Sarcouncil Journal of Engineering and Computer Sciences

ISSN(Online): 2945-3585

Volume- 03| Issue- 9| 2024

Research Article

Received: 27-07-2024| **Accepted:** 15-08-2024 | **Published:** 28-09-2024

An Effective Graph Database-Centric Patient Healthcare Data Management Using A Robust HGCRN and LIM2DCE

Vinodkumar Reddy Surasani[1], Srinivasa Reddy Mukkala[2], Vijaya Kumar Pothireddy[3] and Geethanjali Sanikommu[4]

[1]Sr Software Engineer, RBC Wealth Management, Minneapolis, MN-55401, USA

[2]Director of Statistical Programming and Biostatistics, Episdata inc. Sterling Heights, MI-48313, USA

[3]Software Engineer, Google LLC. Cambridge, MA-02142, USA

[4]Application Developer, Flagstar Bank, Troy, Michigan-48098, USA

Abstract: In basic, effective data management in the healthcare domain is a difficult task due to the complex relationship between the patients and the healthcare sector. However, none of the prevailing models focused on analyzing the complex relationships and hidden patterns between the client data for identifying the intention of the patients. Therefore, this article proposes an effective graph database-grounded patient healthcare data management using a robust HGCRN and LIM2DCE. Firstly, the patients are registered into the healthcare applications by entering their details, followed by privacy preservation. Subsequently, a patient's unique identifier is generated regarding the patient's information. Next, the patient logs in to the application, and then the healthcare data is encrypted to ensure data privacy. Here, the encrypted data is subjected to the proposed readmission classification model. Initially, the historical dataset is pre-processed. Thereafter, the outlier detection and elimination are done, followed by dimensionality reduction. Next, the graph is constructed by using the neo4j bloom. Also, the client is segmented according to the similar information. Further, the nodes and edges are extracted and then inputted to the proposed HGCRN, where the readmission prediction is done. Accordingly, the intention of the patients is described via the LIM2DCE. The proposed work had an impressive performance with 98.4578% accuracy.

Keywords: Client Data Management (CDM), Healthcare Sector (HS), Complex Client Relationships (CCR), Hidden Patterns (HP), Client Segmentation (CS), Graph Databases (GD), and Local Interpretable Model-agnostic 2D Cigar Explanation (LIM2DCE).

INTRODUCTION

In today's digital era, healthcare institutions produce enormous volumes of data every day (Sharma *et al.*, 2023). This data, encompassing patient records, diagnostic reports, and treatment histories, plays a critical role in enhancing healthcare services (Thantilage *et al.*, 2023) (Lodha *et al.*, 2023). The effective management of this data is important to provide high-quality medical services (Singh *et al.*, 2022). However, efficiently managing such large-scale and complex data is a challenging task, thus leading to data fragmentation, accessibility, and security concerns (Adere, 2022).

Hence, effective healthcare data management is important to ensure seamless patient care, accurate diagnosis, and efficient clinical workflows (Thapa & Camtepe, 2021) (Rani *et al.*, 2023). Healthcare data management refers to the process of collecting, storing, organizing, securing, and analyzing healthcare-related data (Calazans *et al.*, 2024). Proper data handling not only enhances operational efficiency but also supports evidence-based decision-making and personalized treatment plans (Ajagbe *et al.*, 2022). Also, in the context of patient readmission, effective data management allows healthcare providers to identify at-risk patients, implement early interventions, and reduce unnecessary hospital readmission (Khayat *et al.*, 2025).

Various Machine Learning (ML) and Deep Learning (DL)-based methodologies have been implemented for effective patient healthcare data management. ML techniques offer early disease detection and personalized treatment plans by identifying complex patterns in vast datasets (Shah *et al.*, 2024). However, they face several limitations, including high-quality data, potential algorithm bias, and complex model interpretation (Abdulmalek *et al.*, 2022). Similarly, DL methodologies provide significant advantages in patient healthcare data management by analyzing complex medical data, identifying patterns, and predicting disease progression (Yaqoob *et al.*, 2022). Yet, they require large amounts of data, high computational costs, etc., thus hindering overall performance (Rahman *et al.*, 2022). Additionally, some prevailing works offered ineffective data management owing to the presence of duplicate patient details, outliers in healthcare data, and high-dimensional data (Aldwairi *et al.*, 2023). Moreover, none of the traditional methodologies concentrated on analyzing the complex relationships and hidden patterns between the client data for identifying the intention behind the customer using the healthcare

*Corresponding Author: Vinodkumar Reddy Surasani

DOI- https://doi.org/ 10.5281/zenodo.15041597

services. Hence, the proposed framework introduces a graph database-based healthcare data management system using HGCRN and LIM2DCE for accurate patient readmitting prediction.

Problem Statement

The limitations of several traditional methodologies are depicted as follows,

- None of the conventional works analyzed the complex relationships and hidden patterns between the client data for patient intention identification.
- (Taloba *et al.*, 2023) failed to perform patient identity matching, thus leading to duplicate or mismatched records.
- (Gupta *et al.*, 2024) had inaccurate analysis since it did not detect and eliminate the outliers in healthcare data.
- Due to the high-dimensionality data, (Demirbaga & Aujla, 2022) failed to identify the hidden patterns in healthcare data.
- Most of the works had privacy risks owing to the lack of protecting patients' sensitive information from unauthorized access.

OBJECTIVE

The objectives of the proposed framework are illustrated as follows,

- The graph database-based patient healthcare data management is conducted for accurate patient intention prediction.
- To perform patient identity matching, the unique patient identifier is generated using the Merkle tree.
- The EYEVA is employed to detect and eliminate outliers in healthcare data.
- The dimensionality of the data is reduced using the CPHCA.
- By using the CMC-KA, the privacy of the patient's sensitive information is preserved.

The structure of the paper is outlined as follows: Section 2 illustrates the literature survey, Section 3 demonstrates the proposed methodology, Section 4 indicates the result and discussion, and Section 5 concludes the proposed work with future recommendations.

RELATED WORKS

(Taloba *et al.*, 2023) implemented multimedia data processing in the IoT-powered healthcare sector via a blockchain-centric hybrid platform. IoT devices were utilized to record the individual activity of the patients. Then, the IoT data, such as intermediary activities, health records data, and drug shipment phenomena were stored in the blockchain to track the complex behavior of the pathways. Also, illegal activities were constantly monitored in the blockchain. Here, blockchain technology was established to maintain the secrecy and security of the data control systems in real time. However, this framework had considerable transaction duration owing to the hashing process, which consumed maximum time as the number of nodes increased. Also, this model failed to perform effective patient identity management.

(Gupta *et al.*, 2024) demonstrated a robust quantum cyber-security scheme-based healthcare data management. This work established quantum encryption to ensure secure data storage and healthcare data communication over the shared cloud environment. Additionally, a quantum feed-forward neural network unit was employed to recognize the intention of the request, thus minimizing the data breaches. Thus, the developed model had high robustness and consistency. However, this framework had maximum error due to the increased number of qubits and outliers.

(Demirbaga & Aujla, 2022) introduced a blockchain-based healthcare service management in an IoT-enabled big data ecosystem. Here, a scalable computing system (Map-Reduce) was established to provide verifiable data access mechanisms. Also, the blockchain was employed to ensure secure healthcare data storage. This framework had high scalability and dependability. However, this system was ineffective due to the maximum gas consumption and high dimensional data.

