Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 218 (2023) 989–1002

Procedia Computer Science

www.elsevier.com/locate/procedia

International Conference on Machine Learning and Data Engineering

An efficacy analysis of data encryption architecture for cloud platform

Sheenam Malhotra[a], Williamjeet Singh[b]

[a]*Research Scholar, Department of Computer Science and Engineering, Faculty of Engineering and Technology, Punjabi University, Patiala, Punjab, India,*

[b]*Assistant Professor, Department of Computer Science and Engineering, Faculty of Engineering and Technology, Punjabi University, Patiala, Punjab, India*

Abstract

In recent times, cloud computing is being utilized largely for storage and information sharing purposes in several established commercial segments, particularly those where online businesses are prevalent, such as Google, Amazon, etc. Cloud system presents several benefits to users in terms of easy operations, low implementation, and maintenance expenses. However, significant risks are encountered in the data security procedures of cloud systems. Although the area is frequently being analyzed and reformed, the concern of cloud data protection and user reliability remains under uncertainty due to growing cyber-attack schemes as well as cloud storage system errors. To deal with this risk and contribute to the endeavor of providing optimal data security solutions in cloud data storage and retrieval system, this paper proposes a Symmetric Searchable Encryption influenced Machine Learning based cloud data encryption and retrieval model. The proposed model enhances data security and employs an effective keyword ranking approach by using an Artificial Neural Network. The comparative assessment of the proposed model against multiclass SVM and Naïve Bayes has established the better operational potentiality of the model. The effectiveness of the proposed work is justified by the association between high TPR and low FPR. Further, a low CCR of 0.6973 adds up to the success of the proposed work.

Keywords: Cloud Computing; Symmetric Searchable Encryption; Machine Learning.

1. Introduction

Cloud computing is now widely regarded as a highly developed and forward-thinking technology. According to an IHS analysis, global investment in cloud infrastructure and services hit $174.2 billion in 2014, up 20% from $145.2 billion the previous year [1]. More and more enterprises and users are going to outsource their local data to public clouds due to its benefits such as flexibility and low management expenses [2].

Cloud computing offers on-demand network access to a shared pool of configurable computing resources (e.g., networks, servers, storage, applications, and services) that may be quickly supplied and released with minimal administrative effort or service provider contact.

10.1016/j.procs.2023.01.079

All hardware resources are viewed as services and delivered via the Internet in order to provide reliable, quick, and efficient data management and net computing services. Benefits of the cloud include integration, adaptability, flexibility, availability, the ability to adjust to changes in demand, the power to speed up development activity, and the potential for cost savings through simplified and effective processing.

In order to meet users' computing needs, CC combines a number of computational ideas and technologies, including Service Oriented Architecture (SOA), Web 2.0, selection and placement of virtual machine, and others with a dependence on the Internet. Customers' operating system and data are kept on web server, and common business applications are made available online through web browsers. In some aspects, the term "cloud computing" refers to the development of these technologies and serves as a marketing term for both that development and the services they provide [3].

Despite the many advantages of cloud computing, the most crucial area of concern is privacy and security. Cloud computing's key security aspects are data security, access management, data utilization management, and trust. We give a comparative assessment of the existing Encryption Architectures for Cloud platforms in this study, which allows us to determine the best data security and retrieval model to use.

1.1. User side data management concerns in the cloud

The cloud-based information must be protected, especially data protection in the cloud storage, in order to be secure. By prohibiting un-authorized access, this enhances data security [4]. Although it is the cloud based supplier's duty to offer users a stable and reliable storage service, a number of circumstances could compromise the safety and integrity of user data. In such a case, the service provider might be forced to forgo the loss of the user's banking system and reputation in an effort to cover the full extent of the damage.

In order to prevent data loss due to a single storage node's original reliability of the data being subverted, data in cloud services is often separated into slices and encrypted before being stored in multiple storage nodes [2].

Because of these traits, certain standard network and storage security technology are no longer fully relevant in cloud storage environments. For instance, message digital signatures are employed in traditional storage technology to confirm file integrity, but with cloud storage, the data is stored on a remote server, making it impossible to periodically retrieve the data and validate the signature to ensure data integrity.

There is a requirement for a trustworthy and efficient fault - tolerant system that can guarantee that even if a number of slices are lost, the security of files may be restored utilising the remainder slices in addition to cloud storage, which separates data into fixed-size parts.

Agencies are now attempting to avoid focusing on the IT infrastructure. They must focus on their business operations to boost productivity. In comparison to the traditional IT model, cloud computing has several advantages. Questions regarding cloud computing security, on the other hand, are a key barrier to cloud computing adoption from the user's perspective. The availability of computer network services, primary data storage, and processing power, without explicit user active control, is referred to as cloud computing. Using cloud service suppliers' services, cloud data is processed and retrieved on a web server.

