Sigma J Eng Nat Sci, Vol. 43, No. 3, pp. 714–725, June, 2025

Sigma Journal of Engineering and Natural Sciences
Web page info: https://sigma.yildiz.edu.tr
DOI: 10.14744/sigma.2025.00066

Research Article

Cyber-shield machine learning: an intrusion detection system with blockchain-powered secure log storage across network nodes

Tejaswini PAWAR[1,*], Jyoti RAO[2], Pramod PATIL[2]

[1]Dr. D. Y. Patil Institute of Technology, Pune; MVP Karmaveer Adv. Baburao Ganpatrao Thakare College of Engineering KBTCOE, Nashik, 422013, India
[2]Dr. D. Y. Patil Institute of Technology, Pimpri, Pune, 411018, India

ARTICLE INFO

Article history
Received: 02 April 2024
Revised: 02 July 2024
Accepted: 11 September 2024

Keywords:
Blockchain, Collaborative Intrusion Detection; Learning-Based Ensemble Model; Machine Learning; Privacy Preservation

ABSTRACT

Intrusion detection and blockchain technology have been extensively studied to enhance data privacy and identify current and potential cyber attacks. In this approach, machine learning algorithms can identify complicated malicious occurrences simultaneously by protecting data privacy. This model is utilized to offer more security and privacy in the cloud to protect networks. This research proposes an innovative Intrusion Detection System imposing blockchain technology which employs consensus mechanisms and compares different machine learning methods including XGBoost, Random Forest, Decision Tree, and Extra Tree. This system is developed to enhance data privacy, security, and protect Internet-of-Things networks. The utilization of ant colony optimization further improves accuracy during real-time data analysis. The aim of the study is to extend the application of machine learning for network intrusion detection and to conduct a comparative analysis of the proposed ant colony algorithms with other existing studies in terms of accuracy, recall, and precision measures. The high accuracy and strong intrusion detection capability of the model are verified through experimental evaluations using datasets such as NSL-KDD and CICIDS2017. Proposed system shows effectiveness by using creative technology integration for mitigating cyber security risks. The experimental result shows that a machine learning algorithm using ant colony optimization achieves 99% accuracy. This model represents a significant progression towards addressing a resilient solution for cyber security challenges and to tackle new cyber threats.

Cite this article as: Pawar T, Rao J, Patil P. Cyber-shield machine learning: an intrusion detection system with blockchain-powered secure log storage across network nodes. Sigma J Eng Nat Sci 2025;43(3):714–725.

INTRODUCTION

The features of blockchain, such as trust promotion, data privacy guarantee, and transparency improvement, make it a feasible option for securely facilitating the exchange of transactions and information among participants. The immutable record of data sharing is maintained in audit trails and accessible exclusively by authorized entities or securely hosted by cloud service providers under

***Corresponding author.**
*E-mail address: pawar.tejaswini@kbtcoe.org

This paper was recommended for publication in revised form by Editor-in-Chief Ahmet Selim Dalkilic

particular trust standards and permissions. Blockchain validate data transactions using consensus techniques among participating entities by presenting itself as a potential alternative to trustworthy third-party intermediates [1]. These consensus mechanisms play a significant role in cloud a system which is crucial for verifying transactional records. Notably, three widely used consensus methods are prominent in this context [2]: Practical Byzantine Fault Tolerance (PBFT), Proof of Stake (PoS), and Proof of Work (PoW). PoW involves the selection of nodes and transaction verification through computational resource consumption [3], designed to counteract Sybil attacks and the potential threat of a 51% attack. In contrast, PoS selects block creators based on their stake without the requirement for extensive computational power [4]. PBFT, currently implemented in Hyperledger Fabric, simplifies operational complications, ensures energy efficiency, and proves extremely useful in distributed systems [5].

A simple blockchain-based IoT network authentication system is used to solve this problem, Wang et al. propose. introduce this approach, data integrity and non-repudiation are guaranteed since IoT nodes must first approve one another via an intuitive blockchain interface to communicate. This assures that all transactions are traceable and impervious to tampering, secure authentication, and prevention of unauthorized access. This approach's durability and simplicity make it perfect for the dynamic conditions seen in IoT networks, where user-friendliness and security are essential [6]. Tian et al. an innovative approach that combines reinforcement learning with distributed ledger technology is put forth to enhance the efficiency and security of routing in WSNs. To ensure effective data transfer, the best routing paths are found using the reinforcement learning algorithm. Additionally, a layer of security is introduced by using blockchain technology to authenticate network nodes and manage all routing information concurrently. By blocking harmful routing links that can potentially send data through compromised nodes, this blockchain integration serves to protect the network from security risks and guarantee the dependability of data transmission within the WSN [7].

To address the problems of certificate-less key management, Manoj et al. present a decentralized framework of blockchain for node registration, authentication and transitions. This approach leverages the blockchain's capability for constructing an immutable ledger which facilitates active monitoring and rapid identification of the hostile nodes thus enhancing the reliability and security of the constructed network in the face of node breaches [8][9]. Mohanta et al. suggest the use of blockchain to implement secure storage of node authentication and trustful data in network security. Public keys make it safe to authenticate while blockchain guarantees that data is accurate and open. This system describes node trustworthiness based on prior engagement, thus improving network dependability [10]. Blockchain is used in [11] to solve IoT problems. Internet of Things devices self-register on the blockchain. The activity is carried out per the capabilities of the IoT device if it has successfully registered and authenticated. Similarly, to control and administer Internet of Things devices, users must authenticate themselves on the blockchain network. It prevents hostile nodes from joining the network and keeps all proof on a blockchain

Wireless sensor networks face challenges in ensuring fairness and accountability in the identification of malicious nodes and therefore, She et al. proposed a Blockchain Trust Model (BTM) that addresses the issue effectively. To detect and record the unauthorized nodes in 3D space it utilizes a framework and blockchain data structure, smart contracts and a localization method. Actual simulation proves the efficiency and explicability of the present approach [12]. Within the realm of advancing technologies, the generation of vast datasets presents notable challenges in securely transmitting this data across networks. Despite the existence of numerous commercial Intrusion Detection Systems (IDS) and ongoing research efforts, the persistent threat of breaching network systems calls for innovative approaches to strengthen data security [9]. Functioning as a vigilant tool, an Intrusion Detection System monitors every event and packet within a network or computer system, scrutinizing them to recognize malevolent actions. Specifically, anomaly-based IDS identifies data packets in network traffic that deviate from the established normal profile [13,14].

