendations to reduce risk, and develop personalized treatment plans. It is important for patients to provide accurate and up-to-date family medical history information to their healthcare providers, as it can have a significant impact on their health outcomes.

# Chapter 3

# The proposed technique

## 3.1 The idea behind the suggested method for representing data

We provide an elaborate description of a proposed method that aims to represent data in a highly compact manner. The node applies the proposed technique to convert the data matrix and subsequently transmits it. The conversion process is both rapid and straightforward, irrespective of the matrix size, and the output is considerably smaller than the original data size.

### 3.1.1 Description of the model that is being suggested

The proposed model introduces a novel approach to efficiently reduce the size of information by using two vectors, namely the horizontal vector *ho* and the vertical vector *ve* , to replace the matrix. For instance, given a matrix $I$ with $m$ rows and $n$ columns, the vector *ho* contains m values, where the first value denotes the number of "1"s in the first row of the matrix, the second value indicates the number of "1"s in the second row, and so on, until the last value, which represents the number of "1"s in the last row of the matrix. Likewise, the second vector *ve* consists of n values that correspond to the number of "1"s in each column of the matrix I. Specifically, the first value of vector *ve* indicates the number of "1"s in the first column of the matrix, the second value represents the number of "1"s in the second column, and so on, until the last value, which denotes the number of "1"s in the last column of the matrix.

The current method of representing the information is inadequate for retrieving it from the cloud. This is because there can be two distinct matrices M and N that have identical vector representations, meaning $he_M = he_N$ and $ve_M = ve_N$. To address this issue, the proposed model includes the hash of the matrix as a unique identifier for the data, where $Hsh_M \neq Hsh_N$. The following example provides a clear demonstration of this concept.

$$M = \begin{bmatrix} 1 & 0 & 0 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 0 & 0 & 1 \\ 0 & 1 & 1 & 1 \end{bmatrix} \tag{3.1}$$

with $ho_M = (2, 4, 2, 3)$; $ve_M = (3, 2, 2, 4)$.

and

$$N = \begin{bmatrix} 0 & 0 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 0 & 0 & 1 \\ 1 & 1 & 0 & 1 \end{bmatrix} \tag{3.2}$$

with $ho_N = ho_M = (2, 4, 2, 3)$; $ve_N = ve_M = (3, 2, 2, 4)$.

### 3.1.2 Proposed model: recuperation

Once a matrix $I$ has been represented as described in the previous subsection (Subsection 3.1.1), the cloud can use its available computing resources to recover the original matrix l based on the vectors $ho_I$, $ve_I$, and $Hsh_I$. The current algorithm used in this proposed model involves an exhaustive search of all possible matrix configurations. This search begins with a matrix of zeros with $m$ rows and $n$ columns, and the potential configurations are tested based on the number of ones in the $ho_I$and $ve_I$ vectors. For each candidate matrix $I_i$, the cloud computes its hash value and compares it to the provided $Hsh_I$. Algorithm1 provides a general outline of this process.

**Algorithm 1** Recover matrix algorithm

**Require:** $ho_I, ve_I, Hsh_I$
**Ensure:** $matrixI$

```
1: function RecI
2:     I_0 ← zeros
3:     for for each possible composition of the number of 1 do
4:         form I_i
5:         compute ho_i
6:         compute ve_i
7:         compute Hsh_i
8:         if ho_i = ho_I and ve_i = ve_I and Hsh_i = Hsh_I then
9:             I ← I_i
10:            exit
11:        end if
12:    end for
13:    return I
14: end function
```

Let's now discuss the number of feasible matrices (NPI) that the algorithm 1 can derive from the vectors $ho$ and $ve$.

**Lemma 1.** *If* $size(I) = n \times m$ *then* $NPI_{n,m} = \prod_{i=1}^{m} NPR_i$ *where* $m$ *is the number of rows and* $NPR_i$ *denotes number of possibilities in the row* $i$.

*We give* $NPR_i = \alpha \times (\alpha + 1) \times (2 \times \alpha + 4)/12$ *where* $\alpha = n - x + 1$

*Proof.* The *RecI* algorithm 1 will not try all possibilities from 'zeros' to 'ones'; in this case, the number of possibilities is $2^{n \times m}$. This is because the number of possible values for the first line is $2^n$ if the first line size is equal to $n$, and if the matrix has $m$ rows, then $NPI = 2^n \times 2^n \ldots 2^n = 2^{n \times m}$.

The *RecI* algorithm 1 will exploit the number of 'ones' in each line to reduce the total number of possibilities, thus greatly reducing the number of trials.

Suppose the number of "1" is $x$ with $x < n$ where $n$ is the size of a row; the first possibility is to position the $x$ bits of "1" successively (e.g. 11100000000 where $x = 3$ and $n = 11$). The next possibility is to move the last "1" (11010000000; then 11001000000 ..) etc.

Therefore, we have $(n - x + 1)$ possibilities. Next, we move the "1" before the last one and do the same with the last "1" as the first step (10110000000; 10101000000; 10100100000), so there are $(n - x + 1) - 1$ possibilities.

In the third step, we do the same thing with the first "1", we will finally find that $NPR$ is equal to the sum of the series $1 + (1 + 2) + (1 + 2 + 3) \cdots + (1 + 2 + 3 + 4 + \cdots + \alpha)$ where $\alpha$ is $(n - x + 1)$ and $x$ denotes the number of "1" in the first row.

We have $s_i = \sum_{i=1}^{n} i = i \times (i + 1)/2 = (i^2 + i)/2$.

Therefore, $NPR = \sum_{i=1}^{\alpha} s_i = 1/2 \times (\sum_{i=1}^{\alpha} i^2) + \sum_{i=1}^{\alpha} i$.

$NPR = 1/2 \times (\alpha \times (\alpha + 1) \times (2 \times \alpha + 1)/6) + \alpha \times (\alpha + 1)/2$.

So, $NPR = \alpha \times (\alpha + 1) \times (2 \times \alpha + 4)/12$.

□

For example, consider a matrix $I$ of size $8 \times 8$. In an absolutely exhaustive search without using $H$ and $V$, the number of possibilities in the first line $NPR_1 = 2^8$; so, $NPI = 2^8 \times 2^8 \ldots 2^8 = (2^8)^8 = 2^{64} = 18446744073709551616$.

In our proposal using $H$ and $V$ and according to Lemma 1, if we assume that the number of "1" equals 4, $x = 4$ (the average), this gives $\alpha = 8-4+1 = 5$ and $NPR = 5\times(5+1)\times(2\times5+4)/12 = 35$; so $NPM = 35^8 = 2251875390625$.

