Enhancing Security in Big Data Communication: A Novel Approach with Chi-Square Detective Ensembled Cardinal Gradient Bootstrap Aggregating Classifier

S L Swapna[1*]and V Saravanan[2]

[1] Research Scholar, Department of Computer Science, Hindusthan College of Arts and Science (Autonomous), Coimbatore, India
swapnamartin2003@gmail.com,

[2]Professor and Head, Department of Information Technology, Hindusthan College of Arts and Science (Autonomous), Coimbatore, India
vsreesaran@gmail.com

ABSTRACT: The utilization of big data analytics and related technologies, such as the Internet of Things (IoT), supports user intentions, behaviors, and operational decision-making. Security is a major concern when applying big data analytics to safeguard systems and protect information and data. Traditional security techniques have become inefficient in processing and identifying network threats within a reasonable timeframe. To address this issue, we introduce a novel model called Chi-Square Detective Ensembled Cardinal Gradient Bootstrap Aggregating Classifier-based Secured Data Communication (CSDECGBAC-SDC). This model enhances security with improved accuracy and reduced time complexity. The core functions of the CSDECGBAC-SDC model include user registration, data collection, and data communication. During registration, user information is initially recorded, and subsequently, the model collects data from registered users. The Chi-Square Detective Ensembled Cardinal Gradient Bootstrap Aggregating Classifier in the CSDECGBAC-SDC model is employed for user authentication. This ensemble technique utilizes a group of weak learners, represented as a Tversky Indexive Chi-square automatic interaction detection decision tree, to detect authorized users. The results from weak learners are combined, and cardinal voting is applied to determine the majority vote in data classification using the gradient ascent function, thereby enhancing secured data communication. Experimental evaluation considers factors such as classification accuracy, error rate, and classification time across various user numbers. Results indicate that the CSDECGBAC-SDC model significantly improves classification accuracy while minimizing error rates and classification times compared to conventional approaches.

KEYWORDS: Secured Big data Communication; Tversky Indexive Chi-Square Automatic Interaction Detection; Ensembled Cardinal Gradient Bootstrap Aggregating Classifier; Cardinal Voting; Gradient Ascent Function

(Received December 12th, 2023 / Accepted June 28th, 2024)

1. INTRODUCTION

Devices in a smart environment can connect with one another thanks to Internet of Things (IoT). The extensive usage of IoT technology is employed for data collection. Besides, these devices are utilized for a wide range of applications that are interconnected to the IoT environment and are remotely controlled. For IoT security solutions, various factors must be considered. Due to the obvious extensive capabilities made available by big data technologies, user information can be securely transmitted across all communications channels. When information is transmitted without authentication, malicious actors have the opportunity to violate the owner's privacy. Utilizing machine learning techniques, big data technologies have the potential to provide secure communication [1].

Homomorphic Block-Ring Security System (HBRSS) was designed in [2] for the security of data communication. But the access control method was not applied to improve the security level. A lightweight multi-factor authentication and authorization scheme in IoT cloud-based environment (LMAAS-IoT) was developed [3]. However, the IoT-based promising security solutions were not obtained.

A deep learning model was introduced in [4] for enhancing IoT security. But it failed to solve the challenges of improving the secure transmission of a big volume of data. A novel security risk analysis method was introduced in [5] for Big Data environments. However, it failed to apply the learning system in the cloud environment for security risk analysis.

Information Security Monitoring and Management scheme was developed in [6] with large Data for IoT Environment. However, the application of the access control algorithm to implement the massive amount of data did not increase the security level. The network security and protection approach was introduced in [7] for big data. But it failed to design a network security system that accomplishes all big data requirements.

A Secure Authentication and Data Sharing in Cloud (SADS-Cloud) enabled Big Data Environment was introduced in [8]. But, the complexity of data sharing was not reduced by applying SADS-Cloud architecture. A hybrid unauthorized data handling (HUDH) method was introduced in [9] for big data in cloud computing. But the effective learning technique was not applied to further improve the security of big data handling. A data transmission technique was designed in [10] to provide secure communication of IoT infrastructure. The designed technique failed to increase the complexity of device-to-device communication in IoT infrastructure.

A flexible and efficient authentication method was introduced in [11] for heterogeneous IoT devices to provide security and privacy with better storage and computational ability. But the designed authentication method failed to provide better security to a system with communications between the IoT devices to further increase the transmission efficiency.

A new CSDECGBAC-SDC model is developed with the following objectives to address the challenges currently present,

- The proposed CSDECGBAC-SDC model is introduced to facilitate secure big data communication with significantly less time consumption.
- First, the CSDECGBAC-SDC model is applied to accurately classify the authorized user and unauthorized during big data communication for enhancing security. The CSDECGBAC-SDC model uses the Chi-Square Detective Ensembled Cardinal Gradient Bootstrap Aggregating technique to categorize the authorized and unauthorized users based on the Tversky similarity index. After the classification using weak learners, the strong results are obtained by applying the cardinal voting scheme. Then apply the gradient ascent function to find the majority votes of classification results. In this way, authorized users are identified for secure data communication.
- To estimate the performance of the CSDECGBAC-SDC model and other related works, an empirical analysis is performed. The obtained results show that our proposed CSDECGBAC-SDC model performs better in terms of accuracy, error rate, and time.

