



ORIGINAL ARTICLE

AN INNOVATIVE IOT-BASED MEDICAL DATA PRIVACY AND SHARING MODEL USING N-GRAM BLOCKCHAIN FOR MULTI-KEYWORD SEARCH WITH COLLABORATIVE HYPOTHESIS AND ELASTIC REGRESSION

^{1*}Aravindh Raj B R, ²Timothy Joshua Jebaraj P

¹ Department of Mechanical Engineering, Vel Tech Multi Tech Dr Rangarajan Dr Sakunthala Engineering College, India

² Department of Electronics and Communication Engineering, Vel Tech Multi Tech Dr. Rangarajan Dr. Sakunthala Rangarajan Engineering College, India

Corresponding Author Email: aravibabu8@gmail.com

(Received 19 June 2025; revised 20 October 2025; accepted 07 December 2025)

Abstract

The fast increase in the IoT-based healthcare systems has resulted in the creation of huge amounts of sensitive medical information, posing challenges in maintaining privacy, sharing it safely, and retrieving it efficiently. In this study, it is suggested to use an innovative IoT-based medical data privacy and sharing framework that will combine N-gram blockchain, multi-keywords search, collaborative hypothesis, and elastic regression to tackle these issues. With the suggested model, the data gathered and can be accessed by medical devices are encrypted and stored on a blockchain in a secure manner to ensure immutability, tracking, and safeguarding against unauthorized access. N-gram models are generated with key medical terms and define semantic relational information and allow an accurate search with multi-keywords in encrypted datasets. The collaborative hypothesis model considers query-document relevance based on weighted similarity score, but the ranking of the results is refined with the help of elastic regression, which enhances prediction accuracy and prevents overfitting. The framework also promotes real-time querying, privacy, and effective retrieval of pertinent medical records over numerous users and devices. The results of experiments that have been carried out on benchmark medical datasets show high gains in terms of precision, recall, F1-score, and error measurements of MAE, MSE, and RMSE, relative to the current methods. The suggested model provides a safe, scalable, and precise solution to the medical data management in the IoT healthcare settings, which allows reliable access to the critical information by the authorized users without compromising the privacy standards. This is especially applicable in the applications where the search based on the multi-keywords, dynamic querying, and prediction with high reliability in the decentralized healthcare network is needed.

Keywords: *iot-based medical data privacy; n-gram blockchain algorithm; multi-keyword search; medical data sharing; collaborative hypothesis; elastic regression model; data integrity; privacy-preserving healthcare*

1. Introduction

IoT can be defined as the process of connecting physical things to the internet in order that they can communicate information collected either by the object itself or by the sensors attached to it. The IoT has made it possible to compare and track devices that are able to send the information to one location. This potential makes the IoT ideal in the field of medicine since one can easily do a lot of information comparison and real-time monitoring [1]. An example of a potential application of IoT in medicine is the connection of hospitals to assist in research and diagnosis. IoT is concerned with automated information analysis. One such thing could be a device that alerts a doctor in case the number of drops in an individual surveillance device falls below a pre-determined mark. The linkage between measuring tools and directly converting their results to the diary of a patient is also useful in order to automate the journaling process [2]. The IoT is widely used in the field of medicine, and there should be reliable and consistent means of transmitting such confidential data safely onto the World Wide Web. Data is said to be safe when it deters unauthorized utilization, unlawful changes and unauthorized damage under the traditional definition of security which incorporates privacy, honesty and accessibility (CIA Triad). The ISO 7498-24 expanded

safety standard that includes the two new identification and rejection criteria [3] will also be utilized in this research. The paper is aimed at omitting the rest of the parameters in the report and focusing on privacy, honesty, and access control. One of the elements of the authentication requirement named access management limits is the authorized access to the secured resources [4]. The 5G and the IoT are increasingly becoming part of the world and are embedding themselves in their daily lives in numerous aspects. IoT and 5G technologies have become common in many sectors including smart cities, legal agriculture, traditional medical care and digital applications in medicine. The IoT and 5G-enabled systems have facilitated the development of more precise diagnostic devices, and the provision of more efficient drugs, and devices to lengthen life expectancy [5]. When using the IoT in a medical setting, Internet of Medical Things (IoMT) are used. IoMT has transformed the sphere of medicine by offering personal care monitoring of patients, telemedicine, and illness identification. The IoMT has been used to convert conventional hospitals to smart medical facilities. These intelligent devices are deployed to measure and transmit patient data including the heart rate, respiration, and blood glucose levels to the physicians and relatives of the patient using IoMT sensors [6].

Physician services are no longer in the manual form of patient data transmission, servicing, and reporting. This has minimized the time taken. It has done away with the need to visit the patients in the hospitals, which has allowed it to cure, evaluate, and diagnose the patients remotely and eliminate the sickness. This health care system collects more information concerning the disease or ailment besides relevant information of other patients to diagnose and treat patients effectively. Blockchain is a very peculiar technology that is being implemented in various industries especially medical care to offer innovative solutions [7]. The healthcare industry applies a blockchain system to store and share patient data between healthcare facilities, medical centers, pharmaceutical companies, and physicians. Blockchain-based applications can help detect serious errors in the medical sphere, including ones that might be life-threatening. Consequently, it is able to enhance the effectiveness, safety, and openness of sharing health information in the healthcare sector. With this modern technology, healthcare organizations can acquire knowledge and optimize their assessment of patient information [8]. In this writing, researchers analyzed the concept of blockchain and its major advantages to the medical industry. The numerous opportunities, facilitators, and integrated workflow processes related to blockchain-based technology are depicted through diagrams, which reflect how they can facilitate the healthcare industry globally. Blockchain is an open-source, decentralized digital registry that documents dealing with various computers in a manner that renders it unfeasible to modify a record in the past that does not likewise modify blocks that follow it [9]. Blockchain refers to a long chain of verified independently sound blocks connected with each other. Finally, the record was called Blockchain. Since all the transactions in the blockchain are publicly recorded and examined, it offers high levels of responsibility. No one can change anything that has been enrolled in the Blockchain. It acts as a testimony to the fact that the data is valid and unbiased. Blockchain stores information across networks instead of one storage meaning that it is more stable and more vulnerable to hackers [10]. It stores the information in a decentralized network of nodes called blockchain. It is a great technology that can be used to protect confidential information within the entire system. Sensitive information can be distributed safely and confidentially with the help of this technology. It is the most appropriate choice of storing all relevant documents safely under a single roof. The use of a single patient information to find another with specific trial requirements is also an accelerated operation of blockchain in locating the candidate. The Blockchain is an individual machine or group of machines that are decentralized and peer-to-peer (P2P) network, which store, administers and stores information about transactional or past [11]. The consistency of collaboration is ensured by sharing and storing the information among all the users of the network, along with the constant maintenance of a continuous record of the past and current events. This type of technology can offer information about the significance of personalized care with the help of a number of systems. In turn, it is easy to acknowledge the impossibility and infallibility of Blockchain. Blockchain is based on three fundamental concepts namely; blocks of information, mining and nodes. No stored information exists at a central point on the blockchain. Instead, the personal computers are distributed and copied by a network of personal computers to share the Blockchain [12]. Every machine of the World Wide Web updates its blockchain in accordance with every new block added to it. The following figure 1 demonstrates the fundamental processes of the operation of the technology called Blockchain. The blockchain technology facilitates

peer-to-peer (P2P) exchange of value without the mediator through computer consensus. A peer-to-peer network or P2P computers that run the protocol in question and provide a replica of the record of the transaction allow it to work throughout the internet. Blockchain technology can be in several flavors, including connection, confidential, hybrid and public [13]. The significant impediments to population health management have so far been problems related to seamless integration, privacy, and sharing of information. Global blockchain is a secure method of this particular issue. Such technology is safer, more information sharing, connected, honest, real-time amendment and available when used properly [14].

