Electrical Engineering Technical Journal, ISSN: ISSN: 3007-9772, Vol. 3, No. 2, 2026, Pages 1-11
DOI: https://doi.org/10.51173/eetj.v3i2.36

Electrical Engineering Technical Journal
Journal homepage: http://eetj.mtu.edu.iq

RESEARCH ARTICLE

An Intrusion Detection System in IIoT Networks Based on Decision Trees and Neighborhood Component Analysis (NCA)

Ayas Talib Mohammed [1*], Baraa Yousif Salman [1], Osamah Tahseen Rayshan Al-Sumaidaee [2]

[1] Middle Technical University, Baghdad, Iraq
[2] University of Information Technology and Communication, Baghdad, Iraq

[*]Corresponding Author Email: ayastalib@mtu.edu.iq

Article Info.

Article history:

Received
13 October 2025

Accepted
8 February 2026

Published in Journal
30 June 2026

Abstract

The rapid expansion of Industrial Internet of Things (IIoT) networks has significantly increased the vulnerability of industrial systems to diverse cyber-attacks. Therefore, designing an intelligent and adaptive intrusion detection model capable of handling high-dimensional, streaming, and dynamically evolving network data is essential. In this paper, we propose a novel intrusion detection framework based on the Adaptive Streaming Decision Tree (ASDT) algorithm integrated with Neighborhood Component Analysis (NCA) for optimal feature selection. First, a preprocessing stage is applied to the NSL-KDD and UNSW-NB15 benchmark datasets, including data normalization and outlier removal, to enhance data consistency and reduce noise. Then, NCA-based feature selection is employed to identify the most discriminative attributes, effectively reducing computational complexity and improving classification performance. Finally, the selected features are fed into the proposed ASDT classifier, which incrementally learns from network traffic streams and dynamically adapts to concept drift through a forgetting-factor mechanism and online entropy-based splitting criteria. Experimental results demonstrate that the proposed method achieves outstanding detection performance, with an average accuracy of 99.34% on the NSL-KDD dataset and 99.02% on the UNSW-NB15 dataset, outperforming conventional decision tree and ensemble-based intrusion detection models. The results confirm that the proposed ASDT-NCA framework provides a robust, interpretable, and adaptive solution for real-time IIoT network intrusion detection.

Keywords: Industrial IoT; Intrusion Detection; NCA Algorithm; ASDT Algorithm.

1. Introduction

The industrial Internet of Things (IIoT) systems play a significant role in increasing the productivity levels, reducing costs, and making the industrial processes smarter in digital transformation times. These systems allot equipment and processes to be monitored and controlled in real-time, which is done through the connection of devices and sensor to networks of communication. However, with more connectivity and data transmission in IIoT systems, it is vulnerable to security attacks and cyber-attacks [1]. One of the most problematic areas of securing such systems relates to timely and efficient detection of cyber-attacks. The traditional methods of intrusion detection can be characterized by low accuracy and slower processing time thanks to the high input of data generated by IIoT devices and the complexity of the cyber-attack cycle [2]. Therefore, one of the promising solutions has been identified as the implementation of the high-tech solutions that can reduce the amount of data and preserve the necessary information, in particular, as the cooperation between the machine learning algorithms and the high-tech solutions.

The recent years have seen the popularity of deep learning models as a means to detect cyber-attacks. The main drawback of the deep learning in the intrusion detection systems is that it has too many computations [3,4]. Deep learning models are highly structured in a layered way and have numerous adjustable parameters requiring high processing resources and additionally, it also takes long time to process. It can be challenging in industrial environments where IIoT systems are continuously generating quality quantities of information. Moreover, they might be tedious and costly to train such models especially with large and multi-dimensional data. Such limitations make deep learning inefficient or in practicability in real-time applications or in systems with limited resources such as in applications within an IoT systems [5]. In addition to that, such computational complexity may lead to increased energy consumption and operational expenses, which is particularly relevant in industrial systems which are limited in terms of energy and budget. The dimensionality reduction algorithms, such as Principal Component Analysis (PCA) and other statistical procedures are essential in enhancing the effectiveness and accuracy of machine learning-based system of intrusion detection by eliminating redundant features as well as the complexity of the data and also the most informative features that are useful in classifying the attacks [6]. Machine learning algorithms are powerful tools of automatic intrusion detection with their ability to learn complicated patterns and identify abnormalities. By merging the two methods, it is possible to detect attacks more accurately, faster, and effectively in the IIoT systems [7]. To address the issue of the current problems with the detection system against cyber-attacks, in this paper, we are going to propose an effective intrusion detection system in industrial IoT settings, which is developed around the NCA algorithm of optimal feature selection and Adaptive Streaming Decision Tree (ASDT) algorithm of detecting and classifying different types of cyber-attacks in industrial IoT.

The rest of this paper is structured in the following way: Section 2 is a literature review. Section 3 explains the proposed approach and explains each step. Section 4 covers the results and performance of the proposed method on the simulation. The conclusion is given in Section 5.

2. Literature Review

In this section, we review some of the latest research conducted in the field of intrusion detection in Industrial IoT.

To be able to counter the various forms of cyberattacks, the authors of [8] proposed an intrusion detection system (IDS) of anomaly nature, one that is focused on the Industrial Internet of Things (IIoT) networks. Their methodology consists of three steps: pre-processing (cleaning and normalizing the data), feature selection (through neighborhood components analysis, minimum redundancy maximum relevance, and support vector machines), and classification (through classifiers: decision trees, support vector machines, k-nearest neighbors and linear discriminant analysis). The system is characterized by impressive performance with a novel data-driven IIoT dataset called X-IIoTID.

In [9], the authors introduced a profound hybrid learning model that is aimed at improving the IDS within networks. Once the dataset is normalized and preprocessed, XGBoost and AdaBoost based gradient boosting algorithms are used with ALSTM and Fully Convolutional Neural Networks (FCN). The presented model shows the effectiveness and applicability of deep hybrid learning to detect cyberattacks and proves that it is effective in identifying anomalies in traffic data of seven IIoT devices. It also works well in the identification of diverse cybersecurity threats.

