INDIAN JOURNAL OF SCIENCE AND TECHNOLOGY

RESEARCH ARTICLE

Two Level Privacy Preserving Model for Association Rule Mining in Cloud Using Deep-SVM Method

Kanthimathinathan Mangayarkkarasi[1]*

1 Associate Professor, Department of Computer Science, D.G.G.A. College For Women, Mayiladuthurai, Tamil Nadu, India

OPEN ACCESS

Received: 22-03-2023

Accepted: 07-12-2023

Published: 05-01-2024

Citation: Mangayarkkarasi K (2024) Two Level Privacy Preserving Model for Association Rule Mining in Cloud Using Deep-SVM Method. Indian Journal of Science and Technology 17(1): 1-15. https://doi.org/10.17485/IJST/v17i1.663

* **Corresponding author**.

kanthimangai@gmail.com

Funding: None

Competing Interests: None

Published By Indian Society for Education and Environment (iSee)

ISSN
Print: 0974-6846
Electronic: 0974-5645

Abstract

Background/Objectives: It is becoming more common for data owners to outsource data mining tasks and storage to cloud service providers as a result of the rising costs of maintaining IT infrastructures for large-scale data mining. This trend, however, also raises security concerns about unauthorized breaches of data confidentiality and outcome integrity**. Methods:** This research considers this scenario in which cloud user can encrypt their data and store it to cloud environment. In order to perform mining operation, the user needs to outsource the task to cloud servers. Then, the cloud server performs the mining task on the encrypted data and share the encrypted association rule to the cloud user. Yet, existing single cloud server systems have privacy leakage issues since their work focuses on either database privacy or item privacy. To remedy this gap in the literature, this study maintains both database privacy and item privacy during the frequent itemset mining process. For item privacy, it first describes a universal safe multiplication protocol with a single cloud server. We build the inner product rules, comparison rules, frequent itemset protocol, and final association rule mining process that is secure against privacy leaking on top of this multiplication protocol. During this Association Rule Mining (ARM) operation, it provides two level of protection to data privacy. This model is designed with distributed Elgamal cryptosystem and sub-protocols for item and database privacy along with Deep learning-based Support Vector Machine (SVM) for secure rule generation**. Findings:** The proposed method is named as two-level privacy preserving method association with Deep SVM model (2-level:D-SVM), provides guaranteed solutions to the confidentiality of the outsourced cloud data and minimizes the user interaction during association rule mining task. Here, data on breast cancer and heart disease are used, and the effectiveness of the proposed model is demonstrated by comparison to existing models. According to the study, at 25000 transactions, the proposed 2-level:D-SVM model stands for 52% and 50% more efficient than Parallel Processing (PP) and Privacy-preserving Collaborative Model Learning (PCML) techniques in terms of computing cost. Additionally, the proposed model performs 34%, 22%, 6%, and 4% better in terms of execution time

than the PP, Apriori, Eclat, and FP-growth techniques, respectively. **Novelty:** The proposed method is built on a set of well-constructed 2-level secure computation techniques that not only maintains confidentiality of data and query confidentiality, but additionally allows the data owner to operate offline throughout data mining. When compared with previous attempts, this technique provides a higher degree of privacy, in addition, lowers the computation cost for data owners.

Keywords: Association Rule mining; Cloud; Data mining as a service; Deep learning; SVM

1 Introduction

Cloud computing is a shared architecture which is based on distributed processing, parallel computing and grid computing [1]. Cloud computing is performing distribution of incoming tasks to the resource pool which is received from different computers worldwide. Also, it makes sure all the application systems are having enough power, memory and software services as per their need. So, there are mass storage in the cloud and all these data are valuable information. Data mining is the technique to pick up the data and discover this useful knowledge for making decision from the mass number of database.

Association Rule mining (ARM) has widely applied in many fields for data mining applications such as business [2], healthcare service [3], and website analysis [4]. To provide better rule mining, the mass amount of database needs to be analysed, so the cloud can give enough support for making strong computing and power to the ARM applications [5]. In outsourced association rule mining, the cloud server is responsible to make association rule from the uploaded database.

Usually, cloud owner, lacking storage, data, computational resources and expertise stores their data to cloud and the cloud service provider perform the mining task in this outsourcing service. This is called as data mining- as- a-service, it makes the benefits in business intelligence. But this service is having serious privacy issue that the service provider can access the data owner secret data for providing the mining result. Furthermore, converting the database to k-support anonymity or k-privacy is too expensive. If the data owner is having resources to make this service, then it is possible to make ARM [6].

On Medical area, Patients' privacy may be risked if the hosting cloud has access to sensitive patient data that has been uploaded to the cloud by the individual healthcare providers. This may also result in a breach of industry-specific rules, such as the Health Insurance Portability and Accountability Act for healthcare providers in the United States. As a result, current research has focused on privacy-preserving in cloud association rule mining.

In this research model, let us consider, the data owner does not have large storage space for their data and selects the cloud storage for their transaction databases. Also, a new transaction database is also to be added to the cloud continuously. Our proposed model consists of the two-privacy requirement for association rule mining on encrypted transactional database. It may give the encrypted association rules for the encrypted query to the data owner using Elgamal cryptosystem.

Our research maintains both database privacy and item privacy during the frequent itemset mining process. For item privacy, we first describe a universal safe multiplication protocol with a single cloud server. We build the inner product rules, comparison rules, frequent itemset protocol, and final association rule mining process that is secure against privacy leaking on top of this multiplication protocol. Then, Deep learning-based Support Vector Machine (SVM) is implemented for secure rule generation. This

method will provide accuracy and efficient mining result at high level security range. It encrypts the data and upload to cloud server to further protect it from unwanted malicious cloud server. Our proposed system model consists of cloud server, data owner, assessor and rule miner for performing high level security in outsourced association rule mining procedures.

The rest of this research is organized as follows: In Section 2 the literature on outsourced association rule mining techniques is discussed. Section 3 consists of our proposed system model for high level security for frequent pattern mining protocol and the association rule using Deep-SVM procedures. Section 4 provides the performance of our proposed model and security analysis results. The last section of Section 5 concludes our work.

2 Literature Survey

S. Priyadarsini et al.[6] proposed ARM as a method for extracting information features particularly when using the Apriori algorithm for ARM. To make the Apriori formula more appropriate for parallel computing, they updated it. This method's primary drawback is the duration of time needed to hold a large number of candidates sets with frequent item sets, low minimum support, or huge item sets.

