Studying the Security and Privacy Issues of Big Data in the Saudi Medical Sector

Ramy Elnaghy, Hazem M. El-Bakry
Faculty of Computer and Information Sciences, Mansoura University, Mansoura 35516, Egypt

Abstract—In today's era of Big Data, with the integration of data from various systems, devices, and machines used by healthcare service providers, health insurance companies, and their sub-sectors, maintaining privacy and security has become crucial. It is important to uphold the confidentiality and security of data exchanged between data service providers and insurance companies as required by law. The purpose of this paper is to focus on addressing the security and privacy issues associated with healthcare data, particularly concerning medical data in both in- transit and at-rest modes. We aim to provide a proposed solution to enhance data security and maximize privacy protection.

Keywords—Security; privacy; healthcare; medical data; big data

I. Introduction

The healthcare industry has made significant strides in recent years with the adoption of electronic patient records and the digitization of healthcare workflows. However, this transformation has resulted in an explosion of clinical data, which is characterized as big data. The abundance of clinical data has created numerous opportunities for healthcare organizations to leverage data analytics, clinical decision support, and disease surveillance to optimize treatment and improve patient outcomes. Nonetheless, with the benefits of big data come numerous challenges, particularly regarding information security and privacy. Healthcare organizations must adopt proactive measures to safeguard sensitive information and prevent security breaches and other incidents. This paper aims to address the security and privacy concerns associated with medical data, particularly in in-transit and at-rest modes. We propose a solution that enhances data security and maximizes privacy protection. The adoption of big data technologies in healthcare has created numerous opportunities for clinical decision support, disease surveillance, population health management, and treatment optimization. However, with the increased use of electronic patient records and digitization of healthcare workflows come significant security and privacy risks. Medical data is highly sensitive, and its confidentiality and privacy must always be protected. Healthcare organizations must take a proactive, multi-layered approach to data security, including encryption, access control, network security, and employee training. Data encryption is a highly effective way to protect medical information. Encryption uses algorithms to convert plain text into cipher text, adding a layer of security to ensure that only authorized parties can access the data. In-transit encryption is used to protect data that is transmitted over a network or the internet, such as emails or data transfers between healthcare providers. At-rest encryption, on the other hand, is used to protect data that is stored on electronic devices, such as servers or hard drives. To implement data encryption, healthcare organizations must first identify which data needs to be protected and where it is stored. Then, appropriate encryption algorithms and methods must be selected and applied to the data. The encryption keys must be securely managed and stored to ensure that only authorized parties can access the data. While data encryption is an effective security measure, it is not a complete solution. Healthcare organizations must also implement access controls to limit access to sensitive information to authorized personnel only. Network security measures such as firewalls, intrusion detection, and prevention systems can help protect against cyberattacks and data breaches. Employee training is also essential to ensure that all staff members understand their responsibilities in safeguarding patient data and are aware of potential security threats. In summary, healthcare organizations must adopt a comprehensive, multi-layered approach to data security that includes encryption, access control, and network security, through utilizing big data and modern technologies such as blockchain [1].

II. Literature Review

Soumya et al. in [2], reviews the security issues associated with big data in Internet of Things (IoT) and cloud computing. The authors discuss the challenges of securing big data in IoT and cloud computing, including issues related to data privacy, data integrity, authentication, and access control. They also discuss various security solutions that have been proposed to address these challenges. Parsa et al. in [3] proposes a security management framework for big data in smart healthcare. The authors identify various security challenges in smart healthcare, including issues related to data privacy, data integrity, and data availability. They then propose a framework that includes various security measures, such as access control, encryption, and monitoring, to address these challenges. Hayat et al. in [4] presents a security model for the big healthcare data lifecycle. The authors discuss the security challenges associated with healthcare data, including issues related to data privacy, data integrity, and data confidentiality. They then propose a security model that includes various security measures, such as access control, data encryption, and data backup, to address these challenges. Aqeel et al. in [5] reviews the security and privacy issues associated with big data in healthcare applications. The authors discuss the challenges of securing big data in health-care, including issues related to data privacy, data integrity, and data confidentiality. They also discuss various security and privacy solutions that have been proposed to address these challenges.

Musfira et al. in [6] surveys various big data security solutions that have been proposed to address security issues in healthcare. The authors discuss various security solutions, such as access control, data encryption, and data backup, and evaluate their effectiveness in addressing the security challenges associated with big data in healthcare. Isabel et al. in [7] analyzes the security issues associated with big data in healthcare. The authors discuss the challenges of securing big data in healthcare, including issues related to data privacy, data integrity, and data confidentiality. They also discuss various security solutions that have been proposed to address these challenges. Saraladevi et al. in [8] studies the security issues associated with big data and Hadoop. The authors discuss the challenges of securing big data in Hadoop, including issues related to data privacy, data integrity and data confidentiality. They also discuss various security measures that can be implemented to address these challenges. ABHISHEK et al. in [9] analyzes the issues related to healthcare data integrity. The authors discuss the challenges of maintaining data integrity in healthcare, including issues related to data accuracy and data consistency. They also propose various measures to ensure data integrity, such as data validation and data verification.

Kanika et al. in [10] proposes an encryption approach based on the Rivest-Shamir-Adleman (RSA) algorithm to preserve the confidentiality of big data. The authors discuss the challenges of securing big data and propose an encryption approach based on RSA to ensure data confidentiality. Mustafa et al. in [11] presents a systematic review of privacy-preserving healthcare data sharing on blockchain. The authors discuss the challenges of ensuring privacy in healthcare data sharing and evaluate various blockchain-based solutions that have been proposed to address these challenges. Peng Xi et al. in [12] reviews the blockchain-based solutions that have been proposed to secure healthcare data sharing. The authors discuss various blockchain-based solutions, such as smart contracts and permissioned blockchain, and evaluate their effectiveness in ensuring data security and privacy. Mehak et al. in [13] identified potential vulnerabilities and attacks that could compromise the security of Hadoop-based big data systems. The purpose of the paper is to highlight these issues and provide recommendations for mitigating them. Mohan et al. in [14] proposed an efficient and secure big data storage solution using triple data encryption standard (Triple Data Encryption Standard (3DES)) in a cloud environment. The purpose of the paper is to provide a secure storage solution for big data in the cloud. Rakib et al. in [15] presented privacy- preserving k-nearest neighbors (KNN) training scheme over blockchain-based encrypted health data. The purpose of the paper is to propose a secure and privacy-preserving KNN training mechanism for health data in a blockchain-based system.