To tackle the evolving challenges posed by cyber threats and enhance intrusion detection effectiveness, this paper introduces an Intrusion Detection System utilizing machine learning techniques [15], for regression and classification applications, including XGBoost, Random Forest, Decision Tree, and Extra Tree. The efficacy of machine learning is often compromised when applied to real-time and continuous data streams [16,17]. To address this limitation, the system integrates ant colony optimization, leveraging their fitness functions to improve accuracy in real-time data analysis [18]. The primary focus of this research involves delving into the utilization of machine learning in network intrusion detection. It aims to construct a resilient system using modern algorithms and offers a thorough comparative analysis with other existing studies to encompass accuracy, precision, and recall metrics [19,20]. Singh [21] introduces a study which improves WSN's data security using a new blockchain-based authentication method, thus improving aggregation integrity. The method was merged with sensor nodes and a private blockchain, the security of which was checked, and the work was carried out on the WiSeN sensor node. Sun et al. [22] propose an IDS framework of PSO-AdaBoost for the detection of malware in health apps. Based on the preprocessed NSL KDD dataset, the system recognizes the 12 features of the various attacks and separates DoS, U2R, R2L, and Probe attacks. Abualigah et al. [23] propose an IDS model consisting of a Particle Swarm Optimizer and AdaBoost for the detection of malware in health applications. The proposed system adopts 12 features from the NSL KDD dataset and differentiates between DoS, U2R, R2L and Probe attack types.

Table 1. A condensed associated work table

Restrictions	Resolutions	Verification	Identified research gaps
The network contains malicious nodes [6]	A proposed intrusion detection framework	Network connectivity and data integrity	Because the XoR function is weak, attacks by black holes and grey holes could happen
The problem of uncertain nodes' localization [24]	A blockchain-based trust paradigm is applied	Traceability, equity, and feasibility	Large data sets cause the encryption process to lag with the RSA
Networks for crowd sensing and vulnerabilities [7]	Proposed are a confusion mechanism and an incentive mechanism based on the blockchain	Energy usage, lateness	Route acquisition latency and packet delivery ratio did not improve
The registration process uses the centralization method [8]	A model based on a hybrid blockchain is suggested	Time spent processing and delay in transmission	Duplication of data
WSN node localizations; unidentified nodes launch network assaults [10]	A proposed blockchain trust model	Rate of false negatives, precision of detection, and energy usage	There is no usage of hashing algorithms or encryption for security
Routeing dynamically and central registration [11]	Reinforcement learning algorithms and blockchains are utilized	Time lag and energy use	Processing and queue delays
IoT node vendors are at odds on choosing a straightforward central administrator [12]	A suggested BCR protocol	Packet delivery ratio, packet drop ratio, and delay	Route acquisition latency and packet delivery ratio did not improve
To extend the lifespan of WSNs and balance sensor node energy consumption [25]	A suggested dynamic hierarchical protocol employing combinatorial optimization	A hierarchical network structure can be built using a hierarchy-based connecting technique	The computational complexity and processing delay
Lifetime reduction of ultra-dense WSNs [26]	Unsupervised learning methodology	Complexity of computing and residual energy	A rise in the complexity of computation

Through this holistic approach, the proposed model seeks to offer a resilient and adaptive solution to the dynamic landscape of cybersecurity.

1) To explore the utilization of machine learning in the context of the detection of intrusion in networks.
2) To create a system to detect Intrusion employing algorithms like Random Forest, Extra Tree, and Decision Tree, and the XGBoost algorithm.
3) Enhancing accuracy and decreasing computational time, both genetic and ant colony algorithms were tested whichever gives the best result that implementation incorporates Ant Colony algorithms in the proposed model.
4) To conduct a comparative analysis and present outcomes concerning accuracy, precision, and recall metrics for these algorithms.

Methodology for Evaluation

Within the realm of advancing technologies, the generation of vast datasets presents notable challenges in securely transmitting this data across networks. Despite the existence of numerous commercial Intrusion Detection Systems (IDS) and ongoing research efforts, the persistent threat of breaching network systems calls for innovative approaches to strengthen data security. Functioning as a vigilant tool, an Intrusion Detection System monitors every event and packet within a network or computer system, scrutinizing them to recognize malevolent actions. Specifically, anomaly-based IDS identify data packets in network traffic that deviate from the established normal profile.

Illustrated in Figure 1 is the proposed implemented system for intrusion detection, a state-of-the-art amalgamation of blockchain technology, machine learning, genetic algorithms, and ant colony algorithms. Every network node securely records its activities on the blockchain, creating an immutable and transparent ledger. The ant colony and genetic algorithms yielded the most significant results during testing, indicating that the Ant Colony algorithm is utilized in the proposed model. The recorded logs then function as input for a pre-trained machine learning model, enhancing its accuracy in predicting intrusions. Addressing

