*MATTER: International Journal of Science and Technology*
*ISSN 2454-5880*

*Pandey & Bathla, 2020*

*Volume 6 Issue 3, pp. 87-101.*

*Date of Publication: 30th December 2020*

*DOI-https://doi.org/10.20319/mijst.2020.63.87101*

*This paper can be cited as: Pandey, S., & Bathla, R. K., (2020). A Secure and Managed Cloud Storage System Using Encryption with Machine Learning Approach. MATTER: International Journal of Science and Technology, 6(3), 87-101.*

# A SECURE AND MANAGED CLOUD STORAGE SYSTEM USING ENCRYPTION WITH MACHINE LEARNING APPROACH

**Shweta Pandey**
*Research Scholar, Department of Computer Science & Application, Deshbhagat University Mandi, Gobindgarh, Punjab, India*
*shwetapandey.sm@yahoo.com*

**Dr. R. K. Bathla**
*HOD Computer Science & Application, Deshbhagat University Mandi, Gobindgarh, Punjab, India*
*hodcs@deshbhagatuniversity*

## Abstract

*Cloud storage platform is a promising architecture provided by the Cloud Service. However, the use of these services raises many doubts and concerns about the security, confidentiality, reliability and integrity of users' data and information. Because the cloud is based on a per-user payment model, it will take longer to retrieve the required document, which raises the financial burden, and hence affects the satisfaction level of cloud users. This is the major point, where presented research comes to play. The proposed work takes advantage of multi-layered neural network architecture for secure cloud storage system along with the involvement of encryption and similarity approaches such as Cosine Similarity algorithm while encryption check engrosses AES and DSA approaches. Simulation analysis offers a secure cloud platform using cloudsim* *simulator. Experimentation was performed against 700 text documents to evaluate the proposed work in terms of precision, recall, f-score and accuracy with an average accuracy of 92.48%. Simulation results had demonstrated that the designed algorithm proved to offer data storage in a cloud computing environment with high-end security. In future, the authors aim to involve some deep learning approaches to improve the text mining capabilities using cloud storage without challenging data security.*

**Keywords**

Cloud Computing, Cloud Storage, Security, Encryption, Similarity, Classification

## 1. Introduction

Cloud computing is an emerging technology that enables cloud user to share resources on-demand basis and pay for the used resources. Resources might be software, a single program, a platform, bandwidth, which is being utilized by the user through the internet (Gilad-Bachrach et al., 2016). Multiple requests are handled by the cloud service provider; therefore, the cloud must be highly scalable. Any cloud user can access data from the cloud though internet by using PC, tab, mobile or laptop as a platform (Fu, Wu, Guan, Sun, & Ren, 2016.) As cloud computing provides an efficient paradigm, therefore the security and privacy have become a major and significant concern of the cloud user as well as the cloud providers (Zhou, Varadharajan, & Hitchens, 2013).

One important reason is that cloud users rely on the cloud provider and it’s the duty of the cloud provider's to deliver data with high security (Li, Shi, & Zhang, 2017). To provide security, cryptography is one of the widely accepted techniques in industry and academia to resolve the security and privacy issues in cloud computing environments. In the last few years, many cryptography-based techniques have been used for the cloud data security some of the researchers have focused on secure storage (Mahmood, Huang, & Jaleel, 2017; Hur, Koo, Shin, & Kang, 2016; Deng, Li, Li, & Zhou, 2017), reliable computing (Amoon, 2016) and secure service usage (Jouini, & Rabai, 2019; Li, Gai, Qiu, Qiu, & Zhao, 2017). The cloud storage facility is known to be specific Sub-offer within cloud computing platform. Using this cloud storage is provided to the user to a distinct space rather the data is stored on the dedicated cloud providers (Cui, Deng, & Li, 2018). The providers provide data storage services over Internet users and others. The main requirement of today’s cloud storage providers is to store a large amount of data with minimum cost (Mahmood, Huang, & Jaleel, 2019). As the data is moved through radio means, therefore, security and privacy are also needed. To deal with this problem, secure cloud storage has been designed using cryptography approach (Yu, & Wang, 2017). The general strategy of storing data using the internet is shown in Figure 1.

