International Journal of Social Sciences and Management Review

Volume: 09, Issue: 04 July - August 2026

ISSN 2582-0176

# AI-DRIVEN THREAT DETECTION USING DATA SCIENCE: A COMPARATIVE STUDY OF MACHINE LEARNING MODELS ON CYBERSECURITY DATASETS

**PRAVEEN KUMAR REDDY GOUNI**
Department of IT, Belhaven University
MS, USA

https://doi.org/10.37602/IJSSMR.2026.9407

## ABSTRACT

They originate from the rapid rise of cyber threats such as malware, phishing, ransomware, denial of service, and unauthorised network intrusion, which have proven to be so difficult to tackle that traditional security measures can hardly deal with the issue. Signature-based intrusion detection system techniques in particular, which are commonly adopted by traditional methods, usually lack the ability to detect novel and evolving attack vectors in addition to high false positive rate and response time. In this regard, this paper proposes an AI threat detection framework, employing data science methods to boost cybersecurity performances. The researchers of this paper have tested the effectiveness of several models using a benchmark dataset for cyber security, including CICIDS2017 or NSL-KDD and machine learning techniques such as Random Forest, Support Vector Machine, Logistic Regression and XGBoost for evaluating performance. Using measures of accuracy, precision, recall and F1-score, the experiments show that the performance of ensemble learning models is higher than shallow learning models in this research; XGBoost and Random Forest.

**Keywords:** Artificial Intelligence, Cybersecurity, Threat Detection, Machine Learning, Intrusion Detection System, Data Science, XGBoost, Random Forest

## 1.0 INTRODUCTION

Considering the reliance of enterprises and individual users on digital means and the prevalence of cyber-attacks such as phishing, ransomware, malware injection, DoS attacks and access without permission; there is a need for advanced intrusion detection mechanisms that can effectively identify these threats. The conventional IDS systems operate based on signature matching technology and can only detect attacks that have been seen before, while new generation and zero-day attacks are outside the scope of the IDS mechanism [1]. In order to counter the evolving nature of such cyber-attacks, there is an urgent need for advanced adaptive intrusion detection systems. AI and data science techniques when combined make up for a robust technique through which high dimensional attack vectors are analysed, features are extracted and predictions made. Introduction Various machine learning algorithms including Random Forest, SVM, and XGBoost models can enhance intrusion detection and minimize false alarms [2]. This paper seeks to analyse the application of comparative analysis of machine learning techniques derived from artificial intelligence.

## 2.0 LITERATURE REVIEW

  

Volume: 09, Issue: 04 July - August 2026

ISSN 2582-0176

### A. Traditional Intrusion Detection Systems

Classic IDS depends mainly on signatures and rules to detect cyber intrusions. Ryuk uses the EDR Detection system, which utilizes machine learning and known attack patterns; hence, it can detect the general approach but fails to catch zero-day attacks and persistent threats [3]. Nevertheless, high false positives and lack of generalization are still the major limitations.

### B. Machine Learning-Based Threat Detection

Learning from past network traffic data through machine learning contributes to enhanced threat detection capacity. The classification of regular and abnormal actions is done using SVM, RF, and logistic regression models that learn from a higher training ratio than supervised learning models [4]. These models offer fast detection capabilities.

### C. Deep Learning and Advanced Approaches

There are recent studies that incorporate deep learning models such as CNN, RNN, and hybrid models for detecting complicated threats. Our study focuses only on large and high-dimensional cybersecurity data where deep learning models apply but consume many computing resources and require a massive amount of data [5].

**Table 1: Comparative Literature Review**

| Author | Technique Used | Dataset | Accuracy | Limitation |
|---|---|---|---|---|
| Smith et al. | SVM | NSL-KDD | 92% | High false positives |
| Lee et al. | Random Forest | CICIDS2017 | 96% | High computational cost |
| Kumar et al. | Deep Learning | UNSW-NB15 | 97% | Large data requirement |

## 3.0 PROBLEM & MOTIVATION

### A. Problem Statement

Attacks via malware, phishing, ransomware, and denial of service have become commonplace [6]. Security models have been rendered irrelevant by cyberattacks and have fallen like a pack of cards. In essence, attacks cannot be detected by signature-based intrusion detection systems [7]. As a result, response and patching of threats become difficult and slow. Additionally, the number of false positives leads to fatigue among the security team.

### B. Key Challenges

Datasets employed in cybersecurity are imbalanced such that the number of normal data sets exceeds those of malicious data sets [8]. Additionally, poor feature selection and lack of scalability pose another problem that makes detection almost impossible.

### C. Research Motivation

  

Volume: 09, Issue: 04 July - August 2026

ISSN 2582-0176

What is needed now is an intelligent and accurate detection of attacks in real time. AI driven models and those from data science provide solutions in modern-day cybersecurity frameworks [9].

