International Journal of Applied Mathematics
Volume 38 No. 12s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

NETWORK INTRUSION DETECTION PREDICTIVE ANALYSIS USING MACHINE LEARNING ALGORITHMS

Dr. SATTI SUDHA MOHAN REDDY[1],Dr. Darshan B D[2] & Manoj I. Patel[3]

[1]Professor, Department of ECE, SRKR Engineering College, Bhimavaram Andhra Pradesh, India

[2]Assistant Professor, Electronics & Communication Engineering, SJB Institute of Technology, Bangalore, Karnataka, India

[3]Assistant Professor, Faculty of Computer Science and application, Gokul Global University, Sidhpur, Gujarat, India

Abstract:-

The exponential expansion of digital connectivity has intensified the vulnerabilities of networked systems, making intrusion detection a critical component of modern cybersecurity infrastructure. Conventional rule-based intrusion detection systems, though historically effective, struggle to cope with the sophistication, diversity, and velocity of contemporary cyberattacks. This research examines how machine learning–driven predictive models can enhance the detection, classification, and early identification of malicious network behavior. By exploring multiple supervised, unsupervised, and ensemble-learning approaches, the study demonstrates how data-centric techniques can adapt to evolving threat landscapes and offer measurable improvements in detection accuracy and response agility. The research employs a combination of benchmark datasets and real-time traffic traces to capture the complexity of network behavior across benign and attack scenarios. Feature engineering is performed through layered preprocessing steps, including noise filtration, correlation analysis, dimensionality reduction, and protocol-specific feature extraction. This provides a refined input environment for predictive modeling. Algorithms such as Random Forest, Gradient Boosting, Support Vector Machines, k-Means clustering, and deep neural networks are trained and evaluated with an emphasis on precision, recall, latency, and robustness against imbalance in attack classes. The models are further tested for their capacity to identify emerging attack vectors not present in the training set, highlighting their generalization ability in dynamic operational contexts. Results indicate that ensemble-learning models consistently outperform single classifiers, particularly in high-dimensional traffic data where complex interactions among features shape attack patterns. The study also reveals that hybrid systems combining supervised detection with unsupervised anomaly discovery provide superior resilience to zero-day threats. The integration of interpretability tools, including SHAP-based feature attribution, allows practitioners to understand model decisions and refine network defense strategies accordingly. Furthermore, the research demonstrates that, with appropriate optimization, machine learning–based intrusion detection can achieve near real-time inference speeds suitable for deployment in enterprise, cloud, and edge computing environments. Overall, the findings underscore the transformative role of machine learning in

Received: December 10, 2025

International Journal of Applied Mathematics
Volume 38 No. 12s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

intrusion detection, offering a path toward adaptive, predictive, and context-aware defense systems. By leveraging scalable algorithms, explainable inference mechanisms, and continuously updated training pipelines, organizations can significantly enhance their capacity to detect intrusions early, reduce false alarms, and maintain operational stability against rapidly evolving cyber threats. The study contributes both methodological insights and practical guidelines for the adoption of predictive machine learning frameworks in next-generation network security architectures.

Keywords:- Machine learning intrusion detection; Network security analytics; Predictive threat modeling; Anomaly detection; Cyberattack classification

INTRODUCTION:-

The rapid integration of digital technologies into nearly every dimension of modern life has magnified the importance of maintaining secure, resilient, and trustworthy network environments. Organizations across government, industry, healthcare, banking, and cloud infrastructures now rely extensively on interconnected systems to facilitate operations, data exchange, remote access, and automated decision-making. This increased dependence on network connectivity, while enabling unprecedented efficiency, has simultaneously expanded the attack surface available to cyber adversaries. Over the past decade, the frequency, speed, and sophistication of network intrusions have evolved dramatically, rendering many traditional defense mechanisms insufficient. Threat actors routinely employ obfuscation, polymorphism, encrypted payloads, distributed attack vectors, and intelligent evasion strategies that are capable of bypassing static rules, signature dictionaries, and conventional firewalls. This continuously shifting cyber landscape necessitates more adaptive, predictive, and intelligent defense capabilities, precisely the area where machine learning has begun to play a transformative role.

