Journal of Intelligent Systems and Internet of Things
Vol. 16, No. 02, PP. 142-157, 2025
DOI: https://doi.org/10.54216/JISIoT.160211

Deep Learning-based sensitive data detection with optimization-enabled secure encryption model for data privacy preservation in IoT

Mathias Agbeko[1,*], Disha Handa[1]
[1]University Institute of Computing, Department of Computer Applications, Chandigarh University, Punjab, India
Emails: magbeko@uew.edu.gh; disha.e11162@cumail.in

Abstract

The express expansion of the Internet of Things (IoT) has led to an exponential increase for data being generated and transmitted from various connected devices. This poses significant challenges in terms of data privacy and security, as unauthorized access to such sensitive information can have severe consequences like identity theft or financial fraud. This research proposes a model for sensitive data detection and protection in IoT, based on deep learning and optimization-enabled secure encryption. By combining deep learning-based sensitive data detection and optimization-enabled secure encryption, this model offers a comprehensive solution to preserve data privacy in IoT. The proposed model uses a novel and secure encryption algorithm, ensuring the privacy of the data. An algorithm, Improved Skill Optimization Algorithm (ISOA), which enhances the performance of existing optimization algorithms by incorporating the concept of Double Exponential Smoothing (DES), is proposed for the secure key generation for the data encryption. Data Encryption Standard (DES) is a block cipher algorithm that encrypts and decrypts data using a 56-bit key and 64-bit blocks. The proposed model provides a robust solution for data privacy preservation in IoT networks, which is crucial for protecting sensitive information from unauthorized access and data breaches. The proposed algorithm's performance analysis is evaluated using metrics, like computation time, memory, and fitness function. Results indicate that proposed ISOA based encryption model succeeded a greater performance, with a memory of 0.5170 MB, computational time of 1126.47 sec and fitness value of 1.3630.

Received: December 25, 2024 Revised: February 19, 2025 Accepted: March 09, 2025

Keywords: Deep learning; Internet of Things (IoT); Privacy preservation; Optimization; Sensitive data; Encryption

1. Introduction

In recent times, the IoT has acquired noteworthy consideration in both research and practical implementation. It is a model that involves regular objects having the ability to sense and communicate with one another via the internet. With the widespread availability of broadband internet and reduced connectivity costs, more devices and sensors are being connected to it, creating an ideal environment for IoT to flourish. Furthermore, complex chips and sensors embedded in everyday objects are transmitting valuable data. However, to share such vast amounts of data, the devices must communicate securely by means of the IoT podium that incorporated records from several sources and employs analytics to extract the generally useful information for various applications. Nonetheless, it is crucial to note that security and privacy remain paramount concerns, especially in terms of confidentiality, data integrity, and authenticity, as everything will be connected to the internet [1].

The interconnectivity of devices in IoT presents several security and privacy concerns, leading to security challenges in network computing. Anywhere and at any moment, such devices could face attacks including denial of service, identity fabrication, physical threats, and more. Consumption of power resources and computational overheads are among the most significant challenges associated with IoT devices [10]. IoT poses severe security issues when connecting everything to the existing conventional networks, especially concerning securing sensitive data. Additionally, upcoming security and privacy issues prompt further consideration of privacy, authenticity, and data integrity concerns in IoT [12, 13]. A challenge concerning IoT security includes crypto-attack, RF jamming, spoofing, Sybil attack, wormhole attack, tag attack, eavesdropping, among others. Conventional encryption strategies require that third-party services (cloud) decrypt sensitive data before managing and computing the data from clients [7].

Privacy preserving is a top concern for end-clients that use IoT-enabled applications, and cloud computing serve as a foundation for storing and processing IoT data. Cryptographic-based methodologies are one of the most effective approaches introduced to ensure IoT data security, providing components to facilitate data classification and integrity. Maintaining constant encryption of data in the cloud preserves control and eliminates assumptions. The security requirements for IoT data and algorithms have become incredibly stringent over the years. Various options like secure data encryption exist for safely storing and reading data. However, handling encrypted data freely or modifying executable functions while ensuring privacy can be challenging. Homomorphic cryptosystems can be beneficial in preserving data with high security. Despite their potential benefits, there is still much research required to make these systems practical. Recent studies have focused on implementing privacy laws that enforce adherence to data protection principles, including data minimization, organizational and technical measures, such as pseudonymization, are employed. [10]. Privacy involves protecting sensitive information from individuals, and privacy preservation (PP) aims at preventing intruders from accessing additional personal information beyond what is necessary in real-time or statistical data [15]. In the commercial IoT device marketplace, security solutions are often promoted as privacy-preserving solutions, leading to confusion between the two concepts [16]. Existing privacy-preserving solutions focus mainly on securing communication channels and authentication and authorization mechanisms, with minimal attention paid to privacy preservation in the data collection, aggregation, storage, and retrieval processes [11, 17].

The proposed model aims to preserve the privacy of sensitive data in IoT systems through a combination of deep learning-based sensitive data detection and secure encryption. The model utilizes an Improved Skill Optimization Algorithm (ISOA) to generate the key for the encryption process and provides enhanced security. One advantage of this model is that it employs deep learning techniques to detect sensitive data in IoT systems so that the time required to process the entire data can be reduced. Deep learning has shown to be an effective method for detecting complex patterns and anomalies in large datasets. By using this technique, the model can more accurately identify sensitive data, which helps to ensure that the most critical data in the system is best protected. The proposed secure encryption process with the ISOA secures the sensitive data, and results in a more robust security framework for the IoT system. This optimization allows the encryption process to be dynamic and adaptive, making it more challenging for an attacker to compromise the system and access the sensitive data.

This work aims to achieve the following main objectives:

1. To identify the data as sensitive or non-sensitive, a deep learning model is employed so that the computational complexity can be reduced since only the sensitive data are taken into account for the encryption.
2. To protect the sensitive data in IoT network, a novel mathematical model is designed, where the keys for the encryption procedure is generated by proposing an algorithm, Improved Skill Optimization Algorithm (ISOA).
3. To propose ISOA by modifying existing skill optimization algorithm (SOA) using DES, which is capable of quickly adapting to changes in the data, as it assigns more weight to recent observations. This leads to better convergence rates, improved solution quality, and faster computation time when compared to conventional algorithms.
4. To estimate the effectiveness of the presented strategy using metrics such as computation time, memory, and fitness function.

2. Literature Survey

The following segment illustrates the previous research regarding privacy preservation methods in IoT environment and is summarised in Table 1 below.

Muhammad Usman *et al.,* [1] investigated an encryption algorithm, called SIT, which is a lightweight 64-bit block cipher requiring a 64-bit key for data encryption. The algorithm combined feistel and uniform substitution-permutation network architectures and was capable of providing significant security in just five encryption rounds, as evidenced by simulation results. The algorithm was also implemented on an inexpensive 8-bit micro-controller, and the resulting memory utilization, encryption/decryption, and code size execution cycles were evaluated with those of established

benchmark encryption algorithms. Kedir Mamo Besher *et al.,* [2] conducted research on the use of IoT networks for remote medical treatment and identified safety hazards associated to unsafe data transmission, particularly among IoT sensor devices and network routers. They presented a secure solution that involved implementing an encryption algorithm within the sensor device itself, which ensured that patient health data was encrypted before being transmitted. The researchers verified their proposal through a proof-of-concept study, which involved two levels of encryption at the sensor device level and two levels of decryption at the doctor's office. The results were very promising, and indicated that this approach could provide an end-to-end security solution for healthcare data transmission via IoT networks.