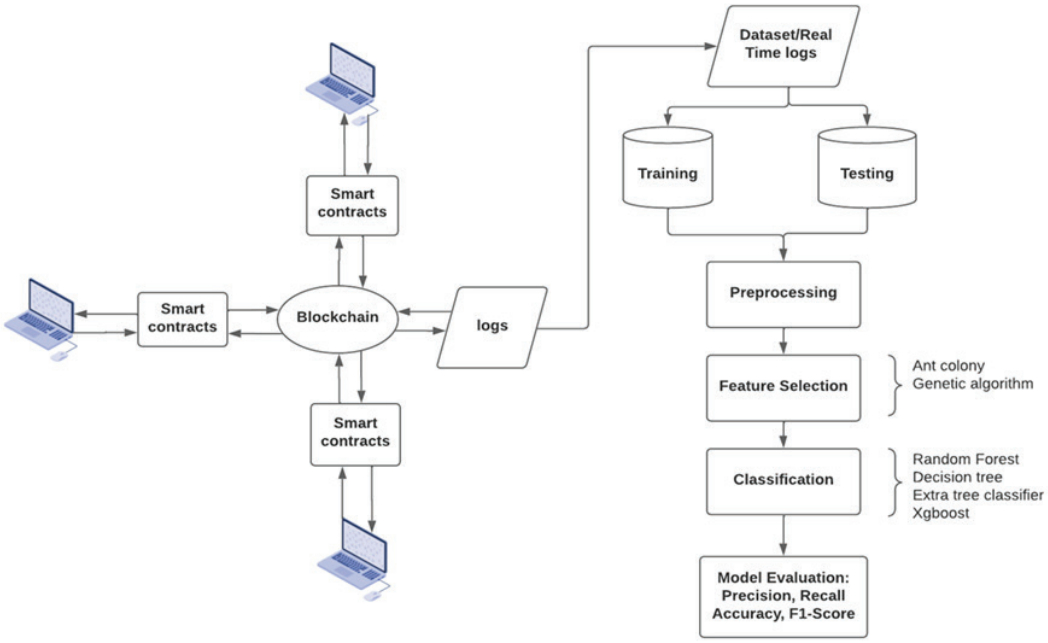

Figure 1. Block diagram of the proposed Intrusion detection model.

the issue of reduced accuracy in real-time systems, they skilfully incorporated ant colony algorithms with fitness functions. This integration allows for the dynamic adaptation and optimization of parameters, ensuring efficient intrusion detection.

The model based on machine learning continuously analyses transactions in real time on the blockchain, recognizing anomalies that may signify potential security threats. Upon receiving an alert, nodes engage in a collaborative decision-making process to confirm the existence of the threat. Following verification, an execution of a response mechanism, such as isolating the impacted node, takes place. Significantly, the entire process of decision-making and response is recorded on the blockchain, guaranteeing transparency and auditability. Although the system demonstrates a resilient and decentralized strategy for intrusion detection and prevention, continuous evaluations of scalability, resource consumption, adaptability to emerging threats, and the security of the foundational blockchain are essential for maintaining its effectiveness in the ever-evolving cyber security landscape.

MACHINE LEARNING ALGORITHMS

A system created to monitor networks for indications of malicious activities or unauthorized access is termed an intrusion detection system (IDS). Employing machine learning techniques for real-time monitoring, the IDS is trained to identify and intercept packets recognized as malicious. The training process of the model involves multiple stages, including pre-processing, data collection, feature selection, validation, and model training. This iterative process ensures that the intrusion detection system remains adept at recognizing evolving patterns of cyber security threats.

Gathering of Data

In Intrusion detection, the CICIDS2017 and NSL-KDD datasets are employed, encompassing characteristics present in both the testing and training datasets.

Pre-processing is necessary for the dataset known as NSL-KDD, which includes categorical traits and duplicated data. Comprising 41 features initially, the dataset underwent feature reduction, resulting in 25 features. The CIC-IDS-2017 datasets employed the concept of profiles to systematically generate well-ordered datasets. Network traffic, captured using CICFlowMeter, was labelled appropriately, providing details about port numbers, source and destination addresses, timestamps, and detected attacks. The testing environment simulations included the generation of network traffic from various protocols, including HTTP, HTTPS, SSH, and email protocols like SMTP and POP3.

Pre-Processing of the Data

Pre-processing is a crucial step in converting unprocessed data into a more standardised format. The CICIDS2017 and NSL-KDD datasets were analyzed, and pertinent variables were found, as the study got underway. They checked the quality of the data to find mistakes and abnormalities. To guarantee dataset completeness, missing data points were removed, and outliers were handled with the proper techniques. The ant colony technique was used

for feature selection, and variable modifications were used to improve model performance or verify assumptions for statistical analyses. There were training and testing sets inside the dataset. The cleaned dataset was confirmed to satisfy study objectives using an IDS model. The pre-processing step is used in this instance to standardise and normalise the data.

i. Categorical feature identification: This entails creating a list of all the categories, as well as the number of them, that can be discovered in the testing and training datasets. This is shown in Figure 2 and Figure 3.

```
Training set:
Feature 'protocol_type' has 3 categories
Feature 'service' has 70 categories
Feature 'flag' has 11 categories
Feature 'label' has 23 categories
```

Figure 2. Categorical characteristics within the training dataset.

```
Test set:
Feature 'protocol_type' has 3 categories
Feature 'service' has 64 categories
Feature 'flag' has 11 categories
Feature 'label' has 38 categories
```

Figure 3. Features categorized in the testing dataset.

In the 'protocol_type' there are three categories of features in the training set. Specifically, there are 70 categories under 'service', 23 categories under 'label,' and 11 categories under 'flag'.

ii. One-Hot Encoding: Categorical features pose compatibility issues with machine learning algorithms. Consequently, a transformation is applied to convert all categorical features into binary vectors, serving both training and testing purposes. In the initial stage, every categorical value is associated with an integer value. Following this, each integer is represented in a binary vector format, where all values are 0 except for the index corresponding to the integer, which is designated as 1.

In the test set's 'protocol_type' attribute has 3 categories. Specifically, there are 64 categories under 'service,' 38 categories under 'label,' and 11 categories under 'flag'.

iii. Including Missing Categories in Testing Data: Six categories were found to be absent from the service feature in the testing data. Zeros have been added to these absent categories.

iv. Dataset Segmentation: The different types of attacks present in the datasets used for testing and training are mapped to DoS, Probe, R2L, and U2R after one-hot encoding is applied. Next, to ease model training across all attack types and forecast outcomes for each, the data is split into four datasets depending on these types of attacks (DoS, Probe, R2L, and U2R).

