International Journal of Computer (IJC)
ISSN 2307-4523 (Print & Online)
http://ijcjournal.org/

An Efficient Method to Enhance Health Care Big Data Security in Cloud Computing Using the Combination of Euclidean Neural Network And K-Medoids Based Twin Fish Cipher Cryptographic Algorithm

Arnav Goyal*

Neerja Modi School Jaipur Rajasthan, 302015
Email: arnavgoyal63774@gmail.com

Abstract

Big data is a phrase that refers to the large volumes of digital data that are being generated as a consequence of technology improvements in the health care industry, e-commerce, and research, among other fields. It is impossible to analyze Big Data using typical analytic tools since traditional data storage systems do not have the capacity to deal with such a large volume of data. Cloud computing has made it more easier for people to store and process data remotely in recent years. By distributing large data sets over a network of cloudlets, cloud computing can address the challenges of managing, storing, and analyzing this new breed of data It's possible for private data to be leaked when it is kept in the cloud, as users have no control over it. This paper proposes a framework for a secure data storage by using the K-medoids-based twin fish cipher cryptographic algorithm. We first normalize the data using the Filter splash Z normalization and then apply the Euclidean neural network to compute similarity, which ensures data correctness and reduces computational cost. As a result, the suggested encryption strategy is used to encrypt and decode the outsourced data, thereby protecting private information from being exposed. The whole experiment was conducted using health data from a large metropolis from the Kaggle database. Using the recommended encryption method, users will be able to maintain their privacy while saving time and money by storing their large amounts of data on the cloud.

Keywords: Big data; cloud storage; Filter splash Z normalization; Euclidean neural network; K-medoids based twin fish cipher algorithm.

1. Introduction

Due to exponential growth of digital information, researchers and scientists are faced with an ever-increasing volume of data in the exabyte range, making it more difficult to construct automated systems.

---

* Corresponding author.

Researchers, on the other hand, must invest a significant amount of time and effort in order to cope with and develop knowledge-based technologies for the future. EHRs (electronic health records) have undergone a tremendous transition, which has permanently transformed the global scenario in which health-care services and technology are used in the field of application. Technology developments in sensor-based, imaging, scanning, and other areas have heralded the end of the era of traditional instruments as a result of modern ICT (Information and Communication Technology).

Figure 1: Big data challenges

Clinical decision-making may be considerably enhanced while retaining the privacy and security of patient data when paired with modern intervention technologies. Use this technique to uncover hidden patterns and information in massive databases. It is envisaged that big data analytics, on the other hand, would play an important role in healthcare applications in a wide range of areas, such as fraud detection and treatment of illnesses and the identification of factors that contribute to healthcare expenditures. Finally, the goal is for the treatment to be effective and efficient so that patients may make their own choices in the future. Big data analytics is a unique IT-based technology that has the potential to revolutionize clinical decision-making in healthcare. Health care and big data security are causing a lot of people to worry. As a result, dealing with data that is both complex and huge is a challenge, and the development and preservation of patient-level data is the next level of concern. New algorithmic tools and standards are being developed in order to achieve such a string of demanding jobs. Medical big data analytics has the ability to increase patient treatment efficiency and efficacy by giving insights into data while preserving the greatest degree of confidentiality and privacy. Health care processes may be made more efficient by using this technology, such as flow of patients' information; total hospital stays; insurance data; and other cost-related elements. In order to lessen the global burden of illness, using big data analytics technologies, it is imperative to achieve remarkable results and concentrate on data

privacy and security. The concept is to build and create a safe and sustainable tool for a wide range of application domains, including healthcare, GIS (Geographic Information System), imaging, industry and banking. As a consequence, data analytics approaches must have a broad vision. a setting that has evolved technologically Knowledge discovery has been created as an exponential tool for data assessment because of the massive developments in electronic databases in terms of volume and complexity. To be clear, when it comes to privacy and security issues, "information discovery" is an explicit term that denotes a radical approach to uncovering patterns and knowledge that may be valuable to end users. The security and privacy issues that big data faces in the healthcare industry originate from the possibility of emerging risks occurring as a result of vulnerabilities and leaks in adaptive information systems. Despite this, researchers and scientists throughout the world have an enormous problem in dealing with the complementary challenges of security and privacy. Patient data records should be kept as private as possible, as we all know, since healthcare is such a delicate topic. HIV, AIDS, cancer, tuberculosis, and other severe illnesses are among the illnesses that patients do not want to disclose information about with any group that might jeopardies their fundamental social requirements and well-being. Consider this constant danger to patient privacy to be one of the most important issues to be considered in the context of patient privacy. Consequently, privacy and security are concerns that need to be addressed to safeguard patient information from cyber-attacks. Afterwards, we sought to cope with the way of big data analytics while protecting the privacy of healthcare records in order to facilitate future research. The K-Medoids-based twin fish cipher optimization technique is thus used to secure data encryption. The current objectives of this work were to,

1. To develop and implement a one-of-a-kind framework that can integrate enormous amounts of data with privacy and security concerns, as well as produce knowledgeably patterns for future decision making.

2. Prediction of the data similarity can be done by using the Euclidean neural network

3. For data encryption K-medoids based cipher twin fish cryptographic algorithm was designed

4. Then for obtaining the optimal key blue bumper key optimization algorithm was used.

The remaining section of the paper can be organized as follows, Section I shows an overview of data security impact and the paper's goal contributions; Section II shows an analysis of other existing technologies; section III illustrates the problem statement, Section IV shows the implementation of a novel methodology for the secure big data cloud storage; and Section V shows the methodology's effectiveness based on its findings. An overall summary was depicted in section VI .