Parvathaneni et al. in [16] proposed a blockchain-based solution for secure healthcare data communication among non-terminal nodes in the Internet of Things (IoT) architecture in 5G networks. The purpose of the paper is to provide a secure and efficient data communication mechanism for healthcare IoT systems. Rafik et al. in [17] proposed fully homomorphic en- cryption (Fully Homomorphic Encryption (FHE)) algorithms for secure big data analysis. The purpose of the paper is to provide a solution for secure big data analysis in a privacy- preserving manner using FHE. Jose et al. in [18] discussed the security and privacy issues of big data in general. The purpose of the paper is to provide an overview of the security and privacy challenges associated with big data. Gousiya et al. in [19] proposed a sandbox security model for the Hadoop file system. The purpose of the paper is to provide a security mechanism for the Hadoop file system, which enables secure execution of untrusted code. Iroju et al. in [20] This paper provides an overview of the prospects, challenges, and solutions related to big data in healthcare. It discusses the potential of big data to improve healthcare outcomes and reduce costs, as well as the challenges related to data privacy, security, and integration. The paper also discusses some of the current solutions to these challenges, such as the use of big data analytics and cloud computing.

MATTURDI et al. in [21 provides a comprehensive review of the security and privacy challenges of big data, with a focus on data protection, access control, and privacy- preserving techniques. The paper also discusses some of the current solutions to these challenges, such as the use of encryption, authentication, and authorization mechanisms, as well as data anonymization and data masking techniques. Karim et al. in [22] provides a review of the security and privacy challenges of big data in healthcare, with a focus on the protection of patient data. The paper discusses the importance of ensuring the confidentiality, integrity, and availability of healthcare data, as well as the challenges related to data privacy, security, and interoperability. The paper also discusses some of the current solutions to these challenges, such as the use of encryption, access control, and data anonymization techniques. Gunasekaran et al. in [22 focuses on the security challenges of big data in the healthcare industry 4.0, which includes the integration of advanced technologies such as IoT, Artificial Intelligence (AI), and cloud computing. The paper discusses the importance of ensuring the security and privacy of healthcare data in this new era, and proposes a framework for big data security intelligence that includes threat intelligence, security analytics, and incident response.

III. Objectives

The purpose of this paper is to address the security and privacy challenges encountered in the Saudi healthcare and medical insurance sector during the storage and transfer of big data. Specifically, this paper aims to achieve the following:

1) Identify potential security and privacy risks associated with storing and transferring big data. This includes identifying the risks of unauthorized access to sensitive information, data breaches, cyber-attacks, hacking, insider threats, and data loss.

2) Analyze current practices and technologies used for securing big data in the healthcare and medical insurance sector. This analysis will provide an understanding of the existing security and privacy measures implemented by the industry and highlight any gaps or deficiencies in current practices.

3) Propose technical solutions by applying best practices and techniques for addressing the security and privacy challenges associated with storing and transferring big data.

The proposed solutions aim to maximize the security and privacy of medical information.

This paper proposes a multi-layered approach to address the security and privacy risks associated with storing and transferring big data in the healthcare and medical insurance sectors. The proposed solution includes implementing encryption techniques, access control measures, and security protocols to minimize the risk of unauthorized access to sensitive information. Additionally, the solution involves the use of advanced technologies such as blockchain to ensure data integrity and prevent data tampering.

The implementation of the proposed solution can provide many benefits, including better protection of patient information, improved data accuracy, and increased trust and confidence among stakeholders. Moreover, it can help healthcare organizations comply with relevant regulations and standards, such as the General Data Protection Regulation (General Data Protection Regulation (GDPR)) and the Health Insurance Portability and Accountability Act (Health Insurance Portability and Accountability Act (HIPAA)).

IV. Methodology

To ensure the security and privacy of big data, it is essential to adopt a comprehensive approach that covers all aspects of data protection. In this paper, we propose a solution that combines multiple techniques to address the various vulnerabilities that could compromise healthcare security and privacy using big data.

The proposed solution will encompass techniques such as data masking, encryption, access controls, and data anonymization. By combining these techniques, the proposed solution provides a comprehensive approach to ensure data security and privacy, particularly in healthcare organizations when processing medical insurance data for insurance companies or claim transaction operators. The implementation of these techniques can significantly reduce the risks of data breaches, unauthorized access, and misuse of data, thereby enhancing data security and privacy.

A. *Medical Insurance Transaction Ecosystem*

To better understand the techniques and solutions involved, it is important to first gain insight into the transaction system of medical insurance. This includes understanding how data is ingested, stored, and processed.

The claim processing system typically comprises three main domains or parts, as illustrated in Fig. 1.

1) Healthcare Provider Integration Tier: A collection of tools and configurations are utilized to extract data from Hospital Management System (HMS) or Practice Management Application (PMA) systems and safeguard the original data before transmitting it to the data processing stage.

2) Data Processing Tier: During this tier, the data Extract, Transform, Load (ETL) process will occur, followed by data analysis. The data will then be structured and prepared for the subsequent phase, which involves transmitting the final claims to the insurance companies as shown in Fig. 2.

3) Insurance Company Integration Tier: During this stage, the integration hub component will direct the data to the appropriate destination based on the configurations and the targeted beneficiary or consumers.