(Gohar *et al.*, 2022) propounded a resilient and effective model based on IoT-Cloud-Blockchain called a patient-centric healthcare system for effective healthcare systems interoperability. Here, a tiered-based architecture (5-tier) with collaboration was introduced to ensure a feasible patient-centric healthcare system. This framework achieved effective data sharing with better interoperability. However, this model had a maximum risk of data breaches and unauthorized access due to the data distribution across multiple providers.

(Ahamed *et al.*, 2023) established effective healthcare data management based on an energy-aware IoT device-assisted wearable sensor platform. Initially, the healthcare data was gathered by using the IoT wearable sensors. Then, the sensed IoT data was subjected to IoT-enabled

Publisher: SARC Publisher

healthcare monitoring. Thereafter, the healthcare outcomes were processed in the energy-efficient data management system. The data management layer involved the resource classifier, blockchain engine, shared database, and data flow indicator. Finally, the output data was updated in the local database. This model had high computational efficiency. However, this approach was inadequate to support real-time data processing.

(Abbas *et al.*, 2024) investigated Internet of Medical Things (IoMT)-based blockchain-assisted secure data management model for healthcare information analysis. Initially, the sensitive healthcare data was sensed from the patients via IoMT devices like smartwatches and smartphones. Thereafter, the healthcare information was transmitted to the blockchain-enabled cloud server through the access point. Finally, the trusted authority and intermediates securely accessed the healthcare data stored in the cloud. This model provided secure data management between personal servers and implantable medical devices. Also, this approach ensured secure data distribution between cloud servers and personal servers. Thus, the experimental outcomes proved that the model had maximum accuracy and minimum response time. Nevertheless, this system had storage overhead issues.

(Islam *et al.*, 2023) depicted a blockchain-assisted decentralized trustworthy system in healthcare management with cyber protection. Here, blockchain was introduced to ensure a secure environment for the privacy-preserving health data management ledger based on hash processing. This framework had high data security and storage immutability. But, this model had the risk of insider threats, where the individuals who have access to the database can manipulate the sensitive data.

(Zala *et al.*, 2022) demonstrated a patient's e-healthcare records management system. Here, a customized steganographic encryption system was employed to store the e-healthcare information at the third-party cloud servers in a reliable and secure way. Also, data anonymization was done to preserve the user's sensitive information. This framework had minimum latency and maximum throughput. However, this model had considerable processing delay and execution time.

(Arul *et al.*, 2021) propounded a multi-modal secure healthcare data distribution scheme based on IoMT-powered blockchain. Primarily, the bio signals were captured from the patients through the IoMT devices. Then, the healthcare data was updated in the blockchain and then securely stored in the cloud database. This model had high data integrity and efficacy. However, this framework had the risk of information loss.

(Almalawi *et al.*, 2023) investigated secure healthcare data management for a modern healthcare system. Here, a Lionized Remora Optimization-based Serpent (LRO-S) encryption method was employed to reduce data breaches and unauthorized access. This framework had minimum encryption time and minimum cost. Nevertheless, this framework had poor convergence and suboptimal results due to the random search criteria.

PROPOSED METHODOLOGY fOR GRAPH DATABASE-CENTRIC PATIENT HEALTHCARE DATA MANAGEMENT

In the proposed work, an effective graph database-grounded patient healthcare data management is implemented using a robust HGCRN and LIM2DCE. The proposed HGCRN significantly predicts the readmission. Accordingly, the proposed LIM2DCE is introduced to explain the intention of the patients. The conceptual framework of the proposed work is shown in the Figure 1

Publisher: SARC Publisher

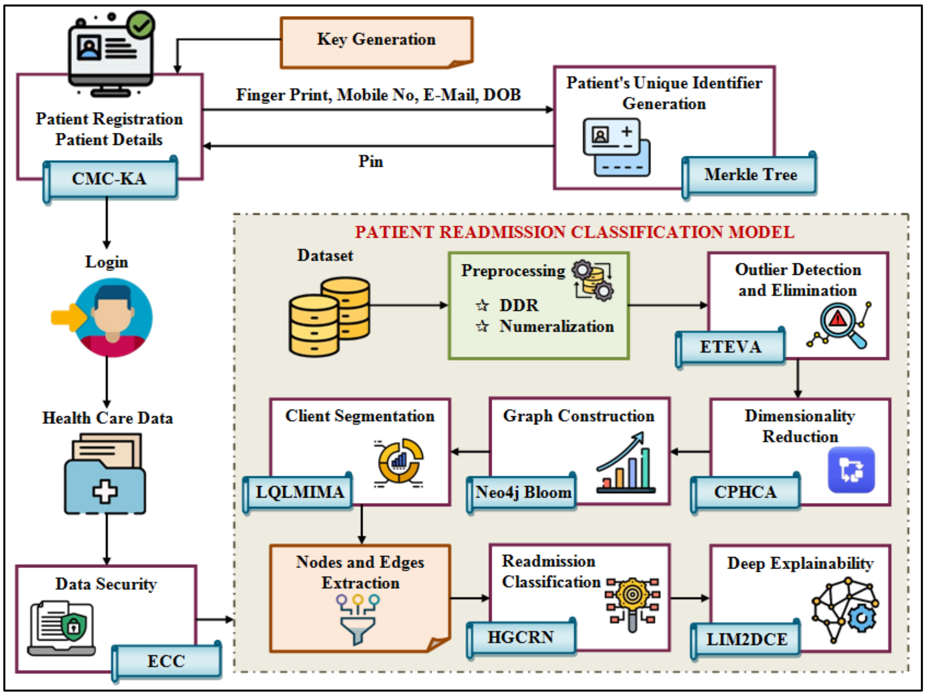

Figure 1: The structural representation of the proposed work

The research methodology includes major steps, such as patient registration, privacy preservation, patient unique identifier generation, data security, outlier detection and elimination, graph construction, client segmentation, readmission classification, and deep explainer. Thus, the mathematical workflow of the proposed approach is explained further.

Patient Registration

The patient registration process in the healthcare sector is a prime step that ensures secure data handling and effective healthcare delivery. In the proposed work, primarily, the patients are registered into the healthcare applications (H_∞) by providing their personal details (name, email, address, fingerprint, contact information, date of birth, and gender), medical conditions, and financial details.

$$\sum_{p=1}^{P}\left(X_1, X_p \dots X_P\right) \xrightarrow[\partial\nabla]{register} H_\infty \qquad (1)$$

Where, P indicates the number of registered users (X_P) and $\partial\nabla$ exhibits the patient's details. At the time of registration, the security keys, including the public key and private key, are created for the registered users based on the Elliptic Curve Cryptography (ECC), which is derived in Section 3.4.