2. Related Works

According to [1,] the author provides a novel technique for optimizing the selection of virtual machines (VMs) in cloud-IoT health services applications in order to properly handle a large volume of data in integrated industrial 4.0 settings. In order to support Industry 4.0 applications, massive amounts of data, such as sensor data, must be processed and assessed automatically. Among the model's stated benefits for improving healthcare system efficiency are shorter request execution times for stakeholders, reduced storage requirements for patient

huge data, and support for real-time data retrieval approaches, amongst others. [2] has suggested an AmIHMS (Ambient Intelligence Assisted Health Monitoring System) that uses Internet of Things devices. To gather the information Ami needs, wireless sensor networks (WSNs) are used. We will be able to manage the ever-increasing volume of health data, disseminate information in new and imaginative ways across health care networks, and make Big Data Analytics a commercially viable long-term endeavor with the help of the cloud. An essential part of the planned study is the real-time alerting of large data sets to student health information. The author of [3] gave a detailed review of the application of machine learning technology for big data analysis in the healthcare industry. A number of research challenges are also addressed, including the strengths and weaknesses of present approaches and the challenges that researchers face. Our study will assist healthcare professionals and government agencies keep on top of the latest breakthroughs in ML-based big data analytics for smart healthcare. In Music and video may be accessed and sent through mobile apps although this often limits the number of apps available. Despite the fact that the cloud saves data, the customer does not have a software storage notion in mind. The cloud and our mobile users were safeguarded when we limited our mobile access capacity. Watermarking was developed by [4] in order to protect data by verifying the cloud and its users' identity. The use of Reed–Solomon coding and water markup coding may reduce transmission errors. Adaptability of access control was strengthened by [5] in cryptographic systems, where check capabilities are crucial. This is due to the high cost and difficulty of computing the key and decrypting in ABE. There was no significant change in the level of performance from the customer or the authority. The computer function had to depend on a third party to get the easy answer and deal with the third party's confirmed findings. Processing and setup of cloud infrastructure necessitates the use of authentication and access control mechanisms. RBAC and context-aware RBACs were not recommended to the customer as viable choices for dynamic access control. Onto-ACM was used to solve the shortcomings of cloud computing [6]. Resource virtualization, global replication, and migration are used to assure the quality of the service provided. Although cloud storage data was beneficial for cloud customers, no consistent outcomes were seen. Using the trusted computer audit methodology provided by [7], batch verification was used to verify storage safety while also streamlining and reducing the expenses associated with the sample collection procedure. The results showed that it was a successful and efficient operation. innovative model of healthcare delivery that puts the patient at the center of all it does. As a consequence of this method, attribute-based encryption delivered transparent and adaptable results that differed from data found in dependable databases in terms of dependability.[8] Some security authorities have made attempts to make key management more manageable for their users because of the diverse data types. The Logistic equation, the Hyperchaotic equation, and DNA encoding are all used to build a security framework. [9]. The secret image is decrypted and turned into a share using a Lossless Computational Secret Picture Sharing (CSIS) technology, which is then broadcast to allow extensive storage on cloud-based servers. Hyperchaotic and DNA encryption methods must be used in order to enhance the overall security of the system. There are two steps in which PRN generated by the logistic equation are XORed with the image sequence in order to generate a true random number generator. [10] Healthcare-as-a-Service is the focus of this study, which introduces a new classifier based on a fuzzy rule set. Initial cluster creation and huge data retrieval/processing in cloud settings are crucial components of this method. This is followed by the creation of a fuzzy rule-based data classification decision-maker. Membership functions are used to draw conclusions and deduce inferences about data throughout the fuzzification and defuzzification stages of the data collection

process. Discusses the security problems and potential cures of cloud computing in [11]. Second, we created the MetaCloudDataStorage Architecture, which is meant to safeguard massive volumes of data in cloud computing settings... It is stated in [12] that the author uses a hybrid red deer-bird swarm technique (RD-BSA) that assures better convergence while also restricting the usage of control components throughout the solution generation process. In [13], research has been done to improve cloud platform stability, dependability, and security by analyzing the virtualization security protection management system from a virtualization security technology viewpoint. [14] studies and analyses privacy threats in modern social networks, as well as offers protection methods for users' personal data based on data mining techniques, in order to really assure that social network users' privacy is not unjustly misused in the age of big data. The data mining technique described in this research protects the user's identify and private information. Using a variety of privacy protection mechanisms for data publication and data mining, social networks may efficiently secure users' personal information and preserve their popularity. The author [15] presents a technique based on artificial intelligence and blockchain technology to enhance data security in smart cities. [16] aims to compare and analyze algorithms like as AES (Rijndael), Blowfish, and RSA. The combination of AES and blowfish encryption/decryption techniques has been shown in [17] to increase cloud data security by minimizing the chance of data theft. The hybrid approach uses AES-256 as the first layer, followed by blowfish as the second layer, as illustrated in the figure 1. The output of the first layer is first sent to the second layer, where it is then analyzed by the second layer. One may combine AES with conventional algorithms; however, the suggested way is more efficient. You may employ an encryption/decryption strategy to keep data in the cloud secure [18]. An inter-cloud data exchange architecture that is both easy and safe for users' personal information is the end result. An article by the author of [19] examines the features of a large number of encryption algorithms and creates a graph for each one. The graph neural network creates a vector representation of the algorithm graph once it is repeatedly incorporated into the graph. In [20], the authors examine the characteristics of a large variety of encryption algorithms and create a graph for each one they take into account. Another step follows, and a vector representation of an encryption algorithm's structure is generated.

3. Problem Statement

User data is protected by a layer of security provided by cloud providers. However, it is still not adequate since the confidentiality of data is often compromised. Various assaults, such as password guessing, man in the middle, insider threats, shoulder surfing, and phishing, are all forms of attacks. In the cloud, there are many threats to data protection:

Data misuse: In the cloud, there is the potential for data abuse, which is a concern for many enterprises. Data repositories must be protected immediately to prevent this possibility. To do this, one may employ authentication and limit access control for the cloud's data in order to accomplish this objective.

Data locality: It might be difficult to locate the actual location of data stored in the cloud since it is typically spread across many locations. There are also compliance difficulties and data protection regulations that come into play when data is transferred from one jurisdiction to another, which affects the cloud storage of data. Data storage rules and the precise location of the data storage server must be disclosed by the cloud service provider

to its consumers.

Integrity: System security and access controls must be implemented in order to protect the system. Only those who have been granted access to certain information should be able to do so. The integrity of data in a cloud environment must be maintained at all times to prevent data loss. Furthermore, only a small number of persons should have the authority to alter the data, in order to prevent a larger issue with access in the future.

Access: Data control and access rules are critical for long-term data security. They must be in place. In order to ensure that everyone has access to just the data they need, data owners are mandated to provide limited access to the data mart. It's possible to exert a great deal of control and data security by limiting and regulating who has access to the system.

