PCA-Chain: A Novel Medical Image Retrieval Blockchain

Abhay Kumar Yadav\textsuperscript{a}, Virendra P. Vishwakarma\textsuperscript{b}

\textsuperscript{a,b} Department of Computer Science and Engineering, Guru Gobind Singh Indraprastha University, Delhi 110078
abhaybbdnitm01@gmail.com \textsuperscript{a}, virendravishwa@rediffmail.com \textsuperscript{b}

1. Introduction

The modern healthcare industry is a complex interconnection of different industrial enterprises, hospitals, research institutions, insurance companies, legal partners, and other governmental regulatory systems. Maintaining the secrecy, privacy and scalability of the patient’s private data is highly essential. Telemedicine and e-health services are an eminent area of the healthcare sector. Health services require data to be transferred to medical professionals for diagnosis and patient treatments remotely through telecommunication technologies. Maintaining security and privacy when monitoring patient data online is a challenging task because of sensitive data (Aste et al., 2019). Another challenge, that of interoperability between different components must be overcome for secure data exchange. Different types of research have been conducted in recent times and a variety of contemporary, technological
solutions have been presented to the healthcare industry, including artificial intelligence (Nguyen and Dang, 2018), computer vision, machine learning, deep learning, Internet of Things (IoT), and many more with the purpose of enhancing the efficiency and robustness of healthcare infrastructure.

1.1. Blockchain

The term “blockchain” refers to a digitally distributed public ledger technology that records and distributes information about all the transactions and events that occur across a peer-to-peer network (Yadav and Vishwakarma, 2022). A transaction on the blockchain is done using consensus algorithms which cannot be tampered with. Each block in blockchain is linked to previous blocks with the help of a digitally verified signature; those are cryptographically chained structures, which make data validation and storage secure (Wenhua et al., 2023). Any attempt to change a record in blockchain would be a costly endeavor, resulting in the information being tamper-proof.

The origin of blockchain can be traced back to 1990s, when two researchers, S. Haber and W.S Stornetta proposed a chain of cryptographically secure blocks for time-stamping digital documents. It gained popularity in 2008 when Satoshi Nakamoto used this technology to create Bitcoin. Blockchain is a fully distributed technology with a decentralized ledger. This ledger contains a detailed chain of blocks responsible for maintaining a transactions list. Each block has three main fragments: the first one is information, the second one is a hash block, and the last one is a hash of the previous block (Nakamoto, 2008). The uniqueness of each block is controlled by hash functions which are specific to each block. A blockchain may be public, consortium, hybrid or private, - based on its size, intended user base and access control (Peres et al., 2023).

Given that data blocks are arranged in a systematic chronological order, blockchain can be used for storing sequential data in the system. It uses a distributed consensus node algorithm to generate and update data, and automated scripting code (smart contracts) for programing and manipulating data (Wang et al., 2019). The different properties of blockchain are listed in Fig. 1.

![Figure 1. Properties of blockchain](image)

Any transactions written on blockchain, would always be present in its database (Xu et al., 2023). The immutability feature solved the major problems in securing data as the network does not have any single point of vulnerability.

Given the innovative property of blockchain, which permits users to transfer data assets in a completely decentralized method, the need of trusted centralized third party is eliminated (Caldarelli, 2020). Additionally, as blockchain provides enhanced security through its use of cryptography, it is suitable for the storage of highly sensitive and private data such as data of financial transactions,
medical records and other related data requiring higher security features. A detailed SWOT analysis of blockchain is shown in Fig 2.

Blockchain has been successful in storing textual data; data such as transaction status, login credentials can be easily stored. More recently, it has been successful in securing surveillance images in IoT and medical images. However, storing images on blockchain requires huge storage which in turn increases the size and complexity of blockchain due to associated computational costs. Hence, images must be compressed before being stored on blockchain. Different image compression techniques are used by researchers- to make images suitable for storage in blockchain. In this paper, the PCA compression method is presented for compressing, storing, and retrieving medical images.

1.2. Image Compression

Images constitutes an important method of data sharing. The human brain can understand and interpret image messages more easily than textual data. Image processing enhances an image’s attributes making it suitable to be stored and transmitted across a network. With the enormous amount of data on the internet, there is a need for securing the images. Various techniques such as watermarking, steganography, visual cryptography, cryptography without sharing the keys etc. These techniques are somehow successful in securing images but during the last 5 years cyber-crimes have increased rapidly because they have been able to bypass these security measures. Hence, there is a strong need for a newer more robust mechanism for securing images.