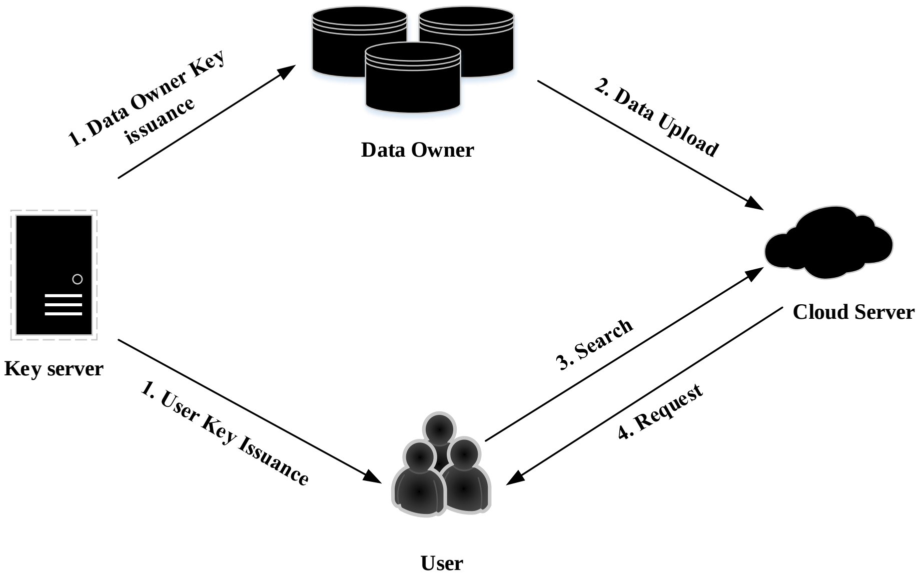

**Figure 1:** *General Structure of Uploading Data in Cloud Server*

This research has been performed to store data using encryption with a classification approach. The entire working process is discussed in section 3.

The rest of the paper is organized in the following way: Section 2 reviews and provided a survey of conventional methods. Section 3 presents the architecture of the proposed work with step by step description, Section 4 explains the computed results of the designed model and part 5 summarizes the entire work followed by the references used.

## 2. Related Work

Currently, most of the trade marketable processes are digitized. The data are of the greatest importance, so any damage or loss of the data can be a major failure for the data owner. Large organizations want to keep their data in a place where there is maximum security and required low cost. One of the most commonly used locations in the cloud where information is stored. Therefore, the foremost duty of cloud service providers is to improve security in the cloud. In the last couple of years, research has been done to enhance the cloud data security using cryptography approach in integration with other techniques (Venkatesh, & Eastaff, 2018). This section deals with the research work provided by the existing authors.

Aghili, (2019) has used DES as a security algorithm for cloud storage, which is being created as a duplicate of the original storage. The results indicated that good data accuracy has been provided against the attacker as the attacker has not to access confidential data. Khalique, Hussain, Alam, & Khan, (2020) have presented a new mechanism to provide security for the cloud data using the RSA approach with 2 prime numbers. Using this approach, data security has been increased to the desired security level based on the user's request or the importance of the data. Abdel-Kader, El-Sherif, & Rizk, (2020) have used cryptography technique twice in the designed model and after that save data into the cloud database. Initially, One Time Password (OTP) has been generated that provides authentication. This process also has not required extra processing time for the identification of the genuine user. Also, the plain text has been classified into two parts and each part is encrypted by a separate key. Using this approach, the complexity, as well as the data size, has been reduced.

Tyagi, Manoria, & Mishra, (2019b) have designed a cloud model using AES and RSA as an asymmetric and asymmetric algorithm. The confidentiality of the data along with the data security has been obtained by integrating these two techniques and the performance has been analyzed with AES, RSA and AES with RSA approach. Tyagi, Manoria, & Mishra, (2019a) have designed a framework that is being defended data based on classified data. The dedicated server has been selected using the fitness function of the cuckoo search algorithm. After the selection of an appropriate server, elliptic curve integrated encryption has been used to encrypt data towards the user side and then transmit data to the cloud service provider. Towards the service provider side, the data is encrypted using AES approach and then stored in the cloud database. At last, the system has provided better confidentiality of data with higher computing efficiency.

## 3. Proposed Work

The entire flow of the work is shown in Figure 2. The entire work is divided into two parts authentication and encryption. For authentication of the appropriate user, the login panel is designed that includes two fields (i) user ID and (ii) password. The password of a maximum of 8 characters (upper case and lowercase) length has been used. For the wrong password following algorithm is used.