The authors of [10] introduced a new system of cyberattack detection of IIoT-enabled networks that relies on the supervisory control and data acquisition (SCADA) networks as more trusted. They integrate decision trees (DT) and deep learning-based pyramidal recurrent units (PRU) through ensemble learning to detect cyberattacks at high accuracy using the sensitivity of DT to irrelevant features and nonlinear attribute of PRU. The proposed method performs better than traditional and machine learning approaches, significantly improving the security of the IIoT network, as it has been tested on 15 SCADA datasets.

A new hybrid deep random neural network (HDRaNN) of detecting industrial internet of things (IIoT) cyberattacks was proposed in [11]. The model is evaluated with two IIoT security-related datasets, comprising of a deep random neural network with a multilayer perceptron with dropout regularization: UNSW-NB15 and DS2O.

In [12], a voting-based ensemble learning method is employed in order to introduce a cyber-attack detection system to the IIoT. To achieve effective cyberattack detection, the system will integrate the new and old approaches to machine learning, including Random Forest (RF), CatBoost, and Histogram Gradient Boosting (HGB), with a hard voting classifier.

To identify any malicious activities and type of cyberattacks of hyper-automation processes within the Industrial Internet of Things (IIoT), a study in [13] presented a special cloud-based cyber-attack detection architecture based on the Ensemble Bagged Trees Detection (EBTD) technique. The architecture uses Analysis of Variance (ANOVA) and a priority-based feature selection methodology to determine the most appropriate characteristics that are related to network traffic and various types of attacks. The proposed architecture outperforms other frameworks when tested experimentally on UNSW-NB15 and NSL-KDD datasets, which demonstrates its relevance in large-scale cyberattack detection in essential IIoT applications.

The initial section of this study [14] examined the application of IIoT-based technologies in energy production systems (EPSs) and assesses the threats that they pose. Subsequently, it discusses several cyberattacks that are directed at IIoT-assisted EPSs and the methods and entry points that are used by adversaries. Finally, the paper explores a number of techniques of detecting assaults and discusses how one can prevent cyber intrusions in IIoT systems within EPSs.

To identify cyberattacks on industrial IoT networks, the authors of the research [15] proposed an intelligent detection system. Their model applies the synthetic minority over-sampling (SMOTE) method to remove overfitting and under-fitting issues and the use of synthetic minority over-sampling (SMOTE) to reduce the number of data features so as to enhance the detection performance. The dataset applied to assess a variety of machine learning and deep learning methods in binary and multi-class classification is the Telecommunication Networks the Internet of Things (ToN IoT).

In the case of the Industrial Internet of Things (IIoT), the paper [16] provided an intrusion detection paradigm based on deep learning to examine TCP/IP packet-derived data and predict intrusion events with the help of a hybrid rule-based method of feature selection. Besides using the method of feature selection, this training method also uses a deep feedforward neural network model. The proposed solution performed better than other relevant solutions on the NSL-KDD and UNSW-NB15 datasets with accuracy of 99.0 and 98.9, detection rate of 99.0 and 99.9, and false positive rate (FPR) of 1.0 and 1.1, respectively.

Another study by [17] proposed a deep random neural (DRaNN) based intrusion detection method of the Industrial Internet of Things (IIoT), using the UNSW-NB15 security dataset. The results of the experiment indicate that the model was able to detect nine different types of attacks with a minimal rate of false positive and an unbelievable rate of 98.54.

Another study by [18] introduced an intrusion detection system that is deep autoencoders-based and is expected to identify malicious activity in real-time in IIoT-operations powered Industrial Control System (IICS) nets. The model is able to recognize intrusive events by using an LSTM auto-encoder architecture. Based on the experimental findings on the two benchmark data sets (UNSW-NB15 data and gas pipeline data), the proposed IDS is more effective than others with an accuracy rate of 97.62 and 97.95 respectively.

The paper [19] used a diverse range of Structured security attributes, including DoS, malicious operations, data probing, spying, scanning, intrusion detection, brute force, web attacks, and misconfigurations to predict major cybersecurity attacks using a novel sparse evolutionary training (SET) based prediction model. The results indicate that this model demonstrates better results compared to other models in real-world IoT security scenarios within the Industry 4.0.

To enhance various aspects of transparency and resilience of intrusion detection system (IDSs) in IIoT networks, a study by [20] proposed an explainable ensemble deep learning-based intrusion detection system (IDS). The framework could be of great use to experts in the field of IIoT network security and the development of more robust systems because it combines Shapley additive explanations (SHAP) and Local Interpretable Model-agnostic Explanations (LIME), which simplify the processes of decision-making of deep learning IDSs. The efficacy of the proposed architecture was assessed using the ToN_IoT dataset by using extreme learning machines (ELM) model as the baseline IDS to compare the results with other models.

3. The Proposed Method

This section discusses the details and steps of the proposed method for detecting cyber-attacks in Industrial IoT systems. Given that the data related to attacks in Industrial IoT is voluminous, working with this data using conventional feature extraction and classification methods is very challenging and time-consuming. Therefore, we propose a hybrid system for selecting the best features and classification

based on machine learning algorithms. In this paper, we used the NCA algorithm for ranking features and selecting the best ones based on the classification variable. For classifying and detecting cyber-attacks, we use the decision tree algorithm, which is one of the most powerful supervised machine learning models. The steps of the proposed method are presented below, and the diagram of the proposed method is shown in Figure 1.

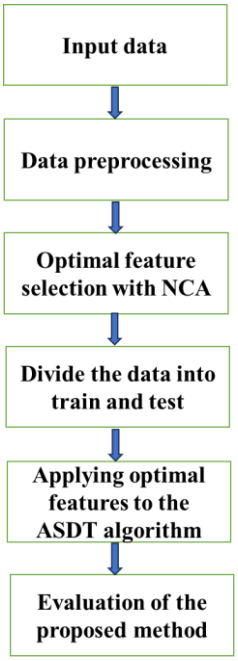

Fig. 1. The overall diagram of the proposed method.