Selection of Features

Features are pivotal in machine learning, representing the measurable properties inherent in the observed phenomenon. The crucial step in the process involves selecting informative and independent features. Feature selection is the process of choosing a portion of useful features from the full collection of dataset is also known as attribute selection. A variety of feature selection strategies are applied to identify important factors and eliminate superfluous attributes. Including or excluding these attributes could reduce model accuracy but it could not effect on prediction model accuracy. It's important to emphasize that feature selection and feature extraction are two different things which represent separate processes. Feature selection selects the optimal subset of features from the initial set using forward or backward selection method, whereas feature extraction produces a new set of features using feature extraction algorithm like Explicit Semantic Analysis (ESA), Singular Value Decomposition (SVD) and Principal Component Analysis (PCA).

Genetic Algorithm

Genetic algorithms (GAs) play important role in optimizing machine learning model for improving adaptability of model to real-time data scenarios. These algorithms are inspired by the principles of natural selection, involving an evolutionary optimization process where solutions to develop over multiple generations and find the best solution to a specific problem.

i) GAs's Fitness Function and Objective Function:

It utilizes the decoded sequence and the ETC matrix to determine the execution time for task completion of each resource. Subsequently it calculates the overall time required to accomplish the resource scheduling task. The objective function is defined as follows:

$$F(x) = max_{r=1}^{n} \sum_{i=1}^{w} (r, i) \quad (1)$$

Where,

Work(r, i) is time required for resource r spends on this resource's subtask i.

w= number of subtasks that allocated to the resource.

The fitness function is used to estimate the strengths and weaknesses of chromosomes. A higher value of the function indicates stronger survivability of the chromosomes and a superior solution for the function. As the fitness function value is the reciprocal of the objective function, a shorter time results in a larger fitness value, increasing the probability of being selected.

The definition of the fitness function is:

$$f(x) = \frac{1}{F(x)} \quad (2)$$

ii) Genetic Manipulation

The genetic manipulation of a genetic algorithm encompasses processes like mutation, crossover, and selection. Through these operations, the algorithm consistently generates new individuals to explore and identify the optimal solution.

Selection

The likelihood of selecting each individual is determined by evaluating the fitness function value. Equation outlines the calculation for determining the probability of selection.

$$P(i) = \frac{f(i)}{\sum_{j=1}^{SCALE} f(j)} \quad (3)$$

Crossover

A higher crossover probability facilitates the exchange of some bits between individuals, helping to prevent premature occurrences. In the later stages of the algorithm, with a reduced crossover probability, there is an increased likelihood of generating new favorable individuals, thereby accelerating the convergence rate.

Mutation

A single-point mutation is employed to alter individual bits within groups with a lower probability, involving changes from "1" to "0" or "0" to "1". During actual operation, it excludes new individuals with fitness function values lower than the average value after multiple recursive iterations. It establishes the optimal solution for specific groups as a foundation for acquiring pheromones.

Ant-Colony Algorithm

The foundation of the ACO (Ant Colony Optimisation) algorithm is a computer framework modelled after the actions of actual ant colonies. The algorithm uses many constructive computational ants. Every ant is directed towards creating a solution, with the outcomes of earlier experiments kept in the ant dynamic memory system serving as guidance. The main goal of the ACO algorithm is to view an issue as the process of looking for the least expensive path through a graph. In this case, nodes stand in for features, and the edges that connect them indicate which feature to choose next.

An ant travels the graph in search of the best feature subset, stopping at the fewest nodes (features) that meet the traversal stopping criterion. Since every node is fully connected, any characteristic can be selected as the next step. The typical ACO algorithms' transition rules and pheromone update rules can be implemented by reformulating the graph representation in this way. In this case, heuristic and pheromone values are not linked to any one feature; rather, they are unique to each feature.

Performance Measures

Confusion matrices are used to illustrate the results of the model's training with machine learning, genetic, and ant colony methods on the provided dataset. Predictions are made for a variety of attacks. One method for summarising the performance of the classification algorithm is to use a confusion matrix. After that, measures for recall, accuracy, and precision are computed with the data in the confusion matrix.

The ratio of normal records that are mistakenly categorized as incursions to all normal records is known as the False Positive Rate or FPR. The number of intrusion records that the Intrusion Detection System (IDS) successfully classifies split by the total amount of test dataset intrusion records yields the detection rate. These metrics are useful performance indicators since they show the percentage of intrusions detected by the system and the quantity of incorrect classifications.

Typically, True Positive Rate (TPR), also known as Recall sensitivity of Detection Rate (DR) and False Positive Rate (FPR), also referred as False Alarm Rate (FAR) and they are employed for performance assessment. Table 2 represents contingency Matrix:

Table 2. Contingency matrix

		Predicted	
		No	**Yes**
Actual	No	∑TN	∑FP
	Yes	∑FN	∑TP

Accuracy from contingency matrix is calculated as follows:

$$\frac{TP+TN}{TP+TN+FP+FN} \quad (4)$$

Precision from contingency matrix is calculated as follows:

$$\frac{TP}{TP+FP} \quad (5)$$

Recall from contingency matrix is calculated as follows:

$$\frac{TP}{TP+FN} \quad (6)$$

$$FDR = \frac{FP}{TN+FP} \quad (7)$$

Another frequently employed metric F-measure is calculated as follows:

$$F - measure = \frac{2*Recall*Precision}{(Recall*Precision)} \quad (8)$$

RESULTS AND DISCUSSION

Datasets

The CICIDS2017 and NSL-KDD datasets are datasets utilized to assess the performance of the machine learning model.

CICIDS2017

Performance of machine learning model was evaluated on the CICIDS2017 dataset by considering evaluation metrics such as Accuracy, Precision, Recall, and F1-score. The results show that XGBoost and Random Forest achieved 98.89% accuracy, with XGBoost scoring 99% in precision and recall, and an F1-score of 98.9%, while Random Forest had 99% precision, 98.89% recall and an F1-score of 98.91%. The Decision Tree model performed best, with 99% accuracy, 99.12% precision, 99% recall and an F1-score of 99%. The Extra Tree Classifier achieved 98.89% accuracy, with 98.95% precision, 98.89% recall, and an F1-score of 98.86%. These metrics provide an in-depth understanding of how each algorithm performed in classifying data within the CICIDS2017 dataset, highlighting their strengths and relative performance levels.