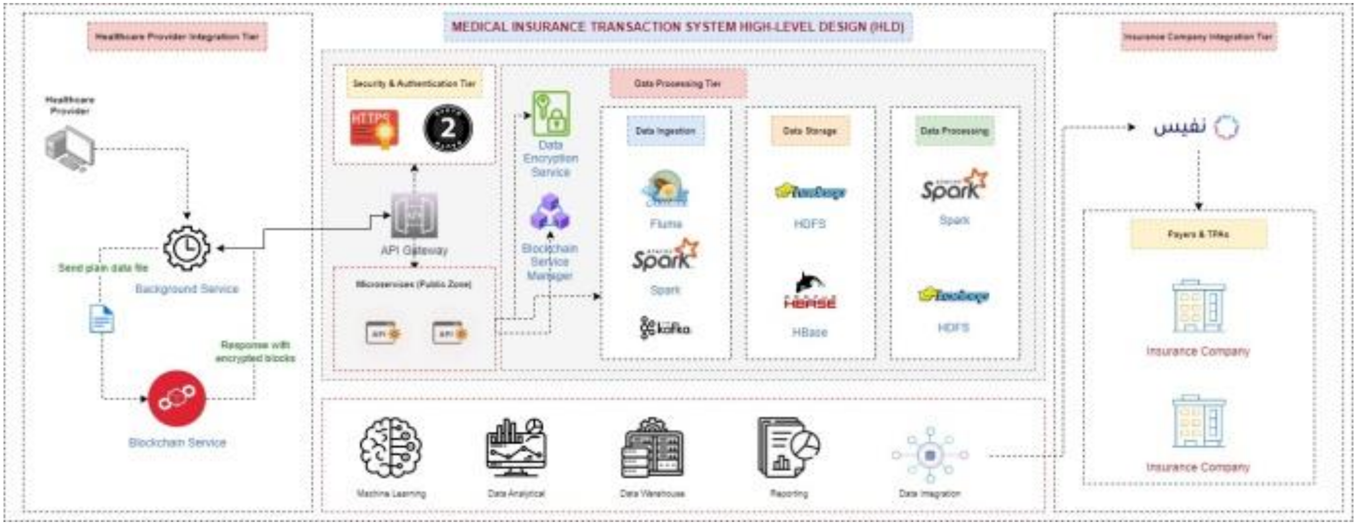

Fig. 1. Claim processing system.

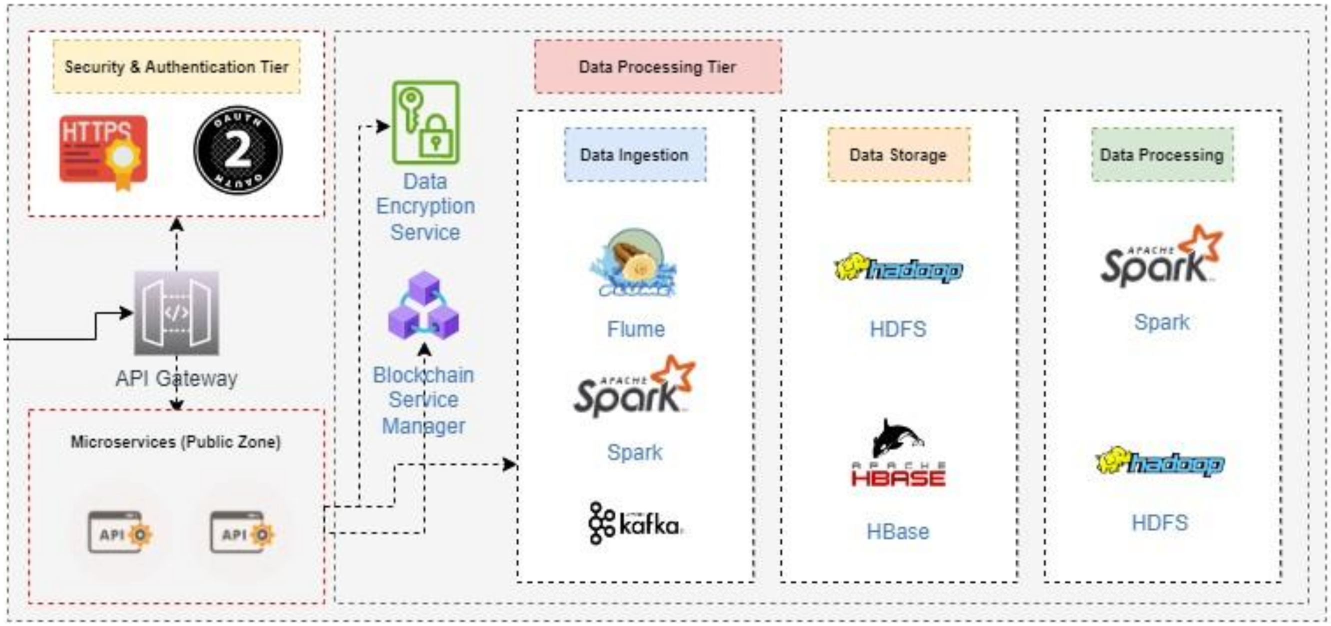

Fig. 2. Data processing tier.

B. A Common Data Processing Pipeline in the Medical Insurance Domain Includes the following:

1) Data Ingestion: The initial stage of the data processing pipeline involves collecting data from diverse sources, such as electronic health records and claims data, and ingesting it into a big data storage platform, such as Hadoop’s Hadoop Distributed File System (HDFS) or other distributed file systems.

2) Data Cleaning: After data ingestion, cleaning and preprocessing are necessary to remove inconsistencies, missing data, or errors, which can involve various techniques like deduplication, normalization, and data imputation.

3) Data Transformation: Here, raw data is transformed into a format suitable for processing and analysis. This involves tasks such as data format conversion, data aggregation at different granularities, and feature engineering.

4) Data Analysis: After data transformation, patterns and trends can be identified through analysis using techniques like statistical analysis, machine learning, and predictive modeling.

5) Data Visualization: Visual aids such as dashboards, charts, and graphs can be utilized to present the analysis results. This helps in conveying the data insights to stakeholders and decision-makers.

6) Data Storage: After completing the analysis, the processed data can be securely stored in scalable data stores like Hadoop’s HDFS or other distributed file systems. This stored data can be further transformed and analyzed in the future as required.

C. Compliance with Regulations

Different regulations and standards, such as HIPAA and GDPR, govern the storage and transfer of medical information that we need to comply with.

1) The Saudi Arabian Standards Organization (The Saudi Arabian Standards Organization (SASO)): SASO sets standards for consumer products and services, including those related to data protection.

2) The Saudi Arabian Monetary Authority (The Saudi Arabian Monetary Authority (SAMA)): SAMA regulates financial institutions in Saudi Arabia, including insurance companies, and has issued regulations related to data protection for the financial sector.

3) HIPAA: This law requires the establishment of national standards for electronic health care transactions [22].

4) The Privacy Law: This law was enacted in 2019 and governs the collection, processing, use, and disclosure of personal data in Saudi Arabia. The law applies to both government and private sector entities and requires organizations to obtain consent for the collection and use of personal data and to implement appropriate security measures to protect personal data.

5) The Cybercrime Law: This law criminalizes a range of activities related to cybercrime, including unauthorized access to computer systems and the theft of personal data.