The remainder of the article is divided into the following sections. The related studies are discussed in Section 2. Section 3 details the proposed CSDECGBAC-SDC technique. The experiment along with the dataset description is shown in Section 4. The performance outcomes of the proposed method and benchmarked works are presented in section 5. Finally, Section 6 brings the paper to the a conclusion.

2. RELATED WORKS

A lightweight smartcard-based secure authentication (LS-BSA) method was developed in [12] for proving the security and minimizing the computation and communication costs. However, the IoT-based big data forensic system was not designed. A trusted collaborative framework was developed in [13] for AI-enabled IoTs, computation security, and transmission security. The big data security access control technique did not, however, increase the level of security.

A novel lightweight authentication method was introduced in [14] for a cloud-based IoT environment. But the accurate authentication was not performed with minimum time consumption. A new authentication method was designed in [15] for IoT Environments to achieve a high-security level and minimum computation cost. However, the method was not improved the security level with big data applications.

A three-factor authentication framework was developed in [16] for IoT-driven security analysis. But it failed to investigate a more refined approach for improving IoT security by applying the learning techniques. Secure lightweight authentication and key agreement protocol is proposed in [17] for healthcare applications. But the optimal experimental designs, evaluation, big data handling were not considered.

A smart healthcare system was designed in [18] based on edge computing architecture for maintaining the privacy of patient data. But, the time complexity of authorized and unauthorized detection was not minimized. A secure and lightweight mechanism was designed in [19] for ensuring the security of device-to-device message transmission for the IoT-cloud system. The designed mechanism failed to preserve the message authentication.

The UUDIS-ECC and LSRHS-CNN algorithms were used by the authors of [20] to build a secure data transfer and effective data balancing strategy for 5G-based IoT data. Using an authentication and key agreement protocol for IoT-based WSNs that overcomes the security issues of earlier protocols is another initiative in [21]. Though the above two works produced satisfactory results, they failed to consider IoT analysis systems and fog layers.

A lightweight security mechanism based on Cipher Block Chaining (CBC) is implemented in [22] for an energy-efficient and secure health monitoring system to reduce energy consumption, monitoring system communication costs, and to ensure secure data transmission. It was claimed that the proposed work would prove to be a viable solution for securing and extending the lifetime of the IoMT network. In [23], researchers propose an efficient information embedding solution for ensuring data security in a cyber-physical network called CLoG. The secret information has been embedded in the detected edges in this case without using any authentication methods.

A share generation model was introduced in [24] for enhancing the privacy of healthcare data among individuals. However, the designed model was not used for various applications in cloud data security. Efficient data distribution and secure data transmission were performed in [25] based on IoT. However, an effective learning system was not implemented to improve secure data transmission.

3. PROPOSED WORK

The Internet of Things (IoT) has emerged as a significant technology that connects a large number of sensor devices in order to collect data based on applications. Communications in IoT environments are carried out on wireless channels that are susceptible to various unauthorized access [26]. IoT devices generated enormous data stored that are accumulated into a storage server and shared between the users. The processing of secure data communication is a significant problem for the development of real-time analysis. Therefore, an efficient technique called the CSDECGBAC-SDC model is introduced to perform secure big data communication based on the ensemble learning technique.

Figure 1 illustrates the cloud-enabled CSDECGBAC-SDC technique used for secured big data communication. The architecture model comprises a number of cloud users '$CU = CU_1, CU_2, \dots, CU_n$' wants to store their big data '$D = D_1, D_2, \dots., D_n$' to a cloud server.

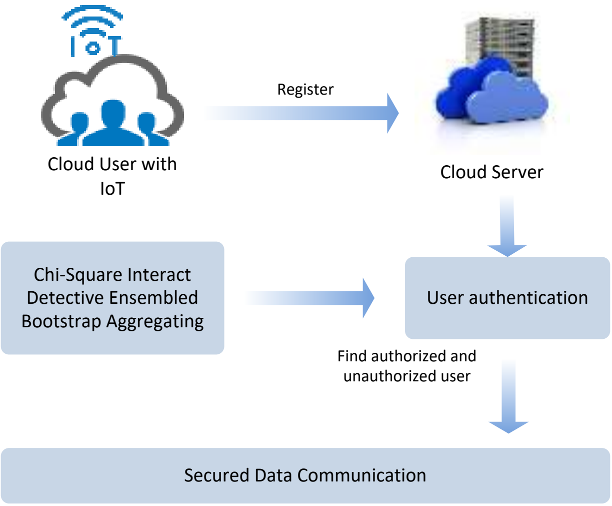
Cloud User with IoT
Register
Cloud Server
Chi-Square Interact Detective Ensembled Bootstrap Aggregating
User authentication
Find authorized and unauthorized user
Secured Data Communication

Figure 1. Architecture of the Proposed Cloud-Enabled CSDECGBAC-SDC Technique.

The cloud users first register their information with the cloud server. Data storage on a cloud server is sought by the user. Every time a user needs access to data, they must first confirm their identity. The cloud server decides if a user is authorized or unauthorized based on user authentication by using the Chi-Square Interact Detective Ensembled Bootstrap Aggregating method. The server grants access to the data to authorized users while denying it to unauthorized ones. The following subsections show how the CSDECGBAC-SDC technique differs from other approaches.