### 3.1.3 Current magnitude of the blockchain

One of the key uses of the proposed technique is in Blockchain, which is presently the foremost data storage system and represents the direction in which modern technology is progressing due to its many benefits. If we view Blockchain as consisting of a cluster of blocks, where each block comprises a set of transactions, we can conceive of these transactions as a data matrix in which each row represents one or more transactions

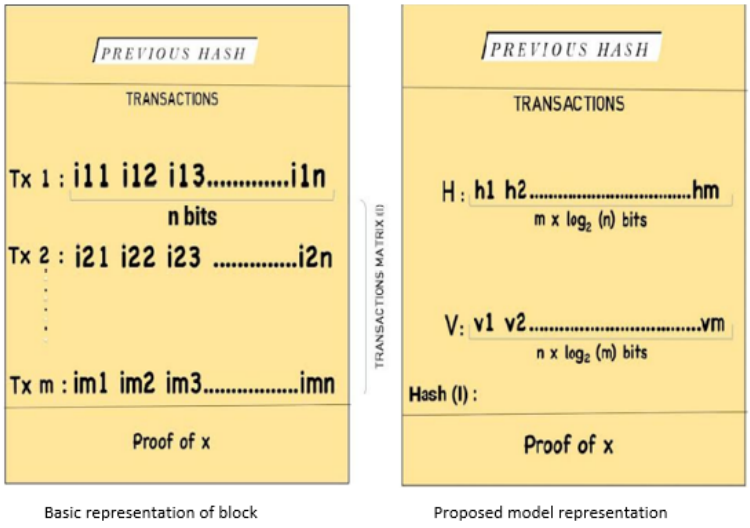


Figure 3.1: Simplified block representations: basic and ours, where $ho_i$ denotes the number of 1 in row $i$ $(tx : i)$ and $ve_i$ denotes the number of 1 in column $i$

Figure Illustrated in the figure is the matrix layout (Basic representation of block) comprising $m$ transactions, with each transaction having a size of $n$ bits; $i_{\alpha,\beta} = \{0, 1\}, \alpha \in [1, m]$ and $\beta \in [1, n]$. Using the proposed method, the block representation changes as shown, as depicted on the right-hand side of Figure (Representation of the proposed model). Since $h_i$ denotes the number of "1" in row $i$ and the size of row $i$ equals $n$ bits, we need $log_2(n)$ bits to represent each $h_i$.

$$\mid ho \mid = m \times log_2(n) \tag{3.3}$$

and

$$\mid ve \mid = n \times log_2(m) \tag{3.4}$$

and

$$\mid I \mid = \mid ho \mid + \mid ve \mid + \mid Hsh_I \mid \tag{3.5}$$

In case any party wishes to access the transactions, they can be obtained by computing from $ho$, $ve$, and $Hsh_I$."

# Chapter 4

# Discussion and analysis

## 4.1 Introduction

Our proposal is a method of compressing big data in the blockchain, which helps in its transfer, use, and ease of storage, especially for devices with low power. Given the large data size of the blockchain, we claim it appropriate to exploit the proposed model for this pioneering storage technology. To have a more comprehensive view, we will discuss in the next section the implementation of the proposed idea.

## 4.2 Development environment :

### 4.2.1 Hardware environment :

To bring our proposed idea to fruition, we used the computer as our primary tool, taking advantage of its good characteristics that supported our work. We will mention the specifications of the device:

- Processor: intel (R) Core (TM) i7-8650U CPU@ 1.90GHz 2.11GHz
- Installed RAM: 16.0GB (15.4GB usable)
- Device ID : 1772284E-0A13-453C-9Ce3-36A183EF40F0
- Product ID :00331-10000-00001-AA615
- System type: 64-bit operating system,x64-based processor

### 4.2.2 Software environment :

**Python**

In our idea, we use the Python programming language as the main implementation language. Python is a popular and versatile programming language known for its simplicity, readability, and extensive ecosystem of libraries and frameworks [30].

Figure 4.1: Python Logo

Here are some of the main reasons why we chose Python for our project:

- Large community and libraries: Python has an extensive and active community of developers, which means there is a wealth of resources, documentation, and support available. In addition, Python offers a rich ecosystem of libraries and modules that speed up development by providing pre-built functions for various tasks.
- Versatility: Python is a versatile language that can be used for various applications, including blockchain development. It supports multiple programming paradigms, such as object-oriented, procedural, and functional programming.
- Integration: Python excels in its integration capabilities, enabling seamless integration with other languages, systems, and platforms. This is useful in blockchain development, where there may be a need for interoperability with existing systems and technologies.
- Solid Foundation of Blockchain Libraries: Python strongly supports blockchain development through numerous libraries and frameworks. These include libraries such as Web3.py for interacting with Ethereum-based blockchains, PyTezos for the Tezos blockchain, and others for developing smart contracts and creating decentralized applications (dApps).

**Visual Studio**

Visual Studio is a versatile and powerful Integrated Development Environment (IDE) that caters to software development needs, including Python programming. With its extensive toolset, Visual Studio offers developers a seamless experience for writing, debugging, and testing Python code. It simplifies the development process by providing code completion, syntax highlighting,

and integrated debugging features. Moreover, Visual Studio facilitates the integration of various Python libraries and frameworks, empowering developers to harness the vast ecosystem available for Python. [31]


Figure 4.2: visual studio Logo[32]

In essence, Visual Studio is a widely embraced IDE that supports Python development, along with other programming languages, and equips developers with a comprehensive set of features and tools to boost productivity and streamline the overall development workflow.

**Blockchain**

A blockchain is an distributed ledger consisting of a series of records, or blocks, that are interconnected using cryptographic hashes. Each block includes a cryptographic hash of the previous block, a timestamp, and transaction data organized in a Merkle tree structure. The inclusion of the previous block's information creates a chain-like structure, where each new block is linked to the preceding ones. As a result, altering the data in a specific block would require modifying all subsequent blocks, making blockchain transactions irreversible.[33]


Figure 4.3: blockchain Logo[34]

Blockchain networks are commonly managed by a peer-to-peer (P2P) computer network with the objective of establishing a distributed public ledger. The network nodes adhere to a consensus algorithm protocol to incorporate and authenticate new transaction blocks. While it is feasible to create branches within the blockchain, which results in non-immutable records, the blockchain is designed to prioritize security and functions as a distributed computing system with significant tolerance to Byzantine faults.

