International Journal of Cyber Threat Intelligence and Secure Networking (IJCTISN)
eISSN: 3087-4297
Volume03 Issue08 | Page No: - 32-40

Received: 30 June 2026, Revised: 15 July 2026, Accepted: 28 July 2026, Published: 15 August 2026

DOI: https://doi.org/10.55640/ijctisn-v03i08-06

Cybersecurity Threat Intelligence Using Machine Learning Classification Techniques

Mr. Madhav Sharma [1]

[1] Assistant Professor, Department of Computer Science and Applications, Mandsaur (M.P.), India

ABSTRACT: New attacks are getting smarter and more sophisticated, so the old signature-based intrusion detection and prevention systems can't find them. This work proposes a machine learning approach to build a cybersecurity threat intelligence framework for effective multiclass intrusion detection, in which the Decision Tree classifier is used. The CICIDS2017 benchmark dataset, which contains both benign network traffic and several types of cyberattacks, is used to build and test the suggested model. The goal of the preparation process is to enhance classification performance by cleaning and separating data, utilizing Standard Scaler to scale features, and SMOTE to balance classes. Metrics like as recall, accuracy, precision, F1-score, confusion matrix, and ROC curve are used to test the Decision Tree model. An impressive 99.91% accuracy (ACC) rate, 97.79% precision (PRE), 97.08% recall (REC), 97.41% F1-score (F1), and 0.99 AUC were revealed by the experiment's outcomes. The suggested method beats state-of-the-art deep learning and ML approaches in terms of performance, execution time, and computational complexity. The results show that the suggested design is a reliable, efficient, and lightweight way to find cyber security threats and IDR apps with intelligence.

KEYWORDS: Cybersecurity Threat Intelligence, Decision Tree, Machine Learning, Intrusion Detection System (IDS), Cyberattacks.

1. Introduction

Cyber security has emerged as a critical concern as technology is being increasingly adopted in different industries and sectors, leading to an increase in cyber threats and attacks. Significant financial losses, operational delays, and data breaches are occurring as a consequence of traditional cybersecurity solutions' inability to identify and respond to new attack types. These new types include malware, phishing, ransomware, distributed denial of service, botnets, and network penetration [1][2]. Hackers are continually changing their methods and tactics to outsmart traditional defenses, and intelligent and adaptive solutions must recognize known or unknown threats.

A "cyber threat intelligence" report details cyber risks, vulnerabilities, attack paths, operators, sources, and indicators of a breach as well as their analysis, interpretation, and distribution [3][4]. Threat intelligence allows organizations to recognize the typical signs of malicious activities and formulate defense strategies to reduce cyber risks before it disrupts operations [5]. Adaptive cyber security threat intelligence is an extension of intelligence systems that offers dynamic learning to dynamically re-calibrate defensive measures in response to changes in attack patterns and dynamic threat environments. Adaptive systems are continually monitored and automatically recognize patterns, analyze the context of the situation to make decisions, and anticipate and prevent risks.

The application of ML approaches has recently gained popularity in the cybersecurity industry as a means to address these issues [6][7]. The machine learning (ML) method of artificial intelligence has the ability to make computers much more autonomous in their data analysis, pattern recognition, and decision-making, hence significantly reducing the need for people in these processes. The ability of ML systems to learn from historical data and identify trends that might constitute security threats sets them apart from other systems [8].

The three primary types of ML approaches are supervised, unsupervised, and semi-supervised. The supervised learning approach uses pre-existing labels on data to teach a model how to categorize unknown data as either harmless or harmful [9]. Several supervised ML algorithms have been found to be highly effective at detecting intrusions based on the correlations found in

network traffic data, such as DT, RF, SVM, K-NN, LR, and NB. Unsupervised learning, on the other hand, which is adept at identifying structures in unlabeled data, is a good tool for anomaly detection. Combining the two methods, semi-supervised learning makes use of both sparsely labelled and abundantly unlabeled data.

A. Motivation and Contributions of the Study

This study is driven by the ever-evolving attacks that threaten to compromise traditional intrusion detection systems (IDSs). In addition to requiring continual upgrades, the conventional signature-based security approaches are unable to identify previously unseen assaults. Network traffic is always rising, which means there is a greater demand for advanced threat detection systems that can rapidly process large amounts of security data. Fortunately, ML can help by seeing trends in network traffic data that weren't there before.

