Journal of Information Systems Engineering and Management
2025, 10(10s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

A Cutting-Edge Security Solution System for Smart Home by Applying the Intelligent AI And BC Framework to Secure Data

*Sameena Shaik[1], Sangeetha Komandur[2]
[1,2]Department of Computer Science, College of Engineering and Computer Science,
Jazan University, Jazan
SaudiArabia
*[*1]sabdualoheed@jazanu.edu.sa, [2]skomandur@jazanu.edu.sa*
*Corresponding author - [*1]sabdualoheed@jazanu.edu.sa*

ARTICLE INFO

Received: 02 Nov 2024

Revised: 26 Dec 2024

Accepted: 04 Jan 2025

ABSTRACT

Revolutionary technologies like blockchain and artificial intelligence (AI) have come together due to the fast Internet of Things (IoT) growth. Thanks to these technologies coming together, complex infrastructures like "smart homes," "smart cities," and "smart industries" have become possible to build. While IoT gadgets may provide excellent connection and convenience, they often use insecure conventional network interfaces. These older communication protocols and interfaces may be vulnerable if they aren't securely upgraded and protected. This research has developed a strong framework that utilizes the AI along with blockchain technology capabilities to work with the security concerns related to smart home systems (SHS). To start, the research used an isolation forest (IF) algorithm with random segmentation, anomaly score computation, route length, as well as thresholding phases to get rid of the weird data in a normal dataset for SHS. To further categorize the data as either attack or non-attack, the dataset is then used to train classification algorithms including Catboost, K-nearest neighbors (KNN), support vector machines (SVM), and linear discriminate analysis (LDA). When protecting sensitive information from data manipulation assaults, it is also stored in an interplanetary file system (IPFS). In order to save non-attack data safely, IPFS functions as an onsite storage system; the generated hash is then sent to the immutable register of the blockchain. Different performance metrics were used to assess the proposed framework.

Keywords: Blockchain, Artificial Intelligence, K-Nearest Neighbors, Linear Discriminate Analysis, Internet of Things

1 INTRODUCTION

In the next 30 years, the global use of renewable energy is predicted to increase by 147%. 2019 saw roughly 10 times as much money spent globally on renewable energy as it did in 2004. Furthermore, from 5.2% in 2007 to 13.4% in 2019, the proportion of renewable electricity in the world's energy production has grown [1]. In order to meet these two needs, two cutting-edge technologies have been developed: AI and blockchain. AI enables the best possible operational management of power systems, while the blockchain offers decentralized energy market trading platforms. The goal of this study is to discuss how to use AI and blockchain technology in smart grids to enable prosumers to trade energy [2].

Smart grids (SGs) are designed to replace conventional grids that rely heavily on fossil fuels with distributed energy resources (DERs). They do this by combining several current and new technologies, such as digital communications and information, to handle a multitude of processes [3]. This study offers a comprehensive overview to enhance energy management systems employing AI techniques. In particular, in the next few years, building energy management systems will need to be improved. These advancements will be largely attributed to the role that AI methods play [4].

In this context, the literature has used blockchain knowledge, which has the automation characteristics, immutability, irreversibility, decentralization, consensus, and security, to address the issues that centralized IoE architecture is now facing. Furthermore, another significant worry for Internet of Energy (IoE) technology is the security and privacy problems resulting from centralization [5]. This study suggests a safe energy trading system for residential properties based on blockchain technology. To choose miners and create blocks, a proof-of-computational closeness (PoCC) consensus mechanism is given. Additionally, an analytical energy pricing strategy aims to address the issue with current energy pricing strategies in a distributed trading environment [6].

In this article, an AI and blockchain-based IoT architecture is proposed, showcasing the combination of both technologies for IoT applications. Both qualitative and quantitative measurements are used to assess the performance of the proposed architecture [7]. Block IoT Intelligence: An AI-powered blockchain-enabled intelligent IoT architecture that offers an effective means of fusing blockchain as well as AI for IoT with existing methods [8].

The remaining tasks will be carried out using surveys. Part II offers a synopsis of a number of recent and ongoing projects. Section III defines the proposed approach. Section IV provides a summary of the results and analysis, and then the sources are given.

2 RECENT WORKS FOR RESEARCH

This research review some of the most current publications on IoT-based SHS in this section. Discover all the information you need in Table 1, which summarizes and identifies the pros and cons of the AI for categorization creation.