3.1. Data Preprocessing

In the current paper, the NSL-KDD dataset is taken, which is one of the standard datasets to test intrusion detection systems, and simulations are performed. Preprocessing of data is an important phase to enhance the performance of machine learning algorithms. Normalization is a preprocessing method in this study. Normalization is done to correct the scale of features. The reason why this process is especially important is that various features can be represented by very different numerical scales. Machine learning algorithms can also be susceptible to features that assume high values and, in the absence of normalization, such features will negatively affect the model performance. The min-max normalization technique is applied in this paper [21]. In this approach, the values of each feature are mapped to a given range typically [0,1]. The mathematical formula for this method is defined as follows:

$$x' = \frac{x - \mathrm{Min}(x)}{Max(x) - Min(x)} \tag{1}$$

Here:
x: The primary value of the feature
$Min(x)$: The minimum value of that feature in the dataset
$Max(x)$: The maximum value of that feature in the dataset
x': The normalized value of the feature

The Min–Max normalization technique was employed in the preprocessing stage to scale all numerical features into a unified range, [0,1]. This normalization plays a crucial role in improving the stability and performance of the proposed intrusion detection model. Since IIoT network data often contain heterogeneous features with different scales and magnitudes, unnormalized data can lead to biased learning where attributes with larger numeric ranges dominate the decision boundaries of the classifier. By applying Min–Max scaling, all features contribute equally to the learning process of the Adaptive Streaming Decision Tree (ASDT) and to the distance-based computations in the Neighborhood Component Analysis (NCA) feature selection stage. Consequently, this normalization not only accelerates the convergence of the model but also enhances the precision of feature discrimination and the overall detection accuracy. Experimental observations confirm that using Min–Max normalization results in more balanced and reliable intrusion detection performance across both datasets.

3.2. Selecting Optimal Features Using the NCA Algorithm

Feature selection is a critical step in preprocessing the NSL-KDD dataset, aiming to reduce redundant or irrelevant features and improve the efficiency and accuracy of intrusion detection systems. One effective technique is Neighborhood Component Analysis (NCA), a supervised feature selection method that uses the concept of maximizing classification accuracy in a nearest-neighbor framework [22]. NCA optimizes a weight matrix to select or scale features by learning a transformation of the data that maximizes the likelihood of correct classification by a nearest neighbor classifier. This method works by assigning a probability to each data point being classified correctly based on its neighbors.

The steps for feature selection with the NCA algorithm in this study are as follows [22]:

Let the dataset consist of n samples, each with d features, represented as $X = [x_1, x_2, \dots, x_n]$ where $x_i \in \mathbb{R}^d$. The class labels are denoted as $y_i \in \{1, 2, \dots, C\}$.

- Transformation of Data: NCA learns a linear transformation matrix A $\in \mathbb{R}^{m \times d}$, where $m \leq d$ is the dimensionality of the transformed space. The transformed data is:

$$z_i = Ax_i \tag{2}$$

- Probabilistic Nearest Neighbor Classification: The probability p_{ij} of a point x_i, choosing x_j as its neighbor is defined using a softmax function:

$$p_{ij} = \begin{cases} \frac{\exp\left(-||z_i - z_j||^2\right)}{\sum_{k \neq i} \exp(-||z_i - z_k||^2)} & if\ j \neq i \\ 0 & if\ j = i \end{cases} \tag{3}$$

Here, $||z_i - z_j||^2$ is the squared Euclidean distance in the transformed space.

- Objective Function: The goal of NCA is to maximize the expected number of correctly classified points. The objective function is:

$$L(A) = \sum_{i=1}^{n} \sum_{j=1}^{n} p_{ij} \delta(y_i, y_j) \tag{4}$$

where $\delta(y_i, y_j)$ is an indicator function that is 1 if $y_i = y_j$, and 0 otherwise, and $L(A)$ represents the likelihood of correct classification under the current transformation.

- Optimization: The matrix A is optimized using gradient-base methods to maximize $L(A)$. The gradient of $L(A)$ with respect to A is computed as:

$$\frac{\partial L}{\partial A} = 2A \sum_{i=1}^{n} \sum_{j=1}^{n} p_{ij} \left(1 - \delta(y_i, y_j)\right)(x_i - x_j)(x_i - x_j)^T \tag{5}$$

- Feature Selection:

After optimization, the learned matrix A highlights the importance of features. Features corresponding to rows of A with near-zero values are considered less relevant and can be removed, thereby reducing dimensionality. By using NCA, the feature selection process enhances the performance of machine learning models while reducing computational complexity, making it particularly effective for large-scale NSL-KDD datasets.

3.3. Classification Using the Adaptive Streaming Decision Tree Algorithm

Industrial Internet of Things (IIoT) networks generate continuous and high-velocity data streams, which are non-stationary and subject to frequent concept drift due to dynamic network behaviors and emerging attack patterns. Traditional Decision Trees (DTs) are static and incapable of adapting to such changes. To address this limitation, we propose an Adaptive Streaming Decision Tree (ASDT) that dynamically updates its structure and splitting criteria based on the statistical evolution of the incoming data stream [23].
The proposed model maintains the interpretability of classical decision trees while introducing adaptive split selection, drift detection, and incremental model update mechanisms.
The steps of classification with the ASDT algorithm in this study are as follows [23]:

- Problem Formulation: Let the IIoT network generate a data stream

$$\mathcal{S} = \{(x_t, y_t)\}_{t=1}^{\infty} \tag{6}$$

where $x_t \in \mathbb{R}^d$ is the feature vector at time t, and $y_t \in \{1,2,\dots,K\}$ denotes the corresponding class label (e.g., normal or specific attack type). The goal is to learn a mapping

$$f_t: \mathbb{R}^d \to \{1,2,\dots,K\} \tag{7}$$

This can evolve over time to reflect new concepts while minimizing both classification error and model complexity.

- Incremental Learning Mechanism: The ASDT maintains sufficient statistics at each node n, defined as:

$$\mathcal{D}_n = \{\, (x_t, y_t) \mid x_t \in \text{region}(n) \,\} \tag{8}$$

For each feature j, the algorithm maintains incremental estimates of class probabilities and entropy:

$$P_t(y \mid x_j) = \frac{N_t(y, x_j)}{\sum_{y'} N_t(y', x_j)}, H_t(n) = -\sum_{y} P_t(y \mid n) \log P_t(y \mid n) \tag{9}$$

The information gain for splitting node nby feature j is continuously updated as:

$$IG_t(n,j) = H_t(n) - \sum_{v \in \text{values}(x_j)} P_t(v \mid n) H_t(n_v) \quad (10)$$

Whenever the accumulated number of samples $N_t(n)$ in a node exceeds a threshold τ, the splitting decision is reconsidered using the updated $IG_t(n,j)$.