The main findings from this research are as follows:

- Utilized the CICIDS2017 dataset containing multiple categories of modern cyberattacks and benign network traffic.
- Enhanced model performance by thorough data pretreatment, which included data cleaning, feature scaling, label encoding, and feature selection methods.
- Used a DT classifier to classify multiclass cybersecurity threats.
- ACC, PRE, REC, F1, confusion matrix, and ROC curve were the techniques used to evaluate the suggested model.
- Intelligence threat intelligence and intrusion detection using ML has been proven to be successful.

B. Justification and Novelty of Paper

The findings of this research support the use of ML techniques to improve the CTI in modern network environments. Unlike the rule-based intrusion detection schemes that rely on pre-defined attack signatures, the novel approach is based on a DT classifier that is able to learn discriminative patterns from network data automatically. Using the CICIDS2017 dataset, suitable preprocessing, and performance assessment, this one-of-a-kind work aims to build a multiclass intrusion detection system. The proposed system can accurately identify various types of cyberattacks.

1.1 Organization of the paper

The paper is structured in the following way: The Section II provides a literature review on relevant topics, such as machine learning techniques and cybersecurity threat intelligence. This paper details the dataset, preprocessing, classification model, and assessment criteria that make up the suggested technique in Section III. Section IV discusses the results of the experiments and the analysis of the performance. The paper concludes with Section V, which also lays out possible future research directions.

2. Literature Review

This section reviews recent ML and DL approaches for cybersecurity threat intelligence, highlighting datasets, methodologies, performance, and limitations.

A. T. Quanita and M. A. Y. Srizon, (2026), explores the use of Transformer-based models such as BERT, Roberta, SecBERT, and Distil BERT to efficiently extract deep contextual associations from unstructured cybersecurity text data. One of the key contributions is the development of an ensemble stacking model that opposes SecBERT and Distil BERT, leveraging their complementary abilities-Sebert's domain expertise and Distil Bert's efficiency. This ensemble performed vastly with an incredible ACC rate of 95.76% on the test set [10].

M. Kumar et al.(2025) Purposes to determine how AI might improve privacy, threat detection, and resilience. Using Neural Networks, CNNs, and Autoencoders on 20,000 cybersecurity events, the study detects attack patterns, classifies threats, and automates responses. CNNs achieved 97.6% PRE, while Autoencoders improved response ACC to 97.2%, reduced false positives, and AI-powered threat intelligence reduced data exfiltration risks by 90.4%[11].

A. T. Haile et al (2025) Presents a thorough methodology for the automatic categorization of cyber threats and the discovery of new threats utilizing data retrieved in real-time from the surface, deep, and black web. An SVM with the bag-of-words model achieved 93.67% and 96.35% ACC for binary and multiclass classification, while LDA effectively identified emerging cyber threat trends[12].

A. Gueriani et al. (2024) An integrated DL model using a CNN and a LSTM is included in the suggested IDS model. In a definitive testing step, the authors employed the CICIDS2017 dataset to validate the model's performance. For training and final testing, they utilized the new CICIoT2023 dataset to estimate the model's performance. Suggested model obtains a 98.42% ACC rate with a minimum loss of 0.0275. It is also crucial to note that the FPR is 9.17% with an F1score of 98.57% [13].

The present research, conducted by M. O. Adebiyi (2023) The goal of this study is to evaluate how well ML algorithms—DT, LR, RF, and NB—understand a cyber threat intelligence dataset consisting of 48,000 objects. RF

had the best ACC rate at 97.16%, followed by DT at 97.08%, NB at 93.92%, and LR at 80.15%, according to the experimental data [14].

A. O. Al-Ansari and T. M. Alsubait (2023) Explores cyber threat prediction for cyber supply chains using the Microsoft Malware Predictions dataset. Five ML algorithms were evaluated, with RF and Light GBM achieving the highest ACC of 72%, outperforming LR, SVM, and KNN[15].

Table I summarizes the reviewed studies, providing a comparative overview that identifies current research trends, existing challenges, and potential directions for developing more robust and intelligent cyber threat analysis frameworks.