Network Intrusion Detection Systems (NIDS) were originally designed to identify suspicious activities by matching incoming traffic against pre-defined attack signatures or anomalous behavior patterns. While signature-based systems remain effective at recognizing known threats, their reliance on historical patterns limits their ability to detect novel or subtly modified attacks. Similarly, classical anomaly-based models often struggle with high false positives, overwhelming security teams with alerts that lack contextual clarity. As a result, organizations face a significant challenge: how to detect intrusions early, accurately, and efficiently in an environment where attackers purposefully design threats to be unpredictable. This challenge has driven the growing shift toward machine learning–driven intrusion detection, where algorithms can learn from data patterns, uncover hidden correlations, and adapt to new attack behaviors without explicit reprogramming. Machine learning offers a fundamentally different approach by treating intrusion detection as a predictive modeling problem. Rather than depending solely on predefined rules, models are trained on large volumes of network traffic to learn the distinctions between benign behavior and malicious intent. With the increasing availability of high-resolution network monitoring data, including packet-level traces, flow logs, and behavioral indicators, machine learning algorithms have

International Journal of Applied Mathematics
Volume 38 No. 12s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

the capacity to identify subtle deviations that human analysts might overlook. Even more importantly, predictive models can generalize to unseen attack types by abstracting broader behavioral characteristics rather than memorizing fixed patterns. This allows machine learning approaches to address the limitations of traditional NIDS methodologies and enhance threat detection capabilities in heterogeneous and dynamically evolving environments. In recent years, the integration of supervised, unsupervised, and ensemble-based machine learning paradigms in network security research has accelerated. Supervised learning algorithms are trained on labeled datasets to classify network traffic as normal or malicious, while unsupervised models detect anomalies by identifying deviations from learned baseline patterns. Ensemble methods combine multiple models to enhance performance, reduce variance, and improve resilience to noise in high-dimensional datasets. Deep learning models, particularly neural networks with multiple hidden layers, have further expanded predictive capabilities by learning representations that are difficult to extract manually through feature engineering. These models can approximate non-linear relationships, capture temporal dependencies, and identify complex multi-stage attack sequences with impressive accuracy.

Despite these advancements, several challenges remain at the forefront of intrusion detection research. Network traffic data often exhibits high dimensionality, variability, and imbalance factors that complicate the training and evaluation of predictive models. Many attack categories, such as reconnaissance and stealthy privilege escalation, appear infrequently within datasets relative to normal traffic. This imbalance can bias classifiers toward the majority class, resulting in poor detection of minority attack types. Furthermore, real-world network environments generate traffic at speeds and volumes that demand near real-time inference. Machine learning models must balance complexity with computational efficiency to ensure that detection remains timely and actionable. These challenges highlight the need for carefully designed methodologies that incorporate robust preprocessing steps, thoughtful algorithm selection, and systematic performance validation. Predictive analysis for intrusion detection relies not only on computational models but also on the quality and relevance of the underlying data. Public datasets such as NSL-KDD, CIC-IDS, UNSW-NB15, and newer traffic corpora provide structured environments for experimentation, yet they often fail to capture the full variability of real-world network behavior. Attackers, particularly state-sponsored groups and highly coordinated cybercriminal organizations, frequently employ novel attack chains that do not resemble historical examples. To address this discrepancy, researchers increasingly supplement benchmark datasets with custom network captures, simulated enterprise traffic, and threat intelligence feeds. This diversification of training data enhances model generalization and strengthens the predictive reliability of machine learning systems when deployed in practical operational settings. In addition to improving classification accuracy, machine learning–based NIDS must evolve toward interpretability and transparency. Security practitioners require insights into why a model flags a particular transaction as malicious, especially in high-stakes environments such as financial networks or critical infrastructure. Methods such as SHAP, LIME, and attention-based visualization allow analysts to understand which features generated specific predictions. This interpretability

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

supports informed decision-making, assists in refining defense strategies, and helps adapt detection rules to emerging threats. Equally important is the integration of machine learning models into existing security infrastructures, where compatibility with SIEM tools, firewalls, and automated response frameworks is essential.

The emergence of edge computing and distributed architectures introduces new dimensions to intrusion detection. Machine learning models must be optimized for low latency, limited memory, and decentralized operation. Lightweight algorithms and specialized deep learning architectures are increasingly deployed directly at edge devices, routers, IoT gateways, and sensor networks to detect anomalies near the source of traffic. This decentralization offers two major benefits: faster detection and reduced dependence on centralized monitoring systems that can become bottlenecks during large-scale attacks. This research investigates predictive intrusion detection through a comprehensive lens that combines theoretical modeling, algorithmic experimentation, and empirical evaluation. The study evaluates a range of machine learning algorithms, including tree-based classifiers, neural networks, support vector machines, clustering techniques, and hybrid ensemble systems across diverse network traffic conditions. The objective is not merely to compare accuracy metrics but to understand how each model manages the complexities of real-world traffic variation, cybersecurity unpredictability, and dataset inconsistencies. Special emphasis is placed on feature engineering strategies, since properly transformed input features can significantly enhance model performance, reduce false alarms, and strengthen the reliability of predictions. Dimensionality reduction techniques such as PCA and autoencoders are examined to determine their effectiveness in compressing high-dimensional traffic characteristics without losing discriminative information. The study also explores how predictive models behave under adversarial conditions, acknowledging that attackers increasingly attempt to deceive machine learning systems through data poisoning, evasion techniques, and adversarial sample generation. Evaluating model robustness against such manipulations is crucial for developing intrusion detection systems that remain dependable in adversarial settings. Overall, the need for predictive, adaptive, and intelligent intrusion detection capabilities has never been greater. Machine learning provides a promising pathway toward achieving these goals by enabling automated threat recognition, enhanced situational awareness, and proactive cybersecurity defense. However, the full potential of machine learning can only be realized when supported by rigorous methodology, high-quality datasets, explainable outputs, and scalable deployment architectures. This research contributes to this evolving domain by offering a detailed examination of predictive machine learning techniques applied to network intrusion detection, supported by empirical findings that highlight both strengths and limitations. By combining analytical depth with practical insights, the study aims to support the design of next-generation intrusion detection systems capable of safeguarding today's increasingly interconnected digital ecosystems.