3.1 Registration

The proposed CSDECGBAC-SDC technique starts to perform the registration process before storing the big data into the cloud server. When the users want to store their data on the server, they first need to perform the registration. During the registration step, the user enters their personal information like the name, date of birth, age, gender, mobile number, and so on. The user's information's collected from the corresponding IoT is stored in the cloud server.

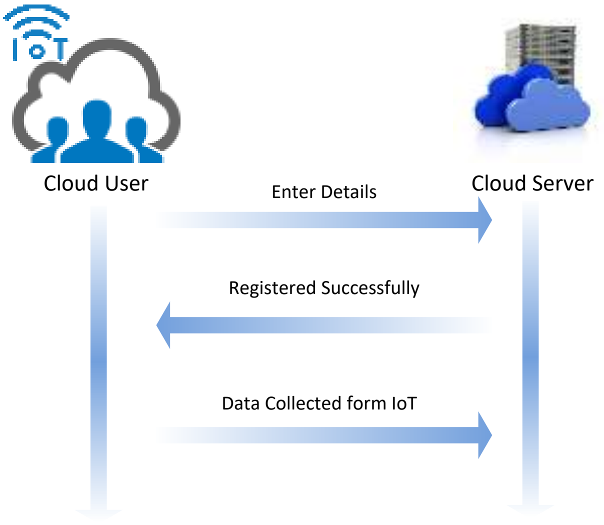

Figure 2. Registration Process Flow

Figure 2 exhibits the registration process. First, the users enter their personal information and the data collected from the IoT. Consider the MHEALTH (Mobile HEALTH) dataset, which includes recordings of ten subjects' body mobility and vital signs while performing varieties of physical activities. The movements encountered by the subject's various body parts, including acceleration, rate of turn, and magnetic field direction, is determined by sensors placed on the subject's chest, right wrist, and left ankle. The data collected is sent to a cloud server for further processing.

3.2 Chi-Square Interact Detective Ensembled Bootstrap Aggregating technique based user authentication

After the registration, the proposed CSDECGBAC-SDC technique performs the data collection at the cloud server by verifying the user authenticity. The authenticity of the user is identified by applying a Chi-Square Interact Detective Ensembled Bootstrap Aggregating technique. Bootstrap aggregating also termed bagging is a machine learning ensemble technique designed to improve the stability and accuracy of machine learning algorithms in statistical classification. The main aim of weak learners is combined to make a strong learner that attains better performance than a single one. Figure 3 demonstrates the Chi-Square Interact Detective Ensembled Bootstrap Aggregating technique for accurate classification.

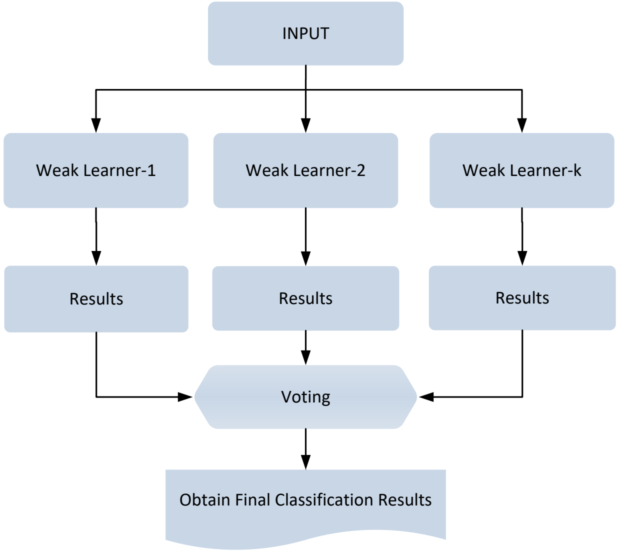

Figure 3. Structure of the Chi-Square Interact Detective Ensembled Bootstrap Aggregating Technique.

The Ensemble technique consists of training sets $\{x_i, Y\}$ where x_i denotes an input sample (i.e. cloud users) and 'Y'represents an ensemble classification results. The Ensembled Bootstrap Aggregating classifier initially constructs a 'k' set of weak learners $\{w_1, w_2, w_3, \ldots . w_k\}$. Here, the Chi-Square Interact Detective Decision Tree is used as a weak learner to categorize the cloud users as authorized or unauthorized. Tversky Indexive Chi-square automatic interaction detection decision tree is a classification tree is in which each internal (non-leaf) node is labeled with input. The arcs coming from a root node are connected to the leaf of the tree that the data set has been classified by the tree into either a specific class. Figure 4 demonstrates the Chi-square automatic interaction detection decision tree that classifies the user into authorized or unauthorized.

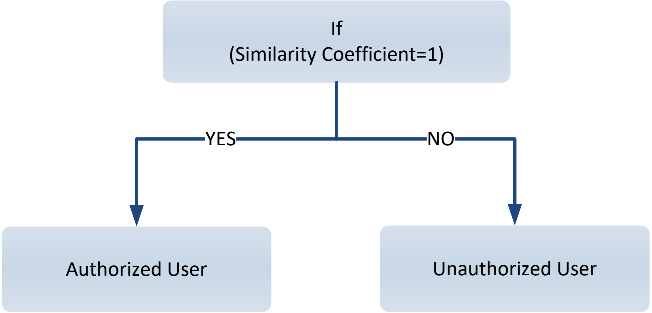

Figure 4. Tversky Indexive Chi-square Automatic Interaction Detection Decision Tree.