Ensuring data security in images is a challenge for researchers across the globe. Blockchain can potentially solve this problem. Implementing images on any blockchain platform would provide advantages such as rewarding the image creators, legally purchasing print photos, securely registering and storing an image, and combating image theft problems (Zheng et al., 2018).

Blockchain consumes a large amount of space and is generally costlier than other data storage databases. Storing all images on blockchain would be extremely costly, moreover, it would drastically slow down the blockchain network. Hence, researchers have proposed solution to this problem- involving the use of using different distributed databases for storing information with a fixed number of hashes and then storing- the hash function in blockchain. This would provide the necessary security for image data without worrying about the space an image would require in blockchain network (Tian et al., 2019).
1.3. Principal Component Analysis

PCA, commonly known as the Karhunen-Loeve approach, is a popular feature selection and dimension reduction method. Turk and Pentland were the first to use PCA to recognize human faces, and Kirth and Sirovich were the first to recognize human faces (Santo, 2012). The eigen face technique for recognition creates a feature space that decreases the dimension of the actual data space. For recognition, the smaller data space is employed.

PCA is a versatile statistical method used in various research areas. PCA implementation in image processing has been proven an efficient method for image compression and image retrieval. Several modifications in the method have been proposed such as kernel PCA, 2D-PCA and others. PCA is widely used in face recognition as it can be used in obtaining principal feature extraction on a large data set. Moreover, due to PCA’s capability of storing only important information using principal components, it has been highly successful in image compression.

1.4. PCA for Image Compression

Reducing data size is the most commonly performed task in image processing. There are different types of algorithms working on different principles for compressing images. Some of the previously developed image compression methodologies based on PCA image compression techniques are listed in Table 1. The compression parameters vary with the type of algorithms used and selected notions.

<table>
<thead>
<tr>
<th>Authors</th>
<th>Application on medical images</th>
<th>Objective evaluation method</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Taur et al., 1996)</td>
<td>✓</td>
<td>Method not specified</td>
</tr>
<tr>
<td>(Dwivedi et al., 2006)</td>
<td>-</td>
<td>$CR = \frac{\alpha}{\beta \mu + \theta + \varphi}$</td>
</tr>
<tr>
<td>(Stolevski, 2010)</td>
<td>-</td>
<td>$CR = \frac{MN}{kn^2 + k(\frac{MN}{n^2} + \frac{MN}{n^2})}$</td>
</tr>
<tr>
<td>(Santo, 2012)</td>
<td>✓</td>
<td>$CR = 1 - \frac{\text{PCA image memory size}}{\text{memory size of original}}$</td>
</tr>
<tr>
<td>(Mofarreh and Mostafa, 2012)</td>
<td>-</td>
<td>NA</td>
</tr>
</tbody>
</table>

Medical images are an important source of the patient’s private data. Most of these images are of large size and must be compressed for transmission and storage. The CR can be calculated using the total number of columns being stored in rows and column matrix. Although, this method makes an assumption about original size being equal for all consideration. Mofarreh and Mostafa (2012) improvised the equation by making operations on a varying number of rows and columns in the matrices. Santo (2012) further modified the equation by using the compression rate, in place of compression ratio, for elaborating compression degree.
1.5. PCA-Chain

PCA application in image compression is well known. We have proposed a blockchain based image compression and retrieval techniques. The proposed PCA-Chain utilizes eigen vectors for storing data values and is implemented on blockchain to solve the problem of security associated with storing images on blockchain (Richardson, 2009). The input of the PCA-Chain is a compressed medical image. The PCA-Chain stores the coefficient and is further added with the blockchain hash functions creating the encrypted image. The process is reversed for decryption and the output image is retrieved. The loss due to compression is calculated by measuring the correlation coefficient to determine the accuracy of the proposed PCA-Chain. The proposed PCA chain reduces the storage cost of storing images on blockchain and also minimize the execution time of transactions taking place in blockchain.