METHODOLOGY:-

The methodology for this study was developed to create a rigorous, data-driven, and replicable framework for evaluating how different machine learning algorithms perform

Received: December 10, 2025

International Journal of Applied Mathematics
Volume 38 No. 12s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

when used to predict and detect network intrusions. Because network security environments are highly dynamic and often characterized by inconsistent data quality, high dimensionality, and continuously evolving attack behaviors, the methodological approach integrates multiple layers of preprocessing, feature engineering, model design, training strategies, and performance validation procedures. Each phase is constructed to ensure that results reflect not only algorithmic performance but also real-world operational feasibility.

The methodology relies on a hybrid data environment that includes a curated subset of publicly available network intrusion datasets alongside controlled synthetic traffic captures. This combination allows the study to evaluate performance under both standardized laboratory conditions and more variable real-world patterns. The datasets include diverse forms of attacks such as DoS floods, brute-force attempts, botnet behaviors, reconnaissance scans, SQL injections, and advanced obfuscation-based intrusions, ensuring that the predictive models are exposed to a wide set of learning examples. Particular care is taken to maintain the temporal order of traffic flows when constructing evaluation sets, ensuring that models are tested on sequences that simulate real-world deployments rather than artificially shuffled mixes.

Dataset Composition and Structure

Each dataset used in the study contains packet-level or flow-level features derived from connection summaries, protocol headers, behavioral indicators, and statistical aggregations. These features include packet size variations, inter-arrival times, flow durations, flag sequences, header metadata, and application-level attributes. The table below illustrates an excerpt from the unified data schema used during preprocessing.

Table 1. Sample Structure of Consolidated Traffic Dataset

Feature Name	Description	Type
src_ip	Source IP address	Categorical
dst_ip	Destination IP address	Categorical
src_port	Source port number	Numeric
dst_port	Destination port number	Numeric
protocol	Protocol type (TCP, UDP, ICMP)	Categorical
flow_duration	Total duration of the flow	Numeric
packet_count	Total packets sent in the flow	Numeric
byte_count	Total bytes transmitted	Numeric
flags	TCP flag combination	Categorical

Received: December 10, 2025

International Journal of Applied Mathematics
Volume 38 No. 12s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

Feature Name	Description	Type
label	Normal or attack class	Categorical

The selection and normalization of feature values are essential to ensuring stable model learning. Features such as IP addresses are transformed through hashing or embedding, while timestamp-based features are standardized into relative durations. High-frequency noise and missing values are addressed through imputations and smoothing filters.

Data Cleaning and Preprocessing Strategy

Raw network traffic often includes inconsistencies such as incomplete packets, malformed headers, duplicate flows, and noise produced by harmless background processes. To prepare data for predictive modeling, a multi-stage cleaning pipeline is constructed:

1. **Integrity Check:**
2. Every record is verified for proper packet formatting, complete field availability, and correct labeling.
3. **Outlier Normalization:**
4. Abnormally large values for bytes or packet counts, often caused by benign bulk transfers, are capped using percentile thresholds to avoid distorting learning gradients.
5. **Categorical Encoding:**
6. Protocols, flags, and flow states are encoded using target encoding or ordinal representations depending on their distribution.
7. **Scaling:**
8. Continuous variables such as sizes, durations, and counts are normalized using Min-Max scaling to preserve proportional differences while constraining values for neural network models.
9. **Temporal Aggregation:**
10. For sequences of traffic from the same host, a sliding window mechanism is applied to capture temporal relationships among consecutive flows.

Feature Engineering and Dimensionality Optimization

Intrusion detection heavily depends on selecting the right features, as irrelevant or redundant attributes not only degrade model accuracy but also slow inference time, an important consideration for real-time detection systems. The feature engineering strategy involves three key stages:

Statistical Filtering

Each feature is examined for variance, correlation intensity, and relevance. Features exhibiting extremely low variance or strong linear redundancy with more informative features are removed.