As a result, the value of cloud computing is increasing, making it a rising market that is attracting a lot of interest from the educational and business sectors. The cloud storage solution, on the other hand, has several drawbacks, including a lack of access and security concerns. Because the cloud storage service is centered on two-way data sharing between the service provider and user, cloud computing security concerns include honesty, integrity, availability, verification, permission, and confidentiality.

As a result, the risk of data compromise is increasing, and it may be classified into two categories: vital data and archival material. Important information is information that a subscriber needs at any given time and would be irritated by any halt or disappearance. Furthermore, archival data is data that is extremely rare in its entirety, and typically in a non-critical moment. As a result, a gap in it won't be able to be regarded as a major issue.

When using the internet-cloud platform, data security and privacy must be prioritized. Data loss or disclosure can have a significant negative impact on a company's brand and confidence. Data leakage prevention is viewed as the most pressing issue, accounting for 88 percent of main concerns. Similarly, data remoteness and privacy have a 92

percent impact on security issues. Data protection, dependability, honesty, availability, authentication, and secrecy, as well as a lack of resources and skills, are among the most serious security challenges in cloud computing [4].

- Integrity: Data in the cloud can be harmed as a result of data transmission to cloud storage. Because the information and calculations are outsourced to a web server, the data's validity should be monitored and maintained at all times to ensure that the information and calculations interact. Data integrity refers to the protection of records from tampering. It is necessary to make some adjustments to the specifics.
- Availability: The ability of a cloud subscriber to receive critical data at any time is referred to as accessibility. The ability to keep access to cloud computing resources at any moment is a crucial concern for each firm. A system is called available when an authorized individual can use and control the device at any time and preserve data.
- Vendor lock-in: Companies that employ cloud-based services frequently opt to change their Cloud Service Provider (CSP) and go with a different one. This could be an explanation since the CSP will no longer adapt to the tenant's needs, regardless of whether there are improvements or upgrades in the services that the client does not expect. They are unable to meet the needs of clients or any other trigger that causes the customer to migrate to a different CSP, but they are unable to exit this situation, which is known as Vendor Lock-in.
- Data Security: To increase the security of cloud computing, it is necessary to provide encryption, certification, and intrusion detection for information maintained in the cloud.
- Interoperability: This is defined as the willingness of two or more processes to collaborate to share and use data. Companies will be unable to combine their information technology network in the cloud to obtain efficiency and cost savings if these systems are not integrated. There are also cloud-computing networks that are designed as closed systems that do not link.

Due to the quick development of information technology, there is a significant amount of data outsourcing to cloud servers, and multiple attacks will jeopardize the secrecy of cloud data. Before being transferred to the cloud, user data is regularly encrypted in order to prevent data leakage and guarantee data security. The fact that ciphertext is kept on external servers prevents the use of standard search algorithms. It is required to apply an appropriate searchable encryption technique to find the target data [5].

To address these issues, searchable encryption (SE) technology can provide data privacy and availability while also allowing ciphertext data to be queried and retrieved. A searchable encryption approach often consists of encryption, tokens, search, and decryption.

• Encryption: After encrypting the data and building the index structure, the user uploads the ciphertext and the index structure to the server.

• Token: Users create a trapdoor for keywords using a key, and the token is required to keep any keyword information private.

• Keyword search: The server runs the search algorithm using the keyword and returns the cypher text with the matching keywords. Just the keyword information in the ciphertext must be obtained by the server.

• Decryption: Users decrypt the server's encrypted files with the help of the key to get the search results. Searchable encryption systems can be divided into two categories:

i) Searchable symmetric encryption (SSE): SSE is a method for retrieving ciphertext that is based on asymmetric encryption. Both data owners and users exchange the vital information.

ii) Searchable asymmetric encryption (SAE), a form of public key encryption that is suitable for one-to-many data sharing scenarios. To guarantee its security, many hypotheses, including Decision Bilinear Diffie-Hellman, is applied (DBH).

The separation of public and private keys makes the SAE technique particularly suitable for multi-user data sharing systems, despite its typical inefficiency and reliance on bilinear pairings, which results in significant algorithm complexity. The main motivation of this research is to develop complete data encryption and retrieval security system in the cloud platform. It is known that the existing cloud system is not systematic. The development of Searchable Encryption (SE), the accepted cloud data protection scheme gives end-to-end security and privacy. The devised encryption system assures privacy and security. Further, research has been done using query encryption for document that encrypts data in the cloud. The document is given a keyword, and during decryption, the ranking has been done using ANN and extracting the best appropriate document based on the keyword.

1.2. Contribution of the Study

Based on contextual understanding and justifying the contemporary importance of the concern, this research is developed as an analysis to identify the shortages and scopes of various Encryption Schemes used for data security in a Cloud Environment.

