IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 15, Issue 3, March 2026

DOI: 10.17148/IJARCCE.2026.15396

Network Intrusion Detection System Using Machine Learning

Anusree P S[1], Aryanandha Anil R[2], Fathima S[3], Gopika K[4] and Prof. Merlin K Thomas[5]

Undergraduate Research Paper, Department of Computer Science, College of Engineering Kottarakkara, Kollam, Kerala, India[1-4]

Assistant Professor, Department of Computer Science, College of Engineering Kottarakkara, Kollam, Kerala, India[5]

Abstract: With the rapid growth of Internet services and digital communication technologies, cyberattacks have become increasingly frequent, complex, and sophisticated. Modern computer networks are continuously exposed to various types of security threats that can compromise the confidentiality, integrity, and availability of data. Traditional security mechanisms such as firewalls and signature-based detection systems are widely used to protect networks; however, these methods are often ineffective in detecting unknown or evolving attack patterns. Therefore, more intelligent and adaptive security solutions are required to ensure effective protection of network infrastructures.

Intrusion Detection Systems (IDS) play an important role in identifying malicious activities by continuously monitoring network traffic and detecting abnormal behaviour. In recent years, machine learning techniques have gained significant attention in the field of cybersecurity due to their ability to analyse large volumes of data and identify hidden patterns associated with cyberattacks. This paper proposes a Machine Learning-based Network Intrusion Detection System designed to improve the detection of cyberattacks in network environments. The proposed system performs several processes including data preprocessing, feature selection, and classification to distinguish between normal and malicious network traffic. Various machine learning algorithms are applied to analyse network behaviour and detect attacks such as Denial of Service (DoS), Probe attacks, and unauthorized access attempts.

The proposed model is trained and evaluated using a standard intrusion detection dataset and implemented using Python. Experimental analysis demonstrates that the proposed approach improves detection accuracy while reducing false alarm rates. The results indicate that the system provides an efficient and reliable solution for enhancing network security and detecting cyber threats in modern computing environments.

Keywords: Intrusion Detection System, Machine Learning, Cyber Security, Network Security, Data Classification

I. INTRODUCTION

The rapid development of internet technologies, cloud computing, and Internet of Things (IoT) devices has significantly increased global network connectivity. Modern organizations rely heavily on networked systems for communication, data storage, and service delivery. While these technological advancements provide many benefits such as faster communication, efficient data sharing, and improved accessibility, they also expose computer networks to a wide range of cybersecurity threats. Cyberattacks such as Denial of Service (DoS), Distributed Denial of Service (DDoS), probing attacks, malware infections, and unauthorized access attempts can severely affect the confidentiality, integrity, and availability of network systems. Traditional security solutions such as firewalls and antivirus software mainly rely on predefined attack signatures to detect malicious activities. These signature-based approaches are effective in detecting known attacks but are often unable to identify new or previously unseen attack patterns. As cyber threats continue to evolve rapidly, attackers frequently modify their techniques to bypass traditional security mechanisms. Therefore, relying solely on conventional security tools is not sufficient to ensure complete network protection. To address these limitations, Intrusion Detection Systems (IDS) have been developed to monitor network traffic and detect suspicious behaviour within a network environment. IDS analyse network activities and identify abnormal patterns that may indicate potential cyberattacks. Intrusion detection systems can be broadly classified into signature-based detection and anomaly-based detection. Signature-based systems detect attacks by comparing network activity with known attack signatures, while anomaly-based systems detect deviations from normal network behaviour.

In recent years, machine learning techniques have gained significant attention in the field of cybersecurity due to their ability to automatically learn patterns from large volumes of data. Machine learning algorithms can analyse complex

DOI: 10.17148/IJARCCE.2026.15396

network traffic patterns and identify anomalies that may represent malicious activities. By learning from historical network data, these models can detect both known and unknown cyberattacks with improved accuracy and efficiency.

Machine learning-based intrusion detection systems offer several advantages, including improved detection accuracy, reduced false alarm rates, and the ability to adapt to evolving attack patterns. These systems can analyse large-scale network datasets and extract meaningful features that help distinguish between normal and malicious network activities. This paper proposes a machine learning based Network Intrusion Detection System that performs data preprocessing, feature selection, and classification to identify malicious network traffic. The proposed system aims to improve detection accuracy while reducing false alarm rates and enhancing overall network security. By integrating machine learning techniques with intrusion detection mechanisms, the system provides an intelligent approach for identifying cyber threats in modern network environments.