[7] They employed an association rule mining technique in the present research to protect the encrypted data stored in the cloud. In order to mine associations between rules, they employed the ECLAT method, and to encrypt data, they used the AES algorithm. In general, Transaction Id sets, also known as tidsets, are used to determine a dataset's Support value and to prevent the production of subsets that are not present in the prefix tree. The Intermediate Tidsets produced by the Eclat method use additional memory.

The FP-Growth method is used in this article[8] to mine and analyze computer large data. An enhanced FP-Growth method is developed to address the existing FP-Growth technique's weak extraction efficiency in environments with vast amounts of data. FP Tree, however, is more complicated and challenging to construct. When the database is big, the method might not fit in the shared memory.

In this[9] study, a data mining platform based on the Hadoop distributed file system was developed, and then the K-means algorithm was enhanced with the concept of max-min distance. This parallel processing can be described as a class of techniques that enable the system to attain concurrently data-processing tasks to increase the computational speed of a computer system, but it also increases the cost of computers because more hardware is required.

This[10] research article examines how data may be dynamically mined in real time for pattern recognition in a secure cloud computing environment utilizing a mix of decision tree and Random Forest through a secure Application Programming Interface (API). Random forest's important limitation is that a huge number of trees could make the process slow and ineffective for real-time predictions. In general, these algorithms are quick to train but slower to generate predictions once trained.

In[11], defined the first privacy outsourced method for measuring privacy on data. They proposed the model, in which the database is partitioned horizontally and vertically to create privacy using Private-Set Intersection (PSI). Despite this, PSI approaches keep depending on strong cryptographic assumptions that, in certain situations, may be impossible. In[12], proposes a unique double encryption and Transaction Splitter approach to alter the database to optimize the data utility and confidentiality tradeoff in the preparation phase. This Parallel Processing (PP) approaches are having higher run time due to parallel processing on Data owner and Cloud.

Wong et al.[13] presented a model to resist background attack in rule mining procedure, it performs one to many mapping to randomly convert the database transactions. The fundamental idea behind this technology is the paillier cryptosystem, which adds noise to the transaction database in order to provide anonymity. Paillier Cryptosystem key creation is a little more difficult.

First outsourced privacy preserving ARM technique with semantic security is proposed by Lai et al.[14], by which predicate encryption is used for data privacy. But, the efficiency of this scheme is unpredictable in practice. Furthermore, working with smaller activities may only minimally increase data access time, while working with a larger number of larger tasks dramatically increases data access time. Then, the Elgamal homomorphic encryption algorithm-based privacy preserving ARM algorithm is proposed in Yi et al.[15]. In this approach, data owner can perform encryption on their data and miner can outsource the cloud sever for computations. Furthermore, this protocol requires 'n' servers to execute homomorphic computations, resulting in increased communication among all cloud servers.

[16] proposed employing Table Partitioning Techniques to provide quick access to data. The cost of storage and memory will rise as a result of this strategy. Users may require more disk space and RAM to store and access partitions and indexes, as well as pay more fees or update their data to take use of advanced partitioning capabilities or choices.

The proposed GHLBO algorithm detects DDoS attacks quickly and efficiently in this research[17]. This improved approach is useful in training DSA to detect attacks efficiently. However, no overhead analysis for optimization is performed in this research.

This study[18] introduces AroSheb_Jo, a hybrid One-time Password generation technique for IoT data, and gives an evaluation of security of that approach. On the other hand, there is an essential link between the complexity of the design of this system. The complexity of this framework increases as researchers include more layers of security mechanisms.

2.1 Challenges

The previous study proposed a frequent itemset mining method on outsourcing encrypted cloud data that preserves confidentiality. They suggested three distinct privacy level procedures throughout all of their literary work. To begin, just the cloud transaction database is encrypted, while the miner's query is in plain text. This protocol is highly efficient; however, it does not secure the privacy of the inquiry. Second, although the miner's query is protected or partially shielded, the mining result is public. Third, the computation costs of data owners are relatively high when using time-consuming homomorphic cryptosystems. In this study, we primarily explore a case in which a higher level of privacy is necessary. In our case, both the initial data outsourced by the data owners and the mining result for the miner must be kept secret by the cloud server. Furthermore, we examine database and item privacy in our system to improve system privacy. Furthermore, we emphasize that frequent itemset mining is the basis of association rule mining. Mining frequent itemsets alone will not provide the strong association rule, which is required to discover the link between itemsets.

In this research, the above-mentioned problem is focused, and privacy preserving association rule mining is proposed in cloud environment. Most of the previous solutions are shared the mined association rules to the other third parties. So, the main difference between our proposed model is the generated rule should be secured and privacy of the query is maintained till the end user with the highest accuracy.

Our proposed model consists of the two-privacy requirement for association rule mining on encrypted transactional database. It may give the encrypted association rules for the encrypted query to the data owner using Elgamal cryptosystem. Second, Deep learning plays a vital role in learning procedure from complication functions. It may consist multiple hidden layers for deep learning on pattern[19]. Also, SVM is the supervised learning algorithm for pattern classification[20]. In this model, the stacked SVM based deep learning procedure is used for association rule mining generation for better performance. When compared with previous outsourcing association rule mining technique, our proposed scheme reduces data storage, reduce execution time and improves the privacy on the data based on cryptography methods. This method will provide accuracy and efficient mining result at high level security range. It encrypts the data and upload to cloud server to further protect it from unwanted malicious cloud server.

2.2 Our Contributions

We propose a 2-level privacy association rule mining methodology for the cloud architecture utilizing machine learning methods in this research. This study makes five contributions, which are as follows:

- This research provides two levels of privacy association rule mining on encrypted data using distinct keys. The method we propose allows distinct data owners to outsource their data to the cloud server for secured storage as well as processing using multiple encryption keys.
- We create a set of cryptographic sections for 2 level secure association rule mining based on the Elgamal cryptosystem, which serves as the system's foundation.
- During frequent item mining, the proposed approach creates a novel protocol for both item set and database privacy.
- We build a stacked SVM-based deep learning framework for an association rule method using the cryptographic blocks given.
- We demonstrate that our approach can provide a better level of privacy than most recent research[12,21,22]. Furthermore, we comprehensively demonstrate the security of our approach using multiple datasets.

3 Preliminaries

3.1 ElGamal Cryptography

The Elgamal security model contains key generation[23], Encryption and Decryption procedures.

Key Generation: if the security key of k given to the key generator, it produces a multiplicative cyclic group G of prime number with order q . Then, it selects the private key x randomly from $Z_q^* = \{1,2,\ldots,q-1\}$ and calculates a public key $y = g^x$. This computed public key is known to all, while the private key of x is kept as secret one.