### 4.2.3 Used data

In our study, we utilized approximately twenty medical images encompassing both individuals without any health complications and patients diagnosed with COVID-19. and we intend to cite a few of them.

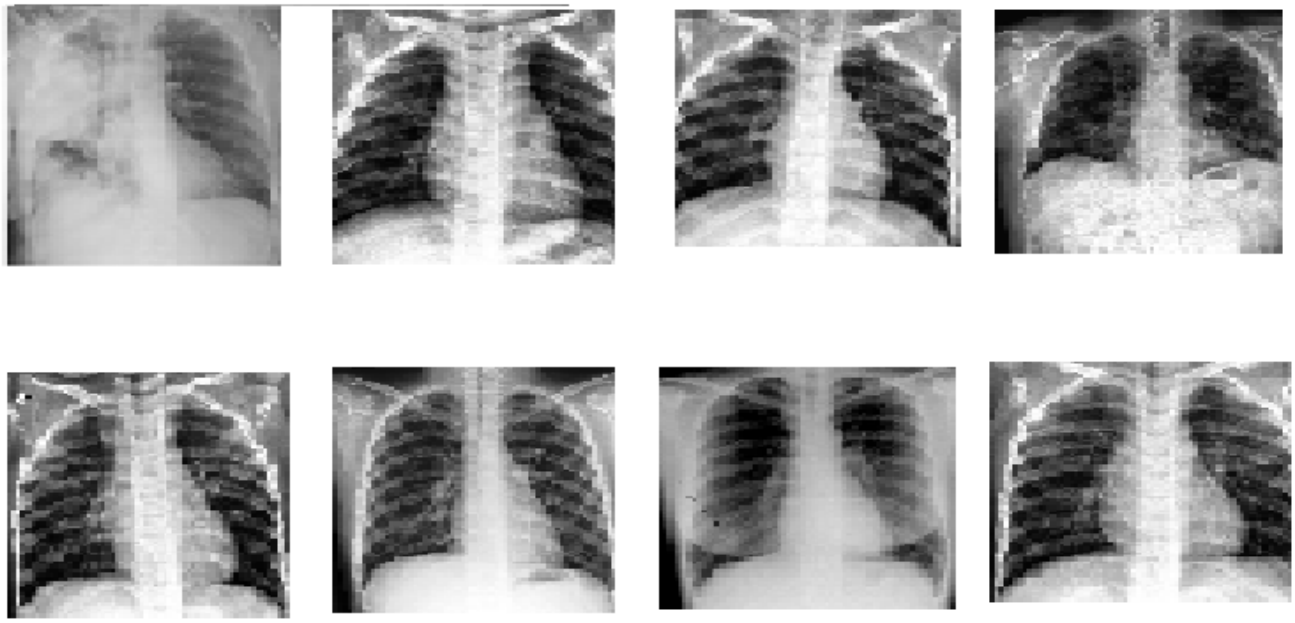

Figure 4.4: Medical pictures used

## 4.3 Implementation

You will apply the suggested method

### 4.3.1 Implementation of the proposed method

- **using Python:**

Code Listing 4.1: Import necessary packages

```
import os
import math
from PIL import image
import itertools
import numpy as np
import hashlib
from sys import argv
from lib import m_hash_hex
```

To begin, we bring in the required libraries to facilitate our coding process 4.1.

Code Listing 4.2: Matrix representation and calculate the MD5 hash.

```
def matrix_to_md5(matrix):
def matrix_to_md5(matrix):
    # Convert matrix to bytes and compute MD5 hash
    matrix_bytes = matrix.tobytes()
    md5_hash = hashlib.md5(matrix_bytes).hexdigest()
    return md5_hash

def total_possible_matrices(rows, cols):
    m = sum(rows)
    n = sum(cols)
    num_matrices = math.factorial(m+n-1) // (math.factorial(m) * math.
                                             factorial(n-1))
    return num_matrices
```

Code Listing 4.3: Matrix regeneration code.

```
    def generate_binary_matrices(rows, cols):
    # Compute the total number of possible matrices
    num_matrices = total_possible_matrices(rows, cols)

    # Generate all possible binary matrices
    matrices = set()
    for mat in itertools.product([0, 1], repeat=num_matrices):
        # Reshape binary vector into matrix with specified dimensions
        binary_matrix = np.reshape(mat, (len(rows), len(cols)))
        # Check if matrix satisfies row and column constraints
        if np.all(np.sum(binary_matrix, axis=1) == rows) and np.all(np.sum(
                                                 binary_matrix, axis=0) ==
                                                 cols):
            matrices.add(binary_matrix)

    return matrices
```

The code 4.2 and 4.3 is to convert an array to bytes and calculate its MD5 hash, and to determine the total number of possible arrays, we followed the following steps.

Code Listing 4.4: Hash and binary matrix generation.

```
    def generate_binary_matrices(rows, cols):
rows = np.int64(tuple(ones_in_rows))
cols = np.int64(tuple(ones_in_cols))

# Compare hashes and call regenerate_matrix() if they are equal
while md5_hash != md5_hash_original:
    #regenerate_matrix()
    # Generate new binary matrix
    binary_matrix = generate_binary_matrices(rows, cols).pop()
    # Compute new hash
    matrix_bytes = np.array(binary_matrix, dtype=np.uint8).tobytes()
    md5_hash = hashlib.md5(matrix_bytes).hexdigest()

# Hashes are equal, print the final matrix
print("Final Matrix:")
print(binary_matrix)
```

- **Using Blockchain:** Libraries and Modules used in Blockchain are:

Code Listing 4.5: Used Blockchain modules.

```
Import hashlib
Import time
Import os
Import secret
Import smartpy as sp
```

In code listing 4.5 code, we define a Block class representing each block in the blockchain. Each block has properties like timestamp, data, previous_hash, and hash. The timestamp stores the current time when the block is created. The data represents the information or transaction data to be stored in the block. The previous_hash is the hash of the previous block in the chain, linking the blocks together.