II. LITERATURE REVIEW

With the rapid growth of internet services and connected devices, network security has become a critical issue. Cyber-attacks such as Denial of Service (DoS), Distributed Denial of Service (DDoS), probing, and unauthorized access can severely affect data integrity, confidentiality, and system availability. Traditional security mechanisms such as firewalls and antivirus systems are not sufficient to detect new or evolving threats. Therefore, Intrusion Detection Systems (IDS) have been widely developed to monitor network traffic and identify malicious activities. Recently, machine learning and deep learning techniques have been applied to IDS to improve detection accuracy and efficiency. Several researchers have proposed different approaches and datasets for building effective intrusion detection models.

Denning (1987) [1] proposed one of the earliest intrusion detection models based on monitoring system activities and detecting anomalies. The model introduced the concept of analysing audit records to identify abnormal behaviour in computer systems. This work laid the foundation for modern intrusion detection systems.

Lee and Stolfo (1998) [2] introduced data mining techniques for intrusion detection. Their approach used machine learning algorithms to analyse network traffic data and identify patterns of malicious activities. The study demonstrated how data mining could improve the effectiveness of IDS by automatically learning attack patterns.

Vinayakumar et al. (2019) [3] proposed a deep learning-based intrusion detection system using a combination of Convolutional Neural Networks (CNN) and Bidirectional Long Short-Term Memory (BiLSTM). The model was designed to analyse network traffic and detect complex attack patterns with higher accuracy compared to traditional machine learning methods.

Lin et al. (2020) [4] developed a Generative Adversarial Network (GAN)-based one-class classifier for early intrusion detection. Their approach focused on detecting anomalies by learning the normal behaviour of network traffic and identifying deviations that indicate potential attacks.

Bhavsar et al. (2023) [5] proposed an anomaly-based intrusion detection system using PCC-CNN for IoT networks. Their method improved detection performance by extracting important features from network traffic and applying deep learning techniques to classify malicious activities.

Chen et al. (2023) [6] introduced a network intrusion detection approach using feature images and a deformable vision transformer model. The study converted network traffic features into images and applied advanced deep learning models to enhance detection accuracy.

Sharafaldin, Lashkari, and Ghorbani (2018) [7] developed the CICIDS2017 dataset, which contains realistic network traffic data and various attack scenarios. This dataset has been widely used for evaluating intrusion detection models.

Moustafa and Slay (2015) [8] introduced the UNSW-NB15 dataset, a comprehensive dataset designed to address the limitations of earlier datasets such as KDD Cup 99. The dataset includes modern attack types and realistic network traffic patterns.

Tavallaee et al. (2009) [9] performed a detailed analysis of the KDD Cup 99 dataset and identified several issues such as redundancy and imbalance. Their work helped researchers better understand dataset limitations and improve IDS evaluation methods.

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 15, Issue 3, March 2026

DOI: 10.17148/IJARCCE.2026.15396

III. PROPOSED METHODOLOGY

Modern computer networks face a wide range of cyber threats such as denial-of-service attacks, unauthorized access, malware infections, and data breaches. Traditional intrusion detection systems mainly rely on signature-based detection methods, which are effective only for known attack patterns. These systems often fail to detect new or unknown attacks and generate high false alarm rates. Therefore, an intelligent intrusion detection system that can automatically analyse network behaviour and detect anomalies is required. The proposed system is a Machine Learning–based Network Intrusion Detection System (IDS) designed to detect malicious activities in network traffic. The system uses machine learning techniques to analyse network data and classify traffic as either normal or malicious. By learning patterns from historical network data, the system can detect both known and previously unseen attacks.

The proposed intrusion detection framework consists of several stages including data collection, data preprocessing, feature selection, machine learning model training, and attack classification. These stages work together to provide accurate and efficient detection of cyber threats.

A. Data Collection:

The first component of the proposed system is data collection. Network traffic data is collected from publicly available intrusion detection datasets such as NSL-KDD or other benchmark datasets used for cybersecurity research. These datasets contain labelled network traffic records representing both normal activities and different types of cyberattacks. The collected data provides the foundation for training machine learning models and evaluating their performance in detecting malicious network behaviour.