Encryption: An input message to this step is taken from $m \in G$, and the public key y is given, then it selects an integer randomly from Z_q^* and produces the cipher text $C = E(m) = (A,)$, where $A = g^r$ and $B = m.y^r$.

Decryption: The input Cipher text (A,B), and the private key x is given, then the plain text is produced by $m = D(c) = B/A^x$.

This Elgamal decryption protocol is correct because

$\frac{B}{A^x} = m.\frac{y^r}{(g^r)^x} = m.\frac{(g^x)^r}{g^{rx}} = m.$

This cryptography method is also having two important privacy properties as follows.

Homomorphic Property- If the encryption of the two messages m_1 *and* m_2 is given, then the encryption of these messages is denoted as $E(m_1) = (A_1,B_1) = \left(g^{r1}, m_1 y^{r1}\right)$ *and* $E(m_2) = (A_2,B_2) = \left(g^{r2}, m_2 y^{r2}\right)$., the the multiplication of these two messages is represented as $E(m_1 m_2) = (A_1A_2, B_1B_2)$. This multiplication property is also correct because of,

$$\frac{B_1B_2}{(A_1A_2)^x} = \frac{m_1 y^{r1} m_2 y^{r2}}{(g^{r1}g^{r2})^x} = m_1 m_2.$$

From this property, we might say that $E(m^n) = (A^n, B^n)$ for all positive integer n. We also denote that $E(m_1)E(m_2) = (A_1A_2, B_1B_2)$ *and* $E(m^n) = (A^n, B^n)$

Re-encryption Property – If the encryption of a message m is given as $E(m) = (A,B)$, Then another encryption of the message is calculated, represented as $RE(E(m)) = (A', B') = (Ag^{r'}, By^{r'})$, where r' is chosen randomly from z_q^*. This property is also correct because

$$\frac{By^{r'}}{\left(Ag^{r'}\right)^x} = \frac{Bg^{xr'}}{\left(A^x g^{r'x}\right)} = \frac{B}{A^x} = m$$

4 Methodology

Figure 1 demonstrates the architectural representation of the disease prediction model. The experimentation is analyzed with datasets on Kaggle. The breast cancer data consist of 518 data under 2 different conditions. The validation and test sets consist of 300 data.

The proposed disease prediction model includes three phases, namely, Frequent item based on item privacy, Frequent item based on database privacy and Association rule mining Deep-SVM. Initially, the input data is subjected to pre-processing and from the attained pre-processed data, the frequent item-based rule generation is carried out. The main role of the proposed disease prediction model trusts on the generation of fine-tuned association rules. For that, four protocols are used with the help of Elgamal cryptosystem and its namely Multiplication protocol, inner product protocol, comparison protocol and frequent itemset mining protocol. After item-based rule generation, the optimal frequency rule based on database privacy is generated. After the database-based association rule selection, classification takes place using hybrid classification with SVM and Depp learning model. From the classification outputs, it can be predicted whether the disease exists or not. The Deep-SVM-based frequent rule generation model is exploited, by which the accurate prediction can be attained.

Data privacy is a concept of guaranteeing the correct use of personal data by providing individuals with decision over how their data is accessed, utilized, or shared. The encryption technique is used to keep items secure. The cloud server contains both real and faulty cloud data. The frequent item is then created from an authentic cloud server utilizing a private comparison protocol. With the Deep-SVM model, both item and dataset privacy are protected in this hybrid secured approach.

When compared with all existing model, our proposed model consists of fine-tuned association rules based on both item privacy and database privacy. So, the security will be high in this model. Also, machine learning based Deep- SVM is introduced to improve the classification efficiency with lower error rate. The comparison of the proposed model with state of art methods is executed in terms of execution time and computational cost.

4.1 Proposed Model

Our proposed model is designed for cloud service with user and cloud service provider. This cloud service provider is responsible for secure storage and providing association rule to the cloud user. The user can store their privacy data to the cloud. To ensure the secure data outsourcing, the Elgamal cryptosystem is used. First, the user can generate the public and private key pair by Elgamal cryptosystem. After that, they encrypt their data with generated public key and uploaded it to CSP for storage.

In our proposed system, the secured association rule mining on cloud server is performed. For privacy, we define five parties on the proposed model named as: Key Generator (KG), Assessor, Cloud Service Provider (CSP), Data Owner (DO) and Rule Miner (RM). The Overall proposed model is described in Figure 1.

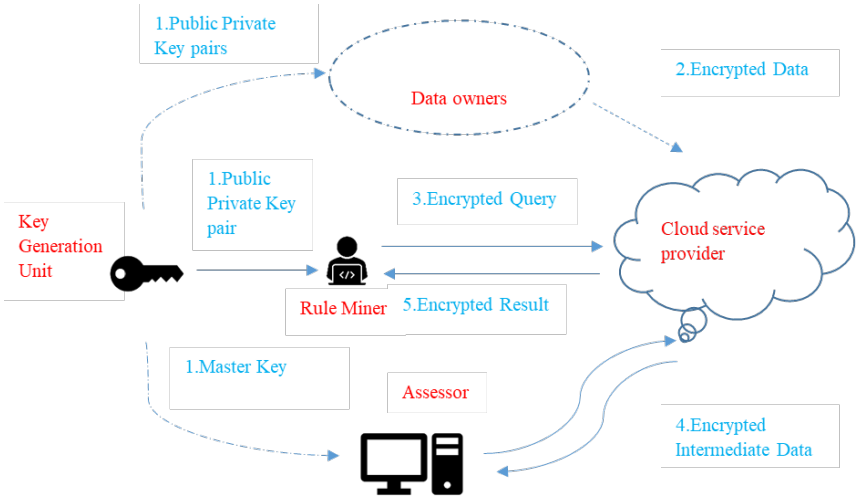

Fig 1. System Model

1. **Key Generator:** This KG is treated as trusted center for both private and public key generation for all users in this system.
2. **Data Owners:** Generally, the Data owner can encrypt their secret data by their public key and upload to CSP.
3. **Cloud Service Provider:** This CSP is having huge storage space for storing and managing the data from all the DOs. Also, it is capable to perform the basic computation over the data. In the proposed model, CSP is performing association rule mining with the help of Rule Miner and Assessor.
4. **Assessor:** The Assessor provides the computation of rule mining in our model. It has the key pair of all user for Elgamal Cryptosystem. This rule has been mined with the help of CSP in our system.
5. **Rule Miner:** In our proposed model, this rule miner provides association rule mining service. DO can also be a rule miner. The aim of this DO is to generate the strong and frequent association rule for their transaction query. To generate the association rule, they can send the encrypted query to the CSP, then it can perform the mining service with the help of assessor and send the frequent association rule in encrypted form. This encrypted rule can also be decrypted by the DO itself.