Code Listing 4.6: Block creation code.

```
# ----------------------  SmartPy  -------------------------

try:
 class MyContract(sp.Contract):
    def __init__(self, matrices, vectors):
        self.init_blockchain()
        self.matrices = matrices
        self.vectors = vectors
        self.init_storage()

    @sp.entry_point
    def run(self, params):
        for i in range(len(self.matrices)):
            matrix_data = {"matrix": self.matrices[i]}
            vector_data = {"vector": self.vectors[i]}
            matrix_hash = self.data.blockchain.append(matrix_data)
            vector_hash = self.data.blockchain.append(vector_data)
            matrix_file = f"./Matrix/binary_matrix_{i+1:03d}.txt"
            vector_file = f"./Vector/vector_{i+1:03d}.txt"
            matrix_size = get_size(matrix_file)/1000 # convert size to
                                                     kilobytes
            vector_size = get_size(vector_file)/1000 # convert size to
                                                     kilobytes
            sp.p(matrix_hash, time.ctime(self.data.blockchain.tail.timestamp),
                                                      matrix_size)
            sp.p(vector_hash, time.ctime(self.data.blockchain.tail.timestamp),
                                                      vector_size)
```

Code Listing 4.7: Hash generation.

```
  for i in range(len(matrices)):
       matrix_data = {"matrix": matrices[i]}
       vector_data = {"vector": vectors[i]}
       # ----------------------------------------
       matrix_value = open(matrices[i], "r").read()
       matrix_hash = hashlib.sha256(matrix_value.encode('utf-8')).hexdigest()
       matrix_data = {"matrix": matrices[i], "value": matrix_value, "hash":
                                                  matrix_hash}
```

```
        vector_value = open(vectors[i], "r").read()
        vector_hash = hashlib.sha256(vector_value.encode('utf-8')).hexdigest()
        vector_data = {"vector": vectors[i], "value": vector_value, "hash":
                                                vector_hash}
        # ----------------------------------------
        matrix_hash = blockchain.append(matrix_data)
        vector_hash = blockchain.append(vector_data)
        matrix_file = f"./Matrix/binary_matrix_{i+1:03d}.txt"
        vector_file = f"./Vector/vector_{i+1:03d}.txt"
        matrix_size = get_size_block(matrix_file)/1000 # convert size to
                                                kilobytes
        vector_size = get_size_block(vector_file)/1000 # convert size to
                                                kilobytes

        # Generate a random nonce (16 bytes)
        matrix_nonce = secrets.token_hex(16)
        vector_nonce = secrets.token_hex(16)

        with open(f"./Blocks/block_{2*i+1}.txt", "w") as f:
            f.write(f"Block {2*i+1}: Hash - {matrix_data['hash']} | Previous
                                                    Hash {blockchain.tail.
                                                    previous_hash} | Time {
                                                    time.ctime(blockchain.
                                                    tail.timestamp)} | Nonce
                                                     - {matrix_nonce} |
                                                    Matrix size - {
                                                    matrix_size:.2f} KB |
                                                    Stored_Data = {
                                                    matrix_data['value']}")
        with open(f"./Blocks/block_{2*i+2}.txt", "w") as f:
            f.write(f"Block {2*i+2}: Hash - {vector_data['hash']} | Previous
                                                    Hash - {blockchain.tail.
                                                    hash} | Time - {time.
                                                    ctime(blockchain.tail.
                                                    timestamp)} | Nonce - {
                                                    vector_nonce} | Vector
                                                    size - {vector_size:.2f}
                                                     KB | Stored_Data = {
                                                    vector_data['value']}")
```

In Code Listing 4.7, the hash is generated by combining the timestamp, data, and previous hash using the SHA-256 hashing algorithm. The code in 4.7 demonstrates the usage by creating a genesis block (the first block) with arbitrary data and a previous hash of '0'. Then, a new block is created with transaction data and the previous hash set to the hash of the genesis block. Finally, the details of the new block are printed.

### 4.3.2 Blockchain Size Comparaison

The proposed technique has a significant application in Blockchain, the leading data storage system. Blockchain as a set of blocks, each block contains transactions, we consider transactions as a data array with rows.

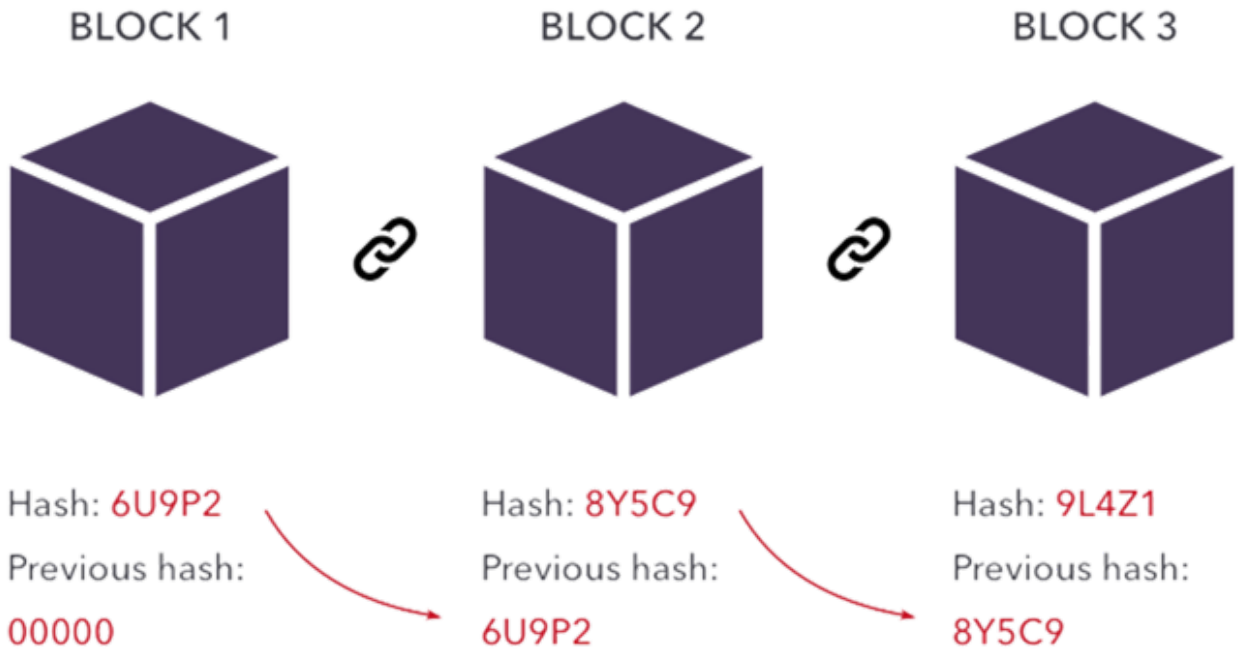


Figure 4.5: Blockchain

Block 1: Hash- 4d75efa55a0b5909896f88de5026b5da28fdbed6298a 51f8f3f99cc469fb1e63 | Previous Hash -b3e8a1c3f4624ea3d1a141930b0343311bdd744 5e2576148f6c8aa95014dc575 | Time - Thu Jun 1 23:14:35 2023 |Nonce - 4ff828b9afddb61e3fc341867fa759c5 | Matrix size - 5660.03 KB | Stored-Data = 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 ...

