JOURNAL OF BASIC SCIENCE AND ENGINEERING

AN ENSEMBLE COGNITIVE LEARNING-BASED INTRUSION DETECTION SYSTEM FOR SECURE IOT ENVIRONMENTS

Kushal Kumar B.N.[1], Dr. Balakrishna R.[2] *, Dr. M.V. Panduranga Rao[3]

[1] Research Scholar, Dept. of Computer Science & Engineering, IoT,
Cyber Security & Blockchain Technology, K.S. Institute of Technology Bengaluru, Karnataka
[2] Principal & Professor, Department of Computer Science & Engineering
Rajarajeswari College of Engineering, Bangalore, Karnataka
[3] Professor, Dept. of Computer Science & Engineering,
Jain University, (Deemed to be University) Bengaluru, India

DOI: https://doie.org/10.0113/Jbse.2025967844

Article History :	Received: July 2024	Revised: Sept 2024	Accepted: December 2024

***Abstract* : -** The Internet of Things (IoT) has significantly transformed various sectors, including smart homes, wearable technology, network applications, and autonomous systems. Despite its advantages, IoT networks face substantial challenges in detecting abnormal traffic patterns, which are crucial for safeguarding network infrastructure from cyber threats. Intrusions that exploit system vulnerabilities can enable attackers to use malicious traffic to gain unauthorized access. Attacks like Distributed Denial of Service (DDoS), Denial of Service (DoS), and Service Scans highlight the necessity of an automated system that can quickly identify anomalies and minimize damage. Although many automated techniques for abnormal traffic detection exist, current Intrusion Detection Systems (IDS) require improvements in efficiency, scalability, and flexibility to handle diverse IoT network environments effectively. This research focuses on addressing these gaps by introducing an Ensemble Cognitive Learning-Based Intrusion Detection System. The proposed system utilizes the Edge-IoT dataset as a benchmark to evaluate machine learning models for detecting network anomalies. A cognitive engine incorporating a meta-classifier dynamically selects the best-performing model for the ensemble classifier, ensuring precise detection of attacks. Experimental findings indicate that the proposed model surpasses existing methods, offering superior multi-class classification accuracy and robust detection capabilities.

***Keywords* :** Internet of Things Intrusion detection, Ensemble learning, Sustainability

1. Introduction

The Internet of Things (IoT) encompasses a vast array of diverse devices, such as sensors and actuators, all interconnected through the Internet. The adoption of IoT technologies across various industries is witnessing significant and rapid growth. Figure 1 highlights the global increase in IoT connections, presenting data from 2022 and 2023 [1], along with projections extending to 2033. The IoT market is also expanding notably in sectors such as consumer electronics, smart grids, and asset management. By 2033, it is estimated that consumer applications will account for approximately 61% of all IoT connections, with major contributions from smart home systems and connected vehicles. In recent years, smart objects, IoT devices, and Wireless Sensor Networks have increasingly incorporated Ambient Intelligence, Research conducted over the past decade highlights a significant rise in the deployment of smart IoT infrastructures, integrating Intelligence to enhance business intelligence and support informed decision-making processes.

The IoT has created a new ecosystem of connected devices that serve as the backbone of smart cities, representing a major shift from the centralized system designs of the past [2]. It plays a crucial role in modern smart city infrastructure. This network of devices enables the exchange of large volumes of sensitive and private data through embedded systems and wireless communication technologies. While IoT offers numerous advantages to individuals, businesses, and service providers, it also brings significant security concerns that affect the reliability of the system. Unlike conventional network systems, IoT relies on embedded systems with communication protocols that vary depending on the specific device and application. The absence of a unified and centralized

security framework makes addressing these vulnerabilities more challenging. As IoT networks handle increasingly large amounts of data, the associated security risks have risen to new and critical levels [3].

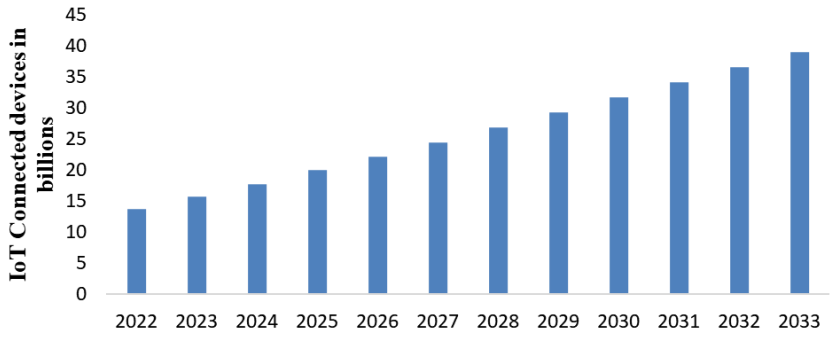

Figure 1. The number of Internet of Things (IoT) connections globally from 2022 to 2023, along with projections from 2024 to 2033

2. Related work

Bibliometric analysis serves as an effective tool for organizing and analysing extensive collections of scientific literature. It employs systematic methodologies similar to those used in structured literature reviews, ensuring the reliability and accuracy of both the information utilized and the outcomes produced. In this study, VOS-viewer software is employed to map and visualize connections between bibliometric data sources, identify leading authors, and extract valuable insights from publications [4]. The VOS-viewer tool used for creating and visualizing bibliometric networks, the following figures shows the network visualization involving organizations, countries and citation sources for the keyword “Ensemble Intrusion Detection in IoT environment”. The figure 2 shows the Network Visualization of Citation Organizations for the given keyword