To add the extra layer of security, the patient's information $(\wp)$ is privacy preserved using the proposed Cramer-von Mises Criterion-KAnonymity (CMC-KA). The K-Anonymity significantly obscures the sensitive information while maintaining usability. However, it struggled with complex datasets, thus resulting in poor scalability and higher cost. Therefore, the Cramér–von Mises (CM) criterion is employed in the proposed model to enhance the anonymization task. The anonymization process includes two steps, such as generalization and suppression. Several values of the elements are replaced by an asterisk (*) according to the suppression $(\delta\mu)$. In the generalization process (φ), the individual values of elements are represented in a wide range based on the generalization level determined by the CM criterion $(C\vartheta)$.

$$C\vartheta = \frac{1}{T}\sum_{t=1}^{T}\left(\hbar(\wp_t) - \mathrm{O}(\wp_t)\right)^2 \qquad (2)$$

Publisher: SARC Publisher

Where, $t = 1 to T$ depicts the number of samples, $\hbar(\wp_t)$ denotes the target data, and $\mathrm{O}(\wp_t)$ indicates the original data.

$$\Omega \xrightarrow{preserve} \left| \begin{array}{l} \delta\mu \xrightarrow{replace} *(\wp) \\ \varphi(C\vartheta) \xrightarrow{define} \langle i \leq \wp \leq I \rangle \end{array} \right. \quad (3)$$

Here, Ω represents the privacy-preserved data and $i \leq \wp \leq I$ denotes the category. The anonymization process helps to mitigate the data theft in the central database.

Patient's unique identifier generation

During patient registration, the unique identifier (PIN) is generated using the Merkle tree by considering the patient's information. The process of a patient's unique identifiers helps to diminish the risk of duplicate or mismatched records. A Merkle tree (hash tree) is a cryptographic data structure, which ensures data security and data integrity in healthcare applications. Here, the patient's information $(\wp)$ like fingerprint $(\wp_1)$, mobile number $(\wp_2)$, email $(\wp_3)$, and date of birth $(\wp_4)$ are considered as the input for the Merkle tree creation.

$$\wp = \{\wp_1, \wp_2, \wp_3, \wp_4\} \quad (4)$$

Then, each patient's information is converted into a hash value using a cryptographic hash function (λ). Subsequently, the hashed values are arranged as leaf nodes $(L^\circ = L^\circ{}_1, L^\circ{}_2, L^\circ{}_3, L^\circ{}_4)$ at the bottom of the Merkle tree.

$$L^\circ \rightarrow \begin{cases} L^\circ{}_1 = \lambda(\wp_1) \\ L^\circ{}_2 = \lambda(\wp_2) \\ L^\circ{}_3 = \lambda(\wp_3) \\ L^\circ{}_4 = \lambda(\wp_4) \end{cases} \quad (5)$$

To form parent nodes $\mathrm{P} = (\mathrm{P}_1, \mathrm{P}_2)$, each pair of hashes is added.

$$\mathrm{P} = \begin{cases} \mathrm{P}_1 = \lambda(L^\circ{}_1 \,\|\, L^\circ{}_2) \\ \mathrm{P}_2 = \lambda(L^\circ{}_3 \,\|\, L^\circ{}_4) \end{cases} \quad (6)$$

Thus, the above-mentioned steps are recursively continued until a single hash value (Merkle root) is obtained at the top of the tree.

$$\Lambda = \lambda(\mathrm{P}_1 \,\|\, \mathrm{P}_2) \quad (7)$$

The Merkle root (Λ) acts as the unique identifier (PIN) for the patient.

Login

In this phase, the patient logs in to the application using their username and password along with the unique identifier (PIN). After successful login, the patients provide healthcare data to communicate with healthcare providers.

$$\mathrm{Z}_n = \{\mathrm{Z}_1, \mathrm{Z}_2, \ldots\ldots \mathrm{Z}_N\} \quad (8)$$

Where, $n = 1 to N$ depicts the number of healthcare data Z_n. At the time of login, the healthcare data is collected from the patients and then subjected to data encryption, which prevents leakage of sensitive healthcare information.

Data Security

Here, the Z_n is fed into the ECC, which transforms the original healthcare data into an unreadable format, thereby protecting the data against security breaches. The ECC is the public key cryptography that follows algebraic structures of elliptic curves over finite fields. The ECC offers better security with a relatively

Publisher: SARC Publisher

smaller key size. The ECC includes three steps, namely key generation, encryption, and decryption, which are explained below,

Key generation: Initially, the elliptic curve $(E\infty)$ is computed based on the selected prime number.

$$E\infty \rightarrow U^2 = V^3 + cV + d \tag{9}$$

Here, U and V denote the plane coordinates and (c, d) depict the curve parameters. Then, the base point (β) is selected from the elliptic curve $E\infty$. Further, the private key $(\Im)$ is randomly chosen within a particular range. Based on the private key and base point, the public key $(\Re)$ is computed as below,

$$\Re = \Im \cdot \beta \tag{10}$$

Encryption: In basic, the public key is used to encrypt the data. To encrypt the healthcare data Z_n, two cipher points (R_1, R_2) are computed as follows,

$$R_1 = rd * \beta \tag{11}$$

$$R_2 = Z_n + (rd \cdot \Re) \tag{12}$$

Here, the pair (R_1, R_2) is considered as the encrypted message.

Decryption: The private key is used to perform data decryption.

$$R_3 = \Im \cdot R_1 \tag{13}$$

$$Z_n = R_2 - R_3 \tag{14}$$

Where, rd indicates the random integer and R_3 depicts the temporary point. Thus, the encrypted data is mentioned as R_e.

Patient readmission classification framework

To predict the patient readmission, the R_e is decrypted in this section. The proposed patient readmission classification model involves data collection, pre-processing, outlier detection and elimination, dimensionality reduction, graph construction, client segmentation, readmission classification, and deep explainer (patient's intention analysis). Patient intention analysis during readmission is fundamental to improving healthcare outcomes and optimizing resource allocation. Thus, the patient readmission classification model is elaborated as follows,

Dataset

To train the patient readmission classification model, the historical healthcare dataset (diabetes+130-us+hospitals+for+years+1999-2008) is collected. The collected dataset includes crucial information like demographic features, admission details, medical features, diagnostic details, health metrics, and readmission status.

$$G_q = \left\| G_1, G_2, \ldots\ldots G_Q \right\| \text{ Where, } q = 1\,to\,Q \tag{15}$$

Here, Q demonstrates the number of input data G_q. During the testing process, the decrypted data Z_n is considered as the input.

Pre-processing

Next, the G_q is subjected to pre-processing, which improves the quality of the data, thus resulting in high computational efficiency. The pre-processing involves two steps, namely Duplicate Data Removal (DDR) and numeralization. Generally, the raw dataset may contain duplicate or redundant information that mainly degrades the classifier's performance. Therefore, the proposed method eliminates the duplicate information from the G_q. In numeralization, the duplicate removed data is converted into the form of numerical vectors. Finally, the pre-processed data is shown as ς_k.

Outlier detection and elimination

After pre-processing, the pre-processed data ς_k is given to the outlier detection and elimination phase to upgrade the model's performance. In common, outliers are data points that entirely deviate from

Publisher: SARC Publisher

the majority of the data. In the context of healthcare data, outliers represent errors in data collection or extreme values. Therefore, outlier identification and handling is crucial to improve healthcare results. The proposed method introduces the Extreme Value Analysis (EVA) to detect the outliers. The EVA had high robustness. However, the random threshold values cause misinterpretation and suboptimal outcomes. Therefore, the proposed method establishes the Tsallis Entropy (TE) technique $(T\delta)$ to set the appropriate threshold values $(\tau\nu)$.

$$T\delta = \frac{1-\sum_{k=1}^{K} Pb(\varsigma_k)^{\mu}}{\mu - 1} \xrightarrow{set} \tau\nu \tag{16}$$

Here, K denotes the number of pre-processed data, Pb indicates the probability factor, and μ denotes the entropic index. Based on $\tau\nu$, the extreme values (outliers) are identified. Data points with entropy values that exceed the threshold values are assumed as outliers (o). Then, the outliers are removed from the pre-processed data ς_k to enhance the model's significance.

$$\Theta = \varsigma_k - o \tag{17}$$

Thus, the outlier-handled data is mentioned as Θ.