Paper and Author	Method	Advantage	Limitation
Kumari et al. [9]	Blockchain (BC) and AI	Integrate the BC skill along with AI in the ECM system	highlight the research issues of the BC-AI-based ECM system
Yang et al. [10]	IoT	develop a blockchain-based transactive energy management system	evaluate the feasibility and performance
Van Cutsem et al. [11]	Blockchain	advanced decentralized mechanisms that balance distributed supply as well as demand	Considers only Smart-Buildings
Ahmed et al. [12]	Blockchain and AI	advanced AI-based technologies and approaches, like, machine learning and deep learning	Does not work for sustainable IoT applications
Khattak et al. [13]	IoT	various new methods have been devised to meet modern society's electricity needs	Smart contracts of hyperledger fabric specify the permissions and resident's access for a dynamic price
Ali et al. [14]	intelligent energy management systems	various appraches along with new solutions proposed for managing the energy resources intelligently	No real tune
Pandiyan et al. [15]	smart energy management	Provides based on their applications.	smart cities and delivers valuable insights for researchers, industry professionals, along with policymakers working towards a more sustainable future

Kumari et al. [9] discuss the SG system has faced many challenges in recent years, the massive expansion of distributed energy generation (EG) with renewable energy sources (RES), the widespread of IoT devices, the emergence of security pressures, and the objective of maintaining the SG efficiency, stability, as well as reliability. The energy cloud management (ECM) system, which integrates energy infrastructure with intelligent energy consumption as well as value-added services based on customer demand, was developed to address these problems. Secure data transfer and effective demand-side forecasting are essential for achieving objectives. The challenges of energy management make it very important to use BC and AI to discover sustainable solutions.

Yang et al. [10] provided a blockchain-based transactive energy management system for IoT-enabled smart homes is developed. A comprehensive range of choices for smart houses to engage in transactive energy. In order to reduce grid load, smart houses may engage in vertical transactions with the grid, such as adding more solar energy to the system as well as offering demand response services. Peer-to-peer energy trading is one example of the horizontal transactions that smart homes may carry out in conjunction with other peers.

Van Cutsem et al. [11] described the renewable energy sources (RES), this study proposes a decentralized cooperative distributed resource (DR) architecture for managing daily energy exchanges. By allowing participants to choose a community power profile for the day ahead, the suggested algorithm takes advantage of their flexibility and guarantees prediction tracking for the next day. In actuality, the algorithm is completely decentralized thanks to BC technology, which also ensures autonomous monitoring along with payment by smart contracts and provides a reliable channel of communication for participants.

Ahmed et al. [12] applied to better understand how blockchain technology as well as AI are combining to create intelligent and sustainable IoT applications, this paper will look at this convergence. Our major discussion topic was how blockchain technology may help improve and establish sustainable IoT applications. A smart as well as sustainable conceptual framework based on the conversation, which makes use of cloud computing, IoT devices, also AI to analyze and gather relevant data.

Khattak et al. [13] detailed the electric cars and renewable energy resources (RER). Large data saves, data loss, manipulation, and modification are among the current issues. Additionally, it does away with the need for middlemen. Its distributed architecture and inherent security make it an excellent choice for enhancing the services as a whole. Upon execution, the smart contract's rules are automatically enforced.

Ali et al. [14] discussed the topic of intelligent energy management systems (IEMSs) has seen tremendous advancements over the last ten years, with new ideas and approaches put forward for intelligent resource management. Unexplored is a crucial problem that is linked to achieving the intended results: how to extract valuable insights from the dearth of scholarly literature in the era of digital publication. By turning the limited literature into visual presentations, this research suggests a unique approach to methodically reviewing the relevant studies in order to lessen the problem.

Pandiyan et al. [15] described the developments in technology that lead to smart energy management. There is a lack of evidence from the aforementioned methods that smart home security systems can be enhanced via the integration of AI and blockchain. This research proposes the following additions as a result of the aforementioned articles on blockchain and anomaly detection implementation in SHS.

- A secure framework with AI and BC to combat network-related attacks. There is an urgent need for strong security measures and cutting-edge security solutions since these systems are far more vulnerable to cyber threats and illegal access.
- The proposed study uses the typical SHS data set to train AI classifiers including Catboost, KNN, SVM, and LDA to categorize attack as well as non-attack data.
- To exclude erroneous information from the initial dataset for the SHS, this work uses anomaly-detection algorithms like IF and local outlier factor (LOF) before classification.
- To address concerns about data integrity, we also implemented the Ethereum blockchain, which is based on the IPFS protocol. For safe data storage, here is where AI classifiers' non-attack data may reside.

- The non-attack data is validated using a variety of user-defined methods in a smart contract that is specifically developed for that purpose. Integrating IPFS improves the blockchain network's responsiveness and scalability.
- Various performance criteria, including accuracy of the blockchain, are used to assess the proposed system.

3 PROPOSED METHODOLOGY

For the IoT-based SHS, this section offers the framework used. The proposed framework consists of many layers, including AI and blockchain, that offer a sequential flow, including data collection, data classification, along with blockchain security. Figure 1 shows the entities of the proposed framework.

3.1 AI Layer

Various AI algorithms, including Catboost, KNN, SVM, and LDA, are used in this part to illustrate the AI layer working. Both "Dataset Description" and "Adoption of AI algorithms" make up this section. The following is an in-depth description of each part.