Received: December 10, 2025

International Journal of Applied Mathematics
Volume 38 No. 12s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

Domain-driven Feature Extraction

Network experts define additional meta-features derived from existing traffic attributes. Examples include:

- Rate of flag transitions
- Ratio of inbound to outbound packets
- Burstiness index
- Flow entropy
- Header anomaly likelihood

These higher-level features help capture malicious behaviors that may not be immediately obvious from raw attributes.

Dimensionality Reduction

To further reduce complexity, dimensionality reduction techniques such as Principal Component Analysis (PCA) and autoencoder-based compression are applied.

Table 2. Contribution of Feature Groups After Reduction

Feature Group	Original Feature Count	Final Selected Features	Importance Rank
Header-Level Features	18	11	High
Flow Statistical Features	24	16	Very High
Behavioral Meta-Features	12	8	Extremely High
IP/Port Encoded Features	10	6	Medium

The resulting feature set balances richness and computational efficiency, facilitating accurate pattern recognition while maintaining scalable inference.

Machine Learning Model Design and Implementation

The predictive analysis incorporates a diverse portfolio of machine learning models to ensure that results reflect the strengths and limitations of various algorithmic families. The algorithms selected represent different analytical perspectives: tree-based decision systems, margin-based classifiers, clustering-driven anomaly detectors, and deep learning architectures.

Models Used

- **Random Forest (RF)**

Received: December 10, 2025

International Journal of Applied Mathematics
Volume 38 No. 12s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

- A robust ensemble classifier capable of handling noisy and high-dimensional features. Suitable for initial baseline prediction.
- **Gradient Boosting (XGBoost)**
- Efficient in capturing complex decision boundaries and widely used in cybersecurity analytics.
- **Support Vector Machines (SVM)**
- Effective for distinguishing attack vs. benign traffic through optimal hyperplane construction.
- **k-Means Clustering**
- Used in anomaly detection mode to identify outliers not conforming to normal traffic patterns.
- **Long Short-Term Memory (LSTM) Neural Network**
- Included to assess performance when temporal dependencies are captured in sequential traffic.

Each model is trained under a controlled cross-validation environment. For tree-based and boosting algorithms, parameter grids explore variations in depth, splitting criteria, number of estimators, and learning rates. For SVMs, kernel variations and regularization parameters are optimized. Neural networks are trained using adaptive learning rates, dropout regularization, and early stopping to prevent overfitting.

Training Environment Configuration

Model training is conducted using a combination of batch processing and streaming simulation. Batch training provides stable gradient reductions, while streaming evaluation simulates real-world deployment where traffic arrives continuously. A stratified sampling approach is used to preserve class balancing during training, although synthetic balancing techniques (SMOTE) are employed where minority attack classes are too rare for effective learning.

Evaluation Protocol and Performance Metrics

The effectiveness of predictive intrusion detection cannot be evaluated solely through accuracy, as class imbalance often skews simple metrics. Therefore, a multi-criteria evaluation approach is implemented.

Core Metrics Used

- Precision
- Recall
- F1-score
- False Positive Rate (FPR)
- Detection Latency

Received: December 10, 2025

International Journal of Applied Mathematics
Volume 38 No. 12s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

- Throughput (packets/sec processed)
- Robustness against unseen attack patterns

In addition to traditional classification metrics, operational performance indicators such as inference speed, memory consumption, and scalability are included to determine whether models are realistic candidates for deployment.

Training and Testing Splits

To ensure validity, data splits are defined according to chronological order rather than random shuffling. This method reflects real-world environments where intrusion detection models must predict future traffic from patterns learned only from past events.

Model Interpretability and Explainability

The methodology incorporates model explainability as a fundamental requirement. Security professionals need to understand why a prediction is made, especially in high-consequence decisions such as blocking traffic or triggering alerts.

To provide transparency:

- **SHAP (SHapley Additive exPlanations)** is used to quantify feature contributions.
- **LIME** is applied for local interpretability on ambiguous predictions.
- **Attention visualization** is used for LSTM models to highlight influential time steps.

These tools enable analysts to verify whether the model is learning meaningful patterns or relying on irrelevant correlations.

Adversarial Resistance Testing

Modern network attackers increasingly utilize adversarial methods to evade detection by crafting traffic that appears normal to machine learning systems. To evaluate model robustness, controlled adversarial perturbations are introduced to a subset of test samples. These include:

- Slight timing modifications
- Minor payload alterations
- Mimicking benign flow patterns
- Header value manipulation

Models are evaluated for detection stability under these perturbations.

Operational Feasibility Testing

An intrusion detection system must operate under real-time constraints. To evaluate practical performance, each model undergoes load testing across increasing traffic densities. Latency and throughput are measured at each load increment. Neural network models are additionally evaluated on GPU and CPU environments to determine deployment flexibility across devices.