Dimensionality Reduction

Afterward, the Θ is fed into the proposed Confluent Principal Hypergeometric Component Analysis (CPHCA) to perform dimensionality reduction. The Principal Component Analysis (PCA) reduces the dimensions of the features while preserving the significant information. However, it was less effective due to the Gaussian distribution-based eigenvalue approximation. Hence, the proposed method introduces the confluent hypergeometric technique to perform eigenvalue approximation. Initially, the Θ is standardized into 0s and 1s.

$$\Phi = \left[\frac{\Theta - mn}{sd}\right] \tag{18}$$

Here, mn and sd indicate the mean and standard deviation, respectively. To analyze the relationship among the standardized features (Φ), the covariance matrix $(\eta\nabla)$ is computed.

$$\eta\nabla = \frac{1}{V-1} \cdot \sum_{v=1}^{V} \begin{bmatrix} (\Phi_1, \Phi_1) & (\Phi_1, \Phi_2) \\ (\Phi_2, \Phi_1) & (\Phi_2, \Phi_2) \end{bmatrix} \tag{19}$$

Regarding the covariance matrix, the eigenvectors and eigenvalues are determined.

$$\eta\nabla \cdot \rho = \ell^{\circ} \cdot \rho \tag{20}$$

Here, V denotes the number of standardized features, ρ and ℓ° depict the eigenvector (non-zero vector) and eigenvalue (scalar value), respectively. Then, the eigenvalue approximation is done by using the confluent hypergeometric technique (Ch_{geo}), which upgrades the system's consistency.

$$Ch_{geo}(u, f, w) = \sum_{v=0}^{\infty} \frac{(Rs)_v w^v}{(\ell^{\circ})_v v!} \tag{21}$$

Where, (u, f, w) denotes the parameters and $(Rs)_v$ depicts the rising factorial. Then, the approximated eigenvalues are sorted in descending order. Here, the eigenvector with the maximum eigenvalue is assumed as the principal component. Subsequently, the dimensionality of the data is reduced with respect to the principal component. Lastly, the dimensionality reduced data is shown as Iu.

Graph construction

Publisher: SARC Publisher

Subsequently, the graph is constructed from Iu based on the Neo4j bloom. Neo4j bloom is a graph visualization technique, which helps to interact with graph data. Once the neo4j bloom setup is designed, then the graph schema is determined as node labels (patient, doctor, and hospital), relationship types (edges), and indexes. Finally, the graph snapshot is exported from the neo4j bloom. Thus, the constructed graph is mentioned as $\partial\kappa$.

Client segmentation

Afterward, the $\partial\kappa$ is subjected to client segmentation, where the patients are grouped based on their similar details. Here, the Louvain Quantum Linear Mutual Information Modularity Algorithm (LQLMIMA) is employed to perform client segmentation in the healthcare sector. The Louvain Modularity Algorithm (LMA) had high scalability and adaptability. Also, the LMA significantly handles both the weighted and unweighted networks. However, the LMA had computational complexity owing to the random initialization. To address this issue, the quantum linear mutual information is used in the proposed model. A graph $\partial\kappa$ involves the nodes, edges, and edge weights. Here, nodes (N_{ode}) represent the patients or healthcare product customers. Likewise, edges (E_{dge}) represent relationships between patients like medical history (shared diagnoses or treatments), purchase behavior (similar medicines), and geographical proximity (hospital and clinic visits). Next, the edge weights are initialized based on the quantum linear mutual information (Q_{mut}), which is expressed below,

$$Q_{mut}(N_{ode1} : N_{ode2}) = N_{ent}(N_{ode1}) + N_{ent}(N_{ode2}) - N_{ent}(N_{ode1}, N_{ode2}) \qquad (22)$$

Where, N_{ent} depicts the Neumann entropy and (N_{ode1}, N_{ode2}) indicates node 1 and node 2, respectively. Based on Q_{mut}, the edge weights are assigned. Then, each node is considered as its own group. Afterward, all the nodes are merged to maximize the modularity. Finally, the segmented clients are mentioned as (A).

Nodes and edges extraction

From A, the nodes and edges are extracted for the purpose of readmission classification. Here, the nodes and edges represent the information, such as patient identifier, gender, age, weight, admission type ID, admission source ID, payer code, medical specialty, number of patients, and number of laboratory procedures.

Patient readmission classification

In this phase, the NE is inputted into the proposed Hexpo Graph Convolutional Ridge Networks (HGCRN), which predict the patient's readmission status. The Graph Convolutional Network (GCN) effectively handles complex relationships between nodes. However, the GCN has computational complexity and low learning efficiency. Therefore, the proposed work introduces ridge regularization to mitigate the computational complexity. Likewise, the hexpo activation function is used to improve learning efficiency. Thus, the architecture of the proposed HGCRN is given in Figure 2.

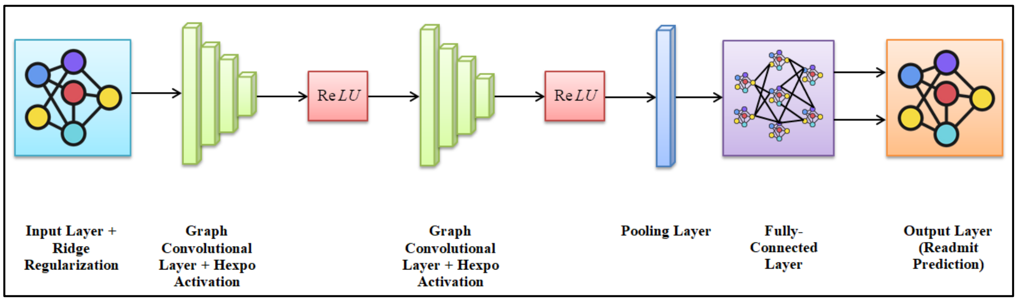

Figure 2: The schematic representation of the proposed HGCRN

Input layer: The input layer $(I\varpi)$ represents the graph as an adjacency matrix $(M\alpha)$ and a feature matrix $(F\varphi)$, which contain node features.

$$I\varpi = NE(M\alpha, F\varphi) \qquad (23)$$

Publisher: SARC Publisher

Ridge regularization: To reduce the computational overhead, ridge regularization $\left(\Diamond_{reg}\right)$ is employed in the proposed approach.

$$\Diamond_{reg} = \sum_{j=1}^{J} \left(Ro - wt^{T} I\varpi\right)^{2} + \phi \sum_{j=1}^{J} WT^{2} \tag{24}$$

Where, $j = 1 \, to \, J$ denotes the number of features, Ro indicates the regularized output, wt depicts the weight factor, ϕ represents the regularization parameter, and WT represents the large weight factor.