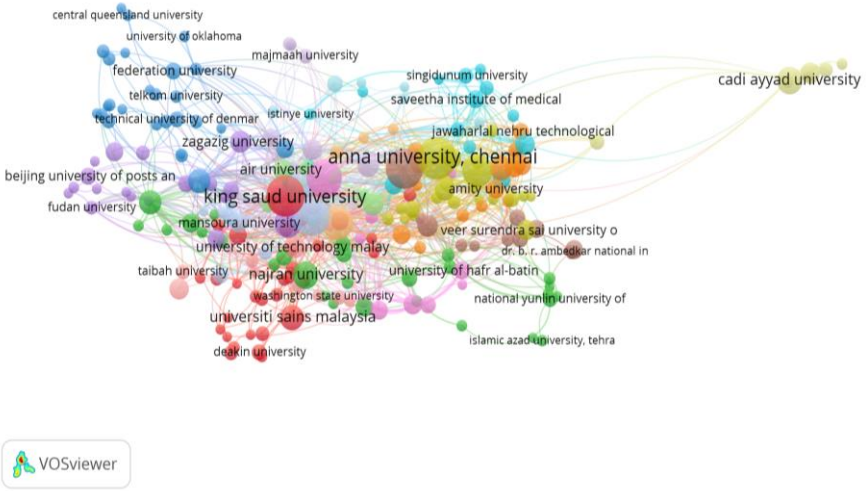

Figure 2. Network Visualization of Citation Organizations

The figure 3 shows the Network Visualization Citation Sources for the given keyword where the sensors, Scientific reports, Computational intelligence and IEEE access are the major publications. The figure 4 shows the Network Visualization of Citation Countries where India, UK Malaysia are the major contributors.

Vol. 21, No. 1, (2024)
ISSN: 1005-0930

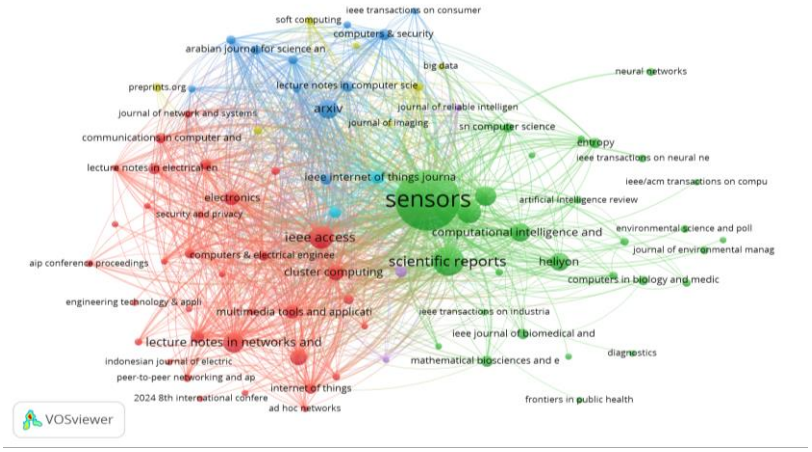

Figure 3. Network Visualization Citation Sources

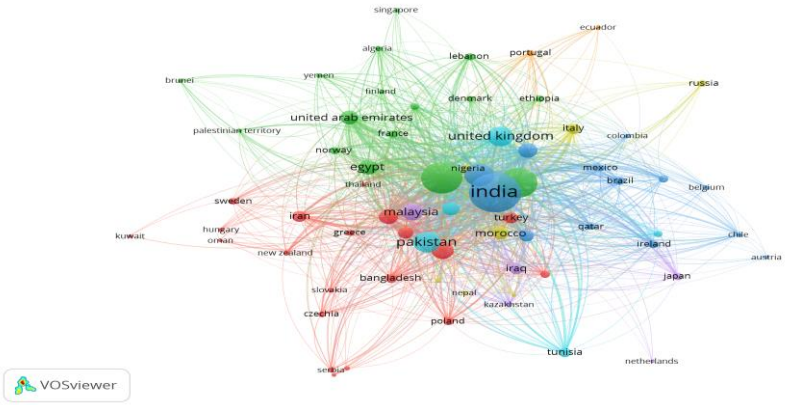

Figure 4. Network Visualization of Citation Countries

In the context of IoT security, reconnaissance attacks involve unauthorized scanning and probing of devices to gather information for potential exploitation. Detecting such activities is crucial to prevent subsequent attacks, including malware infections and botnet formations. Traditional rule-based intrusion detection systems (IDS), like SNORT, have been employed to detect port scanning activities. However, these systems often struggle to adapt to the dynamic nature of modern attacks and may not provide the required accuracy due to their broad focus on various attack types. [5]

Machine learning approaches have been increasingly applied to enhance detection capabilities. For instance, Alani et al. introduced a detection system based on deep neural networks aimed at identifying IoT device scanning and reconnaissance attacks. This system demonstrated the potential of deep learning in improving detection accuracy [6]. The integration of explainable artificial intelligence in security systems addresses the need for transparency in machine learning models. XAI techniques help in understanding model decisions, thereby increasing trust and facilitating the identification of potential weaknesses. This is particularly important in security applications, where understanding the rationale behind a detection can be as critical as the detection itself.