Balasubramanian Prabhu Kavin and Sannasi Ganapathy [3] have developed a technique that ensures safe storage of user data in cloud databases using a data storage mechanism based on the CRT. Along with that, they introduced a new group key management scheme using CRT to provide secure access to encrypted data in cloud databases. Their secure storage scheme includes two encryption methods, both with innovative formulas for the first and second encryption steps, as well as a unique decryption formula for cloud data. The group key generation formula, introduced in their research, grants secure access to encrypted cloud data on a cloud server. They evaluated the performance of their security models through experimental analysis. They proposed that a new lightweight encryption and decryption schemes for reducing the computational complexity over the Cloud and IoT-based applications should be developed. On the other hand, Ntebatseng Mahlake *et al.,* [4] developed a new security algorithm called LSA. It is a combination of the SPINS with the SIT encryption method. The primary objective behind designing LSA was to enhance data security in WSNs while minimizing power consumption. It ensured an improved security level of 99% by significantly reducing the key generation time by 102mS. Moreover, during data transmission, it reduced power consumption by an average of 411.2uJ and lowered the PDR from 90 to 99%, surpassing other techniques like SPN and Feistel.

Khalid Haseeb *et al.*,[5], have presented a protocol, LSDAR, for Next-generation Sensor Networks integrated with IoT, to enhance energy routing performance and node-level data protection against malicious threats. The protocol clusters network nodes independently based on varying radiuses, which helps to avoid energy holes across the base station. LSDAR uses an efficient and loop-free routing path construction according to A-star heuristics algorithm. End-to-end communication links are secured alongside malevolent nodes with the help of an OTP encryption scheme. The simulation outcomes prove significant enhancements in packet drop ratio, end-to-end delay, network lifetime, and energy consumption over state-of-the-art techniques. Furthermore, A. Biswas *et al.,* [6] introduced a lightweight encryption method, named LRBC, for resource-constrained IoT devices, to offer data security at the sensing level. The LRBC algorithm combines the structural benefits of both SPN and Feistel structure to enhance security. The presented algorithm has been implemented on NEXYS 4 DDR FPGA trainer kit (Artix-7) and utilized for the ASIC chip on TSMC 65nm technology. A comprehensive security analysis revealed that the LRBC algorithm is robust in contradiction with various attacks with high strength. Besides, the average avalanche effect of LRBC is 58% and 55.75% for plaintext and key.

In their study, G. Kalyani and Shilpa Chaudhari [7] developed methods to enhance IoT security authentication using cryptography. They focused on securing sensitive IoT data using OHE with high reliability. Firstly, they classified sensitive data using DNN structure. Later, OHE encrypted sensitive data during the encryption and decryption process. The SFF optimization algorithm was employed to authenticate the key and select the best key during encryption, providing the highest privacy-preserving data in IoT. This approach generated an encrypted key, thus achieving maximum key breaking time, with less computational time. Their presented IoT security model was evaluated through experimental analysis, which demonstrated its effectiveness. Additionally, S. Medileh *et al.,* [8] introduced an encryption technique, called FlexenTech, to guard IoT data during storage and transit. This method was scalable and compatible with resource-constrained devices and networks, offering low encryption time and defense against common attacks, including replay attacks. FlexenTech defined a configurable mode that allowed variable numbers of rounds or key sizes. The experimental investigation demonstrated that FlexenTech was robust and provided various benefits, such as multiple configurable confidentiality levels, during resource-constrained device operations. Moreover, FlexenTech reduced encryption computation time by up to 9.7% compared to the best rivals in the literature.

Table 1: Summary of Literature Survey

Author & Year	**Findings**	**Journal**	**Gap**
Muhammad Usman *et al., 2017* [1]	The implementation shows promising results, making the algorithm a suitable candidate to be adopted in IoT applications.	arXiv preprint arXiv:1704.0868, 2017.	It failed to perform detailed evaluation for different possible attacks.
Kedir Mamo Besher *et al., 2020* [2]	Test results are promising for an end-to-end security solution of healthcare data transmission in IoT.	IEEE Sensors Journal, vol. 21, no. 10, pp: 11977-11982, 2020.	The cost of overall system is high.
Balasubramanian Prabhu Kavin and Sannasi Ganapathy [3]	They developed a technique that ensures safe storage of user data in cloud databases using a data storage mechanism based on the CRT.	Computer Networks 151, pp: 181-190, 2019	It did not to consider the fitness and memory.
Ntebatseng Mahlake *et al.,* 2023 [4]	During data transmission, The power consumption is reduced.	Journal of Communication, vol. 18, pp: 47-57, 2023.	Computational or execution time is high.
Khalid Haseeb *et al., 2020* [5]	The simulation results indicate significant improvements with the comparison of the state-of-the-art in terms of energy consumption, network lifetime, an end-to-end delay and packet drop ratio.	Sustainable Cities and Society, vol. 54, pp. 101995, 2020.	It does not consider the fitness function.
A. Biswas *et al., 2020* [6]	Security analysis shows that the scheme guarantees high security with a balanced area and power consumption.	Journal of Ambient Intelligence and Humanized Computing, pp: 1-15, 2020.	It failed to consider the fitness and memory.
G. Kalyani and Shilpa Chaudhari 2020 [7]	The outcome shows that the performance of the IoT security approach achieves maximum key breaking time and less computational time with high security.	International Journal of Computers and Applications, vol. 42, no. 3, pp: 306-314, 2020.	Under extensive scale count situation, it is tedious.
S. Medileh *et al.,2020* [8]	A security analysis of the technique shows its configuration flexibility and feasibility in terms of limited resource consumption and low execution time.	Ad Hoc Networks, vol. 103, no. 1, pp: 102240, 2020.	It does not adapt with various levels of security.

2.1 Challenges

Based on the challenges in the strategies mentioned in the preceding segment as shown in Table 1 above, the identified research gaps are listed below:

The approach developed in [1] is designed specifically for IoT devices, which typically have limited processing power, memory, and energy resources. It is optimized to provide robust security while minimizing the impact on system performance. However, it lacks a comprehensive evaluation of its performance for different possible attacks. Even though the approach presented in [3] has built-in error tolerance capabilities, the computation complexity of the approach is high. Similarly, the method in [4] indicates that the computational or execution time is higher during the encryption and key expansion procedure. Additionally, the method developed in [8] does not address varying levels of security requirements.

3. IoT System model

The system model of IoT as depicted in Figure 1 involves a series of interconnected devices capable of exchanging data through the internet. IoT devices collect and transmit data to the internet, where it undergoes processing and analysis to derive useful insights. To protect user privacy, the system incorporates robust privacy protection mechanisms to make sure that only authorized users have the right to use the data. Once the data is deemed secure, it

is then stored in highly secure storage facilities, preventing unauthorized access by hackers or any malicious actors. Therefore, IoT systems must incorporate state-of-the-art encryption solutions to safeguard user privacy and data integrity.

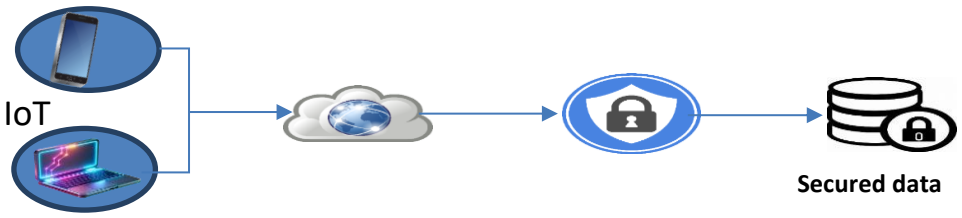

Figure 1. IoT system model

4. Proposed sensitive data protection model using ISOA-based encryption

The aim of the study is to develop an approach to provide security to the sensitive data in IoT. The schematic view of the proposed privacy preservation scheme using ISOA and deep learning model for the data security in IoT is illustrated in Figure 2. Initially, the data is identified as sensitive or non-sensitive [7] using a deep learning technique and the sensitive data are subjected for the privacy preservation so that the time to execute the approach is reduced. Then, a mathematical model is developed to provide security to the sensitive data. The model uses various security operators, like encryption functions, hashing, Face-splitting product, secret keys, etc. Moreover, the secret keys to encrypt the sensitive data are generated using an optimization algorithm, named Improved Skill Optimization Algorithm (ISOA). ISOA is newly developed by modifying Skill Optimization Algorithm [9].