Data protection regulations and laws in some countries as mentioned in study [22] address the growing thicket of applicable data protection legislation.

D. Threats and Risks

The medical insurance sector deals with vast amounts of sensitive data, including patient medical records, financial information, and insurance claims. Storing and transferring this big data poses significant security and privacy risks [23], such as data breaches, loss, and unauthorized access [24].

1) Unauthorized access: Malicious individuals may gain access to medical information in big data, leading to identity theft, fraud, and other malicious activities.

2) Data breaches: Compromised systems or networks can result in the loss or theft of sensitive medical information, which can be exploited for malicious purposes.

3) Insider threats: Misuse or disclosure of medical information by employees or contractors can lead to breaches in confidentiality or privacy.

4) Cyberattacks: Attacks like denial-of-service or malware can compromise the availability or integrity of medical information in big data.

5) Data loss: Accidental deletion or system failure can result in the permanent loss of medical information.

6) Inadequate security measures: Weak passwords or lack of encryption can make medical information in big data vulnerable to attacks and breaches.

7) Lack of regulatory compliance: Noncompliance with regulations and standards like HIPAA or GDPR can lead to legal and financial penalties.

E. Security Measures and Techniques

Various technologies are in use for protecting the security and privacy of healthcare data. The most widely used technologies are [22]:

1) Access control: Restricts medical information access based on user roles and privileges.

2) Encryption: Encodes medical information to prevent unauthorized access.

3) Data masking: Replaces sensitive information with fake values to protect privacy while allowing for research.

4) Data minimization: Reduces the amount of sensitive information stored and transferred.

5) Data backups: Stores copies of medical information in a separate location.

6) Penetration testing: Simulates cyberattacks to identify system vulnerabilities.

7) Security audits: Assessments of security measures and practices to ensure compliance with regulations and standards.

F. Privacy-preserving Techniques

1) Differential privacy: Adds noise to data to protect the privacy and prevent re-identification attacks.

2) Anonymization: Removes identifying information from medical data to protect privacy while allowing data analysis [24].

3) Pseudonymization: Replaces identifying information with pseudonyms to protect privacy while allowing data analysis [25].

4) Secure multi-party computation: Allows multiple parties to compute medical data without revealing it to each other, enabling collaborative analysis while protecting privacy.

5) Homomorphic encryption: Enables computation on encrypted data without decrypting it, protecting privacy while allowing data analysis [15].

6) Secure data sharing: Enables secure sharing of medical data between parties with appropriate permissions, protecting privacy while allowing data analysis [9].

G. Proposal

Several existing techniques have been employed to secure healthcare data, such as access controls, data encryption, and anonymization. Studies have shown that using blockchain can effectively limit the risk of data integrity by restricting unauthorized access to sensitive data [12] (Peng et al., 2022). Similarly, the use of data encryption is an effective method for maintaining data confidentiality and integrity [10] (AB-HISHEK et al., 2019). Furthermore, anonymization, column-based encryption, and two-way authentication techniques are effective in preserving patient privacy (Musfira et al., 2018). These results demonstrate the value of utilizing such techniques to enhance the security and privacy of healthcare data as shown in Table I.

TABLE I. SECURITY AND PRIVACY TECHNIQUES

Ref	Year	Method / Technique
[21]	2004	Using integrated Rule-Oriented Data (Integrated Rule-Oriented Data (iRODS)) is proposed to be the solution to ensure security and privacy in big data.
[8]	2015	Highlighting the big data security issues and how to secure the Hadoop File System (HDFS) by using Kerberos, Algorithm, and Name node.
[22]	2017	The proposed system uses key management security mechanisms to protect big data.
[4]	2018	Discussed the threats and countermeasures of the healthcare data life cycle and its suggested defense.
[6]	2018	Using the proposed security framework uses column-based encryption and two-way authentication for privacy and multi-biometric-based key generation.
[13]	2018	Using Kerberos authentication protocol, and attribute-based encryption.
[10]	2019	Using the RSA algorithm to encrypt sensitive medical data.
[2]	2020	Highlighting the cloud security threats, Fog computing, IoT security threats, and Blockchain security threats and addressing the defense mechanisms of different computing technologies.
[9]	2020	Discuss the blockchain as a data integrity technique.
[11]	2020	Examining blockchain-based techniques used to privacy-preserving healthcare data and discussing smart contracts and PKI as blockchain technologies.
[19]	2020	Using Sandbox security for Map Reduce jobs to scan the jar files and check if it contains any harmful code.
[3]	2021	A proposal for a security framework is being made that utilizes the logistic equation, hyperchaotic equation, and DNA encoding techniques to cipher and encrypt images. The encrypted images will then be stored in shares across multiple cloud-based servers for distributed storage.
[16]	2021	Discussing different encryption algorithms used in blockchain.
[12]	2022	Using blockchain and IOMT technologies to protect healthcare data.
[14]	2022	Using Triple Data Encryption Standard (TDES) to secure big data in the cloud environment.
[17]	2022	Focused on using homomorphic encryption technology to secure Big Data processing.