A Tversky similarity index is used in the stated Chi-square automatic interaction detection decision tree. It is used to calculate the correlation between the two variables (receiver ID and registered ID) as shown below,

$$SI = \frac{RID \cap RegID}{P\ (RID\ \Delta\ RegID) + Q\ (RID\ \cap RegID)} \qquad (1)$$

Where, SI indicates a Tversky similarity coefficient, RID denotes received ID, $RegID$ indicates a registered ID, $RID \cap RegID$ indicates a mutual dependence between the received ID and registered ID, $RID\ \Delta\ RegID$ indicates a variance between the received ID and registered ID. P and Q in (1) represent Tversky index parameters (P,Q 0). The similarity coefficient (SI) returns a value between [0, 1]. The authorized and unauthorized users are correctly identified by their similarities.

$$SI = \begin{cases} if\,[SI = 1]\ ;\ Authorized\ user \\ if\,[SI = 0]\ ;\ Unuthorized\ user \end{cases} \qquad (2)$$

Where, if the two IDs get matched, the user is said to be authorized. The user is considered to be an unauthorized user if not. The weak learner separates the user into many classes in this way.

The output of the weak learner results contains some training errors. Weak learners' classification results are combined in order to produce strong ones.

$$Y = \sum_{i=1}^{k} w_i \qquad (3)$$

Where, Y designates the output of strong classification results, w_i indicates an output of the weak learners. For each weak learner, the training error is computed based on the squared difference between the actual classification results and observed classification results. The error rate is measured as given below,

$$Er = [R_A - R_o]^2$$

(4)

Where, Er represents the error after the classification, R_A represents the actual output,R_o symbolizes the observed results. By applying the cardinal voting scheme, the weak learner results are rated based on the preference order according to the error value.

Table 1. Example of Cardinal Voting Based on Weak learner's Arrangement

Weak Learners	Preference Order
$\boldsymbol{w_1}$	First
$\boldsymbol{w_3}$	Second
$\boldsymbol{w_4}$	Third
$\boldsymbol{w_2}$	Fourth

Table 1 shows the order of the weak learners based on the error value. The weak learner 'w_1' rated in the order of the first position. It means that the weak learner 'w_1' has minimum error among the other weak learner. Likewise, the weak learners are ordered second, three, four according to the error value. After the rating process, the votes are registered to find accurate classification results that have minimum errors. The strong classification results are obtained by finding the majority votes based on finding the local maximum of that function. This process is then called gradient ascent.

$$Y = \arg\max_{b} R\,(w_i) \qquad (5)$$

Where Y represents the strong classification results, $arg\ max$ indicates an argument of the maximum function (i.e. gradient ascent) to find out the majority vote (R) of the classification result whose decision is known to the k$^{\text{th}}$classifier. In this way, accurate classification results are obtained.

Algorithm 1 describes the phased process of secure data communication using the Chi-Square Detective Ensembled Cardinal Gradient Bootstrap Aggregating technique. The user transmits their information to the server during the registration process. Accordingly, the server generates a successfully registered message for the registered user. Then the user stores the details collected from the IoT. The user wants to access the data from the server. The user wishes to obtain data from the server. The cloud server first validates the user's identity using the Tversky similarity coefficient. When the similarity returns '+1,' the user is said to be authorized. Otherwise, the user is considered unauthorized. Finally, the data is delivered to the authorized user by the cloud server. In this way, secure data transmission between server and user is performed.

Algorithm 1: Chi-Square Detective Ensembled Cardinal Gradient Bootstrap Aggregating Classifier based Secured Data Communication

```
Input:  Dataset '$\boldsymbol{DS}$', IoT device '$\boldsymbol{S = S_1, S_2, \dots, S_n}$'
        Cloud user's '$\boldsymbol{CU = CU_1, CU_2, \dots, CU_n}$'
        Big data '$\boldsymbol{D = D_1, D_2, \dots, D_n}$'
        Cloud server '$\boldsymbol{CS}$'
Output: Secured Data Communication

Begin
    For each user '$\boldsymbol{CU}$'
        Enter the details to '$\boldsymbol{CS}$'
        Enter the details to '$\boldsymbol{CS}$'
    Store the details collected from IoT
    End for
    $\boldsymbol{CS}$ generates the successfully registered message
    For each user '$\boldsymbol{CU}$'
        Collect the data from '$\boldsymbol{CS}$'
        $\boldsymbol{CS}$ validate the user detail
        Apply ensemble technique
        Construct 'k' number of weak learners
            Construct the decision tree
            Measure the similarity '$\boldsymbol{SI}$'
        $\boldsymbol{if}$ ($\boldsymbol{SI = +1}$) then
            The cloud user is said to be an authorized user
            Perform secure data communication
        else
            cloud user is said to be an unauthorized user
            No data communication between cloud users
        End if
    End for
End
```

4. EXPERIMENTAL SETUP

Experimental evaluation of the proposed CSDECGBAC-SDC method and the existing HBRSS [2] LMAAS-IoT [3] is implemented in the Java language via CloudSim simulation. In order to experiment, secure big data communication between the cloud users was performed using the MHealth dataset collected from https://www.kaggle.com/datasets/gaurav2022/mobile-health. The dataset consists of 14 attributes and 12,15,745 instances. The main aim of the dataset is to record several physical activities for ten volunteers of diverse profiles. The ten volunteers generate many data (i.e. instances) and these data are communicated to the authorized entity.