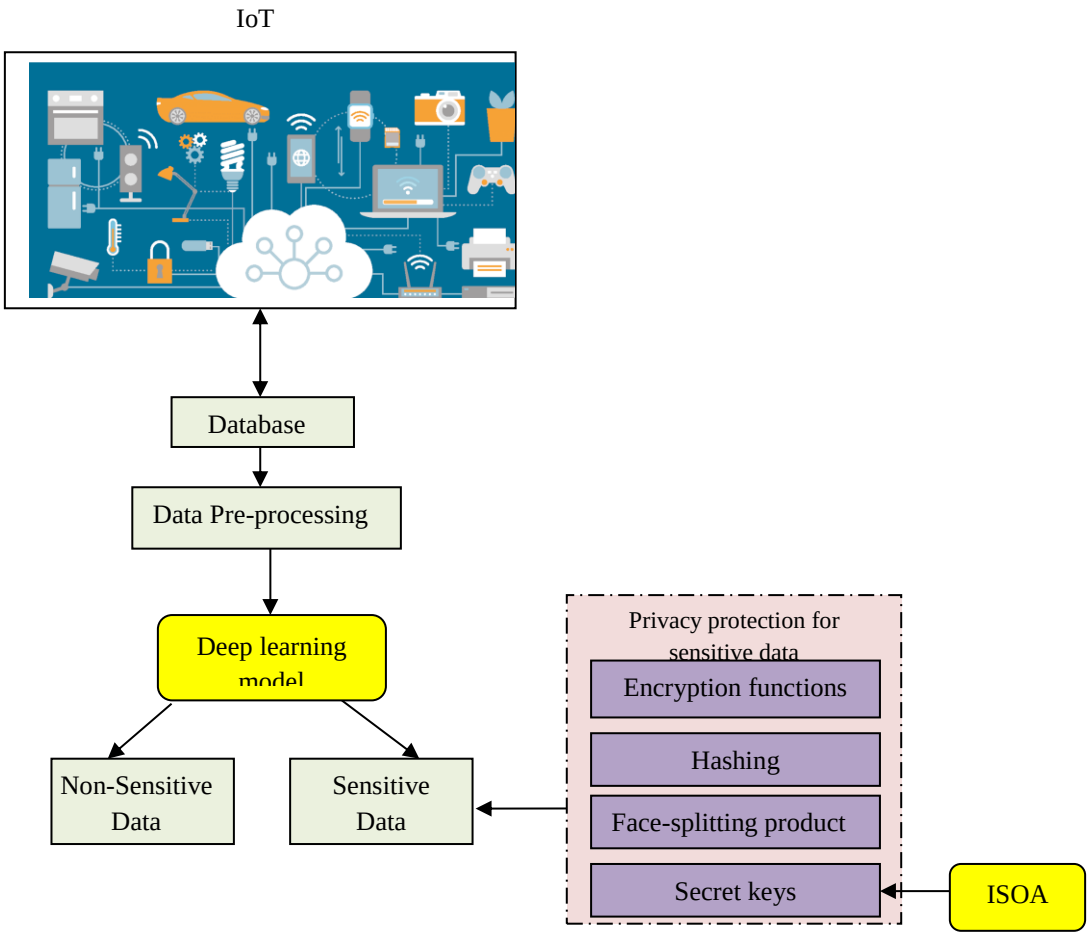

Figure 2. Schematic diagram of the presented ISOA-based encryption model for data security in IoT

4.1 Sensitive Data Identification

The process of identifying sensitive data involves searching through a large dataset to find information that may need to be protected due to its confidential nature. This can be a demanding task, particularly when handling with big data that contains various types of information. Deep learning models have the ability to analyse large datasets, identify patterns, and recognize complex relationships that may not be easily identifiable by humans. Training deep learning models on large volumes of sensitive data can enable them to accurately detect and classify sensitive information, thus improving data security and reducing the risk of data breaches. Sensitive data identification using deep learning has many benefits, including improving data security and reducing the risk of data breaches. It is particularly useful in industries such as healthcare, finance, and cyber security, where safeguarding sensitive data is essential. The approach can help organizations comply with regulatory requirements, protect customer privacy, and mitigate the risks of data breaches.

4.1.1 Pre-processing

Due to data loss, errors, unavailable measuring equipment, equipment failures, and other reasons, a dataset may contain missing values. This can result in biased data and lower the quality of the classification process. To address this issue, missing data imputation is a flexible and general procedure used to fill in missing values with plausible values to achieve precise classification. One commonly used method of missing data imputation is to substitute every single missing value with the mean value of other non-missing values in the same attribute. The formula for calculating the mean value for replacing missing values is expressed as,

$$mean(y) = \sum_{j-0}^{h_\tau} y_j \tag{1}$$

here, h_τ represents the entire dataset, while y_j refers to the data available in a particular row.

4.1.2 Classification using Deep Maxout Network

The preprocessed data is fed into a deep maxout network, which classifies the data as sensitive or non-sensitive. The deep maxout network is a type of DNN that has proven to be effectual in distinguishing complex patterns in data. Deep maxout networks [20] consist of multiple layers of maxout units, with each layer learning increasingly complex features from the input data. These networks are capable of accurately modeling complex functions and can be applied in various domains, including image and speech recognition, text classification, and natural language processing. One of the primary advantages of the deep maxout network is its capability to address the matter of overfitting, which is a widespread challenge in deep learning. Maxout units in the network use dropout regularization to prevent neuron co-adaptation and enhance generality potential. The network includes several layers, such as maxout function, dense layer with activation, maxpooling, dropout, convolution, embedding, and input. Deep maxout networks possess greater generalization ability compared to shallow networks or networks with a large number of parameters, making them better suited for generalizing from limited training data. Furthermore, these networks can be easily scaled to accommodate large and complex datasets. As the network depth increases, it becomes capable of modeling more abstract and higher-level features, which improves the input's understanding and interpretation. This aspect can be especially beneficial in scenarios where the data is noisy or scarce. The deep maxout network produces robust and accurate predictions while achieving state-of-the-art performance on various tasks [21]. Deep Maxout Networks (DMNs) use a flexible maxout activation function, offering higher expressiveness than fixed functions like ReLU. DMNs adapt well to complex patterns that CNNs and Transformers. Transformers dominate NLP and sequential tasks, whereas CNNs excel in vision applications, making them more practical choices for most domains. DMNs are however used despite their training challenges and resource demands. The maxout unit in this network is denoted by Equations (2) and (3):

$$Q(m) = \max_{l \in [1,\eta]} v_{wl,} \tag{2}$$

$$v_{wl} = m.\lambda_{wl} + \gamma_{wl} \tag{3}$$

here, m symbolizes the input dataset, the weight is denoted as λ, and γ represents a bias factor. The feature map is characterized by the symbol η.

4.2 Privacy Preservation of Sensitive Data

Sensitive data, counting but not restricted to personally identifiable information (PII) and medical records, must be preserved to protect individual privacy. This includes data encryption to minimize the risk of unauthorized access.