NSL-kdd dataset

The effectiveness of several machine learning methods was evaluated using NSL-KDD dataset by utilizing criteria such as F1-score, recall, accuracy, and precision. Random Forest yielded the best results out of all the methods examined, with an accuracy of 97.11%. Decision Tree and Extra Tree Classifier also performed well, both reaching 97% accuracy. Even though its accuracy was marginally lower at 96.58%, XGBoost

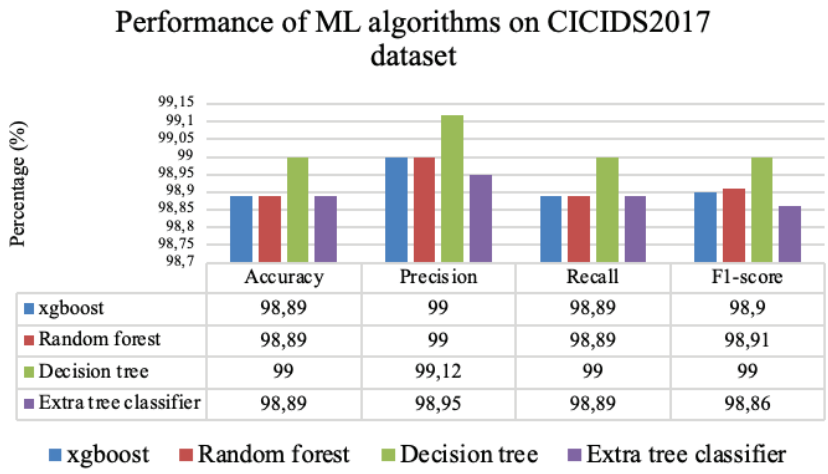

Figure 4. Performance of ML algorithms on CICIDS2017 dataset.

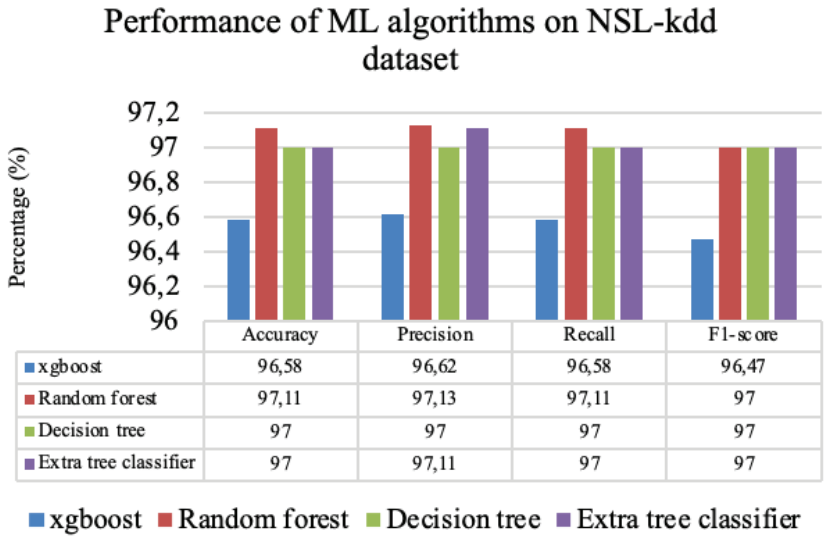

Figure 5. Performance of ML algorithms on NSL-kdd dataset.

still did remarkably well by all measures. With a precision of 97.13%, the Random Forest demonstrated the highest level of accuracy in correctly identifying positive events. The Decision Tree and Random Forest algorithms had the greatest and most constant recall rates, at 97.11% and 97%, respectively. Recall gauges an algorithm's capacity to collect all pertinent events. The Random Forest also has the best F1-score 97%, which measures recall and precision in a balanced manner. This indicates its overall robust performance.

ALGORITHMS

Ant Colony Algorithm Features

ON CICIDS2017 dataset

Using the dataset CICIDS-2017, the evaluation of machine learning methods using ant colony attributes produced remarkable outcomes. Accuracy, precision, recall, and F1-score were all perfect at 99% for XGBoost, Random Forest, and Extra Tree Classifier. With steady scores of 98%, the Decision Tree likewise demonstrated good performance. These results show that machine learning techniques and ant colony features work together to reliably identify and categorize cases in the dataset CICIDS-2017.

On NSL-KDD datasets (ant colony)

To assess machine learning, the NSL-KDD dataset utilized ant colony features, and the results were excellent in terms of accuracy, precision, recall, and F1-score. The detection and classification capabilities of XGBoost, Decision Tree, Random Forest, and Extra Tree Classifier were outstanding, as seen by their flawless scores of 99% in each parameter. These results highlight that ant colony features and machine learning techniques work together to provide reliable analysis within the NSL-KDD dataset.

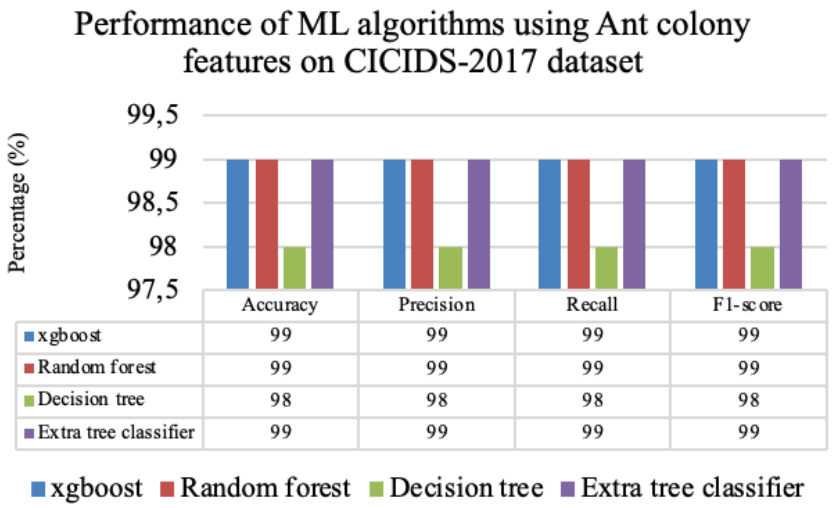

Figure 6. Performance of ML algorithms using Ant colony features on CICIDS-2017 dataset.