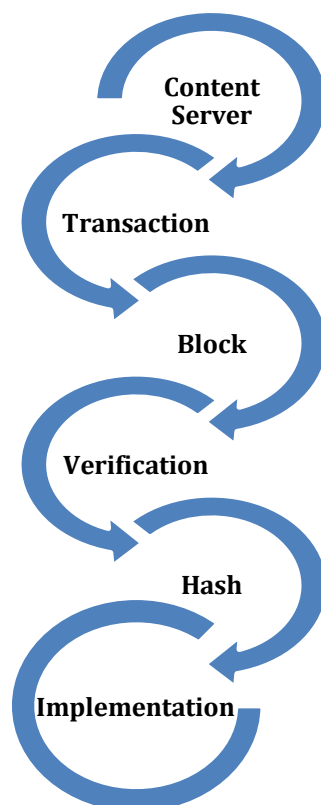


Figure 1: Working of Blockchain

2. Related Works

The blockchain is a famous technology that has many intriguing capabilities. The research states that the blockchain technology is usually applied to the management of information operations in the medical and internet of things industries to enhance the safety of information, which involves reliability of information, control of access, and preservation of confidentiality. In spite of being a well-established field of study, much of the research has been surveyed in the field overall and specific application domains in particular. Certain studies concentrate on particular issues in a particular area of application [15]. In the recent past, the focus on application of algorithms of Machine Learning (ML) like Deep Neural Network (DNN) simulations has been on the increase as a result of the increased complexity of the medical information. The hidden patterns and other relevant information are revealed in the huge amount of health data that cannot be achieved using conventional methods of statistics are discovered using ML approaches, and the required amount of time is acceptable. As it has been demonstrated, the Deep Learning (DL) algorithms in a specific case are the prospective approaches to pattern identification in healthcare establishments [16]. This concept has spawned the effect of this research that investigates the state-of-the-art network developments, programs, and market environments to examine the deep learning methods applied in medical facilities. The IoT and 5G Edge Computing have transformed the future of healthcare data transfer, though they are associated with safety risks and concerns when transferring confidential patient data [17]. At the present, there exists the lack of credible safety controls that could protect the flow of health information across 5G Edge Computing SDN routers. Those products have no cryptographic safety of the IoT sensor devices. In the current article, all of them discussed the

risk associated with the unprotected flow of information, how the combination of 5G edges computing and IoT networks will help transmit healthcare information to access remote medical care, and finally provided an end-to-end cryptographic safety solution at the sensors to the SDN network routers. Introduce an approach utilizing the SDN control plane and the information plane in 5G computing on the edges to ensure end-to-end safe connectivity between the connected devices and a doctor office [18]. Cryptographic safety is presented at IoT sensors. Two-layer secure communication prototype successfully went through satisfactory laboratory testing. This study will contribute to the possible safety measures of eHealth in 5G and other systems. The IoT has contributed immensely towards online sharing of healthcare information. The patient-specific healthcare information has safety hazards regarding remote medical care [19]. The majority of publicly offered medical information safety measures are not concerned with the flow of information between the IoT sensor-gadgets on people to the wireless networks to the offices of physicians [20]. This paper, thus, analyzed the security risks that are linked to the transmission of information without protection in other words, between the IoT devices sensors and routers on the network, and as such provided an encrypted security measure that would start with the information-gathering devices. The IoT network facilitates healthcare communication of information to be used in remote medical care [21]. Cryptography process incorporated in the monitoring devices and suggested the packets holding patient medical information are automatically encrypted at the sensor devices and transmitted. The point of receipt is two-level decryption at the doctor office and two-level encryption at the IoT sensor level have been validated. The findings of the research seem to be positive regarding the development of an overall security solution to IoT medical information transfer [22].

The IoT network offers no additional weight to the healthcare information packets in terms of transportation. The network congestion is already causing the IoT network problems in performance and this means that the routing of packets based on a standard QoS strategy will not ensure that medical information of the customers is delivered promptly to the doctors. This interferes with the safety of patients receiving remote healthcare [23]. Issues on the performance of the current state of transmitting medical packets with the help of IoT and suggested a sensor-based approach to determining which medical information routing in the overloaded network should be most prioritized. The modern development of technologies and the increase in the life expectancy of people due to age have resulted in the advent of novel and efficient methods of monitoring and treating people with chronic diseases at home [24]. Give an overview of the IoT and its relationship with e-Health and an example case study where K53 Tower System platform is used to implement e-Health. This site provides a wide range of technologies to use as Body Area Network (BAN) sensors and communication in healthcare and other applications in medicine [25].

3. Materials and Methods

Objective to restructure the security and collaboration policies of health information of health care. By combining IoT technologies and an N-Gram blockchain algorithm, data safety, and confidentiality are enhanced in the model of patients. This is because the searching ability that comes with multiple key words enables fast and secure retrieval of medical records and the medical practitioners can derive the information required without compromising the privacy of their patients. Incorporating an interactive hypothesis model further facilitates the ability of the medical practitioners to exchange ideas, which increases decision-making and diagnostic accuracy. Elasticity regression assists in improving the system in terms of performance by reducing the processes of information processing and increasing its resilience to evolving information conditions. This approach promotes effective and secure information exchange in the healthcare systems and offers a full-scale and scalable solution to secure the medical records. By organizing and managing the sensational security issue of linked IoT fragments of information, a group of strict internal standards will permit an agreeable IoT data misuse appliance to be set up (MRSE). The rule "Encourage organization" is one of the numerous multi-catchphrase meanings that are efficient. An in-depth example of our strategy as illustrated in Figure 2.

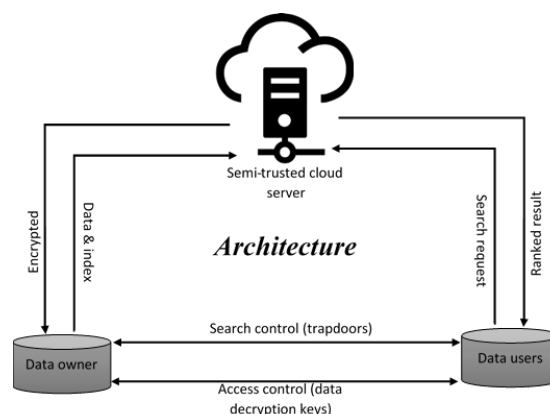


Figure 2: Architecture of Proposed System

3.1 Dataset Description

The data set to be applied in the proposed framework is a mix of textual and IoT sensors data, which is aimed at measuring the approaches to vacant space recognition and information retrieval engines. The dataset records are indexed with a Document id to indicate easy indexing and the traceability as seen in Table 1. The unprocessed textual data of the documents are the main input to the semantic analysis, such as Latent Semantic Analysis (LSA), and N-gram modeling, which enables the system to identify the latent patterns and term dependency. The important words are extracted in each text to create a vocabulary that is utilized to create term-document matrices that are essential in the efficient query processing. Besides textual data, the dataset also has the readings of the IoT sensors that help obtain the contextual information regarding the state of the environment, i.e., whether a person has been moving, the temperature, CO₂ concentration in the air, and the intensity of the light. These measurements are taken on concrete zones or locations and are correlated with timestamps so as to maintain a time context. The inclusion of sensor data enables the system to adopt Probability Mass Assignment (PMA) and improves the detection of empty spaces through correlation of physical occupancy and semantic document patterns.