B. Data Preprocessing:

Data preprocessing is an important step in preparing the dataset for machine learning analysis. Raw network traffic data may contain missing values, redundant records, or inconsistent data formats that can negatively affect model performance. In this stage, data cleaning techniques are applied to remove irrelevant or duplicate entries. Categorical features such as protocol type or service type are converted into numerical form using encoding techniques. Data normalization is also applied to ensure that all features are within a consistent range, improving the efficiency of machine learning algorithms.

C. Feature Selection:

Network intrusion datasets usually contain a large number of features, some of which may not contribute significantly to attack detection. Feature selection techniques are used to identify the most relevant attributes that influence classification performance. By selecting important features, the dimensionality of the dataset can be reduced while preserving meaningful information. This step helps improve model accuracy, reduce computational complexity, and enhance the overall efficiency of the intrusion detection system.

D. Machine Learning Model Training:

In this component, machine learning algorithms are trained using the processed dataset. Classification algorithms such as Decision Tree, Random Forest, and Support Vector Machine (SVM) can be applied to learn patterns in network traffic data. These algorithms analyse relationships between network traffic features and attack labels to build predictive models capable of identifying malicious activities. The trained models are evaluated using performance metrics such as accuracy, precision, recall, and F1-score.

E. Attack Detection and Alert Generation:

The final component of the proposed system is attacking detection and alert generation. Once the machine learning model is trained, it is used to analyse incoming network traffic in real time. The model evaluates traffic features and classifies them as either normal behaviour or malicious activity. If suspicious activity is detected, the system generates alerts to notify network administrators so that appropriate security actions can be taken. This enables early detection and prevention of cyberattacks before they cause significant damage to the network.

F. Hardware Requirements:

- Processor: Intel Core i5 / i7 or equivalent
- RAM: Minimum 8 GB (16 GB recommended for optimal performance)
- Storage: 500 GB HDD or SSD
- Network Adapter: Gigabit Ethernet
- Display: 1920 × 1080 resolution
- Peripheral Devices: Keyboard, mouse, and optional Wi-Fi module

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 15, Issue 3, March 2026

DOI: 10.17148/IJARCCE.2026.15396

G. Software Requirements:

- Operating System: Windows / Linux / macOS
- Programming Language: Python 3.x
- Development Environment: pycharm / Jupyter Notebook / VS Code
- Python Libraries

  Pandas – for reading and processing CSV dataset files

  NumPy – numerical computations

  Scikit-learn – machine learning algorithms (Decision Tree, Random Forest, SVM)

  Matplotlib / Seaborn – visualization of results such as accuracy graph and confusion matrix
- Dataset

  CSV files containing network traffic data used for training and testing the machine learning model.
- Real-Time Monitoring Tools

  Python modules for capturing or analysing real-time network traffic and checking it with the trained model to detect possible intrusions

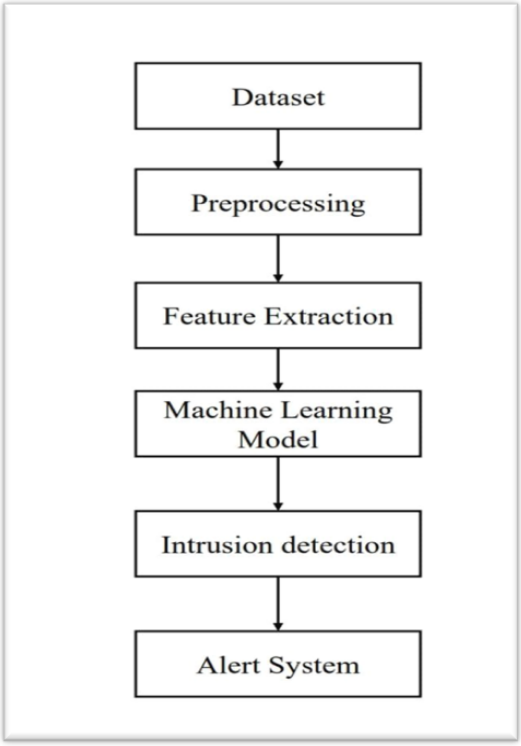

Fig 1. Block Diagram of Intrusion Detection and Alert Framework

IV. DATA FLOW DIAGRAM

A Data Flow Diagram (DFD) is used to visually represent how data moves through a system and how it is processed at different stages. In the proposed Network Intrusion Detection System (IDS), the DFD illustrates how network traffic data is collected, processed, analyzed using machine learning techniques, and how intrusion alerts are generated. It helps in understanding the interaction between different components of the system such as data preprocessing, feature extraction, and classification. The DFD is divided into multiple levels to represent the system with increasing detail.