One effective way of preserving privacy is through differential privacy, which involves adding noise to data to prevent the identification of individuals while still providing useful insights from data analysis. Differential privacy algorithms ensure that the data cannot be traced back to any individual. Initially, the original sensitive data is subjected to data transformation, applied using box cox transformation.

Transformed matrix, $T_{a\times b} = B[D]$ (4)

where, $B[.]$ is the box cox transformation of the original data D, and $a \times b$ denotes the dimension of the data size.

Then, a Polynomial matrix is generated using chebyshev polynomial as

$$c = 4x^3 - 3x \quad (5)$$

where, $x_{a\times b} = L \bullet T$ and $\bullet$ is the face-splitting product.

The transformed matrix is encrypted using a key to generate another matrix,

$$L_{a\times b} = Y(T, K) \quad (6)$$

where, $Y(.)$ is the encryption function that uses ECC encryption. Here, to construct the key, the data is multiplied using a random number, followed by the summation of the elements in the row of the data.

$$K = \sum_{i=1}^{a} (D_i * r), \text{ where } r = [0,1] \quad (7)$$

Multiplying the polynomial matrix with the secret key, which is generated using improved skill optimization algorithm, a matrix of size $a \times b$ is designed as,

$$A_{a\times b} = c \times S_K \quad (8)$$

where, S_K is the key generated from improved skill optimization algorithm. Henceforth, the protected data formulation is given by,

$$P_j = A_j \times D_j \quad (9)$$

To retrieve the original data, the data requester uses a recovery key. A recovery key is a security feature that is used in various encryption systems to enable an authorized third party to recover data in the event of the loss of the original encryption key. The key point of the recovery key is to contribute a fail-safe mechanism for data recovery.

$$R_j = A_j \parallel S_K \quad (10)$$

Thus, the data can be retrieved using the key as,

$$D_j^* = \frac{P_j}{R_j} \quad (11)$$

4.2.1 Secret key generation using improved skill optimization algorithm

Secret key generation aims to address the issue of secure key exchange in modern cryptography. This is done using ISOA, which is proposed by improving SOA using DES. The algorithm optimizes the search space of possible keys to generate an optimal key that fits these requirements. This process continues until the best possible key is generated. Thus, ISOA algorithm is a powerful tool in cryptography and can be used to generate highly secure keys that can withstand attacks from external agents. It represents an important development in the field of cryptography, and has the potential to replace existing methods of key generation.

a) Solution Representation:

In the context of solution representation, an optimal key finder with dimension 1x1 refers to a method or algorithm that effectively identifies and selects the best possible solution from a given set of alternatives. The dimension 1x1

signifies that the key finder only considers a single variable or criterion in the decision-making process. This approach to solution representation can be useful in scenarios where there is a clear and well-defined primary objective or criterion that is crucial in determining the overall quality or success of a solution.

b) Objective function:

By providing information about the fitness values of better solutions, the algorithm can adapt its search strategy and explore more effectively. This helps ISOA to converge faster and find optimal or near-optimal solutions more efficiently, making it a powerful optimization algorithm for various applications. The objective function of ISOA considers two measures, namely Positive Information Disclosure (PID) and entropy, to assure the encrypted data quality. PID is a technique that involves utilizing information about the positive aspects of the population when evaluating the fitness function rather than just focusing on negative aspects. By using PID in the objective function, it is possible to generate secret keys that are not only highly secure but also meet other important criteria such as efficiency and performance. The technique is an important development in the field of cryptography and has the potential to further enhance the security of sensitive information.

$$\text{Fitness, } \underset{(\max)}{F} = \frac{PID + sh}{2} \tag{12}$$

where, PID is the Positive Information Disclosure [22] and sh is the shannon entropy. PID metric quantifies the extent to which the adversary's posterior probability is enhanced and identifies the secret that provides the greatest increase in probability D^*. which is,

$$PID = \sup_{D^* \in D} \frac{P(D^* / P) - P(d^*)}{P(d^*)} \tag{13}$$

where, $P(d)$ is the probability of each member in D. This technique is poised to make a massive impact in the field of information security and safeguard sensitive information from prying eyes.

Shannon entropy serves as the foundation for numerous metrics. Generally, it assesses the amount of uncertainty that is connected with predicting a random variable's value. As a privacy metric, it may be considered the effective magnitude of the anonymity set or the quantity of additional bits of information the adversary must possess to recognize a user. Then, the entropy of D can be expressed as:

$$sh = \frac{1}{1-\varphi} \log_2 \sum_{d \in D} P(d)^2 \text{ , such that } \varphi = 1 \tag{14}$$

c) Improved skill optimization algorithm

ISOA is an algorithm proposed as an enhanced version of SOA, which is widely used for optimizing different types of skills across various fields. This heuristic algorithm has proven to be highly effective in minimizing errors while maximizing performances, making it an essential tool for skill optimization in different applications. ISOA, derived from SOA, takes advantage of Double Exponential Smoothing (DES) to improve the algorithm's accuracy in forecasting relevant future trends. The Data Encryption Standard (DES) improves key generation by using a 56-bit key derived from a 64-bit input, where 8 bits are parity bits for error detection. It applies a key scheduling algorithm to generate 16 unique subkeys, one for each encryption round, ensuring variability and complexity. This process enhances security by introducing round-specific keys, making brute-force attacks significantly harder. By continuously learning how to optimize skills over time, this method has shown to be highly effective in making skillful predictions and directions on a test dataset, outperforming several other alternatives. One significant advantage of the improved SOA is its potential to optimize skill without the need for expert knowledge or experience and its ability to handle non-linear and noisy time-series data, allowing for a more accurate prediction and decision-making process. During the initial stage, each member of the SOA [18] endeavors to acquire a skill with the help of an expert community member's guidance. The individual's quality is proportional to the objective function value gained by that particular population member. The expert member assigned to guide an SOA member is determined based on their superior objective function value. However, one of these experts is randomly chosen to train the population member, implying that the selected expert may not essentially be the optimal candidate solution. The finest candidate solution is a permanent member of the expert set for all SOA members. The expert aids the population member in exploring different positions in the search space by acquiring the skill, thereby improving the algorithm's global search and

exploration capabilities. A new position for each population member is considered acceptable only if it advances the objective function value.

$$Z_q^{p1} : z_{q,V}^{p1} = z_{q,V} + u \times (E_{q,V} - n \times z_{q,V}), E_q = z_k \quad (15)$$

where, $k \neq q$

Here, Z_q^{p1} is the updated status of the n th candidate solution calculated during the first phase, whereas $z_{q,V}^{p1}$ denotes the V th dimension of the solution. The expert member E_q selected to guide and train the q th population member is represented by $E_{q,V}$ in its V th dimension. u is a random number in interval [0 − 1]. Furthermore, n is a randomly selected number from the set {1, 2}, and k is a number that is randomly selected from $\{1,2,...,N\}$.

$$Z_q^{p1} : z_{q,V}^{p1} = z_{q,V} + (u \times E_{q,V}) - (u \times n \times z_{q,V}) \quad (16)$$

$$Z_q^{p1} = z_{q,V}(1 - u \times n) + (u \times E_{q,V}) \quad (17)$$

From Double exponential smoothing (DES) [19],

$$Z(t+1) = a(t) + b(t) \quad (18)$$

$$Z(t+1) = \alpha Z(t) + (1-\alpha)S(t) + \beta[a(t) - a(t-1)] + (1-\beta)V(t-1) \quad (19)$$

$$Z(t+1) = \alpha Z(t) + (1-\alpha)S(t) + \alpha\beta Z(t) + (1-\alpha)\beta S(t) - \alpha\beta Z(t-1) - (1-\alpha)\beta S(t-1) + (1-\beta)V(t-1) \quad (20)$$

$$Z(t+1) = \alpha Z(t)[1+\beta] + (1-\alpha)S(t)[1+\beta] - \beta[\alpha Z(t-1) + (1-\alpha)S(t-1)] - [(1-\beta)V(t-1)] \quad (21)$$

$$Z(t) = \frac{Z(t+1) - [(1-\alpha)S(t)(1+\beta)] + [\beta(\alpha Z(t-1) + (1-\alpha)S(t-1))] - [(1-\beta)V(t-1)]}{\alpha(1+\beta)} \quad (22)$$