3.2 Dataset Description

This work made use of the TON IoT dataset, which is a commonplace collection of data from SHS that includes several IoT sensors, including garage door, fridge, weather, and motion detectors. There are several distinct service profiles inside the whole dataset, such as those for IoT refrigerators, garage doors, location trackers, thermostats, and many more. Consider a dataset that includes weather (D_5), motion (D_4), garage door (D_1), fridge (D_2), GPS tracker (D_3), and other related data. Thus, the representation of a dataset for a SHS is denoted as $D\epsilon\{D_1, D_2, .., D_5\}$. Equation (12) represents the rows (w) and columns (q) count that make up each dataset $(D_i)\ \epsilon D$.

$$D_i^{w\times q}{=}D_i^{6401\times 5} \qquad (1)$$

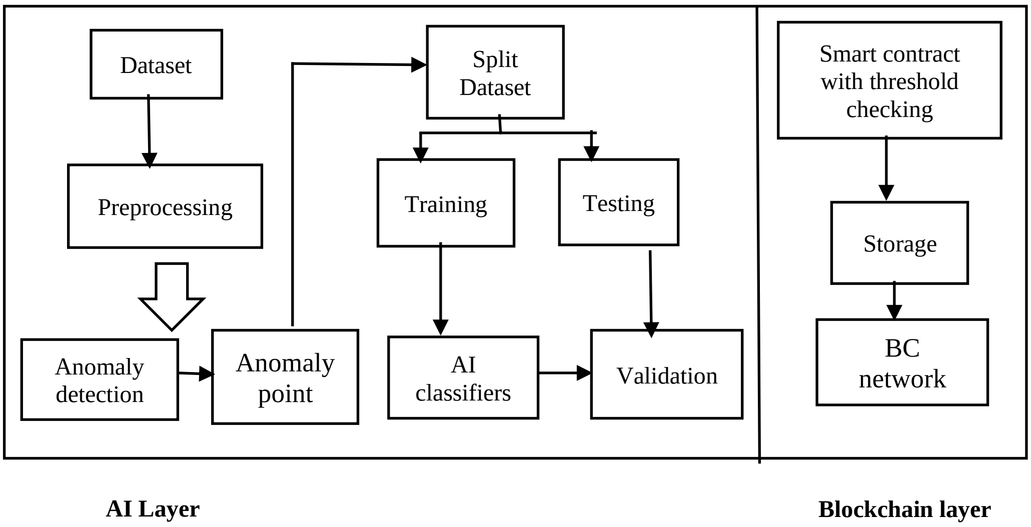

Figure 1. The proposed framework

3.3 Dataset Preprocessing

Here, the dataset $(D_i)\ \epsilon D$ is preprocessed utilizing the following procedures. Missing values, infinite values, not a number (NaN), as well as data type casting are all examples of errors in D_i. Take into consideration the D_i dimension, which is stated as

$$\begin{pmatrix} d_{1,1}\ d_{1,2} & \cdots & d_{1,q} \\ \vdots & \ddots & \vdots \\ d_{w,1}\ d_{w,2} & \cdots & d_{w,q} \end{pmatrix} \xrightarrow{contains} \{1, inf(\infty), NaN\} \qquad (2)$$

where {" –"} are the missing values, NaN is the value as well as ∞ is infinity values that is filled employing the central tendency value, i.e., mean (v).

$$\begin{pmatrix} d_{1,1} -- & \cdots & d_{1,q} \\ \vdots & \ddots & \vdots \\ -inf\ d_{w,2} & \cdots & d_{w,q} \end{pmatrix} \xrightarrow{filled\ with\ v} \begin{pmatrix} d_{1,1}\ \boxed{v} & \cdots & d_{1,q} \\ \vdots & \ddots & \vdots \\ \boxed{v}\ d_{w,2} & \cdots & d_{w,q} \end{pmatrix} \quad (3)$$

The dataset normalization, D_i was also examined. In this case, the ith column values of D_i were not appropriately scaled, for instance, $d_{1,1} \gg d_{2,1}$or$d_{1,1} \ll d_{2,1}$. The data set D_i's columns must thus undergo normalization. We used the min-max scalar, written as in (4)

$$\theta = \frac{d_i - d_i^{min}}{d_i^{max} - d_i^{min}} \quad (4)$$

Where θ the rescaled output for D_i, and falls inside the [0, 1] range. The lowest as well as maximum values of the ith column of D_i are denoted as d_i^{max} and d_i^{min}, respectively, where d_i is the input value. Additionally, the datatype of several columns in the D_i makes them unusable for AI models. For instance, an AI system based on conditional probability cannot adopt the column with the object datatype. Therefore, $D_i$has to undergo an appropriate datatype conversion.

$$int\ d_i{=}(int)d_i \quad (5)$$

For the AI algorithms to train on the data set D_i, explicit data type casting is done in Equation (16). D_i' is the final preprocessed dataset.