Ensemble machine learning has emerged as a highly effective approach for accurately identifying and preventing malicious activities and cyber-attacks. In recent years, security detection systems powered by artificial intelligence have gained prominence as advanced and reliable solutions. AI-driven techniques, supported by extensive data analysis, enable more precise and thorough detection of anomalies and suspicious behavior, enhancing the overall security and resilience of systems [7]. Aburomman et al. [8] developed an innovative method for constructing intrusion detection systems. Their approach employed Particle Swarm Optimization (PSO) to generate weights that improved the integration of classifiers. By combining six support vector machine (SVM) classifiers with six k-nearest neighbor (KNN) classifiers, they showed that their model outperformed traditional methods in terms of accuracy.

Kuncheva [9] introduced a probability-based framework for combining classifiers, focusing on four ensemble learning strategies: majority voting, weighted majority voting, recall combination, and naive Bayesian combination. These approaches relied on the assumptions of class-conditional independence and the accuracy of individual classifiers.

The study presented by the authors [10] introduces an ensemble voting classifier designed to improve intrusion detection accuracy in IoT networks. This methodology integrates multiple traditional machine learning models, leveraging a voting mechanism to consolidate predictions and produce a final decision. The system was evaluated using data from seven distinct IoT devices, covering both binary and multi-class attack detection scenarios. The proposed model demonstrated impressive performance, achieving an accuracy of 96% for GPS sensors and 97% for weather sensors. Results indicated that the ensemble approach surpassed the effectiveness of standalone machine learning classifiers.

A study [11] emphasized that authentication plays a crucial role in ensuring robust security measures. Secure access relies on an authorized key, which is essential for verifying identity and granting permissions. Strengthening the authentication process can further enhance the security of invisible watermarking data, safeguarding sensitive information. In [12] emphasizes the application of ensemble machine learning, neural networks, and kernel-based techniques for identifying anomalies and detecting malicious activities within IoT systems. Among these methods, ensemble machine learning consistently demonstrated superior accuracy and detection performance.

In [13], a novel integrated framework was designed to improve classification effectiveness. This method involved training multiple classifiers using data sampling techniques aimed at maximizing accuracy across various classes. The optimal combination of classifiers was identified through a genetic algorithm paired with a divide-and-conquer approach. The final decision-making process, based on the majority voting method, resulted in enhanced training efficiency and accuracy. Attou et al. [14] developed a hybrid method that integrated graphic visualization with the random forest (RF) algorithm for intrusion detection in cloud security. By working with a minimized feature set of only two attributes, RF outperformed deep neural networks (DNN), decision trees (DT), and support vector machines (SVM) in detecting and categorizing attack types. Nevertheless, the recall rates achieved with the NSL-KDD dataset were not fully satisfactory.

The IoT ecosystem connects a diverse range of devices, incorporating heterogeneous components from various domains. However, the absence of standardized protocols and unified regulations across IoT networks creates significant security challenges. Traditional security protocols, although effective in mitigating internet-based threats, are not entirely sufficient for the dynamic and distributed nature of IoT environments. As a result, developing an Intrusion Detection System powered by advanced Deep Learning techniques offers a promising approach to enhance IoT security.

3. Dataset

3.1 Edge IIoT Dataset

The **Edge IIoT Dataset** is specifically designed to support research and development in securing Industrial Internet of Things (IIoT) environments, particularly in edge computing architectures. Edge computing refers to processing data closer to its source, reducing latency and improving real-time decision-making in IIoT systems. However, this architecture also introduces unique cybersecurity challenges, including increased exposure to cyber threats due to distributed and resource-constrained devices [15]. In the dataset More accurately, they have identified fourteen attacks as showed in figure 5, which are categorized into five major threats.

3.2 Key Features of the Dataset

- **Diverse Attack Types:** Includes a variety of cyber-attacks such as Denial of Service (DoS), Man-in-the-Middle (MITM), and data injection attacks.
- **High Dimensionality:** Contains numerous features capturing both network and device-level metrics.
- **Real-Time Traffic Simulation:** Mimics real-world IIoT network environments and device interactions.

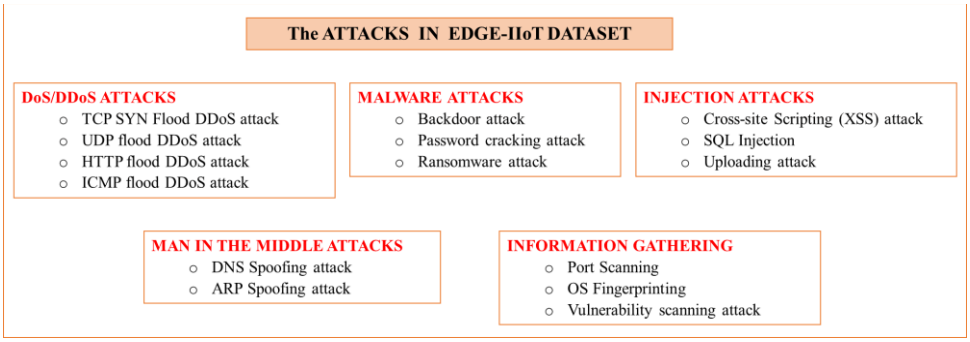

Figure 5. The list of attacks included in Edge-IIoT dataset.