The study's main objective is to determine an optimal data security and retrieval scheme in cloud platform through:

- Comparative analysis of various encryption architectures based on their performances
- Enlist the merits and demerits of the available cloud data encryption and retrieval schemes
- Justify the efficacy of the optimal data encryption architecture for cloud platforms based on the above evaluations
- Using a combination of HAC Tree, Neural Network for index generation, and Cosine Similarity to support the encryption mechanism, recognizes the value of Cosine Multi-Keyword Searchable Index (CMSI).
- Introduction of Neuro-Rank Policy

1.3. Organization of the study

The presented research is arranged as follows: Section 1 introduces the concept of cloud computing followed by the challenges and methods to address the various challenges. Section 2 illustrates the literature review in which the pros and cons of state of art techniques have been described. Further, the research gap is defined in section 3. The next section elaborates on the research methodology and experimental study. Section 5 defines the simulated results and is lastly concluded in section 6.

2. Literature Review

Several scholarly works are done to assess, compare and refine the encryption architectures that are applied for cloud computing data security purposes. Some significant studies are given in the following table 1.

3. Research Gap

The review of literature presented in the previous section shows a significant number of approaches developed and tested over the past years (they are done even before the period covered in this study) to ensure privacy while using cloud-based on trust, encryption and access control. However, none of them formally claimed to be universal and eligible for providing complete data encryption and retrieval security in the cloud platform. Furthermore, the systems are scattered and mostly not systematic.

Searchable Encryption (SE), the accepted cloud data protection scheme offers the privacy of the data when data is transferred from local storage. Although, privacy is assured, it further complicates the taskthe at server level. For traditional query-based search and retrieval processes, content and keywords are revealed. Therefore, encrypted search also inhibits the searching operations.

Trapdoor-based user queries as experimented with within [6], [7], and [8] cannot fully ensure data protection from harmful attackers, where they can guess the keyword through the use of a constant trapdoor. Thus, the attackers can identify the frequent trapdoors. Although [8] has enabled authenticated keyword search as an improvement over the PEKS scheme, the model is vulnerable to deliberate targeted attacks. Also, these models are high on overhead costs.

Multiple keyword-based searches are secure, faster, and low cost than single keyword-based searches [11]. These tools, such as the model used in [9] also reduce query representation and its expression problem (typos, spelling mistakes, etc.).

Table 1. Comparative Analysis of the Existing Literature

Author	Year	Proposed Encryption Architecture	Advantages	Disadvantages
Hui Yin *et al.* [6]	2017	They suggested a search strategy that boosted privacy by allowing the data to generate a different random query trapdoor each time. We create a secure indexing for each data frame using the Bloom filter and the bilinear pairing operation, allowing the cloud to do a search without receiving any helpful information.	Ensures user's privacy of query content; Secure Search Scheme	Time-Consuming to do the search on a cloud server than KNN and SSE
Tahir *et al.* [7]	2017	By precisely describing keyword-trapdoor in distinguishability and trapdoor-index table in distinguishability, they enumerate the characteristics of a "secure" ranked SE system. To counter passive attacks, they developed and demonstrated a novel Ranked based SSE that is entirely based on a probabilistic encryption algorithm.	Lightweight; Efficient in performing large-scale data search.	Security concerns exist on leakages through Index table and search outcome; although no information is exposed of data outsourcing; works ONLY with a single keyword
Huang *et al.* [8]	2017	Presented the idea of Public-key Authenticated Encryption with Term Search (PAEKS), which allowed the data sender to both encrypt and authenticate a keyword. The verification would be persuaded that the sender alone is responsible for creating the encrypted keyword. Based on straightforward and static presumptions, the random oracle model's security was examined using the provided security models.	The scheme is comparable in efficiency with Boneh et al.'s scheme.	Bears security risk in a real-life condition where the attacker chooses a user to collect data information.
Poh *et al.* [9]	2017	Presented a comparative study on available Searchable Symmetric Encryption (SSE) models to classify their features and evaluate the model efficacy	Most models used a combination of index tables and trees for reliable updates and storage; External mechanisms such as Oblivious RAM can be employed to reduce leakages.	Practical use of external devices to reduce leakages is not properly analyzed; Mostly uses Index Tables that are prone to leakage risk; Search is time-consuming; Not a preferred option where I/O access is essential
Ilakiya *et.al*. [10]	2019	OTP-based Secured Information Retrieval from the Cloud Using Human Voice.	Data will be protected by using multi authentications.	High storage is required for this kind of authentication. This kind of authentication can also be extraneously influenced by one sore throat and cold.
Sun [5]	2019	Encryption and decryption of images using classical and quantum cryptography.	More secure in the case of exchanging multimedia data.	Increase the communication range and bit transfer rate.
Shan Jiang *et al.*[11]	2019	Proposed a multi-keyword search protocol with bloom filter support that is more effective and protects user privacy. In the protocol, a multi-keyword search operation filtered the database using a low-frequency term chosen using a bloom filter. It was suggested to utilise pseudorandom tags to make it easier to finish each search operation in only one go	Low in computational cost as most of the data are excluded for the use of low-frequency filter; efficient database with dynamic update option; faster and safer than single-keyword based blockchain models	Unreliable and not feasible for large-scale search; Blockchain data can be mishandled by attackers; Risk of data loss