3.4 Anomaly Detection

After being preprocessed, the data set is passed on to the AI layer, where several AI models are used for anomaly detection. To verify the model's parameters, the preprocessed data set D_i' is split into the training along with testing data sets.

$$\forall D_i' = \begin{cases} D'_{train} \\ D'_{test} \end{cases} \quad (6)$$

The (D_i') training and testing sections, are denoted as D0 train and D0 test, respectively. The train_test_split () technique is used to divide the dataset into two parts: one for training (0.8:80%) and one for testing (0.2:20%). The model's validation encompasses the many parameters used to assess its performance. Rerunning the model on the test data has confirmed its correctness. The (D_i') dataset is checked for anomaly detection before classification. This checks to see whether the attacker has tampered with the dataset. AI models are trained on as well as deliver false results if an attacker has altered the data set values. The functionality of the whole SHS is thereby put at risk. To begin with, under the AI layer, the dataset (D_i') is iterated through anomaly-detection algorithms in order to identify data behavior or if the data are abnormal. The program detects anomalies or outliers in the data and categorizes them as anomaly or nominal data. We discovered that IF outperforms all other anomaly-detection algorithms in terms of how well it identifies outliers as anomalies via our model performance study. Comparable to the random forest method, IF follows a similar logic structure. To determine its behavior (anomaly or nominal), the tuple that is handled at a specific moment in time during the model iteration will be separated. An estimate is the quantity of divisions needed to identify the position of a certain point or tuple. Building an ensemble of isolation trees is how IF functions. For the construction of each isolation tree, a feature along with a split value within that feature's range are chosen at random. This is done in a cyclical fashion until every data point is placed in its own leaf node. According to what was said before, when the tree is built, the nature of that occurrence is determined by calculating the feature value of the anomaly score. A possible formulation for the anomaly score Z is

$$Z(o) = 2^{\frac{-E(h(o))}{c(s)}} \quad (7)$$

where o is the data point that is being used to generate the anomaly score. The term $\left(E\left(h(o)\right)\right)$gives the average path length over all trees for each data point o in the ensemble. In addition, the normalization factor, denoted as $c(s)$, is the isolation trees average path length; here, s is the total data points count. The formula provides a definition of the word $c(s)$,.

$$c(s) = \begin{cases} 2h(s-1) - 2\frac{s-1}{n}, for\ s > 2 \\ 1, for\ s = 2 \\ 0,\ otherwise \end{cases} \quad (8)$$

The point's behavior is determined by the anomaly score Z value. Anomaly is classified as a score that is close to 1. It is classified as a minimum point if it is close to 0.5. The anomaly-free dataset, which contains just nominal data, is the updated dataset D_a. D_a still contains attack as well as non-attack data. To classify the data in D_a, classification algorithms are thus required. A number of performance indicators are used to classify the data once supervised learning algorithms have been created and validated. For classification catbbost algorithm is used in the work and is compared with some other AI algorithms.

Gradient Boosting Decision Tree (GBDT) is the foundation of CatBoost, a machine learning framework. After a weak learner finish learning, the GBDT method calculates the current loss function gradient also utilizes the next weak learner to fit the gradient. Decision trees are utilized as weak learners in this process. The sum of these weak learners becomes a strong learner in due time. Some of CatBoost's most notable characteristics are the usage of an enhanced GBDT algorithm and the following:

(1) Catboost solves the issue of having too much categorical data in intrusion detection system by adding a way to process categorical variables that uses both numerical encoding and one-hot encoding. This approach is able to effectively handle categorical features. This is the encoding technique that CatBoost employs for category features when their unique value count is higher than the threshold. To be more specific, let's pretend that we have a dataset of observations denoted as $D = (X_i Y_i)$, where i ranges from 1 to n. In this dataset, X_i is a vector containing m characteristics, some of which are numerical and some of which are categorical. Y is a label value belonging to the set R. The encoded value of $x_{\sigma p,k}$ is given by the permutation $\sigma = (\sigma_1, \ldots, \sigma_n)$.

$$\frac{\sum_{j=1}^{p-1}[x_{\sigma j,k}=x_{\sigma p,k}]Y_{\sigma j}+a.P}{\sum_{j=1}^{p-1}[x_{\sigma j,k}=x_{\sigma p,k}]Y_{\sigma j}+a} \quad (9)$$

Iverson brackets, denoted by [·], are defined as follows: $\left[x_{j,k} = x_{i,k}\right]$ =1 if $x_{j,k}$= $x_{i,k}$ and 0 otherwise. Prior values are P and a > 0. Records in the dataset will have various encodings based on their placements, even if they have the same feature value, as shown in the calculation.

(2) To handle highly unbalanced data, Catboost offers a weighted cross-entropy loss function that makes it simple to modify the weight of various variables in the loss function. It has the potential to rectify the systemic issue of category imbalance [17].