Table 1. Statistics of Edge-IIoT Dataset

Normal	**1,12,23,940**
DDoS-UDP	32,01,626
DDoS-ICMP	29,14,354
SQL-injection	51,203
DDoS-TCP	20,20,120
Vulnerability scanner	1,45,869
Password	10,52,385
DDoS-HTTP	2,29,022
Uploading	37,633
Backdoor	24,862
Port-scanning	22,564
XSS	15,915
Ransomware	10,925
Fingerprinting	1,001
MITM	1,229
Total count	2,09,52,648

The Edge IIoT dataset serves as a foundation for advancing IIoT security, allowing researchers to develop and benchmark anomaly detection and classification models. Explore dimensionality reduction techniques for handling large feature sets and test the resilience of IDS under varying levels of cyber threats. The table 1 gives the statistics of the Edge IIoT dataset comprising of both normal and attack classes. The Edge-IoT dataset is considered an ideal resource for evaluating the performance of machine learning models. It offers a comprehensive and realistic cybersecurity dataset specifically created to support the training of machine learning-based intrusion detection systems [16].

4. Methodology

The proposed methodology for a Secure IoT framework employs an Ensemble Cognitive Learning Model to improve intrusion detection in IoT ecosystems. This method is tailored to effectively adapt to emerging threats while achieving high accuracy in identifying various cyber-attacks. Figure 6 depicts the cognitive ensemble learning model-based intrusion detection system (IDS) designed for IoT networks. The process begins with pre-processing the Edge-IIoT dataset, which comprises both normal and attack traffic from IoT environments. In this stage, data cleaning is carried out to eliminate missing or corrupted entries, ensuring the dataset is suitable for analysis.

To address the dataset's imbalance, the Synthetic Minority Oversampling Technique (SMOTE) is applied. In cases where one class significantly outweighs another, models tend to favor the majority class. SMOTE alleviates this by creating synthetic examples for the minority class, using interpolation between existing instances rather than mere duplication. This approach balances the dataset, helping the model generalize more effectively and improving its accuracy in recognizing minority class instances.

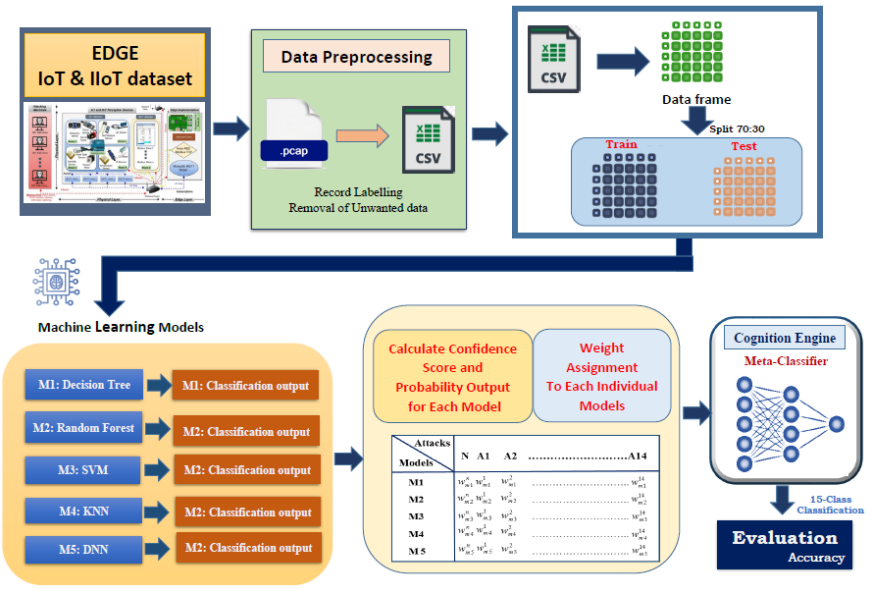

Figure 6. Cognitive ensemble learning model based IDS

The study considers several machine learning models, each tailored for specific classification and regression tasks:

Support Vector Machine (SVM): It is a supervised algorithm primarily used for classification. It determines the best hyperplane to separate data into distinct categories by maximizing the margin between data points from different classes.

Decision Tree (DT): It is a hierarchical model used for classification and regression. It partitions the dataset into smaller groups based on feature values, creating a tree-like structure where internal nodes represent feature-based tests, branches indicate test outcomes, and leaf nodes signify predicted classes or values.

Random Forest (RF): It is an ensemble technique that combines multiple decision trees. For classification, it predicts the class based on the majority vote from all trees, while for regression, it computes the average of predictions from the trees. Random Forest leverages random sampling and feature selection to improve accuracy and reduce the risk of overfitting.

K-Nearest Neighbors (KNN): It is a distance-based algorithm that assigns a class to a data point by analyzing the majority class of its K closest neighbors. It uses distance metrics, such as Euclidean or Manhattan distance, to identify the nearest neighbors and classify the data accordingly.

Deep Neural Networks (DNN): It consist of multiple hidden layers positioned between the input and output layers. Each layer's neurons process input data using learned weights and activation functions to identify patterns. DNNs are particularly effective for analyzing large datasets and uncovering complex relationships within the data.