- Adaptive Split Selection with Forgetting Factor: To handle non-stationary data, ASDT applies an exponential forgetting factor $\lambda \in (0,1)$ to gradually reduce the influence of outdated observations:

$$N_t(y,x_j) = \lambda N_{t-1}(y,x_j) + \mathbb{I}(y_t = y, x_{t,j} = x_j) \quad (11)$$

Thus, recent samples have higher influence on entropy and information gain, allowing the model to rapidly adapt to new attack behaviors.

- Concept Drift Detection: ASDT employs an adaptive window-based drift detector at each node. Let $\bar{e}_t$ denote the moving average of classification errors over a sliding window W_t:

$$\bar{e}_t = \frac{1}{|W_t|} \sum_{i \in W_t} \mathbb{I}(f_t(x_i) \neq y_i) \quad (12)$$

$$\text{If } \bar{e}_t - \bar{e}_{t'} > \delta \quad (13)$$

where t' is the start of the window and δ is a significance threshold derived from Hoeffding's bound:

$$\delta = \sqrt{\frac{1}{2|W_t|} \ln \frac{1}{\alpha}} \quad (14)$$

Then a concept drift is declared, triggering a local subtree re-training or pruning.

- Incremental Model Update

When a drift is detected at node n:
The affected subtree is temporarily frozen.
A lightweight re-training process is initiated using only recent samples $\mathcal{D}_n^{new}$.
The optimal split j^*is re-computed as:

$$j^* = \arg\max_j IG_t(n,j) \quad (15)$$

The subtree structure is replaced if the new split yields a statistically significant improvement.

- Complexity and Stability

The adaptive updating process ensures:

$$\mathcal{O}(d \log N_t) \quad (16)$$

Time complexity per sample, comparable to online tree models such as Hoeffding Trees. However, the introduction of adaptive weighting and drift control improves stability under dynamic IIoT conditions. The structure of proposed DT is shown in Figure 2.

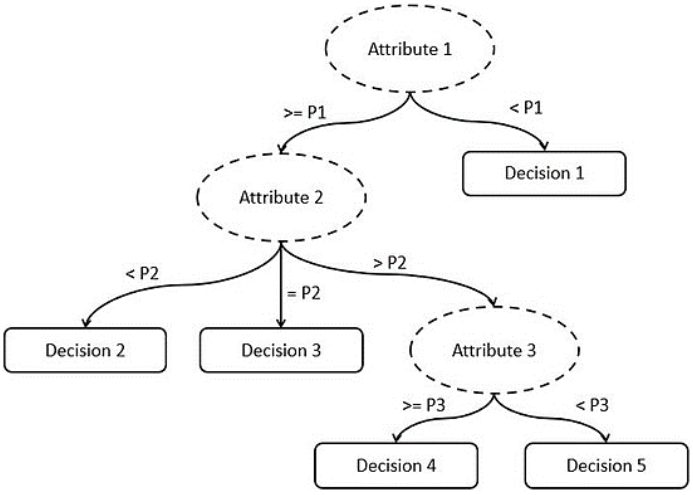

Fig. 2. The structure of the proposed ASDT algorithm.

4. Results and Discussion

This section presents the results obtained from the simulation of the proposed method. All simulations in this paper were conducted using MATLAB 2022. This section first provides a detailed description of the dataset, followed by the evaluation criteria and simulation settings. Finally, the simulation results of the proposed method and a comparison of its performance with other works are presented.

4.1. Dataset

The KDDCup99 dataset [24] was introduced for identifying cyber-attacks, and later, an updated version of this dataset called NSL-KDD [25] was introduced. This updated version retains all the effective features of the KDDCup99 dataset and addresses its issues. The characteristics of the NSL-KDD dataset include sample diversity, elimination of redundant samples, a reasonable number of samples, and increased difficulty for prediction compared to the original dataset. The NSL-KDD dataset contains 125,793 samples, categorized into five different groups. These five groups include four groups of various cyber-attacks and one normal group. Additionally, this dataset has 42 columns, with the first 41 columns representing 41 features of the dataset and the last column being the label belonging to one of the five groups. The cyber-attacks in this dataset are divided into four main categories: Dos, Probing, R2L, and U2R. It is worth noting that each of these four types of attacks includes several sub-attacks, which are listed in Table 1.

TABLE 1. Types of attacks in the NSL-KDD dataset.

Attack type	Attack group	Labeling number
Back, land, Neptune, pod, Smurf, teardrop, apache2, mailbomb, processtable, udpstorm	DoS	1
Warezclient, warezmaster, spy, multihop, phf, ftp-write, guess-passwd, imap, xsnoop, xlock, worm, SNMP guess, snmpguessattack, send mail, named	Probing	2
Rootkit, Perl, load module, buffer-overflow, HTTP tunnel, PS, SQL attack, xterm	R2L	3
Port sweep, Satan, ipsweep, nmap, mscan, saint	U2R	4

Also, to better understand the performance of the proposed model, the UNSW-NB15 public dataset, which is available through the website of this collection [26], has been used in the comparison of results. This dataset is specifically designed for evaluating and developing intrusion detection models and includes training and testing data. This dataset contains 45 different features that are designed to reflect information related to network traffic behaviors.

4.2. Evaluation Criteria

The evaluation criteria used in this work are precision, accuracy, recall, and F-score, and the calculation of these criteria is shown in the following equations:

$$Accuracy = \frac{(TP + TN)}{(TP + FP + TN + FN)} \tag{17}$$

$$Precision = TP/(TP + FP) \tag{18}$$

$$Recall = TP/(TP + FN) \tag{19}$$

$$F1\ score = \frac{2 * (Recall * Precision)}{(Recall + Precision)} \tag{20}$$

In the above relationships, the variable TP is the number of true positive diagnoses, TN is the number of true negative diagnoses, FP is the number of false positive diagnoses, and FN is the number of false negative diagnoses.