Data confidentiality: It's possible that sensitive data will be kept on the cloud. In order to prevent data breaches and phishing attempts, additional layers of security must be put in place by both the service provider and the company. As a precaution, sensitive data should be protected with the greatest care.

Data breach: There have been reports of breaches inside the cloud. Cloud security may be breached by hackers, allowing them to access data that would otherwise be protected as secret by the enterprise. While a breach might be external, firms must pay special attention to monitoring staff behaviors in order to prevent any unwanted assaults on saved data.

Storage: Virtual data storage and retrieval is a common practice among companies. However, service providers must store data in physical infrastructures, making it susceptible to physical assaults.

These are just a few of the security concerns that cloud computing has to contend with. However, given today's level of technology resources, they are not insurmountable obstacles. Many efforts are being made to ensure that stored data is as secure as possible in order to comply with all applicable laws and regulations as well as internal compliance standards of the company. Cloud providers and customers must have mutual trust in order to ensure that the cloud provider is not a hostile invader and to preserve the coherence of data storage. Therefore, trust models and procedures must be built.

4. Proposed work

Figure 2 depicts the data securing framework and the overall implementation method, as well as the suggested cloud data security architecture. At this point, we've acquired 26 of the nation's biggest and most metropolitan cities' health data from Kaggle, which will be used to apply our proposed technique.

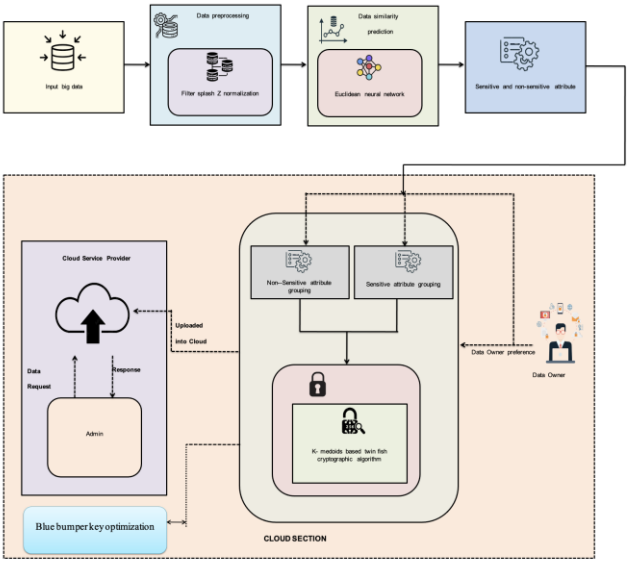

Figure 2: Schematic representation of the suggested methodology

a. Preprocessing

By converting the original data to a common scale, the filter splash Z-normalization technique may make modifications . Each raw data value is subtracted from the population mean of each regime, and the difference is divided by the data standard deviation to get the z-score, a dimensionless number. The formula for determining a raw value's standard score is presented as follows;

$$M\,(I^s) = \frac{I^s - \mu^s}{\sigma^s} \qquad (1)$$

I represent the raw data values for S, σ represents the standard deviation of the data population, and μ represents the population mean or the value of that feature. This dataset contains m scalar observations, and as a result, the standard deviation for the whole sample population is calculated as,

$$\sigma^s = \sqrt{\frac{\sum_{a=1}^{m}(I_a^s - \mu_a^s)^2}{m}}, \forall s \qquad (2)$$

and the population mean is:

$$\mu^s = \frac{1}{m}\sum_{a=1}^{m} i_a^s \qquad (3)$$

This method gives a linear transformation of the data. In accordance with the normalizing method,

$$G' = \left(\frac{G - minvalue\ of\ G}{max\ value\ of\ G - minvalue\ of\ G}\right) * (K - A) + A \qquad (4)$$

Where,

G' includes a flaw Data that has been normalized

Where G is the initial data range

& B is the pinted data

Z-score Mean and standard deviation are used to create normalized values or data ranges from unstructured data using the normalization process. Thus, using the equations supplied, it is possible to normalize unstructured data using the z-score parameter:

$$w_x' = \frac{w_x - \bar{R}}{std(R)} \quad (5)$$

Where, wx' is the normalized one values.

wx is the scaled value of the row R of ith column

$$std(R) = \sqrt{\frac{1}{(m-1)} \sum_{x=1}^{m} (w_x - \overline{R})^2} \quad (6)$$

$$\overline{ER} = \frac{1}{m} \sum_{x=1}^{m} w_x \; or \; mean \; value \quad (7)$$

Consider the following: Five rows of variables or columns, each starting with the variable 'm,' are shown in this table. Because of this, the z score method may be used to determine the normalized values in each column above. The standard deviation of a row is 0 if all of the values in that row are the same, as seen in the following example. The z-score, like Min-Max normalization, displays the range of potential values between 0 and 1. To generate numbers between -1 and 1, you may use Decimal Scaling techniques. It generates integers between -1 and 1. These are the outcomes of using decimal scales to represent data:

$$w^x = \frac{w}{10^y} \quad (8)$$

Where,

w is the range of values

y is the smallest integer Max(|wx|)<1

The proposed filter splash Z-normalization technique is given below with explanation;

$$H = \frac{(|S|) - (10^{m-1}) * (|G|)}{10^{m-1}} \quad (9)$$

Where,

M, is the total number of variables in S

G, is the initial variable of S

H, is the error free value with was in a range between 0 and 1

b. Similarity prediction

Euclidean networks are feed-forward models that receive a set of input vectors (raw data) and produce a second set of output vectors (HI) as the target data. This information is evaluated by neurons, the basic building blocks of the brain. The following code is responsible for managing the nodes in the hidden layer.

$$b = K_i(v + \sum_{a=1}^{m} t_a i_a) \quad (10)$$

Fixed real valued weights (t_a) are multiplied by the state variables (i_a), and then bias v is added to the results. The activation of neurons is controlled by the nonlinear activation function of neurons and nodes b. The generic network equation for the network structure is as follows:

$$u_{(r)} = K\left[i_{(r)}\right]$$

$$= K_o\left\{v + \sum_{l=1}^{ml} t_l k_l\left(v_l + \sum_{a=1}^{m} t_{al} i_{(r-a)}\right)\right\} \quad (11)$$