The multi-class regression model has been constructed for various attack types, distinct weights can be assigned to individual classifiers based on their sensitivity to specific attack types. The weight allocation depends largely on how effectively each classifier distinguishes between different types of attacks. For instance, if classifier M_i demonstrates higher accuracy in identifying normal traffic, it indicates that M_i is more sensitive to normal traffic patterns. Consequently, this classifier is given a higher weight in the ensemble model for normal traffic, contributing to improved overall model accuracy. Conversely, if M_i exhibits lower accuracy, its weight is reduced accordingly. This approach also enhances the precision of abnormal class regression models, leading to more accurate detection and classification of different attack types. The weight assigned to each individual classifier is used to assess the posterior probability of various attack categories to determine the final outcome. While this method improves detection accuracy to some degree, it becomes evident that different classifiers excel at detecting specific types of attacks. For example, a particular classifier might achieve better results with attack type A1 but perform less effectively with attack type A2. As a result, applying uniform decision weights across all attack types for a single classifier may not deliver the best performance.

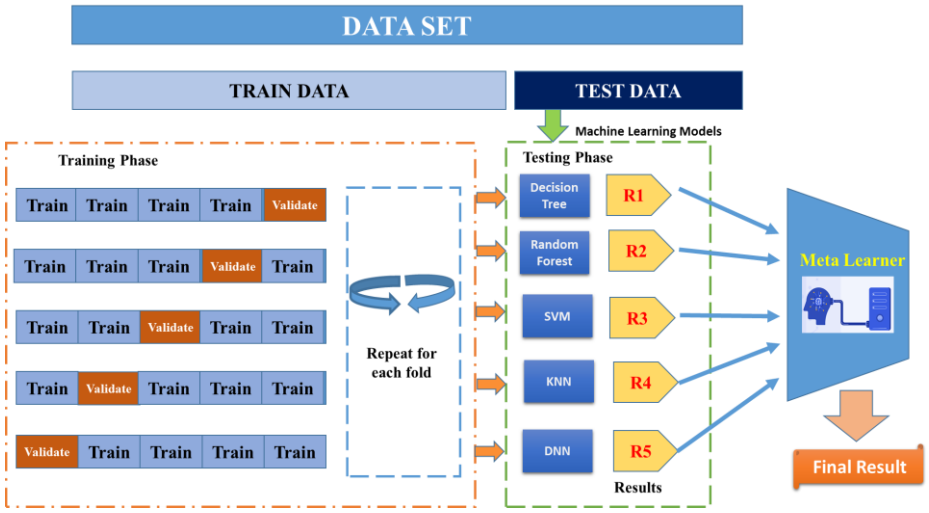

Figure 7. The elements of the proposed ensemble detection model

In dynamic and complex network environments, attack methods continuously evolve over time, requiring detection models to be regularly updated to effectively respond to emerging threats. Traditional machine learning-based intrusion detection models often rely solely on dataset updates for retraining, neglecting valuable insights from historical data. This oversight can result in reduced detection accuracy when the updated model encounters patterns present in historical data. Training a detection model using only a single dataset D may lead to inconsistencies in detection accuracy across different datasets, undermining model stability. Furthermore, the lack of integration between updated models and historical knowledge results in inefficient use of prior insights and hampers the sustainability of model updates.

5. Results

This section outlines the experimental findings of the proposed Ensemble Cognitive Learning-Based Intrusion Detection system, analyzed through standard evaluation metrics. The study incorporates various machine learning techniques, such as Decision Tree, Random Forest, Support Vector Machine, K-Nearest Neighbor, and Deep Neural Network, in addition to the proposed Ensemble Cognitive Learning method. The models' effectiveness in identifying cyber-attacks was tested using the Edge IIoT dataset. The models were tested and evaluated using the test dataset with the following detection metrics:

Accuracy: It Refers to the proportion of correctly classified instances out of the total number of samples, determined by

$$Acc = \frac{TP_{Attack} + TN_{Normal}}{TP_{Attack} + TN_{Normal} + FN_{Attack} + FP_{Normal}} \quad (1)$$

Precision: It Represents the ratio of correctly identified attack instances to the total number of predicted attack cases, calculated as:

$$Prec = \frac{TP_{Attack}}{TP_{Attack} + FP_{Normal}} \quad (2)$$

Recall: It Denotes the percentage of correctly detected attack instances relative to the total actual attack cases that should have been identified, given by:

$$Rec = \frac{TP_{Attack}}{TP_{Attack} + FN_{Attack}} \quad (3)$$

Figure 8 illustrates the Precision and Recall values for the Normal class in a 15-class classification scenario using various machine learning models. The results demonstrate that all the machine learning models, including the proposed model, achieved a Precision and Recall value of 1 for the Normal class of traffic.

Vol. 21, No. 1, (2024)
ISSN: 1005-0930

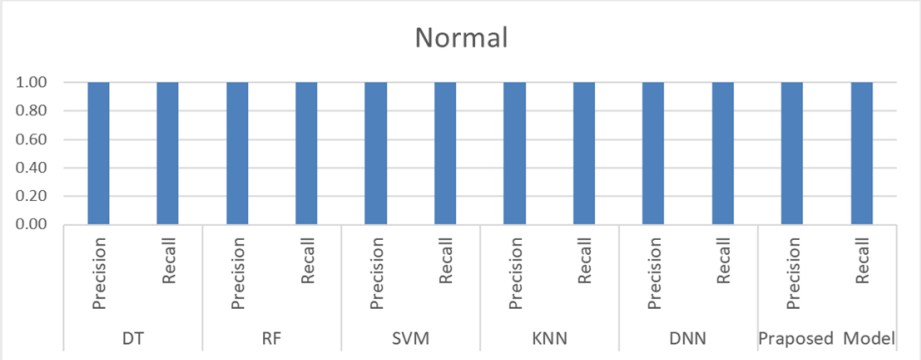

Figure. 8 Precision and Recall for Normal Class Using Machine Learning Models

The Figures 9 through 12 present the Precision and Recall values for different attack categories in the 15-class classification using machine learning models, including the proposed model. Specifically, Figure 9 illustrates results for DoS/DDoS attacks, Figure 10 for Injection attacks, Figure 11 for Information Gathering attacks, and Figure 12 for Man-in-the-Middle attacks.