$U_{ID} = Panel.Input(Document)$

$U\_pass = Panel.Input(Document)$;

$Captcha = Generate.System.Captca(8,1)$;

$Generated\ Code\ from\ System$

Captcha Code: In the age of digital technology and the Internet, each of us had to become an “advanced” user of the network. And for sure, almost everyone came across such an undesirable thing as captcha.

After the user authentication process, the next step is to apply, similarity measures that are Cosine similarity to check the similarity among the uploaded documents. Earlier, Gupta, Dutta & Kumar (2017) had also implemented Cosine similarity to identify frequent terms used in the text documents. The algorithm used in the current work is as follows:

**Algorithm: Cosine Similarity**

```
Input: Data→Raw data in which similarity needed
Output: Cos-sim→Similarity between data
Create an empty array to store similarity, Cos-sim = []
Sim-count = 0
For m = 1 → Length (Data)
Current_Data = Data (m)
    For n = m+1 →Length (Data)
Calculate the cos similarity using the given equation
        L= |Cos (Current_Data) - Cos (Data (n))|
Cos-sim[sim_count, 1] = Current_Data
Cos-sim[sim_count, 2]= Data(n)
Cos-sim[sim_count, 3]=L
Increment in array, Sim-count = Sim-count + 1
     End
End
Return: Cos_sim as an output in terms of similarity between data
End
```

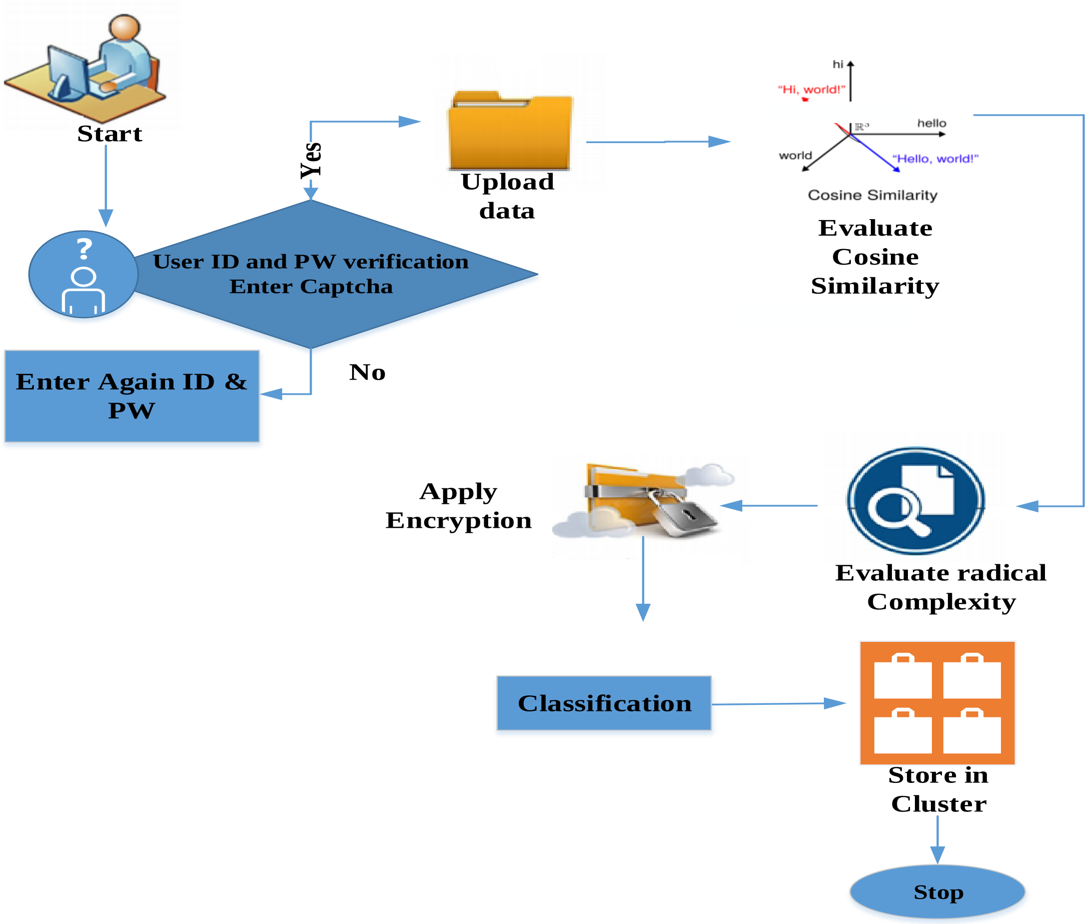

**Figure 2:** *Proposed Work*

After determining the similarity between the documents, the next step is to evaluate the radical complexity of the documents based on the level of complexity; the selection of an encryption algorithm has been done. Using radical complexity finder, a lot of execution time has been saved during the transmission of the document towards the encryption block. An appropriate selection of encryption technique saves time as well as reduces the chances of data loss. The radical complexity has been calculated based on two different conditions such as:

Apply

i. DSA algorithm, if $R\_complexity \leq Th_{complexity}$

ii. AES algorithm, if $R_{complexity} > Th_{complexity}$

The designed cloud security model has used AES or DSA as an encryption technique based on the examined $R\_complexity$ value.

The algorithm that is to be followed to determine complexity probability is written as below;

$Complexity\_prob\ = find\_c\_prob(stored\_$
$documents, uploaded\_document)$

$Int\ doc\_length\ = length(stored\_documents);$

```
$Int\ user\_length = length(uploaded\_document);$
$Cp = (doc/user);$
$Complexity\_probablity = cp * 100;$
$Return\ Complexity\_probablity$
End
```

The encrypted data is then passed to the neural network as a classification approach, which is used to select the appropriate space in the cloud server for the data storage by the cloud provider.

**Advanced Encryption Standard (AES):** It is a symmetric key encryption algorithm, each cypher having 128-bit size block. It makes sure that the encryption has been performed safely.

**Diffi-Hellman (DSA):** This is used to provide a secure key exchange between the cloud user and the cloud service providers.

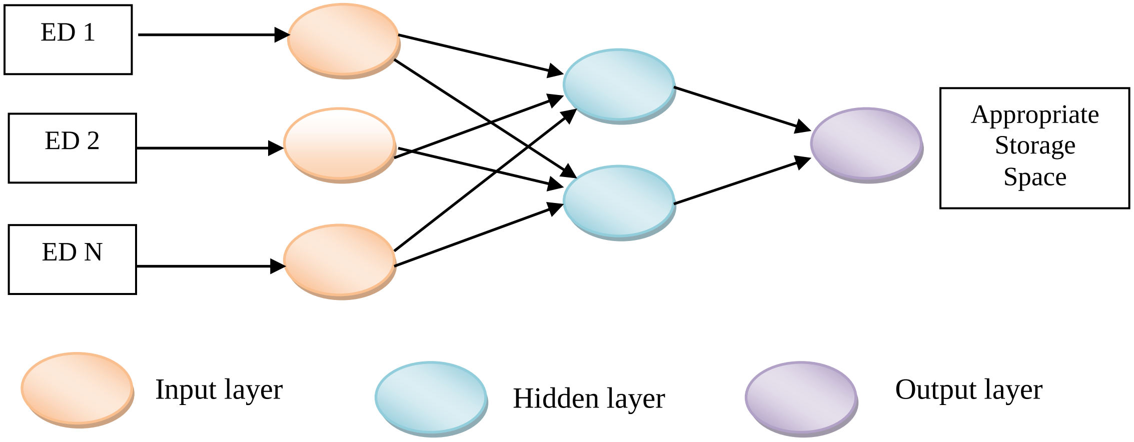

**Figure 3:** *General Structure of ANN*

The main contribution of this paper is providing security of cloud data using similarity with encryption and ANN as a machine learning approach. ANN is an artificial intelligence approach used in machine learning whose general architecture is illustrated in Figure 3. As the name "neuron" indicates that this is related to brain stimulation systems that aim to replicate what people have learned. Hassanzadeh, Nguyen, Karimi, & Chu, (2018) had adapted ANN architecture for text-based categorization of documents for evaluation across hospitals.

Here, the ED is an encrypted document. The neural networks consist of a secret layer (hidden layer) along with inputs and output layer. This is an excellent tool for a human programmer to find out many difficult problems or the problem that is too large to be taught to drive and recognize.

## 4. Results and Discussions

The results of the proposed secure cloud model have been identified using the parameters precision, recall, F-score and accuracy.

$$Precision = \frac{True_{selected}}{True_{selected}+False_{selected}} \quad (1)$$

$$Recall = \frac{True_{selected}}{True_{selected}+True_{left_{selected}}} \quad (2)$$

$$F-score = 2 * \frac{Precision*Recall}{Precision+Recall} \quad (3)$$

$$Acuuracy = \frac{True_{selected}+True_{rejectcted}}{True_{selected}+False_{selected}+True_{rejectcted}+False_{rejectcted}} \quad (4)$$

Where, $True_{selected}$→ It is the total count of features, which are selected during the classification process and similar to the classified results.