Modern day hospitals use PACS services for storing medical images digitally. These services have many vulnerabilities and are prone to being tampered with. Millions of medical images are exposed annually due to such vulnerabilities in the system. In 2019 only, nearly two million U.S residents’ medical images were exposed. Blockchain can solve the security aspects of medical images but the major problem in storing images in blockchain is the size of the image. Using blockchain to directly store images leads to an increase in cost of storage and time. This would eventually slow down the system. Hence, research has focused on breaking the images into smaller fragments and then storing them on blockchain or by putting hash functions in blockchain and storing images on any other distributed database. Due to technological constraints and storage space issues, many researchers have focused on storing the image data in other distributed databases such as IPFS and storing the hash value on blockchain (Koptyra et al., 2020). The performance evaluation of compression is done using state of the art techniques, including correlation, MSE (Mahendiran and Deepa, 2021), PSNR (Deshpande et al., 2018) and SSIM (Dosselmann and Yang, 2011) performance parameters.

The paper follows a generalized systematic approach and is organized into the following sections. Section 2 reviews state-of-the-art research in the healthcare sector. The proposed methodology for our PCA based blockchain for storing medical images is discussed in Section 3. Section 4 highlights the implementation and retrieval of the compressed image. The performance evaluation based on different performance parameters is also discussed in this section. Section 5 concludes the paper and discusses the scope of future research.

2. Related Works

The use of blockchain has proven useful in enhancing security in different areas of healthcare. However, very limited work has been done on the use of blockchain for the storage of digital images. The earliest implementation of blockchain was in cryptocurrencies applied in banking and financial sectors, then extended to the areas of biomedicines and healthcare. Blockchain has been utilized in the healthcare sector in different forms, as shown in Fig. 3. This section discusses a previously implemented blockchain in healthcare management.

2.1. Electronic Health Records (EHR) Management

Healthcare records must be digitized so that the medical workers and healthcare providers can access patient data easily for enhanced decision making. The digitized version of health records is stored online for access at any time, with enhanced security features called EHR. These records store the patient’s
medical history, diagnostic details, medicine reports, treatment plan, laboratory and test reports (Xiao et al., 2021). Blockchain is usually used in EHR to make digital health records more secure and reliable.

The existing drawback of EHR are distributed patient data amongst different healthcare providers, making it difficult to gather past data, especially if it has been obtained from other hospitals at different geographical locations (Vora et al., 2018). Blockchain can possibly solve this problem as it is capable of storing the patient’s present and past medical data securely for lifetime and can be accessed anytime by an authenticated healthcare provider.

MedRec is one such prototype model for maintaining an up-to-date and immutable patient history (Azaria et al., 2016). It utilizes the advantages of blockchain, such as providing authenticity, integrity, security, confidentiality and data sharing. MedRec works as a decentralized data management system for providing immutability and easy access of medical history to different components in healthcare sectors. Although some issues, such as record loss control, data origination issues and data dependency were still present. These limitations were eliminated in MedShare which provided a secure and independent data exchange along with maintaining privacy and EHR were shared amongst different serverless cloud service providers, physicians, and researchers (Chouhan et al., 2023).

2.2. Medicinal Supply Chain Management

The medical supply chain is an important aspect in pharmacology sectors, and it can benefit from blockchain’s unique characteristics. The pharmaceutical supply chain is an essential supply chain which uses a range of use cases and solutions and has a turnover of over $1200 billion from the global transportation of authentic medical drugs. (Mukherjee et al., 2023). However, compromised, fake, and forged medicines also have a market cap of $200 billion per year. These counterfeit drugs are great threats on the global market as unethical and potentially useless drugs are provided to people. Blockchain technology has the potential to solve the problem of counterfeit in this sector.

2.3. Clinical Research and Development

In clinical trials, problems associated with maintaining the privacy of data, providing record keeping features, maintaining integrity and secure data sharing may be encountered. (Sharma et al., 2023). Blockchain can solve these issues easily. An Ethereum based permissioned model with a smart contract
can be used for clinical data management for addressing patient enrollment issues. Since executing a transaction on an Ethereum protocol is faster than Bitcoin, authors used Ethereum smart contracts to maintain data transparency and security. Hence, managing patient admission through blockchain is more reliable than existing applications. Similar research implemented blockchain framework to securely track and store the patients’ informed consent (Benchoufi et al., 2017). These records were made publicly verifiable and were found to be unfalsifiable and secure.

2.4. Genomic Market

In areas of biotechnology and medical research, genomic data generated from human genomic projects are on large scale. The growth is expected to increase exponentially, with 15% of people predicted to have genome sequencing done by 2025, generating zettabytes of data. Unfortunately, at present, no such secure and reliable tool is present that can store such an amount of clinical and genomic data. Therefore, there is a strong need for such tools and technologies that can process and analyze genomic data easily.

