International Journal of Recent Development in Engineering and Technology
Website: www.ijrdet.com (ISSN 2347 - 6435 (Online) Volume 14, Issue 8, August 2025)

Review of Machine Learning Approachs for Network Intrusion Detection Systems using Image Data

[1]Km. Esha Singh, [2]Prof. (Dr.) Vishal Kohli
[1]Research Scholar, Department of CSE, Neelkanth Institute of Technology, Meerut, India
[2]Director, Department of CSE, Neelkanth Institute of Technology, Meerut, India

***Abstract*— As cyber threats continue to evolve in complexity and frequency, safeguarding network infrastructures has become a critical concern for organizations worldwide. Traditional Network Intrusion Detection Systems (NIDS) primarily rely on signature-based or statistical anomaly detection methods, which often struggle to detect novel or sophisticated attacks. Recently, machine learning (ML) techniques have emerged as powerful tools for enhancing intrusion detection capabilities, offering adaptability and improved performance over traditional methods. A particularly innovative trend is the transformation of network traffic data into image representations, enabling the application of advanced image-based ML models, including deep learning techniques such as Convolutional Neural Networks (CNNs). This review systematically explores the current landscape of machine learning approaches that leverage image data for network intrusion detection. We discuss various methods of converting network traffic into image form, the ML models applied, and the performance benchmarks achieved. Furthermore, we identify challenges such as feature representation, scalability, dataset limitations, and real-time deployment issues. Finally, we highlight promising directions for future research, advocating for more robust, interpretable, and generalizable NIDS based on image analysis.**

***Keywords*— ML, NIDS, CNN, Image, DL.**

I. INTRODUCTION

In the digital era, the ubiquity of internet-connected systems and the rapid expansion of networked services have dramatically increased the vulnerability of networks to a wide range of cyber-attacks. From data breaches and ransomware attacks to sophisticated Advanced Persistent Threats (APTs), the growing sophistication and frequency of cyber threats necessitate the continuous advancement of network security measures [1]. Among these measures, Network Intrusion Detection Systems (NIDS) play a crucial role by monitoring network traffic for suspicious activities and unauthorized access attempts. However, traditional intrusion detection approaches—primarily signature-based detection and anomaly-based statistical techniques—face significant challenges in adapting to the dynamic and evolving landscape of cyber threats [2].

Signature-based systems, while highly effective against known attacks, falter when encountering zero-day exploits or slightly modified attack vectors. Meanwhile, statistical anomaly detection methods, although capable of identifying novel attacks, often suffer from high false positive rates, leading to alert fatigue among security analysts [3]. To overcome these limitations, researchers have increasingly turned towards machine learning (ML) techniques, which offer the ability to learn complex patterns from data and adapt to previously unseen attack types [4].

In recent years, an innovative paradigm has emerged: the use of image-based representations of network traffic data for intrusion detection. The fundamental idea is to convert structured network traffic features into two-dimensional or even three-dimensional image formats, allowing the exploitation of the powerful capabilities of image-oriented ML models, especially deep learning architectures such as Convolutional Neural Networks (CNNs), Residual Networks (ResNets), and Vision Transformers (ViTs) [5]. This transformation leverages the success of computer vision methodologies in extracting intricate spatial patterns and relationships within image data, applying these strengths to the domain of cybersecurity[6].

International Journal of Recent Development in Engineering and Technology
Website: www.ijrdet.com (ISSN 2347 - 6435 (Online) Volume 14, Issue 8, August 2025)

The rationale behind converting network traffic into images stems from the realization that certain patterns indicative of malicious behavior may be more easily captured through spatial relationships in an image than through raw numerical or categorical data [7]. For example, sequences of bytes, packet header fields, or even flow-level statistics can be mapped into pixel values, creating visual textures and structures that are characteristic of different types of network activities, whether benign or malicious. This method opens up new opportunities for using pre-trained vision models, transfer learning, and novel deep learning architectures in the cybersecurity space[8].

Several methods have been proposed to accomplish the data transformation process. Some approaches directly map packet-level or flow-level features into grayscale or color images, while others employ more sophisticated techniques such as time-series imaging, Gramian Angular Fields (GAF), or Markov Transition Fields (MTF) to capture temporal dynamics and correlations. Once the image representations are generated, a wide array of machine learning models can be trained to distinguish between normal and malicious traffic patterns. The application of CNNs, in particular, has shown great promise due to their inherent ability to detect local patterns and hierarchical features within images[9].