Using a sigmoid transfer function (k_l) in the hidden layer and a linear output transfer function (K_o) in the output layer, this two-layer feed forward network is utilized to fit the function. However, even if the network is trained with specific data and produces the expected results, estimates based on weight and bias values that are different from those used in training the network might lead to unwanted consequences. A network library (P) stores all the learned functions of several neural networks (M) that include inputs and outputs from various paths.

$$P_{\{a\}} = kMM_{\{a\}}, \quad a = 1,2 \ldots., mp \quad (12)$$

$$P_{\{a\}} = u_a^{MM}, \quad a = 1,2 \ldots., mp \quad (13)$$

In this case, mp denotes the number of previously trained functions that exist in the library. This library of new trajectories is made up of many distinct network functions that have been trained using various datasets. These functions provide comparable outputs, $u_{a:mp}^{MM}$ is used in their trajectory training, no matter what input data is given to the functions.

After several estimations for a single input, the library uses a single estimate for dimension reduction. According to Eq. 14, moving averages of all these potential network outputs are utilized to filter a final error in the next phases.

$$w_s = \frac{1}{d}\sum_{f=s}^{s+d}\frac{1}{mp}\sum_{o=1}^{mp} u_{of}^{MM} \quad (14)$$

where the mean of the d-moving average of the mp number of HIs is used to generate the random sequence w_s. Average windows are supplied as numeric duration matrices for following steps, so they contain both current parameters and any future neighbours that fall inside the window. The centered average equation may be stated as follows when the window is extended prior to the w_s,

$$w_s = \frac{1}{2d+1}\sum_{f=s-d}^{s+d}\frac{1}{mp}\sum_{o=1}^{mp} u_{of}^{MM} \quad (15)$$

The whole matrix of library estimates is produced by;

$$w_a = \begin{cases} \dfrac{1}{d+a}\displaystyle\sum_{f=1}^{d+a}\frac{1}{m}\sum_{o=1}^{m} u_{of}^{MM} & if\ a-d<0 \\ \dfrac{1}{d+(p-a+1)}\displaystyle\sum_{f=i}^{d+(l-a+1)}\frac{1}{m}\sum_{o=1}^{m} u_{of}^{MM} & if\ a+d>l \\ \dfrac{1}{2d+1}\displaystyle\sum_{f=s-d}^{s+d}\frac{1}{m}\sum_{o=1}^{m} u_{of}^{MM} & \text{if a} - \text{d} >= 0\ and\ if\ a + \text{d} <= l \end{cases}$$

(16)

where a is the estimations length.

After training Euclidean neural networks to develop a generalization of the input-output connection, all HIs of the training and test trajectories in the same dataset are calculated using the network library. A similarity model that measures how closely two trajectories resemble one other is what supervised similarity learning is all about. It's a close relative of this strategy, which defines learning as the gap between the trajectories of training and testing. Similarity learning ignores the fact that indiscernible must be identified in order to learn the best-fitting items.

An actual period may be compared to the test case by using the similarity estimate of test trajectories. A similarity measure is calculated by comparing the two paths. The distance between the two vectors used to describe this similarity is:

$$n_{(rs,rx)} = \sqrt{\sum_{a=1}^{m}(rx_a - rs_a)^2} \quad (17)$$

This includes the test trajectory (rx), the associated component (rs), and the test trajectory length (m). It is possible, however, that the best-matched training units may be found later on the curve. A step-by-step approach to moving the curve is therefore taken. The following equation is used to compute and store the pairwise distance for each step;

$$n_{(rs,rx)_{(o)}} = \sum_{o=n_{rx}}^{m_{rs}} \sqrt{\sum_{a=1}^{m_{rx}} (rx_a - rs_{a+o})^2} \quad (18)$$

where m_{rs} and m_{rx} denote the length of the training and testing trajectories, respectively. By finding the least pairwise distance value, the testing curve may be advanced to the end of a baseline case and the best-matching units can be determined.

$$GM_{(rx,rs)} = min\left(n_{(rx-rs)_1}, n_{(rx-rs)_2}, \dots\dots\dots n_{(rx-rs)_{mrs-mrx}}\right) \quad (19)$$

Using the training baseline, the best-matching characteristic is determined:

$$Pz_{rx,rs} = \arg\ find\ \ (n_{(rs,rx)_{(o)}} = Gm_{(rs,rx)}) \quad (20)$$

Where $o = 1,2 \dots\dots (m_{rs} - mrx)$

c. Data security

K medoids clustering may be used to classify c1 and c2 as two distinct groups of attriutes for the purpose of clarity. Assume that R is the total number of features and that Sx, cx, is the metric that compares the xth feature (the dimension of the input space). The distance between the two clusters is determined as follows for the R-th feature.

$$S_x(c_1, c_2) = \sqrt[p]{\frac{\sum \alpha\varepsilon c_1, \beta\varepsilon c_2 [s_x(y_x, z_x)]^p}{|c_1||c_2|}} \quad (21)$$

where the element's clustering function is denoted by the subscript x, the cardinality of cluster c is given by |c|, and p $\in \mathbb{R}$ is equal to 1. p is most often seen to be 2. The distance between the two qualities c1 and c2 is calculated as follows:

$$Q(c_1, c_2) = \sum_{x=1}^{R} [a_x(g_1, g)]^{\lambda} R_x(g_1, g_2) \quad (22)$$

Correlation exists between the inclusion of property x in the selection of c 1,c 2 It's now possible to group all of the non-sensitive and non-sensitive attributes into one category. Encryption of group attributes using the Twin Fish Cipher technique is therefore feasible, and the key length may be varied between 128 bits and 256 bits. When it comes to encrypting data, the cypher uses a 4-by-4 maximum distance separable matrix that is employed in the key schedule, as well as an algorithm called the pseudo-Hadamard transform. Consider the diagram in figure 3 to have a better understanding of the K medoids twin fish cipher.