4.2 Frequent item based on item privacy

In this section, the frequent itemsets are generated based on item privacy. For that, four protocols are used with the help of Elgamal cryptosystem and its namely Multiplication protocol, inner product protocol, comparison protocol and frequent itemset mining protocol.

- **Multiplication protocol**

Let us consider RM contains data as bit $x \in \{0,1\}$ and also the DO consists of data as bit $y \in \{0,1\}$. Then, CSP is responsible for privacy multiplication result of $x.y$. This multiplication is the result of AND operation between x and y. Then the garbled circuit is combined with the AND gate with Elgamal encryption. This multiplication protocol is described as below.

- **RM initialization:** At first, RM generated three cipher texts C_1, C_2, C_3 for message 0 and one cipher text C_4 for message 1 with the support of the public key pk of Elgamal cryptography.

$[0] = \{C_1, C_2, C_3\}, [1] = C_4,$

And select the random number from below series

$$k_a^0, k_a^1, k_b^0, k_b^1 \in Z_{N^2}$$

Select a random secret value from $r_A \in Z_N$ and public hash function from : $\{0,1\}^* \rightarrow \{0,1\}^N$. Then RM is computed

$$\left\{ C_i^{'} = C_i \, X \, r_A^N \mod N^2 \right\}_{i \in \{1,2,3,4\}.}$$

- **RM to DO:** Here, RM chooses k_a^x , $x \in \{0,1\}$ and calculates

$$\{T_i^{'} = k_a^x \oplus T_i\}_{i \in \{1,2,3,4\}}$$

And then share k_b^0, k_b^1 , $\left\{T_i^{'}, C_i^{'}\right\}_{i=\{1,2,3,4\}}$ to the DO.

- **DO to CSP:** Here, DO selects k_b^y , $y \in \{0,1\}$ and calculates

$$\{\, k_b^y \oplus T_i^{'}\}_{i \in \{1,2,3,4\}}$$

And then shuffle the result randomly to get the result $\left\{T_i^{''}\right\}_{i \in \{1,2,3,4\}}$. After that share the result to CSP.

- CSP to RM. A random number R is chosen from $R \in Z_N$ and four cipher texts are generated and denoted as below

$$[R] = C_{R,1,}\ C_{R,2}, C_{R,3}, C_{R,4,}$$

And calculates

$$\left\{\begin{array}{ll} C_i^{''} = C_i^{'} X\ C_{R,i} & mod\ N^2 \\ h_i = H\left(C_i^{''}\right) & \\ \bar{T_i} = T_i^{''} X\ C_{R,i} & mod\ N^2 \end{array}\right\}_{i \in \{1,2,3,4\}}$$

Then CSP sends the $\bar{T_i}$ to the DM.

- **DM to CSP:** RM calculates U_i from the received $\bar{T_i}$ and upload it to the CSP.

$$\left\{U_i = \bar{T_i} X\ r_A^N\ \ mod\ N^2\right\}_{i \in \{1,2,3,4\}}$$

- **At CSP:** Then, the CSP computes

$$\{u_i = H(U_i)\}_{i \in \{1,2,3,4\}}$$

Then compare $\{u_i\}$ *and* $\{h_i\}$ to find the duplicate data $u_k, k \in [1,4]$, finally calculates

$$W = U_k X\ g^{-R}\ mod\ N^2$$

This W is an output of multiplication protocol.

- **Inner Product protocol**

If the CSP is having transaction record $[y] = ([y_1], [y_2], \ldots [y_l])$then the CSP send the mining query in cipher text format as, $[x] = ([x_1], [x_2], \ldots [x_l])$.

To calculate the inner product of *x and y*:

- First CSP calculates the multiplication protocol for *x and y* in order to obtain the cipher text $w_i = [x_i X\ y_i]$,where $i = \{1,2,\ldots l\}$.
- Second the Elgamal cryptosystem is used to calculate the inner product operation.

$$v = [x.y]$$

- **Comparison Protocol**

In association rule mining, the comparison operation is performed between $supp(q)$ and $_{min}$, $conf(X \rightarrow Y)$ and $conf_{min}$. Our proposed comparison protocol is discussed below.

Let consider the CSP is having two cipher texts $[s]$, $[t]$ to obtain the comparison result of s and t without performing decryption.

• **CSP to RM.** In CSP the following operations are computed.

$$[f] = [s-t] = [s]X\ [t]^{N-1}$$

Where a, b is selected randomly $a, b, \in Z_N$, $c \in \{0,1\}$ and satisfying $a \gg b$ and $B(a) < B(N)/4$. After that it is computed by below expression.

$$[\alpha] = [(-1)^c\ a(2f+1)+b] = [f]^2\ X\ [1]^{a\,X\,(N-1)^c}\ X\,C_b$$

Finally, it is sent to RM.

• **RM to CSP.** In RM, the decryption operation is performed with the key of sk to get α.

When $B(\alpha) < \frac{B(N)}{2}$, RM assign $u = 1$

Otherwise,$u = 0$

Encryption operation is performed by RM on u to obtain $[u]$ with the help of public key pk and the result sent to CSP.

• At CSP The following operation is performed

$$when\ c = 0,\ the\ output\ of\ CSP\ B = [u]$$

$$othrewise\ output\ B = [1-u] = [1]X\ [u]^{N-1}$$

- **Frequent itemset mining protocol**

Let us consider the transaction record $T = \{t_1, \ldots t_m\}$,where $t_i = \{t_{i1}, \ldots t_{il}\}$, l represents the total length of the dataset and m represents the number of records in the database. Then the RM encrypts the query q for mining with the $supp_{min}$ and send it to CSP. Finally, the CSP checks q is frequent or not with $supp(q)$.

• **RM to CSP.** In RM, the mining query $q = (q_1, \ldots q_l)$ has been encrypted by Elgamal cryptosystem and share the query to CSP with minimum support value $supp_{min}$.

• **CSP to RM.** Inner product is performed here to calculate the $v_i = [q.t_i]\ \ for\ i \in \{1,2,\ldots,m\}$ and get

$$A = \{v_1 \ldots v_m\}$$

n dummy cipher texts are computed in CSP with the messages belong to $[0,l]$ and concatenate the value with inner product cipher text to get updated A value

$$A = \{v_1 \ldots v_m, v_{m+1}, \ldots v_{m+n}\}$$

Then CSP randomly select $d_i \in Z_N$, where $i \in 1, m+n$ and calculates,

$$[w_i] = [d_i\ (v_i - \|q\|_l)] = ((v_i \times\ \|q\|_l)^{N-1})^{d_i}$$

And get

$$W = \{[w_1], \ldots [w_m], \ldots [w_{m+n}]\}$$

$the\ order\ of\ W$ has been shuffled by secret permutation value ψ

$$W' = \psi(W) = (\left[w'_1\right], \ldots \left[w'_m\right], \ldots [w'_{m+n}])$$

Finally, the W' will be shared to RM.