Blockchain technology can tackle this issue and can provide a suitable solution for the storage and exchange of genomic data safely. Blockchain is being used extensively on the genomic market and its use is expected to increase by 66.4% between 2019-2029 (Gürsoy et al., 2022). A privacy-preserving and decentralized method was proposed for generating and accessing genomic data and also highlighting some present-day challenges (Erol et al., 2023). Another study showed the advancements that blockchain has made in solving genomic data issues and the future scope of blockchain was also mentioned (Ozercan et al., 2018).

2.5. Pharmaceutical Medicine

The pharmaceutical and medical sectors are introducing newer drugs and medicines with a high potential to help patients in fighting diseases. These medicines must be administered on time for effective treatment. Real-time tracking of these medicines is a troublesome task as some compromised or counterfeit medicines may have crept into the supply chain (Attaallah et al., 2022). These counterfeit drugs have no effect or may even have a negative effect on the patient’s health. Hence, there is a strong need for efficient supply chain management to tackle these issues.

2.6. Neuroscience Studies

Neuroscience has emerged as a new branch of healthcare studies. Many new models of neural networks have been proposed for supervised data and devices working based on the basis of brain commands help without using mechanical devices (Swan, 2015). They work by interpreting neural patterns of brain and translating them into neural devices, the obtained patterns return useful commands as instructions used in controlling devices externally. These devices can also help in monitoring the brain condition of subjects based on data patterns in brain cells. Complex algorithms along with big data analysis is responsible for managing complex neural process and can be executed by using blockchain technology for storing signals in neural link. Blockchain is capable of storing this information to be implemented in various areas of neuroscience.

3. Proposed Methodology

The overall framework of our proposed PCA-Chain is shown in Fig. 4. PCA follows a step-by-step approach from image acquisition to image retrieval as listed below.
3.1. Image Acquisition

The “Acute Lymphoblastic Leukemia Image Database” (ALL-IDB1) has been selected as dataset for images that are intended to be stored in blockchain (Labati et al., 2011). The dataset consists of 108 blood cell slide images having 39000 blood elements with marked lymphocytes. Fig. 5 shows some sample images from the database. The dataset was acquired in September, 2005 and was subjected to various magnifications ranging from 300 to 500 percent.
3.2. RGB to Grey Conversion

The color images in ALL-IDB1 are converted into grey scale images for efficient preprocessing as shown in Fig. 6. Since the images are already marked with lymphocytes, their conversion into grey-scale would not affect the image characteristics.

![Figure 6. RGB to grey conversion](image)

3.3. Image Shuffling

The leukemia image was shuffled further to minimize the biasing effect in image acquisition. The random sequence generator has been used to shuffle the images. Shuffling was done individually on each image to minimize the biasing effect. Fig 7 shows a shuffled image from the dataset.

![Figure 7. Shuffled image](image)

3.4. Applying PCA

Here, a total of 108 images, each with \( N^2 \) pixels can be used to form dimensional \( N^2 \) vectors. Each vector contains the intensity values of the same pixel from each picture. PCA would further be implemented for image compression. We obtained 108 different eigen vectors as the covariance matrix was of order 108×108. To compress this data, we further chose the transformation of 32 eigen vectors. If each image was of size 256×256, it could only be easily represented by 32 coefficients. A final data set with 64 dimensions was obtained, thus, we were able to save space. Fig. 8 highlights the whole step by step PCA implementation of our proposed PCA-Chain.
Figure 8. Steps involved in principal component analysis
3.5. PCA Coefficient Storage

PCA coefficient were further stored in the PCA-Chain and were treated with blockchain hashes. PCA-Chain stores the obtained PCA coefficients for further propagation in the blockchain. This minimized the data size of storing images on blockchain. Furthermore, image encryption minimized the possibility of unauthorized access. A single image of 108 pixels was subdivided into a block size of 32.

3.6. Storing Encrypted Image on Blockchain

The image was intended to be stored on blockchain. Since storing the image directly would not be feasible in data storage. A 256-bit simple hashing algorithm was further applied in the encrypted image for inclusion in blockchain. The first occurring image after image shuffling was considered as the genesis block. Other images followed in a linear increasing order with each block of PCA-Chain storing elements in a sequential manner and with the PCA coefficients arranged in the subsequent blocks of the blockchain. Fig. 9 depicts the method of data storage on blockchain. The data stored is the encrypted image along with SHA-256 encryption, furthermore, every subsequent block would follow a similar data storing methodology.