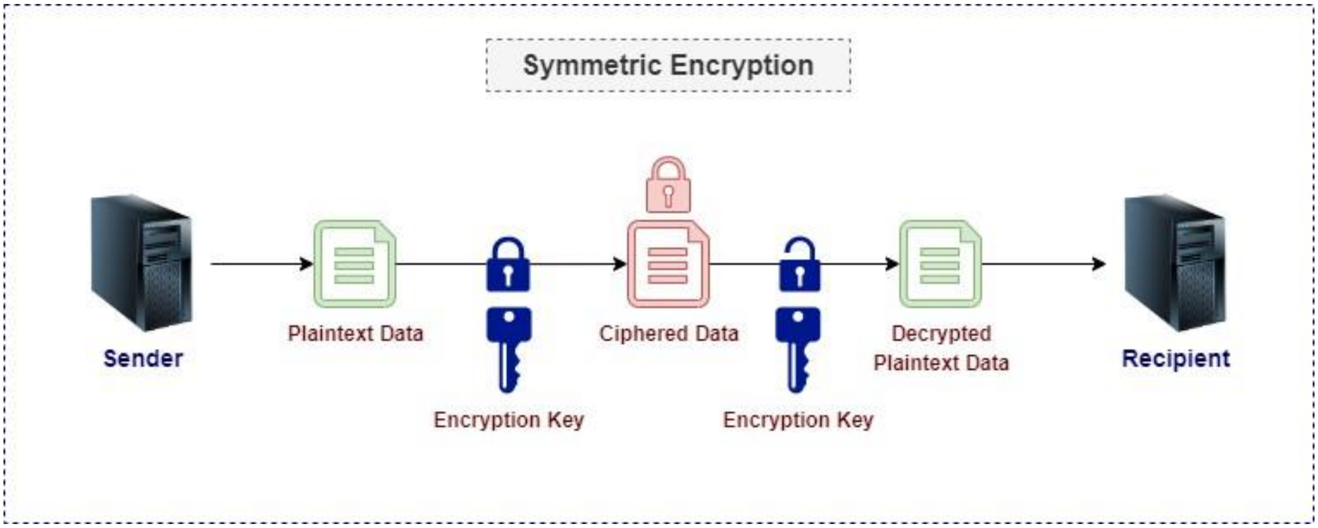

Fig. 3. Symmetric encryption.

Proposed Techniques: Ensuring the security of patient information is critical in the healthcare industry. The following suggested implementations can maximize the security and privacy of the patient's electronic medical records (Electronic Medical Records (EMR)), at-rest, and in-transit modes.

1) Data Encryption: Encryption is a method of encoding information so that only authorized parties can access it.

Symmetric encryption – Advanced Encryption Standard (Advanced Encryption Standard (AES)). Fig. 3 shows the symmetric encryption.

The RSA or Advanced Encryption Standard is one of the safest and most secure types of encryption. As a matter of fact, RSA is used by the United States Government to secure classified information as well as other hardware and software products. RSA, on the other hand, makes use of block ciphers that encrypt information one fixed-size chunk at a time. RSA has three key sizes: 128-bit, 192-bit, and 256-bit. The larger the key size, the stronger the encryption, therefore, we are going to use the 256-bit key as shown in the below algorithm [26].

Algorithm 1 GenerateAESKey(password)

```
function GENERATEAESKEY(password)
    sha256 ← SHA256.CREATE()
    passwordBytes←ENCODING.UTF8.GETBYTES(password)
    hash ← SHA256.COMPUTEHASH(passwordBytes)
    aes ← AES.CREATE()
    AES.KEYSIZE ← 256
    AES.KEY ← hash
    return AES.KEY
end function
```

Hashing SHA-256 Hashing is a one-way encryption technique that generates a unique fixed-length string of characters from a given input. There are several hashing algorithms, including MD5 and SHA-1, and SHA-256 is the most commonly used algorithm in blockchain technology, and it is recommended for its security and performance, we are going to use the SHA-256 as shown in the below algorithm [16].

Algorithm 2 CalculateHash(publicKey)

```
function CALCULATEHASH(publicKey)
    sha256 ← create SHA256 object
    inputString ← concatenate PreviousHash, Data and publicKey using ”-” separator
    inputBytes ← convert inputString to bytes using ASCII encoding
    outputBytes ← compute hash of inputBytes using sha256
    return outputBytes as Base64 encoded string
end function
```

2) Data Encryption Flow stepsenumerate1 [steps,1]label=Step 0:

3) Hash the original data using the SHA-256 RSA algorithm: This generates a unique, fixed-length hash value that represents the original data.

Algorithm 3 EncryptAES256(plainText, publicKey)

```
function ENCRYPTAES256(plainText, publicKey)
    publicKeyBytes ← publicKey
    iv ← create new byte array of size 16
    plainBytes ← convert plainText to bytes using UTF-8 encoding
    aesAlg ← create new AES object
    aesAlg.KeySize ← 256
    aesAlg.Key ← publicKeyBytes
    aesAlg.IV ← iv
    aesAlg.Mode ← CBC
    encryptor ← create encryptor object using aesAlg Key and IV
    memoryStream ← create new memory stream object
    cryptoStream ← create new crypto stream object using memoryStream, encryptor and Write mode
    cryptoStream.Write(plainBytes, 0, plainBytes.Length)
    cryptoStream.FlushFinalBlock()
    cipherBytes ← convert memoryStream content to byte array
    return cipherBytes as Base64 encoded string
end function
```

4) Encrypt the original data using the RSA 256-bit public computed key: This ensures that the data is secure during transmission and can only be decrypted by someone who has access to the public computed key, following is the encryption algorithm.

Algorithm 4 DecryptAES256(cipherText, publicKey)

```
function DECRYPTAES256(cipherText, publicKey)
    privateKeyBytes ← publicKey
    iv ← create new byte array of size 16
    cipherBytes ← convert cipherText to bytes using Base64 decoding
    aesAlg ← create new AES object
    aesAlg.KeySize ← 256
    aesAlg.Key ← privateKeyBytes
    aesAlg.IV ← iv
    aesAlg.Mode ← CBC
    decryptor ← create decryptor object using aesAlg Key and IV
    memoryStream ← create new memory stream object using cipherBytes
    cryptoStream ← create new crypto stream object using memoryStream, decryptor and Read mode
    plainBytes ← create new byte array of size cipherBytes.Length
    decryptedByteCount ←cryptoStream.Read(plainBytes, 0, plainBytes.Length)
    return plainBytes as UTF-8 encoded string from index 0 to decryptedByteCount
end function
```

5) Concatenate the hash value with the encrypted data: This ensures that the hash value is bound to the encrypted data and cannot be altered without detection.

6) Add the concatenated data on a blockchain: This ensures that the data is secured by the distributed ledger and cannot be altered without detection.

7) Send the concatenated data: using a secured channel such as a Token-based Representational State Transfer (REST) Application Programming Interface (API), as shown in the process flow below.

8) Retrieve the concatenated data: from the blockchain: The data can be retrieved using a specific transaction ID or other identification methods.

9) Verify the integrity of the data: using the hash value: The hash value can be recalculated from the retrieved data and compared to the original hash value to ensure that the data has not been tampered with.

10) Decrypt the encrypted data: using the public computed key: This ensures that the original data can be recovered.

11) Convert the decrypted data back into the original format: This ensures that the data can be used in its original form. Using Secure Hash Algorithm (SHA)-256 and RSA 256-bit key pair with blockchain provides an additional layer of security and integrity to the data being transmitted.