$True_{rejectcted}$→It is the count of the total feature which is matched with other categories during the classification process and it is opposite to the classified results.

$False_{selected}$→It is the total number of irrelevant feature set which are used for the classification and it should be minimum for good classification results.

$False_{rejectcted}$→It the total counts of the relevant feature which are considered as other class as compared to the classified output.

The examined results with different techniques are described below:

**Table 1:** *Precision*

| **Number of Uploaded Text Documents** | **Cosine similarity** | **ANN** |
|---|---|---|
| 100 | 0.716 | 0.742 |
| 200 | 0.697 | 0.7289 |
| 300 | 0.689 | 0.7068 |
| 400 | 0.671 | 0.6918 |
| 500 | 0.654 | 0.6879 |
| 600 | 0.648 | 0.6728 |
| 700 | 0.635 | 0.6575 |

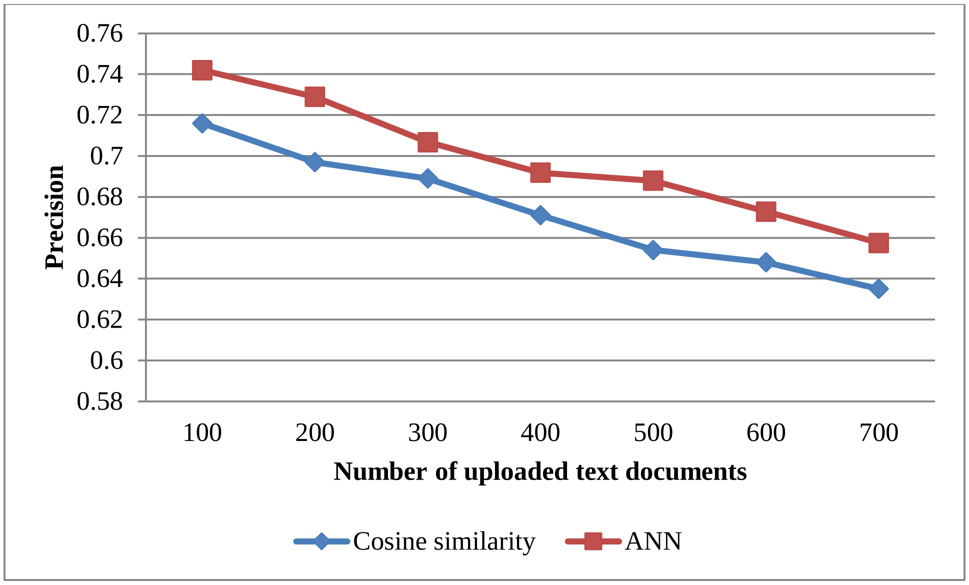

**Figure 4:** *Precision*

The precision values analyzed using cosine similarity and ANN as classification approach is listed in Table 1 with the graphical representation illustrated in Figure 4. The training of the encrypted data has been performed using the ANN technique. The results have been analyzed during the testing processes while the document is uploaded. From the figure, it is seen that the documents are uploaded with higher security (precision) value while using Ann approach compared to the similarity index approach.

**Table 2:** *Recall*

| Number of Uploaded Text Documents | Cosine similarity | ANN |
|---|---|---|
| 100 | 0.637 | 0.612 |
| 200 | 0.622 | 0.6057 |
| 300 | 0.613 | 0.5924 |
| 400 | 0.609 | 0.5817 |
| 500 | 0.597 | 0.5769 |
| 600 | 0.586 | 0.5647 |
| 700 | 0.579 | 0.5526 |