Hexpo activation function: Similarly, the hexpo activation function $\left(\lambda_{hex}\right)$ is introduced in the proposed method to increase the classifier performance.

$$\lambda_{hex}\left(Ro\right) = \tanh\left(\iota Ro\right) * \exp\left(\psi^{\circ} Ro\right) \tag{25}$$

Here, $\left(\iota, \psi^{\circ}\right)$ indicates the trainable parameters.

Graph convolutional layer: In this layer, information is distributed across neighboring nodes using graph convolution operations.

$$F\varphi'' = \lambda_{hex} \cdot \Diamond_{reg}\left(\partial^{mx^{-1/2}} M\alpha^{-1/2} F\varphi wt\right) \tag{26}$$

Where, $F\varphi''$ denotes the updated feature matrix and ∂^{mx} represents the degree matrix.

ReLU: Here, the ReLU activation $\left(R_{unit}\right)$ is applied to capture complex patterns in the graph data.

$$R_{unit}\left(F\varphi''\right) = \max\left(0, F\varphi''\right) \tag{27}$$

Pooling layer: Next, the global pooling operation $\left(Pl+\right)$ is applied to $R_{unit}\left(F\varphi''\right)$ to represent the node embedding into a single graph representation.

$$Pl + \xrightarrow{pooling} R_{unit}\left(F\varphi''\right) \tag{28}$$

Fully connected layer: Then, the fully connected layer $\left(Fc\right)$ converts the graph embedding into predictions.

$$Fc = \lambda_{hex} * \left(wt \cdot Pl\right) + bs \tag{29}$$

Where, bs indicates the bias value.

Output layer: Finally, the output layers present the classified outcome $\left(P_{out}\right)$.

$$P_{out} \rightarrow \left\{re, no\right\} \tag{30}$$

The proposed HGCRN predicts whether the patient will be readmitted within 30 days $\left(re\right)$ or not $\left(no\right)$. The pseudo-code of the proposed HGCRN is given below,

Input: Extracted nodes and edges NE
Output: Classified outcome P_{out}

Begin
Initialize NE, λ_{hex}, $F\varphi''$ and Fc
For 1 to each nodes do,
Perform input layer,

$$I\varpi = NE\left(M\alpha, F\varphi\right)$$

Apply ridge regularization $\Diamond_{reg} = \sum_{j=1}^{J} \left(Ro - wt^{T} I\varpi\right)^{2} + \phi \sum_{j=1}^{J} WT^{2}$

Publisher: SARC Publisher

Activate hexpo function,

$$\lambda_{hex}(Ro) = \tanh(\iota Ro) * \exp(\psi^{\circ} Ro)$$

Evaluate graph convolutional layer,

$$F\varphi'' = \lambda_{hex} \cdot \Diamond_{reg}\left(\partial^{mx^{-1/2}} M\alpha^{-1/2} F\varphi wt\right)$$

Determine ReLU activation

Perform pooling operation,

$$Pl + \xrightarrow{pooling} R_{unit}\left(F\varphi''\right)$$

Execute fully-connected layer,

$$Fc = \lambda_{hex} * (wt \cdot Pl) + bs$$

Determine output layer

End For

Return $P_{out} \rightarrow \{re, no\}$

End

Afterward, the intention behind the readmission is analyzed via the deep explainer.

Deep explainer

In this phase, the intention behind the readmission P_{out} based on the complex relationship and hidden pattern is explained based on the proposed LIM2DCE. The Local Interpretable Model-agnostic Explanation (LIME) offers valuable insights into how the model makes decisions. However, it assigns a weight to each instance using the kernel function, causing overfitting issues. Therefore, the 2D Cigar Formula (2DCF) is introduced to assign the weight value. Initially, the instances (Is) are selected from the P_{out}.

$$P_{out} \xrightarrow{select} Is \tag{31}$$

Then, the selected instances are increased to generate augmented instances (aug).

$$Is \xrightarrow{augment} aug \tag{32}$$

Further, the 2DCF (Cg) is used to assign the weight value (ξ_{wt}) for each instance in an augmented set.

$$Cg(\xi_{wt}) = aug_1^{\,2} + 100 aug_2^{\,2} \tag{33}$$

Thereafter, the weighted dataset (Wgt) is used to train the interpretable model (Ξ).

$$Wgt \xrightarrow{train} \Xi \tag{34}$$

$$\Xi \xrightarrow{Interpret} P_{out} \tag{35}$$

The local model significantly analyses the intention behind the patient's readmission status according to the coefficients of the local model. The pseudo-code of the proposed LIM2DCE is presented below,

Input: Classified outcome P_{out}

Output: Intention analysis

Begin

Initialize P_{out}, Is and Cg

For 1 to each instances do,

Select instances $P_{out} \xrightarrow{select} Is$

Augment selected instances,

$$Is \xrightarrow{augment} aug$$

Assign weight values using the 2DCF

$$Cg(\xi_{wt}) = aug_1^{\,2} + 100 aug_2^{\,2}$$

Publisher: SARC Publisher

Train local mode using the (Wgt)
Interpret complex model

$$\Xi \xrightarrow{Interpret} P_{out}$$

End For
Return Analyzed intentions
End

Overall, the proposed method improves healthcare data management via graph-based databases.

RESULT AND DISCUSSION

In this section, the proposed framework's performance in healthcare data management is demonstrated. The proposed framework's experimental analysis is implemented in the working platform of PYTHON.

Dataset Description

The proposed framework collects the patient healthcare historical data from the Diabetes 130-US Hospital for Years 1999-2008 dataset for accurate patient readmit prediction. The source link to access this dataset is depicted under the reference section. This dataset includes hospital records of patients diagnosed with diabetes who underwent laboratory tests, received medications, and had hospital stays of up to 14 days. Also, this dataset helps to predict early readmission within 30 days of discharge. From this dataset, this framework uses 80% data for training and 20% data for testing.

Performance Validation for the Proposed LIM2DCE

Here, the LIM2DCE's performance is validated by comparing it with several conventional methodologies like LIME, Shapely Additive exPlanations (SHAP), Partial Dependence Plot (PDP), and Gerchberg-Saxton (GS).

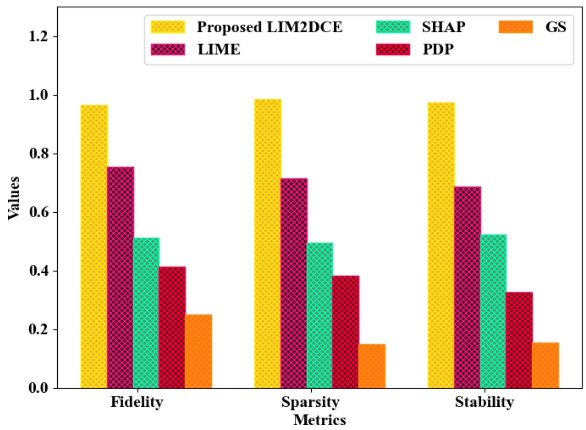

Figure 3: Efficiency analysis for the proposed LIM2DCE

Figure 3 depicts the LIM2DCE's performance in patient intention explanation about their readmission. As per the result, the LIM2DCE attains higher fidelity (0.9658), sparsity (0.9854), and stability (0.9745) than the other conventional methodologies. By using the 2D Cigar formula for weight assignment, the interpretability of the proposed framework is enhanced while reducing overfitting, thus ensuring more reliable patient readmission intent explanations.