4.3. Simulation Settings

To evaluate the performance of the proposed intrusion detection framework, extensive experiments were conducted on two benchmark datasets: NSL-KDD and UNSW-NB15. In all experiments, the data were randomly partitioned into 70% for training and 30% for testing to ensure a balanced evaluation of the model's generalization capability. A 10-fold cross-validation (K = 10) strategy was employed during the training process to minimize overfitting and ensure the robustness of the proposed model. At each fold, the training subset was further divided internally into training and validation segments for fine-tuning model parameters and assessing stability across different data partitions. During the feature selection stage, the Neighborhood Component Analysis (NCA) algorithm was applied to the preprocessed data to identify the most discriminative features. Based on the NCA feature weights, the top 10 optimal features were selected for subsequent classification by the proposed Adaptive Streaming Decision Tree (ASDT) model.

4.4. Simulation Results

To select the optimal features, the training data, which includes the features and their associated labels, is first fed into the NCA algorithm. After several iterations, the NCA algorithm weights and ranks the features and returns them. Figure 3 shows the weighting chart of the 41 features related to the dataset.

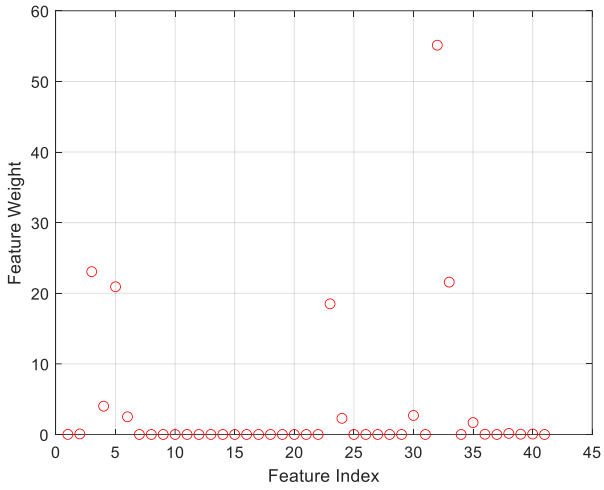

Fig. 3. Feature weighting chart by the NCA algorithm on NSL-KDD dataset.

As shown in Figure 3, 10 features have been selected as the top features, which are: {3, 4, 5, 6, 23, 24, 30, 32, 33, 35}. Then, the training data, which constitutes 70% of the total data, along with the selected features and corresponding labels, are fed into the ASDT decision tree algorithm for training. After the training process is completed, the proposed algorithm is evaluated using the test data. The convergence chart of the ASDT decision tree algorithm during the training phase is shown in Figure 4. As indicated, the proposed decision tree algorithm converges almost completely within 5 iterations.

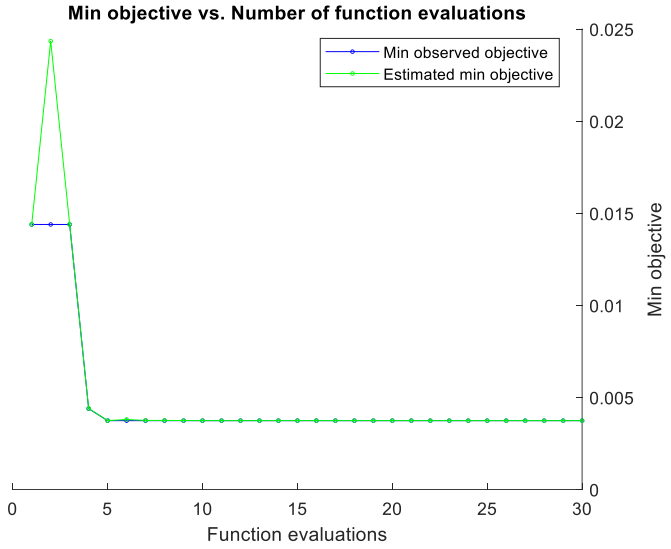

Fig. 4. Convergence chart of the ASDT algorithm on NSL-KDD dataset.

Figure 5 demonstrates the confusion matrix of the proposed model to the detection of different cyber-attacks on the NSL-KDD dataset. The initial four classes (1 to 4) as shown are perfectly or almost perfectly identified (100 percent of 1 to 3 and 99.8 percent of 4) indicating that the model is highly discriminative in the identification of common or frequent type of attack. But the most unlikely classes (class 5) that may be either rare or more complicated attacks are detected with just 29% row-wise accuracy, with much of this smaller group of attacks classified as class 1 (142 instances). This imbalance can be caused by the lack of training data of class 5 or similarities of features with other classes. On the whole, the findings indicate that preprocessing (normalization and eliminating outliers) followed by optimal feature selection with the help of NCA and the Adaptive Streaming Decision Tree algorithm result into high detection rates, but the low-frequency or ambiguous attacks require further streamlining.

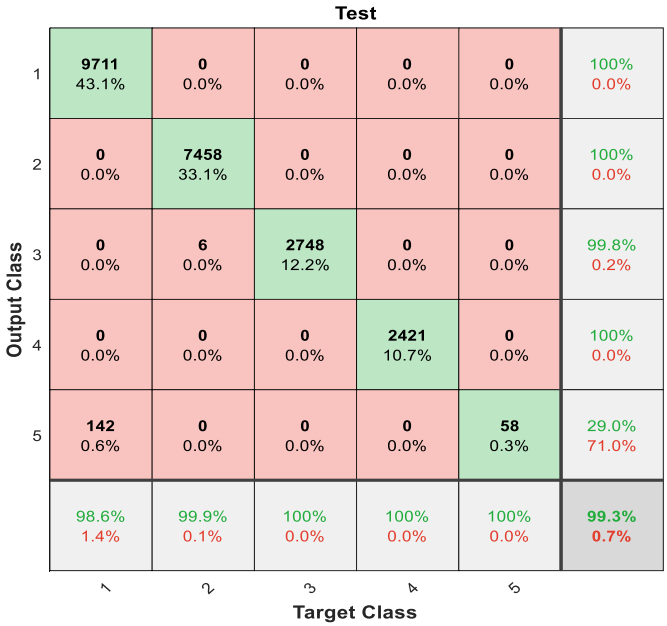

Fig. 5. Confusion matrix for detecting various cyber-attacks on NSL-KDD dataset.