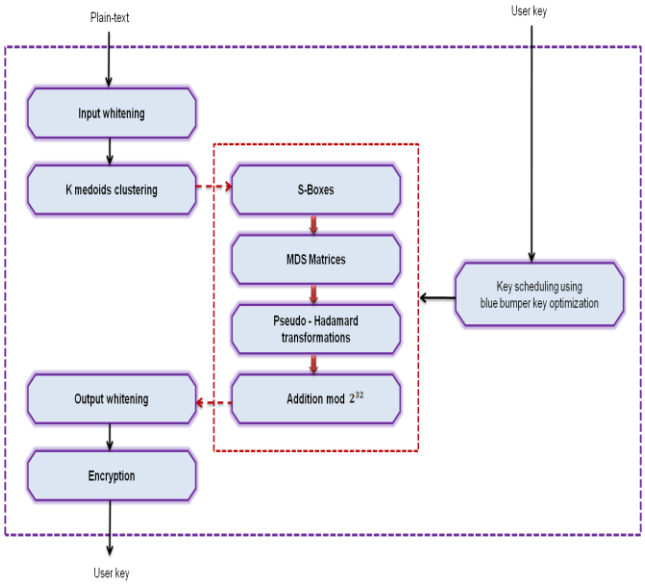

Figure 3: Process of cryptography

The input is whitened via XORing four keys on 128-bit plain text (divided into four 32-bit portions). Despite extensive cryptanalysis, the K medoids twin fish cypher technique can only be cracked after five rounds. Eight subkeys R0...R7 are used in the K medoids twin fish method to XOR input and output data. You might think of it as a kind of the XOR operation known as "input and output whitening." Component operations of the F-function include fixed left rotation by 8 bits, key dependent S-boxes (KDS), MDS, and a pseudo-Hadamard Transform (PHT). Additionally, there are four forms of key dependent S-boxes in addition to the g-function and the MDS matrix form. G-function is employed twice in the cypher construction, which results in duplicate information. There are 16 cycles in the twin fish algorithm for K medoids twins. P is broken into four 32-bit chunks and XORed with four subkeys R0, R1, R2, and R3, as seen in the following graphic, then sixteen rounds of iteration, before the outputs are Xored with four extra keys. Text encoded in 128 bits is the end result of this technique. As a consequence, this technique protects users in open spaces by encrypting their data.

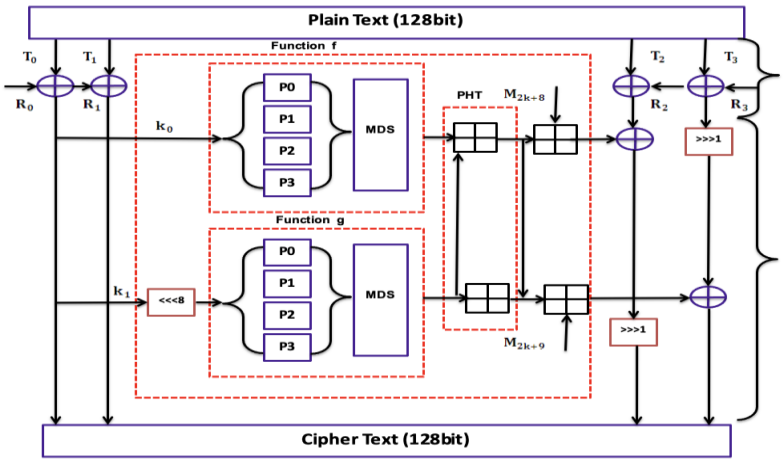

Figure 4: Data security

Thus, the owner encrypts the health record and keeps the encrypted health data in the cloud by employing this encryption approach. The user's identity must first be validated in order to see files. If the user is not a member of the associated group, the user cannot decrypt the health record using the secret key provided by the system. That is why data uploaded to the cloud are protected using the K medoids twin fish encryption mechanism.

d. Key optimization

The user must then choose an ideal key for the blue bumper key optimization once the encryption procedure has been completed. It is possible for bumper whales to find and encircle prey they have sensed. For the BBKO algorithm's search space, the current best candidate solution is the target prey or is close to the optimum. As a result, other search agents will work to improve their positions in order to take on the top search agent. This new technique allows us to tackle the following problems:

$$B = |A.R^*(i) - R(i)|.R(i+1) = R^*(i) - V.B \qquad (23)$$

The current iteration is I. V and A denotes the vector, R* and R stand for the position vectors of the best-so-far solution and the absolute value, respectively. The vectors C and V are created as a result of this technique.

$$V = 2cs - c \qquad (24)$$

As the number of repetitions increases (in both the exploration and exploitation phases), c decreases linearly from (0, 1).

Blue - net attacking method (exploitation phase)

Both of the following improved approaches for simulating blue bumper whale bubble-net activity may be used in mathematical modeling.:

.**1. Prey encircling mechanism**

By lowering the Eq's value of an (12). A also restricts V's range of motion. Since an is lowered from two to zero over time, V may be considered a random integer in the range [a, a]. To find a new location for a search agent, set random values for A in [1, 1]. The new location may be discovered anywhere in this range.

2. Spiral updating position

Spiral equations between the locations of the prey and the whale are used to replicate the helix-shaped movement of blue bumper whales:

$$R(i+1) = B'.e^{d\int}.\cos(2\Pi \int) + R^*(i) \qquad (25)$$

where B =$|R^*(i) - R(i)|$ and denotes the distance of the b^{th} whale to the prey (best solution obtained so far), To determine the form of the spiral, d is a constant that must be multiplied by an element-by-element

multiplication of a random integer in the range [1, 1]. Observe that blue bumper whales swim around their victim in a decreasing circle and in a spiral-shaped manner at the same time. We assume that there is a 50% chance that either the shrinking encircling mechanism or the spiral model will be used to update the whales' positions throughout optimization. The following is the mathematical model:

$$R(i+1) = \begin{cases} R^*(i) - V.B & if\ S < 0.5 \\ B'.e^{d\int}.\cos(2\Pi \int) + R^*(i) & ifS \geq 0.5 \end{cases} \quad (26)$$

where R is an integer drawn at random from the range [0,1]. In addition to using bubble nets, blue bumper whales forage for meals on their own.