In anomaly detection, catboost thus evaluate the risk related with a certain data point or location. This allows to make educated judgments. The proposed system produces an alert if the model's behavior is discovered to be an attack after classification. Conversely, the data is saved in the BC network defined in the BC layer if the behavior is determined to be non-attack.

3.5 Blockchain Layer

This layer is responsible for transmitting as well as securely storing the AI layer's non-attack data. An attacker may theoretically launch many security attacks against a SHS since its non-attack data is saved in a online storage. That is why it's crucial to have transparent, secure storage that can deal with data integrity concerns. One significant answer to this problem is blockchain technology. We implemented it by creating a smart contract that verifies incoming data that is not related to attacks. After the AI classifier validates the incoming data for non-attack purposes, IPFS permits the data to be kept secure. An application programming interface (API) for Filebase that enables programmatic interaction with IPFS enables this. A unique content identifier (CID) is obtained in order to access the material later, when the smart contract's certified data is published to IPFS via Filebase. In addition, IPFS may calculate data hashes and send them to an immutable blockchain ledger. The blockchain network becomes visible when all SHS entities are required to register with it. Enlightening the security of the SHS is possible thanks to the transparency aspect of blockchain, which allows one to identify the person responsible for data modification [16].

4 RESULTS

4.1 Analysis of Results

This research uses the Jupyter Notebook for implementation task. Also, the proposed analysis results is done with diverse performance parameters, including statistical measures. This research employs the anaconda distribution 6.3.0 version.

4.2 Discussion

This section displays the outcomes of anomaly detection in SHS. When looking at the accuracy of identifying abnormalities from the SHS, Figure 2 shows how well the proposed framework has done. The x-axis shows the detection accuracy, while the y-axis shows the chosen anomaly-detection algorithms (i.e., IF as well as LOF).

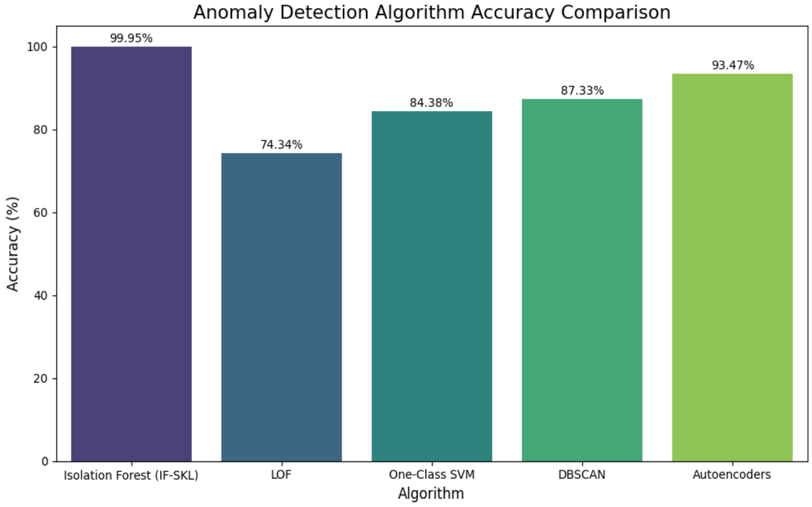

Figure 2. Anomaly-detection algorithms accuracy

With a remarkable 99.95% accuracy rate, the IF-SKL model reveals anomalies more effectively than any of the competing algorithms. Because of the vital importance of identifying even small security breaches in a smart home context, this shows that the algorithm is excellent at separating unusual activity. On a consistent basis, IF-SKL outperforms LOF (74.34%), One-Class SVM (84.38%) [18], DBSCAN (87.33%) [19], and autoencoders (93.47%) [20]. IF-SKL excels in handling high-dimensional data and massive datasets. The data produced by smart home devices and sensors is often high-dimensional, and the number of these devices and sensors might be large. Instead of autoencoders or DBSCAN, which may be inefficient with big, noisy datasets or demand a lot of computational resources, IF-SKL can handle this kind of data very well. In Internet of Things (IoT) settings where data streams are continuous and large-scale anomaly detection is required in real time, scalability and speed are crucial.

Unpredictability and diversity in data supplied by various IoT devices are major obstacles for smart home systems. By removing the very out-of-the-ordinary data points, Isolation Forest is able to successfully deal with noisy data and outliers. On the other hand, the accuracy of LOF, which stands at 74.34%, is lower because it struggles with sparse or highly variable datasets. Although it works well, One-Class SVM isn't as resilient as IF-SKL when it comes to handling the specific data features of smart homes. Unlike DBSCAN, which depends on establishing distance thresholds and the minimum number of points, IF-SKL does not need sophisticated parameter adjustment. This facilitates its implementation and adjustment for real-world uses. IF-SKL outperforms more complicated models like autoencoders due to its efficiency and simplicity, making it suitable for real-time deployment in settings with limited resources like smart homes. For smart homes, where quick actions are required to reduce security concerns, IF-SKL's real-time anomaly detection efficiency is crucial. For real-time detection in real-world settings, algorithms such as autoencoders are efficient but may be computationally costly. IF-SKL is the top option for smart home security because it combines excellent accuracy with little computational overhead and speed.