Efficiency Analysis for the Proposed HGCRN

This section analyzes the proposed HGCRN's efficiency in predicting patient readmission by comparing it with several conventional methodologies like GCN, Gated Recurrent Unit (GRU), Bidirectional Long Short-Term Memory (BiLSTM), and Recurrent Neural Network (RNN).

Publisher: SARC Publisher

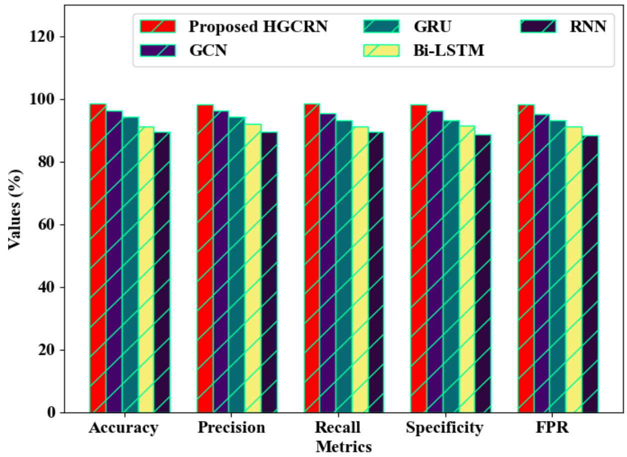

Figure 4: Performance validation for the proposed HGCRN

The proposed HGCRN's performance regarding the accuracy, precision, recall, specificity, and False Positive Rate (FPR) in patient readmit prediction is depicted in Figure 4. Here, the HGCRN uses ridge regularization and hexpo activation function for accurate classification, thus resulting in improved accuracy (98.45%), precision (98.12%), recall (98.56%), specificity (98.20%), and FPR (98.32%) by reducing computational overhead and low learning rate. Hence, the HGCRN generalizes better than the other prevailing approaches.

Effectiveness Evaluation for the Proposed CMC-KA

In this section, the CMC-KA's efficiency in privacy preservation for accurate readmit prediction is depicted by comparing it with several conventional methodologies like K-anonymity, L-Diversity, T-Closeness, and Randomization.

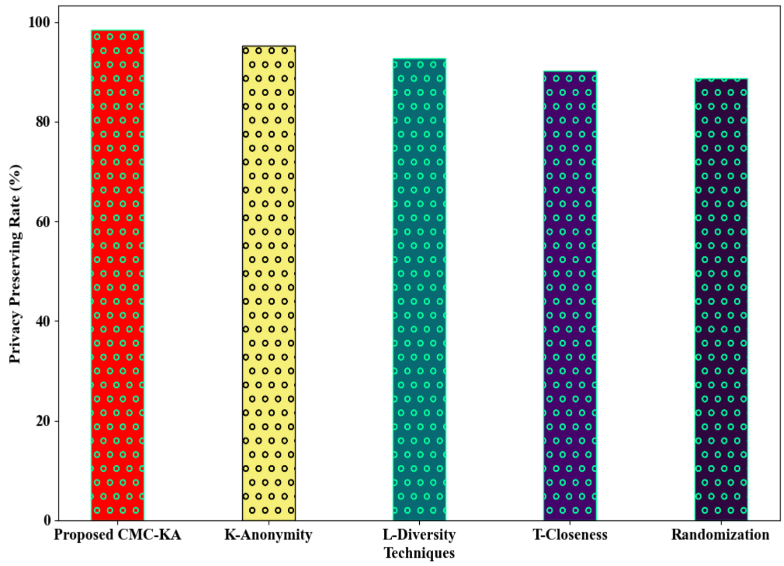

Figure 5: Privacy-preserving rate analysis

Table 1: Information loss evaluation

Techniques	Information Loss
Proposed CMC-KA	124
K-Anonymity	356
L-Diversity	758
T-Closeness	1023
Randomization	1457

Figure 5 and Table 1 illustrate the CMC-KA's performance on privacy preservation for efficient patient healthcare data management based on its privacy-preserving rate and information loss. As per the results, the CMC-KA attains a higher privacy preservation rate and lower information

Publisher: SARC Publisher

loss of 98.35% and 124, respectively. Also, the CMC-KA outperforms the other existing approaches while minimizing computational costs and information loss. This is because of the integration of the Cramér–von Mises criterion with K-Anonymity. Therefore, the CMC-KA improves privacy preservation by effectively handling complex datasets.

Performance Analysis for the Proposed LQLMIMA

By comparing the LQLMIMA's performance with several conventional techniques like LMA, Greedy Modularity Algorithm (GMA), Girvan-Newman Algorithm (GNA), and Stochastic Block Model (SBM), the proposed LQLMIMA is validated to highlight its efficiency in customer segmentation for accurate patient readmit prediction.

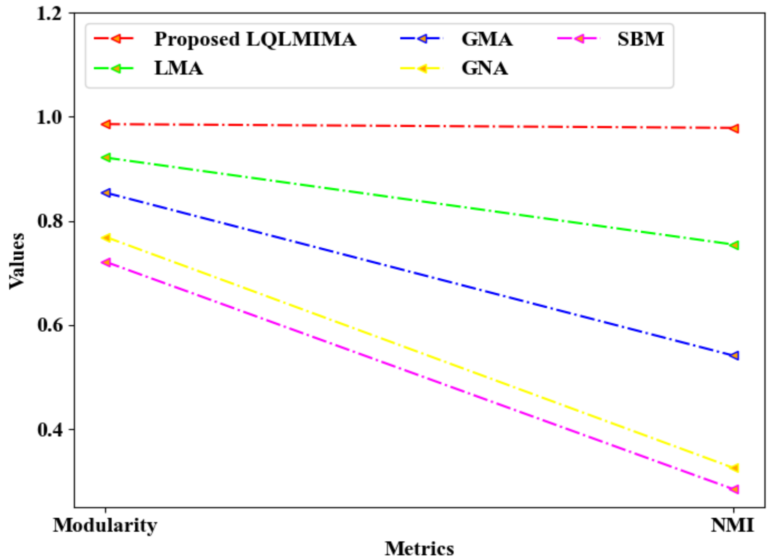

Figure 6: Modularity and NMI validation

The modularity and Normalized Mutual Information (NMI) of the LQLMIMA in customer segmentation is demonstrated in figure 6. Here, the LQLMIMA uses the quantum linear mutual information for addressing the randomness in community detection, thus ensuring more stable and consistent segmentation results. Thus, the proposed framework outperforms the other conventional methodologies with improved modularity (0.9856) and NMI (0.9784) by ensuring consistent community structures.

Table 2: Segmentation time analysis

Methodologies	Segmentation Time (ms)
Proposed LQLMIMA	17241
LMA	24518
GMA	28745
GNA	34158
SBM	38657

Table 2 demonstrates the time taken by the proposed LQLMIMA for customer segmentation. Here, the LQLMIMA takes a minimum time of 17241ms for customer segmentation. However, the traditional methodologies like LMA, GMA, GNA, and SBM took more times of 24518ms, 28745ms, 34158ms, and 38657ms for customer segmentation, respectively. As per the results, the LQLMIMA outperforms the other conventional methodologies by providing efficient and accurate segmentation, thereby attaining faster convergence.

