

Article

# Blockchain and Machine Learning-Based Hybrid IDS to Protect Smart Networks and Preserve Privacy

Shailendra Mishra 

Department of Computer Engineering, College of Computer and Information Sciences, Majmaah University, Al Majma'ah 11952, Saudi Arabia; s.mishra@mu.edu.sa or skmishra1970a@gmail.com

**Abstract:** The cyberspace is a convenient platform for creative, intellectual, and accessible works that provide a medium for expression and communication. Malware, phishing, ransomware, and distributed denial-of-service attacks pose a threat to individuals and organisations. To detect and predict cyber threats effectively and accurately, an intelligent system must be developed. Cyber-criminals can exploit Internet of Things devices and endpoints because they are not intelligent and have limited resources. A hybrid decision tree method (HIDT) is proposed in this article that integrates machine learning with blockchain concepts for anomaly detection. In all datasets, the proposed system (HIDT) predicts attacks in the shortest amount of time and has the highest attack detection accuracy (99.95% for the KD99 dataset and 99.72% for the UNBS-NB 15 dataset). To ensure validity, the binary classification test results are compared to those of earlier studies. The HIDT's confusion matrix contrasts with previous models by having low FP/FN rates and high TP/TN rates. By detecting malicious nodes instantly, the proposed system reduces routing overhead and has a lower end-to-end delay. Malicious nodes are detected instantly in the network within a short period. Increasing the number of nodes leads to a higher throughput, with the highest throughput measured at 50 nodes. The proposed system performed well in terms of the packet delivery ratio, end-to-end delay, robustness, and scalability, demonstrating the effectiveness of the proposed system. Data can be protected from malicious threats with this system, which can be used by governments and businesses to improve security and resilience.

**Keywords:** cyber security; machine learning; blockchain; machine intelligence; smart network



**Citation:** Mishra, S. Blockchain and Machine Learning-Based Hybrid IDS to Protect Smart Networks and Preserve Privacy. *Electronics* **2023**, *12*, 3524. <https://doi.org/10.3390/electronics12163524>

Academic Editor: Ali Mehrizi-Sani

Received: 14 July 2023

Revised: 29 July 2023

Accepted: 3 August 2023

Published: 21 August 2023



**Copyright:** © 2023 by the author. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

The potential of the cyberspace to transform our lives is tremendous, but our access to and use of this powerful tool must be carefully measured and managed to leverage its most advantageous benefits and protect individuals from potential misuse and abuse [1]. Due to the continuous growth of communication and networking technologies, a massive number of devices are connected to the Internet, which introduces the concept of the Internet of Things (IoT). In the past few decades, the IoT has witnessed a similar spike in interest due to the automation benefits that it provides [2]. Because the IoT is connected via the Internet, it has seen good growth over the years. Due to this growth, some crucial security issues can help intruders gain access to network resources [3]. There has been an increase in the number of IoT devices as IoT networks have been implemented in various systems. The number of IoT devices is predicted to increase from 7.74 billion in 2019 to 25.44 billion in 2030 [4]. IoT endpoints are not smart and have limited resources, allowing cyber threats to be exploited [5].

Blockchains, cybersecurity, AI, and ML are closely intertwined and are essential components of a comprehensive digital transformation strategy. A blockchain provides a secure data storage and sharing system, while cybersecurity enables secure data protection from malicious activities. Using blockchain (BC) technology, cybersecurity, AI, and ML together, organisations can enhance security, harness the power of data, reduce costly

operational expenses, and optimise their operations [6]. The majority of cyberattacks target dark web data theft, damaging brands' reputations and exploiting e-commerce sites and the stock market [7]. The use of machine learning (ML) techniques can improve both the effectiveness of the IoT infrastructure and the performance of cybersecurity systems [8]. Blockchain technology and artificial intelligence have the potential to create smarter, safer, more efficient, and more secure systems. Quantum technology, however, has made most existing blockchain systems vulnerable to quantum attacks. Quantum cryptography can be used to protect personal information and protect privacy in blockchain, artificial intelligence, and big data applications [9].

Implementing BCs in IoT systems has numerous advantages, such as decentralisation to eliminate a single point of failure, proof of security, traceability, and immutability [10].

A blockchain can be used to generate insights based on shared data and then to make predictions using artificial intelligence. Via mutual agreements between nodes, blockchains form chains that link existing blocks stored in nodes chronologically with the new blocks. Artificial intelligence and blockchain power can be combined to provide a strong defence against attacks [11]. Many existing approaches have developed the concept of efficient data communication between devices and the storage of these communicated data on either a cloud or blockchain network [12]. However, numerous obstacles exist to effectively communicating and storing data in a smart network, and significant challenges for smart networks have been discussed [13]. One vulnerability is the corruption of the data stored in a BC. The immutability of BCs is the root cause of this problem; therefore, corrupted data must be detected before they are transferred to and stored in a BC [14].

The study aims to achieve the following:

- Develop a secure, tree-based intrusion detection system (HIDT) that predicts and detects threats based on ranking security features by importance based on security features.
- Analyse the proposed model's performance using two main datasets, KDD99 and UNSW-NB 15.
- Develop a secure decentralised blockchain reputation system (SDBCRS) based on machine learning.
- Conduct a performance metric-based comparison between the proposed HIDT model and other existing approaches.
- Evaluate the blockchain-based machine learning model by analyzing packet delivery ratios, end-to-end delays, throughput, and scalability.

The proposed system combines the strengths of machine learning and artificial intelligence to provide the best possible detection accuracy. The blockchain-based machine learning framework provides further assurance that the system is scalable, reliable, and secure. By developing such a system, organisations can protect themselves from cyberattacks, ensuring their information systems' security and continuity. Additionally, the system also provides a proactive approach to security, with its ability to predict potential threats and take steps to mitigate them. Furthermore, this research is also helpful in reducing the time and effort spent detecting threats and responding to threats, as well as reducing the costs associated with cyberattacks.

The proposed system (HIDT) predicts attacks in the shortest amount of time and has the highest attack detection accuracy (99.95% for the KD99 dataset and 99.72% for the UNBS-NB 15 dataset). To ensure validity, the findings of the binary classification test were compared to those of earlier research [15–19]. The HIDT's confusion matrix features low FP/FN rates and high TP/TN rates compared to earlier models [20]. In the network, malicious nodes are quickly and instantaneously identified. The throughput increased as the number of nodes increased, reaching its peak at 50 nodes. The proposed system demonstrated effectiveness by performing well in terms of its packet delivery ratio, end-to-end delay, resilience, and scalability.

The following points of focus form the basis of this research.

1. The challenges, threats, and countermeasures facing security and privacy in smart networks;
2. How machine learning and blockchain technology might enhance security and privacy;

3. The development of a secure tree-based intrusion detection system (HIDT) that uses ranked security features to predict and detect threats;
4. The development and evaluation of a secure decentralised blockchain reputation system based on machine learning.

The sections of this article are as follows: the second section discusses recent state-of-the-art studies conducted by various researchers, and the third section describes the research methodology and framework. Experimental analysis, results, and discussions are discussed in Section 4. Conclusions and future work are presented in Section 5.

## 2. Related Work

In a rapidly evolving network environment, there is not much time to develop new statistical models, so they are not well suited to the new workload. By integrating concepts from edge computing, machine learning, and artificial intelligence, a cognitive engine can be developed [21]. Machine learning is capable of learning without much human assistance. Therefore, paying more attention to security issues and related defences in machine learning is important. With the development of machine learning (ML) and deep learning (DL) models, security in the IoT cloud environment has been enhanced [22]. The use of AI in user access authentication, network situation detection, malicious behaviour monitoring, and abnormal traffic identification is discussed in [23]. In [24], the authors proposed an intrusion detection system based on neural network clustering (IDS) that can help administrators detect and reduce the risk of early-stage attacks, thereby reducing power consumption.

Dong and Sarem [15] proposed a detection algorithm called DDAML. This study aimed to identify DDoS attacks by applying machine learning algorithms and MLP. The DDAML algorithm outperformed all the other algorithms (SVM, RF, KNN, and LR) with the same ROC curve. The DDAML algorithm has an AUC of 0.912, as do the NB, SVM, CIC-SVM, and DDADA algorithms. The NB algorithm has an AUC of 0.891, the SVM algorithm has an AUC of 0.893, the CIC-SVM algorithm has an AUC of 0.895, and the DDADA algorithm has an AUC of 0.899 [15]. Gradient-boosted machine (GBM) technology is proposed in [16] as a means of improving the detection performance of anomaly-based intrusion detection systems (IDSs). The effectiveness of the GBM technology is then evaluated in terms of performance metrics and contrasted with well-known classifiers. The NSL-KDD, UNSW-NB15, and GPRS datasets' full features were applied to yield the highest results to date using either the hold-out approach or tenfold cross-validation.

A detection approach named OGBDT, which combines genetic algorithms (GAs) with optimised gradient boost decision trees (OGBDTs), was proposed in [18]. Enhanced African buffalo optimisations (EABOs) were used to increase categorisation. The proposed IDS (OGBDT) was used to compare conventional MLTs. To evaluate the performance of these approaches, accuracy, precision, recall, and F-score were compared across the UNBS-NB 15, KDD 99, and CICIDS2018 datasets. The suggested IDS has the fastest attack prediction speeds across all datasets and the highest attack detection rates. By replicating message queuing telemetry transport (MQTT) via a virtual network, IoT anomalies were found and discussed in [19]. To detect and stop DDoS attacks, a few machine learning algorithms, including the multilayer perceptron (MLP), naive Bayes (NB), and decision tree (DT) algorithms, as well as an artificial neural network, were analysed. A dataset comprising 4998 records, 34 characteristics, and eight kinds of network traffic was used in the suggested method. With an accuracy rate of 99.94%, the classifier RF displayed the best performance.