Despite these advancements, significant challenges remain. The high dimensionality of network traffic data, the risk of overfitting, the scarcity of labeled attack datasets, and the need for real-time performance all pose considerable hurdles to the widespread adoption of image-based ML techniques for NIDS. Moreover, the interpretability of deep learning models—often considered 'black boxes'—is a critical concern in security applications where explainability is vital[10].

This review aims to provide a comprehensive survey of the current state-of-the-art machine learning approaches that utilize image data for network intrusion detection. We systematically categorize the different strategies used for data transformation, model training, and evaluation. In addition, we critically assess the strengths and weaknesses of existing studies, identify research gaps, and suggest potential avenues for future exploration. Our goal is to offer valuable insights for researchers and practitioners seeking to enhance NIDS capabilities through the innovative integration of image-based machine learning techniques.

II. LITERATURE SURVEY

In 2024, Chen et al. introduced VGGIncepNet, a novel model that transforms non-image network traffic data into image representations to leverage the feature extraction capabilities of convolutional neural networks (CNNs). By integrating VGG16 and Inception modules, the model achieved superior performance on NSL-KDD and CICIoT2023 datasets, outperforming traditional models like BERT and XLNet in accuracy, precision, recall, and F1-score. This approach underscores the potential of combining non-image-to-image conversion techniques with deep learning architectures for effective intrusion detection[1].

A recent study proposed a Self-Supervised Intrusion Detection (SSID) framework that enables fully online deep learning-based intrusion detection without human intervention or prior offline learning. Utilizing an Auto-Associative Deep Random Neural Network, the system analyzes and labels incoming traffic based on its own decisions and statistical trustworthiness estimates. Experimental evaluations demonstrated the framework's adaptability to time-varying network traffic characteristics, making it a promising solution for real-time intrusion detection in IoT systems[2].

Kim and Pak presented an optimized method for processing NIDS datasets by converting them into two-dimensional images using various image transformers and integrating them into three-channel RGB images. This approach significantly improved intrusion detection performance compared to methods using grayscale images or non-image data. The study highlights the effectiveness of applying vision-based deep learning models to NIDS by leveraging color image representations[3].

Talukder et al. proposed a hybrid model combining machine learning and deep learning techniques to enhance detection rates and dependability in NIDS. The model employs SMOTE for data balancing and XGBoost for feature selection, achieving remarkable accuracy on KDDCUP'99 and CIC-MalMem-2022 datasets. This study emphasizes the importance of integrating multiple learning approaches to address the challenges of large and complex network data[4].

International Journal of Recent Development in Engineering and Technology
Website: www.ijrdet.com (ISSN 2347 - 6435 (Online) Volume 14, Issue 8, August 2025)

Mane and Rao addressed the interpretability challenges of deep learning models in NIDS by incorporating explainable AI (XAI) techniques such as SHAP, LIME, and ProtoDash. Applying these methods to the NSL-KDD dataset, the study provided insights into feature influences on model predictions, enhancing transparency and trust in automated intrusion detection systems. This work underscores the necessity of explainability in deploying deep learning models for cybersecurity applications[5].

Wu, Guo, and Buckland explored the application of transfer learning in NIDS by developing a model consisting of two concatenated convolutional neural networks (ConvNets). The model first learns from a base dataset and then transfers the acquired knowledge to a target dataset, improving detection accuracy on both known and novel attacks. This approach demonstrates the potential of transfer learning in enhancing the generalization capabilities of intrusion detection models[6].

Zhang et al. proposed a framework that combines deep adversarial learning with statistical learning to address data scarcity and imbalance in intrusion detection. The model generates synthetic intrusion data using a Poisson-Gamma joint probabilistic generative model and augments it through adversarial learning, improving the training of supervised classifiers. Experiments on the KDD Cup 99 dataset showed enhanced accuracy, precision, recall, and F1-score compared to existing IDS models[7].

Research in 2018 focused on hybrid models combining convolutional neural networks (CNNs) and recurrent neural networks (RNNs) to capture both spatial and temporal features in network traffic data. These models demonstrated improved detection rates and reduced false positives, highlighting the effectiveness of integrating different deep learning architectures for comprehensive intrusion detection[8].

In 2017, studies explored the direct application of CNNs to NIDS by converting network traffic features into image-like representations. These approaches leveraged the pattern recognition capabilities of CNNs to identify anomalies in network behavior, laying the groundwork for subsequent research in image-based intrusion detection[9].

Early research in 2015 investigated the feasibility of representing network traffic data as images for intrusion detection purposes. These pioneering studies set the stage for the integration of computer vision techniques in cybersecurity, inspiring a new direction in NIDS research that combines image processing with machine learning [10].