Assuming at the current iteration, the solutions of both DES and SOA are equivalent, i.e., $z_{q,V} = Z(t)$, equation (19) and (24) can be substituted,

$$Z_q^{p1}(t+1) = \left\{ \frac{z_{q,V}(t+1) - [(1-\alpha)S(t)(1+\beta)] + [\beta(\alpha z_{q,V}(t-1) + (1-\alpha)S(t-1))] - [(1-\beta)V(t-1)]}{\alpha(1+\beta)} \right\} (1 - u \times n) + (u \times E_{q,V}) \quad (23)$$

Thus, the update equation of ISOA is derived as,

$$Z_q^{p1}(t+1) = \frac{\alpha(1+\beta)}{\alpha(1+\beta) - (1+un)} \times \frac{1-un}{\alpha(1+\beta)}$$
$$\left\{ \left[\beta \left(\alpha Z_q^{p1}(t-1) + (1-\alpha)S(t-1) \right) \right] - [(1-\beta)V(t-1)] - [(1-\alpha)S(t)(1+\beta)] \right\} + (u \times E_{q.V}) \quad (24)$$

where, α, β, u were the values in interval [0,1], n were the values in interval [1,2], and $S(t)$ were the estimation of the value at time step $t+1$ made at the time t.

5. Results and discussion

5.1 Performance Measures

To assess the efficacy of the presented ISOA based encryption strategy, its computation time, memory usage and fitness function were analyzed, as detailed below:

i) Computation Time (sec): This specifies the duration needed by a system to complete the analysis procedure.
ii) Memory (MB): This pertains to the system's memory usage in processing data and ensuring authentication.
iii) Fitness Function: A fitness function is a mathematical function used in evolutionary algorithms to evaluate the quality of a potential solution. In a genetic algorithm, the fitness function determines how well a set of candidate solutions performs the task by assigning each of them a fitness score. The function calculates an objective value that represents how well the solution solves the problem.

5.2 Dataset Description

(i) BoT-IoT: The BoT-IoT dataset [23] is a publicly available dataset comprising of diverse types of IoT botnet attacks created in a simulated IoT atmosphere. It comprises of 5 different types of attack scenarios such as Mirai-like malware, HTTP-based command, DDoS attack, IoT protocol-based attack, and P2P-based attack. It also comprises of normal data traffic generated by different types of IoT devices. The dataset is publicly available and is intended to assist researchers in the development and evaluation of new intrusion detection and prevention techniques for IoT botnet attacks. The purpose of the dataset is to improve the security of IoT devices, which are known to be potential vulnerabilities in a network.

(ii) ToN-IoT: ToN-IoT dataset [24] comprises a protocol stack designed to enable secure and efficient communication for IoT devices. Created as the next generation of IoT and Industrial IoT (IIoT) datasets, ToN-IoT is instrumental in assessing the efficacy and fidelity of various cybersecurity applications based on AI. The dataset gathers heterogeneous data from Telemetry datasets of IoT and IIoT sensors, operating systems, such as Windows 7 and 10, Ubuntu 14, and 18 TLS, as well as Network traffic datasets. To facilitate data collection, the researchers developed a new testbed at the IoT lab that employed virtual machines, physical systems, hacking platforms, cloud and fog platforms, IoT and IIoT sensors. This helped to mimic the complexity and scalability of IIoT and Industry 4.0 networks effectively. Several normal and cyber-attack events from IoT networks were collected in parallel processing. The testbed utilized multiple virtual machines, including hosts of Linux, Kali and Windows Linux operating systems, to establish interconnections between three layers of IIoT, Cloud, and Edge/Fog systems. The initial statistical evaluation of the datasets demonstrated their effectiveness in evaluating cybersecurity applications such as threat intelligence, intrusion detection, adversarial machine learning, and privacy-preserving models. ToN-IoT is an innovative testbed that collects telemetry, operating system and network data to create new IIoT-developed datasets.

5.3 Performance Analysis

Figure 3 presents the performance assessment of the suggested ISOA-based encryption method in terms of accuracy, sensitivity, and specificity based on the analysis of the training data. The results indicate that the technique obtained an accuracy rate of 0.9335 with 70% of the training data. When trained with 80% of the data, the method had a sensitivity rate of 0.9630. Additionally, the method obtained a specificity rate of 0.899 when trained with 90% of the data.

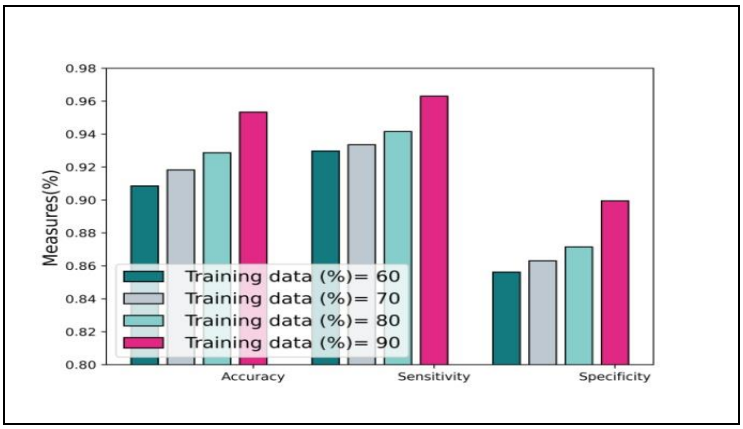

Figure 3. The performance assessment of the suggested ISOA based encryption method based on the training data in BoT-IoT

Figure 4 shows the performance assessment of the suggested approaches in terms of sensitivity, specificity, and accuracy using ToN-IoT and based on the investigation of training data. The presented ISOA-based encryption model achieved an accuracy rate of 0.9359 with 80% of the training data. The sensitivity of the method was found to be 0.9376 with 70% of the training data. Moreover, when trained with 90% of the data, the specificity of the method was found to be 0.8904.

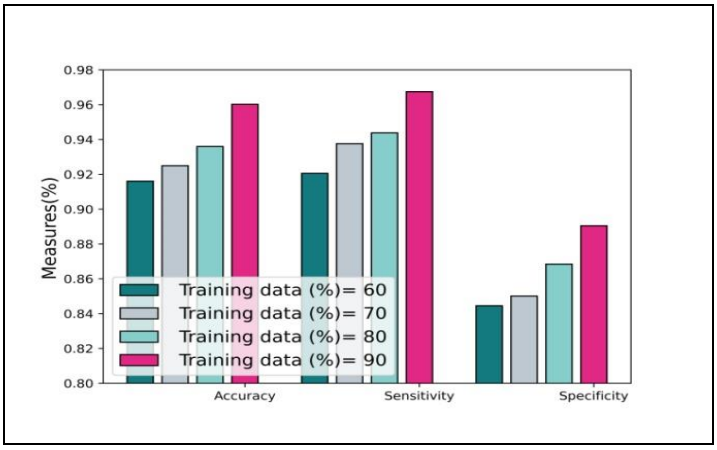

Figure 4. The performance assessment of the presented ISOA based encryption method based on the training data in ToN-IoT

Figure 5 depicts the training accuracy and training loss of the proposed ISOA-based encryption model in a graphical format. In Figure 5 (a), the training accuracy of the model is illustrated. The strategy obtained a training accuracy of 0.9278 at iteration 70 and a validation accuracy of 0.9186. Additionally, at iteration 100, the model achieved a training accuracy value of 0.9278 and a validation accuracy of 0.9247. Figure 5 (b) shows the training loss of the suggested technique. The model obtained a training loss of 0.2546 and a validation loss of 0.2885 at the 100th iteration.