The three primary technologies for addressing security issues in the Internet of Things (IoT)—machine learning (ML), (AI), and BCs—were the subject of a thorough analysis. A study describing the IoT architecture and its supporting technology presented issues [25]. In [26], Derhab et al. proposed the RSL-KNN intrusion detection system, a method of detecting forgeries intended to manipulate industrial control systems that uses random subspace learning (RSL) and the K-nearest neighbour (KNN) algorithm. A blockchain-based

integrity checking system (BICS) protects industrial IoT systems with SDN capabilities from misrouting attacks that alter OpenFlow rules.

As a means of improving and securing the overall security of a system and evaluating its performance in terms of its end-to-end delay, routing overhead, packet delivery ratio, throughput, and confusion matrix, Malik et al. (2022) proposed a solution called the detection and prevention of a BHA (DPBHA) [20]. The proposed model was tested on the benchmark dataset KDD99 (NSL-KDD). The KDD99 (NSL-KDD) dataset [27] includes 494021 records in its training dataset, while its testing dataset contains 311 029 records. A study of the UNSW-NB15 dataset [28] revealed 42 features divided into ten classes (normal, fuzzers, analysis, backdoors, DoS, exploits, generic, reconnaissance, shellcode, and worms). In [29], the authors described a state-of-the-art technique for assessing database damage after a hostile attack on a healthcare system; healthcare systems require fast recovery to minimise downtime, and such an algorithm can also be used to protect healthcare systems [29].

Systems using blockchains are susceptible to quantum assaults. For initiatives including blockchains, artificial intelligence, large data, and privacy protection, quantum cryptography offers a potent security tool [9]. A vast number of complicated operations can be computed using quantum computing in an exponentially short amount of time for its quick, effective, and scalable computing resources. To protect against arbitrary source defects when using current technology, such as state preparation flaws, side channels caused by mode dependencies, Trojan horse attacks, and pulse correlations, a major framework known as a reference technique has been developed [30].

The potential uses of BC technology and its drawbacks in fields like human rights have direct societal effects. SMEs, corporations, organisations, businesses, government institutions, and the general public confront a variety of hurdles while adopting, promoting, and using blockchain technology. The security of decentralised networks is a major challenge because the nodes are not physically protected. Without centralised management and collaboration between nodes, data security is compromised across the network. The current decentralised system has multiple nodes, all of which function properly. However, if one of the nodes fails to complete the user authentication process, a denial-of-service (DoS) attack can occur. In this type of attack, spoofed traffic and data requests are sent to the attacked resource to flood it with requests and prevent real users from accessing it. The attacker exploits the vulnerability in the resource's network by constantly sending information packets that require authentication. If the system shares a spoofed address, it can prevent resources from authenticating and thus shut down without further interaction.

This leads to an increase in traffic on the routing path, which is filled with spoofed data requests. The authentication process has no benefit, and malware activity is recorded during network transmission. A decentralised ledger system should restrict user access. Each user should be verified before accessing the network. Integrated approaches should be developed in conjunction with tactics and techniques used to close these gaps. Security for the Internet of Things (IoT) is becoming increasingly concerned with machine learning (ML) and blockchain technology. These technologies can be applied specifically to intrusion detection systems (IDSs). Despite this, there are still some gaps in the existing research.

Previous research has shown that ML algorithms can effectively detect anomalous behaviour in IoT devices, making them suitable for IDS applications. However, one challenge is that ML algorithms require large amounts of data to for effective training. This can be a problem in the context of the IoT as devices may have limited processing and storage capabilities. Additionally, ML algorithms may be vulnerable to attacks such as adversarial attacks, which can be used to fool the algorithm into making incorrect predictions.

Blockchain technology has also been proposed as a way to enhance the security of IoT devices. One approach is to use a blockchain to create a decentralised and tamper-proof ledger of all device transactions, which can help prevent unauthorised access to IoT devices. However, there are still some challenges that need to be addressed. For example, the

overhead of using blockchain can be significant, which can be a problem in the context of resource-constrained IoT devices.

The existing research on combining ML and blockchain technology for IoT IDS applications has some gaps as well. One challenge is to develop a system that can efficiently and securely store the large amounts of data required for ML algorithms to work effectively. Additionally, there is a need for further research on how to effectively integrate ML algorithms with blockchain technology in the context of the IoT.

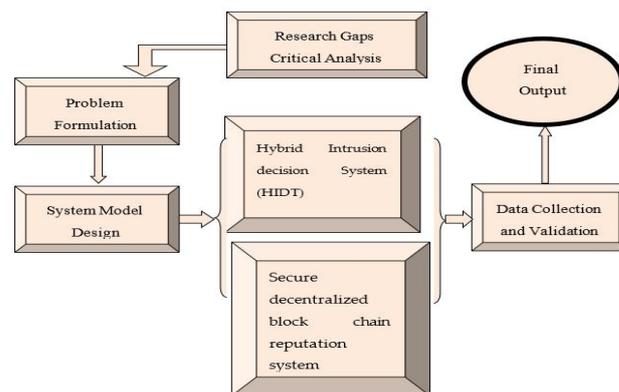
While there has been some promising research on using ML and blockchain technology for IDS in the IoT, there are still some gaps that must be addressed. Future research should focus on developing efficient and secure systems for storing data, as well as exploring ways to integrate ML algorithms with blockchain technology in the context of the IoT.

To bridge the research gap, my aims are as follows:

- The proposed work will use an analysis based on both intelligent systems (CPSs) and computational methods.
- In connection with the cloud, the developed framework will be more efficient due to the advantage of the IoT.
- As part of this work, I propose a secure tree-based intrusion detection and prevention system incorporating selected security features ranked by importance that effectively predicts and detects cyberattacks.
- A machine learning-based, secure decentralised blockchain reputation system is also proposed.
- With the help of various performance metrics, I compare the proposed model with other existing approaches. Blockchain-based machine learning frameworks demonstrate their effectiveness by analysing packet delivery ratios, end-to-end delay, throughput, and scalability.

### 3. Research Methods

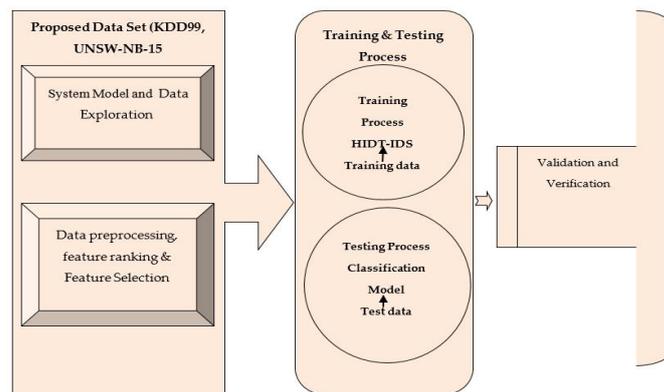
The objectives mentioned in the previous section can be achieved via the following research methodology, which is shown in Figure 1 below.



**Figure 1.** Research process.

#### 3.1. Proposed Hybrid Decision Tree-Based IDS Model

We propose a hybrid decision tree (HDT) method for predicting and categorise malicious cyberattacks in networks. MLTs are proposed to evaluate the importance of security attributes. A genetic algorithm is used to select relevant features within a network security technique based on MLTs. Figure 2 illustrates the proposed framework, and each phase is evaluated in detail in the following sections.



**Figure 2.** Hybrid decision tree-based IDS model.

### 3.1.1. Security Dataset

One of the most important factors in the predictive accuracy of the proposed predictive models’ is the data quality. In this framework, data exploration examines the integrity of the data and reveals more about their characteristics, followed by data analysis. The types of features are explored, including numerical and categorical data. There are two datasets available on Kaggle: KDD99 (NSL-KDD [27] and UNSW-NB 15 [28]. Activities are classified as normal or malignant based on the class characteristics of the dataset.

### 3.1.2. Hybrid Intrusion Detection Tree (HIDT) Design

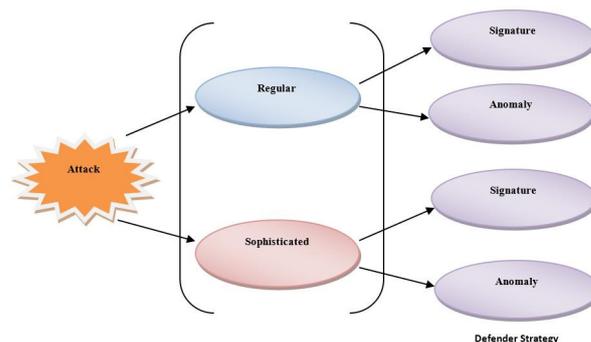
In general, DTs perform well when the training data are known, but they do not perform well when unknown data are tested and thus cannot effectively circumvent the overfitting problem. The proposed IDS method includes data exploration, pre-processing, standardisation, ranking, and feature selection. A tree-based IDS method that selects features by rank requires these phases. Finally, the data are trained and tested to determine the method’s effectiveness at categorising cyber-attacks.