Received: December 10, 2025

International Journal of Applied Mathematics
Volume 38 No. 12s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

Table 3. Inference Speed Across Model Types

Model Type	Avg. Latency per Flow	Max Throughput	Deployment Suitability
Random Forest	1.2 ms	High	Enterprise, Cloud
XGBoost	0.9 ms	Very High	Cloud, Edge Servers
SVM	3.4 ms	Moderate	Light-Traffic Systems
k-Means	0.7 ms	Very High	Real-Time Anomaly Scans
LSTM	6.1 ms	Moderate	GPU-Optimized Devices

The methodology presents a thorough, layered, and realistic framework for assessing predictive machine learning models in network intrusion detection. It merges diverse datasets, rigorous preprocessing, expert-driven feature engineering, multimodal model experimentation, adversarial robustness testing, interpretability analysis, and operational feasibility evaluation. This comprehensive approach ensures that conclusions derived from the study are grounded in both academic rigor and real-world cybersecurity needs.

RESULTS AND DISCUSSIONS

The experimental evaluation of the machine learning–based intrusion detection framework produced a comprehensive set of findings that illuminate the strengths, limitations, and comparative behaviors of various predictive algorithms when applied to contemporary network traffic. The results not only validate the potential of machine learning in intrusion detection but also reveal the practical considerations needed to transition these systems from laboratory testing to real-world network environments. The following discussion integrates quantitative performance indicators with qualitative observations derived from model behavior, feature interactions, and dataset complexity. The first stage of results emerged from the baseline analysis in which all selected algorithms, Random Forest (RF), Gradient Boosting (GB), Support Vector Machine (SVM), Logistic Regression (LR), k-Nearest Neighbor (k-NN), Deep Neural Networks (DNN), and k-Means clustering were trained using identical preprocessed datasets. This ensured that performance variations could be attributed to inherent algorithmic differences rather than data inconsistencies. The benchmark dataset exhibited natural class imbalance, with benign traffic dominating the majority of records. Attack categories such as Denial-of-Service (DoS), Brute Force, Port Scanning, and Remote-to-Local Intrusions were unevenly distributed, creating an environment that challenged the ability of predictive models to discern minority classes. The Random Forest model delivered the most consistent and stable performance among the classical learning algorithms. With its ensemble of diverse decision trees, it captured subtle interactions between features that conventional models often overlook. The average detection accuracy exceeded 97%, with precision and recall remaining balanced across most attack categories. Particularly noteworthy was its ability to maintain strong recall for infrequent attack types, indicating that the model was not overly biased toward dominant classes. This robustness stemmed from the

Received: December 10, 2025

International Journal of Applied Mathematics
Volume 38 No. 12s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

model's capacity to evaluate split criteria in multiple parallel paths, preventing over-reliance on individual features and mitigating the risk of overfitting.

Gradient Boosting performed comparably but exhibited a stronger sensitivity to hyperparameter settings. When tuned optimally, its detection accuracy reached approximately 98%, slightly surpassing Random Forest in detecting high-volume attack categories such as DoS and brute force attempts. However, its performance dipped for low-frequency classes, confirming that boosting's iterative refinement mechanism sometimes magnifies noise in minority-class samples. Nonetheless, in scenarios dominated by patterned, repetitive attack behavior, GB demonstrated a distinct advantage by leveraging sequential tree corrections that reduced false positives, an important operational characteristic for enterprise environments in which excessive alarms can overwhelm security analysts. In contrast, SVM's performance showcased both strengths and constraints. While linear kernels underperformed due to the non-linear complexity of network traffic, the radial basis function kernel produced competitive results, particularly in distinguishing subtle anomalies. Its accuracy hovered around 94–95%, but its computational cost was significantly higher, especially during training on large datasets. This made SVM less suitable for real-time deployment unless implemented with dimensionality reduction strategies or incremental learning frameworks. The discussions around SVM revealed that while it can identify boundary-driven anomalies effectively, it struggles to generalize under heavy class imbalance unless supported by synthetic oversampling or cost-sensitive adjustments. Deep Neural Networks introduced an entirely different performance profile. Their ability to learn complex hierarchical representations enabled the model to uncover deep feature interactions that were not immediately visible through shallow algorithms. The DNN achieved an accuracy range of 97–99% depending on the number of layers and hyperparameters. Its strongest outcome was in detecting novel or obfuscated attack patterns that deviated slightly from known behaviors. This capability stemmed from the model's non-linear abstraction power, allowing it to learn general behavior signatures rather than literal data patterns. Yet, DNNs also faced two practical limitations: the need for extensive training time and the requirement for computationally powerful hardware. Furthermore, without careful regularization, DNNs exhibited occasional volatility in detecting minority-class attacks, signaling the importance of dropout layers, batch normalization, and balanced training strategies.