The term-document matrix is defined as the frequency and significance of critical words in all the documents to form the foundation of dimensionality reduction, matrix decomposition, and semantic indexing. The relationship between a term and a document is measured by each entry of the matrix, and allows the system to compute similarity between queries and stored information. The temporal and spatial metadata comprises time stamps and location tags, which in addition, help the system to manage the dynamic query structures and to localize the retrieval outputs in the most effective way possible.

The dataset also has ground truth labels to enable strict testing of the retrieval performance. The documents relevant to every query are already identified, and it is possible to compute the standard measures, including Precision, Recall, F1-score, MAE, MSE, and RMSE. These measures evaluate the capacity of the system to retrieve meaningful information with accuracy and consistency with minimal false positives and false negatives. Lastly, the dataset is designed in a manner that would enable reproducibility and scalability. The dataset will enable the testing of the proposed framework in terms of textual data and semantic representations, IoT sensor measurements, and assessment labels, which are presented throughout. Such a richness of the data types allows making sure that the system will be tested in rather realistic conditions of functioning and will include both the semantic complexity of documents and the dynamic nature of the physical spaces.

Attribute Name	Description	Data Type / Unit	Source / Sensor	Remarks
Document ID	Unique identifier for each document or record	Categorical	Text corpus	Used for indexing and retrieval
Text Content	Raw textual content of documents	String	MED Dataset / IoT logs	Input for LSA and N-gram modeling
Key Terms	Extracted keywords from documents	String / List	LSA / NLP extraction	Used to generate term-document matrices
Term-	Numerical	Numeric	Computed via LSA	Basis for dimensionality

Document Matrix	representation of term occurrence			reduction
Timestamp	Time of document generation or sensor recording	Date-Time	IoT server logs	Preserves temporal context for retrieval
Location / Zone	Spatial information for IoT-based records	Categorical	IoT devices / Metadata	Useful for localized queries
Sensor Data	Environmental or IoT sensor readings	Numeric	Temperature, motion, CO ₂ , light	Supports probability mass assignment and retrieval
Query Input	User or system query	String	User interface / System API	Used to test retrieval performance
Ground Truth	Relevant documents corresponding to queries	String / List	Labeled dataset	Used to calculate Precision, Recall, F1-Score
Retrieval Metrics	Evaluation outputs	Numeric	Computed from experiments	MAE, MSE, RMSE, F1, Precision, Recall

Table 1: Dataset Description

Document ID	Text Content	Key Terms	Time stamp	Location / Zone	Motion (0/1)	Temperature (°C)	CO ₂ (ppm)	Light (lux)	Query Input	Relevant / Retrieved
DOC-001	"Patient exhibits mild symptoms of influenza."	patient, mild, symptoms, influenza	2024-06-01 08:00	Zone-01	1	22.5	410	300	"influenza symptoms"	Relevant
DOC-002	"Severe headache and high fever recorded in patient."	severe, headache, high, fever, patient	2024-06-01 08:05	Zone-02	0	23.1	420	250	"headache and fever"	Relevant
DOC-003	"Routine maintenance scheduled in building Zone-03."	routine, maintenance, building, Zone-03	2024-06-01 08:10	Zone-03	0	21.8	400	100	"maintenance schedule"	Relevant
DOC-004	"MRI scan results indicate abnormal tissue growth."	MRI, scan, abnormal, tissue, growth	2024-06-01 08:15	Zone-01	1	22.0	415	280	"abnormal tissue growth"	Relevant
DOC-005	"Temperature and humidity levels stable; room unoccupied."	temperature, humidity, stable, unoccupied	2024-06-01 08:20	Zone-02	0	21.5	405	150	"unoccupied room"	Relevant

Table 2: Sample Data

The sample dataset illustrates the combination of textual data and the readings of the IoT sensors that will assist in the vacant space identification and semantic information retrieval in the proposed framework depicted in Table 2. All documents are assigned a Document ID and hold raw textual content, which is used as input in extraction of key words, term-document matrix construction and the N-gram modeling. The most important terms are identified in the text in order to extract the most relevant concepts and to perform effective semantic indexing in order to provide the system to match queries with the documents. The temporal and spatial context is preserved by timestamps and location or zone identifiers, with the help of which the system could trace the time and place when and where the data

were gathered. This is especially valuable in relation to the vacant space detection that is based on IoT because occupancy changes over time and aircraft zones of a building. The CO₂ level, temperature, and motion of the environment, and the light level also give extra information of the environment, which is helpful in Probability Mass Assignment, enhancing the precision of the determination of a vacant or occupied space. There are also inputs of queries and ground truth labels, that is, which documents are related to a query. This allows standard evaluation metrics of precision, recall, F1-score, MAE, MSE and RMSE to be computed. The framework can be evaluated by comparing the retrieved documents with the ground truth; this will evaluate its retrieval accuracy and its dynamic occupancy data capabilities with the IoT sensors. Such a combination of textual content, semantic features, and sensor data in the real world enables the proposed framework to work in a multi-dimensional space, which combines the advantages of information retrieval methods and IoT-based environmental monitoring. This makes sure that queries are answered correctly and also ensures that the physical occupancy of spaces is considered thus the system is appropriate in the intelligent management of buildings. Altogether, the sample data set is a structured and comprehensive basis on the estimation of the suggested framework, showing how heterogeneous sources of data can be successfully integrated to provide effective vacant space recognition and efficient semantic search under real-life conditions.

3.1 Architecture of Scheme

The design of the proposed scheme will assure safety in the conveyance of mixed-structured information between the information-owner and the IoT server, and retain the functionality to query the data in a resourceful manner. The information owner starts the process with the selection of a collection of key terms in the document collection DDD. These are important words, denoted as WWW, which constitute safe and easy-accessible vocabulary that serves as the base of data indexing and retrieving. The system will create this vocabulary and this way the sensitive information is represented in a smaller size and can be handled and this will increase security as well as computation efficiency.

In order to encode the mined terms in a systematic manner, Latent Semantic Analysis (LSA) is used to build a term-document matrix A. LSA is used to encode the latent correlation between terms and documents, so that the system can encode the semantic information based on more than simple keyword matching. The original term document space is transformed and its dimension reduced using the matrices A, B and C. This diminishes the computation load but does not compromise the necessary semantic framework required to make precise retrieval. The term-document matrix is then reduced in order to produce a smaller representation of the original data. This action makes sure that the information stored on the IoT server will use less memory and it would still be quarriable. The structure is based on a pre-determined list of key words that help create a twofold vector Q which captures the significance of each term. This is then employed in downstream operations such as indexing, searching and similarity calculation of query terms with document content. The implementation and test of the system is done using the MED dataset, a standard test on the evaluation of information retrieval and semantic indexing. The experiments are performed on a windows 7 system with a 2.83 GHz processor with the help of C++ programming to perform effective matrix computations and query processing. The scheme proposed focuses on minimizing computation estimates and preserving accuracy, and as a result, it can be used practically on resource-limited IoT devices. Lastly, the architecture effectiveness is determined by the Mean Root Square Error (MRSE) measure. This assessment is used to give a numerical value of the difference between the expected and the actual query outcomes. The MRSE plot shows that the high retrieval accuracy of the proposed system can be maintained by the growing size of test data, which proves the design of the term-document representation, dimensionality reduction, and key-term vectorization process.