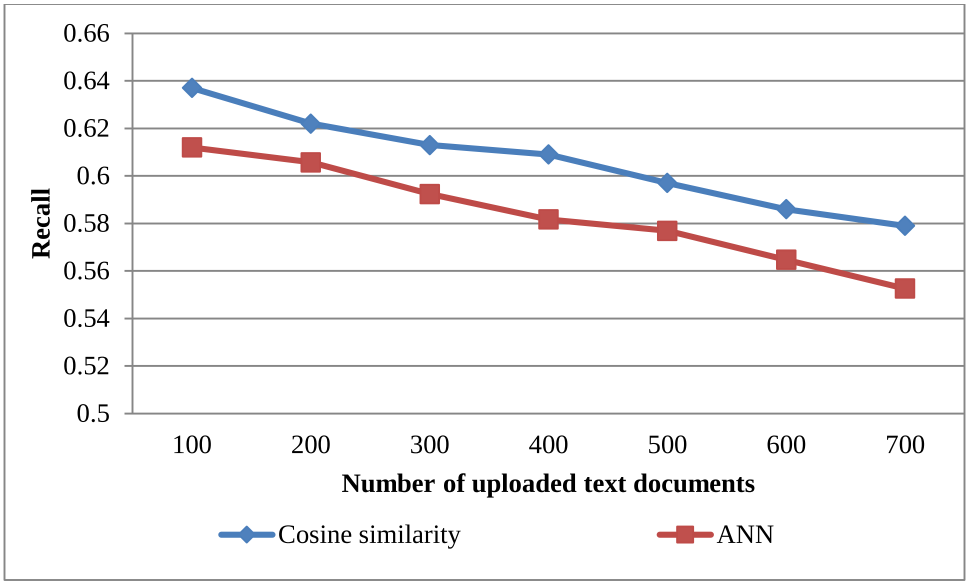

**Figure 5:** *Recall*

The recall values analyzed after applying similarity measures on the uploaded text and the ANN as a classification approach is listed in Table 2 with the graphical representation illustrated in Figure 5. The graph has shown that the recall rate using the ANN approach is less compared to the Cosine similarity index. This is possible only because the similarity approach along with encryption and ANN approach performed well to find the similarity between the uploaded text documents as well as the features that are used during the training time are exactly used by the classifier and hence reduce the falsely detected documents.

**Table 3:** *F-score*

| Number of Uploaded Text Documents | Cosine similarity | ANN |
|---|---|---|
| 100 | 0.674194 | 0.670759 |
| 200 | 0.657368 | 0.661614 |
| 300 | 0.648782 | 0.644563 |
| 400 | 0.638498 | 0.623512 |
| 500 | 0.624201 | 0.614743 |
| 600 | 0.615442 | 0.605562 |
| 700 | 0.59377 | 0.596298 |

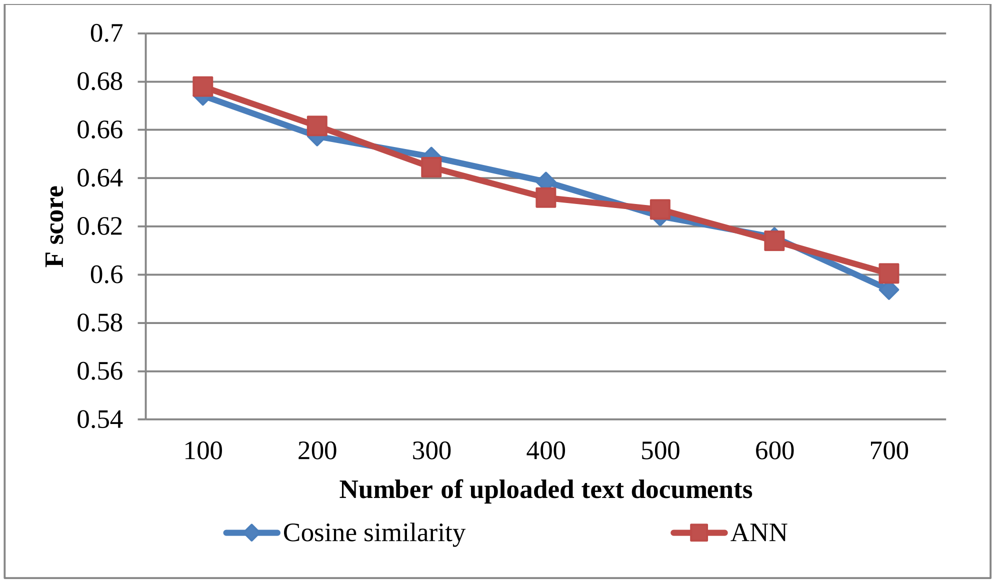

**Figure 6:** *F-score*

F score represents the average selection of truly matched features with classified output based on training as well as a testing feature set. From Figure 6, the F score analyzed using ANN approach is higher than that of the Cosine similarity technique.