4.3 Classification-Based Result Discussion

Various AI classifiers, including catboost, KNN, SVM, and LDA were used to implement the findings in this subsection. AI classifiers, which fall under the umbrella of supervised learning, are therefore more effective in these situations for classification tasks. In Figure 3, we can see how the accuracy used on the dataset, which contains the class labels "1" and "0" for anomaly and nominal, respectively, is compared. An AI classifier's accuracy may be expressed in the following way: Where μ, γ, θ, along with ϑ stand for the true positive, true negative, false positive, as well as false negative values. Accuracy is equal to the sum of these four variables plus one.

$$Acc = \frac{\mu+\gamma}{\mu+\gamma+\theta+\vartheta} \qquad (10)$$

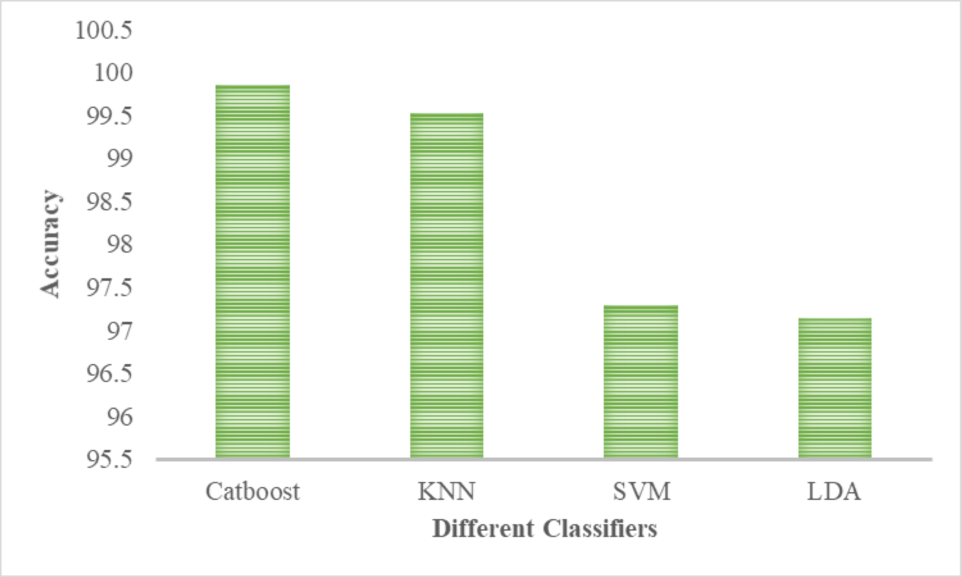

Figure 3. Accuracy of AI classifiers.

To further understand how well the catboost algorithm worked, additional outcome parameters are shown in Figures 3. Summarizing a classifier's performance, a confusion matrix is a statistical performance metric that relies on matrices. The confusion matrix incorporates the following parameters: One important indicator is the true positive (μ) value, which indicates how many positive outcomes were accurately classified as positive according to the data. The false-positive category is the sum of all the negative results that the algorithm falsely iterates as positive outcomes. 3. True Negative (): This group includes all outcomes that were correctly classified as negative. 4. The fourth parameter is the false negative (ϑ) value, which is the sum of all the negative outcomes that were mistakenly forecast as positive.

When dealing with tabular or structured data, the gradient boosting technique CatBoost shines. Its foundation in decision trees allows it to efficiently and quickly deal with categorical data directly, without requiring their preprocessing into one-hot encoding. The 99.85% accuracy rate is a testament to CatBoost's prowess in huge datasets with intricate feature interactions and in data classification tasks with non-linear correlations. Smart home systems often collect complex, multi-dimensional data, which includes multiple categorization variables such as the types of devices, their on/off states, and more. The model's exceptional performance is largely attributable to its regularization-based ability to avoid overfitting and its efficient handling of categorical data. Internet of Things (IoT) settings, such as smart houses, often include complicated, non-linear correlations and interactions between features; CatBoost outperforms more conventional algorithms in this regard.

For instance-based learning, there is KNN, a straightforward algorithm. This method is able to categorize data points by taking into account the feature space majority class of their K nearest neighbors. This method is known as a lazy learner, as it doesn't require any specific training. With an accuracy of 99.53%, KNN is clearly doing well on your dataset. KNN excels when you properly partition the data and select the appropriate K value. When compared to other algorithms, it performs better when using the correct distance metric, such as the Euclidean distance. Because IoT devices in smart home systems often produce data that is relatively close in feature space, the KNN algorithm is a suitable match. With a discernible pattern among the data points, the algorithm's simplicity makes it useful. For smaller to medium datasets and easier challenges, KNN usually works effectively. But when working with high-dimensional data or massive datasets, it could be sluggish. Because of its superior performance compared to more complicated models like SVM or LDA, its accuracy in this case implies that the data may not be too sparse or high-dimensional.