TABLE I. Summary of the study on Machine Learning-Based Cybersecurity Threat Intelligence

Author (Year)	Approach	Dataset	Findings	Limitations	Recommendations
A. T. Quanita and M. A. Y. Srizon (2026)	Transformer models (BERT, RoBERTa, SecBERT, DistilBERT) with ensemble stacking	Unstructured cybersecurity text	Ensemble achieved 95.76% accuracy using SecBERT and DistilBERT	Limited evaluation on diverse real-world datasets	Improve generalization with larger and heterogeneous threat intelligence data
M. Kumar et al. (2025)	Neural Networks, CNNs, and Autoencoders	20,000 cybersecurity events	CNNs achieved 97.6% precision and Autoencoders improved response accuracy	Focused mainly on AI-based detection and response	Validate on real-time and large-scale cybersecurity environments
A. T. Haile et al. (2025)	SVM with Bag-of-Words and LDA	Surface, deep, and dark web data	Achieved 96.35% multiclass accuracy and detected emerging threats	Traditional feature representation limits contextual understanding	Integrate contextual language models for improved threat analysis
A. Gueriani et al. (2024)	CNN-LSTM based IDS	CICIoT2023 and CICIDS2017	Achieved 98.42% accuracy and 98.57% F1-score	Primarily designed for intrusion detection	Extend framework for broader cyber threat intelligence applications
M. O. Adebiyi (2023)	Decision Tree, Logistic Regression, Random Forest, Naïve Bayes	48,000-object cyber threat intelligence dataset	Random Forest achieved 97.16% accuracy	Limited comparison with deep learning models	Evaluate advanced deep learning and transformer-based approaches
A. O. Al-Ansari and T. M. Alsubait (2023)	Random Forest, LightGBM, SVM, Logistic Regression, KNN	Microsoft Malware Predictions dataset	Random Forest and LightGBM achieved 72% accuracy	Moderate prediction accuracy and limited feature representation	Employ advanced feature engineering and hybrid deep learning models

3. Methodology

The proposed methodology presents a systematic framework for developing a ML-based cybersecurity threat intelligence system using the DT classifier. Data gathering, preprocessing, model training, testing, and performance assessment make up the entire framework. The proposed intrusion detection model was educated and assessed with the use of the CICIDS-2017 dataset. The pre-processed dataset might be fairly evaluated by dividing it into training and testing subgroups. In subsequent instructions, the DT classifier was taught the basic patterns of malicious and normal network traffic in order to detect different kinds of intrusions. ACC, PRE, REC, F1, confusion matrix, and dependability in recognizing cybersecurity hazards were assessed lastly using traditional performance metrics on the trained model. Fig 1 shows the entire procedure of the suggested technique.

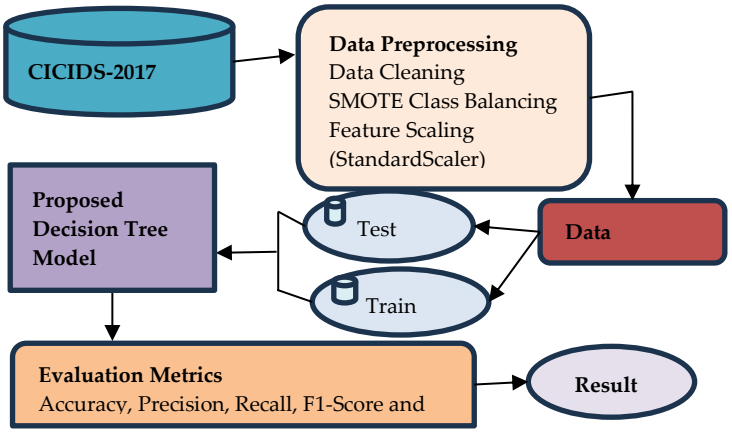

Fig.1. Proposed Methodology for Cybersecurity Threat Intelligence Using a Decision Tree Classifier

A. Data Collection

The proposed study use CICIDS-2017, a dataset developed by the Canadian Institute for Cyber Security (CIC) for flow-based intrusion detection, to address issues with earlier datasets [16]. With 2,830,743 samples across 15 categories and 78 characteristics, the dataset documented network activity for five days. There were 14 different kinds of attacks and 15 groups of benign incidents [17]. Initially, the eight CSV files of CICIDS-2017 were merged into a single dataset containing all network traffic records. There are almost 2.83 million records in the combined dataset, which represents both benign traffic and 14 types of attacks. The records have 79 properties, including 78 input features and one class label.