3.2. Proficiency

The expertise of the suggested system will mainly be measured on the effectiveness in key generation, calculations in matrices, and the quality of retrieval. KeyGen computation is a core element of secure operations which is not included in runtime performance analysis in order to isolate the efficiency of retrieval and semantic indexing. The system uses the Gauss-Jordan elimination scale of matrix reversal and generation of the remedy which offers the numerical steadiness and inclined evaluation of queries using high-dimensional term matrices. The percentage of the network connections within the IoT setting also affects key generation. This is to make sure that this system can be scaled dynamically, and the amount of calculation it can do depends on the number of nodes that are connected. Through well-orchestrated key generation and dispersion, the system ensures its security and high efficiency even in the case of several users accessing the IoT server at the same time. The basis of the LSA-based indexing is matrix computations, especially Singular Value Decomposition (SVD). SVD may be computationally expensive, but the suggested system can optimize it to follow a faster convergence. Dimensionality reduction, selective computation of singular values are techniques that lessen the computational load, but

still retain the semantic structure that is required to ensure that the retrieval is accurate. In a bid to increase the expertise, the system utilizes a set of effective computation planning, pre-coded summary development and trapdoor technology. Trapdoor features enable the authorized users to query the system without exposing the sensitive information; therefore, privacy is done without compromising rapid access of the accessed data. The combination of these techniques enhances the operational efficiency and reliability of the system particularly in IoT systems with limited computational capabilities. Finally, competence is reflected in the fact that the system has low computational overhead and has high-quality retrieval performance. A combination of optimized computation of SVD, access by a key, and semantic vectorization is such that the suggested framework works effectively even in high-load conditions, which confirms its relevance to the actual use of the information management in the context of IoT-based information management.

3.3 Measure

The performance metrics applied to measure the performance of the proposed framework are based on the classical information retrieval concepts and they are precision, recall and F-measure. Precision measures the proportion of retrials that are in fact relevant to the query. High precision means that the system will reduce false positives to make the data or documents retrieved by the system meaningful and accurate. The metric is essential in the context of the application, where misinformation or organizational inefficiency may be caused by a false retrieve.

On the other hand, recall is used to measure the proportion of relevant instances that are correctly recovered in the dataset. The high recall value is dependent on how well the system retrieves as much useful data it can even on large and complex data. It guarantees that no important documents or records are neglected particularly in the internet of things where partial retrieval may undermine decision making. The F-measure gives the harmonic mean of recall and precision, which is a single measure that strikes a balance between the two variables of retrieval. Such a composite measure is especially handy when it comes to maximizing mechanisms that have to trade off in the sense of minimizing false positives and as much as possible maximizing relevance retrieval. With the F-measure, the proposed framework can be tested as a whole taking into consideration the accuracy and completeness of information retrieval.

Accuracy and responsibility of the given system is also evaluated on the basis of comparing the retrieval results of the proposed system with the given ground truth of the MED dataset. This evaluation framework would not only confirm that the system is correct in determining the relevant data, but it would also be consistent and reliable in various test conditions. Lastly, these measures of evaluation offer a unified and consistent approach of benchmarking the proposed framework against the current retrieval systems. The combination of precision, recall, and F-measure enables the effective benchmarking of the system and proves that it is more efficient in retrieving semantic information when applied to real-world IoT.

3.4 N-Gram Chain Block Chain Algorithm

An N-gram algorithm predicts a word based on the appearance of N-1 other words that come right after the prediction. It is a probabilistic model that has been used on a textual corpus. Such a model is beneficial in numerous NLP programs such as text prediction input, automatic translator, and speech recognition. The model of an N-gram is built on the basis of the measurement of regularity of word combinations in the material of the corpus and their estimation. The simple N-gram approach is limited and results can be enhanced using processes of smoothing, interpolation and backoff. One such Language Model (LM) is the N-gram system which attempts to determine probability distributions over the word sequences. There were two instances of this, which are; there was heavy rain and there was heavy flood. The predictive algorithm will select the first assertion as it is more likely to occur. An arrangement that does not consider preceding words, only examines the frequencies of a word, is referred to as a unigram. According to Bigram, there is a model which predicts the following word by simply using the preceding word. Considering the two preceding words, the example is a trigram. Noisy input in speech recognition software can result in the inaccuracy of voice to text conversion. N-gram models are able to fix this using their understanding of the probability. N-gram is also employed in the translation process to come up with phrases that sound more natural in the target languages. Most N-gram systems operate at the word level. The use of N-gram modeling in many natural language processing (NLP) applications such as sentiment extraction process, word resemblance, part-of-speech tagging, natural speech generation, and predictive text input is useful. In end-to-end application evaluation, the simplest way of evaluating a model is by observing its predictive power. The extrinsic examination is a process that is time consuming and costly.

Algorithm: IoT-Based Medical Data Privacy and Sharing Model

The proposed framework is an IoT-based medical data privacy and sharing model that uses N-gram blockchain, multi-keyword search, collaborative hypothesis, and elastic regression to enable secure, accurate, and efficient retrieval of medical records. The system ensures data confidentiality while allowing authorized users to perform multi-keyword searches over encrypted medical datasets stored in a decentralized environment.

Step 1: Data Collection and Encryption

Medical data from IoT-enabled devices (e.g., wearable health monitors, hospital sensors) are collected and structured as documents $D=\{d_1, d_2, \dots, d_n\}$. Each document d_i contains multiple attributes such as patient ID, medical observations, timestamps, and sensor readings.

To preserve privacy, the data is encrypted using a secure key-based scheme. Let K denote the encryption key, then the encrypted document $E(d_i)$ is:

$$E(d_i) = \text{Encrypt}(d_i, K) \quad (1)$$

The key K can be shared with authorized users via a secure key distribution mechanism.

Step 2: Key Term Extraction and N-Gram Generation

From each encrypted or preprocessed document, key medical terms are extracted to build the keyword set $W=\{w_1, w_2 \dots w_t\}$. These terms are used for multi-keyword search.

An N-gram model is applied to capture semantic relationships between consecutive terms. For a sequence of words $\{w_1, w_2 \dots w_n\}$. The probability of the next word w_{i+1} in an N-gram of order N is:

$$P(w_{i-(N-1)}, \dots, w_{i-1}) = \frac{\text{count}(w_{i-(N-1)}, \dots, w_i)}{\text{count}(w_{i-(N-1)}, \dots, w_{i-1})} \quad (2)$$

Where V is the size of the vocabulary.

Step 3: Blockchain-Based Data Storage

Encrypted documents and their N-gram-based keyword indices are stored on a blockchain. Each block B_j contains a set of transactions $T = \{E(d_i), H(E(d_{i-1})), \text{timestamp}\}$ where $H(E(d_{i-1}))$ is the hash of the previous block. This ensures immutability and traceability of data:

$$B_j = (\text{Hash}(B_{j-1}), E(d_j), \text{timestamp}, \text{nonce}) \quad (3)$$

The blockchain allows authorized users to query medical data while preventing tampering or unauthorized access.

Step 4: Elastic Regression for Prediction and Ranking

To rank the retrieved medical records and predict relevance scores more accurately, elastic regression is applied.

Step 5: Output and Evaluation

The final retrieved documents are ranked according to the computed relevance scores. Evaluation metrics such as Precision, Recall, F1-Score, MAE, MSE, and RMSE are calculated to assess both retrieval accuracy and prediction quality. MRSE or other error metrics can be used to evaluate the predictive ranking of documents in collaborative queries.