Table 2. Attributes Description

S. No	Attributes	Description
1	alx	Acceleration from left-ankle sensor (X axis)
2	aly	Acceleration from the left-ankle sensor (Y axis)
3	alz	Acceleration from left-ankle sensor (Z axis)
4	glx	Gyro from the left-ankle sensor (X axis)
5	gly	Gyro from the left-ankle sensor (Y axis)
6	glz	Gyro from the left-ankle sensor (Z axis)
7	arx	Acceleration from right-ankle sensor (X axis)
8	ary	Acceleration from right-ankle sensor (Y axis)
9	arz	Acceleration from right-ankle sensor (Z axis)
10	grx	Gyro from the right-ankle sensor (X axis)
11	gry	Gyro from the right-ankle sensor (Y axis)
12	grz	Gyro from the right-ankle sensor (Z axis)
13	Activity	Corresponding activity

14	Subject	Volunteer subjects (1 – 9)

5. RESULTS AND DISCUSSIONS

In this section, the performance of the proposed CSDECGBAC-SDC method and the existing HBRSS [2] LMAAS-IoT [3] are presented. The proposed CSDECGBAC-SDC method's performance is analyzed based on the following parameters as Classification Accuracy, error rate, and Classification time. The performance of these parameters is analyzed with the help of a table and graphical representation.

5.1 Impact of Classification Accuracy

Classification accuracy is measured as the ratio of the number of cloud users that are accurately classified as authorized or unauthorized for secure data communication. This classification accuracy is formulated as follows,

$$CA = \left[\sum_{i=1}^{n} \frac{CU_{AC}}{CU_i}\right] * 100 \qquad (6)$$

From the above equation (6), the classification accuracy 'CA' is measured based on the number of cloud users accurately classifiedCU_{AC}, CU_i denotes the number of cloud users in the simulation. It is measured in terms of percentage (%).

Table 3. Classification Accuracy

Number of Cloud Users	Classification accuracy (%)		
	CSDECGBAC-SDC	HBRSS	LMAAS-IoT
10000	99.1	98.1	97
20000	98.6	96.5	94.5
30000	97.83	94.03	91.73
40000	97.52	93.37	89.11
50000	96.96	92.84	88.30
60000	96.25	91.41	87.42
70000	95.71	90.64	86.07
80000	95.31	89.32	85.56
90000	94.68	88.05	85.38
100000	94.25	87.54	84.56

Table 3 shows how classification accuracy is measured in relation to the number of cloud users ranging from 10,000 to 100,000. The observed table values represent the classification accuracy performance results of three different methods, namely the CSDECGBAC-SDC method and the existing HBRSS [2] LMAAS-IoT [3]. The obtained experimental results confirm that the classification accuracy is found to be higher using the CSDECGBAC-SDC method when compared to existing methods. Let us consider the 10000 loud users for conducting the experiments in the first iteration. By applying the CSDECGBAC-SDC method, 9910 cloud users are correctly identified as authorized or unauthorized hence the classification accuracy is found to be 99.1%. The classification accuracy of HBRSS [2] LMAAS-IoT [3] is 98.1% and 97% respectively. For each method, varied experimental findings are seen. The CSDECGBAC-SDC method's results are compared to those of other methods. The CSDECGBAC-SDC approach improves the classification accuracy by 9% and 5% when compared to [2] and [3], respectively, according to the average of ten comparison findings.

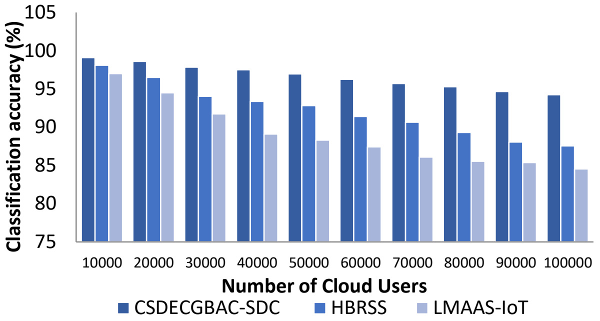

Figure 5. Performance results of classification accuracy versus the number of cloud users.

Figure 5 illustrates the performance results of classification accuracy versus the number of cloud users taken in the range from 10000 to 10000. In the graphical representation, the number of cloud users is taken as input in the horizontal axis i.e. 'x' axis but the output results of classification accuracy are obtained at the vertical axis i.e. 'y' axis. The graphical plot indicates that the classification accuracy of three different methods namely the CSDECGBAC-SDC method and the existing HBRSS [2] LMAAS-IoT [3] are represented by the three various colors such as blue, red, and green respectively. The observed illustrates that the CSDECGBAC-SDC method improves the classification accuracy when compared to existing methods. This improvement of the CSDECGBAC-SDC method is achieved by applying the Chi-Square Interact Detective Ensembled Bootstrap Aggregating technique. The proposed ensemble technique verifies the user-ID with the registered ID by using the weak learner as Tversky Indexive Chi-square automatic interaction detection decision tree. If these two IDs get matched, then the authorized user is identified. The ensemble bootstrap aggregating technique combines the weak classification results to make a strong output result by applying a cardinal voting scheme.