12) Blockchain: Blockchain technology has the potential to revolutionize the healthcare industry by enabling secure and decentralized storage and sharing of patient data. Blockchain technology can help by enabling encrypted, distributed storage of data that can be securely shared for public health purposes [27]. In addition, blockchain networks can be used in the healthcare system to preserve and exchange patient data through hospitals, diagnostic laboratories, pharmacy firms, insurance companies, and other healthcare providers. This can help improve the security and integrity of healthcare data management. Blockchain-based healthcare data management system between multiple stakeholders (nodes) within a healthcare ecosystem [28].

Algorithm 5 DICOM Encryption Algorithm

```
function ENCRYPTDICOMFILE(inputFile, outputFile, key, iv)
    originalFileBytes ← read all bytes from the inputFile
    aes ← create new AES object
    aes.KeySize ← 256
    aes.Key ← convert key from Base64 string to byte array
    aes.IV ← convert iv from Base64 string to byte array
    aes.Padding ←PKCS7
    encryptor ← create encryptor object using aes
    encryptedStream ← create new memory stream object
    cryptoStream ← create new crypto stream object using encryptedStream, encryptor and Write mode
    cryptoStream.Write(originalFileBytes, 0, originalFileBytes.Length) cryptoStream.FlushFinalBlock()
        encryptedFileBytes ← convert encryptedStream to byte array
        write all bytes from encryptedFileBytes to the outputFile
end function
```

13) Secure Sockets Layer (Secure Sockets Layer (SSL)) and Transport Layer Security (Transport Layer Security (TLS)): SSL and TLS are cryptographic protocols that provide secure communication over the internet. These protocols use a combination of symmetric and asymmetric encryption, along with digital certificates, to secure data transmission [22].

- Secure File Transfer Protocols: Medical patient files can contain a variety of file types depending on the type of information being recorded and the needs of the healthcare provider. Here we are focusing on the information that might be fetched from machines or third-party systems in the form of files:

- Medical images: Medical imaging files such as X-rays, CT scans, MRIs, and ultrasounds are often included in patient files.

- Lab results: Reports from laboratory tests such as blood tests, urine tests, and biopsies are commonly included in patient files.

- Consent forms: Documents that record patient consent for medical procedures, treatment plans, and research studies.

- DICOM files are created by medical imaging equipment, and they contain both image data and information about the patient and the imaging procedure. This information includes patient demographics, imaging protocols, acquisition parameters, and image annotations [6].

Here we are going to secure the file transfer channel by using Secure File Transfer Protocol (SFTP) protocol and Hypertext Transfer Protocol (HTTP), to ensure that data is transferred securely. The file itself should be encrypted while transferring from the providers to the receivers (payers and claim transaction system), below the Digital Imaging and Communications (DICOM) encryption and decryption algorithm.

Algorithm 6 DICOM Decryption Algorithm

```
function DECRYPTDICOM-FILE(inputFile, outputFile, key, iv)
    encryptedFileBytes ← read all bytes from the inputFile
    aes ← create new AES object
    aes.KeySize ← 256
    aes.Key ← convert key from Base64 string to byte array
    aes.IV ← convert iv from Base64 string to byte array
    aes.Padding ← PKCS7
    decryptor ← create decryptor object using aes
    decryptedStream ← create new memory stream object
    cryptoStream ← create new crypto stream object using encryptedFileBytes, decryptor and Read mode
    cryptoStream.CopyTo(decryptedStream)
    decryptedFileBytes ← convert decryptedStream to byte array
    write all bytes from decryptedFileBytes to the output File
end function
```

14) Anonymization and Aggregation: Anonymization removes identifiable information from data to protect patient privacy, using techniques like hashing, tokenization, and generalization, with tools such as MapReduce and Pig from Hadoop assisting in this process. Aggregation combines and summarizes data, helping minimize identification risks by presenting broader statistics, as seen when insurance companies aggregate health claims by zip code or age range. While these methods reduce data breach risks and maintain data utility for research, there's still some re-identification risk with anonymized and aggregated data, necessitating additional safeguards depending on data sensitivity and re-identification risks [25].

15) Data Accessibility: Due to the sensitivity of patient information, it is crucial that access to such data is restricted and not available to everyone who has access to the system, including healthcare providers and insurance companies. Once a doctor is providing service to a patient, they are allowed access to the EMR, but the patient's medical and personal information should be safeguarded with a high level of security after the service is provided. If necessary, access to the patient's EMR can be requested with the patient's consent through an electronic approval process. The patient can grant access through one of the following options:

- OTP: One-Time Password (OTP) stands for "one-Time Password." An OTP is a unique code or password that is valid for a single use, typically to verify a user's identity or for securing an online transaction. OTP is often used as an additional layer of security in multi-factor authentication, where users are required to provide more than one form of authentication before being granted access to a system or performing a sensitive action.
    - OTP: can be delivered to the user via various methods, such as:
    - Short Message Service (SMS): The OTP is sent to the user's mobile phone as a text message.
    - Email: The OTP is sent to the user's email address.
    - Authenticator app: The user installs an authenticator app on their mobile device, which generates a new OTP every few seconds.
    - Voice call: The OTP is delivered to the user via an automated voice call.

Here once the patient receives the OTP, he must provide it to the requester within a limited period before it expires. If the OTP is entered correctly and within the valid period, the doctor or physician is granted access or permission to perform the requested action. If the OTP is not entered correctly or within the valid period, the doctor may be locked out of the system or required to request new access.

- EMR Authenticator: Here we can integrate with any of the identity authenticators that supports the Open Authentication (OATH) Time-Based One-Time Password (TOTP) standard such as Google Authenticator, or Microsoft Authenticator. Additionally, Microsoft Authenticator also supports push notifications, which can provide an additional layer of security and convenience for users.

16) Big Data (Hadoop Security)

- Authentication: Authentication is the process of verifying the identity of users and ensuring that they have the necessary permissions to access the Hadoop cluster. Hadoop supports several authentication mechanisms such as Kerberos and Lightweight Directory Access Protocol (LDAP) [13].
- Authorization: Authorization is the process of determining what users are allowed to do once they have been authenticated. Hadoop provides mechanisms for fine-grained access control, such as Access Control Lists (Access Control Lists (ACL)s) and Role-Based Access controls (Role-Based Access Controls (RBAC)).
- Encryption: Encryption is the process of encoding data so that it cannot be read by unauthorized users. Hadoop provides encryption mechanisms for data at rest and in transit [29]. For data at rest, Hadoop supports encryption at the file system level using tools like HDFS Transparent Encryption [30]. For data in transit,

Hadoop supports encryption using Secure Sockets Layer (SSL) and Transport Layer Security (TLS).