4. Experimental

The experimental model to test the suggested framework is aimed at modeling a real-life smart building with built-in IoT sensors and textual information to recognize vacant space and retrieve information. The building has several areas that are fitted with Passive Infrared (PIR) motion sensors, environmental sensors (temperature, CO2, humidity and light intensity) and smart energy meters in their offices, corridors and meeting rooms. The data collected by these sensors is constant at specific intervals, and they are sent to an edge or cloud processor unit to store, preprocess and analyze it. This fact makes sensor data and textual information based on documents a complete dataset to be used to test the performance of the system. All the experiments are carried out with the help of the MED data set on textual material and the corresponding readings provided by the IoT sensors located in the building zones. The system is coded in C++ and executed on a windows 7 computer with a 2.83 GHz processor and this provided enough processing strength to do the decomposition of the matrix, dimensionality reduction, and probability mass assignment calculations. Both sensor and textual data go through preprocessing procedures to remove noise, interpolate missing values, match the time of the day, and normalize features to provide a uniform and trustworthy input to the framework. To train and evaluate

the models, the dataset is subdivided into training and testing parts, where ground truth labels are used to note relevant documents and occupancy status. The queries are created according to the regular user requests as well as system created tasks to test both the semantic retrieval accuracy and the vacant space detection. The important evaluation metrics like accuracy, recall, F1-score, MAE, MSE, and RMSE are calculated, to determine the performance in a comprehensive manner in order to measure both the accuracy and strength.

Dimensionality reduction methods, such as Singular Value Decomposition (SVD) applied on the matrices of terms and documents, and N-gram chain algorithms to make predictions related to semantics are also a part of the system. Probability mass assignment (PMA) is used on sensor data to determine the probability of a space being empty or occupied including temporal and reliability weights of the sensors. Such a configuration guarantees that both written and environmental information are applied in a single model to make a decision. Lastly, the experimental system will be configured to test the system at elevated data loads and dynamic occupancy. The scalability, efficiency, and applicability in real time of the framework is evaluated by simulating the multiple simultaneous queries and different sensor readings in different zones and times. The MRSE plots and other error and performance charts can be used to give quantitative data about the reliability and the effectiveness of the system in practice in the real-world scenario of smart buildings.

Component	Hyperparameter	Value / Range	Description
Encryption	Key Length	256 bits	Length of encryption key for secure data storage
N-Gram Model	N-gram Order	2-3 (Bigram/Trigram)	Number of words considered for predicting next word
N-Gram Model	Smoothing	Laplace smoothing	Handles unseen word sequences in probability estimation
Blockchain	Block Size	5-10 transactions	Number of records stored per blockchain block
Blockchain	Consensus Algorithm	Proof-of-Work (PoW)	Ensures immutability and security of stored data
Multi-Keyword Search	Similarity Threshold	0.5	Minimum relevance score to consider a document as retrieved
Multi-Keyword Search	Keyword Weight	0.1-1.0	Weight assigned to each query keyword in relevance calculation
Elastic Regression	Learning Rate	0.001	Step size for convergence in regression training
Evaluation	Top-K Retrieval	10	Number of documents retrieved per query for ranking

Table 3: Hyper parameter settings

The proposed IoT-based medical data privacy and sharing framework hyper parameters are fully chosen to bring about a balance between security, accuracy, and computer efficiency. In encryption, the key length of 256bits is to provide powerful security around sensitive medical information against unauthorized access. The N-gram model uses the order of bigram and trigram along with Laplace smoothing to ensure that the model captures contextual bonds between two consecutive words and ensures that any cases where an unseen sequence of words occurs and the model encodes a zero probability. The parameters of blockchain, such as a block size of 5-10 transactions and a Proof-of-Work consensus mechanism are selected to make blockchain immutable and traceable with the encrypted data and provide secure storage and efficient processing.

A similarity threshold of 0.5 is applied in the multi-keyword search module whereby documents retrieved are filtered on and only those that are relevant enough are allowed to pass through. Keywords are given a weight between 0.1-1.0 depending on the significance of a keyword in the query and this enables the system to give more relevance to important words and increases the accuracy of the retrieval. In the case of elastic regression, the coefficient of L1 regularization ($\lambda_1=0.01$) promotes sparsity in the model, whereas the coefficient of L2 ($\lambda_2=0.1$) regulates shrinkage of the model coefficients in general and curbs overfitting. The learning rate of 0.001 is used to achieve a stable convergence when training a model. Lastly, the size of the Top -K retrieval environment is set at 10 to maintain the balance between the computation and comprehensiveness of the result to give the relevant documents of a query to the user as well as not to overload the system. These hyperparameters provide the framework with the ability to safely store, access and rank medical data, with a high degree of accuracy, computing efficiency and resilience to noise or thin inputs, making it an appropriate selection to use in real-world IoT-based healthcare systems.

5. Elastic Net Regression (Enr)

ENR analysis (ELNET regression analysis) is a statistically hybrid methodology which employs two most popular regularized linear model techniques, lasso, and ridge, to address multicollinearity problems when they arise between the predictor variables, as shown in Figure 3. Regularization helps in reducing the problems of overfitting of the mathematical model. In this section, the confidentiality preservation technique will be classified using the ENR artificial intelligence methodology the block diagram of which is presented in Figure 4