		to conduct extensive studies and implemented the protocol in a local, simulating block chain network.		
Malhotra *et al.* [12]	2019	The study demonstrates brand-new secure storage and the document ranking system for the cloud. The data is encrypted based on the correlation between the data files determined by cosine similarity because no previous reference for any data is retained on the server. Through the use of supervised machine learning, the retrieved data is ranked.	Multi-keyword-based model and so better in search flexibility; Can be enhanced with other machine learning algorithms, such as SVM, ANN	Model accuracy may vary based on real-time database configuration
Islam *et al.* [13]	2019	Set up a secure authentication system and appropriate cryptography for cloud computing. Auto encryption and KEYs changing processes were part of the operation in the cloud end. Customers were initially not sent newly generated KEYs. In order to authenticate users, three actions would be taken. The encryption procedure can be started manually by Cloud Service Providers (CSP) at any time or automatically once users log out.	Better data security from hacking	Time-consuming
Suneetha *et al.* [14]	2019	ANN was integrated to enhance the security and confidentiality of the cloud computing environment. The work involved a dynamic hashing component for the storage of the sensitive data.	Focussed on ensuring data confidentiality	Can face the risk of data loss in real practice with unskilled/improper handling
E. Nirmala *et al.* [15]	2021	Constructed a keyword searching Binary Tree algorithm added with multiple corrections features based on ranking. Fuzzy Gramm was used to address spelling errors while ancestral relations can be found based on stemming procedure.	Better in security and reliability; less time consuming; more compatible than traditional methods; workable with single/multiple keywords	Model chiefly focussed on text-based search
Tyagi *et al.* [16]	2021	AES and Fernet were used to present to double the encryption level. Along with CNN auto-encoders were also used to protect the data that is available in the form of images in the cloud.	Effective in Image Encryption in Cloud Environment	AES, when adopted/utilized on Solid-State-Drives (SSDs) it's identified as less foolproof and thus adopts a hybrid model is recommended for confidential datasets. For Fernet, the key disadvantage is that the key could be obtained by third parties while transferring to the receiver's end which is highly-risky and a huge drawback in Fernet and other symmetric cryptography.

Sana *et al.* [17]	2021	A safe and secure data communication was assured with the secure design based on ANN and encryption techniques. Here, third could access the data in encrypted form and thus data is not disclosed to ensure privacy. The work involved Matrix Operation-based Randomization and Encipherment (MORE) along with neural network architecture.	The model can work for speech and voice recognition; less time-consuming; better in data retrieval accuracy	The model is operated with homomorphic encoding and so needs to take care of the noise reduction measures; it is costly; high in computational expenses; MORE encryption used in the model may pose a security risk
Zulifqar [3]	2021	The work have presented a Verifiable public key encryption in a multi-user cloud platform to represent a homomorphic encryption.	Less complexity of computation.	Perform better only with fewer search keywords.
Bernardo Pulido-Gaytan *et al.* [18]	2021	Analysed the fundamental concepts of Fully Homomorphic Encryption (FHE) revolving around a cloud environment with discussion on practical implementations, advantages, limitations, practical implications in concern to neural networks.	Easy to deploy; Enhances reliability	Difficulties exist in performance analysis, bootstrapping, and overhead.
Ma *et al.* [19]	2022	Designing the hybrid encryption technique using the encrypted images and DenseNet model has been used for a fine-tuning ad a feature extractor was constructed to improve the performance.	The proposed model improved DenseNet model is 8 to 9 times smaller than the standard convolution method.	There is a small reduction in accuracy due to the adoption of the encryption model.
Zhang *et al.* [20]	2022	An encrypted retrieval scheme was proposed considering the multiuser search encryption model. The authors used the LSTM model to extract the semantic features.	The proposed system protects data privacy.	There is a leakage of cipher text location and a fuzzy keyword search query was not used for efficient encryption,
Wang *et al*. [21]	2022	A novel searchable technique has been developed using the Linear Secret Sharing scheme. The authors used the various search mechanisms and used 0,1 coding theory.	The system's computational and storage efficiency is high.	The time to execute the ciphertext transfer is more in comparison to other search techniques.

Single Keyword-based searches are usually found to be time-consuming, costly (from peer-stakeholder charges, such as computational resources, bandwidths, etc.), and prone to attacks when repeatedly used. The attacker can retrieve them by guessing the common encryptions.

Index table-based traditional searchable symmetric encryption tools in [9] have a high risk of data leakage. Ranked Searches are considered efficient and usable for large-scale data over cloud platforms as [7] and [15] based on their reliability and privacy preservation. Asymmetric Key Encryption tools [10] are considered more secure than Symmetric Encryption tools. However, they are time-consuming and costly to develop and maintain.

Data retrieval systems with Bloom Filter reduce the search time as the method can directly search the index in place of scanning the whole cipher text. However, bloom filter is also associated with a limitation [11] that they have false positive probability and so are unreliable.