### 3.1.3. Computational Modelling

Two main attack types are considered for analysis: signature-based and anomaly-based attacks. The defender available in IDS applies its best strategy to defend a system. The defender’s strategy can be represented as a vector,  $S_d = (S_{d1}, S_{d2})$ , and the strategy is provided in Table 1 and illustrated in Figure 3.

**Table 1.** Defender Strategy.

Symbol	Meaning
$S_{d1}$	System monitoring based on signatures
$S_{d2}$	System monitoring based on anomaly



**Figure 3.** Defender strategy.

The payoffs ( $m_{ij}$ ) of the defender can be represented by a matrix  $M$ , which is shown in Equation (3). The nomenclature for the terminology are stated in Table 2. Here, ( $E$ ) is the energy consumption of the IDS, ( $G_d$ ) is the benefit that a defender receives after the successful detection of an attack, and ( $V$ ) is the asset value that is being attacked at a time ( $t$ ). In the case of a regular attack, the attacker attacks using traditional techniques, and the IDS is detected based on the signature method and successfully traces the attack. Here, the payoff of the IDS for a regular attack depends on the energy consumed ( $E$ ) and the benefit gained ( $G_d$ ) and is shown in Equation (1). For an anomaly attack, the payoff is provided in Equation (2). The complete payoff matrix for anomaly attacks is provided by Equation (4).

$$P_{def} = G_d(t) - E(t) \tag{1}$$

$$P'_{def} = \mu G_d(t) - (1 - \mu)V(t) - E(t) \tag{2}$$

$$M = \begin{bmatrix} m_{11} & m_{12} \\ m_{21} & m_{22} \end{bmatrix} = \begin{bmatrix} G_d(t) - E(t) & -E(t) - v(t) \\ -E(t) & G_d(t) - E_{ids}(t) \end{bmatrix} \tag{3}$$

$$M' = \begin{bmatrix} \mu G_d(t) - (1 - \mu)V(t) - E(t) & -E(t) - V(t) \\ -E(t) - V(t) & \alpha G_d - (1 - \alpha)V(t) - E(t) \end{bmatrix} \tag{4}$$

**Table 2.** Nomenclature.

Symbol	Meaning
$E$	IDS energy consumption
$G_d$	Gain in terms of detection
$V$	Asset value that is attacked
$\mu$	Misuse of IDS detection
$\alpha$	Anomaly detection rate
$\beta$	False positive rate of defender

Let  $a$  and  $(1 - a)$  be the probability of the defender defending from a regular attack and an anomaly attack, and let  $b$  and  $(1 - b)$  be the probability of the attacker performing a regular attack and an anomaly attack. The defender payoff for these attacks can be illustrated using Equation (5).

$$U_d = a b M'_{11}(D) + a(1 - b)M'_{21}(D) + (1 - a)b M'_{12}(D) + (1 - a)(1 - b)M'_{22}(D) \tag{5}$$

Taking the partial derivatives of Equation (5), I can compute the values of  $a$  and  $b$ .

$$\frac{\partial U_d}{\partial a} = ab\mu G_d(t) + ab\alpha G_d(t) + ab\mu V(t) + ab\alpha V(t) - a\alpha G_d(t) - a\alpha V(t) - b\alpha G_d(t) - b\alpha V(t) \tag{6}$$

$$a = \mu / (\mu + \alpha) \text{ and } (1 - a) = \mu / (\mu + \alpha) \tag{7}$$

$$b = \alpha / (\mu + \alpha) \text{ and } (1 - b) = \mu / (\mu + \alpha) \tag{8}$$

### 3.1.4. Hybrid Intrusion Detection Tree Generation (HIDT)

The proposed hybrid decision tree algorithm (HIDT) is used for cybersecurity intrusion detection to predict and classify potentially dangerous network breaches. MLTs are a mechanism for classifying the importance of security features. This MLT-based network security method constructs trees using derived attribute ranks and uses evolutionary

algorithms to select key features. The hybrid intrusion detection tree generation is shown in Algorithm 1.

---

**Algorithm 1:** Hybrid Intrusion Detection Tree Generation

---

```

1. Begin
2. Input:  $G_d(t), V, E, \mu, \alpha$ 
3.  $d = \{a_1, a_2, a_3, \dots, a_n\}$  // a dataset with  $n$  values
4.  $f = \{f_1, f_2, f_3, \dots, f_n\}$  // a set of feature list
5.  $c = \{c_1, c_2, c_3, \dots, c_n\}$  // intrusion class information
6. Compute matrices  $M$  and  $M'$ 
7. if there is pure strategy Nash-Equilibrium, then go to Step 1
8. else
   create  $IDS\_decision\_tree(d, f, c)$ 
9. find the mixed strategy Nash Equilibrium solution
10. endif
11. function  $IDS\_decision\_tree(d, f, c)$ 
   begin
      $f\_score < -compute\ score\ on\ f$ 
     //select appropriate features
      $sel\_feature < -ShortListFeature(f, f\_score, n)$ 
      $GenTree(d, sel\_feature, c)$ 
   end
12.  $GenTree(d, sel\_feature, c)$ 
   begin
      $root < -newNode();$  //add root node
     if ( $\forall d \in c$ ) //same instance of  $c$ 
       return ( $root$ ) //return the root as a leaf node
     elseif ( $sel\_feature = NULL$ )
       return ( $root$ ) //return the root as a major class of  $d$ 
     end
   end
13. Find  $F$  the features with high precedence
14. for ( $val \in F$ )
     create  $d\_sub$  of  $d$  using  $val$ 
     if ( $d\_sub \neq \emptyset$ )
        $child\_node < -GenTree(d\_sub, F, c)$  //find the new leaf node
       attach  $child\_node$  to  $N$  //attach to the parent node
     end
15. end

```

---

### 3.1.5. Feature Ranking

Information gains and Gini indices are commonly used for trait classification. In binary splits (decisions for nodes), the attributes with the lowest Gini indices (GIs) are set as root nodes [18]. In this work, I add ranks to the features before developing the trees. Gini indices are used to detect inaccuracies in the feature ranks. To compute Gini indices, a value is subtracted from the squared probabilities of the classes.

### 3.2. Proposed System for IDS in the IoT Using ML and Blockchain

The proposed system for IDS in the IoT using ML and blockchain is designed to enhance the security of IoT devices by detecting and preventing intrusions and preserving the integrity of the data generated by these devices. The system comprises four main components: IoT devices, an intruder detection system (IDS), blockchain (BC) nodes, and the BC network. Machine learning algorithms fuel the IDS, which can detect anomalous patterns in data supplied by IoT devices. IoT devices are responsible for collecting data from the environment and sending it to the IDS. The IDS is the first line of defence against intrusions, and its primary function is to identify any compromised devices and remove any

corrupt data. The IDS is powered by machine learning algorithms that can detect unusual patterns in the data generated by the IoT devices. This allows the IDS to differentiate between normal and abnormal data and to identify devices that are compromised or behaving suspiciously. When the IDS has filtered the regular data, they are forwarded to the BC nodes. These nodes encrypt the data and send them to the BC network. The BC network is a public blockchain that accepts signed data from BC nodes and creates a block containing the data. A blockchain network consists of five element nodes, a ledger, a wallet, a nonce, and a hash. The nodes store a full copy of transactions. In a ledger, information is stored digitally; every node has a wallet, and cryptographic keys are used to maintain the privacy of wallets. Once a number is placed in a hashed or encrypted block and only used once, the hash ensures security and integrity.

The BC network assures that data are adulterate and irrevocable, and it provides a decentralised architecture for IoT devices, improving data preservation security. The BC networks are in charge of validating the data and ensuring that only legitimate data are uploaded to the BC network. This is accomplished through the use of cryptography algorithms and a hash. The BC servers are also in the position to safeguard the BC network by verifying new blocks and keeping the network safe. The suggested solution for intrusion prevention in the IoT when utilizing machine learning and a blockchain is intended to improve the security of IoT devices by identifying and blocking intrusions and protecting the integrity of the data generated by these devices. The BC network assures that data are interfered with and unchanging, and it provides a mesh network for IoT devices, improving data preservation security. The proposed hybrid IDS integrated into existing smart networks and IoT devices is shown in Figure 4.

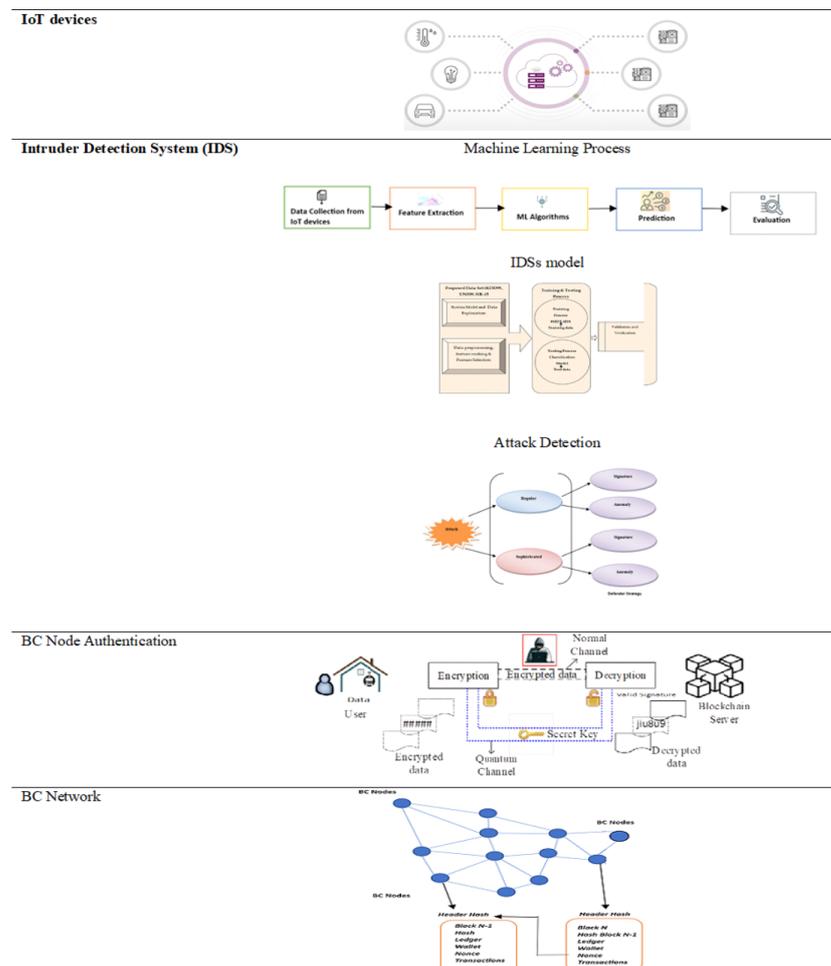


Figure 4. Hybrid IDS integrated into existing smart networks.