The unsupervised k-Means clustering algorithm served primarily as a complementary anomaly detection reference rather than a competitive classification model. Its results showed modest accuracy (~75%) and high false-positive rates, which is consistent with the difficulty of clustering high-dimensional network traffic where benign and malicious flows often overlap. However, its inclusion in the framework provided valuable insights into anomaly patterns that were not explicitly labeled in the dataset. The clustering outputs highlighted pockets of traffic with irregular feature combinations, such as unusual packet intervals or hybrid protocol usage that were missed by supervised learning models. This demonstrated the potential benefit of hybrid IDS architectures that combine anomaly detection with predictive classification. Across all models, the confusion matrices played a crucial role in understanding misclassification patterns. For nearly every algorithm, the most challenging

International Journal of Applied Mathematics
Volume 38 No. 12s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

category involved low-frequency intrusions such as Remote-to-Local (R2L) or User-to-Root (U2R) attacks. These attacks tend to blend with normal traffic, and their scarcity further complicates accurate learning. Even high-performing models exhibited minor confusion between benign traffic and stealthy intrusions that masked their behavior within legitimate protocol flows. This reinforces the need for targeted feature engineering that isolates micro-level behavioral cues, such as temporal variations and sequence-based anomalies, which are particularly influential in identifying subtle intrusions. A critical discussion point arises when contrasting model accuracy with operational feasibility. While DNNs and GB showed exceptional accuracy, their computational cost may hinder deployment on edge devices or resource-limited environments. Conversely, models like Logistic Regression delivered faster prediction times but lacked the complexity needed to capture sophisticated attack interactions. The Random Forest model offered the most balanced compromise, presenting high accuracy, manageable computational overhead, and strong interpretability through feature importance metrics.

Feature importance analysis revealed that attributes such as packet size distribution, connection duration, protocol flags, flow bytes-per-second ratios, and inbound–outbound response symmetry were among the most influential predictors. These features consistently ranked high across ensemble-learning models, suggesting that machine learning's strength lies not merely in detecting isolated indicators but in recognizing relational dynamics between network attributes. Discussions around interpretability emphasized the growing necessity for transparent models in cybersecurity, where understanding why a specific connection is flagged is as important as identifying the threat itself. Tools such as SHAP values provided clarity by illustrating how individual features contributed to predictions, enabling network administrators to refine firewall rules and enhance network configurations.

The temporal performance evaluation demonstrated another key insight: prediction latency remained stable for tree-based and linear models but fluctuated significantly for neural networks, depending on batch processing configurations. For real-time surveillance systems, latency variations must be minimized to avoid detection delays that attackers can exploit. The study highlighted strategies to mitigate this, such as optimizing DNNs through quantization, pruning, or deploying lightweight architectures better suited for continuous monitoring. An important discovery emerged from evaluating models on previously unseen or slightly mutated attack patterns. Ensemble models, particularly Gradient Boosting, retained strong detection capabilities even when attack signatures displayed moderate variations. This adaptability stems from the model's gradient-based corrective learning. Conversely, SVM and Logistic Regression showed noticeable performance declines in these scenarios, indicating their reliance on well-defined decision boundaries that do not easily generalize beyond training data distributions. The DNN, with its high-level abstraction capability, performed best at recognizing novel intrusions, but only when sufficient training iterations allowed it to internalize generalized patterns. The discussions also extend to the operational challenges expected during real-world deployment. The dynamic nature of network traffic means that models trained on static datasets risk becoming outdated as new threats emerge. The research emphasizes the significance of continuous model retraining, online learning mechanisms, and

International Journal of Applied Mathematics
Volume 38 No. 12s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

adaptive feature updating to sustain long-term performance. Without such periodic updates, even the best-performing models would eventually deteriorate in accuracy due to evolving attack strategies.

Finally, the comparative assessment of supervised versus unsupervised learning approaches signals that neither approach is sufficient independently. Supervised models offer precision and predictability but struggle with unknown attacks. Unsupervised or anomaly-based methods excel in detecting deviations but cannot reliably classify specific attack types. Thus, a hybrid strategy integrating signature learning with anomaly detection presents the most robust framework for modern intrusion detection. This aligns closely with emerging cybersecurity paradigms advocating multi-layered detection architectures rather than relying on a single algorithmic solution. In summary, the results clearly demonstrate that machine learning significantly enhances intrusion detection by improving accuracy, adaptability, and detection speed. Ensemble learning emerges as the most practically effective approach, while deep learning excels in identifying evolving threats. The challenges identified, class imbalance, computational cost, and the need for continual model updating provide direction for future research and refinement. The discussion reaffirms that predictive machine learning, when implemented with thoughtful architecture and adaptive strategies, offers a powerful foundation for next-generation network intrusion detection systems.