Homomorphic Encryptions [17], [18] that are currently implemented in cloud data security are reliable applications. But they are costly and most variable in their performances in real-life conditions. The Machine Learning cloud data storage and retrieval tools as [12] – [17] are efficient and more flexible in terms of search types and can be utilized for image and voice recognition. However, searchable encryption gaining attention in terms of flexibility [22].

They reduce computational time and privacy-preserving. But, with the rise of deep learning-based cyber-attacks, these tools are found to be exposed to the latest malicious intrusions. Thus, for real-world applications these tools

need to be equipped. Further, the latest technological advances involving neural based processing architectures and block chain assure promising future.

A thorough analysis is required on the existing research and suitable adaptability to prevent the data-stealing/damage that is commonly occurring due to cloud platform limitations or cyber-attacks.

4. Methodology

The primary goal of this research is to provide a methodology for document or query encryption that encrypts data in the cloud. The document is given a keyword, and during decryption, ANN is used to rank and extract the best appropriate document based on the keyword. Certain parameters are computed to determine whether a document is approved or denied for a certain keyword.

4.1. HAC index:

Term frequency (TF) and Inverse document frequency (IDF) are utilized for the computation of the HAC index in the present section. Here, TF represents the ratio of the occurrence of the term to the total number of terms in the dataset. And IDF is the ratio of one term occurrence in the present document to the total number of occurrences in another document. HAC is calculated with two components M in the numerator and Q in the denominator and it is mathematically represented by the following equations

$$M = \frac{TF}{\sqrt{\sum_{t=1}^{k} (TF_e)^2}} \tag{1}$$

Here, $TF = \frac{\sum_{i=1}^{n} J_i \times count}{n}$ (2)

J=current term for the processing

$$Q = \frac{IDF_e}{\sqrt{\sum (TF_e)^2}} \tag{3}$$

Here, $IDF = \frac{TF_{present}}{\sum_{t=1}^{k} TF_{other}}$ (4)

e= number of elevations

$$HAC_{index} = \frac{M}{Q} \tag{5}$$

4.2. Neural index

Machine learning is utilized for the ranking of SE keywords. This Machine Learning architecture perform training based on the data extracted using TF, IDF, M, and Q. The three layered neural network architecture is shown in Fig. 1. The four input values are presented by P1, P2, P3 and P4 to generate the output using Machine Learning index which ranks the document.

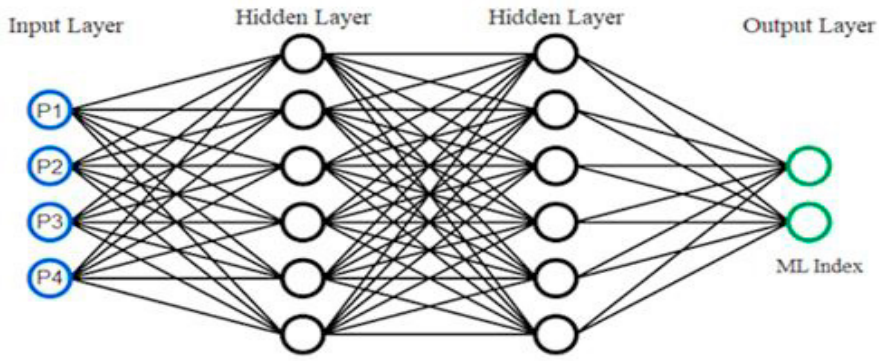

Fig.1. Artificial Neural Network

The ground truth for the output ML index is given by the K-means based clustering technique. K-means is understood as a repetitive process which is used for the clustering of the similar records into k number of clusters. Here, each cluster is represented by a central point labelled as a Cluster Head (CH). The data is clustered into, same cluster if the Euclidean distance between term and the CHs is minimum. Mathematically, the equation is as below:

$$E_d(docx_i, ch_k) = \sum_{j=1}^{m} (docx_{i,j}, ch_{k,j})^2 \quad (6)$$

Where,

$$docx_i = docx_{i,1}, docx_{i,2}.....docx_{i,m}$$

$$ch_k = ch_{k,1}, ch_{k,2}, ..., ch_{k,m}$$

4.3. Search Index

$$Search\ Index = \log(HAC + Neural\ Index) \quad (7)$$

Here,

Upper Boundary (UB)= $Nei + Nei \times .30$ / / where Nei is the stored index, A 30% upper margin is considered.

Lower Boundary (LB) = $Nei - Nei \times .30$ / / 30% lower margin

If $SI \geq LB, and\ SI \leq UB$ then, RL.append which is adding the recommendation value otherwise reject the document. The process of the proposed model is as in the below displayed flow chart in Fig. 2.