5.2. Impact of Error Rate

The error rate is measured as the ratio of cloud users wrongly classified to the total number of cloud users. This formula for calculating the error rate is expressed as follows,

$$ER = \left[\sum_{i=1}^{n} \frac{CU_{wC}}{CU_i}\right] * 100 \qquad (7)$$

From the above equation (7), 'ER' denotes an error rate, CU_{wC} denotes a cloud user wrongly classified, CU_i number of cloud users It is measured in terms of percentage (%).

Table 4. Error Rate

Number of Cloud Users	Error rate (%)		
	CSDECGBAC-SDC	HBRSS	LMAAS-IoT
10000	0.9	1.9	3
20000	1.4	3.5	5.5

30000	2.16	5.96	8.26
40000	2.47	6.62	10.88
50000	3.04	7.16	11.69
60000	3.75	8.58	12.58
70000	4.28	9.35	13.92
80000	4.68	10.68	14.43
90000	5.31	11.95	14.61
100000	5.74	12.45	15.43

Table 4 exhibits the performance results of the error rate using three methods namely the CSDECGBAC-SDC method and the existing HBRSS [2] LMAAS-IoT [3] with respect to the number of cloud users considered from 10000 to 100000. The above-observed results indicate that the performance of error rate using the CSDECGBAC-SDC model is significantly minimized than the two existing HBRSS [2] LMAAS-IoT [3]. By considering '10000'cloud users for experimentation and the overall error rate was observed to be '0.9%', '1.9 %', and '3%' using CSDECGBAC-SDC, HBRSS [2] LMAAS-IoT [3]. Similarly, other performance results are observed with respect to various numbers of cloud users. The observed results indicate that the error rate is considerably reduced by 57% using the CSDECGBAC-SDC method when compared to [2] and 70% compared to [3] respectively.

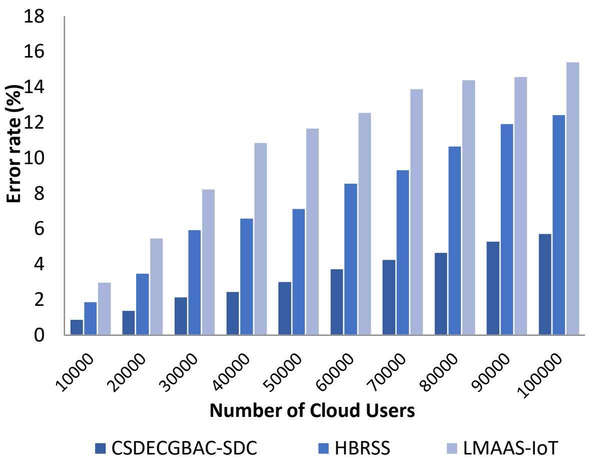

Figure 6. Performance results of error rate versus the number of cloud users.

Figure 6 given above exhibits the graphical illustration of the error rate using three different methods and a different number of cloud users. The above graphical representation illustrates that the error rate using the CSDECGBAC-SDC method is found to be relatively lesser when compared to conventional methods. The reason behind this significant improvement was the application of cardinal voting and gradient ascent function. By applying cardinal voting, the weak learner results are ordered based on the training error. The weak learner with a higher error rate is removed. Finally, the gradient ascent function is applied to find the majority votes of the samples. Therefore, the obtained ensemble results improve the accuracy and minimize the error rate.

5.3. Impact of Classification Time

The classification time is defined as the amount of time consumed by the algorithm for classifying the authorized and unauthorized users. The formula for classification time is measured as given below,

$$CT = \sum_{i=1}^{n} CU_i * Time\ [classification] \quad (8)$$

Where CT denotes a classification time 'CU_i denotes the number of cloud users and the time consumed in classification '$Time\ [classification]$'. It is measured in terms of milliseconds (ms).

Table 5. Classification Time

Cloud Users	Classification Time (ms)		
	CSDECGBAC-SDC	HBRSS	LMAAS-IoT
10000	4600	6100	7500
20000	4800	6800	8000
30000	5100	6900	7500
40000	6400	8000	9200
50000	7100	9500	10000
60000	9120	10800	11400
70000	9800	12600	14000
80000	10400	13600	15200
90000	11250	14400	17100
100000	13200	15000	18000

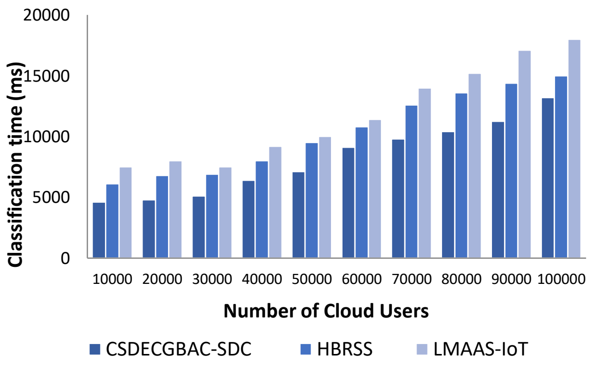

Figure 7. Performance results of classification time versus the number of cloud users.