Block 2: Hash - e7a500d9140b1e810d33810bde535685a445197803ccd0093e87b 42080a323b9 | Previous Hash - 4fd423ee32b8b80fafd9c3132af29aed63a15551 04d99c0ed5ee1b7bbb25a922 | Time - Thu Jun 1 23:14:35 2023 | Nonce - 1ce4b1539f44815d4d6d540558ca6e91| Vector size - 14.01 KB | Stored-Data = L: 478 477 476 469 467 465 467 466 464 .. C: 460 454 450 448 448 443 443 444 .. 543 460 454 450 448 448 443 443 444 ...

## 4.4 Results analysis

The main interface of the application through which control is managed over various functions such as compression, blockchain creation, and displaying results.

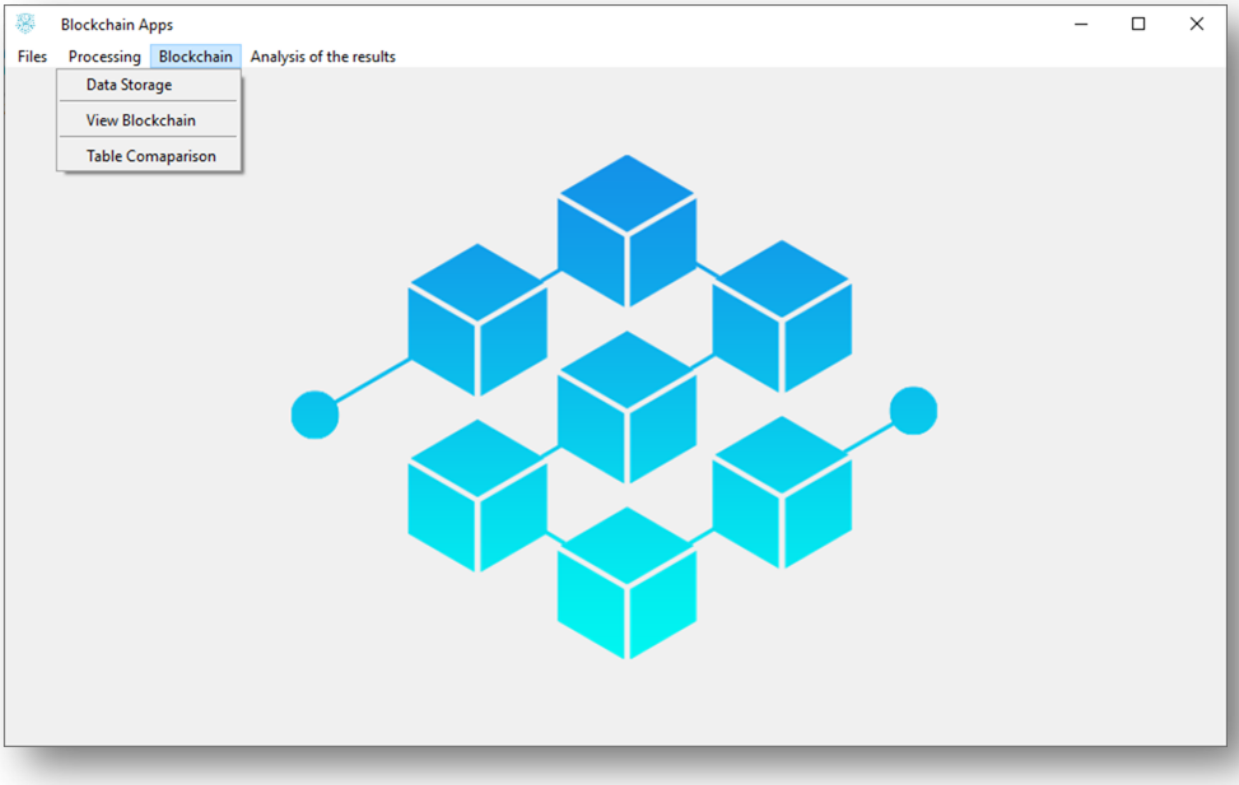


Figure 4.6: The main interface of the application

Stages of code execution and obtaining these results.