• **RM to CSP.** In RM, the $[w_i']$ are decrypted by sk to obtain the value of w_i', where $i \in \{1,2,\ldots m+n\}$. When the decryption value of $w^{'} = 0$, then $s_i = 1$ otherwise $s_i = 0$. Then this obtained s_i has been decrypted,

$$S = \{[s_1],\ldots[s_m],\ldots[s_{m+n}]\}$$

Then, it will be shared to CSP.

• **CSP to RM.** It performs the below mentioned steps.

The inverse permutation of ψ^{-1} is performed to obtain

$$S^{'} = \psi^{-1}(S)$$

then, the last n components have been removed,

$$S^{''} = \{[s_1'],\ldots[s_m']\}$$

After that,

$$s_i' = \begin{cases} 0, & v_i < \|q\|_l \\ 1, & v_i = \|q\|_l \end{cases}$$

Then the support value of the mining query is

$$[supp(q)] = \prod_{i=1}^{m}[s_i^{'}]$$

Finally, the comparison results of B is calculated by performing comparison protocol between $supp(q)$ and $supp_{min}$. This result has been shared to RM for evaluation.

• **At RM.** The comparison results of B has been obtained after performing decryption operation.

$$q = frequent\ ,\ B = 1$$

$$q = not\ frequent,\ otherwise$$

4.3 Frequent item based on database privacy

In this section, the frequent item is generated based on database privacy. Let us assume, the cloud is having original dataset D and noisy database $D^{''}$.

Consider, our database (D) is having k itemset and all are encrypted,

$$\alpha = = \{E(a_1),E(a_2),\ldots E(a_k)\}$$

Then, the Assessor calculates the support value of $k-item$ in D, and denoted as F_i. Also, the support value of noisy transaction database (D'') is calculated and denoted as f_i. These two-support value is only known to assessor. The original support value is denoted by

$$F = \sum_{i=1}^{n}(F_i - f_i)$$

Consider, the client is having the encrypted version of binary dataset items

$$E(\delta) = \{E(g^{w_{m-1}}),\ldots E(g^{w_1}),E(g^{w_0})\}$$

And $F_i - f_i = (v^i_{m^{'}-1}\ldots v^i_1 v^i_0)(i = 1,2,\ldots n)$ for the itemset α, Here $m^{'} = [log_2|T''|]+1$ and $T^{''}$ is the noisy database, and RM is co-operated with assessor for calculating frequent itemset. The detailed protocols are described in Algorithm 1.

————————————————————————————————————

Algorithm 1: Private Comparison

Input: $E(\delta), y(public), (F_i - f_i, x_i)(i = 1, 2, \ldots, n)$

Output: 1 *if* α *is frequent and* 0 *otherwise*

———————————————————————————————————

1: n RM servers cooperate to compute

$$E(-\bar{\delta}) = \left(E\left(g^{1-w'_m}\right), \ldots, E(g^{1-w_1}), E(g^{1-w_0})\right.$$

2: $E(Z) \leftarrow E - \bar{\delta}$

3: for $i = 1$ *to* n {

4: Server S_i lets $v^i_{m'} = 0$ if $F_i - f_i \geq 0$ and 1 otherwise.

$$E(F_i - f_i) = (E\left(g^{v^i_{m'}}\right), E\left(g^{v^i_{m'}-1}\right), \ldots E\left(g^{v^i_1}\right), E\left(g^{v^i_0}\right))$$

5: n RM servers cooperate to compute

$E(Z) = E(Z) \boxplus E(F_i - f_i)$ as described in Section 3.4

6:}

7: n RM servers cooperate to decrypt $E(g^{z'_m})$

8: if $z'_m = 0$,then $\beta = 1$ else $\beta = 0$

9: return β

———————————————————————————————————

4.4 Association rule mining Deep-SVM

4.4.1 Deep-SVM Model

Deep neural networks and Support Vector Machines have recently become important in machine learning decision-making for feature extraction and categorization. In this research, we propose a Deep -Support vector machine (D-SVM) model for association rule mining in medical science. Two innovative approaches for feature extraction and learning are included in our proposed technique. 1. SVM is utilized as the top layer instead of softmax, which improves learning quality through gradient backpropogation. 2. Stacked SVM is intended to extract lower-level features, increasing the classification margin and ensuring quality rule generation.

From SVM rule generation, maximum hyperplane generalization has been achieved. In this section, the novel deep architecture of SVM is proposed. Here, softmax layer is replaced by stacked SVM section in order to extract the higher order feature for patient classification. These mined features are more helpful to make decision on incoming dataset.

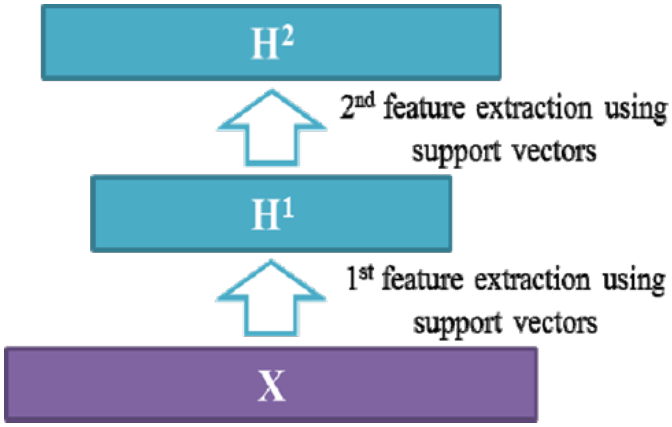

Fig 2. The structure of the proposed model

Figure 2 shows that the proposed method for feature learning and extraction. Here, multiple hidden layers are formed using SVM vectors, so each layer is responsible to extract discriminative features with corresponding multiplier. The hyperplane

between different classes is nonlinear in our cases. So, here the kernel radial basis function is used to make a hyper plane between training and testing samples. The kernel RBF is given by,

$$k(x,x') = exp\left(-\frac{\|x-x'\|}{2\sigma^2}\right)$$

In this section the association rule mining is calculated in RM with the help of assessor. After getting frequent itemset from section 4.1 and 4.2 the final frequent item is calculated by Algorithm 2.