Efficacy Validation for the Proposed CPHCA

This section elaborates the CPHCA's efficacy in dimensionality reduction for accurate patient readmit prediction by comparing it with several traditional approaches like PCA, Independent Component Analysis (ICA), Linear Discriminant Analysis (LDA), and Quadratic Discriminant Analysis (QDA).

Publisher: SARC Publisher

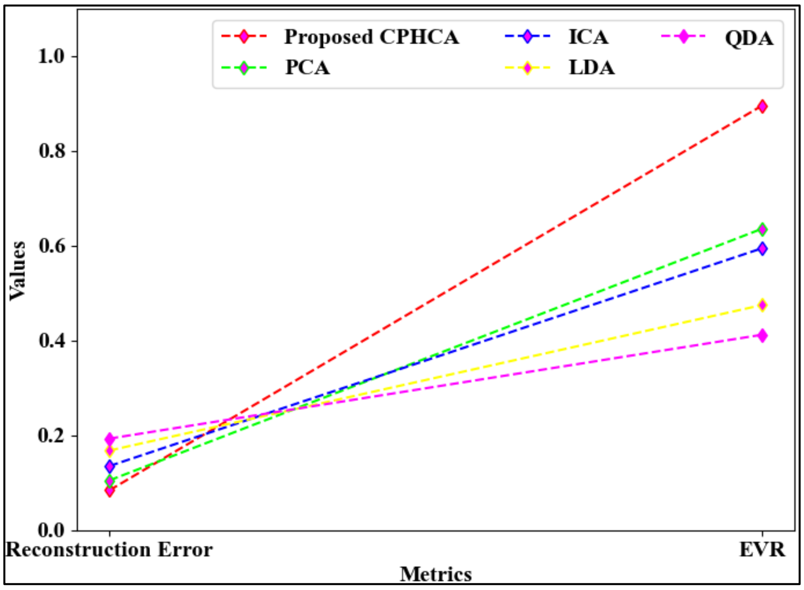

Figure 7: Reconstruction error and EVR validation

Figure 7 depicts the CPHCA's performance in dimensionality reduction regarding its reconstruction error and Explained Variance Ratio (EVR). As per the results, the CPHCA generalizes better in reducing the dimensionality than the other prevailing approaches by attaining lower reconstruction error and higher EVR of 0.0847 and 0.8954, respectively. This is because of the integration of the confluent hypergeometric technique. And, the dimensionality of the data is efficiently reduced by mitigating the reconstruction rate and increasing EVR while improving data distribution modeling.

Comprehensive Analysis

The comparative analysis of the proposed framework against various conventional approaches in efficient patient healthcare data management for accurate patient readmit prediction is estimated.

Table 3: Comparative analysis

Authors' Name	Objectives	Methodologies	Advantages	Limitations
Proposed Framework	Graph database-based patient healthcare data management	LIM2DCE	Helped in accurately predicting patient readmits by describing their intentions	Failed to find the disease outbreak threats
(Nauman *et al.*, 2025)	Big data analytics-based diabetes management revolution and healthcare decision making	Ensemble ML	Improved predictive performance with high accuracy	Owing to the reliance on a single dataset, this model failed to achieve more generalized results.
(Hovorushchenko *et al.*, 2023)	Blockchain-based medical data management	Blockchain-based transaction algorithm	Ensured systematic validation and supplementation of medical data	This model was not suitable for real-world applications.
(Azbeg *et al.*, 2022)	IoT, Blockchain, and InterPlanetary File System (IPFS)-based healthcare system for data management security	Proof of Authority consensus with Proxy Re-encryption	Reduced processing time	The framework introduced centralization concerns.
(Kala & Priya,	Smart IoT-	Shuffled Random	Enhanced security	Yet, this model

Publisher: SARC Publisher

2024)	Blockchain-based sensitive personal medical data security enhancement	Starvation Link Encryption (SRSLE)	and authentication and reduced time complexity for encrypted healthcare data sharing.	faced scalability and interoperability issues.
(Rathee & Iqbal, 2025)	Healthcare data management and decision-making based on blockchain.	Hybrid Ensemble Learning	Provided better efficiency and enhanced integrity and availability.	This model failed to learn patterns of heterogonous records from various sources.

Table 3 depicts the comprehensive analysis of the proposed framework against various convention works to highlight its efficiency in healthcare data management. As per the validation, the proposed framework outperforms the other existing works by achieving better results. Although conventional work offers high accuracy, low processing time, high data privacy, and integrity, they face several limitations like scalability issues, centralization concerns, interoperability challenges, etc., thus affecting their performance in healthcare data management.

CONCLUSION

In this paper, an efficient graph database-based patient healthcare data management framework was implemented using HGCRN and LIM2DCE for predicting patient readmission. Here, the patient's details were successfully registered using the CMC-KA with a 98.35% privacy preserving rate. Further, the patient readmit classification model was efficiently trained by collecting the patient's healthcare historical data. The EYEVA was employed to detect and eliminate the outliers from the data. Also, the data dimensionality was reduced within a minimum time of 27458ms using the CQHDA, which attained a 0.08457 reconstruction error. Further, by using the LQLMIMA, the customers were segmented accurately with higher modularity and NMI of 0.9856 and 0.9784, respectively. Moreover, the HGCRN provided the classified outcome, whether the patients were readmitted or not, with an accuracy of 98.45%. Finally, the intention of the patients was accurately described using the LIM2DCE by attaining 0.9858 fidelity. Thus, the proposed framework outperformed the other conventional works by efficiently managing the patient healthcare data based on the graph database.

Future Scope: Yet, the proposed framework failed to identify the disease outbreak trends. Hence, in the future, the work will focus on identifying the disease outbreak trends and suggesting personalized treatment plans by ensuring overall system performance.

REFERENCE

1. Strack, B., DeShazo, J. P., Gennings, C., Olmo, J. L., Ventura, S., Cios, K. J., & Clore, J. N. "Impact of HbA1c measurement on hospital readmission rates: analysis of 70,000 clinical database patient records." *BioMed research international* 2014.1 (2014): 781670. https://archive.ics.uci.edu/dataset/296/diabetes+130-us+hospitals+for+years+1999-2008
2. Abbas, A., Alroobaea, R., Krichen, M., Rubaiee, S., Vimal, S., & Almansour, F. M. "Blockchain-assisted secured data management framework for health information analysis based on Internet of Medical Things." *Personal and ubiquitous computing* 28.1 (2024): 59-72.
3. Abdulmalek, S., Nasir, A., Jabbar, W. A., Almuhaya, M. A., Bairagi, A. K., Khan, M. A. M., & Kee, S. H. "IoT-based healthcare-monitoring system towards improving quality of life: A review." *Healthcare*. Vol. 10. No. 10. MDPI, 2022. https://doi.org/10.3390/healthcare10101993
4. Adere, E. M. "Blockchain in healthcare and IoT: A systematic literature review." *Array* 14 (2022): 100139. https://doi.org/10.1016/j.array.2022.100139
5. Ahamed, B., Sellamuthu, S., Karri, P. N., Srinivas, I. V., Zabeeulla, A. M., & Kumar, M. A. "Design of an energy-efficient IOT device-assisted wearable sensor platform for healthcare data management." *Measurement: Sensors* 30 (2023): 100928.
6. Ajagbe, S. A., Misra, S., Afe, O. F., & Okesola, K. I. "Internet of Things (IoT) for secure and sustainable healthcare intelligence: Analysis and challenges." *International Conference on Applied Informatics*. Cham: Springer International Publishing, 2022. https://doi.org/10.1007/978-3-031-19647-8_4