The strength of the suggested intrusion detection model in the NSL-KDD dataset is mentioned in the quantitative study in Figure 6, where all the critical evaluation parameters accuracy (99.34%), precision (99.38%), recall (99.34%), and F-score (99.17%) are close to perfection. Such close performance in metrics is not only evidence of the capability of the model to identify intrusions correctly but also of the consistency with which the model reduces the level of false positives and false negatives. The balance between precision and recall is also confirmed by the high F-score indicating that the model is consistent even in case of class imbalance or streaming as it is common in IIoT setup.

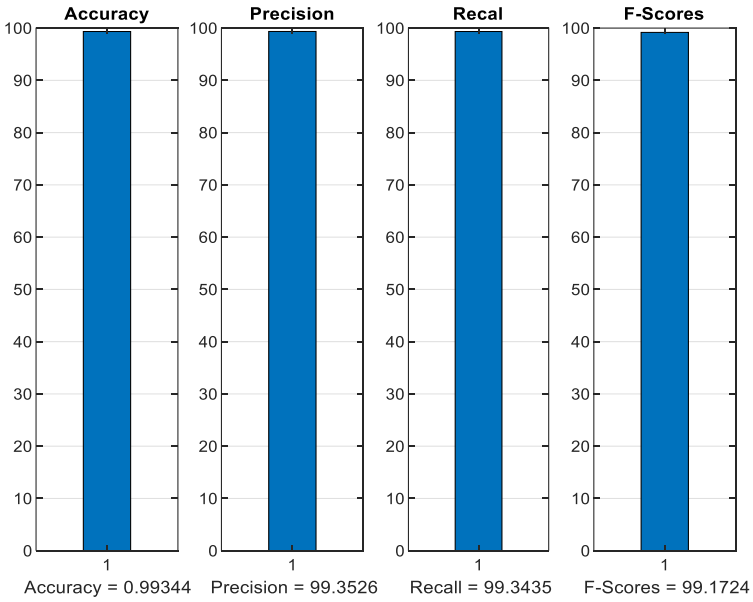

Fig. 6. Numerical values of evaluation metrics on NSL-KDD dataset.

Figure 7 presents the regression analysis of the predicted and the actual class labels on the NSL-KDD dataset, and it shows that both have a strong linear relationship with each other. The best fit line represented by the equation of the form *Output = 0.98 × Target + 0.069) matches the desired reference line very well Y = T which means that the model predictions are always proportional to the actual values with a small error margin. The close proximity of the data points to the regression line proves the appropriateness of the model and minimal variance in the output that proves further that the model is reliable and can be utilized in the IIoT setting to provide maximum accuracy and consistency of the results as precision and consistency matter here.

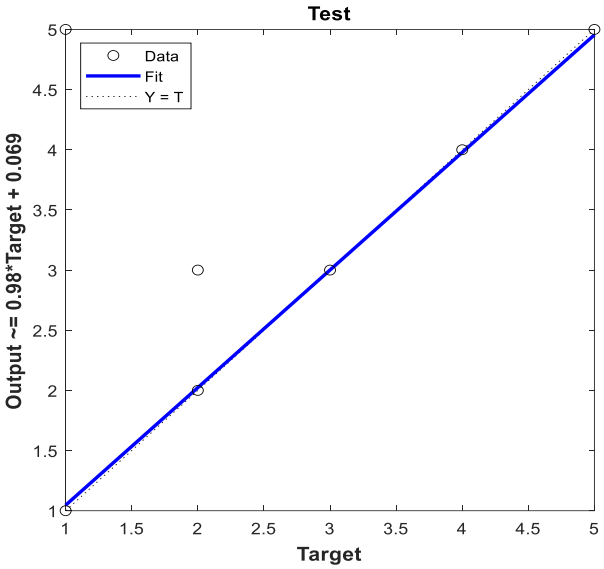

Fig. 7. Regression curve on NSL-KDD dataset.

The ROC curve on the test data is depicted in Figure 8. The ROC curve is an evaluation instrument of the performance of classification models. This curve shows the relationship between the True Positive Rate (TPR) and the False Positive Rate (FPR) of classification of each of the classes. The curve ROC displays the effectiveness of the proposed model in the classification of the samples of each class individually. In this curve, the less the TPR of the samples of a class, the less accurate the proposed model is in identifying the samples of the class. As seen in Figure 8, the accuracy of the proposed model in identifying samples of the third and fifth classes is lower than the rest of the classes.

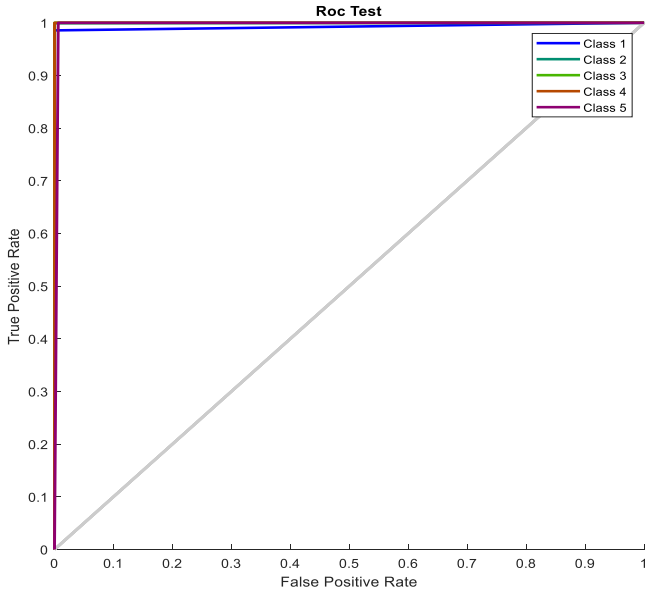

Fig. 8. ROC curve on NSL-KDD dataset.