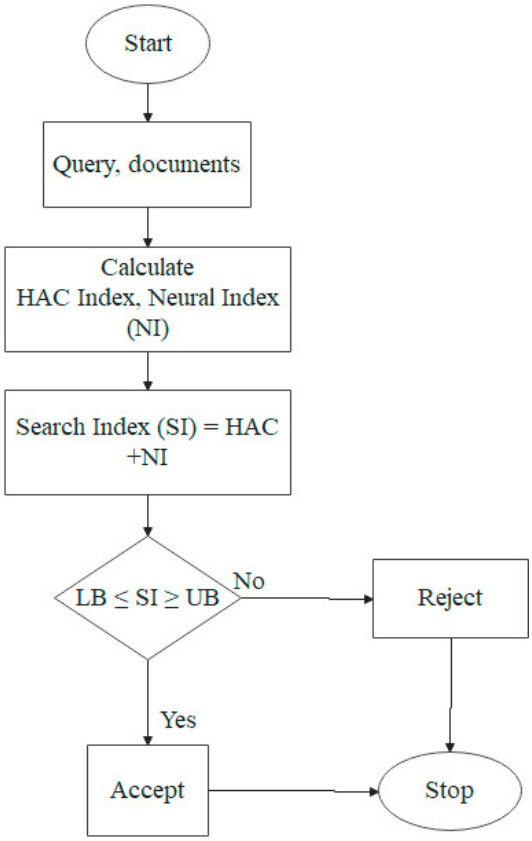

Fig. 2. Work flow of Proposed Algorithm

5. Results

In this section, the list of keywords that are mentioned for the search word is displayed and all the parameters get calculated. The results for recommendation are also tabulated below.

The tweets that are used as the keywords are summarized in Table 2. The search term chosen is “Sheenam requires a dentist with knowledge of tooth gums”.

Table 2. List of Keyword

‘dentist’	‘crown’	‘dental’	‘Insurance’	‘face’	‘appointment’	‘payment’	‘tooth’	‘ache’	Watch	time
‘tooth’	‘dentist’	‘haircut’	‘bones’	‘teeth’	‘pain’	‘salt’	‘orajel’	‘ache’	Afford	Time
‘canal’	‘afford’	‘dentist’	‘root’	‘teeth’	‘advil’	‘cavity’	‘solution’	Work	Root	Endurance
‘tooth’	‘dentist’	‘afford’	‘glasses’	Ignore	‘pain’	‘jobs’	‘money’	‘hurt’	Whiskey	Teeth
‘filling’	‘afford’	‘insurance’	‘dentist’	Appointment	‘teeth’	‘genes’	Afford	Fee	Consumption	Gum
Cricket	dentist	pain	teeth	yard	list	runs	wicket	toss	win	loose

Table 3. Result for the Recommendation

Max TF	Max Idf	HAC	Neural Index	Neutral Index= log (HAC+ Neural Index)	NI-.30	NI+.30	SEARCH INDEX	OUTCOMES
0.75758	0.090909	0.833333	5.334134	1.724801	1.207361	2.242241	2.064346	ACCEPTED
0.75758	0.090909	0.833333	5.400214	1.734745	1.214322	2.255169	2.159225	ACCEPTED
0.75758	0.090909	0.833333	6.537944	1.917689	1.342382	2.492996	1.80181	ACCEPTED
0.75758	0.090909	0.833333	6.802597	1.955842	1.369089	2.542594	2.375813	ACCEPTED
0.75758	0.090909	0.833333	6.713969	1.943226	1.360259	2.526194	1.816691	ACCEPTED
0.75758	0.090909	0.833333	5.857237	1.812298	1.268609	2.355988	2.410054	REJECTED

The generated indexes have been used for the storage and processing of the documents against their ground truth using Neural Networks (NN) having feed forward orientation methods. To evaluate the performance, true positive rate (TPR) along with false positive rate (FPR) has been calculated to determine the significance of the document to its relative class. In addition to that, a computation Cost Ratio (CCR) is also calculated for the proposed algorithm to the indexes generated by HAC only and Neural only. The CCR is calculated with the NN classification algorithm. The evaluation is done based on several relevant documents searched in the given interval of time that is being utilized by the proposed algorithm. To check the performance, a time of 60 seconds is supplied to search and list data from various categories.

Table 4. Classified Results

Number of Searches	TPR proposed	TPR multi-class SVM	TPR-Naïve Bayes	FPR proposed	FPR multi-class SVM	FPR-Naïve Bayes
100	0.95845908	0.85590128	0.8846419	0.04154092	0.14409872	0.1153581
200	0.93943807	0.8134781	0.85424481	0.06056193	0.1865219	0.14575519
300	0.95493366	0.86945992	0.88373954	0.04506634	0.13054008	0.11626046
400	0.92610297	0.88529536	0.85307919	0.09389703	0.11470464	0.14692081
500	0.92986068	0.80521886	0.85671023	0.07213932	0.19478114	0.14328977
600	0.94462483	0.80696161	0.87114682	0.05537517	0.19303839	0.12885318
700	0.94788054	0.8050978	0.89461094	0.05211946	0.1949022	0.10538906
800	0.92019485	0.83528332	0.8952823	0.08980515	0.16471668	0.1047177
900	0.9197219	0.85870628	0.8863668	0.0852781	0.14129372	0.1136332
1000	0.93950663	0.8295922	0.88777715	0.06049337	0.1704078	0.11222285