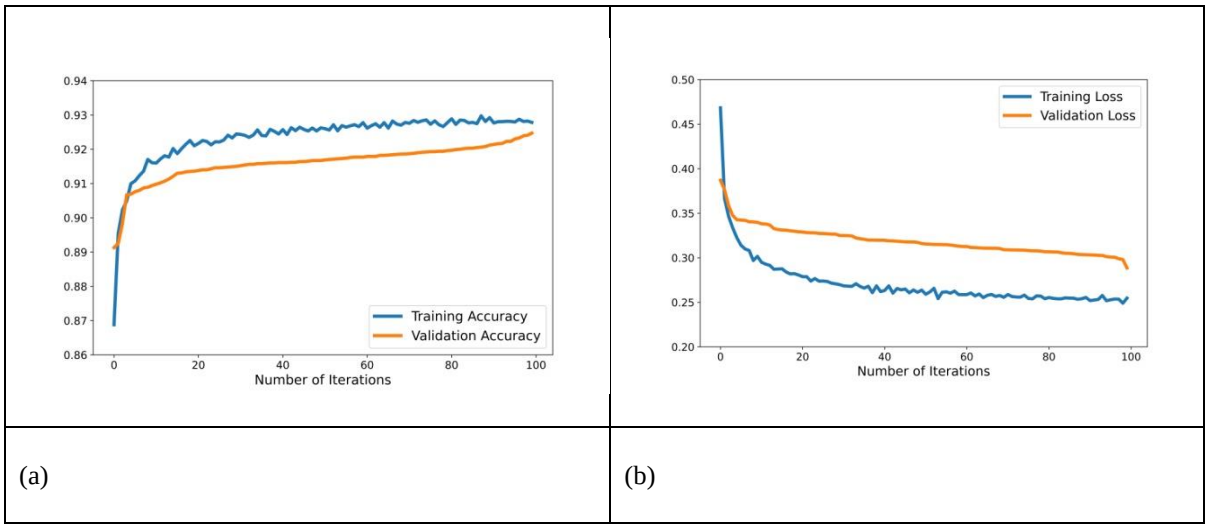

Figure 5. The performance examination of the presented ISOA based encryption technique based on the (a) training accuracy and (b) training loss

Figure 6 represents the graphical illustration of the training loss and training accuracy using ToN-IoT. Figure 6 (a) exhibits the training accuracy of the suggested ISOA based encryption method. The presented approach attained a training accuracy of 0.9321 in iteration 70 and a validation accuracy of 0.9231. Moreover, the proposed model attained 0.9333 as the training accuracy for iteration 100 and the validation accuracy is 0.9264. Figure 6 (b) illustrated the training loss of the presented ISOA-based encryption technique. The presented approach attained a training loss of 0.2340 and a validation loss of 0.2799 of 100th iteration.

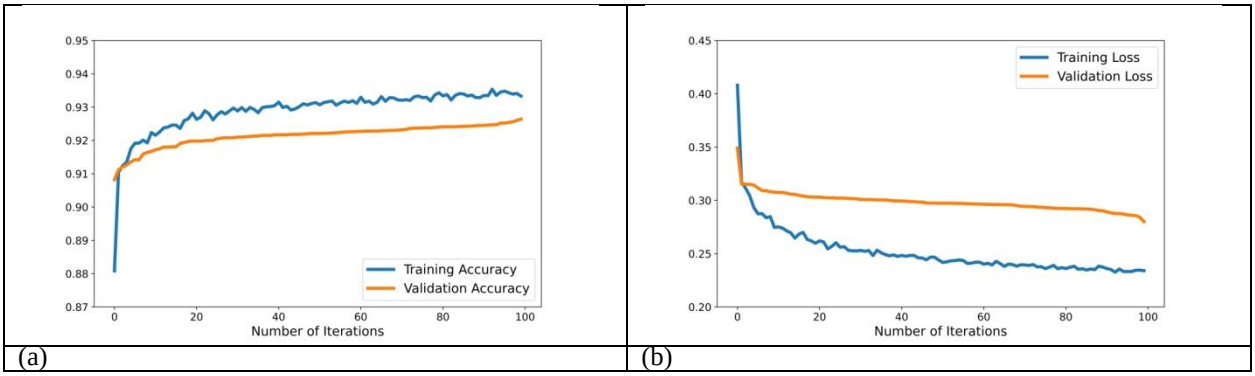

Figure 6. The performance examination of the presented ISOA based encryption approach according to the (a) training accuracy and (b) training loss in ToN-IoT

5.4 Comparative Analysis

The performance of the examined ISOA-based encryption model is compared with existing methods, like LSA [4], LRBC [6], and SFF optimization algorithm [7] to show its supremacy.

A comparative investigation of the suggested technique and other methods is illustrated in Figure 7. Figure 7 (a) demonstrates the comparative study of dissimilar techniques according to memory usage. The proposed ISOA-based encryption model obtained a memory usage of 0.4136 MB for a file size of 90, while the LSA, LRBC, and SFF optimization algorithm achieved memory values of 0.6535 MB, 0.5750 MB, and 0.5168 MB, respectively. Figure 7 (b) shows the comparative analysis based on computational time. The introduced ISOA-based encryption method obtained a computational time of only 913.17 sec for an iteration of 120. In contrast, the LRBC and SFF optimization algorithm achieved computational times of 1385.94 sec and 1231.38 sec, respectively. Figure 7 (c) presents the comparative evaluation according to fitness. The suggested ISOA-based encryption technique gained a higher fitness value of 1.3249 for an iteration of 90, while the LSA and LRBC achieved fitness standards of 0.9791 and 1.0650, respectively.