A. Data Flow Diagram (Level 0)

Fig. 2.1 represent the Level 0 Context Diagram, provides a high-level overview of the entire intrusion detection system. In this level, the Network Intrusion Detection System (IDS) is represented as a single process that interacts with external entities. The system receives network traffic data as input from the network environment. This data may include packets, connection information, and communication patterns between devices in the network. The IDS analyzes this incoming traffic to identify suspicious or malicious activities. After processing the data, the system generates alerts or notifications whenever an intrusion or abnormal behavior is detected. These alerts are then sent to the admin or user, allowing them to

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 15, Issue 3, March 2026

DOI: 10.17148/IJARCCE.2026.15396

take appropriate action to secure the network. Thus, the Level 0 diagram mainly shows the overall interaction between the network traffic source, the intrusion detection system, and the administrator or user who receives the alerts.

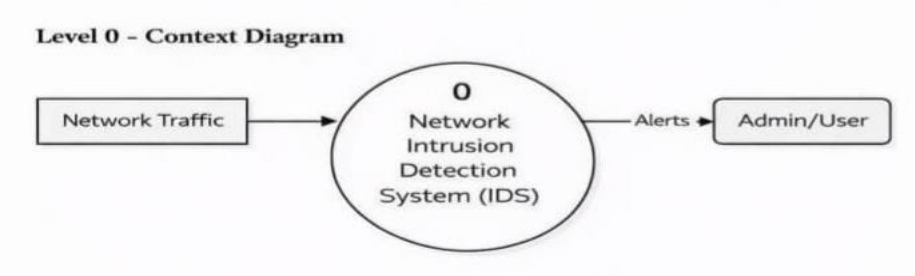

Fig. 2.1 Level 0 Data Flow Diagram of the Proposed IDS

B. Data Flow Diagram (Level 1)

Fig 2.2 represent the Level 1 Data Flow Diagram provides a detailed representation of the internal processes involved in the Network Intrusion Detection System. Initially, the system receives a network traffic dataset, which contains information collected from network packets and system logs. This dataset forms the input for the intrusion detection process.

1. Data Preprocessing**:**

   The first stage of the system is data preprocessing. In this step, the raw network data is cleaned and prepared for analysis. Preprocessing includes tasks such as removing missing values, eliminating redundant records, and normalizing the dataset. These operations help improve the quality of the data and reduce noise, which ultimately enhances the performance of the machine learning model.

2. Feature Extraction**:**
   After preprocessing, the system performs feature extraction. In this step, important attributes are selected from the dataset that represent the behavior of network traffic. These features may include packet size, protocol type, connection duration, and number of failed login attempts. Feature selection helps reduce the dimensionality of the dataset and removes irrelevant features, improving computational efficiency and detection accuracy.

3. Machine Learning Classification**:**

   The extracted features are then passed to the machine learning classification module. In this stage, a trained machine learning model analyzes the network traffic patterns and classifies them into two main categories:

   - Normal network activity
   - Intrusion or malicious activity

   The classification model learns patterns from the dataset and identifies unusual behavior that may indicate cyberattacks.

4. Alerts and Logs**:**

   Finally, the results generated by the classifier are stored in the alerts and logs database. If malicious activity is detected, the system generates alerts that are sent to the system administrator. These alerts allow administrators to monitor network security and respond quickly to potential threats.