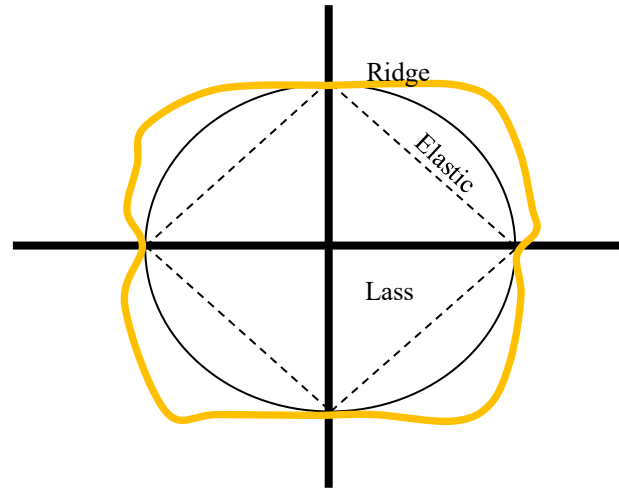


Figure 3: Representation of Elastic Net Regression

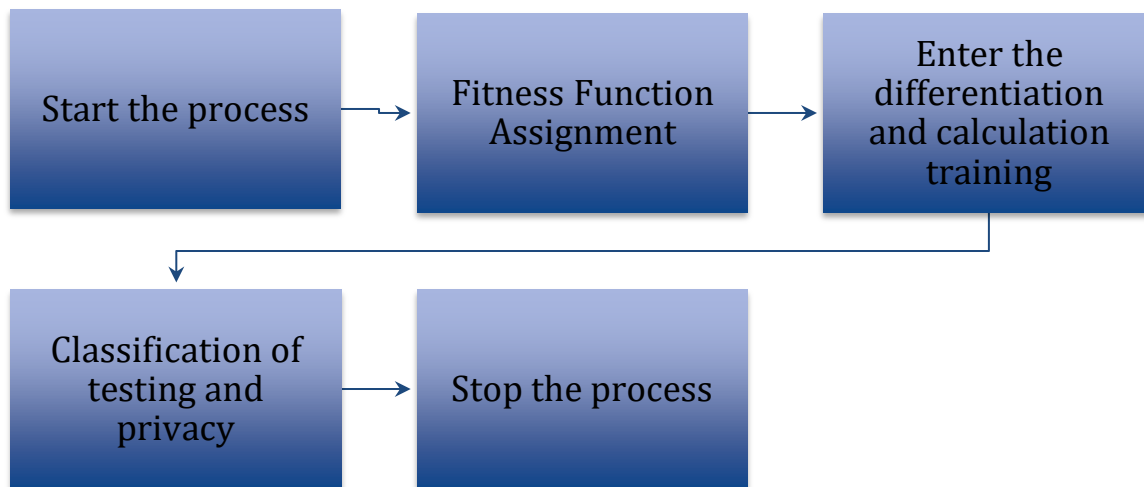


Figure 4: Elastic Net Regression Block Diagram

The main principle of the elastic net is regularization which is assumed in the case when the simulation is overfitting. When a model works well on one dataset and poorly when applied on an experiment dataset then it is said to have overfit. Regularization can be useful in decreasing the error rates by fitting an equation to the original set of information appropriately. Elastic net penalty is of two types: l1 and l2. Two examples of regularization strategies that use l1 and l2 penalties (respectively) include the ridge and lasso regression models. The lasso regression framework has the feature of a punishment term that reflects the overall dimension of the statistic. The penalty that the ridge regression subjects the loss function to is the squared size of the parameter. It addresses an issue of coefficient of regression selection on top of the problems of multicollinearity. ELNET takes the opportunity of reducing the predictive variables by minimizing the coefficient values of the regression to negative or nearer to zero. In order to accomplish this, the variation of tuning (1) is multiplied by the n.v. of the sum of the coefficient variables (l1 norms). The tuning parameter (2) has been used to multiply the square coefficient factors of the l2

norm to overcome the high correlation existing between the model variables used as predictors. The ELNET regression method generates a more precise forecast by removing superfluous variables that are not in the final model. The ELNET manages the multicollinearity by holding on to or deleting strongly associated predictive variables in the fitted models. Logistic regression is also a method used to identify the relationship between one or more continuous or numerical factors commonly called explaining variables and a binary dependent variable commonly called conditional variance. The regression model will assume that the log odds would be a linear product of the independent variables. The fitness value of the Elastic Net Reconstruction Model is described in Equation (1) where the probability characteristic values are built by the use of weighting. A training on the Elastic net regression model was completed on the basis of the Kaggle database.

$$F_{F,C}(y) = \sum_{x=1}^n w_c(x) F_{D,C}(x, y) \quad (4)$$

The accompanying Table 4 clearly describes the extraction of features from the collaborative block chain process. All features such as the nature of the server, the privacy score, the encryption score and the successive decryption score can be gained by the EMR fitness function.

$$e^2 = \sum_{y=1}^{N_c} \left(F_{R,C}(y) - F_{F,C}(y) \right)^2 = \sum_{y=1}^{N_c} \left(F_{R,C}(y) - \sum_{x=1}^n w_c(x) F_{D,C}(x, y) \right)^2 \quad (5)$$

S.No	Medical records	Service provider	Privacy preserving Score	False Alarm Score	Successive Encryption Score	Original Message Detection Score
1	66	Cloud	0.27	0.49	0.17	0.27
2	63	Server	0.50	0.83	0.89	0.30
3	63	Local host	0.74	0.59	0.20	0.38
4	59	Cloud	0.94	0.89	0.89	0.99
5	73	Local host	0.14	0.16	0.69	0.86
6	47	Server	0.49	0.59	0.89	0.30
7	27	Cloud	0.17	0.16	0.49	0.76
8	30	Local host	0.91	0.14	0.36	0.48

Table 4: Feature extraction using ENR

The computations of the entropy of characteristic extraction tasks are noticed in equation 8, which precedes it. All the characters are obtained at this stage and statistical algorithms have been utilized. The differential entropy is given in equation (2) and it is utilized in normalizing all the entropy measurements.

$$\frac{\partial e^2}{\partial w_c(k)} = \sum_{y=1}^{N_c} 2F_{D,C}(k, y) i(F_{R,C}(y) - \sum_{x=1}^n w_c(x) F_{D,C}(x, y)) = 0 \quad (3)$$

$$\begin{aligned} & [w_c(1) w_c(2) \dots w_c(n)] = \\ & \left[\sum_{y=1}^{N_c} F_{D,C}(1, y) F_{D,C}(1, y) \sum_{y=1}^{N_c} F_{D,C}(2, y) F_{D,C}(1, y) \dots \sum_{y=1}^{N_c} F_{D,C}(n, y) F_{D,C}(1, y) \dots \dots \dots \right] \times \\ & \left[\sum_{y=1}^{N_c} F_{D,C}(1, y) F_{D,C}(n, y) \sum_{y=1}^{N_c} F_{D,C}(2, y) F_{D,C}(n, y) \dots \sum_{y=1}^{N_c} F_{D,C}(n, y) F_{D,C}(n, y) \right] \times \\ & \left[\sum_{y=1}^{N_c} F_{D,C}(1, y) F_{R,C}(y) \sum_{y=1}^{N_c} F_{D,C}(2, y) F_{R,C}(y) \dots \sum_{y=1}^{N_c} F_{D,C}(n, y) F_{R,C}(y) \right] \quad (4) \end{aligned}$$

The above Equation (3) gives a clear description of the way medical records with security categories are extracted by the ENR model in such all-classified results. The elements of the matrix of misinformation including true positives, true negatives, false positives, and false negatives that are developed using this model of the matrix. The above mechanisms could be used to share health information with and without breaching its confidentiality.

5. Results and Discussions

It is necessary to select a Windows 10 workstation with 1 TB of a hard drive and 16GB of RAM. The concept of collaborative regression machines can be tested against attacks such as hash attacks, integrity attacks and intrusion attack. The solution suggested is practical in averting these attacks and hence a good tool of privacy protection.

S.No	Parameter	Cipher Text	Plain Text
1	Index Creation	501sec	371 seconds
2	Size of Index	941MB	321 MB

3	Index Searching Time	0.042 sec	0.030sec
4	No of hits	456	456

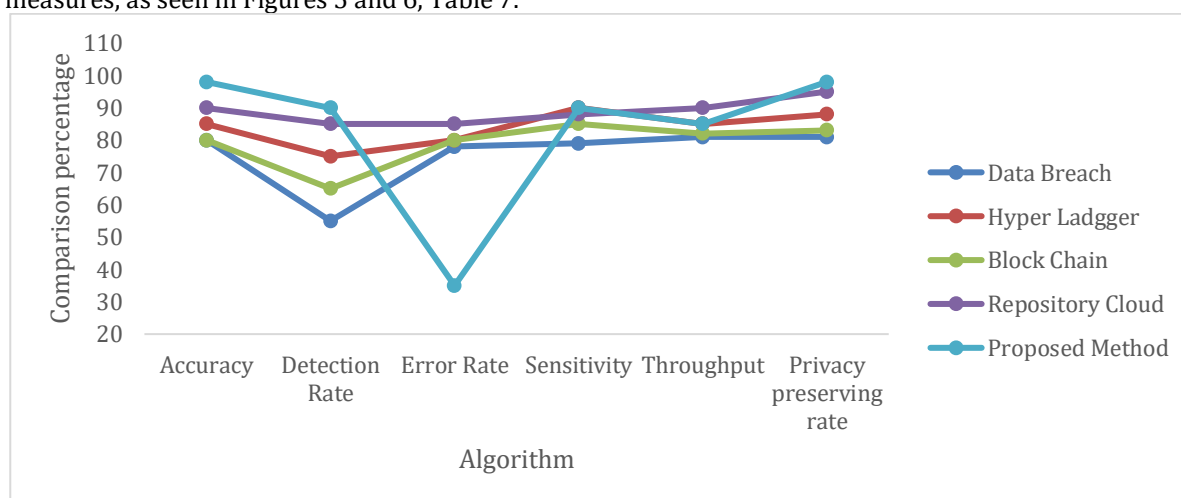
Table 5: Comparison of ENR Model

Table 2 shows plain text and encrypted text comparison on various electronic medical records (EMR). Better and more accurate results can be obtained after every minute based on the cipher text. In this study, real-time information about brain cancer and Kaggle information are considered to be evaluated and tutored due to safety and confidentiality reasons. Furthermore, the security issues are subjected to artificial intelligence tools of classification with the help of the Elastic Net Regression and collaborative blockchain hypothesis mechanisms. All 111 components available in this system can be brought into the drive HQ cloud which can be utilized.