3. Search for prey (exploration phase)

To find prey, the same technique based on the vector variant may be used (exploration). However, blue bumper whales seek in a random manner, taking into account the other whales' whereabouts. A with random values bigger than 1 or less than zero is used to make the search agent travel away from a reference whale. For comparison's sake, in the exploration phase, we don't use the best search agent yet; instead, we use a randomly selected search agent. Because of this mechanism and |V| > 1, the WOA algorithm may do a global search. How it works in math is as follows:

$$B = |A.R_{rand} - R| \quad (27)$$

$$B = |A.R_{rand} - R| \quad (28)$$

$$R(i+1) = R_{rand} - V.B \quad (29)$$

R_{rand} is a random position vector (a random whale) taken from the current population, where R_{rand} equals R rand. If a random agent is chosen or a solution is found, the search agents' locations are updated at each new iteration. For both exploration and exploitation, the a parameter is reduced from 2 to 0. Search agents' positions are updated when |V| is more than 1 and less than 1, in which case the best solution is selected. It's time to use the BBO to find the finest possible n-key combination in the shortest amount of time.

After then, the data is encrypted and stored in a third-party cloud provider. The service provider first checks the user's identity before allowing access to the files. The service provider receives a request from the user to get the files after the authentication has been validated. Final step: The owner provides a decryption key for usage by the user. This approach is incredibly effective and secure since the user can only access the data once they have been authorized, verified, and decrypted.

Algorithm 1: KMBTFCA_BBKO

Input: Classified attributes

Output: Encrypted attributes

```
Key Expand (key byte[4*SK],word k[Sm+1,Sa], Sa, Sk, Sr)
Begin
Set the encryption algorithm=”twinfish”
Crypt.put_CryptAlgorithm(“twinfish”)
Crypt.put_Cipher Mode (“ada”);
j=0
while (i<SK)
K[j]=word [key[4*j+3], key[4*j+2], key[4*j+1], key[4*j]
j=J+1
// Depending on the padding technique {add bytes}
// block size encryption process
j=SK
while(J<Sa*(Sa+1))
word temp=K[j-1]
if(j mod Sk=0)
temp=(SubByte(Rot Word(temp))Xor
Mcon[j/Sk]))
end if
crypt.put_Padding Scheme(0);
K[j]=k[j-SK] xor temp
```

```
If(j mod SK =3)

// Apply the new approach “ShiftRow” transformation

crypt.put_EncodingMode(“hex”);

Word temp 1 [4] [4]

For (a=3;a>=0;a-1)

Temp1 [a] [m]=K[Sk+a]

ECB mode does not use an IV

ShiftRow (temp 1);

System.out.println(enc Str);

//now decrypt:

String decStr=crypt.decryptStringENC(“HEART BEAT ABNORMAL.””);

System.out.println(decStr);

end if

j=j+1

end while

end
```

---

5. Performance analysis

In this part, we conduct experiments for the purpose of assessing performance. The suggested method is implemented in the MATLAB environment. Securing public and individual health is made easier using the technique described here. In addition, the proposed cryptographic technique is used to encrypt this procedure. A comparison of the proposed approach's efficacy with an existing technique is made in this section. The suggested system's performance may be evaluated using the evaluation metrics of encryption and decryption time, clustering accuracy, precision, recall, and memory utilization.

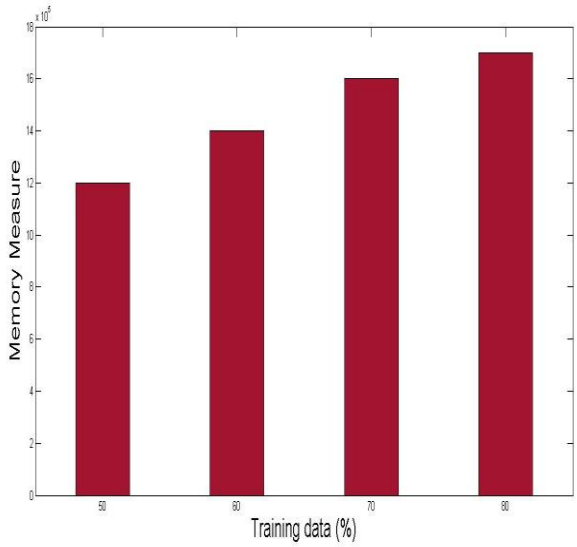

Figure 5: Training data Vs. Memory measure

Memory is defined as the amount of space used to store data in the cloud. Figure 5 shows that 74.1 bits of memory are required for data of 50 kilobytes, 75.51 bits are required for data of 60 kilobytes, 76.2 bits are required for data of 70 kilobytes, and 76.910 bits of memory are required for data of 80 kilobytes.

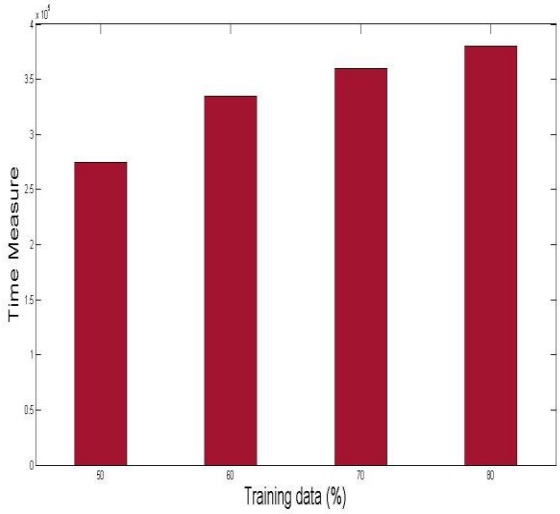

Figure 6: Training data Vs. Time measure

As shown in Figure 6, data with a size of 50 kb takes 264,850 ms to process, followed by 60 kb at 312,450 ms, 70 kb at 336,000 ms, and an 80 kb file at 364987 ms, all in descending order of processing time.