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 15, Issue 3, March 2026

DOI: 10.17148/IJARCCE.2026.15396

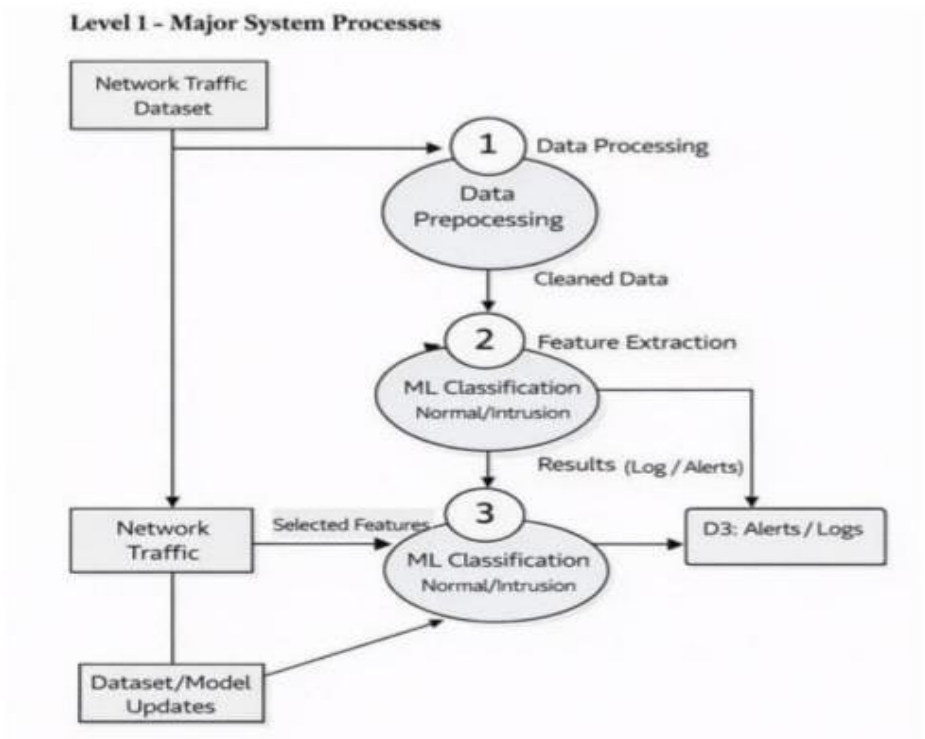

Fig.2.2 Level 1 Data Flow Diagram of the Proposed IDS

V. RESULTS AND DISCUSSION

The proposed machine learning-based Network Intrusion Detection System was evaluated using a standard intrusion detection dataset. The model was trained to analyse network traffic patterns and classify them into normal or malicious activities. The system achieved effective intrusion detection performance by applying preprocessing techniques, feature selection methods, and machine learning classification algorithms. Experimental results demonstrate that the model can successfully detect various cyber-attack patterns while maintaining a low false alarm rate.

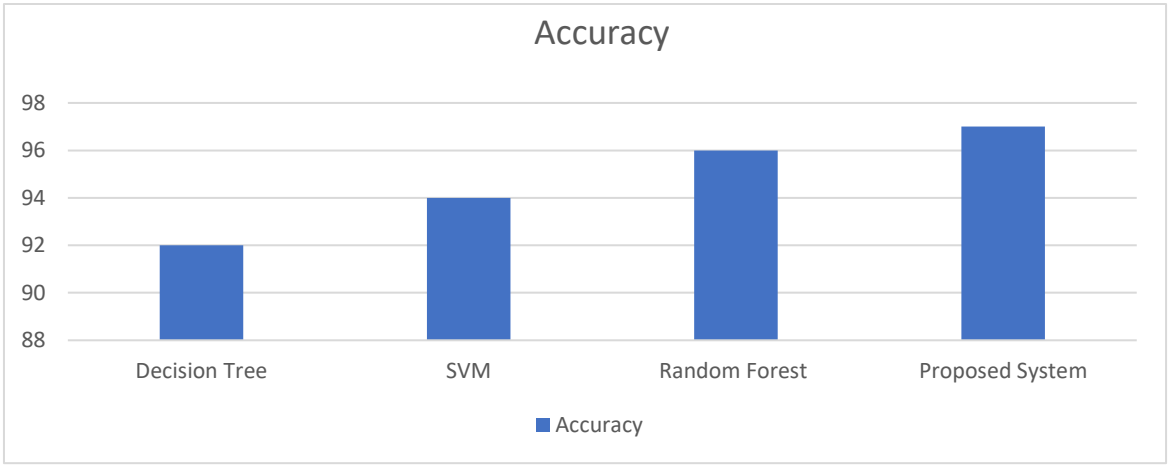

Fig.3.1 Accuracy Comparison Graph

Fig 3.1 illustrates the performance of the Proposed System against three standard machine learning baseline models: Decision Tree, SVM, and Random Forest. In Performance Analysis, Decision Tree shows the lowest accuracy, hovering around 92%. While fast, it often struggles with complex, high-dimensional network data. SVM (Support Vector Machine) shows an improvement at approximately 94%. Random Forest: A strong ensemble performer reaching roughly 95.5% accuracy by combining multiple decision tree. Proposed System outperforms all baseline models with a peak accuracy of approximately 96%.