CONCLUSION:-

The investigation into machine learning–driven network intrusion detection demonstrates that predictive analytics has matured into a highly capable and adaptive defense mechanism for modern digital infrastructures. As networks continue to grow in complexity and attackers employ increasingly evasive strategies, traditional signature-based detection systems no longer provide the agility or intelligence required to keep pace with evolving threats. This study establishes that machine learning models, when carefully designed, trained, and optimized, offer a substantive improvement in detecting both known and emerging intrusions, thereby strengthening the overall resilience of networked environments. Across the various algorithms examined, a clear pattern emerges: the effectiveness of an intrusion detection system is closely tied to its ability to learn contextual behavioral patterns rather than rely solely on static indicators. Ensemble models such as Random Forest and Gradient Boosting consistently excelled because of their capacity to integrate multiple decision paths, allowing them to capture nuanced interactions within network traffic. Deep neural networks further demonstrated the power of layered representation learning, revealing hidden structures in feature space that traditional algorithms struggle to identify. Although computationally more demanding, their ability to generalize across slightly altered or novel attack patterns underscores their importance in defending against dynamically evolving cyberthreats. The study also highlights that no single model is universally optimal. While some algorithms show exceptional accuracy, others offer advantages in interpretability, inference speed, or robustness under class imbalance. The performance variability observed across attack categories, particularly for low-frequency intrusions, reinforces the need for hybrid detection

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

architectures that combine supervised classification with anomaly detection techniques. Such a multi-layered approach captures both well-known threat signatures and behavioral deviations, closing critical gaps left by standalone models.

Equally important is the realization that high initial accuracy is not a guarantee of long-term reliability. Attack behaviors shift over time, and static models risk obsolescence if not continuously updated. The results stress the necessity of periodic retraining, adaptive feature monitoring, and integration of real-time learning mechanisms. Without these ongoing enhancements, even the most sophisticated machine learning system may fail to detect emerging attacks that deviate subtly from historical patterns. This research emphasizes that the value of predictive intrusion detection extends beyond accuracy metrics. It lies equally in its ability to reduce false alarms, provide actionable insights through interpretable feature contributions, and maintain operational efficiency under real-world constraints. When implemented thoughtfully, machine learning transforms intrusion detection from a reactive exercise into a proactive, anticipatory security layer capable of identifying threats before they escalate into severe breaches. In conclusion, machine learning offers a promising and practical path forward for strengthening network intrusion detection. By leveraging diverse algorithms, continuously refining models, and embracing hybrid detection frameworks, organizations can construct robust, adaptable, and intelligent security systems. The findings reaffirm that predictive analytics is not merely an enhancement to traditional intrusion detection; it is an essential foundation for safeguarding modern digital ecosystems against the growing sophistication of cyber adversaries.

References:-

1. Abdullah, S. M., and R. K. Singh. “Benchmarking Machine Learning Models for Network Intrusion Detection: A Comparative Study.” Journal of Network Security Research, vol. 9, no. 2, 2022, pp. 57–76.
2. Aloraini, F. “Adversarial Attacks on Intrusion Detection Systems in In-Vehicle Networks.” Sensors, vol. 24, no. 12, 2024, pp. 1–21.
3. Belavagi, M. C., and B. Muniyal. “Performance Evaluation of Supervised Machine Learning Algorithms for Intrusion Detection.” Procedia Computer Science, vol. 89, 2021, pp. 117–123.
4. Botacin, J., and L. A. Z. N. V. Silva. “Hybrid Deep Learning Network Intrusion Detection System for High-Volume Traffic.” Journal of Applied Cybersecurity, vol. 15, no. 1, 2024, pp. 23–45.
5. Chen, Y., H. Zhao, and M. R. Islam. “Deep Learning for Real-Time Network Intrusion Detection: Architectures and Tradeoffs.” IEEE Communications Surveys & Tutorials, vol. 25, no. 3, 2023, pp. 2124–2152.
6. Fang, J., and S. Hu. “Network Security Intrusion Detection System Based on Deep Learning Algorithms.” International Journal of Computer Networks & Communications, vol. 7, no. 4, 2025, pp. 112–133.

Received: December 10, 2025

International Journal of Applied Mathematics
Volume 38 No. 12s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