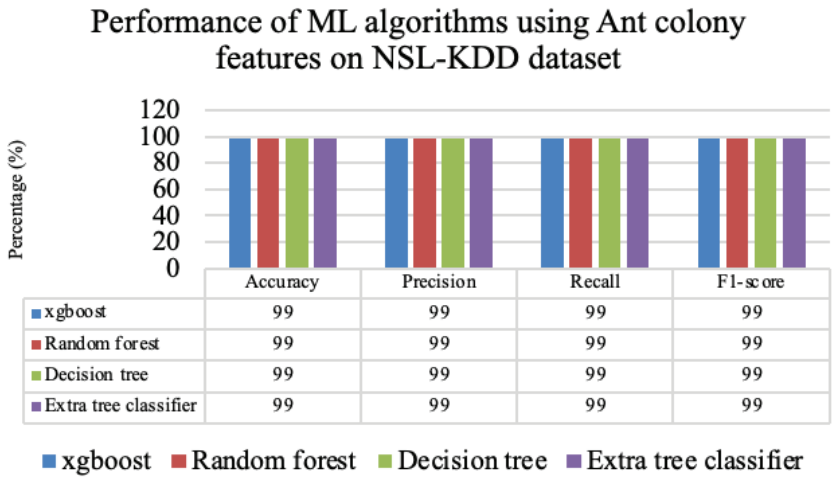

Figure 7. Performance of ML algorithms using Ant colony features on NSL-KDD dataset.

Genetic Algorithm

Using the CICIDS2017 dataset

Machine learning techniques on the CICIDS-2017 dataset performed well on a range of measures by utilizing features from genetic algorithms. 98% accuracy, 97% precision, 98% recall, and a 97% F1 score put Random Forest in the first place. Both the Decision Tree and the Extra Tree Classifier demonstrated 97% accuracy; the Decision Tree displayed 97% precision and recall with a 70% F1 score, and the Extra Tree Classifier displayed 96% precision, 97% recall, and a 96% F1 score. 96% accuracy, 95% precision, 96% recall, and a 95% F1-score were attained via XGBoost. These results show that using machine learning and features from genetic algorithms together can effectively achieve accurate identification and classification in the dataset CICIDS-2017.

Using NSL-KDD dataset

NSL-KDD dataset machine learning performance was significantly improved by features of genetic algorithms. With 99% performance in accuracy, precision, recall, and F1-score, XGBoost and Decision Tree ensured balanced performance and accurate categorization. Furthermore, Random Forest performed well, scoring 98% on all criteria, demonstrating high categorization ability. Extra Tree Classifier, however, underperformed slightly with 95% across all categories. In the NSL-KDD dataset, these results show the utility of genetic algorithm features for improving accuracy and performance, especially when combined with XGBoost and Decision Tree algorithms.

Using the CICIDS-2017 and NSL-KDD datasets, machine learning algorithms utilizing genetic and ant colony algorithm features were compared. Ant colony algorithm features consistently showed better performance. While genetic algorithm features produced somewhat lower

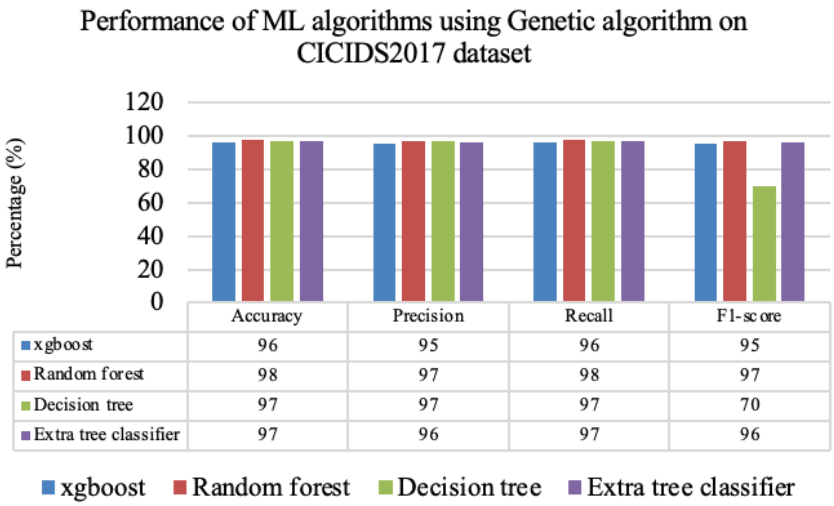

Figure 8. Performance of ML algorithms using Genetic algorithm on CICIDS2017 dataset.

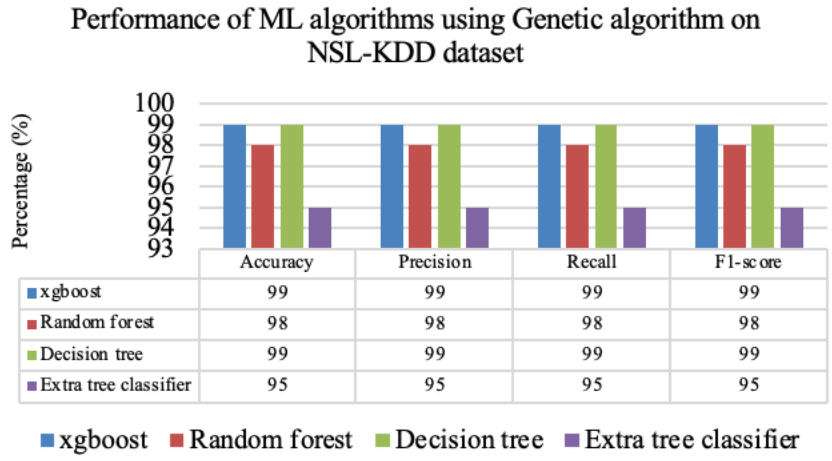

Figure 9. Performance of ML algorithms using Genetic algorithm on NSL-KDD dataset.