One use of SVMs is in classification tasks that need supervised learning. To do this, it searches the feature space for a hyperplane that effectively divides the classes. In order to improve generalization to new data, the model strives to increase the margin between the two classes. SVM demonstrates good performance on the dataset, with an accuracy of 97.29 percent. When there is a large gap between the classes, SVM performs well. It is well-suited for complicated data because, thanks to kernels, it can successfully handle non-linear boundaries. By identifying the limits between typical and anomalous data points, SVMs are able to successfully categorize data in smart home systems. Despite their effectiveness, SVMs are sensitive to kernel and parameter choices (such as C and gamma). Data that is high-dimensional, like in smart homes equipped with several sensors, might make SVM inefficient compared to CatBoost, which is designed to handle such complexity.

The goal of linear combination of features (LDA), a linear classification approach, is to identify the optimum way to distinguish between classes. To establish a decision border between the classes, LDA optimizes the ratio of inter-class variance to within-class variance, in contrast to logistic regression that employs probability for classification. When the data shows a linear separation between the classes, LDA performs well (97.15% accuracy). It becomes quite helpful when the classes are well defined and the data follows a normal distribution. By reducing the dimensionality, LDA makes the classification problem simpler and is therefore well suited to high-dimensional data. When dealing with complicated interactions (non-linearity), LDA may not function as well as it does with simpler, linearly separable datasets. However, more complicated systems, such as smart homes, benefit from the performance of CatBoost and KNN because of their superior handling of non-linear connections.

Because it is so good at dealing with complicated feature interactions and non-linear correlations, CatBoost gets the best accuracy (99.85%). It has built-in support for categorical data and is resistant to overfitting. Using its gradient boosting approach, it may build a powerful model by combining the strengths of several weak learners, such as decision trees. Smart home systems, dealing with complex, high-dimensional data, benefit from CatBoost's superior processing power on large datasets, which includes categorical characteristics, and its ability to prevent overfitting. KNN achieves an impressive 99.53% accuracy using a straightforward and easy-to-understand method. Some datasets involving smart homes may include data points that are comparable in feature space, making it an effective choice. However, performance could suffer with high-dimensional data (the curse of dimensionality), and computational costs might increase with dataset size. SVM achieves respectable results (97.29% accuracy) because it establishes distinct decision boundaries across classes. When there is a distinct boundary between the two variables, it works well. However, it may not always outperform alternatives such as CatBoost and may struggle with data with numerous dimensions. When data is linearly separable, LDA does a decent job (97.15 percent accuracy), but it has a challenging time when feature correlations aren't linear. Algorithms like CatBoost outperform LDA when dealing with IoT data, which might include intricate patterns. Once the validation is complete, the secure storage system based on IPFS receives the data that was not a victim of an attack.

5 CONCLUSION

To address security risks connected with SHS, the article proposed a secure as well as intelligent framework. To eliminate SHS anomalous data, the proposed AI framework makes use of the automation along with intelligence of AI algorithms. This is performed by training an IF algorithm that employs an ensemble strategy to identify as well as remove anomalous data points from the data set. The AI classifiers are trained to classify data once the anomalous data have been deleted. Only non-attack data is permitted to help improve the performance of the smart framework, which scarifies the attack data. In addition, the non-attack data is sent to the immutable blockchain nodes to fortify the SHS security. By storing smart home non-attack data on blockchain nodes, the likelihood of data tampering is reduced. After comparing the proposed framework to the existing work, the findings clearly demonstrate greater performance. To lessen the burden on participants and the environment, we want to use proof-of-stake (PoS) and hybrid methods in the future work.

REFERENCES

[1] Li, Joey, Munur Sacit Herdem, Jatin Nathwani, and John Z. Wen. "Methods and applications for Artificial Intelligence, Big Data, Internet of Things, and Blockchain in smart energy management." Energy and AI 11 (2023): 100208.

[2] Hua, Weiqi, Ying Chen, Meysam Qadrdan, Jing Jiang, Hongjian Sun, and Jianzhong Wu. "Applications of blockchain and artificial intelligence technologies for enabling prosumers in smart grids: A review." Renewable and Sustainable Energy Reviews 161 (2022): 112308.

[3] Kumar, Nallapaneni Manoj, Aneesh A. Chand, Maria Malvoni, Kushal A. Prasad, Kabir A. Mamun, F. R. Islam, and Shauhrat S. Chopra. "Distributed energy resources and the application of AI, IoT, and blockchain in smart grids." Energies 13, no. 21 (2020): 5739.