To prove the efficiency of the suggested mechanism it can be compared with the existing approaches [20],

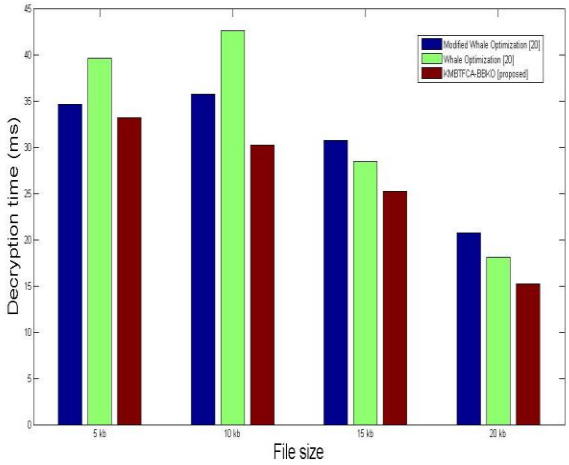

Figure 7: File size Vs. Decryption time

Decryption timings for various file sizes are displayed in Figure 7, as well as a comparison of the recommended method's performance. To put it another way, the findings are in line with those of other prominent algorithms, such as Whale Optimization and modified Whale Optimization. It has been found that a proposed technique outperforms other methods in terms of performance.

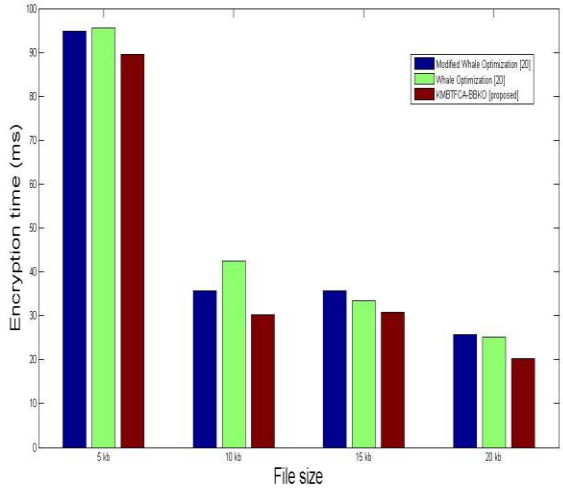

Figure 8: File size Vs. Encryption time

When it comes to encrypting large files, the efficiency of the proposed and present approaches is shown in Figure 8 by comparing the encryption times for different file sizes. Using approaches like whale optimization and modified whale optimization, the results are reviewed and contrasted. Results showed that the suggested methodology outperformed than currently used approaches in terms of encryption execution performance.

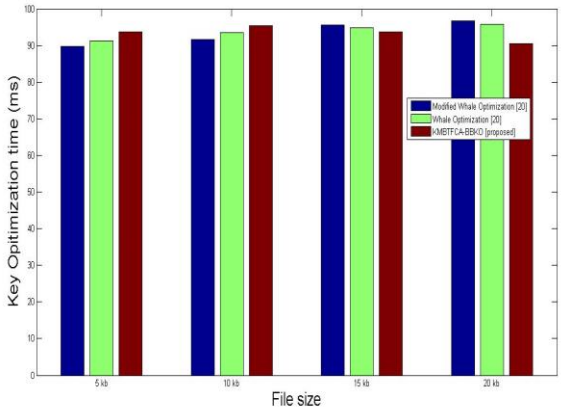

Figure 9: File size Vs. Key Optimization time

Figure 9 compares the proposed and current methods' critical optimization times. The proposed method has a lower key optimization time than the current techniques.

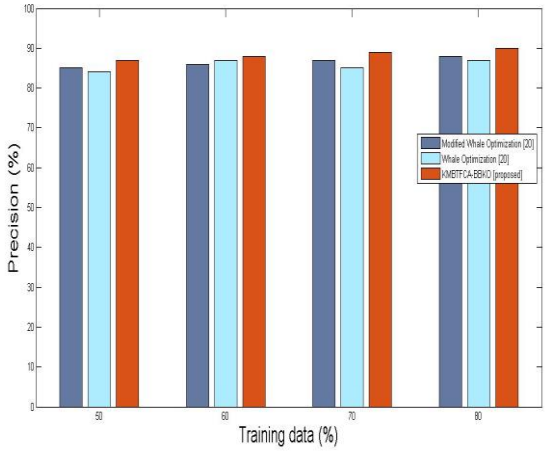

Figure 10: Training data Vs. Precision

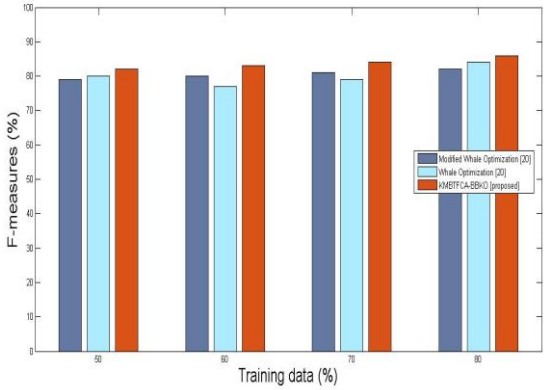

Figure 11: Training data Vs. F-measure

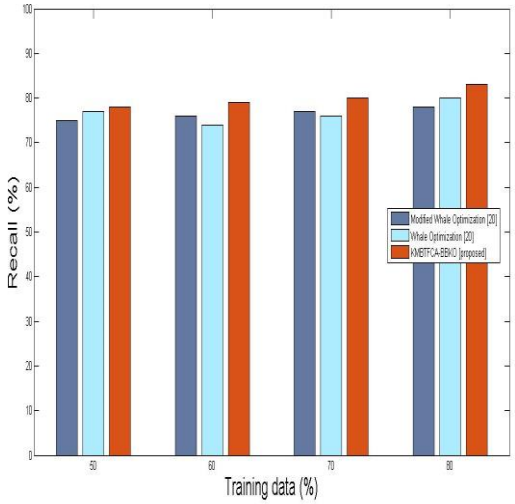

Figure 12: Training data Vs. Recall

A plot of the F-measures for the training data set is shown in Figures 10,11,12. The precision, recall, and F-measure of our proposed prediction-based secure data storage study are used to evaluate its performance. A better fit for securing and predicting E-health data storage and prediction may be inferred from the outcomes of our research. The accuracy values for the training data are 84.16, 87.0, 88.5,91.0, and 92, respectively. Recall values of 82, 85.0, 88.5, and 89.1 were found in the training dataset. F-measures values for the training dataset are discovered by our recommended secure data storage to be 79.1, 81.1, 82.4, and 80.3 of f-measures values respectively.

From the result obtained it was revealed that the suggested mechanism expresses satisfied results over secure cloud storage when compared to other existing mechanisms.