- Auditing: Auditing is the process of recording all user activities on the Hadoop cluster. Hadoop provides auditing tools such as Apache Ranger and Apache Atlas, which can be used to track user activity and detect any security breaches.
- Network security: Hadoop clusters should be deployed in a secure network environment, with firewalls and other security measures in place to prevent unauthorized access. Network security can also include isolating different parts of the cluster and ensuring that only authorized users have access to sensitive data.[2]

17) Data Monitoring and Auditing Healthcare facilities and insurance companies need robust monitoring and auditing systems to ensure patient care and data quality, safety, and security. These systems monitor user access, track changes to records, and detect unauthorized or suspicious activity, ensuring data is used only for authorized purposes by authorized individuals [15]. They provide real-time alerts on security threats, enabling immediate action to protect patients and their data. Monitoring and auditing are crucial for maintaining data confidentiality, integrity, and availability, and ensuring compliance with regulations and best practices in patient data security [15] [31-41].

V. Discussion

In conclusion, the security techniques discussed in this paper provide a comprehensive approach to addressing the security and privacy concerns associated with healthcare data. By using data encryption, blockchain technology, secure API, and Hypertext Transfer Protocol Secure (HTTPS)/SFTP protocols, we can ensure the security and integrity of medical data both at rest and in transit. Additionally, the use of Kerberos can help ensure big data security. To ensure compliance with HIPAA standards and other regulations, it is important to limit data access to authorized parties only. Furthermore, in order to maintain a balance between performance and security, multi-layered security approaches have been employed, including big data security. It is important to note that the complexity of data processing can lead to poor performance and high latency, which is why it is crucial to carefully design and implement security measures that do not compromise the performance of the system. Overall, this proposed ecosystem provides a robust and secure solution for managing healthcare data and ensuring the privacy and security of patients' sensitive information.

VI. Conclusion

In conclusion, safeguarding healthcare data security is crucial to ensure patient privacy and uphold the confidentiality of their information. While advanced technologies like cloud computing bring numerous benefits to the healthcare industry, they also give rise to new security challenges that demand effective management. In this paper, we have applied a range of techniques, such as a customized Blockchain encryption mechanism, a combination of asymmetric and symmetric algorithms for data encryption, hashing, secure network protocols like HTTPS and SFTP, protection of APIs and resources within a cloud environment, and the incorporation of HDFS security.

The security strategies discussed in this paper form a solid foundation for healthcare organizations to establish a secure infrastructure that adheres to pertinent regulations and safeguards sensitive data. This paper highlights a multi-layered security approach that balances performance and data processing without compromising robust protection against cyber threats. After applying these techniques and integrating them, a highly secure data streaming system was achieved. During testing, neither passive nor active attacks were able to penetrate the protected data, demonstrating the effectiveness of the combined security measures.

It is essential for healthcare organizations to prioritize data security and consistently evaluate and improve their security measures to keep pace with emerging risks. By embracing a proactive and comprehensive approach to security, which includes the techniques outlined above, healthcare organizations can ensure the safety and privacy of their patients' data, instilling trust and confidence among all stakeholders.

References

[1] X. Yan, S. Feng, Y. Tang, P. Yin, and D. Deng, “Blockchain-based verifiable and dynamic multi-keyword ranked searchable encryption scheme in cloud computing,” J. Inf. Secur. Appl., vol. 71, no. 103353, p. 103353, Dec. 2022.

[2] K. Soumya and S. Nath Mishra, “Big data security issues from the perspective of IoT and cloud computing: A review,” Recent Advances in Computer Science and Communications, vol. 12, no. 1, pp. 1–22, 2020.

[3] S. A. Parsa, G. M. Parah, and K. Bhat, “A security management framework for big data in smart healthcare,” Big Data Research, vol. 25, 2021.

[4] H. Khaloufi, K. Abouelmehdi, A. Beni-hssane, and M. Saadi, “Security model for big healthcare data lifecycle,” Procedia Comput. Sci., vol. 141, pp. 294–301, 2018.

[5] Aqeel-ur-Rehman, I. U. Khan, and S. u. Rehman, “A review on big data security and privacy in healthcare applications,” in Big Data Management. Cham: Springer International Publishing, 2017, pp. 71–89.

[6] M. Siddique, M. A. Mirza, M. Ahmad, J. Chaudhry, and R. Islam, “A survey of big data security solutions in healthcare,” in Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering, ser. Lecture notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering. Cham: Springer International Publishing, 2018, pp. 391–406.

[7] B. Isabel and M. Garc´ıa-Zapirain, “Analysis of security in big data related to healthcare,” Journal of Digital Forensics, Security and Law, vol. 12, no. 3, 2017.

[8] B. Saraladevi, N. Pazhaniraja, P. V. Paul, M. S. S. Basha, and P. Dhavachelvan, “Big data and hadoop-a study in security perspective,” Procedia Comput. Sci., vol. 50, pp. 596–601, 2015.

[9] K. Pandey, A. I. Khan, Y. B. Abushark, M. M. Alam, A. Agrawal, R. Kumar, and R. A. Khan, “Key issues in healthcare data integrity: Analysis and recommendations,” IEEE Access, vol. 8, pp. 40 612–40 628, 2020.

[10] K. Sharma, A. Agrawal, D. Pandey, R. A. Khan, and S. K. Dinkar, “RSA based encryption approach for preserving confidentiality of big data,” J. King Saud Univ. - Comput. Inf. Sci., vol. 34, no. 5, pp. 2088–2097, May 2022.

[11] M. Tanrıverdi, A Systematic Review of Privacy-Preserving Healthcare Data Sharing on Blockchain, 2020.

[12] P. Xi, X. Zhang, L. Wang, W. Liu, and S. Peng, "A review of blockchain-based secure sharing of healthcare data," Appl. Sci. (Basel), vol. 12, no. 15, p. 7912, Aug. 2022.

[13] M. Choudhary, A. Singh Yadav, D. Kumar Yadav, and V. Pawar, "A review on hadoop security issues," A Review on Hadoop Security Issues.