[4] Aguilar, J., Alberto Garces-Jimenez, M. D. R-moreno, and Rodrigo García. "A systematic literature review on the use of artificial intelligence in energy self-management in smart buildings." Renewable and Sustainable Energy Reviews 151 (2021): 111530.

[5] Miglani, Arzoo, Neeraj Kumar, Vinay Chamola, and Sherali Zeadally. "Blockchain for Internet of Energy management: Review, solutions, and challenges." Computer Communications 151 (2020): 395-418.

[6] Samuel, Omaji, Nadeem Javaid, Turki Ali Alghamdi, and Neeraj Kumar. "Towards sustainable smart cities: A secure and scalable trading system for residential homes using blockchain and artificial intelligence." Sustainable Cities and Society 76 (2022): 103371.

[7] Sharma, Ashutosh, Elizaveta Podoplelova, Gleb Shapovalov, Alexey Tselykh, and Alexander Tselykh. "Sustainable smart cities: convergence of artificial intelligence and blockchain." Sustainability 13, no. 23 (2021): 13076.

[8] Singh, Sushil Kumar, Shailendra Rathore, and Jong Hyuk Park. "Blockiotintelligence: A blockchain-enabled intelligent IoT architecture with artificial intelligence." Future Generation Computer Systems 110 (2020): 721-743.

[9] Kumari, Aparna, Rajesh Gupta, Sudeep Tanwar, and Neeraj Kumar. "Blockchain and AI amalgamation for energy cloud management: Challenges, solutions, and future directions." Journal of Parallel and Distributed Computing 143 (2020): 148-166.

[10] Yang, Qing, and Hao Wang. "Privacy-preserving transactive energy management for IoT-aided smart homes via blockchain." IEEE Internet of Things Journal 8, no. 14 (2021): 11463-11475.

[11] Van Cutsem, Olivier, David Ho Dac, Pol Boudou, and Maher Kayal. "Cooperative energy management of a community of smart-buildings: A Blockchain approach." International Journal of electrical power & energy systems 117 (2020): 105643.

[12] Ahmed, Imran, Yulan Zhang, Gwanggil Jeon, Wenmin Lin, Mohammad R. Khosravi, and Lianyong Qi. "A blockchain-and artificial intelligence-enabled smart IoT framework for sustainable city." International Journal of Intelligent Systems 37, no. 9 (2022): 6493-6507.

[13] Khattak, Hasan Ali, Komal Tehreem, Ahmad Almogren, Zoobia Ameer, Ikram Ud Din, and Muhammad Adnan. "Dynamic pricing in industrial internet of things: Blockchain application for energy management in smart cities." Journal of Information Security and Applications 55 (2020): 102615.

[14] Ali, Muhammad, Krishneel Prakash, Md Alamgir Hossain, and Hemanshu R. Pota. "Intelligent energy management: Evolving developments, current challenges, and research directions for sustainable future." Journal of Cleaner Production 314 (2021): 127904.

[15] Pandiyan, Pitchai, Subramanian Saravanan, Kothandaraman Usha, Raju Kannadasan, Mohammed H. Alsharif, and Mun-Kyeom Kim. "Technological advancements toward smart energy management in smart cities." Energy Reports 10 (2023): 648-677.

[16] Shah, Khush, Nilesh Kumar Jadav, Sudeep Tanwar, Anupam Singh, Costel Pleșcan, Fayez Alqahtani, and Amr Tolba. "AI and Blockchain-Assisted Secure Data-Exchange Framework for Smart Home Systems." Mathematics 11, no. 19 (2023): 4062.

[17] Wei, Zizhong, Fanggang Ning, Kai Jiang, Yang Wang, Zixiang Bi, Qiang Duan, Jichen Zhang, and Rui Li. "CatBoost-based Intrusion Detection Method for the Physical Layer of Smart Agriculture." In ITM Web of Conferences, vol. 60, p. 00009. EDP Sciences, 2024.

[18] Vos, Kilian, Zhongxiao Peng, Christopher Jenkins, Md Rifat Shahriar, Pietro Borghesani, and Wenyi Wang. "Vibration-based anomaly detection using LSTM/SVM approaches." Mechanical Systems and Signal Processing 169 (2022): 108752.

[19] Jain, Praphula Kumar, Mani Shankar Bajpai, and Rajendra Pamula. "A modified DBSCAN algorithm for anomaly detection in time-series data with seasonality." Int. Arab J. Inf. Technol. 19, no. 1 (2022): 23-28.

[20] Zhou, Yingjie, Xucheng Song, Yanru Zhang, Fanxing Liu, Ce Zhu, and Lingqiao Liu. "Feature encoding with autoencoders for weakly supervised anomaly detection." IEEE Transactions on Neural Networks and

Learning Systems 33, no. 6 (2021): 2454-2465.