Algorithm 2: Frequent itemset mining

Input: Frequent item by item privacy F_{item}, frequent item by database privacy $F_{database}$

Output: Final frequent itemset $Final_{item}$

1: $F_{item} = F_{item1,}F_{item2},\ldots F_{itemm}$, where m be the length of the frequent item in Section 4.1
2: $F_{database} = F_{database1,}F_{database2},\ldots F_{databasen}$, where n be the length of the frequent item in Section 4.2
3: $F_{final} = F_{item} \cup F_{database}$ %merge the whole dataset
4: when n==0 % until the length of the database
5: For i==length(m)
6: For j==length(n)
7: $F_{item} = Unique(F_{item}(i))$
8: $F_{base} = Unique(F_{database}(i))$
9: $F_{final} = F_{item} \cup F_{database}$ %merge the whole dataset

---

After calculating final frequent itemset for the incoming encrypted query, the association rule has been generated using Deep-SVM model. Algorithm 3 describes the protocol for mining association rule based on support and frequent itemset.

Algorithm 3: Rule Generation based on Deep-SVM

---

1. Rule-Generation (Training-data, n, minimum-support
2. {
3. S'= empty set
4. k=1
5. Do {
6. Si= generate all support vector k-itemset
7. For each item x in Si
8. Support (r) = suppcount(r)n
9. Lagrange multipliers ai
10. Target Label ti
11. Feature Extraction h^1(i) = $a_i t_i K(s_i, x)$,
12. If h >= support,
13. S'= S'+ r
14. End if
15. End for
16. k=k+1
17. }
18. While (Sk-1! = empty)
19. Return S'
20. }

This section contains the deep-SVM based rule mining. The detailed protocol is described in Algorithm 3. SVM rule procedure is used for association rule mining, rile extraction using trained support vectors. This procedure is very useful in data mining application for rule generation. This extracted rule has been enhanced by stacked SVM architecture. In this model, the softmax layer of deep learning procedure is replaced by stacked SVM. This new architecture will provide deep feature extraction from the frequent itemset for effective rule mining.

5 Results and Discussion

5.1 Security Analysis

In this section, the performance analysis of our proposed algorithm is discussed with different parameters. We used MATLAB software 2020 version with intel core i5- processor of 8 GB RAM for implementing two stage privacy solutions for ARM generation. In this encryption section, the Elgamal encryption method is used with 2160 and 1024-bits modulus form. For privacy comparison [21] and [22] papers encryption method is used. Our proposed method is having high level security when compared with double crypto encryption [21] and some other non- privacy methods [22].

5.1.1 Traditional Non-privacy Algorithm Vs Proposed Algorithm

Computational complexity is calculated from the running time of the algorithm. In order to highlight the computational complexity of our proposed method, comparison with non-privacy methods is carried out in this section. This comparison is performed on the dataset from https://www.kaggle.com/datasets/uciml/breast-cancer-wisconsin-data and https://data.world/datasets/health. The symbols used in this section is illustrated in Table 1.

Table 1. Symbols and Meaning

S.NO	Symbols	Details
1	t	Number of DO
2	k	Number of Itemset
3	c	Number of CSP
4	m	Number of Attributes

Table 2. Execution Time Comparison of Proposed Method t=4 and k=10

Number of Transactions	2-level:D-SVM (ms)	PP(ms) [12]	Apriori(ms) [12]	Eclat(ms) [12]	Fp-Growth(ms) [12]
500	5	39	27	11	9
1000	3	40	25	12	5
3000	6	38	19	9	8
10000	3	41	23	8	4
15000	6	33	17	9	7

Table 2 says that, our proposed algorithm is having minimum execution time when compared with traditional non-privacy algorithm [12]. Due to high computational process on data set, the execution time will be increased in existing methods. An efficiency of an algorithm is calculated by how many operations are executed by an algorithm with different transaction set. Based on this, our proposed algorithm is having minimum computation to generate the frequent itemset. So, it yields lesser execution time.

5.1.2 Privacy Algorithms Vs Proposed method

In this section, the familiar privacy algorithm of PCML is compared with our proposed methods. Here, the same software and dataset is used as previous section.

[21]'s frequent itemset mining applying privacy-preserving collaborative model learning (PCML) approach is one of the most astounding existing privacy-preserving technologies that do not leak sensitive data of raw information. From Figure 3 says that the Computation cost of the proposed algorithm is lower than the existing traditional algorithms. All the existing algorithm is

having higher execution time due to its costly operation. Figure 3 says that, the proposed algorithm is compared with privacy-preserving collaborative model learning PP [12] and (PCML) [21] scheme on different transaction range with different criteria (t,k,c) for mining the frequent item set.

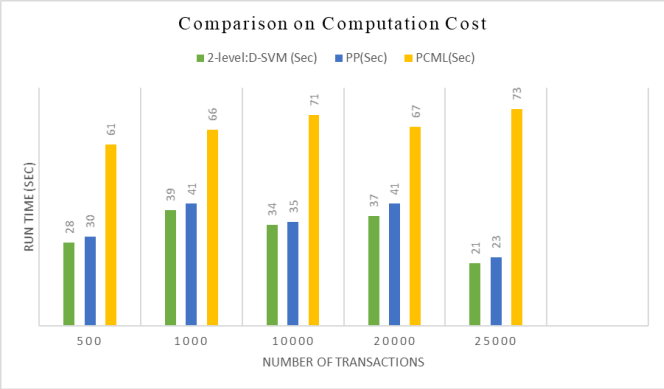

Fig 3. Performance Comparison of proposed work with non-Privacy methods

From the computation complexity, our proposed work is having minimum execution complexity on different transactions.

5.2 User Performance

In this section, the encryption time is compared with existing method of PCARM [12] by varying the transactions and attributes. This evaluation is performed with the time requirement of the outsourced data from user to cloud. From the Figure 4, we can say the encryption time is liner with respect to "m". i.e., when m=20 the encryption time of the data varies from 0.26 to 0.34 seconds if the transaction varied from 5000 to 10,000. When compared to PCARM, the proposed solution has a better computation time due to the elimination of redundancy in the generated rule.