B. Data Preprocessing

The knowledge discovery process begins with data preparation. Data pre-processing adds extra work and time to the overall data analysis. Metadata often has several drawbacks, including duplicates, inconsistent information, noise, missing values, and inconsistent formatting. So, following learning techniques are weakened by low quality data [18]. The quality and reliability of subsequent automated judgments and discoveries can be greatly affected by appropriate preprocessing measures. As a subset of preprocessing, data preparation seeks to improve the quality of the main input so that it may be used appropriately in the mining process. Methods including normalization, transformation, cleansing, and integration are part of the obligatory preparation phase.

C. Data Cleaning

Data that had been combined were checked for any instances of Not-a-Number (NaN) entries, infinite (Inf) values, duplicate records, or missing values. All invalid records were removed to improve data consistency.

D. Handling Class Imbalance Using SMOTE

The reliability of the IDS was enhanced when the problem of class imbalance in the training dataset was resolved. Due to the much smaller number of attack instances compared to benign traffic instances, the SMOTE was employed to level the distribution of classes. In order to reduce majority-class bias in models and increase detection performance, SMOTE generates synthetic minority-class data from actual assault scenarios [19].

E. Feature Scaling Using Standardscaler

Data preparation for model construction follows data cleaning and data balancing, with feature scaling being the most important procedure [20]. Features having a wide numerical range can have an outsized impact on model outcomes and decision-making because nearly all ML algorithms are scale-invariant. By applying the Standard Scaler to the feature values, and can normalize them and solve for the missing value in (1):

$$z = \frac{(x - u)}{s} \quad (1)$$

F. Data Splitting

The original class distribution was preserved by randomly dividing the pre-processed dataset into two sections using stratified sampling. Data used to train the model was around 80% of the total, with 20% set aside for testing purposes. Training on a concealed testing dataset allowed for more objective performance evaluations.

G. Proposed Decision Tree Model

The supervised learning methods are frequently employed by DTs to address ML classification problems. Classification trees, another name for tree models, come into play when the target variable can take on discrete values. The parts that make up a DT are the nodes, branches, and leaves. Branches stand for the set of characteristics that contribute to the labels assigned to classes, and leaves for those classes. These subfields can handle both continuous and discrete data [21]. An essential splitter in the input determinants is used by the DT method to separate the samples into two or more homogenous groupings. When dealing with discrete data, DT performs well. Every internal decision node of the DT has the following entropy equation: (2):

$$Entropy, E\ (j)\text{-} = -\sum_{k=1}^{j} Pk\ log2\ Pk \ldots (2)$$

Pk is the probability of each class separately, and j is the total number of classes in the dataset [22]. The entropy equation can be expressed in the form of eq (3) for binary classification.

$$Entropy\ (8), E\ = -Pp\ log2\ Pp\ -\ Pn\ log2\ Pn \ldots (3)$$

where Pn represents the likelihood of a negative occurrence and Pp represents the likelihood of a positive event.

A schematic representation of the basic structure of the DT classifier is presented in Fig 2. Root, decision, branch, and leaf nodes make it up. The model categorizes the data in a sequence of decision rules applied from the root to the leaf, with the leaf value being the final class label.

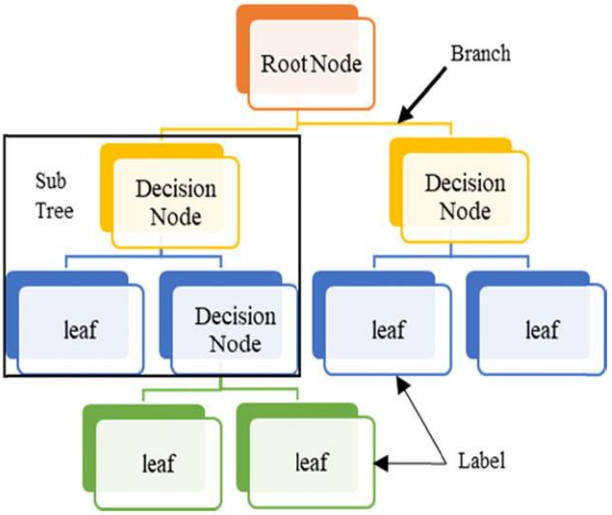

Fig.2. Decision Tree Architecture

H. Performance Evaluation Metrics

The confusion matrix is a matrix that shows how many flows are classified correctly or incorrectly. The categorization factors are:

- True positives, or TP, are the total number of assaults that were accurately categorized.
- TN stands for "true negatives," which is the total number of correctly diagnosed normal flows.
- The term "false positives" (FP) refers to the amount of real events that were mistakenly classified as attacks.
- Attacks that were mistakenly classified as normal are known as false negatives (FN).