Publisher: SARC Publisher

7. Aldwairi, M., Jarrah, M., Mahasneh, N., & Al-khateeb, B. "Graph-based data management system for efficient information storage, retrieval and processing." *Information Processing & Management* 60.2 (2023): 103165.
8. Almalawi, A., Khan, A. I., Alsolami, F., Abushark, Y. B., & Alfakeeh, A. S. "Managing security of healthcare data for a modern healthcare system." *Sensors* 23.7 (2023): 3612.
9. Arul, R., Al-Otaibi, Y. D., Alnumay, W. S., Tariq, U., Shoaib, U., & Piran, M. J. "Multi-modal secure healthcare data dissemination framework using blockchain in IoMT." *Personal and Ubiquitous Computing* (2021): 1-13. https://doi.org/10.1007/s00779-021-01527-2
10. Azbeg, K., Ouchetto, O., & Andaloussi, S. J. "BlockMedCare: A healthcare system based on IoT, Blockchain and IPFS for data management security." *Egyptian informatics journal* 23.2 (2022): 329-343. https://doi.org/10.1016/j.eij.2022.02.004
11. Calazans, M. A. A., Ferreira, F. A., Santos, F. A., Madeiro, F., & Lima, J. B. "Machine learning and graph signal processing applied to healthcare: A review." *Bioengineering* 11.7 (2024): 671. https://doi.org/10.3390/bioengineering11070671
12. Demirbaga, U., & Aujla, G. S. "MapChain: A blockchain-based verifiable healthcare service management in IoT-based big data ecosystem." *IEEE Transactions on Network and Service Management* 19.4 (2022): 3896-3907.
13. Gohar, A. N., Abdelmawgoud, S. A., & Farhan, M. S. "A patient-centric healthcare framework reference architecture for better semantic interoperability based on blockchain, cloud, and IoT." *IEEE access* 10 (2022): 92137-92157.
14. Gupta, K., Saxena, D., Rani, P., Kumar, J., Makkar, A., Singh, A. K., & Lee, C. N. "An intelligent quantum cyber-security framework for healthcare data management." *IEEE Transactions on Automation Science and Engineering* (2024).https://arxiv.org/pdf/2410.03217
15. Hovorushchenko, T., Moskalenko, A., & Osyadlyi, V. "Methods of medical data management based on blockchain technologies." *Journal of Reliable Intelligent Environments* 9.1 (2023): 5-16. https://doi.org/10.1007/s40860-022-00178-1
16. Islam, M. S., Ameedeen, M. A. B., Rahman, M. A., Ajra, H., & Ismail, Z. B. "Healthcare-chain: blockchain-enabled decentralized trustworthy system in healthcare management industry 4.0 with cyber safeguard." *Computers* 12.2 (2023): 46. https://doi.org/10.3390/computers12020046
17. Kala, M. K., & Priya, M. "Smart IoT-Blockchain Security to secure Sensitive Personal Medical Data using Shuffled Random Starvation Link Encryption." *IEEE Access* (2024). https://ieeexplore.ieee.org/iel8/6287639/6514899/10721466.pdf
18. Khayat, M., Barka, E., Serhani, M. A., Sallabi, F., Shuaib, K., & Khater, H. M. "Empowering Security Operation Center with Artificial Intelligence and Machine Learning–A Systematic Literature Review." *IEEE Access* (2025).
19. Lodha, L., Baghela, V. S., Bhuvana, J., & Bhatt, R. "A blockchain-based secured system using the Internet of Medical Things (IOMT) network for e-healthcare monitoring." *Measurement: sensors* 30 (2023): 100904.
20. Nauman, M., Almadhor, A. S., Albekairi, M., Ansari, A. R., Fayyaz, M. A., & Nawaz, R. "The Role of Big Data Analytics in Revolutionizing Diabetes Management and Healthcare Decision-Making." *IEEE Access* (2025).https://ieeexplore.ieee.org/iel8/6287639/6514899/10829563.pdf
21. Rahman, M. S., Islam, M. A., Uddin, M. A., & Stea, G. "A survey of blockchain-based IoT eHealthcare: Applications, research issues, and challenges." *Internet of Things* 19 (2022): 100551.
22. taria, A., & Khang, A. "IoT equipped intelligent distributed framework for smart healthcare systems." *Towards the integration of IoT, cloud and big data: services, applications and standards*. Singapore: Springer Nature Singapore, 2023. 97-114. https://arxiv.org/pdf/2110.04997
23. Rathee, G., & Iqbal, R. "Enhancing Decision-Making and Data Management in Healthcare: A Hybrid Ensemble Learning and Blockchain Approach." *Technologies* 13.2 (2025): 43.
24. Shah, D., Rani, S., Shoukat, K., Kalsoom, H., Shoukat, M. U., Almujibah, H., & Liao, S. "Blockchain factors in the design of smart-

Publisher: SARC Publisher

media for e-healthcare management." *Sensors* 24.21 (2024): 6835.

25. Sharma, D. K., Chakravarthi, D. S., Shaikh, A. A., Ahmed, A. A. A., Jaiswal, S., & Naved, M. "The aspect of vast data management problem in healthcare sector and implementation of cloud computing technique." *Materials Today: Proceedings* 80 (2023): 3805-3810.
26. Singh, S., Pankaj, B., Nagarajan, K., Singh, N. P., & Bala, V. "Blockchain with cloud for handling healthcare data: A privacy-friendly platform." *Materials Today: Proceedings* 62 (2022): 5021-5026. https://doi.org/10.1016/j.matpr.2022.04.910
27. Taloba, A. I., Elhadad, A., Rayan, A., Abd El-Aziz, R. M., Salem, M., Alzahrani, A. A., ... & Park, C. "A blockchain-based hybrid platform for multimedia data processing in IoT-Healthcare." *Alexandria Engineering Journal* 65 (2023): 263-274.
28. Thantilage, R. D., Le-Khac, N. A., & Kechadi, M. T. "Healthcare data security and privacy in Data Warehouse architectures." *Informatics in Medicine Unlocked* 39 (2023): 101270.
29. Thapa, C., & Camtepe, S. "Precision health data: Requirements, challenges and existing techniques for data security and privacy." *Computers in biology and medicine* 129 (2021): 104130.
30. Yaqoob, I., Salah, K., Jayaraman, R., & Al-Hammadi, Y. "Blockchain for healthcare data management: opportunities, challenges, and future recommendations." *Neural Computing and Applications* (2022): 1-16.
31. Zala, K., Thakkar, H. K., Jadeja, R., Singh, P., Kotecha, K., & Shukla, M. "PRMS: design and development of patients' E-healthcare records management system for privacy preservation in third party cloud platforms." *IEEE Access* 10 (2022): 85777-85791.

Source of support: Nil; **Conflict of interest:** Nil.

Cite this article as:

Surasani, V.R., Mukkala, S.R., Pothireddy, V.K and Sanikommu, G." An Effective Graph Database-Centric Patient Healthcare Data Management Using A Robust Hgcrn And LIM2DCE" *Sarcouncil Journal of Engineering and Computer Sciences* 3.9 (2024): pp 1-17.

Publisher: SARC Publisher