### 3.2.1. Data Collection

IoT devices collect data from their environment and send them to the intruder detection system (IDS). These devices generate a vast amount of data, including sensor readings, network logs, and device behaviour information.

### 3.2.2. Intrusion Detection

The IDS acts as the first line of defence against intrusions. Powered by machine learning algorithms, it analyses the collected data to detect unusual patterns and identify compromised or suspicious devices. By differentiating between normal and abnormal data, the IDS can effectively detect and prevent intrusions in real time.

### 3.2.3. Blockchain Integration

After the IDS filters the regular data, they are forwarded to the blockchain (BC) nodes. These nodes encrypt the data to maintain confidentiality and send them to the public blockchain network. The blockchain network accepts the encrypted data, verifies their authenticity, and creates a block containing the data.

### 3.2.4. Data Integrity and Security

The blockchain network ensures the integrity and immutability of the data. Cryptographic algorithms and hashing are employed to validate the data. The decentralised nature of the blockchain network ensures that the data are tamper-proof and secure. BC nodes and servers play a crucial role in safeguarding the blockchain network by verifying new blocks and maintaining the overall security of the system.

In this context, the proposed system uses ML or BC technology to uncover exploited IoT devices and improve data storage security in a decentralised fashion. The IoT devices, which gather data and transfer them to the IDS for analysis, are the platform's initial component. The IDS is in charge of identifying infected devices and removing any damaged data. When the IDS has filtered the regular data, they are forwarded to the BC nodes. These nodes sign the data and transfer them to the BC network, which receives the data and generates a block containing the data. The adoption of BC technology enables decentralised data storage and increases data security by reducing the need for a central authority to govern the data.

Compared to traditional security measures, the proposed framework offers several advantages. ML algorithms can detect compromised devices with high degrees of accuracy and efficiency, minimising false positives and negatives. In addition, BC technology enhances data storage security in a decentralised manner, increasing its resilience to attacks, and the use of feature selection techniques allows for the identification of the most important features for detecting compromised devices, improving the efficiency of the IDS.

## 3.3. Proposed Machine Learning-Based Blockchain Decentralised Reputation System

Reputation transactions require a combination of the decentralised system and a BC. As part of the data request, the network device asks for authentication from the resource. Devices can access resources based on authentication and network confirmation. The system achieves better exhibitions when in-range storage is used. A blockchain has been connected to it instead of a traditional IP system. The secure distribution of content is the goal of some communication systems that are exhibited in networks. As a result of the diverse types of traffic and the secure information in the network, analysts focus on probabilistic and content-based fame-based communication. In the next stage, different organised routers are linked to switches that have specified frequencies.

A reputation transfer value certification authority has been adopted in blockchain technology. Data requesters' privacy requirements must be recognised by the blockchain reputation module, and the reputation value must be accessed. The blockchain network is enabled by the fact that each node maintains an up-to-date copy of the ledger. The scalability of the BC network can be affected by the increasing storage requirements of

the blockchain nodes [31]. Using decentralised IoT devices for high-speed data storage and processing could solve security issues. When used as blockchain nodes, such high-speed storage devices can realise their full potential thanks to a machine learning analytics system. Due to the high cost of node configuration, this technology is economically viable. Moreover, the collapse of decentralised storage systems can lead to the success of machine learning services. Due to developments in IoT applications, burgeoning machine learning models have shown limitless capabilities and potential in numerous industries. The decentralised reputation system is an online transaction method through which one can initiate blockchain transactions [32].

The decentralised blockchain ledger system deals with reputation transactions to determine the reputation value of the user. The reputation value defines the credibility and reputation of the user. The attacker initiates the transaction request and follows a reputation transfer system to activate the reputation value. A block is then published based on the reputation value. In the decentralised reputation online transaction system, the shared reputation from old transactions is accepted, and new transactions with additional reputation values are enabled. The transaction process is significantly estimated using merged old reputation values. Figure 5 shows the blockchain system based on machine learning with a decentralised reputation.

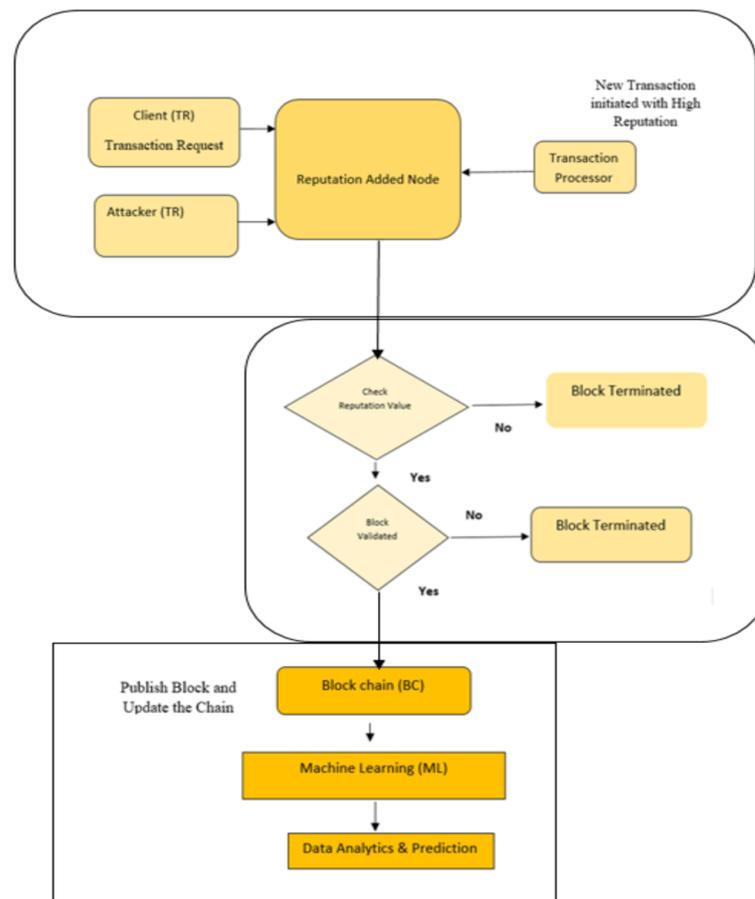


Figure 5. Machine learning- and blockchain-based decentralised reputation system.

The proposed system represents a successful packet delivery ratio, improved throughput, and scalability during routing. Since blockchain data are immutable, it prevents data leakage, protecting deep learning models and data from potential attacks. As a result, this model produces more accurate and reliable forecasts. Blockchain technology and machine learning can now be used together to automate reputation transfer systems that require careful handling and high data security. Blockchain technology provides a secure,

permanent, and distributed way to collect, analyse, and utilise the essential data collected and analysed via machine learning applications. In addition to facilitating existing security and privacy systems, it can also improve them. As a result of machine learning-specific reputation transfer systems, services, application domains, deployment goals, and data transmissions have been improved.

### 3.4. Performance Evaluation

Cross-validation (CV) procedures were used to evaluate the predictive models. Both K-fold cross-validation and leave-one-out cross-validation were utilised. In the K-fold method, 80% of the test set and 20% of the training set were used for cross-validation without replacing the training data [33]. An evaluation of the performance measures (sensitivity, specificity, accuracy, precision, and recognition) was performed [34]. Precision is a good measure, but only if the proportion of false negatives to false positives is low. F1 scores are best when class distributions are not uniform. Precision indicates confidence in true positives, while recall indicates confidence in not missing a positive result.

Choosing precision ensures that you have a higher confidence level in the results, and choosing specificity avoids false alarms and false positives. A classifier correctly classifies attacks when its ability to determine the class in which an attack is detected determines the true positive (TP) and false positive (FP) results. The classifier correctly rejected attacks when a true negative (TN) was determined. When a false negative (FN) was detected, the classifier incorrectly classified attacks as normal traffic.