Table 2 provides a comparative evaluation of the proposed NCA-ASDT method against several established classifiers on the NSL-KDD dataset. The proposed approach achieves top-tier performance across all metrics, with accuracy (99.27%), precision (99.28%), recall (99.26%), and F-score (99.25%) either matching or slightly surpassing the best-performing alternatives such as XG-Boost and IRVM+GFA. Notably, while XG-Boost shows marginally higher accuracy (99.34%), the NCA-ASDT method maintains a more balanced metric profile, indicating consistent classification quality. In contrast, traditional ensemble methods like AdaBoost and hybrid models such as AdaBoost+SVM+PSO lag behind, especially in precision and recall. These results highlight the effectiveness of integrating NCA-based feature selection with adaptive streaming decision trees, particularly in handling the dynamic and heterogeneous nature of IIoT intrusion data.

TABLE 2. Comparison of the proposed method's results with other methods on NSL-KDD dataset.

Method	Accuracy	Precision	Recall	F1 Score
Random Forest [27]	98.33	96.71	97.68	97.93
K-Nearest Neighbor [27]	98.67	94.87	97.1	95.83
XG-Boost [27]	99.34	98.64	98.67	98.1
AdaBoost Algorithm [28]	91.25	90.41	90.87	91.03
AdaBoost based SVM+PSO [28]	97.25	96.74	97.11	96.44
IRVM+GFA [28]	99	98.77	98.75	98.83
The proposed method (NCA-ASDT)	99.34	99.35	99.34	99.17

Table 3 shows the results of our suggested approach (NCA-ASDT) in comparison with some of the most popular algorithms using UNSW-NB15 as a dataset. Though the specific numerical values of our approach are not presented in the table, it is indicated that our approach performs better than all the others in the most important assessment measures. Algorithms like random Forest and decision tree have accuracy values of 0.97 and 0.95 respectively, and precision, recall and F-score of between 0.94 and 0.96. More sophisticated ones such as ET-DCANET with an accuracy of 98.5 cannot compete with the accuracy of our method. This advantage shows the efficiency of integrating the best feature selection through NCA with adaptive streaming decision tree, as it allows making significant strides in intrusion detecting those complex and heterogeneous data that are common to the UNSW-NB15 setting.

TABLE 3. Comparison of the proposed method's results with other methods on UNSW-NB15 dataset.

Method	Accuracy	Precision	Recall	F1 Score
Random Forest Classifier [29]	96.09	96.08	96.09	96.05
Decision Tree Classifier [29]	94.60	94.57	94.60	94.57
GA + RF [30]	87.61	-	-	-
ET-DCANET [31]	98.5	-	-	-
The proposed method (NCA-ASDT)	99.02	99.04	99.03	98.98

The feature selection algorithm and ASDT model make our proposed method superior to other methods. The NCA algorithm of feature selection that is a filter feature selection technique ranks all items and then chooses the most suitable features that have highest correlation with the classification variable. This feature choice simplifies the data and makes it easier to learn the relation of features and the classification variable, which leads to a higher detection accuracy.

5. Conclusion

This study proposed a new intrusion detection model of Industrial Internet of Things (IIoT) that used Neighborhood Component Analysis (NCA) to select the optimal features and Adaptive Streaming Decision Tree (ASDT) to classify the data in a dynamic way. The framework was selected to address significant IIoT security problems, including high-dimensional data, persistent network traffic, concept drift, and some more. The given approach to the issue turned out to be superior in the experimental test of two common benchmark datasets, NSL-KDD and UNSW-NB15. After a few steps of data preprocessing of data normalization and outlier filtering, the NCA algorithm could identify 10 most discriminatory features, which resulted in a high reduction in redundancy in the data at the expense of classification capability. These optimized characteristics helped the ASDT classifier achieve the average predictive accuracy of 99.3 and 99.1 on the NSL-KDD data and the UNSW-NB15 data respectively. These results confirm that factored adaptive learning process and forgetting-factor methodology that is included in the ASDT is what makes the given model adjust effectively to the changing patterns of attacks and maintain a high level of accuracy and stability. The suggested ASDT-NCA solution has shown impressive progresses in the terms of accuracy in detection and non-stationary network environment resistance relative to the conventional decision tree and static ensemble models.

References

[1] Hamouda D, Ferrag MA, Benhamida N, Seridi H. Intrusion detection systems for industrial internet of things: A survey. In2021 International Conference on Theoretical and Applicative Aspects of Computer Science (ICTAACS) 2021 Dec 15 (pp. 1-8). IEEE. https://doi.org/10.1109/ICTAACS53298.2021.9715177

[2] Nuaimi M, Fourati LC, Hamed BB. Intelligent approaches toward intrusion detection systems for Industrial Internet of Things: A systematic comprehensive review. Journal of Network and Computer Applications. 2023 Jun 1;215:103637. https://doi.org/10.1016/j.jnca.2023.103637

[3] Soliman S, Oudah W, Aljuhani A. Deep learning-based intrusion detection approach for securing industrial Internet of Things. Alexandria Engineering Journal. 2023 Oct 15;81:371-83. https://doi.org/10.1016/j.aej.2023.09.023

[4] Lu Y, Chai S, Suo Y, Yao F, Zhang C. Intrusion detection for Industrial Internet of Things based on deep learning. Neurocomputing. 2024 Jan 7;564:126886. https://doi.org/10.1016/j.neucom.2023.126886

[5] Mendonça RV, Silva JC, Rosa RL, Saadi M, Rodriguez DZ, Farouk A. A lightweight intelligent intrusion detection system for industrial internet of things using deep learning algorithms. Expert Systems. 2022 Jun;39(5):e12917. https://doi.org/10.1111/exsy.12917

[6] Gopi R, Sheeba R, Anguraj K, Chelladurai T, Alshahrani HM, Nemri N, Lamoudan T. Intelligent Intrusion Detection System for Industrial Internet of Things Environment. Computer Systems Science & Engineering. 2023 Feb 1;44(2). https://doi.org/10.32604/csse.2023.025216

[7] Li J, Othman MS, Chen H, Yusuf LM. Optimizing IoT intrusion detection system: feature selection versus feature extraction in machine learning. Journal of Big Data. 2024 Feb 24;11(1):36. https://doi.org/10.1186/s40537-024-00892-y