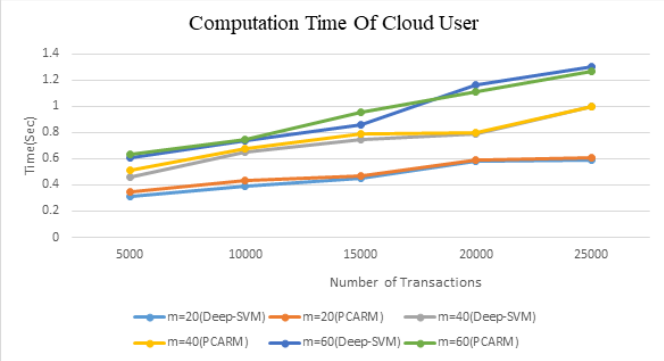

Fig 4. Computation Time Analysis of Cloud Users

Performance of the sub-protocols

In this Section, the subprotocols computations complexity and communication cost is evaluated. All three subprotocols are evaluated 200 times to evaluate this section. Tables 4 and 5 illustrated the obtained result for our proposed model. First, Rule Miner performs multiplication protocol, it will take 46.18ms time for computation, then it shares the data about 3.540 KB to the cloud server and DO. After that, the DO needs 23.24micro Sec of time for computation of 0.987KB of data and share it to CSP. Again, cloud server performs 0.987KB of data and share it to RM, it will take 42.54 ms. Next, the inner product protocol needs 75 times of time requirement when compared to multiplication protocol. So, 2.15,2.15s and 1.64 ms are the computation time cost of the RM, DO and CSP respectively. Note that, the RM needs much higher computation cost than DO. In inner product protocol, the RM, CSP and DO are having 259.541 kB,71.621KB and 71.621KB of computation cost. In comparison protocol,

the RM and CSP are the participants. The time cost and communication cost of the RM and CSP are 28.53ms,0.214KB and 43.47ms,0.198KB respectively.

Table 3. Time Cost of Sub-protocol

Sub-Protocols	Rule Miner	CSP	Data Owner
Multiplication Protocol	46.18 ms	42.54 ms	23.24μs
Inner Product Protocol	2.15s	2.13s	1.64ms
Comparison Protocol	28.53ms	43.47ms	N/A

Table 4. Communication Cost of Sub-protocol

Sub-Protocols	Rule Miner	CSP	Data Owner
Multiplication Protocol	3.540KB	0.987KB	0.987KB
Inner Product Protocol	259.541KB	71.621KB	71.621KB
Comparison Protocol	0.214KB	0.198KB	N/A

5.2.2 Efficiency of Cryptographic Blocks

We begin by evaluating the performance of the basic cryptographic blocks, as shown in Table 5. To attain 80-bit security levels, we represent N as 1024 bits for the proposed and BCP-variant algorithms. Table 6 shows that in the MP algorithm, the computation of CSP takes 0.158 seconds and he transmits 1.998KB data while connecting with Assessor, whereas Assessor takes 0.147 seconds to finish the computation and the transmission costs 1.498 KB. Furthermore, in the CAD procedure, the CSP requires 0.187 seconds to compute and transmit 1.498KB data to the Evaluator, whereas the Evaluator requires 0.154 seconds to estimate and deliver 0.999KB data. About the SC algorithm, the CSP computes and transmits 0.499KB data to the Assessor in 0.068 seconds, whereas the Assessor computes and delivers 0.498KB data in 0.048 seconds. We also tested IP across two 10-bit vectors; as shown in Table 6, the expense of CSP and Evaluator is about ten times that of a single MP. This calculation cost of CSP and assessor communication cost is smaller than the state of the art approach of variation BCP (V-BCP) in [(24)].

Table 5. Performance of cryptographic blocks based on the variant BCP Vs Proposed (100-times for average, 80-bits security level)

Algorithm	CSP Computation		Assessor Computation		CSP Communication		Assessor Communication	
	V-BCP	Proposed	V-BCP	Proposed	V-BCP	Proposed	V-BCP	Proposed
Multiplication Protocol(MP)	0.297 s	0.158s	0.251s	0.147s	1.998KB	1.998KB	1.498KB	1.498KB
Comparison across domain(CAD)	0.254	0.187	0.171s	0.154	1.499KB	1.498KB	0.999KB	0.999KB
Comparison Protocol(CP)	0.083	0.068	0.063s	0.048s	0.499KB	0.499KB	0.499KB	0.498KB
Inner Product protocol(IP)	2.301s	2.007s	3.102s	2.998s	19.981KB	19.981KB	14.989KB	14.987KB

6 Conclusion

In this research, the two-level item and database privacy preserving ARM is proposed. The proposed model is having two phases of (1) Secure frequent itemset mining (Item and Database privacy) and (2) Secure Association rule mining using Deep-SVM model. This approach ensures for the ARM task secured when the cloud user outsourced mining task over collaborated cloud environment by encryption operation. Item privacy has been performed by the sub-protocols of Multiplication, inner product and comparison protocols and private comparison protocol for database privacy. As a result, the proposed method provides two levels of unbreakable security to ensure anonymity. Such two-level security solutions outperform conventional typical security methods. As a result, the research shows the proposed approach provides very good privacy and security for health information kept on cloud servers.

Finally, the proposed model was evaluated with some existing privacy and non-privacy methods. It is concluded that the proposed method is more secure and accounting for lesser execution time and computation cost than the others. However, some additional modifications are also needed for enhanced ARM rule in various applications. The proposed method's security can be improved in the future by increasing key length or including alternative key generation techniques to avoid replay attacks.