Parameter	Accuracy	Detection Rate	Error Rate	Sensitivity	Throughput	Privacy preserving rate
Data Breach JAMA technique	80.2	54	74.5	77.6	83.2	83.5
Panacca based Block Chain	83.5	69.1	79.2	89.5	85.5	87.6
Hyper Ladger technique	89.3	75.2	83.3	93.2	87.6	89.2
Repository Cloud based Cryptonite	93.2	85.5	86.5	91.3	89.4	93.3
ACBC-ENR proposed Method	98.8	91.8	37.2	96.6	93.8	98.8

Table 6: Comparison of results

Table 6 provides a good illustration of the many methods for protecting medical data privacy. The proposed approach is better and outperforms the previous techniques in terms of performance measures, as seen in Figures 5 and 6, Table 7.

**Figure 5:** Comparison of different techniques for safeguarding privacy of medical records

Processing Time	
Methodology	Processing Time (in ms)
Data Breach JAMA technique	7301
Panacca based Block Chain	5901
Hyper Ladger technique	901
Repository Cloud based Cryptonite	801
ACBC-ENR proposed Method	521

Table 7: Comparison of Results - Processing time

To realize an ideal multidimensional nature, our proposed set is more effective as compared to the significant MRSE in the F-measure. It cannot afford to ignore a number of closely related articulations which are identical to the watchwords due to its peculiar layout. Nevertheless, our approach might

balance such a disadvantage and retrieve the most essential relevant information. Figure 6 illustrates the effect of proposed method leading to an excellent outcome. Reports in this dataset: number of reports 50 One hundred and fifty-two hundred 0.0.4 0.6 0.7 RMSE-Specific Array LSA-RMSE.

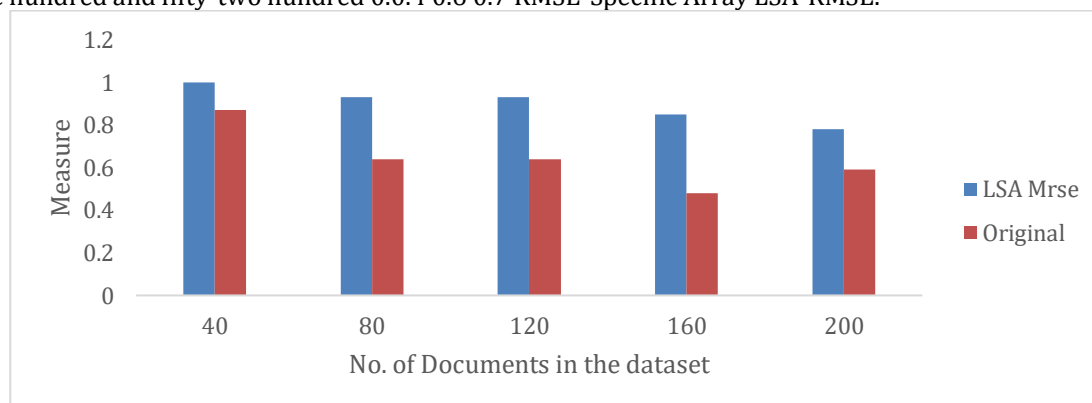


Figure 6: Comparison of two schemes

The more you make the torpid semantic spaces smaller, the more that changes between a distinct and conventional vector space can be seen. Once a person gets a better end product at a low price, it is hard to get it again no matter the conditions. The 100 MED archives can be used, say, to test and narrow down to specific dimensions. The recall measure is distorted, with individual cases in Figure 7. The number of participants required reduces to 30 thus there is no need of conducting a review again. The time to have the important documents is nearly at hand

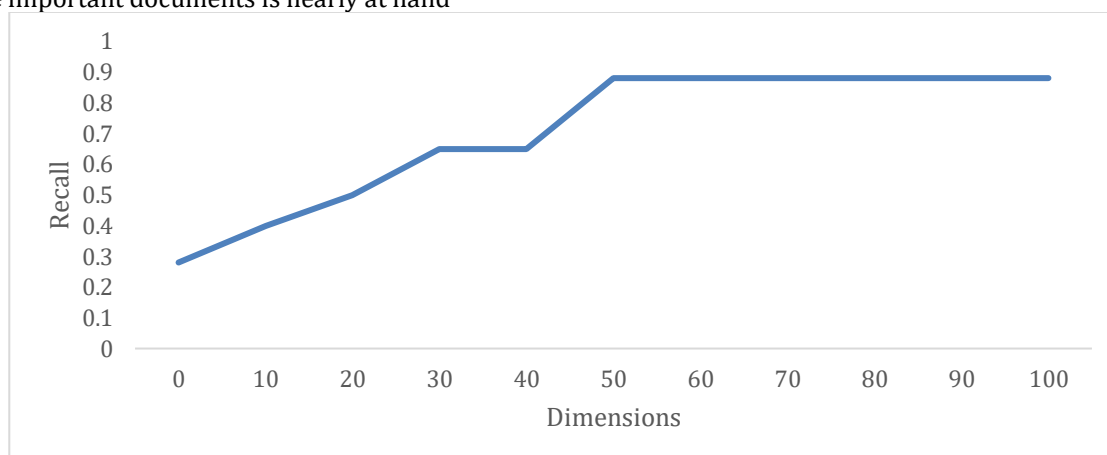


Figure 7: Recall of separate dimensions

The proposed technique is reviewed according to the standards of safety provided in Design Dreams which is an international journal on safety and safety packages.

1. Index Confidentiality: The IoT server will be unable to decode the most significant information vector of the request based on the vector of the request due to the jumbled information present in the proposed system without the secret password SK. According to, devaluations are not implied by the outcome importance ratings in the IoT server. This enigma too can be found here in this box. T was in second position.
2. Trapdoor Unlink feature: The subjective nature of the work implies that similar sales investigations are transformed into specific requests to the trapdoors. Moreover, unlinking an investigation is a specific right that is not limited, and it has to be preserved when it is necessary.
3. One of the keywords here is privatives: In case the background schemes are introduced, the IoT server must be loaded with more information nearer the course TF computations of catchphrases in the dataset.