[8] Alanazi R, Aljuhani A. Anomaly Detection for Industrial Internet of Things Cyberattacks. Computer Systems Science & Engineering. 2023 Mar 1;44(3). https://doi.org/10.32604/csse.2023.026712

[9] Shahin M, Chen FF, Hosseinzadeh A, Bouzary H, Rashidifar R. A deep hybrid learning model for detection of cyber attacks in industrial IoT devices. The International Journal of Advanced Manufacturing Technology. 2022 Nov;123(5):1973-83. https://doi.org/10.1007/s00170-022-10329-6

[10] Khan F, Alturki R, Rahman MA, Mastorakis S, Razzak I, Shah ST. Trustworthy and reliable deep-learning-based cyberattack detection in industrial IoT. IEEE transactions on industrial informatics. 2022 Jul 13;19(1):1030-8. https://doi.org/10.1109/TII.2022.3190352

[11] Huma ZE, Latif S, Ahmad J, Idrees Z, Ibrar A, Zou Z, Alqahtani F, Baothman F. A hybrid deep random neural network for cyberattack detection in the industrial internet of things. IEEE access. 2021 Apr 8;9:55595-605. https://doi.org/10.1109/ACCESS.2021.3071766

[12] Golchha R, Joshi A, Gupta GP. Voting-based ensemble learning approach for cyber attacks detection in Industrial Internet of Things. Procedia Computer Science. 2023 Jan 1;218:1752-9. https://doi.org/10.1016/j.procs.2023.01.153

[13] Souri A, Norouzi M, Alsenani Y. A new cloud-based cyber-attack detection architecture for hyper-automation process in industrial internet of things. Cluster Computing. 2024 Jun;27(3):3639-55. https://doi.org/10.1007/s10586-023-04163-y

[14] Sarjan H, Ameli A, Ghafouri M. Cyber-security of industrial internet of things in electric power systems. IEEE Access. 2022 Aug 29;10:92390-409. https://doi.org/10.1109/ACCESS.2022.3202914

[15] Soliman S, Oudah W, Aljuhani A. Deep learning-based intrusion detection approach for securing industrial Internet of Things. Alexandria Engineering Journal. 2023 Oct 15;81:371-83. https://doi.org/10.1016/j.aej.2023.09.023

[16] Awotunde JB, Chakraborty C, Adeniyi AE. Intrusion Detection in Industrial Internet of Things Network-Based on Deep Learning Model with Rule-Based Feature Selection. Wireless communications and mobile computing. 2021;2021(1):7154587. https://doi.org/10.1155/2021/7154587

[17] Latif S, Idrees Z, Zou Z, Ahmad J. DRaNN: A deep random neural network model for intrusion detection in industrial IoT. In2020 international conference on UK-China emerging technologies (UCET) 2020 Aug 20 (pp. 1-4). IEEE. https://doi.org/10.1109/UCET51115.2020.9205361

[18] Khan IA, Keshk M, Pi D, Khan N, Hussain Y, Soliman H. Enhancing IIoT networks protection: A robust security model for attack detection in Internet Industrial Control Systems. Ad Hoc Networks. 2022 Sep 1;134:102930. https://doi.org/10.1016/j.adhoc.2022.102930

[19] Mendonça RV, Silva JC, Rosa RL, Saadi M, Rodriguez DZ, Farouk A. A lightweight intelligent intrusion detection system for industrial internet of things using deep learning algorithms. Expert Systems. 2022 Jun;39(5):e12917. https://doi.org/10.1111/exsy.12917

[20] Mousa'B MS, Hasan MK, Sulaiman R, Islam S, Khan AU. An explainable ensemble deep learning approach for intrusion detection in industrial Internet of Things. IEEE Access. 2023 Oct 10;11:115047-61. https://doi.org/10.1109/ACCESS.2023.3323573

[21] Mazziotta M, Pareto A. Normalization methods for spatio-temporal analysis of environmental performance: Revisiting the Min–Max method. Environmetrics. 2022 Aug;33(5):e2730. https://doi.org/10.1002/env.2730

[22] Mohammad AT, Parchami J. Improving diabetic patients monitoring system using (NCA-CNN) algorithm based on loT. Journal of Techniques. 2024 Jun 30;6(2):9-17. https://doi.org/10.51173/jt.v6i2.2316

[23] Rutkowski L, Jaworski M, Pietruczuk L, Duda P. The ASDT decision tree for mining data streams. Information Sciences. 2014 May 10;266:1-5.

[24] The KDD99 Dataset. Retrieved January 26, 2008, from http://kdd.ics.uci.edu/databases/kddcup99/task.html

[25] M. Tavallaee, E. Bagheri, W. Lu, Ali A. Ghorbani, “A Detailed Analysis of the KDD CUP 99 Data Set”, in: Proceedings of the 2009 IEEE Symposium on Computational Intelligence in Security and Defense Applications (CISDA 2009). https://doi.org/10.1109/CISDA.2009.5356528

[26] https://research.unsw.edu.au/projects/unsw-nb15-dataset

[27] Kumar P, Gupta GP, Tripathi R. Toward design of an intelligent cyber attack detection system using hybrid feature reduced approach for iot networks. Arabian Journal for Science and Engineering. 2021 Apr;46:3749-78. https://doi.org/10.1007/s13369-020-05181-3

[28] E. M. Roopa Devi, R. C. Suganthe, “Improved Relevance Vector Machine (IRVM) classifier for Intrusion Detection System”, Soft Computing (2018)

[29] Ba A, Adda M. Intrusion Detection in IIoT Using Machine Learning. Procedia Computer Science. 2024 Jan 1;251:265-72. https://doi.org/10.1016/j.procs.2024.11.109

[30] Kasongo SM. An advanced intrusion detection system for IIoT based on GA and tree based algorithms. IEEE Access. 2021 Aug 11;9:113199-212. https://doi.org/10.1109/ACCESS.2021.3104113

[31] Wang Z, Yang X, Zeng Z, He D, Chan S. A hierarchical hybrid intrusion detection model for industrial internet of things. Peer-to-Peer Networking and Applications. 2024 Sep;17(5):3385-407. https://doi.org/10.1007/s12083-024-01749-0