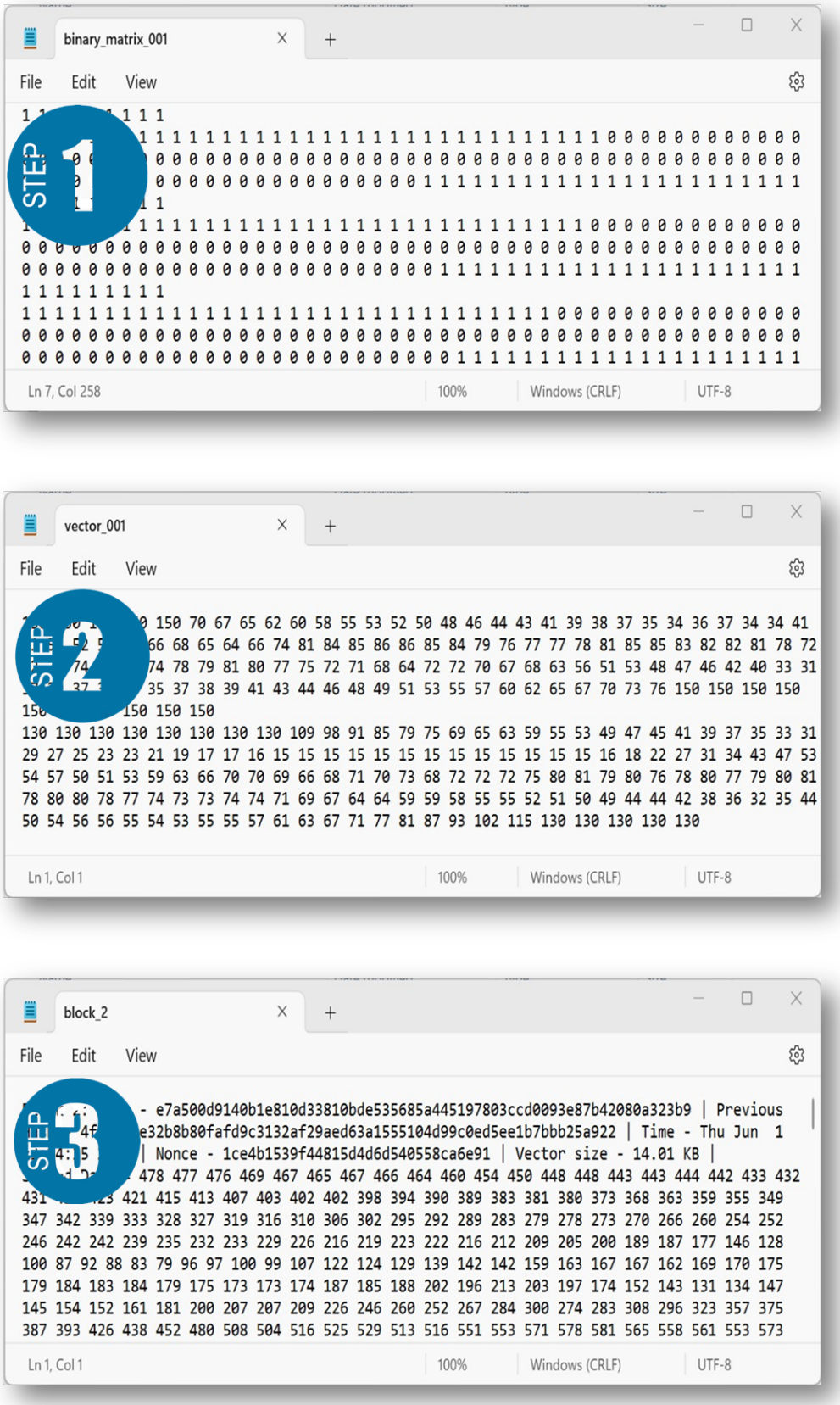


Figure 4.7: Stages of code execution

An image illustrating the block size after storing the data once as a vector and once as a matrix. The difference in block size is evident only when tested on the sample data used.

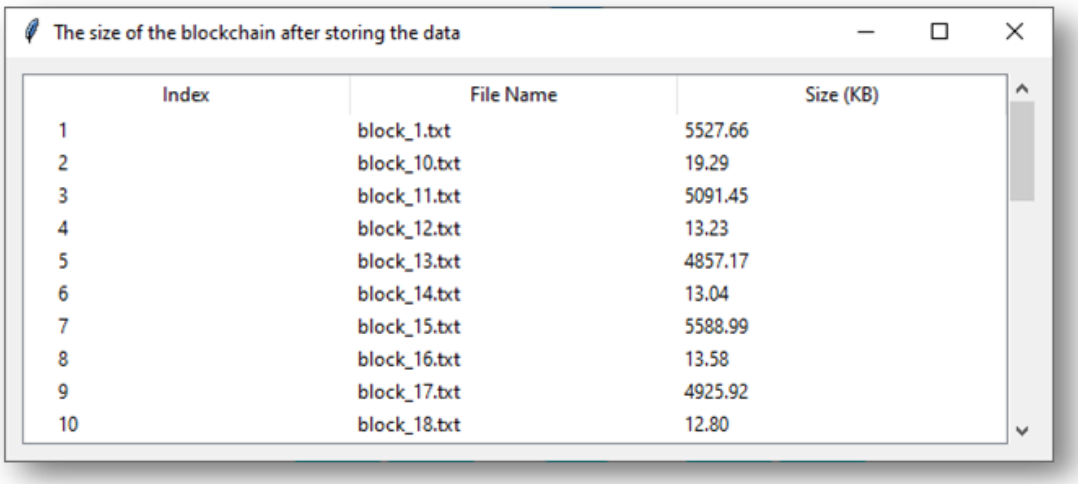


Figure 4.8: The size of the blockchain after storing the data

The content of the block showcases the various data it contains, such as hash, nonce, and creation timestamp.

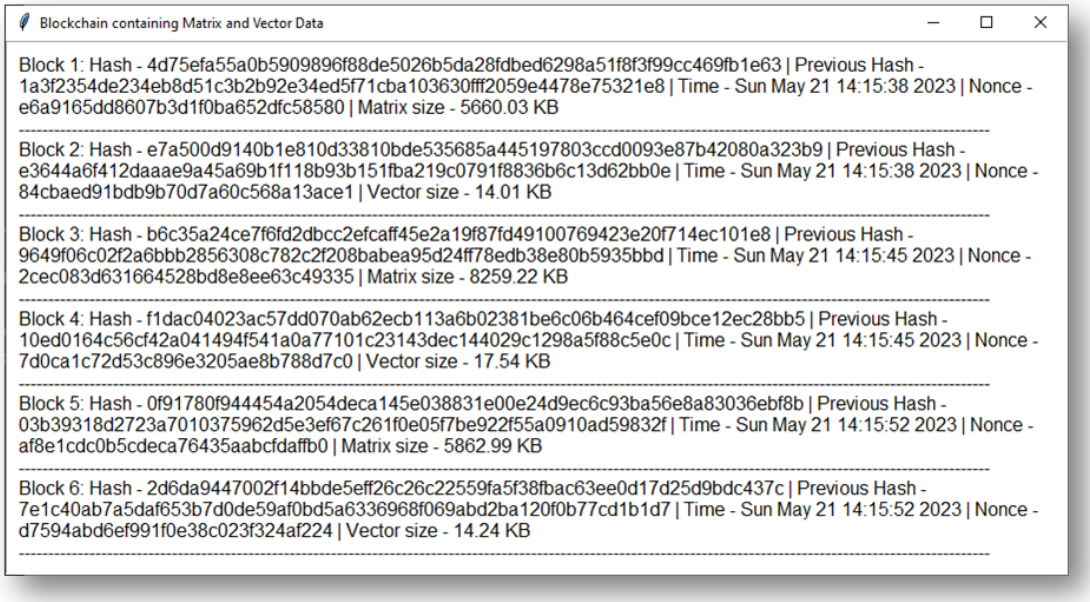


Figure 4.9: Blockchain containing Matrix and Vector data

A Diagram demonstrates the strength of the algorithm, where regardless of the increasing size of the matrix, the vector size remains constant.