The performance of a blockchain-based machine learning system is evaluated based on packet delivery ratio, end-to-end delay, throughput, and scalability. The end-to-end delay measures the delay that packets experience while travelling across a network. Accuracy rate, precision, recall, and the F1 score were computed to validate the accuracy and reliability of the proposed system. If there was a suspicion of class imbalance, all metrics were also weighted and averaged.

$$Sensitivity = \sum_{i=1}^n (Tpi) / \sum_{i=1}^n (Tpi + Fni) \quad (9)$$

$$Specificity = \sum_{i=1}^n (Tni) / \sum_{i=1}^n (Tni + Fpi) \quad (10)$$

$$Accuracy = \sum_{k=0}^n (Tpi + Tni) / (Tpi + Tni + Fpi + Fni) / n \quad (11)$$

$$Precision = \sum_{i=1}^n (Tpi) / \sum_{i=1}^n (Tpi + Fpi) \quad (12)$$

$$Recall = \sum_{i=1}^n (Tpi) / \sum_{i=1}^n (Tpi + Fni) \quad (13)$$

$$F\_Score = \frac{2(\sum_{i=1}^n (Tpi) / \sum_{i=1}^n (Tpi + Fpi) * \sum_{i=1}^n (Tpi) / \sum_{i=1}^n (Tpi + Fni))}{\sum_{i=1}^n (Tpi) / \sum_{i=1}^n (Tpi + Fpi) + \sum_{i=1}^n (Tpi) / \sum_{i=1}^n (Tpi + Fni)} \quad (14)$$

$$end\ to\ end\ delay = \frac{\sum_{i=1}^n (Rpt - Spt) * 1000\ ms}{Tpds} \quad (15)$$

$$Routing\ overhead = \sum_{i=1}^n (Tcpt) / Td \quad (16)$$

$$Throughput = \sum_{i=1}^n (Rp * Ps) / St \quad (17)$$

Packet timers are represented as (Rpt), sent packet timers are represented as (Spt), total packets successfully delivered as (Tpds), total control packets are represented by (Tcpt), total data packets are indicated by (Td), received packets are indicated by (Rp), packet size is indicated by (Ps), and simulation time is indicated by (St).

#### 4. Experimental Analysis

This research was carried out using the Python programming language. Jupyter notebooks were used to combine code, text, and visualisations. Several widely used Python libraries (Scikit-learn, Pandas, Numpy, Sklearn, Matplotlib, Seaborn, TensorFlow, and Keras) were used in the analysis process [35]. Pandas is a data manipulation library that allows the researcher to read, manipulate, and analyse structured data in a tabular format. There is a wide range of algorithms available in Sklearn for supervised and unsupervised learning. Numpy is a numerical computing library that allows the researcher to perform complex mathematical operations on large datasets efficiently. Matplotlib is a plotting library that provides researchers with a range of options for visualising their data, while Seaborn is a data visualisation library that specialises in creating visually appealing and informative statistical graphics. To remove redundant columns, I used a correlation value of 0.9. A K-fold cross-validation procedure was used in this study. As test data, an observation from sample one was used, and training data were generated via leave-one-out cross-validation.

In order to detect intrusions in IoT networks, we developed machine learning models using these libraries. By utilising Python and these libraries, I was able to develop a flexible and powerful analysis framework that can be easily extended and modified to suit my needs.

##### 4.1. Implementation

To implement the proposed intrusion detection system, these steps were followed:

###### 4.1.1. Collect and Preprocess the Dataset

This step involves collecting data from IoT devices and pre-processing them to remove any noise or outliers. The pre-processed data are then split into training and testing sets.

###### 4.1.2. Design the Neural Network Architecture

This step involves selecting the appropriate neural network architecture, such as a feedforward neural network or a convolutional neural network. The number of layers and nodes in each layer is also determined in this step.

###### 4.1.3. Train the Neural Network

In this step, the neural network is trained using the training set. To minimise the error between the predicted and actual output, the weights and biases of the network are adjusted.

###### 4.1.4. Evaluate the Neural Network

The trained neural network is evaluated on the testing set in this step. Its performance can be measured using metrics such as accuracy, precision, recall, and F1-score.

##### 4.2. Import Libraries

To implement machine learning- and blockchain-based intrusion detection in IoT using the NSL-KDD dataset, I imported several libraries into Python. As part of the implementation process, these libraries provided the necessary tools and functionalities. For machine learning, I imported libraries such as Pandas, Scikit-learn, Numpy, and Matplotlib. Pandas was used for data manipulation and preparation, Scikit-learn was used for machine learning algorithms and evaluation, Numpy was used for numerical computations, and Matplotlib was used for visualisation. For blockchain-based intrusion detection, I imported libraries such as web3, Ethereum, and Py-solc. The library web3 allowed me to interact with the Ethereum blockchain, while Ethereum provided me with the necessary tools to deploy smart contracts on the blockchain. Py-solc is a library that allowed me to compile Solidity smart contracts. By combining the machine learning and blockchain libraries, I built a system that utilises machine learning algorithms to detect intrusions in IoT devices and then records these events on a blockchain for secure and tamper-proof storage. This

system can provide enhanced security and privacy for IoT devices as blockchain-based storage ensures that intrusion detection data cannot be altered or deleted.

#### 4.3. Exploring the Dataset

A widely used datasets in network intrusion detection systems are NSL-KDD [27] and UNSW-NB 15 [28]. There are several types of attacks in these datasets, including denial of service (DoS), probe, remote to local (R2L), and user to root (U2R). The neural network implementation for IoT intrusion detection requires preprocessing the data first, which includes converting categorical data into a numerical form, normalising the numerical data, and splitting the data into training and testing sets. For intrusion detection, neural networks were built and trained using Keras and TensorFlow. The compilation and training of the neural network could then be performed using TensorFlow, with the datasets NSL-KDD and UNSW-NB15 (Table 3).

**Table 3.** Dataset: intrusion detection system in IoT.

Dataset	Total	Normal	DoS	Probe	R2L	U2R
KDD Train+	125,973	67,343	45,827	11,456	995	49
KDD Test+	25,192	13,449	9234	2289	209	11
KDD Test-21	22,542	12,709	7749	1867	175	42
UNSW-NB15 Train+	175,341	56,000	12,264	11,450	985	48
UNSW-NB15 Test+	82,332	37,000	4089	2012	201	14

To explore the datasets, several Python libraries were used, such as Pandas, NumPy, and Matplotlib. Pandas can be used to load the dataset into a data frame, which allows for the easy manipulation and analysis of the data. NumPy can be used for numerical computations, and Matplotlib can be used for data visualisation.

The first step in exploring the dataset was to load it into a data frame using Pandas. Then, Pandas functions such as `describe()` and `info()` were used to obtain summary statistics and information about the dataset. This includes the number of samples, number of features, and data types of each feature.

Next, I used Matplotlib to visualise the distribution of each feature in the dataset. This can help identify potential issues such as missing values, outliers, or imbalanced classes. For example, if a feature has a highly skewed distribution, this could indicate that the feature may not be useful for prediction.

After identifying potential issues, various preprocessing techniques can be used, such as data normalisation, feature scaling, and handling missing values to clean and prepare the data for machine learning or blockchain-based models.

#### 4.4. EDA and Data Preprocessing

It is necessary to transform each data characteristic before inputting it into the algorithm. Data preprocessing is a crucial stage in this process. Outlier removal is a technique used to reduce the size of the dataset. It involves replacing outlier values or reducing their impact by modifying outlier weights.

#### 4.5. Potential Challenges in Real-World Deployment

Data play a major role in machine learning. When data are noisy and erratic, they can be extremely challenging to analyse. Underfitting occurs when training data cannot establish a relationship between inputs and outputs accurately. Whenever a machine learning model performs poorly after being trained on a large amount of data, it is said to be overfit. As a result, the algorithm's performance will be negatively affected due to noisy and biased data. Machine learning is a relatively new field that is rapidly evolving. Learning is complicated: there are many opportunities for error since the process is constantly changing. Training the data is the most crucial step in the machine learning process. In the absence

of sufficient training data, predictions will be excessively biased or inaccurate. One of the problems frequently experienced by machine learning experts is slow implementation. Although it takes a long time, machine learning models are very effective at producing correct results. As the amount of data increases, the algorithm may become flawed.

## 5. Results

### 5.1. Model Evaluation

#### 5.1.1. Performance Metrics (Secure Tree-Based IDS INTRUSION Detection System (HIDT))

The proposed HIDT performance was evaluated using Equations (9)–(14), shown in Table 4. In comparison to other MLTs, HIDT is highly accurate in detecting DDoS attacks. HIDT shows a higher detection rate for the KDD99 (NSL-KDD) and UNSW-NB-15 datasets (Table 5). A comparison of the predicted values for sensitivity, specificity, accuracy, precision, detection, and F1 measure metrics is presented in Figure 6. The proposed IDS, based on a hybrid detection tree, performs better than other algorithms in terms of sensitivity, specificity, accuracy, precision, detection, and F1. Classifiers classify attacks correctly when true positives and false positives are determined via their ability to determine the class in which an attack is detected. The classifier correctly dismissed attacks when a true negative was determined. When a false negative was determined, the classifier incorrectly classified attacks as normal traffic.