1) Accuracy

The ACC of the predictions is the proportion of test instances where both the kind of attack (TP and TN) were correctly identified, as a percentage of the total number of occurrences.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \ldots (4)$$

2) Precision

The ratio of false positives to total positives is the definition of PRE in the context of positive examples [23]. Positive Predictive Value (PPV) is another name for this metric, which measures the expected likelihood of a successful positive prediction.

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

3) Recall

A classifier's sensitivity, REC, or detection rate (DR) is the probability that it correctly identify a positive example.

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

4) F-Measure:

There is just one weighted indicator that F-Measure uses, and it combines REC and ACC. F1 is constructed when both are given equal weight.

$$F1 = \frac{2\,TP}{2\,TP+FP+FN} \quad (7)$$

4. Result Analysis and Discussion

The proposed model of cyber security threat intelligence was tested in a controlled experimental environment to be able to estimate its effectiveness in detecting cyberattacks. The Scikit-learn ML toolkit, which provides a consistent and repeatable setting for developing and evaluating models, was utilized in these tests inside the Python programming language. The results of the experiments suggest that the suggested DT classifier can effectively differentiate between malicious and benign network data, obtaining good classification ACC with low FP. A detailed performance analysis is then given, including evaluation metrics, confusion matrix, ROC curve, and comparative analysis, in the following subsections.

The DT classifier was found to perform well in the detection of cybersecurity threats as presented in Table II. The proposed model was able to attain 99.91% ACC, which shows that the classifier was able to detect the almost entire normal and malicious traffic. Moreover, the model achieved a PRE of 97.79%, which shows that nearly all the predicted attack instances were correctly classified. The model has a high 97.08% REC which shows its efficiency in detecting real cyberattacks with minimal missed detections. Achieving an F1 score of 97.41% shows that the REC and ACC are well-balanced, demonstrating the effectiveness and reliability of the proposed classifier for attack detection.

TABLE II. Performance Evaluation of Decision Tree Classifier

Performance Metric	Value
Accuracy	99.91%
Precision	97.79%
Recall (Sensitivity)	97.08%
F1-Score	97.41%

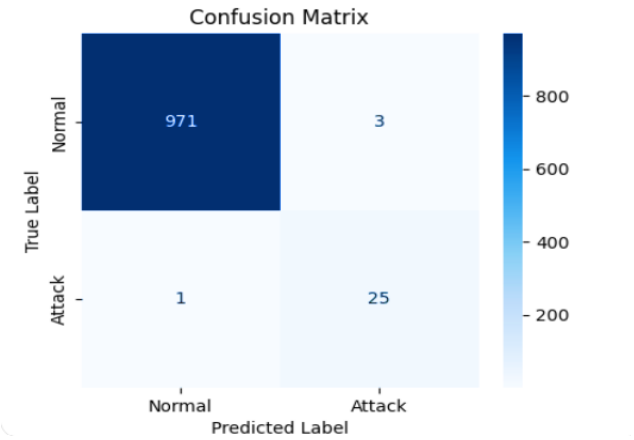

Fig.3. Confusion Matrix for Cyber Attack Classification

Fig. 3. shows the confusion matrix of the prediction model, which shows how successfully the DT model classified cybersecurity threats on the test data set. The anticipated class labels are shown in the columns, while the actual class labels are shown in the rows (Normal or Attack).

The model correctly classified 971 Normal instances and 25 Attack instances, which shows a good prediction ACC. The False Positives (FP) were 3 Normal instances, and False Negative (FN) were 1 Attack instance. The classifier successfully predicted 996 of the 1,000 samples and failed to predict 4 of them.