**Table 4:** *Accuracy*

| Number of Uploaded Text Documents | Cosine similarity | ANN |
|---|---|---|
| 100 | 68.536 | 0.955 |
| 200 | 67.98 | 0.9422 |
| 300 | 66.12 | 0.938 |
| 400 | 65.37 | 0.926 |
| 500 | 64.47 | 0.917 |
| 600 | 63.68 | 0.904 |
| 700 | 62.92 | 0.892 |

The accuracy of the designed secure cloud system is shown in Table 4 and Figure 7. From the figure, it has been observed that with the increase in the number of uploaded text documents the accuracy of the designed system decreases. This is because with the increase in the number of text documents, the chances of relevant features as well as irrelevant features increases, which results in the reduction of detection accuracy.

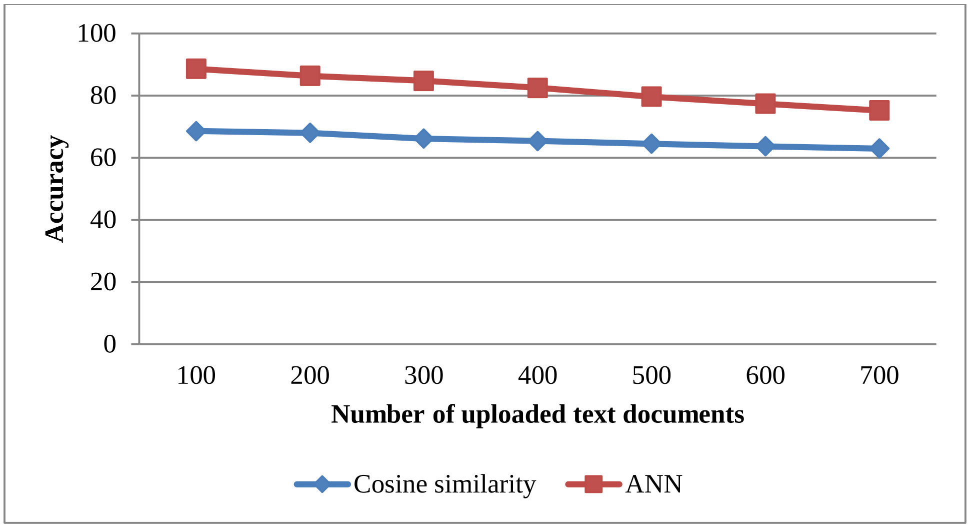

**Figure 7:** *Accuracy*

## 5. Conclusion

Presently, in the IT sector everything is moving to adopt new and trending technology. Most people around the world are switching to cloud storage for data storage so that they can access their data from anywhere and anytime. This paper has presented a neural network-based secure cloud storage architecture that takes advantage of encryption and cosine similarity approaches.

The designed model offers data security with the high-performance rate in terms of precision, recall, F-score and accuracy evaluated using large number of text documents ranging from 100 to 700. Based on the AES and DSA as an encryption approach, the document has been encrypted and the results have been computed in cloudsim as a simulator. The comparative analysis with and without involvement of ANN architecture is also performed to justify the essence of neural architecture in the proposed work. Overall, simulation analysis concluded that the proposed work proved to be successfully offering a secure data storage platform for cloud users with an average accuracy of 92.48%.

## REFERENCES

Abdel-Kader, R. F., El-Sherif, S. H., & Rizk, R. Y. (2020). Efficient two-stage cryptography scheme for secure distributed data storage in cloud computing. *International Journal of Electrical & Computer Engineering (2088-8708), 10*. https://doi.org/10.11591/ijece.v10i3.pp3295-3306

Aghili, H. (2019). Improving Security Using Blow Fish Algorithm on Deduplication Cloud Storage. In *Fundamental Research in Electrical Engineering* (pp. 723-731). Springer, Singapore. https://doi.org/10.1007/978-981-10-8672-4_54

Amoon, M. (2016). Adaptive framework for reliable cloud computing environment. IEEE Access, 4, 9469-9478. https://doi.org/10.1109/ACCESS.2016.2623633

Cui, H., Deng, R. H., & Li, Y. (2018). Attribute-based cloud storage with secure provenance over encrypted data. Future Generation Computer Systems, 79, 461-472. https://doi.org/10.1016/j.future.2017.10.010

Deng, Z., Li, K., Li, K., & Zhou, J. (2017). A multi-user searchable encryption scheme with keyword authorization in a cloud storage. Future Generation Computer Systems, 72, 208-218. https://doi.org/10.1016/j.future.2016.05.017