**Table 4.** Performance parameters for different MLTs and the proposed algorithm (HIDT).

Algorithms	Sensitivity	Specificity	Accuracy	Recall	F1	Precession
SVM	96.2	99.65	99.47	98.19	96.27	97.09
RF	99.3	99.89	99.83	99.87	99.76	99.88
KNN	98.7	99.8	99.65	99.82	98.69	98.75
LR	96.7	99.67	99.46	97.28	96.72	96.97
MLP	99.08	99.13	99.12	99.31	99.46	99.39
HIDT	99.92	99.96	99.95	99.94	99.89	99.92

**Table 5.** Attack detection accuracy using the KDD99 (NSL-KDD) and UNSW-NB-15 datasets for different MLTs.

Algorithms	Sensitivity	Specificity	Accuracy	Recall	F1	Precession
SVM	96.2	99.65	99.47	98.19	96.27	97.09
RF	99.3	99.89	99.83	99.87	99.76	99.88
KNN	98.7	99.8	99.65	99.82	98.69	98.75
LR	96.7	99.67	99.46	97.28	96.72	96.97
MLP	99.08	99.13	99.12	99.31	99.46	99.39
HIDT	99.92	99.96	99.95	99.94	99.89	99.92

In all datasets, the tree-based intrusion detection system (HIDT) was capable of predicting attacks in the shortest amount of time and had the highest attack detection accuracy (99.95% for the KD99 dataset and 99.72% for the UNBS-NB 15 dataset). The binary classification test results were compared to those of earlier studies to ensure the validity of the results [15–19] (Table 6). Several studies have been conducted to evaluate the effectiveness of various machine learning techniques for intrusion detection in IoT systems. In this paper, I compare and analyse the reported accuracies of different machine learning algorithms in the literature for intrusion detection in IoT systems.

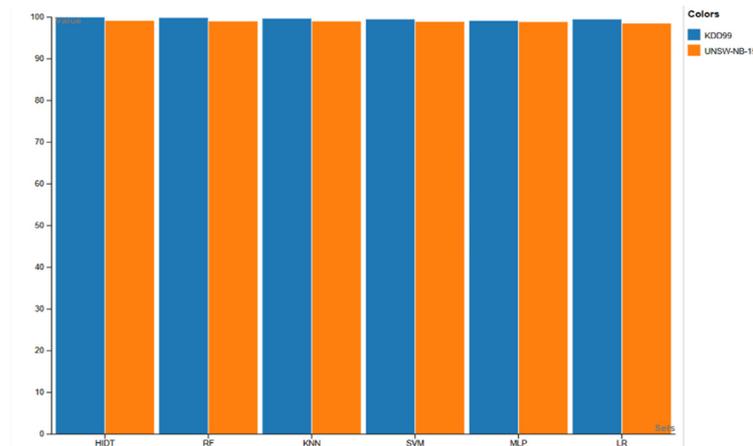


Figure 6. Attack detection accuracy.

Table 6. Proposed IDS accuracy within the existing literature.

Reference	Method	Accuracy (%) (KD99 Dataset)
[15]	DDAML algorithm	91.2
[16]	Optimised gradient boost decision tree	91.82
[17]	Decision tree	93.3
[18]	Gradient boost decision tree using enhanced African buffalo optimisation method	99.81
[19]	MQTT	99.94
Proposed	HIDT	99.95

5.1.2. Confusion Matrix

The confusion matrix by itself is insufficient for creating a suitable visual representation since classes are not equally represented in the data [20]. The confusion matrix obtained for the HIDT algorithm is shown in Figure 7 when all 34 features and eight classes were used. The correct classifications are along the first diagonal, while all other entries are misclassifications. The overall accuracy is shown in the lower suitable cell. In contrast to previous models, the HIDT’s confusion matrix (Figure 7) has low FP and FN rates while having high TP and TN rates. In this instance, the model’s accuracy for the KDD 99(NSL-KDD) dataset is 99.95%, and for the UNBS-NB 15 dataset, it is 99.72%.

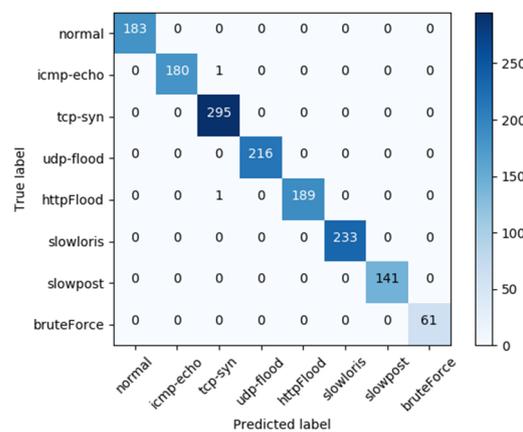
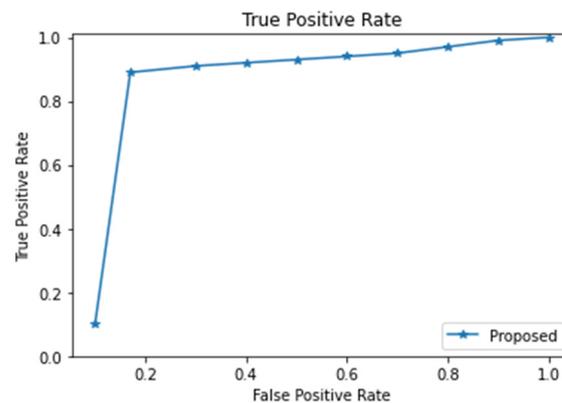


Figure 7. Confusion matrix for HIDT.

### 5.1.3. Receiver Operating Characteristic Curve (ROC)

Any predictive model can be understood using the ROC curve. The false-positive rate is plotted against the true-positive rate in Figure 8. ROC graphs and confusion matrices are additional tools for assessing classifier performance. At various thresholds, the ROC graph compares the true-positive rate (TPR) and false-positive rate (FPR). The point (0, 1) is the best classifier because it correctly classifies good and bad examples. The test results were compared to those of earlier studies to ensure the validity of the results [15].



**Figure 8.** ROC graph for false-positive rate vs. true-positive rate.

### 5.2. Secure Decentralised Blockchain Reputation System (SDBCRS)

A machine-learning-based secure decentralised blockchain reputation system was designed using an Ns3 simulator. Simulations were based on the reputation values and parameters displayed. Using transaction requests, a radius of 800 m was set for communication between the nodes. The simulation area was  $2000 \times 2000$ , and the total number of nodes was 10–100. To generate authentic traffic, I simulated an IoT scenario with multiple Internet of Things sensors; the simulation parameters are shown in Table 7. DoS attackers of the content request flooding variety were created to spread malicious traffic. The simulation was then run, after which I ran a blockchain-based machine learning system. TCP flows were retrieved from the PCAP file using the Ns3 tool after network packets were captured using Wireshark. The decentralised reputation system extracted legitimate user requests from the flooding to create the source response. I labelled the flow using multi-valued classification. The following performance metrics were used to evaluate the proposed secure decentralised blockchain reputation system (SDBCRS):

1. End-to-end delay;
2. Routing overhead;
3. Packet delivery ratio;
4. Scalability;
5. Throughput.

**Table 7.** Simulation parameters.

Parameter	Values
Tool	Ns3 (ns Network simulator)
Simulation area	$2000 \times 2000$
Nodes	10, 20, 30, 40, 50, 100
Simulation time	600 s
Standard protocol	802.11 g
Network	Peer to peer
Tx range	800 m
Packet size	512 bit/s
Channel bandwidth	11 Mbps

The following performance metrics were used to evaluate the reputation system of the proposed HIDT via the SDBCRS system:

### 5.2.1. End-to-End Delay

Using Equation (15), end-to-end delay can be calculated by considering the time between the source of a packet and its destination across a network. An end-to-end delay for reputation transfer is shown in Figure 9 as a function of the number of nodes involved. Both delays decrease as the number of nodes increases, which is one of the best features of the algorithms. The attacker was ranked based on its value and removed from the verification system.

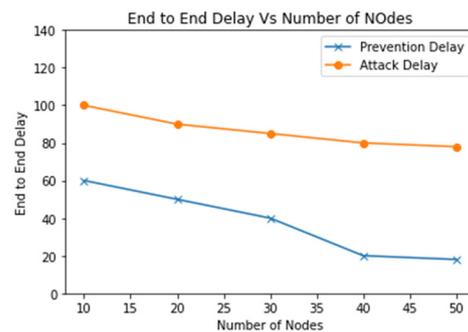


Figure 9. End-to-end prevention and attack delay.

An end-to-end delay of attack and prevention are based on the number of nodes and packets. With fewer packets, there is a lower end-to-end delay, but with more packets, the delay increases [20]. The increase in packets in the network increases traffic, increasing the time it takes for packets to reach their destinations, delaying the attack and prevention from the start to the finish (ms). Compared to other schemes [20], the proposed method showed lower average end-to-end delays. There was a lower end-to-end delay and optimal throughput because many packets were delivered to the destination node in a shorter period. The proposed method had an average end-to-end delay of 60 ms. Compared to the attack delay, this is a shorter delay. As a result, the proposed method reduced the overall average end-to-end delay.

### 5.2.2. Routing Overhead (ROH)

Using Equation (16), the router overhead can be calculated as the ratio between the number of control packets sent and the number of data packets sent. Figure 10 shows the packet delivery ratio versus the routing overhead. The results clearly describe the attacker node response and the effectiveness of the proposed system.

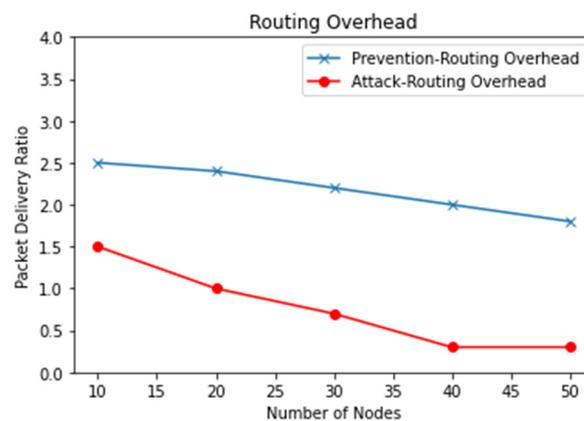


Figure 10. Packet delivery ratio vs. routing overhead.