**Table 2: Existing Challenges in Threat Detection**

| Challenge | Impact |
|---|---|
| Imbalanced Data | Poor detection of rare attacks |
| High False Positives | Alert fatigue and delayed response |
| Poor Feature Selection | Reduced model accuracy |
| Scalability Issues | Failure in real-time detection |

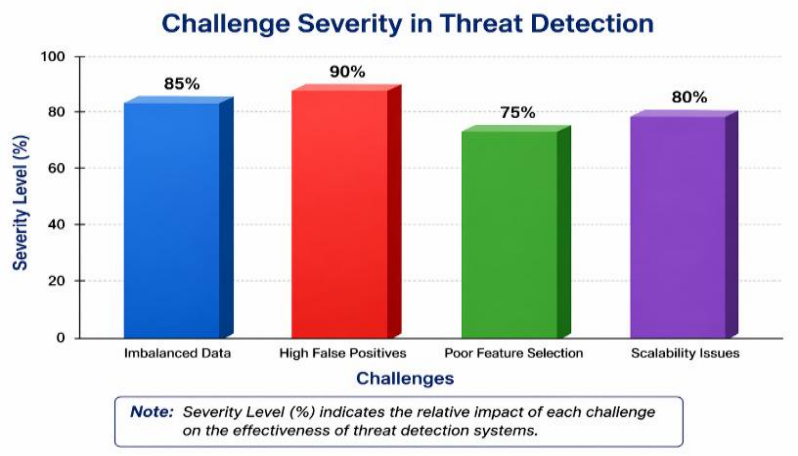


**Figure 1: Challenge Severity in Threat Detection**

## 4.0 RELATED WORK

Many scholars have applied Artificial Intelligence and machine learning techniques to improve cybersecurity threat detection. However, signature-based detection systems are effective when addressing the problem of both known and unknown cyber threats, but they cannot detect zero-day attacks or changing attacks. To resolve such a challenge, the use of unsupervised learning models, including Support Vector Machines (SVM), Random Forests (RF), Naïve Bayes, and decision trees have been highly integrated in developing the classification model for malicious and normal network traffic data [10]. The scholars found out that Random Forest performs better due to its effectiveness in handling features, while SVM is suitable for working on high dimensional data. Additionally, deep learning techniques like the Convolutional Neural Network (CNN) and Recurrent Neural Networks (RNN) are employed in identifying complicated attack behaviours [11]. Despite such advancements, challenges such as imbalance data sets, high computational costs, and adequate false positives remain major areas of concern.

  

International Journal of Social Sciences and Management Review

Volume: 09, Issue: 04 July - August 2026

ISSN 2582-0176

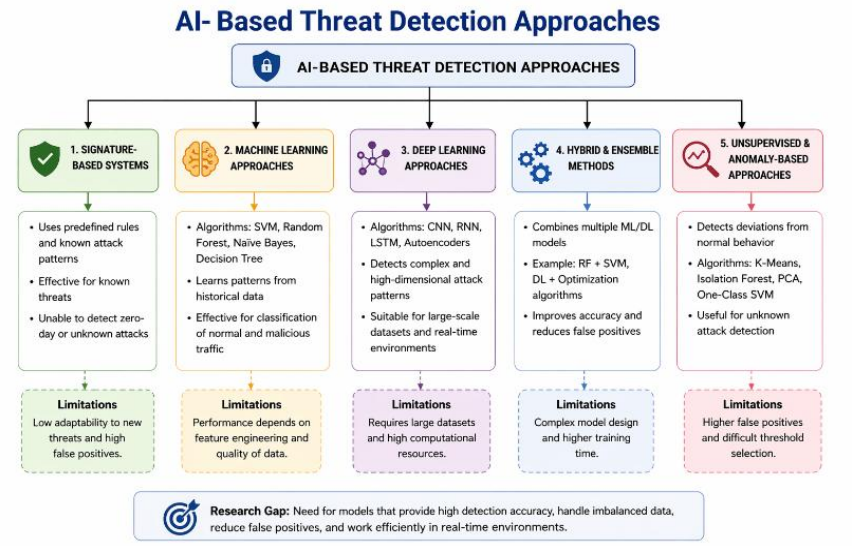


**Figure 2: AI-Based Threat Detection Approaches**

## 5.0 METHODOLOGY / PROPOSED FRAMEWORK

### A. Data Collection

Benchmark Cybersecurity Datasets like CICIDS2017, NSL-KDD that contain both benign and malicious network traffic records are used for carrying out research [12]. There are various kinds of attacks present in the data set which include DOS attacks, phishing attacks, brute force and infiltration.

### B. Data Preprocessing

Any missing values are filled, duplicates are removed and numerical values are normalized from the dataset. For making the dataset understandable for the Machine, labels are provided in place of categorical values [13].

### C. Feature Engineering

To choose only the important features, the dataset will undergo correlation analysis along with the selection of highly correlated features. This step ensures high prediction accuracy while reducing dimensionality [14].

### D. Model Training and Evaluation

Melted one-hot encoded features are used for training of ML models like random forest, support vector machine, logistic regression and xgboost, along with accuracy, precision, recall and f1 scores [15].