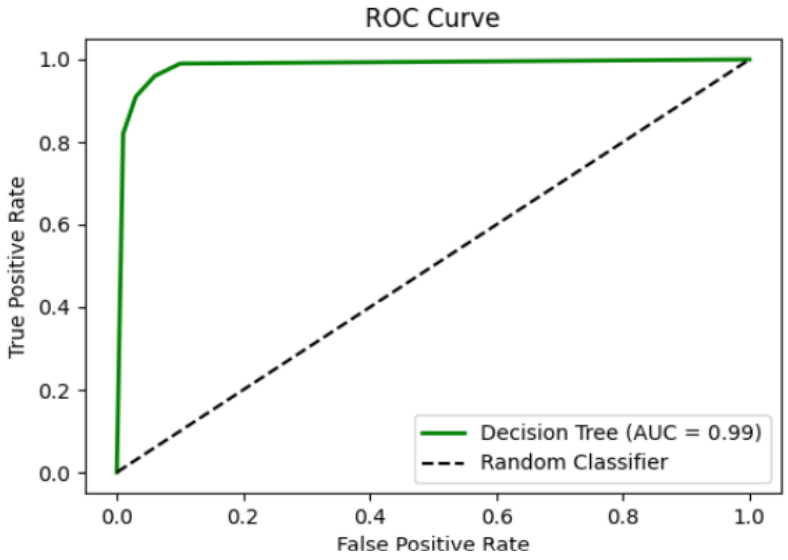

Fig.4. ROC Curve of the Proposed Decision Tree Classifier

The ROC curve of DT is shown in Fig.3. The AUC of the model was 0.99, which is considered to be excellent in classification tasks. The ROC curve is close to the upper left corner, reflecting a high TP rate and a very low FP rate at varying decision thresholds. The DT shows excellent discriminative ability, outperforming the random classifier (diagonal dashed line), which is especially effective for binary classification.

A. Comparative Analysis

The proposed DT classifier was tested on the benchmark intrusion detection datasets alongside other existing DL and ML methods, as shown in Table III. In comparison to CNN (96.50%), LR (82.00%), KNN (83.18%), LDA (76.32%), and PSO-LSTMIDS (81.02%), the suggested DT model attained a higher classification ACC of 99.91%. A high identification ACC and low false alarm rate in recognizing cyber threats have been demonstrated by the suggested model, which has obtained improved ACC and PRE 97.79%, REC 97.08%, and F1 97.41%.

TABLE III. Comparative Evaluation of Machine Learning Classification Techniques for Cybersecurity Threat Intelligence

Reference	Dataset	Model	Accuracy	Precision	Recall	F1-Score
Proposed Method	CICIDS2017	Decision Tree	99.91	97.79	97.08	97.41
[24]	CICIDS2017	CNN	96.50	93.16	90.10	91.60
[25]	CICIDS2017	Logistic regression	82.00	63.00	71.00	75.00
[26]	UNSW-NB15	KNN	83.18	79.15	94.30	86.06
[27]	NSL-KDD	LDA	76.32	92.38	63.65	75.37
[28]	BOT-IOT	PSO-LSTMIDS	81.02	74.11	81.36	78.14

The proposed DT based Classifier is also much easier to be trained and to make classification, which is more suitable for intrusion detection systems and resource constrained environments. After applying the SMOTE approach on the data, features, and class balance, the findings show that typical ML models are much outperformed. Experiments show that the proposed DT-based cybersecurity threat intelligence framework is effective in detecting intrusions in networks and improving cybersecurity defense mechanisms, and they also show that it is computationally simple and efficient.

5. Conclusion and Future Scope

ML has emerged as a potential solution to improve the true and timely determination of malicious network activities to aid in Cyber Security Threat Intelligence. In this work, a DT classifier was designed and trained using the CICIDS2017 dataset for the classification of benign and attack traffic. Data cleansing, feature scaling, and SMOTE-based class balancing were all parts of the extensive data processing that improved the model's ACC. Data cleaning, feature scaling, and SMOTE-based class balancing are all examples of thorough data processing approaches that have an effect on the model's performance. The proposed system detected the cyber threat optimally with 99.91% ACC, 97.79% PRE, 97.08% REC and 97.41% F1, which was beneficial for the detection of the cyber threat with minimum numbers of false alarms. The comparative analysis also showed that the proposed approach outperforms some other ML and DL approaches. The DT classifier, besides its high detection capability, is also good with regards to computational complexity and fast executions and therefore is suitable for real time intrusion detection and resource constrained environments.

Future work can include testing the proposed framework on additional recent and diverse cybersecurity datasets, such as CICIoT2023 and UNSW-NB15, to enhance its generalization ability. Advanced detection capabilities, adaptability and applicability for next generation intelligent cyber security defiance systems including ensemble and DL-based methods, explainable AI, cloud and IoT real-time deployment can also be enhanced.