As shown in Figure 10, the routing overhead increased as the number of nodes increased [9]. The path break rate and packet drop rate increased in a congested network. By dropping packets in the desired route, routing overhead increased. By detecting malicious nodes instantly from the network, the proposed method reduced routing overhead compared to benchmark schemes. According to Figure 11, the average preventing routing overhead was lower than the average attack routing overhead; this resulted in a reduction in the overall routing overhead as a result of the proposed method.

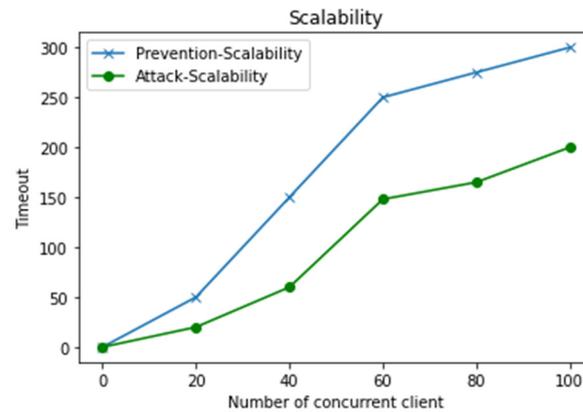


Figure 11. Scalability of the proposed system.

### 5.2.3. Scalability

Figure 11 illustrates the scalability of the proposed system. Preventing and attacking multiple clients increases scalability simultaneously.

Congested networks are more prone to packet drop rates and path breaks. In the desired route, malicious nodes dropped more packets, increasing the timeout. With the proposed method, malicious nodes were detected instantly in the network as opposed to the benchmark scheme, which reduces timeout. Figure 11 shows that most points in the proposed method timeout of prevention scalability were higher than in attack scalability.

### 5.2.4. Throughput

Throughput is computed using the formula provided in Equation (17). Figure 12 shows the throughput of the proposed system. As the number of users increases, the throughput also increases. The throughput indicates the efficiency of a system.

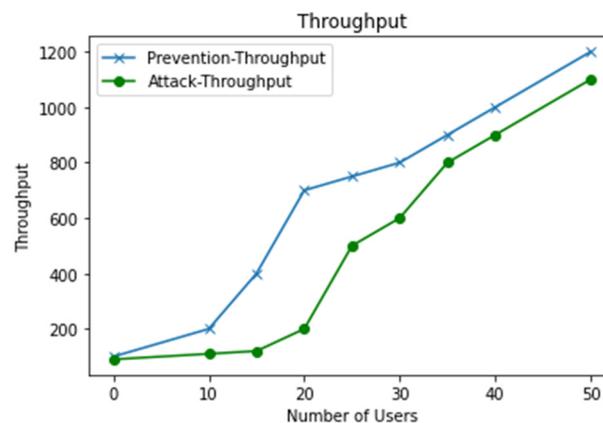


Figure 12. Throughput of the system.

The throughput also degraded due to the malicious nodes in the network, which drastically suffered from the increase in malicious nodes [20]. In the proposed schemes, the highest throughput was recorded at 50 nodes, and the throughput increased as the number of nodes increased. The proposed system performed well in terms of throughput.

### 5.3. Discussion

In the Internet of Things context, a blockchain-based machine learning framework is used for secure data transmission. The performance of the system is evaluated in terms of the successful packet transmission rate, increased throughput, and scalability. In addition, a blockchain is inherently a fault-tolerant system that secures data, while machine learning focuses on using those data to train models and make accurate predictions. The proposed models thus generate more accurate and reliable predictions. Thanks to blockchains and machine learning, reputation transfer systems that require careful handling and high levels of data security can now be automated. Each device stores data requests in the reputation transfer system along with confirmed user credentials and user IP information. After successfully verifying the IoT device, the data request is approved, and the extracted data are retrieved. User IP verification enables end-to-end data security at the transport layer by providing authentication, authorisation, and anonymity. One of the main limitations of this proposed work is the parameters that will affect performance while training the model. As a result of time constraints, Edge devices cannot be distributed differently when the number of users is added and execution time increases. However, I would like to take this task of improving this limitation as a future problem.

The use of ML and BCs technology in IoT security can provide effective solutions to identify and prevent intrusions. However, the security and privacy threats associated with these technologies need to be carefully considered and addressed through a comprehensive and holistic approach that leverages the strengths of different countermeasures and technologies. The adoption of hybrid approaches that combine ML algorithms with blockchain-based data management and access control mechanisms can provide a robust and effective solution to IoT security challenges.

The findings of this research highlight the potential of machine learning algorithms in boosting the security of IoT systems. Nonetheless, choosing a suitable algorithm depends on various factors such as the data type and volume, processing speed, and the resource constraints of the IoT system. Hence, it is crucial to evaluate different machine learning approaches carefully and select the one that fits the IoT system's requirements optimally.

Another challenge for ML-based intrusion detection systems is the possibility of data poisoning attacks in which attackers inject malicious data into the training dataset to manipulate the ML model's behaviour. Data poisoning attacks can compromise the system's integrity and reliability by inducing false positives or negatives. To address these threats, various countermeasures can be implemented, such as data sanitisation and validation, model robustness testing, and model interpretability and transparency. Furthermore, integrating blockchain technology can improve the security and privacy of ML-based intrusion detection systems by providing a decentralised and tamper-proof data management framework.

Blockchain-based intrusion detection systems can also tackle other security and privacy threats in IoT systems, such as device spoofing, data tampering, and unauthorised access. By facilitating secure and transparent data sharing and access control across various IoT devices and networks, blockchain technology can improve the intrusion detection system's resilience and transparency.

The findings of this study may be used to guide the creation of a secure data transmission framework for IoT that employs blockchain technology and machine learning. The suggested system provides a high packet delivery ratio and increased throughput and is scalable. As a result, these results have the potential to greatly improve the security and privacy of IoT devices.

## 6. Conclusions and Future Scope

This research proposes a hybrid decision tree method for automatic anomaly detection that integrates machine learning and artificial intelligence concepts. By using artificial intelligence and machine learning, the proposed system provides secure and useful real-time information for cyberattack detection. The proposed approach outperforms the

existing algorithms in terms of sensitivity, specificity, accuracy, detection, precision, and F1 score. The results were obtained via hyper-parameter tweaking with a grid search and a five-fold CV. The results show the high (99.95%) accuracy of the HIDT in detecting DDoS attacks. Moreover, the multilayer perception performance is generally ideal and very similar to the HIDT. Receiver operating characteristic curve plots and confusion matrices show the overall performance of the classifiers. The system's performance is measured in terms of successful packet transmission rate, improved throughput, scalability, and overall efficiency. Machine learning focuses on using data to train models and make precise predictions, whereas the blockchain is a naturally fault-tolerant system that secures data. Thus, the suggested models predict forecasts that are more precise and trustworthy. The blockchain certification module must confirm the data requester's privacy information, and the certificate is issued by the decentralised blockchain reputation for the transaction. Blockchain-based machine learning systems have meaningful conversations. within addition to end-to-end prevention and attack delay, routing overhead, scalability, and throughput, the proposed system shows its overall performance.

The research presented in this paper opens several avenues for future work in the field of IDS in IoT using machine learning and blockchain technology. Firstly, further investigations can be conducted to enhance the development of machine learning algorithms for intrusion detection and prevention in IoT systems. Although the HIDT algorithm achieved a high level of accuracy, new techniques and approaches can be explored to improve the performance and efficiency of intrusion detection systems. Secondly, it is crucial to evaluate the performance of different machine learning algorithms and blockchain-based solutions for IoT security in real-world scenarios. This requires testing the effectiveness of these solutions under various conditions and scenarios, including different types of IoT devices, network configurations, and data volumes. Moreover, integrating machine learning and blockchain technology with other security mechanisms, such as quantum cryptography and access control, can create a more comprehensive security solution for IoT systems. This integration can be extended to address issues such as secure communication and device management in the IoT ecosystem.

Future research will address various security issues associated with cloud security and explore state-of-the-art advancements. Multi-classification also needs to be explored to investigate the detection rate of various attacks rather than anomalies. Soon, it might be possible to further improve the results via hyperparameter tuning coupled with dimensionality reduction and/or attention mechanism techniques.

Additionally, the ethical and legal implications of using machine learning and blockchain technology for IoT security must be explored. This includes investigating issues such as privacy, bias, and fairness in machine learning algorithms, as well as the regulatory and legal frameworks for blockchain-based solutions. Finally, there is a need for research on developing user-friendly and accessible tools and platforms for implementing machine learning and blockchain-based solutions for IoT security. This involves developing easy-to-use interfaces and tools that can enable non-experts to deploy and manage these solutions effectively. In conclusion, the research in this paper suggests several possibilities for future work in the field of security and privacy in IoT using machine learning and blockchain technology. Developing these solutions will provide a secure and trustworthy IoT ecosystem, helping to address the security challenges faced by IoT systems.

**Funding:** This research was funded by the Deanship of Scientific Research at Majmaah University for supporting this work under Project Number No: R-2023-531.

**Data Availability Statement:** The data are included in the article/referenced in the article.

**Acknowledgments:** The author would like to thank Deanship of Scientific Research at Majmaah University for supporting this work under Project Number No. R-2023-531.