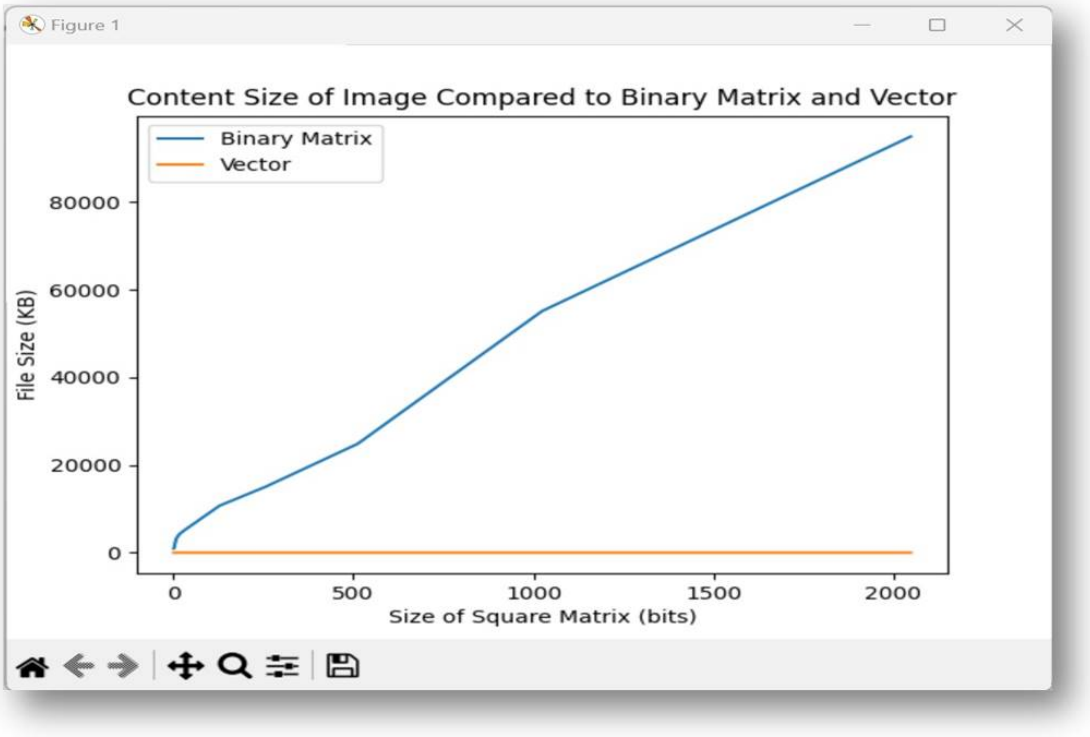


Figure 4.10: Content Size of Image Compared to Binary Matrix and Vector

Reducing the data size in a blockchain can result in improved efficiency and scalability. When the amount of data to be stored and processed is reduced, blockchain applications can become faster, more efficient, and more scalable. By minimizing the data footprint, the overall performance of the blockchain system can be enhanced, enabling it to handle larger transaction volumes and increasing its capacity to scale effectively.
Recovering the original data is a hard process certainly when the matrix size becomes big. Figure 4.11 and 4.12 show the complexity of the proposed technique as a function of the data size. We can exploit this technique for data integrity verification using blockchain, wherein the verification process differs from classical methods. The verification relies on reconstructing the original data from vectors and hashes, which can be computationally demanding and time-consuming. This approach involves reconstructing the entire data, making it potentially inconvenient compared to alternative methods. However, in numerous domains, the need for blockchain verification is infrequent. Additionally, with the substantial computational power of mainframe machines, these machines can handle the reconstruction process, while network nodes can perform the verification.

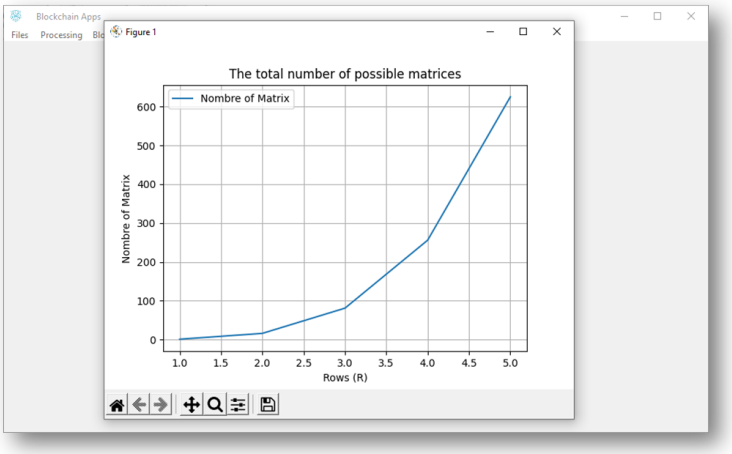


Figure 4.11: The total number of possible matrices

A Diagram shows the relationship between the length of both the row and column and the number of possible matrices, represented by the following function. (R * C) * log2(2**(R * C))

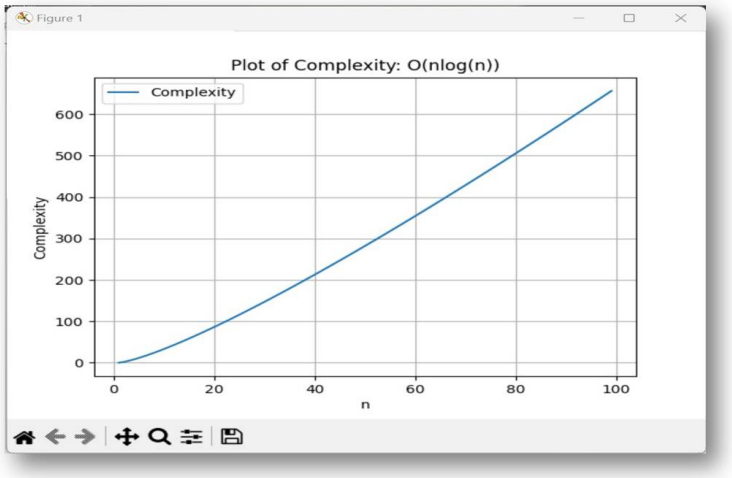


Figure 4.12: Plot of Complexity:O(nlog(n))

A Diagram illustrates the complexity and computational requirements for retrieving the original matrix from the vector and comparing the hash.

# General Conclusion

This manuscript introduces an innovative approach to represent large-scale biomedical datasets, surpassing traditional warehousing techniques efficiently. Our proposal centers around compressing big data within a blockchain framework, streamlining its transfer, utilization, and storage, particularly for low-power devices.