Table 3. A comparison between the suggested model and previous research

References	Techniques	Accuracy
The proposed model.	ACO using ML algorithms	99%
Tejbir Singh and Rohit Vaid. [21]	ACO	94.68%
Zhenyang Sun et al. [22]	Adaboost	98.5%
Laith Abualigah et al. [23]	SVM	98.76%

results, XGBoost, Random Forest, and Extra Tree Classifier with ant colony features earned perfect scores of 99% across accuracy, precision, recall, and F1-score on CICIDS-2017. Ant colony features again outperformed genetic algorithm features on NSL-KDD, where they again received flawless 99% scores across all measures, especially with the Extra Tree Classifier at 95%. Therefore, the Ant Colony algorithm added features to the proposed model due to their superior performance on both datasets. This will ensure reliable and strong detection and classification skills for complex datasets such as CICIDS-2017 and NSL-KDD.

Comparative Analysis

The proposed model gets the maximum accuracy at 99% by combining machine learning algorithms with Ant Colony Optimization (ACO). By contrast, the accuracy of Tejbir Singh and Rohit Vaid's. ACO technique is 94.68%. The Adaboost technique by Zhenyang Sun et al. reaches 98.5%, while the Support Vector Machine (SVM) method by Laith Abualigah et al. obtains 98.76%. The proposed model's better accuracy demonstrates how good it is in identifying cybersecurity threats.

The integration of ant colony algorithm with machine learning algorithms in proposed intrusion detection system presents a comprehensive strategy to address the challenges of maintaining high accuracy in real-time scenarios. ML algorithms are proficient in pattern analysis and anomaly prediction they often face performance degradation in the dynamic and rapidly changing nature of real-time data. The static nature of pre-trained models becomes a limitation to evolving patterns in real time. To enhance the adaptability and responsiveness of ML algorithms, ant colony optimization is employed in algorithm. It is recognized for their prowess in handling optimization problems and dynamic environments. Ant colony algorithms contribute to overcoming the limitations posed by static ML models. Additionally, the introduction of the ant colony algorithm further strengthens ability system to optimize parameters and adapt to real-time data dynamics.

Ant colony algorithms function as key optimization tools by fine-tuning ML model parameters based on fitness functions designed for efficient handling of real-time data. The genetic algorithm enables the ML model to dynamically evolve and adjust to changing patterns in real-time data streams. They iteratively refine parameters through selection, crossover, and mutation processes. The findings show that robust performance of machine learning algorithms using ant colony and genetic algorithm features on both the CICIDS-2017 and NSL-KDD datasets. Ant colony algorithm features consistently achieved perfect scores of 99% across all metrics, highlighting their effectiveness in accurately classifying instances. In contrast, genetic algorithm features showed slightly lower performance, particularly with the Extra Tree Classifier on the dataset of NSL-KDD. This comparative analysis shows the superior performance of ant colony algorithm features in enhancing model accuracy and reliability for complex dataset analysis. The accuracy achieved by the proposed model using ACO with machine learning was 99% which outperforms the precision attained by Tejbir Singh and Rohit Vaid's ACO (94. 68%), Zhenyang Sun et al. 's Adaboost (98. 5%), and Laith Abualigah et al. 's SVM (98. 76%).

This adaptive approach ensures the intrusion detection system remains accurate and effective in real-time scenarios, continuously learning and optimizing its performance. ACO possesses the capability to rapidly converge, demonstrating a robust search capability within the problem space, and efficiently identifying minimal feature subsets. The experimental results showcase competitive performance. Further investigation and additional experimentation into this technique are imperative. The collaborative synergy between machine learning algorithms and the ant colony algorithm empowers the intrusion detection system to uphold a high level of accuracy. This adaptive and dynamic approach proves crucial for countering emerging security threats in real-time environments, providing a robust defense mechanism against the unpredictability inherent in network activities.

CONCLUSION

This study emphasizes that how blockchain technology may boost confidence, protect personal information, and increase transaction transparency. Traditional middlemen can be replaced by blockchain's safe consensus techniques, incorporating Practical Byzantine Fault Tolerance, Proof of Work, and Proof of Stake and its unchangeable record. By integrating blockchain with machine learning algorithms (Decision Tree, Random Forest, Extra Tree, XGBoost) and continuously monitoring network traffic for anomalies an Intrusion Detection System greatly enhances cybersecurity.

The use of blockchain technology combined with machine learning and ant colony optimization to create a better real-time mechanism for detecting intrusions that improve network security and data privacy. This novel approach compares several machine learning models, demonstrating improvements in cybersecurity and detection accuracy. The real-time flexibility of the Intrusion Detection System is further improved using ant colony optimization, which increases detection accuracy and scalability in dynamic networks. High accuracy and strong detection capabilities are demonstrated in experimental findings using the CICIDS2017 and NSL-KDD datasets which confirms the model's effectiveness in reducing cybersecurity threats.

The study is limited by its reliance on the CICIDS2017 and NSL-KDD datasets, potentially excluding newer attack vectors and network intrusion scenarios. Simulations used to replicate network traffic may not fully capture real-world complexities. The effectiveness of machine learning and optimization algorithms varies based on settings and dataset characteristics. Anomaly detection struggles with defining normal behaviour in evolving networks. Standard metrics like accuracy may not fully reflect intrusion detection system efficacy. Findings may not generalize beyond the specific datasets and conditions studied, limiting applicability to broader network environments or different attack types.

Future Work

For future work it may prioritize utilizing streaming algorithms and adaptive learning strategies to efficiently integrate real-time data streams to increase the system's capabilities. To capture a greater variety of network behaviours and attack patterns dataset need to be diversified. Machine learning methods may be improved to lower computational burden and increase detection accuracy. Different methods may develop to changing network conditions and adapting new attack strategies on their own. To ensure user privacy and resilience in blockchain deployment, address security problems. Verify using field experiments or simulations in the real world, and work across disciplines to gain comprehensive insights and future developments.

AUTHORSHIP CONTRIBUTIONS

Authors equally contributed to this work.