Fu, Z., Wu, X., Guan, C., Sun, X., & Ren, K. (2016). Toward efficient multi-keyword fuzzy search over encrypted outsourced data with accuracy improvement. IEEE Transactions on Information Forensics and Security, 11(12), 2706-2716. https://doi.org/10.1109/TIFS.2016.2596138

Gilad-Bachrach, R., Dowlin, N., Laine, K., Lauter, K., Naehrig, M., & Wernsing, J. (2016, June). Cryptonets: Applying neural networks to encrypted data with high throughput and accuracy. In International Conference on Machine Learning (pp. 201-210).

Gupta, V. K., Dutta, M., & Kumar, M. (2017, December). Frequent term based text document clustering using similarity measures: A novel approach. In *2017 Fourth International Conference on Image Information Processing (ICIIP)* (pp. 1-6). IEEE. https://doi.org/10.1109/ICIIP.2017.8313704

Hassanzadeh, H., Nguyen, A., Karimi, S., & Chu, K. (2018). Transferability of artificial neural networks for clinical document classification across hospitals: A case study on abnormality detection from radiology reports. *Journal of biomedical informatics*, *85*, 68-79. https://doi.org/10.1016/j.jbi.2018.07.017

Hur, J., Koo, D., Shin, Y., & Kang, K. (2016). Secure data deduplication with dynamic ownership management in cloud storage. IEEE Transactions on Knowledge and Data Engineering, 28(11), 3113-3125. https://doi.org/10.1109/TKDE.2016.2580139

Jouini, M., & Rabai, L. B. A. (2019). A security framework for secure cloud computing environments. In Cloud security: Concepts, methodologies, tools, and applications (pp. 249-263). IGI Global. https://doi.org/10.4018/978-1-5225-8176-5.ch011

Khalique, A., Hussain, I., Alam, M. A., & Khan, T. A. (2020). Scalable Security Based on Data Classification Using Generalized RSA in Cloud Storage. In *Proceedings of ICETIT 2019* (pp. 973-983). Springer, Cham. https://doi.org/10.1007/978-3-030-30577-2_86

Li, J., Shi, Y., & Zhang, Y. (2017). Searchable ciphertext-policy attribute-based encryption with revocation in cloud storage. International Journal of Communication Systems, 30(1), e2942. https://doi.org/10.1002/dac.2942

Li, Y., Gai, K., Qiu, L., Qiu, M., & Zhao, H. (2017). Intelligent cryptography approach for secure distributed big data storage in cloud computing. Information Sciences, 387, 103-115. https://doi.org/10.1016/j.ins.2016.09.005

Mahmood, G. S., Huang, D. J., & Jaleel, B. A. (2017). Data security protection in cloud using encryption and authentication. Journal of Computational and Theoretical Nanoscience, 14(4), 1801-1804. https://doi.org/10.1166/jctn.2017.6508

Mahmood, G. S., Huang, D. J., & Jaleel, B. A. (2019). A Secure Cloud Computing System by Using Encryption and Access Control Model. Journal of Information Processing Systems, 15(3).

Tyagi, M., Manoria, M., & Mishra, B. (2019a). A framework for data storage security with efficient computing in cloud. In *International Conference on Advanced Computing Networking and Informatics* (pp. 109-116). Springer, Singapore. https://doi.org/10.1007/978-981-13-2673-8_13

Tyagi, M., Manoria, M., & Mishra, B. (2019b). Analysis and Implementation of AES and RSA for cloud. *International Journal of Applied Engineering Research*, *14*(20), 3918-3923. https://doi.org/10.37622/IJAER/14.20.2019.3918-3923

Venkatesh, A., & Eastaff, M. S. (2018). A study of data storage security issues in cloud computing. *International Journal of Scientific Research in Computer Science, Engineering and Information Technology*, *8*.

Yu, J., & Wang, H. (2017). Strong key-exposure resilient auditing for secure cloud storage. *IEEE Transactions on Information Forensics and Security*, *12*(8), 1931-1940. https://doi.org/10.1109/TIFS.2017.2695449

Zhou, L., Varadharajan, V., & Hitchens, M. (2013). Achieving secure role-based access control on encrypted data in cloud storage. IEEE transactions on information forensics and security, 8(12), 1947-1960. https://doi.org/10.1109/TIFS.2013.2286456