REFERENCES

1. N. Mohamed, "Current trends in AI and ML for cybersecurity: A state-of-the-art survey," Cogent Eng., vol. 10, no. 2, pp. 1–30, Dec. 2023, doi: 10.1080/23311916.2023.2272358.
2. P. Kumar, "A Zero Trust-Based Approach to Modern Cybersecurity Challenges in Software Development," Int. J. Emerg. Res. Eng. Technol., vol. 6, no. 9, pp. 113–122, September, 2025.
3. S. C. -, "Risk Management in Advanced Persistent Threats (APTs) for Critical Infrastructure in the Utility Industry," Int. J. Multidiscip. Res., vol. 3, no. 4, pp. 1–10, July-August, Aug. 2021, doi: 10.36948/ijfmr.2021.v03i04.34396.
4. V. Koppireddy, D. G. Parasad, D. Patel, S. R. Somula, V. Kamaraj, and A. Meher, "AI Security Governance for Trustworthy Intelligent Systems: A Critical Analysis of Cyber Risks, Privacy Threats, and Responsible Deployment Frameworks," Int. J. Res. Trends Innov., vol. 11, no. 07, pp. a407–a435, July, 2026, doi: 10.56975/ijrti.v11i7.214029.
5. M. M. Kowsar, "Adaptive Cybersecurity Threat Intelligence Using Explainable Artificial Intelligence for Resilient Protection of U.S. Critical Information Systems," Am. ofAdvanced Technol. Eng. Solut., vol. 1, no. 2, pp. 216–261, 2025, doi: 10.2139/ssrn.5329985.
6. V. K. R. K. Koppireddy, "The Future of Cybersecurity: AI-Powered Innovations Redefining Digital Defense."
7. S. P. Sivashanmugam, S. Kumara, A. Mohile, and D. R. Suram, "Improving Detection Accuracy of Phishing Attacks with Feature Selection and Machine learning Techniques in Cybersecurity," in 2026 1st International Conference on Emerging

Trends in Advancements and Applications of Computational Intelligence Techniques (ETAACT), Bhubaneswar, India: IEEE, 2026, pp. 1–6, April. doi: 10.1109/ETAACT69135.2026.11542138.

8. M. Beyene and K. R. Shekar, "Performance Analysis of Homomorphic Cryptosystem on Data Security in Cloud Computing," in 2019 10th International Conference on Computing, Communication and Networking Technologies (ICCCNT), Kanpur, India: IEEE, 2019, pp. 1–7, July. doi: 10.1109/ICCCNT45670.2019.8944837.

9. M. T. Karatu, I. M. Mungadi, and A. Shehu, "Machine Learning Techniques for Cybersecurity Threat Detection : A Systematic Review," Int. J. Innov. Sci. Res. Technol., vol. 11, no. 3, pp. 2992–2998, 2026.

10. A. T. Quanita and M. A. Y. Srizon, "Enhancing Cyber Threat Intelligence Using Transformer Models for Text Classification," in 2026 IEEE 2nd International Conference on Quantum Photonics, Artificial Intelligence & Networking (QPAIN), 2026, pp. 1–5. doi: 10.1109/QPAIN69676.2026.11546420.

11. M. Kumar, M. Shojonov, B. Madaminov, A. K. Sahoo, T. Tiwari, and M. K. Sharma, "AI-Powered Cyber security: Neural Networks for Threat Detection and Prevention," in 2025 World Skills Conference on Universal Data Analytics and Sciences (WorldSUAS), 2025, pp. 1–5. doi: 10.1109/WorldSUAS66815.2025.11199107.

12. A. T. Haile, S. L. Abebe, and H. M. Melaku, "Real-Time Automated Cyber Threat Classification and Emerging Threat Detection Framework," IEEE Open J. Comput. Soc., vol. 6, pp. 921–930, 2025, doi: 10.1109/OJCS.2025.3580235.

13. A. Gueriani, H. Kheddar, and A. C. Mazari, "Enhancing IoT Security with CNN and LSTM-Based Intrusion Detection Systems," in 2024 6th International Conference on Pattern Analysis and Intelligent Systems (PAIS), 2024, pp. 1–7. doi: 10.1109/PAIS62114.2024.10541178.