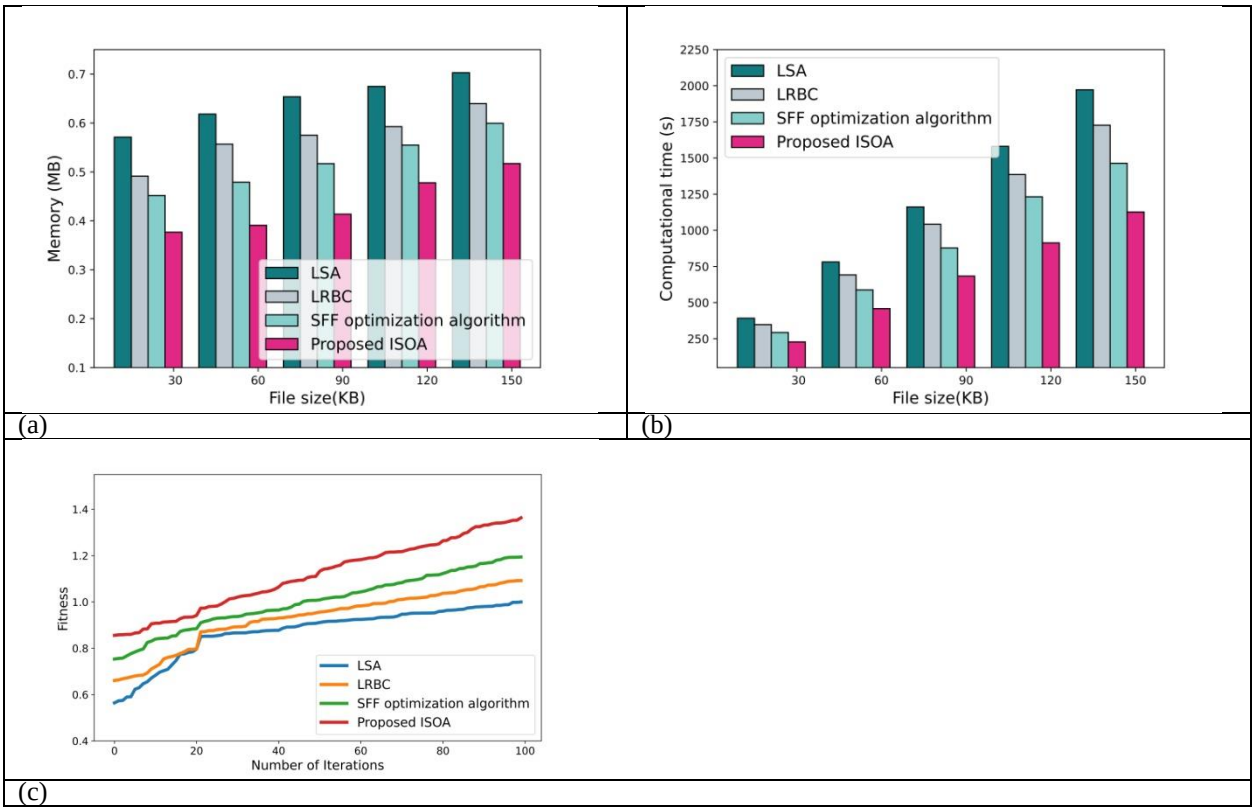

Figure 7. Comparative analysis of the proposed model based on BoT-IoT, (a) memory, (b) computational time, (c) fitness function

A comparative investigation of the recommended technique with other techniques is presented in Figure 8. Figure 8 (a) depicts the comparative study of dissimilar methods based on memory, with the projected ISOA based encryption model achieving a memory of only 0.9351 MB for the file size of 20. By contrast, the LSA, LRBC, and SFF optimization algorithm attained the memory values of 1.2625 MB, 1.1495 MB, and 1.07424 MB. Figure 8 (b) illustrates the comparative evaluation according to the computational time, enlightening that the examined ISOA based encryption approach attained a less computational time of 2088.97 sec for the iteration of 120. Meanwhile, the LRBC, and SFF optimization algorithm achieved the computational time of 2733.8 sec, and 2333.8 sec. Figure 8(c) elucidates the comparative examination related with fitness, enlightening that the presented ISOA based encryption model reached a higher fitness value of 1.3234 for iteration of 90. By contrast, the LSA, and LRBC reached the fitness standards of 0.9585, and 1.0525.

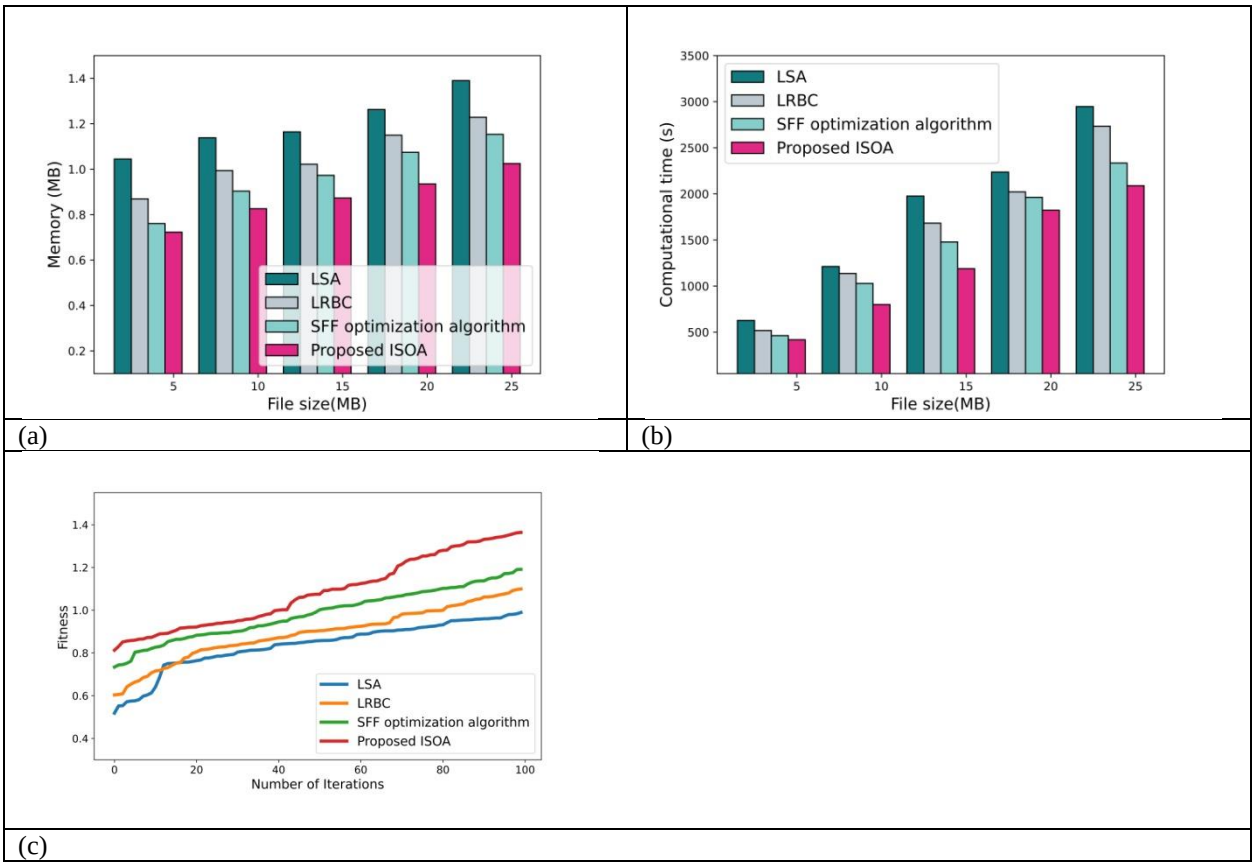

Figure 8. Comparative analysis of the proposed model based on ToN-IoT, (a) memory, (b) computational time, (c) fitness

In Table 2, a comparison is given among the presented ISOA based encryption strategy and other approaches based on memory, computational time, and fitness for BoT-IoT and ToN-IoT. The examination expose that the ISOA based encryption strategy achieved the fewest memory usage, least computation time and highest fitness value. Compared to the LSA model, the suggested technique exhibits a 35% less usage of memory. Additionally, the proposed ISOA based encryption model uses 29.9% faster computational time than the SFF optimization algorithm. The presented method attained a fitness of 19.8% when compared to LRBC.

Table 2: Comparative analysis

Measures	LSA	LRBC	SFF optimization algorithm	Proposed ISOA-based encryption model
Memory	0.7026	0.6397	0.5994	**0.5170**
Computation Time (sec)	1972.12	1727.16	1463.34	**1126.47**
Fitness function	0.9998	1.0920	1.1939	**1.3630**

6. Conclusion and future work

This work developed a model for protecting data in IoT by identifying the sensitive data using deep learning. Deep learning-based sensitive data detection can be more accurate and efficient than traditional methods, as the models can learn to recognize patterns and features in data that may be difficult for humans to spot. This approach has become increasingly important in today's data-driven world, where protecting sensitive information is critical for privacy and security reasons. The ISOA-based encryption model is designed to improve the security of data in the IoT environment. This method has proven to be effective and has outperformed existing methods because; the keys are generated using the ISOA algorithm, which ensures that the keys are unpredictable and secure. The algorithm also optimizes the number of encryption keys, as compared to other encryption algorithms, thus reducing the risk of successful attacks. Moreover, the ISOA-based encryption model provides high scalability, as IoT devices can generate and exchange encryption keys efficiently. Furthermore, it is lightweight and fast, making it suitable for resource-constrained IoT devices. The combination of Deep Maxout Network and secure encryption make certain the confidentiality and integrity of sensitive data. The implementation of this method can contribute significantly to the emergence of secure and reliable IoT systems, particularly in the areas where sensitive data is involved. Overall, the proposed method shows promising results and sets the stage for more advanced research in this area. Results indicate that proposed ISOA based encryption model accomplished superior performance, with a memory of 0.5170 MB, computational time of 1126.47 sec and fitness value of 1.3630. Future work can be done in this direction by exploring real-time application of this model or further optimizing the ISOA for various IoT environments.