International Journal of Advanced Research in Computer and Communication Engineering
Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 15, Issue 3, March 2026
DOI: 10.17148/IJARCCE.2026.15396

Predicted Attack	Predicted Normal	
FP	TN	Actual Normal
TP	FN	Actual Attack

Fig.3.2 Confusion Matrix

Fig 3.2 shows the Confusion Matrix is a critical tool for evaluating the performance of a classification model, specifically for identifying how often the system confuses two classes: Attack and Normal. For an Intrusion Detection System (IDS), the goal is to maximize TP and TN while keeping FN as close to zero as possible to ensure no threats bypass the system.

To evaluate the effectiveness of the model, several performance metrics were used, including:

- Accuracy
- Precision
- Recall
- F1-score
- False Positive Rate

These evaluation metrics help measure how accurately the system detects attacks and how effectively it distinguishes between normal and malicious traffic.

VI. CONCLUSION

The increasing dependence on computer networks and internet-based services has made network security a major concern in modern information systems. Traditional security mechanisms such as firewalls alone are not sufficient to detect sophisticated cyber-attacks. Therefore, an effective Intrusion Detection System (IDS) is essential to monitor network activities and identify malicious behaviour in real time. In this a machine learning-based Network Intrusion Detection System was developed to analyse network traffic and detect potential cyber threats.

The proposed system processes network traffic data through several stages including data preprocessing, feature extraction, and machine learning classification. These steps help in preparing the dataset, selecting relevant attributes, and accurately identifying patterns associated with normal and malicious activities. The preprocessing stage improves the quality of the dataset by removing redundant and irrelevant data, while feature extraction helps reduce dimensionality and focuses on the most significant features that influence intrusion detection. The processed data is then passed to the classification module, where machine learning algorithms analyse the traffic patterns and classify them into normal behaviour or intrusion activity. The experimental results demonstrate that machine learning techniques significantly enhance the detection capability of intrusion detection systems. The system is able to identify various types of attacks while maintaining good accuracy and minimizing false alarms. By continuously monitoring network traffic and generating alerts, the system helps administrators take timely action against suspicious activities.

Overall, the proposed intrusion detection model provides an efficient approach for improving network security and protecting sensitive information from cyber threats. It highlights the importance of applying intelligent data analysis techniques in cybersecurity applications. In the future, the system can be further enhanced by incorporating advanced machine learning or deep learning algorithms, improving feature selection techniques, and implementing real-time intrusion detection for large-scale networks. These improvements can make the IDS more scalable, accurate, and capable of handling complex and evolving cyberattacks.

REFERENCES

[1]. XE-AI: Evaluating Black-Box Explainable AI Frameworks for Network Intrusion Detection, A. Arreche et al.,2024.
[2]. SecIDS-CNN-WF: A Trust-Aware Edge-Efficient CNN for Real-Time Wi-Fi Intrusion Detection A. Almalkawi et al., 2024.
[3]. An Anomaly-Based Intrusion Detection System Using PCC-CNN for IoT Networks Scientific Reports, A. Bhavsar et al., 2023.
[4]. Machine Learning and Deep Learning Techniques for Network Intrusion Detection: A Survey- Discover Artificial Intelligence, A. Arreche et al., 2023.
[5]. Anomaly-Based Intrusion Detection for IoMT Networks: Design and Evaluation, A. Alrashdi et al.,2022.
[6]. Optimization of Network Intrusion Detection Model Based on Big Data Analysis, Y. Zhang et al., 2021.

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 15, Issue 3, March 2026

DOI: 10.17148/IJARCCE.2026.15396

[7]. Early Detection of Network Intrusions Using a GAN-Based One-Class Classifier-Journal of Information Security, W. Lin et al., 2020.
[8]. An Intrusion Detection Model Based on SMOTE and Convolutional Neural Network Ensemble, T.Tian and Y. Lu, 2020.
[9]. Deep Learning Approach for Network Intrusion Detection Using CNN-BiLSTM, R. Vinayakumar et al., 2019.
[10]. An Evolutionary Computation Based Feature Selection Method for Intrusion Detection- Security and Communication Networks, Y. Xue et al., 2018.
[11]. A Hybrid Intrusion Detection Method Integrating Anomaly Detection with Misuse Detection - Computers & Security, Elsevier, G. Kim, S. Lee, and S. Kim, 2014.
[12]. Data Mining Approaches for Intrusion Detection-, USENIX Security Symposium, W. Lee and S. J. Stolfo,1998.