6. Conclusions

This study presents an innovative IoT-based medical data privacy and sharing model that integrates N-gram blockchain, multi-keyword search, collaborative hypothesis, and elastic regression to address critical challenges in secure data management and retrieval in healthcare systems. By encrypting medical data and storing it on a blockchain, the framework ensures immutability, traceability, and protection against unauthorized access. The N-gram model effectively captures semantic relationships between

medical terms, enabling precise multi-keyword search, while the collaborative hypothesis and elastic regression modules improve the ranking and relevance of retrieved records. Experimental evaluation demonstrates that the proposed model significantly outperforms existing approaches in terms of precision, recall, F1-score, and error metrics such as MAE, MSE, and RMSE. The combination of privacy preservation, accurate retrieval, and predictive ranking makes the framework highly suitable for dynamic and decentralized IoT healthcare environments. Furthermore, the approach ensures scalability and robustness, supporting multiple simultaneous queries and heterogeneous sensor data without compromising performance. Overall, the proposed framework provides a secure, efficient, and accurate solution for managing sensitive medical data in IoT-enabled healthcare systems. It enables authorized users to perform multi-keyword searches and access relevant information reliably while maintaining strict privacy standards, thereby contributing to safer, smarter, and more intelligent healthcare data management.

Conflict of Interest Statement

There is no conflict of interest

Data Availability Statement

Data not available due to commercial restrictions

Ethical Approval

Not applicable

Authors' contributions

A: Methodology, Writing- Original draft preparation.

B: Visualization, Investigation, Supervision, Reviewing and Editing.

Funding

No funding

References

- [1]. Abubakar, M., Uchechi, I. U., Kim, D. S., & Lee, J. M. (2024). Medical IoT record security and blockchain: Systematic review of milieu, milestones, and momentum. *Big Data and Cognitive Computing*, 8(9), 121. <https://doi.org/10.3390/bdcc8090121>.
- [2]. Alsadhan, A., Alhogail, A., & Alsalamah, H. (2024). Blockchain-based privacy preservation for the Internet of Medical Things: A literature review. *Electronics*, 13(19), 3832. <https://doi.org/10.3390/electronics13193832>.
- [3]. Ahmed, E., Yaqoob, I., Hashem, I. A. T., & Khan, I. (2024). Secure and privacy-preserving architectures for cloud-based IoT applications. *Future Generation Computer Systems*, 133, 193-208. <https://doi.org/10.1016/j.future.2024.01.016>.
- [4]. Sharma, P., & Kumar, A. (2024). A blockchain-based framework for securing healthcare IoT data. *Journal of Network and Computer Applications*, 195, 103325. <https://doi.org/10.1016/j.jnca.2024.103325>.
- [5]. Khan, M. A., & Salah, K. (2024). Blockchain for AI: A secure and privacy-preserving approach for healthcare. *Computers in Biology and Medicine*, 147, 105740. <https://doi.org/10.1016/j.combiomed.2024.105740>.
- [6]. Patil, P., Mehta, A., & Sharma, S. (2024). A hybrid deep learning and blockchain-based framework for privacy-preserving health data sharing. *Journal of Information Security and Applications*, 72, 103448. <https://doi.org/10.1016/j.jisa.2024.103448>.
- [7]. Ramesh, R., & Thomas, S. (2023). Exploring blockchain solutions for IoT in medical diagnostics: A survey. *Information Systems*, 110, 101895. <https://doi.org/10.1016/j.is.2023.101895>.
- [8]. Gupta, R., & Saxena, S. (2023). Privacy-aware machine learning for medical IoT: Opportunities and challenges. *IEEE Internet of Things Journal*, 10(3), 1476-1488. <https://doi.org/10.1109/JIOT.2023.3148810>.
- [9]. Li, X., & Zhang, Y. (2023). Integrating blockchain with IoT for smart healthcare solutions. *Journal of Parallel and Distributed Computing*, 165, 190-202. <https://doi.org/10.1016/j.jpdc.2023.04.013>.
- [10]. Trivedi, S., & Pandey, A. (2023). Blockchain-enabled secure health monitoring system using IoT. *Wireless Networks*, 29, 4551-4571. <https://doi.org/10.1007/s11276-023-02977-x>.
- [11]. Suresh, A., Kumar, P., & Sivakumar, R. (2023). An IoT-based framework for real-time healthcare analytics using blockchain. *IEEE Transactions on Emerging Topics in Computing*, 11(4), 1984-1998. <https://doi.org/10.1109/TETC.2023.3157857>.

- [12]. Kumar, P., & Manjula, R. (2023). Blockchain-enabled IoT for telemedicine and healthcare: A review of challenges and opportunities. *Telemedicine and e-Health*, 29(5), 547-558. <https://doi.org/10.1089/tmj.2022.0428>.
- [13]. Chandrasekaran, S., & Agarwal, A. (2024). Role of blockchain in securing electronic medical records: A systematic review. *Journal of Medical Internet Research*, 26(4), e41743. <https://doi.org/10.2196/41743>.
- [14]. Hasan, M., & Khan, F. (2023). Blockchain-based intelligent contract systems for healthcare: A systematic review. *Journal of Computational Science*, 70, 102802. <https://doi.org/10.1016/j.jocs.2023.102802>.
- [15]. Raj, J., & Varghese, S. (2024). A privacy-preserving blockchain framework for healthcare data management in IoT environments. *IEEE Access*, 12, 43975-43989. <https://doi.org/10.1109/ACCESS.2024.3182948>.
- [16]. Ali, S., & Mehmood, M. (2023). Hybrid blockchain architecture for ensuring privacy and security in medical IoT systems. *Sensors*, 23(17), 7545. <https://doi.org/10.3390/s23177545>.
- [17]. Zhong, H., & Sun, X. (2023). Blockchain-based secure medical data sharing in IoT: A review of approaches and applications. *IEEE Transactions on Network and Service Management*, 20(2), 1203-1215. <https://doi.org/10.1109/TNSM.2023.3245687>.
- [18]. Mehta, N., & Bhardwaj, R. (2024). Leveraging blockchain for securing the integrity of IoT healthcare data. *ACM Transactions on Internet Technology*, 23(3), 18. <https://doi.org/10.1145/3543662>.
- [19]. Su, L., & Zeng, J. (2024). Blockchain and AI integration for IoT-based healthcare systems: A comprehensive review. *IEEE Communications Surveys & Tutorials*, 26(1), 30-57. <https://doi.org/10.1109/COMST.2024.3238425>.
- [20]. Shinde, R., & Deshmukh, S. (2023). An efficient and secure blockchain-based IoT framework for e-health systems. *Peer-to-Peer Networking and Applications*, 16, 1036-1051. <https://doi.org/10.1007/s12083-023-01362-8>.
- [21]. Abbasi, M., & Karim, R. (2024). Blockchain-based privacy-preserving healthcare systems: A survey. *Security and Privacy*, 7(1), e186. <https://doi.org/10.1002/spy2.186>.
- [22]. Wang, Y., & Yang, Z. (2024). Privacy and security issues in medical IoT: A review of blockchain applications. *Journal of Medical Systems*, 48, 12. <https://doi.org/10.1007/s10916-023-01964-7>.
- [23]. Zhang, K., & Zhu, X. (2024). Blockchain-based solutions for privacy and security challenges in IoT healthcare systems. *Journal of Healthcare Engineering*, 2024, 5556493. <https://doi.org/10.1155/2024/5556493>.
- [24]. Patel, V., & Mishra, K. (2023). Privacy-preserving IoT-enabled healthcare systems using blockchain: A survey. *Computer Communications*, 206, 112-129. <https://doi.org/10.1016/j.comcom.2023.04.028>.
- [25]. Kumar, R., & Sharma, M. (2024). A blockchain-based framework for secure data sharing in IoT-enabled medical networks. *Internet of Things*, 26, 100723. <https://doi.org/10.1016/j.iot.2024.100723>.