![Figure 9. Hash function-based storage in blockchain](image)

3.7. Image Retrieval

In the end, the image was retrieved after transmitting blockchain into the greyscale format. The image was obtained after removing the hash values from the eigen values (PCA coefficient). The PCA coefficients stored in blockchain were retrieved from the different blocks of the blockchain after removing hash values. The shuffled image was retrieved from the PCA-Chain. The reverse process of shuffling was applied on this outcome to get original images. The retrieved image of the input image (Fig.5) after being processed with PCA-Chain is shown in Figure 10.
4. Result and Discussion

We have constructed a PCA based framework for blockchain, on which the medical images of leukemia patients were stored. It has been implemented because of the compression properties provided by principal component analysis. The images were stored on a distributed cloud, namely, IPFS and their coefficient were stored in blockchain. The framework works by compressing the images and storing their eigen vectors. The eigen vectors were further treated with SHA algorithms to encrypt the data and its results along with element vectors and were stored as a block in blockchain. Fig. 11 shows the original and retrieved image. The process was repeated for all 108 images present in the dataset creating a blockchain of 108 blocks. Thus, the blockchain that was created has been named PCA-Chain. After storing the image on blockchain the image was retrieved back in original form by providing the correct eigen values which were known only to authenticated users. The loss due to compression was negligible and has been calculated by measuring different performance parameters as shown in Table 3. The higher the compression ratio, the lower the loss in compression. The enhanced security makes our proposed PCA-Chain a reliable image compression and retrieval method. Thus, it can be used as an alternative method of storing digital images as only the eigen vectors of images must be stored on blockchain.

![Figure 10. Retrieved image](image_url)

Figure 10. Retrieved image

![Figure 11. a) Original input Image. b) Retrieved output Image](image_url)

Figure 11. a) Original input Image. b) Retrieved output Image
5. Conclusion and Future Work

This paper has presented a new image storage and retrieval technique for storing images on blockchain. The proposed PCA-Chain uses PCA coefficients to store data values and was implemented on blockchain to solve the problem of secure image storage on blockchain. The use of blockchain to directly store images leads to an increase in the cost of storage and time. Our proposed PCA-Chain compresses the images and represents them in terms of eigen vectors with fixed block size. Only the coefficients along with the hashing values were stored in blockchain with images being stored on a distributed database. This paper also highlights the contribution of different researchers to the area of image security through blockchain. Similar to our work, many researchers have focused on storing image data in other distributed databases, such as IPFS and storing its hash value in blockchain. In the near future, focus will be placed on providing blockchain with the ability to detect medical diseases on the images stored in PCA-Chain. Moreover, a comparison with other dimension reduction techniques for determining the best approach of compressing the image will also be performed in the near future. Furthermore, research will also focus on integrating machine learning techniques in our blockchain for automated disease detection in a secured environment.

Declarations

Conflict of Interest: The authors declare no competing interests.

Acknowledgement: Both authors acknowledge All India Council of Technical Education (AICTE) for the Junior Research Fellowship of first author.

6. References


<table>
<thead>
<tr>
<th>Image from dataset</th>
<th>Correlation Coefficient (r)</th>
<th>Mean Squared Error (MSE)</th>
<th>Peak Signal to Noise Ratio (PSNR)</th>
<th>Structural Similarity Index Measure (SSIM)</th>
</tr>
</thead>
<tbody>
<tr>
<td>lm001_1</td>
<td>0.997</td>
<td>0.96</td>
<td>48.36</td>
<td>0.994</td>
</tr>
<tr>
<td>lm008_1</td>
<td>0.998</td>
<td>0.95</td>
<td>49.05</td>
<td>0.995</td>
</tr>
<tr>
<td>lm017_1</td>
<td>0.996</td>
<td>0.96</td>
<td>48.78</td>
<td>0.994</td>
</tr>
<tr>
<td>lm062_1</td>
<td>0.999</td>
<td>0.97</td>
<td>49.81</td>
<td>0.995</td>
</tr>
<tr>
<td>lm089_0</td>
<td>0.989</td>
<td>0.99</td>
<td>48.20</td>
<td>0.993</td>
</tr>
<tr>
<td>lm103_0</td>
<td>0.990</td>
<td>0.99</td>
<td>48.23</td>
<td>0.993</td>
</tr>
</tbody>
</table>