III. CHALLENGES

Despite the promising results achieved by machine learning approaches leveraging image data for network intrusion detection systems (NIDS), several significant challenges must be addressed to fully realize their potential in practical and large-scale deployments. These challenges span across data preparation, model performance, real-world applicability, and interpretability.

1. Data Transformation and Representation Complexity

Transforming raw network traffic into meaningful image representations is non-trivial. Choosing an effective mapping strategy that preserves critical features related to network behavior while enabling meaningful visual patterns is challenging. Poor transformations may obscure important anomalies or introduce noise that confuses the model. Additionally, there is no standardized method for converting network data into images, making it difficult to compare results across different studies.

2. High Dimensionality and Resource Consumption

Converting network traffic to images often results in high-dimensional data, especially when working with large packet captures or flow data. Processing these images requires substantial computational resources in terms of memory, processing power, and storage. Training deep learning models on such high-dimensional image data can be time-consuming and resource-intensive, posing difficulties for real-time deployment in production environments.

International Journal of Recent Development in Engineering and Technology
Website: www.ijrdet.com (ISSN 2347 - 6435 (Online) Volume 14, Issue 8, August 2025)

3. Imbalanced and Insufficient Datasets

Intrusion detection datasets often suffer from class imbalance, where malicious traffic is vastly outnumbered by benign traffic. This imbalance is further magnified in image datasets created from network flows. Additionally, publicly available datasets like NSL-KDD, CICIDS2017, and others may not accurately represent current threat landscapes or network traffic patterns, limiting the generalizability of models trained on them.

4. Overfitting and Generalization Issues

Deep learning models, particularly those dealing with complex image data, are prone to overfitting, especially when trained on small or unbalanced datasets. Overfitted models perform well on training data but fail to generalize to unseen traffic patterns or new types of attacks. Developing models that can maintain high detection accuracy across varied network environments remains a major hurdle.

5. Real-Time Detection and Scalability

In real-world networks, intrusion detection systems must operate in real-time to prevent attacks proactively. However, processing and analyzing high-dimensional image data in real-time is computationally expensive. The latency introduced by data transformation, image generation, and deep learning inference can make it impractical for real-time or near-real-time applications without significant optimization.

6. Interpretability and Trust

Security applications require not only accurate predictions but also explainable outcomes. Deep learning models, particularly CNNs and other complex architectures, are often viewed as "black boxes," making it difficult for cybersecurity professionals to understand or trust their decisions. This lack of transparency hampers adoption, as security analysts need clear justifications for alerts and automated actions.

7. Robustness Against Adversarial Attacks

Adversarial examples—specially crafted inputs designed to fool machine learning models—pose a serious threat to image-based NIDS. Small perturbations to the input data (or its image representation) could cause misclassification, allowing attackers to evade detection. Building robust models that can resist adversarial manipulation is crucial but remains an open research problem.

8. Dataset Standardization and Benchmarking

The lack of standardized datasets and evaluation metrics for image-based NIDS impedes progress in the field. Current research often uses different datasets, preprocessing techniques, and evaluation criteria, making it difficult to perform fair comparisons. Establishing benchmark datasets and evaluation frameworks is necessary to measure true advancements in the field.

IV. PROPOSED STRATEGY

To address the identified challenges and to advance the effectiveness of machine learning approaches for network intrusion detection systems (NIDS) using image data, a comprehensive and systematic strategy is proposed. This strategy integrates improvements at multiple stages—ranging from data preprocessing and image generation to model design, training, evaluation, and real-world deployment.

1. Standardized and Optimized Data Transformation Techniques

A critical first step is the development and adoption of standardized methods for converting network traffic into images. Instead of arbitrary mappings, domain-specific transformations should be designed that preserve crucial spatial and temporal features of network behavior. For example, techniques such as Gramian Angular Fields (GAF), Recurrence Plots (RP), or Time-Frequency Representations could be leveraged to create consistent, information-rich images. Benchmarking different transformation techniques systematically would ensure the selection of the most effective representation for various types of network attacks.

2. Dimensionality Reduction and Efficient Image Encoding

To combat high-dimensionality issues, dimensionality reduction techniques such as Principal Component Analysis (PCA), t-Distributed Stochastic Neighbor Embedding (t-SNE), or Autoencoders can be applied before the image generation stage. Moreover, utilizing compact image formats and focusing on key features instead of all packet fields will reduce the computational burden, making real-time deployment more feasible.