As shown in table 4, the evaluations have been done using various data patterns that are associated with the dataset. A total of 1000 data elements have been supplied and the proposed work algorithm performs significantly well when

it comes to TPR and FPR. The maximum TPR attained by the proposed algorithm is .95845908 at 100 data elements viz. 100 searches containing tweets from various categories. The overall TPR of the proposed algorithm lies between .92-.96 and the average is 0.938. The other state of art algorithms like multi-class SVM if combined with the proposed index structure, result way behind the proposed algorithm. The maximum attained TPR with multi-class SVM is .8852 whereas the Naïve Bayes algorithm performs marginally better than multi-class SVM. The maximum attained TPR for Naïve Bayes is 0.8952. As illustrated earlier, a total of 60 simulation seconds have been utilized against passed relevant queries and a total number of true responses have been noted which is shown in table 5. The count values obtained using the proposed, HAC and Neural Index are shown in Table 5.

Table 5. Count Value table

PROPOSED COUNT	HAC Count	Neural Index Count
760	530	630

For the same simulation period, a total of 760 true positives i.e true data against its supplied label is indexed whereas it goes down with Hac and Neural Index alone. To calculate the CCR, the proposed algorithm divides the minimum attained count by self-attain count viz. the CCR for the proposed will be min (All count viz. 530)/(Self Count viz. 760)=0.6973. A graphical representation for the same is shown in Fig. 3.

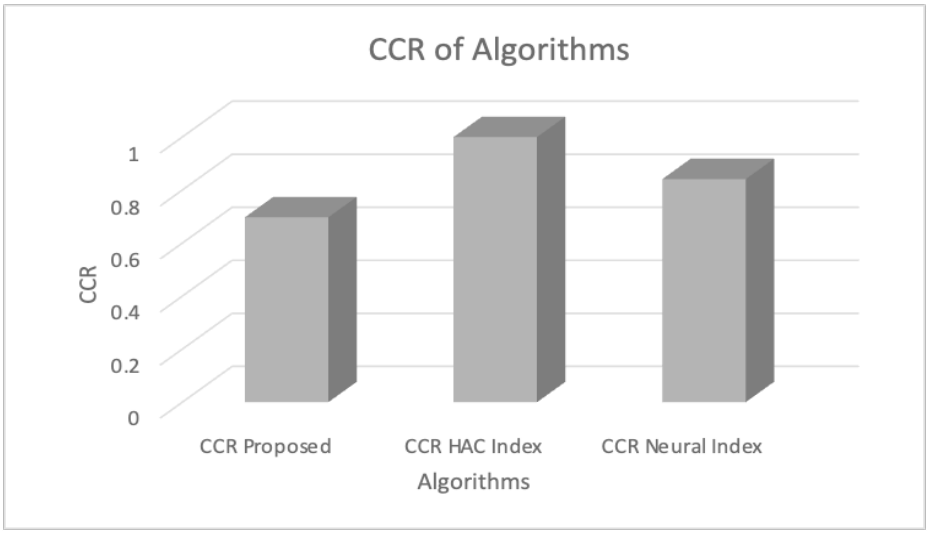

Fig. 3. CCR Comparison

Table 6. Comparison of state of art techniques for TPR

Technique	**TPR(%)**
Proposed Approach	93.81%
Zhang et al. 2022	92.03%
Wang et al. 2016	93%

6. Conclusion

This research provides a cloud data encryption and retrieval model based on Symmetric Searchable Encryption using Machine Learning. The proposed model is used to mitigate this risk and contribute to the data security solutions in cloud data storage and retrieval systems. The suggested model uses Artificial Neural Networks that improve data security in conjunction with an effective keyword ranking technique. Further, the authors have calculated the HAC index and the Neural index for the query or the documents, and the search index will be created based on these two factors. If the Search index of a document falls within the set boundaries, the document is approved; if the Search index of a document falls outside or below the stated limits, the document is refused. The comparative analysis of the proposed work is performed against multi-class SVM and Naïve Bayes. It was observed that even for 100 searches the proposed work exhibited high TPR along with low FPR in comparison to multi-class SVM and Naïve Bayes along with a minimal CCR of 0.6973. These outcomes contribute to the encryption of documents in the cloud with great security, as well as the retrieval of data based on a specific keyword that has been assigned. As a result, this study aids in the prevention of uncertainties such as cyber-attack plans and cloud storage system errors. In the future, an attempt has been made to improve the retrieval performance and enhance the data security model using the fuzzy model.

Acknowledgments

Acknowledgments have been done where ever applicable.

References

[1] Tari, Yi, Premarathne, Bertok, and Khalil. (2015) “Security and privacy in cloud computing: vision, trends, and challenges.” *IEEE Cloud Computing*, **2**(**2**): 30-38.