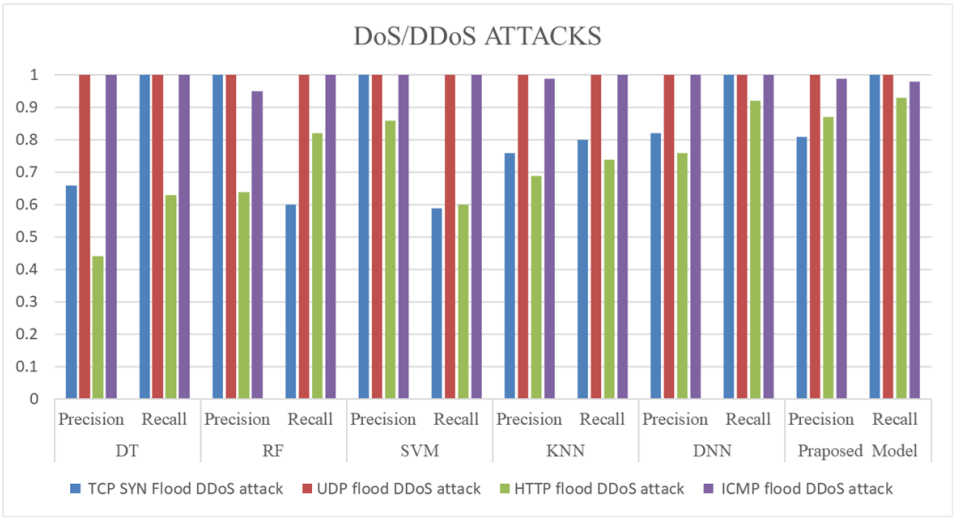

Figure 9 Precision and Recall for DoS/DDoS Attacks in 15-Class Classification

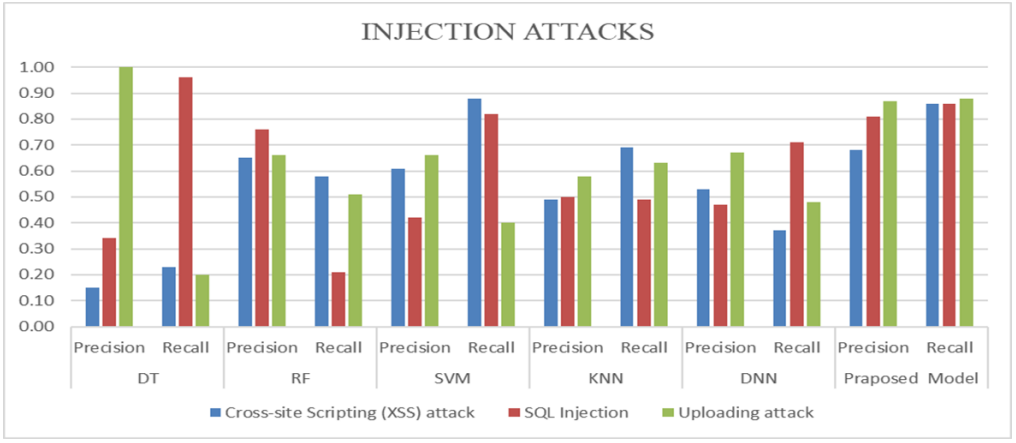

Figure 10 Precision and Recall for Injection Attacks in 15-Class Classification

Vol. 21, No. 1, (2024)
ISSN: 1005-0930

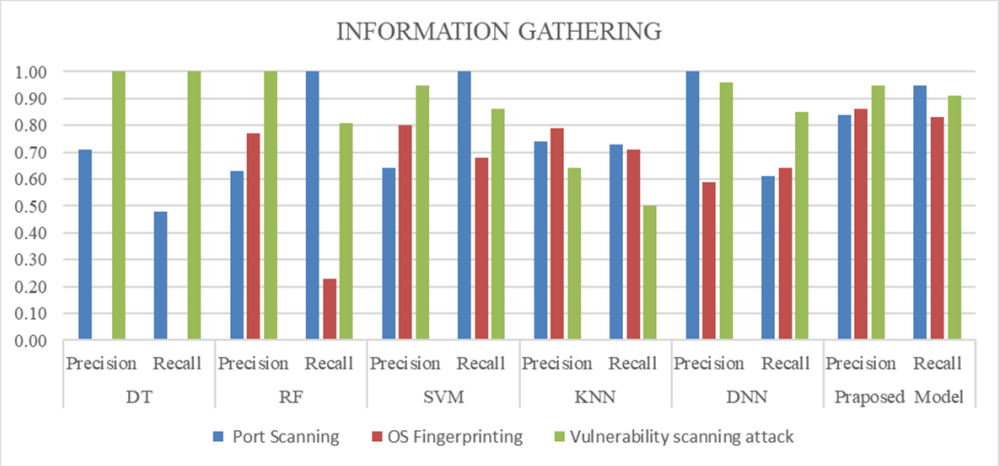

Figure 11 Precision and Recall for Information Gathering Attacks in 15-Class Classification