3. Data Augmentation and Synthetic Data Generation

To address dataset imbalance and insufficiency, extensive use of data augmentation techniques should be employed. Augmentations like rotation, flipping, and cropping can increase dataset diversity without compromising information integrity. Additionally, Generative Adversarial Networks (GANs) can be utilized to create synthetic but realistic network traffic images, helping balance minority attack classes and enhancing model generalization.

4. Hybrid and Ensemble Deep Learning Models

Rather than relying on a single deep learning architecture, hybrid models that combine Convolutional Neural Networks (CNNs) for spatial feature extraction and Recurrent Neural Networks (RNNs) or Transformers for temporal sequence analysis should be adopted. Ensemble strategies that aggregate predictions from multiple diverse models can significantly boost robustness, detection rates, and resilience against overfitting.

5. Explainable and Interpretable Models

Given the critical need for transparency in cybersecurity applications, integrating Explainable AI (XAI) frameworks with NIDS is vital. Techniques such as Local Interpretable Model-Agnostic Explanations (LIME), SHapley Additive exPlanations (SHAP), or Saliency Maps should be incorporated to highlight which parts of the input images contributed to the intrusion detection decision. This will not only foster trust among security analysts but also help in refining models based on human feedback.

7. Adversarial Robustness Mechanisms

To safeguard against adversarial attacks, adversarial training should be incorporated into the model training pipeline. Additionally, defensive distillation and input pre-processing techniques like randomization or denoising can be employed to make models more robust against adversarial manipulations designed to evade detection.

V. CONCLUSION

Machine learning approaches that leverage image data for network intrusion detection systems (NIDS) have shown remarkable potential in enhancing detection accuracy and addressing sophisticated cyber threats. By transforming network traffic into visual representations, these systems can exploit the powerful feature extraction capabilities of deep learning models, particularly convolutional architectures. However, challenges such as data transformation complexity, resource demands, model interpretability, and real-world scalability must be carefully addressed to fully realize their benefits. Through standardized data processing, hybrid model designs, explainable AI integration, and robust real-time deployment strategies, the future of image-based NIDS appears promising. Continued research and collaboration are essential to bridge existing gaps, ensuring that these systems evolve into practical, trustworthy, and scalable solutions for next-generation cybersecurity defenses.

REFERENCES

1. J. Chen, J. Xiao, and J. Xu, "VGGIncepNet: Enhancing Network Intrusion Detection and Network Security through Non-Image-to-Image Conversion and Deep Learning," Electronics, vol. 13, no. 18, p. 3639, 2024. (MDPI)

2. E. Caville, W. W. Lo, S. Layeghy, and M. Portmann, "Online Self-Supervised Deep Learning for Intrusion Detection Systems," IEEE Transactions on Information Forensics and Security, vol. 19, pp. 1-12, 2024. (IEEE)

3. T. Kim and W. Pak, "Deep Learning-Based Network Intrusion Detection Using Multiple Image Transformers," Applied Sciences, vol. 13, no. 5, p. 2754, 2023. (MDPI)

4. M. A. Talukder, M. R. Rahman, A. Alzubi, and S. R. Islam, "A Dependable Hybrid Machine Learning Model for Network Intrusion Detection," arXiv preprint arXiv:2212.04546, 2022.

5. S. Mane and D. Rao, "Explaining Network Intrusion Detection System Using Explainable AI Framework," arXiv preprint arXiv:2103.07110, 2021.

6. P. Wu, H. Guo, and R. Buckland, "A Transfer Learning Approach for Network Intrusion Detection," arXiv preprint arXiv:1909.02352, 2019.

7. H. Zhang, L. Sun, Y. Huang, and X. He, "Deep Adversarial Learning in Intrusion Detection: A Data Augmentation Enhanced Framework," arXiv preprint arXiv:1901.07949, 2019.

8. A. Kumar, S. Patel, and Y. Zhao, "Hybrid CNN-RNN Models for Intrusion Detection," Proceedings of the IEEE International Conference on Big Data Security on Cloud (BigDataSecurity), pp. 120-125, 2018.

9. B. Li, K. Wang, and J. Zhou, "Application of CNNs in Network Intrusion Detection Systems," Proceedings of the IEEE International Conference on Computer and Communications (ICCC), pp. 1937-1941, 2017.

10. L. Zhang and X. Li, "Early Exploration of Image-Based Network Intrusion Detection Systems," Proceedings of the International Conference on Information Security and Privacy Protection (IFIP SEC), pp. 67-78, 2015.