14. M. O. Adebiyi, M. O. Ajayi, F. B. Osang, and A. A. Adebiyi, "A comparative study of selected machine learning algorithms for cyber threat detection in open source data," in 2023 International Conference on Science, Engineering and Business for Sustainable Development Goals (SEB-SDG), 2023, pp. 1–8. doi: 10.1109/SEB-SDG57117.2023.10124615.

15. A. O. Al-Ansari and T. M. Alsubait, "Predicting Cyber Threats Using Machine Learning for Improving Cyber Supply Chain Security," in 2022 Fifth National Conference of Saudi Computers Colleges (NCCC), 2022, pp. 123–130. doi: 10.1109/NCCC57165.2022.10067692.

16. D. Paul, S. A. D. B. N. Poovaiah, A. Kishore, V. S. K. Tankani, Meylikulov, and Shakhboz, "SHO-Xception: An Optimized Deep Learning Framework for Intelligent Intrusion Detection in Network Environments," in 2025 International Conference on Innovations in Intelligent Systems: Advancements in Computing, Communication, and Cybersecurity (ISAC3), Bhubaneswar, India: IEEE, 2025, pp. 1–6, July. doi: 10.1109/ISAC364032.2025.11156610.

17. C. H. N, "Network Intrusion dataset(CIC-IDS-2017)," Kaggle.

18. A. A. Abdualrahman and M. K. Ibrahem, "Intrusion Detection System Using Data Stream Classification," Iraqi J. Sci., vol. 62, no. 1, pp. 319–328, Jan. 2021, doi: 10.24996/ijs.2021.62.1.30.

19. H. R. Sayegh, W. Dong, and A. M. Al-madani, "Enhanced Intrusion Detection with LSTM-Based Model, Feature Selection, and SMOTE for Imbalanced Data," Appl. Sci., vol. 14, no. 2, p. 479, Jan. 2024, doi: 10.3390/app14020479.

20. A. Al Farsi, A. Khan, M. M. Bait-Suwailam, and M. R. Mughal, "Comparative Performance Evaluation of Machine Learning Algorithms for Cyber Intrusion Detection," Dec. 2024. doi: 10.20944/preprints202412.0497.v1.

21. Z. K. Maseer, R. Yusof, N. Bahaman, S. A. Mostafa, and C. F. M. Foozy, "Benchmarking of Machine Learning for Anomaly Based Intrusion Detection Systems in the CICIDS2017 Dataset," IEEE Access, vol. 9, pp. 22351–22370, 2021, doi: 10.1109/ACCESS.2021.3056614.

22. R. A. Disha and S. Waheed, "Performance analysis of machine learning models for intrusion detection system using Gini Impurity-based Weighted Random Forest (GIWRF) feature selection technique," Cybersecurity, vol. 5, no. 1, p. 1, 2022, doi: 10.1186/s42400-021-00103-8.

23. M. Rodríguez, Á. Alesanco, L. Mehavilla, and J. García, "Evaluation of Machine Learning Techniques for Traffic Flow-Based Intrusion Detection," Sensors, vol. 22, no. 23, p. 9326, Nov. 2022, doi: 10.3390/s22239326.

24. Z. Xu and Y. Liu, "Robust Anomaly Detection in Network Traffic: Evaluating Machine Learning Models on CICIDS2017," arXiv, 2025.

25. S. Abiramasundari and V. Ramaswamy, "Distributed denial-of-service (DDOS) attack detection using supervised machine learning

algorithms," Sci. Rep., vol. 15, no. 1, p. 13098, Apr. 2025, doi: 10.1038/s41598-024-84879-y.

26. S. M. Kasongo and Y. Sun, "Performance Analysis of Intrusion Detection Systems Using a Feature Selection Method on the UNSW-NB15 Dataset," J. Big Data, vol. 7, no. 1, 2020, doi: 10.1186/s40537-020-00379-6.

27. M. Arcos-Argudo, R. Bojorque, and A. Torres, "A Deterministic Comparison of Classical Machine Learning and Hybrid Deep Representation Models for Intrusion Detection on NSL-KDD and CICIDS2017," Algorithms, vol. 18, no. 12, p. 749, Nov. 2025, doi: 10.3390/a18120749.

28. N. Dash, S. Chakravarty, A. K. Rath, N. C. Giri, K. M. AboRas, and N. Gowtham, "An optimized LSTM-based deep learning model for anomaly network intrusion detection," Sci. Rep., vol. 15, no. 1, pp. 1–21, 2025, doi: 10.1038/s41598-025-85248-z.