**Conflicts of Interest:** The author declares no conflict of interest.

## References

1. Dwivedi, Y.K.; Hughes, L.; Baabdullah, A.M.; Ribeiro-Navarrete, S.; Giannakis, M.; Al-Debei, M.M.; Dennehy, D.; Metri, B.; Buhalis, D.; Cheung, C.M.; et al. Metaverse beyond the hype: Multidisciplinary perspectives on emerging challenges, opportunities, and agenda for research, practice and policy. *Int. J. Inf. Manag.* **2022**, *66*, 102542. [CrossRef]
2. Khan, A.A.; Laghari, A.A.; Li, P.; Dootio, M.A.; Karim, S. The collaborative role of blockchain, artificial intelligence, and industrial internet of things in digitalization of small and medium-size enterprises. *Sci. Rep.* **2023**, *13*, 1656. [CrossRef]
3. Othman, S.B.; Almalki, F.A.; Chakraborty, C.; Sakli, H. Privacy-preserving aware data aggregation for IoT-based healthcare with green computing technologies. *Comput. Electr. Eng.* **2022**, *101*, 108025. [CrossRef]
4. Orumwense, E.F.; Abo-Al-Ez, K. Internet of Things for smart energy systems: A review on its applications, challenges and future trends. *AIMS Electron. Electr. Eng.* **2023**, *7*, 50–74. [CrossRef]
5. Paul, B.; Rao, M. Zero-Trust Model for Smart Manufacturing Industry. *Appl. Sci.* **2022**, *13*, 221. [CrossRef]
6. Ahmad, T.; Zhu, H.; Zhang, D.; Tariq, R.; Bassam, A.; Ullah, F.; AlGhamdi, A.S.; Alshamrani, S.S. Energetics Systems and artificial intelligence: Applications of industry 4.0. *Energy Rep.* **2022**, *8*, 334–361. [CrossRef]
7. Dawadi, B.R.; Adhikari, B.; Srivastava, D.K. Deep Learning Technique-Enabled Web Application Firewall for the Detection of Web Attacks. *Sensors* **2023**, *23*, 2073. [CrossRef]
8. Mumtaz, G.; Akram, S.; Iqbal, W.; Ashraf, M.U.; Almarhabi, K.A.; Alghamdi, A.M.; Bahaddad, A.A. Classification and Prediction of Significant Cyber Incidents (SCI) using Data Mining and Machine Learning (DM-ML). *IEEE Access* **2023**. [CrossRef]
9. Yin, H.L.; Fu, Y.; Li, C.L.; Weng, C.X.; Li, B.H.; Gu, J.; Lu, Y.S.; Huang, S.; Chen, Z.B. Experimental quantum secure network with digital signatures and encryption. *Natl. Sci. Rev.* **2023**, *10*, nwac228. [CrossRef]
10. Niranjnamurthy, M.; Nithya, B.N.; Jagannatha, S.J.C.C. Analysis of Blockchain technology: Pros, cons and SWOT. *Clust. Comput.* **2019**, *22*, 14743–14757. [CrossRef]
11. Ferrag, M.A.; Maglaras, L.; Benbouzid, M. Blockchain and Artificial Intelligence as Enablers of Cyber Security in the Era of IoT and IIoT Applications. *J. Sens. Actuator Netw.* **2023**, *12*, 40. [CrossRef]
12. Ma, Z.; Zhang, J.; Guo, Y.; Liu, Y.; Liu, X.; He, W. An efficient decentralized key management mechanism for VANET with blockchain. *IEEE Trans. Veh. Technol.* **2020**, *69*, 5836–5849. [CrossRef]
13. Banafaa, M.; Shayea, I.; Din, J.; Azmi, M.H.; Alashbi, A.; Daradkeh, Y.I.; Alhammad, A. 6G mobile communication technology: Requirements, targets, applications, challenges, advantages, and opportunities. *Alex. Eng. J.* **2022**, *64*, 245–274. [CrossRef]
14. Jiang, T.; Shen, G.; Guo, C.; Cui, Y.; Xie, B. BFLS: Blockchain and Federated Learning for sharing threat detection models as Cyber Threat Intelligence. *Comput. Netw.* **2023**, *224*, 109604. [CrossRef]
15. Dong, S.; Sarem, M. DDoS attack detection method based on improved KNN with the degree of DDoS attack in software-defined networks. *IEEE Access* **2019**, *8*, 5039–5048. [CrossRef]
16. Tama, B.A.; Rhee, K.H. An in-depth experimental study of anomaly detection using gradient boosted machine. *Neural Comput. Appl.* **2019**, *31*, 955–965. [CrossRef]
17. Tuan, T.A.; Long, H.V.; Son, L.H.; Kumar, R.; Priyadarshini, I.; Son, N.T.K. Performance evaluation of Botnet DDoS attack detection using machine learning. *Evol. Intell.* **2020**, *13*, 283–294. [CrossRef]
18. Mishra, S. An optimized gradient boost decision tree using enhanced African buffalo optimization method for cyber security intrusion detection. *Appl. Sci.* **2022**, *12*, 12591. [CrossRef]
19. Mishra, S.; Albarakati, A.; Sharma, S.K. Cyber Threat Intelligence for IoT Using Machine Learning. *Processes* **2022**, *10*, 2673. [CrossRef]
20. Malik, A.; Khan, M.Z.; Faisal, M.; Khan, F.; Seo, J.T. An efficient dynamic solution for the detection and prevention of black hole attack in vanets. *Sensors* **2022**, *22*, 1897. [CrossRef]
21. Radanliev, P.; De Roure, D.; Page, K.; Van Kleek, M.; Santos, O.; Maddox, L.T.; Burnap, P.; Anthi, E.; Maple, C. Design of a dynamic and self-adapting system, supported with artificial intelligence, machine learning and real-time intelligence for predictive cyber risk analytics in extreme environments—cyber risk in the colonisation of Mars. *Saf. Extrem. Environ.* **2020**, *2*, 219–230. [CrossRef]
22. Sarker, I.H. Machine learning for intelligent data analysis and automation in cybersecurity: Current and future prospects. *Ann. Data Sci.* **2022**, *1*–26. [CrossRef]
23. Zhang, Z.; Ning, H.; Shi, F.; Farha, F.; Xu, Y.; Xu, J.; Zhang, F.; Choo, K.K.R. Artificial intelligence in cyber security: Research advances, challenges, and opportunities. *Artif. Intell. Rev.* **2022**, *55*, 1029–1053. [CrossRef]
24. Shirafkan, M.; Shahidienjad, A.; Ghobaei-Arani, M. An autonomous intrusion detection system for the RPL protocol. *Peer-Peer Netw. Appl.* **2022**, *15*, 484–502. [CrossRef]
25. Mohanta, B.K.; Jena, D.; Satapathy, U.; Patnaik, S. Survey on IoT security: Challenges and solution using machine learning, artificial intelligence and blockchain technology. *Internet Things* **2020**, *11*, 100227. [CrossRef]
26. Derhab, A.; Guerroumi, M.; Gumaiei, A.; Maglaras, L.; Ferrag, M.A.; Mukherjee, M.; Khan, F.A. Blockchain and random subspace learning-based IDS for SDN-enabled industrial IoT security. *Sensors* **2019**, *19*, 3119. [CrossRef]
27. KDD99 (NSL-KDD) Dataset, Intrusion Detection Dataset. Available online: <https://www.kaggle.com/datasets/hassan06/nslkdd> (accessed on 10 March 2023).
28. UNSW-NB 15 Dataset Was Created by Cyber Range Lab of the Australian Centre for Cyber Security. Available online: <https://www.kaggle.com/datasets/mrwellsdavid/unsw-nb15> (accessed on 10 March 2023).

29. Kaddoura, S.; Haraty, R.; Al Kontar, K.; Alfandi, O. A parallelized database damage assessment approach after cyberattack for healthcare systems. *Future Internet* **2021**, *13*, 90. [[CrossRef](#)]
30. Gu, J.; Cao, X.Y.; Fu, Y.; He, Z.W.; Yin, Z.J.; Yin, H.L.; Chen, Z.B. Experimental measurement-device-independent type quantum key distribution with flawed and correlated sources. *Sci. Bull.* **2022**, *67*, 2167–2175. [[CrossRef](#)]
31. Bellini, E.; Iraqi, Y.; Damiani, E. Blockchain-based distributed trust and reputation management systems: A survey. *IEEE Access* **2020**, *8*, 21127–21151. [[CrossRef](#)]
32. Almadani, M.S.; Alotaibi, S.; Alsobhi, H.; Hussain, O.K.; Hussain, F.K. Blockchain-based multi-factor authentication: A systematic literature review. *Internet Things* **2023**, *23*, 100844. [[CrossRef](#)]
33. Khan, M.N.R.; Ara, J.; Yesmin, S.; Abedin, M.Z. Machine learning approaches in cybersecurity. In *Data Intelligence and Cognitive Informatics: Proceedings of ICDICI 2021*; Springer Nature: Singapore, 2022; pp. 345–357.
34. Ghorri, K.M.; Imran, M.; Nawaz, A.; Abbasi, R.A.; Ullah, A.; Szathmary, L. Performance analysis of machine learning classifiers for non-technical loss detection. *J. Ambient. Intell. Humaniz. Comput.* **2020**, 1–16. [[CrossRef](#)]
35. Susilo, B.; Sari, R.F. Intrusion detection in IoT networks using deep learning algorithm. *Information* **2020**, *11*, 279. [[CrossRef](#)]

**Disclaimer/Publisher’s Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.