The classification times for three alternative methods—the CSDECGBAC-SDC technique, the current HBRSS [2] and LMAAS-IoT [3]—are shown in Table 5 and Figure 7. The observed results show that the CSDECGBAC-SDC model's classification time is significantly shorter than that of the other two existing techniques. The table 5 shows that for each of the ten distinct runs, the classification time is significantly decreased. For each run, the input counts of samples get increased as a result the classification time of three different methods also gets increased. By considering the '10000' samples in the first iteration, the classification time of the CSDECGBAC-SDC model was found to be '4600ms', '6100ms' using [2], and '7500ms' using [3] respectively. Likewise, different counts of input data were used for the several runs. The overall results indicate that the classification time of the CSDECGBAC-SDC model is considerably reduced by 22% and 31% when compared to HBRSS [2] and LMAAS-IoT [3] respectively. This is because of performing the user registration, before the data collection and secure data communication. Initially, the users' details are registered and then the server collects the data from the registered user. Then the classification is performed using the ensemble technique with the help of the Tversky similarity index. The similarity function matches the ID of the user to find the authorized or unauthorized.

6. CONCLUSION

The technology of big data is integrated into the cloud to facilitate the real-time applications for data security. This paper presents a CSDECGBAC-SDC model using the IoT for a cloud environment, which aims to improve big data communication security. The proposed CSDECGBAC-SDC model reduces the possibility of unauthorized data access, enhances accuracy, and provides more protection. The CSDECGBAC-SDC model first performs the registration process that includes user information. Then the registered user allows storing their data collected from the IoT device to server. Then any user who wants to access the data, first they verify the authenticity using Chi-Square Detective Ensembled Cardinal Gradient Bootstrap Aggregating technique. The ensemble technique accurately finds the authorized user or unauthorized user based on the cardinal voting and gradient ascent function. This process enhances the classification accuracy and minimizes the error rate. The performance of the CSDECGBAC-SDC model is analyzed with different metrics such as classification accuracy, error rate, and classification time. Therefore, the quantitative results and discussion conclude that the presented CSDECGBAC-SDC model is highly promising to provide higher classification accuracy with a lesser time as well as error rate than the conventional methods.

REFERENCES

[1] Swapna, S. L., and V. Saravanan. "Survival Analysis on Secured Data Communication in Cloud." *International Journal of Computer Applications*, vol. 183, no. 46, Foundation of Computer Science, Jan. 2022, pp. 31–35. *Crossref*, https://doi.org/10.5120/ijca2022921864.

[2] Xie, Hui, et al. "HBRSS: Providing High-secure Data Communication and Manipulation in Insecure Cloud Environments." *Computer Communications*, vol. 174, Elsevier BV, June 2021, pp. 1–12. *Crossref*, https://doi.org/10.1016/j.comcom.2021.03.018.

[3] Alsahlani, Ahmed Yaser Fahad, and Alexandru Popa. "LMAAS-IoT: Lightweight Multi-factor Authentication and Authorization Scheme for Real-time Data Access in IoT Cloud-based Environment." *Journal of Network and Computer Applications*, vol. 192, Elsevier BV, Oct. 2021, p. 103177. *Crossref*, https://doi.org/10.1016/j.jnca.2021.103177.

[4] Amanullah, Mohamed Ahzam, et al. "Deep Learning and Big Data Technologies for IoT Security." *Computer Communications*, vol. 151, Elsevier BV, Feb. 2020, pp. 495–517. *Crossref*, https://doi.org/10.1016/j.comcom.2020.01.016.

[5] Rosado, David G., et al. "MARISMA-BiDa Pattern: Integrated Risk Analysis for Big Data." *Computers & Security*, vol. 102, Elsevier BV, Mar. 2021, p. 102155. *Crossref*, https://doi.org/10.1016/j.cose.2020.102155.

[6] Liang, Wuchao, et al. "Information Security Monitoring and Management Method Based on Big Data in the Internet of Things Environment." *IEEE Access*, vol. 9, Institute of Electrical and Electronics Engineers (IEEE), 2021, pp. 39798–812. *Crossref*, https://doi.org/10.1109/access.2021.3064350.

[7] El Alaoui, Imane, and Youssef Gahi. "Network Security Strategies in Big Data Context." *Procedia Computer Science*, vol. 175, Elsevier BV, 2020, pp. 730–36. *Crossref*, https://doi.org/10.1016/j.procs.2020.07.108.

[8] Narayanan, Uma, et al. "A Novel System Architecture for Secure Authentication and Data Sharing in Cloud Enabled Big Data Environment." *Journal of King Saud University - Computer and Information Sciences*, vol. 34, no. 6, Elsevier BV, June 2022, pp. 3121–35. *Crossref*, https://doi.org/10.1016/j.jksuci.2020.05.005.

[9] Razaque, Abdul, et al. "Big Data Handling Approach for Unauthorized Cloud Computing Access." *Electronics*, vol. 11, no. 1, MDPI AG, Jan. 2022, p. 137. *Crossref*, https://doi.org/10.3390/electronics11010137.

[10] Sharma, Rohit, and Rajeev Arya. "Secure Transmission Technique for Data in IoT Edge Computing Infrastructure." *Complex & Intelligent Systems*, vol. 8, no. 5, Springer Science and Business Media LLC, Nov. 2021, pp. 3817–32. *Crossref*, https://doi.org/10.1007/s40747-021-00576-7.