Considering the substantial volume of blockchain data, we assert that our model aligns perfectly with this advancing storage technology. The described methodology is straightforward to implement, eliminating the need for intricate data compression operations, resulting in reduced power consumption.

Our analysis of the approach demonstrates its effectiveness, as we observed a noteworthy reduction in volume, particularly when handling extensive amounts of original data. In essence, the proposed technique becomes increasingly effective as the data volume expands. For future investigations, we aim to harness the potential of the proposed approach to develop a consensus system that facilitates the recovery of the original data. This involves distributing the computed matrix problem ($ho, ve$, and $Hsh_I$) across the network, prompting nodes to compete for acquiring the original matrix $I$. Additionally, we propose establishing a centralized system responsible for storing the original data, subject to integrity checks. Overall, our proposed technology presents numerous advantages and potential solutions.

# Bibliography


[1] "What is IoT?" In: (). URL: https://www.oracle.com/internet-of-things/what-is-iot.

[2] Abhishek Khanna and Sanmeet Kaur. "Evolution of Internet of Things (IoT) and its significant impact in the field of Precision Agriculture". In: *Computers and electronics in agriculture* 157 (2019), pp. 218–231.

[3] Mircea Georgescu and Daniela Popescul. "The importance of internet of things security for smart cities". In: *Smart Cities Technologies* 8.12 (2016), pp. 23–33.

[4] Vlada Brilliantova and Thomas Wolfgang Thurner. "Blockchain and the future of energy". In: *Technology in Society* 57 (2019), pp. 38–45.

[5] John Kosakowski. "The Benefits of Information Technology. ERIC Digest." In: (1998).

[6] Robert E Crossler et al. "Future directions for behavioral information security research". In: *computers & security* 32 (2013), pp. 90–101.

[7] Katherine P Andriole. "Security of electronic medical information and patient privacy: what you need to know". In: *Journal of the American College of Radiology* 11.12 (2014), pp. 1212–1216.

[8] Somayeh Nasiri et al. "Security requirements of internet of things-based healthcare system: a survey study". In: *Acta Informatica Medica* 27.4 (2019), p. 253.

[9] Johannes Sametinger et al. "Security challenges for medical devices". In: *Communications of the ACM* 58.4 (2015), pp. 74–82.

[10] Jean-Paul A Yaacoub et al. "Securing internet of medical things systems: Limitations, issues and recommendations". In: *Future Generation Computer Systems* 105 (2020), pp. 581–606.

[11] George J Annas. "HIPAA regulations: a new era of medical-record privacy?" In: *New England Journal of Medicine* 348 (2003), p. 1486.

[12] Rafik Hamza et al. "A privacy-preserving cryptosystem for IoT E-healthcare". In: *Information Sciences* 527 (2020), pp. 493–510.

[13] Ashutosh Dhar Dwivedi et al. "A decentralized privacy-preserving healthcare blockchain for IoT". In: *Sensors* 19.2 (2019), p. 326.

[14] Ulfah Nadiya, Kusprasapta Mutijarsa, and Cahyo Y Rizqi. "Block summarization and compression in bitcoin blockchain". In: *2018 International Symposium on Electronics and Smart Devices (ISESD)*. IEEE. 2018, pp. 1–4.

[15] Chunlin Li et al. "Lightweight blockchain consensus mechanism and storage optimization for resource-constrained IoT devices". In: *Information Processing & Management* 58.4 (2021), p. 102602.

[16] Prem Prakash Jayaraman et al. “Privacy preserving Internet of Things: From privacy techniques to a blueprint architecture and efficient implementation”. In: *Future Generation Computer Systems* 76 (2017), pp. 540–549.

[17] Beltran Borja Fiz Pontiveros, Robert Norvill, and Radu State. “Recycling smart contracts: Compression of the ethereum blockchain”. In: *2018 9th IFIP International Conference on New Technologies, Mobility and Security (NTMS)*. IEEE. 2018, pp. 1–5.

[18] Challapalli Suryanarayana et al. *X-rays and Diffraction*. Springer, 1998.

[19] “MEDICAL TESTS Hand x-ray”. In: (). URL: https://www.ucsfhealth.org/medical-tests/hand-x-ray.

[20] Jørgen Arendt Jensen. “Medical ultrasound imaging”. In: *Progress in biophysics and molecular biology* 93.1-3 (2007), pp. 153–165.

[21] “Twins: A Closer Look at Your Developing Babies”. In: (). URL: https://www.parents.com/pregnancy/stages/fetal-development/twins-a-closer-look-at-your-developing-babies/.

[22] Thorsten M Buzug. *Computed tomography*. Springer, 2011.

[23] T Dill. “Contraindications to magnetic resonance imaging”. In: *Heart* 94.7 (2008), pp. 943–948.

[24] Tatsuo Togawa. “Body temperature measurement”. In: *Clinical physics and physiological measurement* 6.2 (1985), p. 83.

[25] T Deutsch et al. “Time series analysis and control of blood glucose levels in diabetic patients”. In: *Computer Methods and Programs in Biomedicine* 41.3-4 (1994), pp. 167–182.

[26] M Dominique Ashen and Roger S Blumenthal. “Low HDL cholesterol levels”. In: *New England Journal of Medicine* 353.12 (2005), pp. 1252–1260.

[27] Timothy G Chow, Lauren E Franzblau, and David A Khan. “Adverse Reactions to Biologic Medications Used in Allergy and Immunology Diseases”. In: *Current Allergy and Asthma Reports* (2022), pp. 1–13.

[28] Irena Spasić et al. “Medication information extraction with linguistic pattern matching and semantic rules”. In: *Journal of the American Medical Informatics Association* 17.5 (2010), pp. 532–535.

[29] Hans O Mauksch. “A social science basic for conceptualizing family health”. In: *Social Science & Medicine (1967)* 8.9-10 (1974), pp. 521–528.

[30] “What is Python? Executive Summary”. In: (). URL: https://www.python.org/doc/essays/blurb/.

[31] Peter Ritchie. *Practical Microsoft Visual Studio 2015*. Springer, 2016.

[32] “Visual Studio”. In: (). URL: https://logowik.com/microsoft-visual-studio-vector-logo-4965.html.

[33] “Blockchain”. In: (). URL: https://en.wikipedia.org/wiki/Blockchain.

[34] In: (). URL: https://www.freepik.com/premium-vector/blockchain-line-icon-logo-concept-dark-background_30092891.htm#query=blockchai.