Declarations of competing interest

The authors declares that they had no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Availability of data and materials

The dataset for the work is freely available online at https://research.unsw.edu.au/projects/bot-iot-dataset

Funding

This paper has not received any specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

CRediT authorship contribution statement

Mathias Agbeko: Conceptualization, Writing - original draft, Methodology, Formal analysis, Investigation, Resources, Visualization. **Disha Handa:** Writing - Reviewing and Editing, Software, Supervision, Data curation, Validation.

Acknowledgement

The authors of the paper sincerely acknowledge anonymous reviewers who reviewed this manuscript and provided constructive feedback and comments.

List of Abbreviations

Security Protocol for Sensor Networks	SPINS
Secure IoT	SIT
Light-weight structure based Data Aggregation Routing	LSDAR
Substitution-Permutation Network	SPN
Application-Specific Integrated Circuit	ASIC
Optimal Homomorphic Encryption	OHE
Deep Learning Neural Network	DNN
Flexible Encryption Technique	FlexenTech
Chinese Remainder Theorem	CRT
Lightweight Security Algorithm	LSA
Packet Drop Ratio	PDR

References

[1] M. Usman, I. Ahmed, M. I. Aslam, S. Khan, and U. A. Shah, "SIT: a lightweight encryption algorithm for secure internet of things," *arXiv preprint arXiv: 1704.08688*, 2017.

[2] K. M. Besher, Z. Subah, and M. Z. Ali, "IoT sensor initiated healthcare data security," *IEEE Sensors Journal*, vol. 21, no. 10, pp. 11977–11982, 2020.

[3] S. Ganapathy, "A secured storage and privacy-preserving model using CRT for providing security on cloud and IoT-based applications," *Computer Networks*, vol. 151, pp. 181–190, 2019.

[4] N. Mahlake, T. E. Mathonsi, D. Du Plessis, and T. Muchenje, "A lightweight encryption algorithm to enhance wireless sensor network security on the Internet of Things," *Journal of Communication*, vol. 18, pp. 47–57, 2023.

[5] K. Haseeb, N. Islam, T. Saba, A. Rehman, and Z. Mehmood, "LSDAR: A light-weight structure based data aggregation routing protocol with secure internet of things integrated next-generation sensor networks," *Sustainable Cities and Society*, vol. 54, p. 101995, 2020.

[6] A. Biswas, A. Majumdar, S. Nath, A. Dutta, and K. L. Baishnab, "LRBC: a lightweight block cipher design for resource constrained IoT devices," *Journal of Ambient Intelligence and Humanized Computing*, pp. 1–15, 2020.

[7] G. Kalyani and S. Chaudhari, "An efficient approach for enhancing security in Internet of Things using the optimum authentication key," *International Journal of Computers and Applications*, vol. 42, no. 3, pp. 306–314, 2020.

[8] A. Laouid et al., "A flexible encryption technique for the Internet of Things environment," *Ad Hoc Networks*, vol. 103, p. 102240, 2020.

[9] H. Givi and M. Hubalovska, "Skill optimization algorithm: A new human-based metaheuristic technique," *Computers, Materials & Continua*, vol. 74, no. 1, 2023.

[10] Z. Iftikhar et al., "Privacy preservation in resource-constrained IoT devices using blockchain—A survey," *Electronics*, vol. 10, no. 14, p. 1732, 2021.

[11] D. Torre, A. Chennamaneni, and A. Rodriguez, "Privacy-preservation techniques for IoT devices: A systematic mapping study," *IEEE Access*, 2023.

[12] M. Terrovitis and N. Mamoulis, "Privacy preservation in the publication of trajectories," in *Proceedings of the Ninth International Conference on Mobile Data Management (MDM 2008)*, pp. 65–72, IEEE, 2008.

[13] R. Manjula and R. Datta, "A novel source location privacy preservation technique to achieve enhanced privacy and network lifetime in WSNs," *Pervasive and Mobile Computing*, vol. 44, pp. 58–73, 2018.

[14] R. P. Priyadarsini, M. L. Valarmathi, and S. Sivakumari, "Gain ratio based feature selection method for privacy preservation," *ICTACT Journal on Soft Computing*, vol. 1, no. 4, pp. 201–205, 2011.

[15] N. Li, N. Zhang, S. K. Das, and B. Thuraisingham, "Privacy preservation in wireless sensor networks: A state-of-the-art survey," *Ad Hoc Networks*, vol. 7, no. 8, pp. 1501–1514, 2009.

[16] M. Keshk, N. Moustafa, E. Sitnikova, and G. Creech, "Privacy preservation intrusion detection technique for SCADA systems," in *Proceedings of the 2017 Military Communications and Information Systems Conference (MilCIS)*, pp. 1–6, IEEE, 2017.

[17] M. Terrovitis, J. Liagouris, N. Mamoulis, and S. Skiadopoulos, "Privacy preservation by disassociation," *arXiv preprint arXiv: 1207.0135*, 2012.

[18] G. Dhiman, M. Garg, A. Nagar, V. Kumar, and M. Dehghani, "A novel algorithm for global optimization: Rat Swarm Optimizer," *Journal of Ambient Intelligence and Humanized Computing*, vol. 12, no. 6, pp. 6125–6146, 2021.en.wikipedia.org+1en.wikipedia.org+1

[19] Z. Guo, S. Malakooti, S. Sheikh, C. Al-Najjar, and B. Malakooti, "Multi-objective OLSR for proactive routing in MANET with delay, energy, and link lifetime predictions," *Applied Mathematical Modelling*, vol. 35, no. 3, pp. 1413–1426, 2011.

[20] J. Peta and S. Koppu, "An IoT-based framework and ensemble optimized deep maxout network model for breast cancer classification," *Electronics*, vol. 11, no. 24, p. 4137, 2022.

[21] Y. Miao, F. Metze, and S. Rawat, "Deep maxout networks for low-resource speech recognition," in *Proceedings of the IEEE Workshop on Automatic Speech Recognition and Understanding*, pp. 398–403, 2013.

[22] I. Wagner and D. Eckhoff, "Technical privacy metrics: a systematic survey," *ACM Computing Surveys (CSUR)*, vol. 51, no. 3, pp. 1–38, 2018.

[23] The BoT-IoT Dataset, "https://ieee-dataport.org/documents/bot-iot-dataset," accessed on Oct. 2023.

[24] The ToN_IoT Datasets, "https://ieee-dataport.org/documents/toniot-datasets," accessed on Oct. 2023.

[25] M. Abadi et al., "Deep learning with differential privacy," in *Proceedings of the ACM SIGSAC Conference on Computer and Communications Security*, pp. 308–318, 2016.

[26] G. E. Dahl, D. Yu, L. Deng, and A. Acero, "Context-dependent pre-trained deep neural networks for large-vocabulary speech recognition," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 20, no. 1, pp. 30–42, 2011.