6. Conclusion

In the age of information security in the cloud, the most important issue to concentrate on is security. In this post, we discuss how to protect and forecast E-health data stored in the cloud using the KMBTFCA BBKO encryption method, which we demonstrate in this video. The primary purpose of our article in cloud computing is to increase the security of data. Using the Euclidean neural network classification technique, we are able to split qualities into multiples of data in the first step of our research. For the second step, using the K-medoids-based twin fish cipher encryption technique, it is possible to group and encrypt the categorized data sets. The encryption procedure necessitates the use of an optimum key. In order to do this, we developed an algorithm known as the blue bumper optimization method . In the cloud server, the encrypted data is thus kept safe and sound. Cloud servers will ask for verification of the user's identity whenever the user tries to get data from one of them. A request will be sent to the server for processing and response if all of the authentications are successful. There is no doubt that the user decrypts the file using the decryption key that the owner gave. This method is very secure since the data can only be delivered to the user after they have been authorized, confirmed, and decrypted. Each of these aspects of the proposed technique is approximated, as is the overall

performance of the approach. For cloud computing security concerns, our suggested encryption method provides improved solutions while keeping high accuracy, speed, and memory usage at the same time.

References

[1] M. Elhoseny, A. Abdelaziz, A. S. Salama, A. M. Riad, K. Muhammad, and A. K. Sangaiah, "A hybrid model of internet of things and cloud computing to manage big data in health services applications," *Future generation computer systems*, vol. 86, pp. 1383-1394, 2018.

[2] L. Hong-Tan, K. Cui-hua, B. Muthu, and C. Sivaparthipan, "Big data and ambient intelligence in IoT-based wireless student health monitoring system," *Aggression and Violent Behavior*, p. 101601, 2021.

[3] W. Li, Y. Chai, F. Khan, S. R. U. Jan, S. Verma, V. G. Menon, *et al.*, "A comprehensive survey on machine learning-based big data analytics for IoT-enabled smart healthcare system," *Mobile Networks and Applications*, vol. 26, pp. 234-252, 2021.

[4] H. Wang, S. Wu, M. Chen, and W. Wang, "Security protection between users and the mobile media cloud," *IEEE Communications Magazine*, vol. 52, pp. 73-79, 2014.

[5] S. S. Kumar, S. Prasad, M. Parimala, and G. M. Someswar, "Scalable and Secure Sharing of Personal Health records in Cloud Computing Using Attribute Based Encryption," *COMPUSOFT: An International Journal of Advanced Computer Technology*, vol. 5, 2016.

[6] C. Choi, J. Choi, and P. Kim, "Ontology-based access control model for security policy reasoning in cloud computing," *The Journal of Supercomputing*, vol. 67, pp. 711-722, 2014.

[7] L. Wei, H. Zhu, Z. Cao, X. Dong, W. Jia, Y. Chen, *et al.*, "Security and privacy for storage and computation in cloud computing," *Information sciences*, vol. 258, pp. 371-386, 2014.

[8] S. Belguith, N. Kaaniche, M. Laurent, A. Jemai, and R. Attia, "Phoabe: Securely outsourcing multi-authority attribute based encryption with policy hidden for cloud assisted iot," *Computer Networks*, vol. 133, pp. 141-156, 2018.

[9] P. Sarosh, S. A. Parah, G. M. Bhat, and K. Muhammad, "A security management framework for big data in smart healthcare," *Big Data Research*, vol. 25, p. 100225, 2021.

[10] A. Jindal, A. Dua, N. Kumar, A. K. Das, A. V. Vasilakos, and J. J. Rodrigues, "Providing healthcare-as-a-service using fuzzy rule based big data analytics in cloud computing," *IEEE Journal of Biomedical and Health informatics*, vol. 22, pp. 1605-1618, 2018.

[11] G. Manogaran, C. Thota, and M. V. Kumar, "MetaCloudDataStorage architecture for big data security in cloud computing," *Procedia Computer Science*, vol. 87, pp. 128-133, 2016.

[12] B. Balashunmugaraja and T. Ganeshbabu, "Privacy preservation of cloud data in business application enabled by multi-objective red deer-bird swarm algorithm," *Knowledge-Based Systems*, vol. 236, p. 107748, 2022.

[13] Z. Li, G. Liu, Y. Dang, Z. Shang, and N. Lin, "Research on New Virtualization Security Protection Management System Based on Cloud Platform," in *Journal of Physics: Conference Series*, 2022, p. 012010.

[14] J. Du and Y. Pi, "Research on Privacy Protection Technology of Mobile Social Network Based on Data Mining under Big Data," *Security and Communication Networks*, vol. 2022, 2022.

[15] A. S. Rajawat, P. Bedi, S. Goyal, R. N. Shaw, A. Ghosh, and S. Aggarwal, "AI and Blockchain for Healthcare Data Security in Smart Cities," in *AI and IoT for Smart City Applications*, ed: Springer, 2022, pp. 185-198.

[16] R. S. Cordova, R. L. R. Maata, A. S. Halibas, and R. Al-Azawi, "Comparative analysis on the performance of selected security algorithms in cloud computing," in *2017 International Conference on Electrical and Computing Technologies and Applications (ICECTA)*, 2017, pp. 1-4.

[17] M. N. Ul Haq and N. Kumar, "A novel data classification-based scheme for cloud data security using various cryptographic algorithms," *International Review of Applied Sciences and Engineering*, 2021.

[18] Z. Kartit, A. Azougaghe, H. Kamal Idrissi, M. E. Marraki, M. Hedabou, M. Belkasmi, *et al.*, "Applying encryption algorithm for data security in cloud storage," in *International Symposium on Ubiquitous Networking*, 2015, pp. 141-154.

[19] X. Li, Y. Chang, G. Ye, X. Gong, and Z. Tang, "GENDA: A Graph Embedded Network Based Detection Approach on encryption algorithm of binary program," *Journal of Information Security and Applications*, vol. 65, p. 103088, 2022.

[20] S. V. Karuppiah and G. Gurunathan, "Secured storage and disease prediction of E-health data in cloud," *Journal of Ambient Intelligence and Humanized Computing*, vol. 12, pp. 6295-6306, 2021.