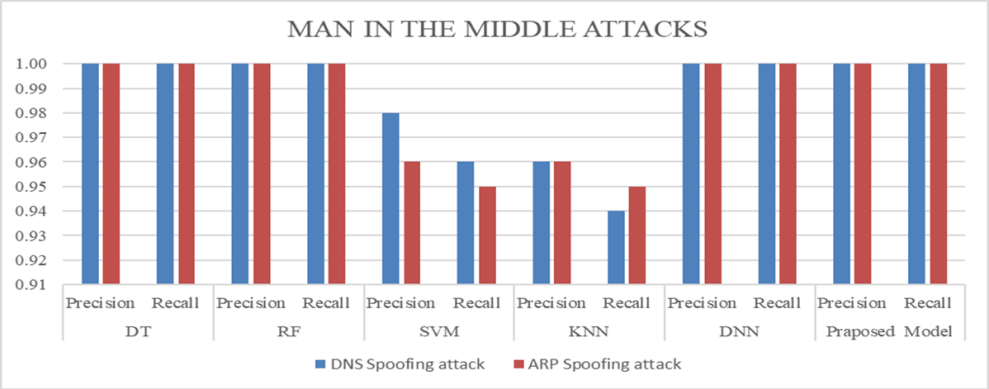

Figure 12 Precision and Recall for Man in the Middle Attacks in 15-Class Classification

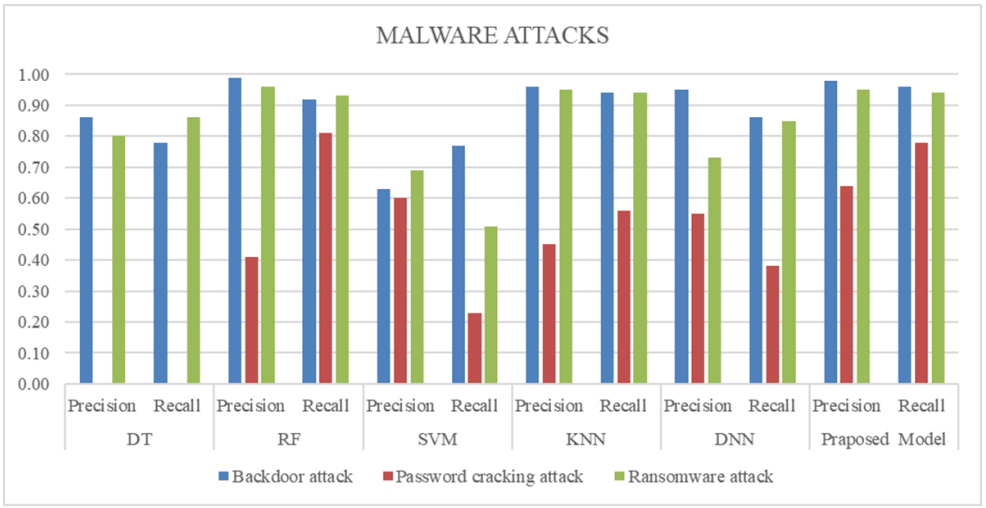

Figure 13 Precision and Recall for Malware Attacks in 15-Class Classification Using Machine Learning Models

In the 15-class classification problem, the DNN classifier achieved the accuracy of 94.67%, while the DT classifier recorded the lowest accuracy at 68.12%. The RF classifier attained an accuracy of 80.13%, followed by the SVM classifier with 76.91%, and the KNN classifier with 80.01%. Notably, the proposed cognitive ensemble learning model surpassed all these classifiers, achieving an accuracy of 96.62%. as shown in the figure 14.

Vol. 21, No. 1, (2024)
ISSN: 1005-0930

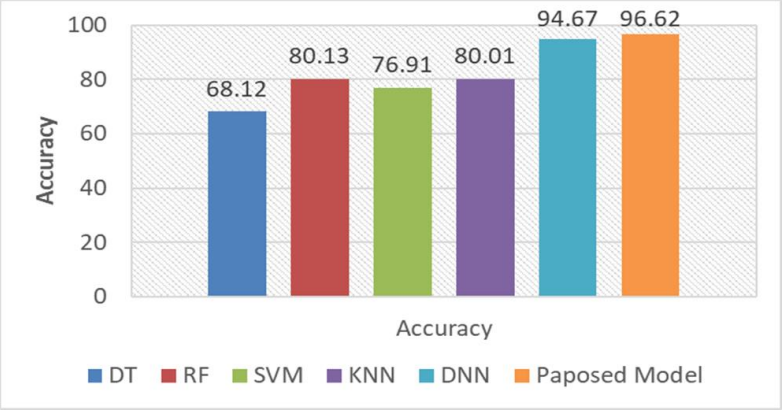

Figure 14 Accuracy of the Machine Learning Models 15 class classification

The Edge-IIoT dataset was employed to assess the effectiveness of the proposed ensemble machine learning based intrusion detection systems in a multiclass classification scenario involving 15 unique classes. This dataset served as a robust benchmark for evaluating the capability of different models to accurately detect and classify various types of cyber-attacks. The results revealed that the cognitive ensemble machine learning model outperformed the other models in terms of accuracy and reliability, showcasing its ability to manage the challenges associated with multiclass intrusion detection more efficiently.