[11] Fang, Dongfeng, et al. "A Flexible and Efficient Authentication and Secure Data Transmission Scheme for IoT Applications." *IEEE Internet of Things Journal*, vol. 7, no. 4, Institute of Electrical and Electronics Engineers (IEEE), Apr. 2020, pp. 3474–84. *Crossref*, https://doi.org/10.1109/jiot.2020.2970974.

[12] Deebak, B. D., and Fadi AL-Turjman. "Lightweight Authentication for IoT/Cloud-based Forensics in Intelligent Data Computing." *Future Generation Computer Systems*, vol. 116, Elsevier BV, Mar. 2021, pp. 406–25. *Crossref*, https://doi.org/10.1016/j.future.2020.11.010.

[13] Zhang, Qingyang, et al. "A Trusted and Collaborative Framework for Deep Learning in IoT." *Computer Networks*, vol. 193, Elsevier BV, July 2021, p. 108055. *Crossref*, https://doi.org/10.1016/j.comnet.2021.108055.

[14] Wazid, Mohammad, et al. "LAM-CIoT: Lightweight Authentication Mechanism in Cloud-based IoT Environment." *Journal of Network and Computer Applications*, vol. 150, Elsevier BV, Jan. 2020, p. 102496. *Crossref*, https://doi.org/10.1016/j.jnca.2019.102496.

[15] Son, Seunghwan, et al. "A Secure, Lightweight, and Anonymous User Authentication Protocol for IoT Environments." *Sustainability*, vol. 13, no. 16, MDPI AG, Aug. 2021, p. 9241. *Crossref*, https://doi.org/10.3390/su13169241.

[16] Saqib, Manasha, et al. "A Lightweight Three Factor Authentication Framework for IoT Based Critical Applications." *Journal of King Saud University - Computer and Information Sciences*, vol. 34, no. 9, Elsevier BV, Oct. 2022, pp. 6925–37. *Crossref*, https://doi.org/10.1016/j.jksuci.2021.07.023.

[17] El Zouka, Hesham A., and Mustafa M. Hosni. "Secure IoT Communications for Smart Healthcare Monitoring System." *Internet of Things*, vol. 13, Elsevier BV, Mar. 2021, p. 100036. *Crossref*, https://doi.org/10.1016/j.iot.2019.01.003.

[18] Singh, Ashish, and Kakali Chatterjee. "Securing Smart Healthcare System With Edge Computing." *Computers & Security*, vol. 108, Elsevier BV, Sept. 2021, p. 102353. *Crossref*, https://doi.org/10.1016/j.cose.2021.102353.

[19] Al Sibahee, Mustafa A., et al. "Lightweight Secure Message Delivery for E2E S2S Communication in the IoT-Cloud System." *IEEE Access*, vol. 8, Institute of Electrical and Electronics Engineers (IEEE), 2020, pp. 218331–47. *Crossref*, https://doi.org/10.1109/access.2020.3041809.

[20] Yadav, Kusum, et al. "A Secure Data Transmission and Efficient Data Balancing Approach for 5G-based IoT Data Using UUDIS-ECC and LSRHS-CNN Algorithms." *IET Communications*, vol. 16, no. 5, Institution of Engineering and Technology (IET), Jan. 2022, pp. 571–83. *Crossref*, https://doi.org/10.1049/cmu2.12336.

[21] Ostad-Sharif, Arezou, et al. "Three Party Secure Data Transmission in IoT Networks Through Design of a Lightweight Authenticated Key Agreement Scheme." *Future Generation Computer Systems*, vol. 100, Elsevier BV, Nov. 2019, pp. 882–92. *Crossref*, https://doi.org/10.1016/j.future.2019.04.019.

[22]Mondal, Sanjoy, et al. "Energy Efficient and Secure Healthcare Data Transmission in the Internet of Medical Things Network." *Microsystem Technologies*, Springer Science and Business Media LLC, Nov. 2022. *Crossref*, https://doi.org/10.1007/s00542-022-05398-2.

[23] Jan, Aiman, et al. “Secure Data Transmission in IoTs Based on CLoG Edge Detection.” *Future Generation Computer Systems*, vol. 121, Elsevier BV, Aug. 2021, pp. 59–73. *Crossref*, https://doi.org/10.1016/j.future.2021.03.005

[24] Rani, S. Sheeba, et al. “Optimal Users Based Secure Data Transmission on the Internet of Healthcare Things (IoHT) With Lightweight Block Ciphers.” *Multimedia Tools and Applications*, vol. 79, no. 47–48, Springer Science and Business Media LLC, May 2019, pp. 35405–24. *Crossref*, https://doi.org/10.1007/s11042-019-07760-5.

[25] Pampapathi, B. M., et al. “Data Distribution and Secure Data Transmission Using IANFIS and MECC in IoT.” *Journal of Ambient Intelligence and Humanized Computing*, vol. 13, no. 3, Springer Science and Business Media LLC, Jan. 2021, pp. 1471–84. *Crossref*, https://doi.org/10.1007/s12652-020-02792-4.

[26] R. John Martin. “IoMT Supported COVID Care – Technologies and Challenges.” *International Journal of Engineering and Management Research*, vol. 12, no. 1, Vandana Publications, Feb. 2022, pp. 125–31. *Crossref*, https://doi.org/10.31033/ijemr.12.1.16.