References

1) Pengwei M, Kai W, Chunyu J, Junyi L, Jiafeng T, Siyuan L, et al. Research on Evaluation System of Relational Cloud Database. In: 2021 IEEE 20th International Conference on Trust, Security and Privacy in Computing and Communications (TrustCom). IEEE. 2021;p. 1369–1373. Available from: https://doi.org/10.1109/TrustCom53373.2021.00191.
2) Jangra G, Jangra M. Role of Artificial Intelligence in Online Shopping and its Impact on Consumer purchasing behaviour and Decision. In: 2022 Second International Conference on Computer Science, Engineering and Applications (ICCSEA). IEEE. 2022;p. 1–7. Available from: https://doi.org/10.1109/ICCSEA54677.2022.9936374.
3) Sarıyer G, Taşar CÖ. Highlighting the rules between diagnosis types and laboratory diagnostic tests for patients of an emergency department: Use of association rule mining. *Health Informatics Journal*. 2020;26(2):1177–1193. Available from: https://journals.sagepub.com/doi/pdf/10.1177/1460458219871135.
4) Asadianfam S, Kolivand H, Asadianfam S. A new approach for web usage mining using case based reasoning. *SN Applied Sciences*. 2020;2(7):1–11. Available from: https://doi.org/10.1007/s42452-020-3046-z.
5) Besiekierska A. Legal Assessment of the National Cybersecurity System in Poland in the Light of the New Developments in the NIS2 Directive. In: 2023 46th MIPRO ICT and Electronics Convention (MIPRO). IEEE. 2023;p. 1474–1477. Available from: https://doi.org/10.23919/MIPRO57284.2023.10159958.
6) Priyadarsini S, Sangeerthana B, Maheswari S, Prasanth A. An Efficient Privacy-Preserving in Frequent Item Set for Cloud Environment Using Apriori. *Annals of the Romanian Society for Cell Biology*. 2021;25(6):2934–2946. Available from: https://www.annalsofrscb.ro/index.php/journal/article/view/5991.
7) Purbey L, Samhitha N, Shreyam K, Teja PK, Ganesh DR. ECLAT Algorithm for Encrypted Files in the Cloud for Fast Association Rule Mining. *International Journal of Engineering Research & Technology (IJERT) NCAIT*. 2020;8(15):49–52. Available from: https://www.ijert.org/research/eclat-algorithm-for-encrypted-files-in-the-cloud-for-fast-association-rule-mining-IJERTCONV8IS15012.pdf.
8) Zhang B. Optimization of FP-Growth algorithm based on cloud computing and computer big data. *International Journal of System Assurance Engineering and Management*. 2021;12(4):853–863. Available from: https://doi.org/10.1007/s13198-021-01139-2.
9) Bu L, Zhang H, Xing H, Wu L. Research on parallel data processing of data mining platform in the background of cloud computing. *Journal of Intelligent Systems*. 2021;30(1):479–486. Available from: https://doi.org/10.1515/jisys-2020-0113.
10) Ige T, Sikiru A. Implementation of Data Mining on a Secure Cloud Computing Over a Web API Using Supervised Machine Learning Algorithm. In: Computer Science On-line Conference, CSOC 2022: Artificial Intelligence Trends in Systems;vol. 502 of Lecture Notes in Networks and Systems. Springer, Cham. 2022;p. 203–210. Available from: https://link.springer.com/chapter/10.1007/978-3-031-09076-9_20.
11) Nomura K, Shiraishi Y, Mohri M, Morii M. Secure Association Rule Mining on Vertically Partitioned Data Using Private-Set Intersection. *IEEE Access*. 2020;8:144458–144467. Available from: https://doi.org/10.1109/ACCESS.2020.3014330.
12) Dhinakaran D, Prathap PMJ, Selvaraj D, Kumar DA, Murugeshwari B. Mining Privacy-Preserving Association Rules based on Parallel Processing in Cloud Computing. *International Journal of Engineering Trends and Technology*. 2022;70(3):284–294. Available from: https://doi.org/10.14445/22315381/IJETT-V70I3P232.
13) Mankar RY, Gade A, Babu R. Association Rules Generation of Outsourced Transaction Data with Privacy-Preserving using Paillier Encryption. In: 2020 International Conference on Smart Electronics and Communication (ICOSEC). IEEE. 2020;p. 898–903. Available from: https://doi.org/10.1109/ICOSEC49089.2020.9215242.
14) Wu J, Mu N, Lei X, Le J, Zhang D, Liao X. SecEDMO: Enabling Efficient Data Mining with Strong Privacy Protection in Cloud Computing. *IEEE Transactions on Cloud Computing*. 2022;10(1):691–705. Available from: https://doi.org/10.1109/TCC.2019.2932065.
15) Hong Z, Zhang Z, Duan P, Zhang B, Wang B, Gao W, et al. Secure Privacy-Preserving Association Rule Mining With Single Cloud Server. *IEEE Access*. 2021;9:165090–165102. Available from: https://doi.org/10.1109/ACCESS.2021.3128526.
16) Šalgová V, Matiaško K. Reducing Data Access Time using Table Partitioning Techniques. In: 2020 18th International Conference on Emerging eLearning Technologies and Applications (ICETA). IEEE. 2021;p. 564–569. Available from: https://doi.org/10.1109/ICETA51985.2020.9379231.
17) Balasubramaniam S, Joe CV, Sivakumar TA, Prasanth A, Kumar KS, Kavitha V, et al. Optimization Enabled Deep Learning-Based DDoS Attack Detection in Cloud Computing. *International Journal of Intelligent Systems*. 2023;2023:1–16. Available from: https://doi.org/10.1155/2023/2039217.
18) Shantha RMJ, Mahender K, Jenifer AJM, Prasanth A. Security analysis of hybrid one time password generation algorithm for IoT data. In: International Conference on Research In Sciences, Engineering & Technology, 12–13 February 2021, Warangal, India;vol. 2418, Issue 1 of AIP Conference Proceedings. AIP Publishing. 2022. Available from: https://doi.org/10.1063/5.0081958.
19) Guo H, Peng L, Zhang J, Qi F, Duan L. Fooling AI with AI: An Accelerator for Adversarial Attacks on Deep Learning Visual Classification. In: 2019 IEEE 30th International Conference on Application-specific Systems, Architectures and Processors (ASAP). IEEE. 2019;p. 136–136. Available from: https://doi.org/10.1109/ASAP.2019.00-16.
20) Saravanan MS, Charan S. Prediction of Insufficient Accuracy for Human Activity Recognition using Convolutional Neural Network in Compared with Support Vector Machine. In: 2022 5th International Conference on Contemporary Computing and Informatics (IC3I). IEEE. 2023;p. 1915–1919. Available from: https://doi.org/10.1109/IC3I56241.2022.10072905.
21) Wang F, Zhu H, Liu X, Lu R, Hua J, Li H, et al. Privacy-Preserving Collaborative Model Learning Scheme for E-Healthcare. *IEEE Access*. 2019;7:166054–166065. Available from: https://doi.org/10.1109/ACCESS.2019.2953495.
22) Rajab A, Aqeel S, Reshan MSA, Ashraf A, Almakdi S, Rajab K. Cryptography based Techniques of Encryption for Security of Data in Cloud Computing Paradigm. *International Journal of Engineering Trends and Technology*. 2021;69(10):1–6. Available from: https://doi.org/10.14445/22315381/IJETT-V69I10P201.
23) Qiao Z, Yang Q, Zhou Y, Yang B, Xia Z, Zhang M, et al. An Efficient Certificate-Based Aggregate Signature Scheme With Provable Security for Industrial Internet of Things. *IEEE Systems Journal*. 2023;17(1):72–82. Available from: https://doi.org/10.1109/JSYST.2022.3188012.
24) Liu L, Su J, Chen R, Liu X, Wang X, Chen S, et al. Privacy-Preserving Mining of Association Rule on Outsourced Cloud Data from Multiple Parties. In: Australasian Conference on Information Security and Privacy, ACISP 2018;vol. 10946 of Lecture Notes in Computer Science. Springer, Cham. 2018;p. 431–451. Available from: https://doi.org/10.1007/978-3-319-93638-3_25.