7. Conclusion

An ensemble learning-based model designed in this paper to address the limitations of individual machine learning classifiers. The weaknesses of one classifier are mitigated by leveraging the strengths of another. The Edge IIoT Dataset is utilized to assess the system's performance, and the SMOTE technique is applied to resolve class imbalance issues. The ensemble model incorporates multi-class regression techniques to dynamically allocate decision weights to individual classifiers based on specific attack types. The primary objective of this ensemble intrusion detection approach is to build a robust model capable of identifying attacks by integrating multiple learning models. Experimental results demonstrate that the proposed ensemble model effectively harnesses the advantages of various classifiers, leading to improved performance across diverse scenarios. The findings reveal significant enhancements in overall system performance, including reduced false positive rates and increased accuracy. The ensemble learning approach proves to be highly effective for intrusion detection systems, showcasing its ability to accurately classify data while addressing the inherent weaknesses of standalone classifiers.

References

[1]. Number of Internet of Things (IoT) connections worldwide from 2022 to 2023, with forecasts from 2024 to 2033 Statista. 2024 https://www.statista.com/statistics/1183457/iot-connected-devices-worldwide/

[2]. Wang, W.; Xia, F.; Nie, H.; Chen, Z.; Gong, Z.; Kong, X.; Wei, W. Vehicle Trajectory Clustering Based on Dynamic Representation Learning of Internet of Vehicles. IEEE Trans. Intell. Transp. Syst. 2020.

[3]. Wang, W.; Chen, J.; Wang, J.; Chen, J.; Liu, J.; Gong, Z. Trust-Enhanced Collaborative Filtering for Personalized Point of Interests Recommendation. IEEE Trans. Ind. Inf. 2020, 16, 6124–6132.

[4]. Tang, M.; Liao, H.; Wan, Z.; Herrera-Viedma, E.; Rosen, M.A. Ten Years of Sustainability (2009–2018): A Bibliometric Overview. Sustain. J. Rec. 2018, 10, 1655, doi:10.3390/su10051655.

[5]. Alani MM, Damiani E. XRecon: An Explainbale IoT Reconnaissance Attack Detection System Based on Ensemble Learning. Sensors. 2023; 23(11):5298.

[6]. Alani, M.M. “Detection of Reconnaissance Attacks on IoT Devices Using Deep Neural Networks.” Advances in Nature-Inspired Cyber Security and Resilience Springer Innovations in Communication and Computing. 2022.

[7]. Luo, C.; Tan, Z.; Min, G.; Gan, J.; Shi, W.; Tian, Z. A novel web attack detection system for internet of things via ensemble classification. IEEE Trans. Ind. Inform. 2020, 17, 5810–5818.

[8]. A. Aburomman and M. B. I. Reaz, “A novel SVM-KNN PSO ensemble method for intrusion detection system,” Applied Soft Computing, vol. 38, pp. 360–372, 2016.

[9]. T. Chen and C. Guestrin, “Xgboost: A scalable tree boosting system,” in Proceedings of the 22nd ACM SIGKDD International conference on knowledge discovery and data mining. ACM, 2016, pp. 785–794.

[10]. Khan, M.A, Khan Khattk, M.A., Latif, S.; Shah, A.A., Ur Rehman, M., Boulila, W., Driss, M.; Ahmad, J. Voting classifier based intrusion detection for IoT networks. In Advances on Smart and Soft Computing: Proceedings of ICACIn 2021; Springer: Berlin/Heidelberg, Germany, 2022; pp. 313–328.

[11]. Gutub, A. Boosting image watermarking authenticity spreading secrecy from counting-based secret-sharing. CAAI Trans. Intell. Technol. 2022.

[12]. Abu Al-Haija, Q.; Al-Badawi, A. Attack-Aware IoT network traffic routing leveraging ensemble learning. Sensors 2021, 22, 241.

[13]. M. Asafuddoula, B. Verma, and M. Zhang, “A divide and conquer-based ensemble classifier learning by means of many-objective optimization,” IEEE Transactions on Evolutionary Computation, vol. 22, no. 5, pp. 762–777, 2018.

[14]. Attou H, Mohy-eddine M, Guezzaz A, Benkirane S, Azrour M, Alabdultif A, Almusallam N. 2023. Towards an intelligent intrusion detection system to detect malicious activities in cloud computing. Applied Sciences 13(17):9588 DOI 10.3390/app13179588.

[15]. M. A. Ferrag, O. Friha, D. Hamouda, L. Maglaras and H. Janicke, "Edge-IIoTset: A New Comprehensive Realistic Cyber Security Dataset of IoT and IIoT Applications for Centralized and Federated Learning," in IEEE Access, vol. 10, pp. 40281-40306, 2022.

[16]. B. N. Kushal Kumar, R. Balakrishna, M. V. Panduranga Rao and P. S. Ashok Kumar, "Comprehensive Insights into Machine Learning for Intrusion Detection Systems in IoT and its Datasets," 2024 4th International Conference on Data Engineering and Communication Systems (ICDECS), Bangalore, India, 2024.