[2] Ratanghayra. (2017) “Review on Dynamic Multi-Keyword Ranked Search over encrypted mobile cloud data.” *IJNRD-International Journal of Novel Research and Development (IJNRD)*, **2**(**12**): 8-10.

[3] Hashizume, Rosado, Fernández-Medina, and Fernandez. (2013) “An analysis of security issues for cloud computing.” *Journal of internet services and applications***, 4**(**1**): 1-13.

[4] Zulifqar, Anayat, Kharal, (2021) “A Review of Data Security Challenges and their Solutions in Cloud Computing.” *International Journal of Information Engineering & Electronic Business*, **13(3)**: 32-41.

[5] Sun. (2019) “Privacy protection and data security in cloud computing: a survey, challenges, and solutions.” *IEEE Access*, **7**: 147420-147452.

[6] Hui Yin, Zheng Qin, Lu Omang, and Keqin Li. (2017) “A Query Privacy-Enhanced And Secure Search Scheme Over Encrypted Data In Cloud Computing.” *Journal of Computer and System Sciences*, 2 (**90)**: 14-27.

[7] Tahir, Ruj, Rahulamathavan, Rajarajan, and Glackin. (2017) “A new secure and lightweight searchable encryption scheme over encrypted cloud data.” *IEEE Transactions on Emerging Topics in Computing*, **7(4)**: 530-544.

[8] Huang, and Li. (2017) “An efficient public-key searchable encryption scheme secure against inside keyword guessing attacks.” *Information Sciences*, **403**: 1-14.

[9] Poh, Chin, Yau, Choo, and Mohamad. (2017) “Searchable symmetric encryption: designs and challenges.”*ACM Computing Surveys (CSUR)*, **50**(**3**): 1-37.

[10] Ilakiya, Vijithra, Kuppusamy, and Mahalakshmi. (2019) “Impact of Asymmetric Encryption in Cloud Computing: A Study.” *International Journal of Computer Sciences and Engineering*, **7(3)**: 32-43.

[11] Jiang, Cao, McCann, Yang, Liu, Wang, and Deng. (2019) “Privacy-preserving and efficient multi-keyword search over encrypted data on a blockchain.” *In 2019 IEEE International Conference on Blockchain (Blockchain*), IEEE: 405-410.

[12] Malhotra and Singh. (2019) “An Optimized Solution for Ranking Based On Data Complexity.*”* *International Journal of Innovative Technology and Exploring Engineering(IJITEE),* **8**(**11**): 41-49.

[13] Islam, Chaudhury, and Islam. (2019) “A simple and secured cryptography system of cloud computing.” *In 2019 IEEE Canadian Conference of Electrical and Computer Engineering (CCECE)*, IEEE: 1-3.

[14] Suneetha, Kishore, Singh, (2019) “A Security Model Using Artificial Neural Networks and Database Fragmentation in Cl Environment”, *International Journal of Recent Technology and Engineering (IJRTE)* **8**(**2**): 34-43.

[15] Nirmala, Muthurajkumar, and Subitha. (2021) “An Efficient Privacy-Preserving Ranked Keyword Search Method.” *In IOP Conference Series: Materials Science and Engineering*, **104**(**1**): 102-112.

[16] Tyagi. (2021) “Enhancing Security of Cloud Data through Encryption with AES and Fernet Algorithm through Convolutional-Neural-Networks (CNN).” *International Journal of Computer Networks and Applications*, **8**(**4**): 288-299.

[17] Sana, Li, Javaid, Liaqat, and Ali. (2021) “Enhanced Security in Cloud Computing Using Neural Network and Encryption.” *IEEE Access*, **9**: 145785-145799.

[18] Pulido-Gaytan, Tchernykh, Cortés-Mendoza, Babenko, Radchenko, Avetisyan, and Drozdov. (2021) “Privacy-preserving neural networks with Homomorphic encryption: Challenges and opportunities.” *Peer-to-Peer Networking and Applications*, **14(3)**: 1666-1691.

[19] Ma, Zhou, Qin, Xiang, Tan, and Cai, (2022) “A privacy-preserving content-based image retrieval method based on deep learning in cloud computing.” *Expert Systems with Applications***, 2(3):**117508.

[20] Zhang, Qiuyu, Minrui Fu, Yibo Huang, and Zhenyu Zhao (2022) "Encrypted Speech Retrieval Scheme Based on Multiuser Searchable Encryption in Cloud Storage." Security and Communication Networks.

[21] Wang, Haiyan, Yuan Li, Willy Susilo, Dung Hoang Duong, and Fucai Luo(2022) "A fast and flexible attribute-based searchable encryption scheme supporting multi-search mechanism in cloud computing." Computer Standards & Interfaces: **82,** 103-115.

[22] Wang, Q., He, M., Du, M., Chow, S.S., Lai, R.W. and Zou, Q., (2016) “Searchable encryption over feature-rich data”, *IEEE Transactions on Dependable and Secure Computing*, **15(3)**, 496-510.