[14] M. N. Ramachandra, M. Srinivasa Rao, W. C. Lai, B. D. Parameshachari, J. Ananda Babu, and K. L. Hemalatha, "An efficient and secure big data storage in cloud environment by using triple data encryption standard," Big Data Cogn. Comput., vol. 6, no. 4, p. 101, Sep. 2022.

[15] R. U. Haque, A. S. M. T. Hasan, Q. Jiang, and Q. Qu, "Privacy-preserving k-nearest neighbors training over blockchain-based encrypted health data," Electronics (Basel), vol. 9, no. 12, p. 2096, Dec. 2020.

[16] P. N. Srinivasu, A. K. Bhoi, S. R. Nayak, M. R. Bhutta, and M. Woz´niak, "Blockchain technology for secured healthcare data communication among the non-terminal nodes in IoT architecture in 5G network," Electronics (Basel), vol. 10, no. 12, p. 1437, Jun. 2021.

[17] R. Hamza, A. Hassan, A. Ali, M. B. Bashir, S. M. Alqhtani, T. M. Tawfeeg, and A. Yousif, "Towards secure big data analysis via fully homomorphic encryption algorithms, Entropy (Basel), vol. 24, no. 4,p. 519, Apr. 2022.

[18] J. Moura and C. Serra˜o, "Security and privacy issues of big data," in Cloud Security. IGI Global, 2019, pp. 1598–1630.

[19] G. Begum, S. Z. U. Huq, and A. P. S. Kumar, "Sandbox security model for hadoop file system," J. Big Data, vol. 7, no. 1, Dec. 2020.

[20] Olaronke and O. Oluwaseun, "Big data in healthcare: Prospects, challenges and resolutions," in 2016 Future Technologies Conference (FTC). IEEE, Dec. 2016.

[21] Matturdi, X. Zhou, S. Li, and F. Lin, "Big data security and privacy: A review," China Commun., vol. 11, no. 14, pp. 135–145, 2014.

[22] K. Abouelmehdi, A. Beni-Hssane, H. Khaloufi, and M. Saadi, "Big data security and privacy in healthcare: A review," Procedia Comput. Sci., vol. 113, pp. 73–80, 2017.

[23] H. Patil and R. Kupwade, "Big data security and privacy issues in healthcare," in 2014 IEEE international congress on big data. IEEE, 2014, pp. 762–765.

[24] Y. Himeur, S. S. Sohail, F. Bensaali, A. Amira, and M. Alazab, "Latest trends of security and privacy in recommender systems: A comprehensive review and future perspectives," Comput. Secur., vol. 118, no. 102746, p. 102746, Jul. 2022.

[25] M. Paul, L. Maglaras, M. A. Ferrag, and I. Almomani, "Digitization of healthcare sector: A study on privacy and security concerns," ICT Express, Feb. 2023.

[26] P. Matta, M. Arora, and D. Sharma, "A comparative survey on data encryption techniques: Big data perspective," Mater. Today, vol. 46, pp. 11 035–11 039, 2021.

[27] L. Guo, H. Xie, and Y. Li, "Data encryption based blockchain and privacy preserving mechanisms towards big data," J. Vis. Commun. Image Represent. vol. 70, no. 102741, p. 102741, Jul. 2020.

[28] W. Y. Ng, T. E. Tan, P. V. H. Movva, A. H. S. Fang, K. K. Yeo, D. Ho, F. S. S. Foo, Z. Xiao, K. Sun, T. Y. Wong, A. T. H. Sia, and D. S. W. Ting, "Blockchain applications in health care for covid 19 and beyond: a systematic review," Lancet Digit. Health, vol. 3, no. 12, Dec. 2021.

[29] S. Gattoju and N. Vadlamani, "A survey on security of the hadoop framework in the environment of BigData," J. Phys. Conf. Ser., vol. 2089, no. 1, p. 012031, Nov. 2021.

[30] W. Rajeh, "Hadoop distributed file system security challenges and examination of unauthorized access issue," J. Inf. Secur., vol. 13, no. 02, pp. 23–42, 2022.

[31] J. J. Hathaliya and S. Tanwar, "An exhaustive survey on security and privacy issues in healthcare 4.0," Comput. Commun., vol. 153, pp. 311–335, Mar. 2020.

[32] H. M. El-Bakry, "An Efficient Algorithm for Pattern Detection using Combined Classifiers and Data Fusion," Information Fusion Journal, vol. 11, issue 2, April 2010, pp. 133-148.

[33] H. M. El-Bakry, and Nikos Mastorakis "New Fast Normalized Neural Networks for Pattern Detection," Image and Vision Computing Journal, vol. 25, issue 11, 2007, pp. 1767-1784.

[34] H. M. El-Bakry, and Nikos Mastorakis, "A New Fast Forecasting Technique using High Speed Neural Networks," WSEAS Transactions on Signal Processing, vol. 4, Issue 10, Oct. 2008, pp. 573-595.

[35] H. M. El-Bakry, "Fast Virus Detection by using High Speed Time Delay Neural Networks," Journal of Computer Virology, vol.6, no.2, 2010, pp.115-122.

[36] H. M. El-Bakry, and Nikos Mastorakis, "Realization of E-University for Distance Learning," WSEAS Transactions on Computers, vol. 8, issue 1, Jan. 2009, pp. 48-62.

[37] N. El-Rashidy, L. Alarabi, H. M. El-Bakry, S. Abdelrazek, T. ABUHMED, F. Ali and S. El-Sappagh "Sepsis Prediction in Intensive Care Unit based on genetic feature optimization and stacked deep ensemble learning," Neural Computing and Applications, vol. 34, pp. 3603–3632, 2022.

[38] H. El-Bakry, "Face Detection Using Neural Networks and Image Decomposition," Proc. of INNS-IEEE International Joint Conference on Neural Networks, 12-17 May, 2002, Honolulu, Hawaii, USA.

[39] H. El-Bakry, "Fast Face Detection Using Neural Networks and Image Decomposition," Proc. of the 6th International Computer Science Conference, AMT 2001, Hong Kong, China, December 18-20, 2001, pp.205-215.

[40] H. M. El-Bakry and M. Hamada, "A New Implementation for High Speed Neural Networks in Frequency Space," Lecture Notes in Artificial Intelligence, Springer, KES 2008, Part I, LNAI 5177, pp. 33-40.

[41] H. M. El-Bakry, and Q. Zhao, "Fast Time Delay Neural Networks," International Journal of Neural Systems, vol. 15, no.6, December 2005, pp.445-455.