DATA AVAILABILITY STATEMENT

The authors confirm that the data that supports the findings of this study are available within the article. Raw data that support the finding of this study are available from the corresponding author, upon reasonable request.

CONFLICT OF INTEREST

The author declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

ETHICS

There are no ethical issues with the publication of this manuscript.

REFERENCES

[1] Sakraoui S, Ahmim A, Derdour M, Ahmim M, Namane S, Dhaou IB. FBMP-IDS: FL-based Blockchain-powered Lightweight MPC-secured IDS for 6G networks. IEEE Access 2024;12:105887. [CrossRef]

[2] Xiao Y. Blockchain and distributed consensus: From security analysis to novel applications [Doctorial thesis]. Blacksburg (VA): Virginia Tech; 2022.

[3] Cao B, Zhang Z, Feng D, Zhang S, Zhang L, Peng M, et al. Performance analysis and comparison of PoW, PoS and DAG based blockchains. Digit Commun Netw 2020;6:480–485. [CrossRef]

[4] Gupta KD, Rahman A, Poudyal S, Huda MN, Mahmud MAP. A hybrid POW-POS implementation against 51 percent attack in cryptocurrency system. In: Proc Int Conf Cloud Comput Technol Sci CloudCom. 2019. p. 396–403. [CrossRef]

[5] Irannezhad E. The architectural design requirements of a blockchain-based port community system. Logistics 2020;4:1–31. [CrossRef]

[6] Wang T, Shen H, Chen J, Chen F, Wu Q, Xie D. A hybrid blockchain-based identity authentication scheme for mobile crowd sensing. Future Gener Comput Syst 2023;143:40–50. [CrossRef]

[7] Tian Y, Wang Z, Xiong J, Ma J. A blockchain-based secure key management scheme with trustworthiness in DWSNs. IEEE Trans Ind Inform 2020;16:6193–6202. [CrossRef]

[8] Manoj T, Makkithaya K, Narendra V. A blockchain based decentralized identifiers for entity authentication in electronic health records. Cogent Eng 2022;9:2035134. [CrossRef]

[9] Hameed K, Garg S, Amin MB, Kang B. A formally verified blockchain-based decentralised authentication scheme for the internet of things. J Supercomput 2021;77:14461–14501. [CrossRef]

[10] Mohanta BK, Jena D, Ramasubbareddy S, Daneshmand M, Gandomi AH. Addressing security and privacy issues of IoT using blockchain technology. IEEE Internet Things J 2021;8:881–888. [CrossRef]

[11] Zafar S, Bhatti KM, Shabbir M, Hashmat F, Akbar AH. Integration of blockchain and Internet of Things: Challenges and solutions. Ann Telecommun 2022;77:13–32. [CrossRef]

[12] She W, Liu Q, Tian Z, Sen Chen J, Wang B, Liu W. Blockchain trust model for malicious node detection in wireless sensor networks. IEEE Access 2019;7:38947–38956. [CrossRef]

[13] Srivastava A, Sinha D. PSO-ACO-based bi-phase lightweight intrusion detection system combined with GA optimized ensemble classifiers. Clust Comput 2024;27:14835–14890. [CrossRef]
[14] Rehman E, Hasseb-ud-Din M, Malik AJ, Khan TK, Abbasi AA, Kadry S, et al. Intrusion detection based on machine learning in the internet of things, attacks and counter measures. J Supercomput 2022;78:8890–8924. [CrossRef]
[15] Dina AS, Manivannan D. Intrusion detection based on machine learning techniques in computer networks. Internet Things 2021;16:100462. [CrossRef]
[16] Pervez MS, Farid DM. Feature selection and intrusion classification in NSL-KDD cup 99 dataset employing SVMs. In: SKIMA 2014 - 8th Int Conf Softw Knowl Inf Manag Appl 2014. [CrossRef]
[17] Maseer ZK, Yusof R, Bahaman N, Mostafa SA, Foozy CFM. Benchmarking of machine learning for anomaly-based intrusion detection systems in the CICIDS2017 dataset. IEEE Access 2021;9:22351–22370. [CrossRef]
[18] Liu F, Li G, Zhao Y, Jin Z. Multi-task learning-based pre-trained language model for code completion. In: Proc 2020 35th IEEE/ACM Int Conf Autom Softw Eng ASE 2020. 2020. p. 473–485. [CrossRef]
[19] Hamdia KM, Zhuang X, Rabczuk T. An efficient optimization approach for designing machine learning models based on genetic algorithm. Neural Comput Appl 2021;33:1923–1933. [CrossRef]
[20] Ziegel E. Genetic algorithms and engineering optimization. Technometrics 2002;44:95. [CrossRef]
[21] Singh T, Vaid R. Preserving security in terms of authentication on blockchain-based wireless sensor network (WSN). Int J Comput Netw Appl 2024;11:390–406. [CrossRef]
[22] Sun Z, An G, Yang Y, Liu Y. Optimized machine learning enabled intrusion detection system for internet of medical things. Franklin Open 2024;6:100056. [CrossRef]
[23] Abualigah L, Ahmed SA, Almomani MH, Abu Zitar R, Alsoud AR, Hanandeh ES, et al. Modified Aquila optimizer feature selection approach and support vector machine classifier for intrusion detection system. Multimed Tools Appl 2024;83:59887–59913. [CrossRef]
[24] Yang J, He S, Xu Y, Chen L, Ren J. A trusted routing scheme using blockchain and reinforcement learning for wireless sensor networks. Sensors 2019;19:970. [CrossRef]
[25] Chang Y, Tang H, Cheng Y, Zhao Q, Li B, Yuan X. Dynamic hierarchical energy-efficient method based on combinatorial optimization for wireless sensor networks. Sensors 2017;17:1665. [CrossRef]
[26] Chang Y, Yuan X, Li B, Niyato D, Al-Dhahir N. A joint unsupervised learning and genetic algorithm approach for topology control in energy-efficient ultra-dense wireless sensor networks. IEEE Commun Lett 2018;22:2370–2373. [CrossRef]