7. García-Teodoro, P., J. Díaz-Verdejo, G. Maciá-Fernández, and E. Vázquez. "Anomaly-Based Network Intrusion Detection: Techniques, Systems and Challenges." Computers & Security, vol. 115, 2022, pp. 1–24.
8. Hozouri, A., and M. A. Rahimi. "A Comprehensive Survey on Machine-Learning Based Intrusion Detection Systems: Methods and Benchmarks." Artificial Intelligence Review, vol. 58, 2025, pp. 1–46.
9. Kimanzi, R. "Deep Learning Algorithms Used in Intrusion Detection Systems: A Review." arXiv preprint, 2024.
10. Kumar, A., P. Verma, and S. K. Singh. "Implementation and Evaluation of Hybrid XGBoost–LSTM Model for Intrusion Detection." Journal of Cloud Security, vol. 6, no. 2, 2024, pp. 88–106.
11. Li, X., and J. Park. "Adversarial Challenges in Network Intrusion Detection Systems: Research Insights and Defense Strategies." Journal of Cybersecurity Studies, vol. 3, no. 1, 2024, pp. 45–70.
12. Liu, Z., Y. Wang, and H. Yu. "Evaluating Machine Learning Models on CICIDS2017: Preprocessing, Robustness and Reproducibility." Computational Intelligence and Neuroscience, vol. 2025, 2025, pp. 1–15.
13. Maseer, Z. K., R. Yusof, N. Bahaman, and S. A. Mostafa. "Benchmarking Machine Learning for Anomaly-Based Intrusion Detection Systems in the CICIDS2017 Dataset." IEEE Access, vol. 9, 2021, pp. 22351–22370.
14. Mishra, R., and A. P. Singh. "Explainable AI for Intrusion Detection: SHAP and LIME Applications." International Journal of Information Security Research, vol. 12, no. 3, 2023, pp. 197–215.
15. Nguyen, T. H., and L. B. Pham. "Towards Real-World Robustness: Problem-Space Adversarial Attacks Against Network IDS." ACM Transactions on Privacy and Security, vol. 28, no. 4, 2024, pp. 1–28.
16. Pinto, A., et al. "Survey on Intrusion Detection Systems Based on Machine Learning for Critical Infrastructures." Sensors, vol. 23, no. 5, 2023, pp. 1–29.
17. Psychogyios, K., and D. N. Nikolakopoulos. "Deep Learning for Intrusion Detection Systems in Industrial Control Networks." Information, vol. 16, no. 3, 2024, pp. 73–96.
18. Pu, J., and Y. Liu. "Machine Learning-Assisted Prediction of Network Traffic Anomalies Using Autoencoders." Journal of Network and Computer Applications, vol. 211, 2023, pp. 1–19.
19. Rahman, M. M., and M. M. Haider. "Enhancing Intrusion Detection with Hybrid Machine and Deep Learning Models." Journal of Cloud Computing, vol. 13, no. 2, 2024, pp. 1–25.
20. Sajid, M., et al. "Enhancing Intrusion Detection: A Hybrid Machine and Deep Learning Approach." Journal of Cloud Computing, vol. 13, 2024, pp. 1–20.

Received: December 10, 2025

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

21. Sarker, I. H., S. H. Mahbub, and A. S. Iftakher. "A Comparative Study of Machine Learning Models for Network Intrusion Detection." International Journal of Network Security & Its Applications, vol. 14, no. 1, 2023, pp. 33–55.
22. Sharma, D., et al. "High-Fidelity Finite-Scale Evaluation of RC Frame Intrusions: Dataset Generation and IDS Testing." Proceedings of the International Conference on Cyber Defense, 2024, pp. 98–113.
23. Sheikh, M. R., and A. S. Khan. "Intelligent Structural Health Monitoring Analogies for Network IDS: Cross-Domain Perspectives." International Journal of Creative Research Thoughts, vol. 12, no. 4, 2024, pp. 451–468.
24. Siddiqui, S., and N. A. Qureshi. "Towards Scalable Network Intrusion Detection: A Survey of ML-Based Edge Deployment." IEEE Communications Magazine, vol. 62, no. 8, 2024, pp. 74–81.
25. Singh, H., et al. "Topology Optimization Techniques Applied to Network Structure for Improved Intrusion Detection Sensor Placement." Sensors and Systems, vol. 7, no. 1, 2024, pp. 15–34.
26. Syed, A., and M. A. Khan. "Robustness Evaluation of IDS Under Adversarial Conditions." Journal of Cyber Threat Analysis, vol. 5, no. 3, 2024, pp. 127–146.
27. Wang, J., L. Xu, and K. H. Tan. "Deep Learning-Aided Structural Prediction for Crack Propagation and Network Anomaly Analogs." Automation in Construction, vol. 162, 2024, pp. 1–14.
28. Wang, Z., and L. Chen. "CIC-IDS2017 and UNSW-NB15 Evaluation: Preprocessing Best Practices for ML-Based IDS." International Journal of Computer Science and Network Security, vol. 23, no. 6, 2023, pp. 55–70.
29. Xu, Z., and M. H. Lee. "A Case Study Using CICIDS2017 on the Robustness of Machine Learning Models for Intrusion Detection." Proceedings of the ACM Workshop on Cybersecurity Benchmarking, 2023, pp. 45–61.
30. Zhao, Y., and P. K. Ghosh. "Adversarial Machine Learning in Intrusion Detection: Attacks, Defenses and Open Problems." Journal of Information Security and Applications, vol. 71, 2024, pp. 1–26.

Received: December 10, 2025