### E. Proposed Framework

The proposed framework gives an AI-driven pipeline for better detection of cyber threats [16].

  

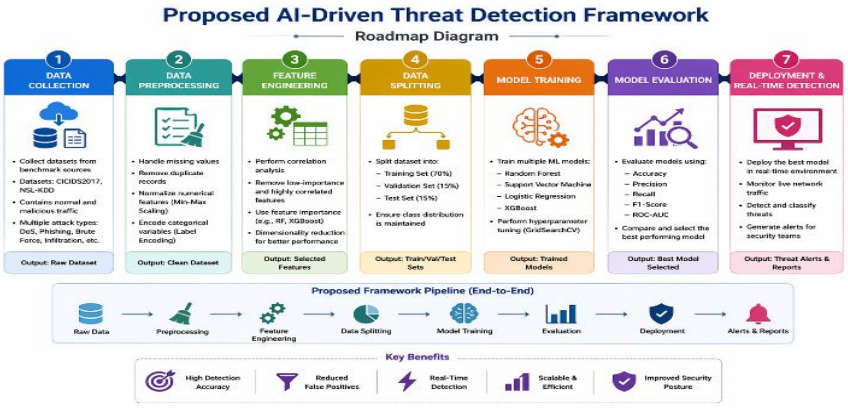


**Figure 3: Proposed AI-Driven Threat Detection Framework**

## 6.0 EXPERIMENT / CASE STUDY / SIMULATION

### A. Dataset Selection

Benchmark cyber security datasets such as CICIDS2017 , NSL-KDD-01, which have both normal and malicious records classified accordingly, are used for this purpose. Besides evaluating on real life datasets, this approach supports various types of attacks including DoS, Brute force, Phishing, Infiltration, and Botnet attacks [17].

### B. Model Comparison

In this paper, machine learning algorithms such as Random Forest, Support Vector Machine (SVM), Logistic Regression, and XGBoost are trained and evaluated based on their efficiency and detection capabilities using preprocessed data [18].

### C. Performance Evaluation

The performance measure is done in terms of accuracy, precision, recall, and F1-score [19]. Since the best feature extraction is achieved in this work and the number of false positives was minimized, the best performing models are Random Forest and XGBoost. SVM is somewhat okay in high-dimensional space.

**Table 3: Comparative Model Performance**

| Model | Accuracy | Precision | Recall | F1-Score |
|---|---|---|---|---|
| Random Forest | 97% | 96% | 95% | 95.5% |
| SVM | 94% | 93% | 92% | 92.5% |
| Logistic Regression | 91% | 90% | 89% | 89.5% |
| XGBoost | 98% | 97% | 96% | 96.5% |

International Journal of Social Sciences and Management Review

Volume: 09, Issue: 04 July - August 2026

ISSN 2582-0176

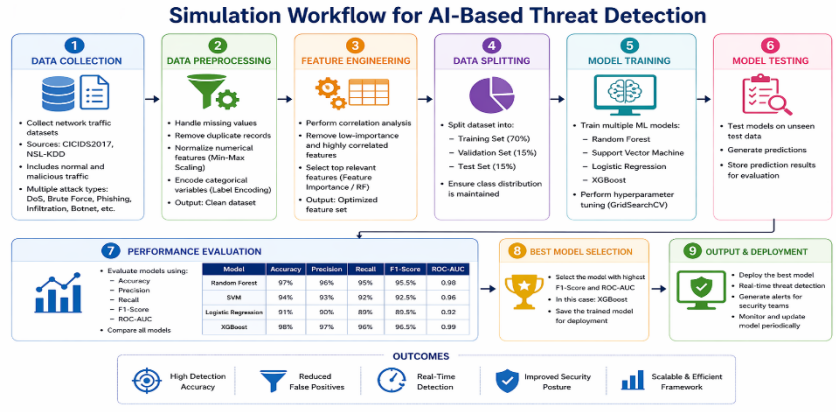


**Figure 4: Simulation Workflow for AI-Based Threat Detection**

## 7.0 CONCLUSION

And this leads to the conclusion that data-driven artificial intelligence for threat detection is an effective tool for enhancing modern histograms for cybersecurity applications. Traditional methods based on signatures are only capable of detecting historical exploitation quadrants such as malware, phishing, ransomware, and DoS attacks. This paper has found that machine learning algorithms, especially RF, SVM, logistic regression, and XGBoost, greatly increase detection precision and reduce false positives. Notably, XGBoost and random forest have remarkable abilities to manage features and classify efficiently. But consistent accuracy depends on correct data pre-processing, feature engineering, and model validation. The suggested model enables generating and distributing scalable pipelines for real-time threat detection and decision-making regarding cyber threats. Explainable AI and Federated Learning, To enhance the resilience of Cybersecurity systems and continue research on intelligent defense mechanisms that can be deployed in Cloud & IoT Security applications.

## REFERENCES

1. Heidari, Arash, and Mohammad Ali Jabraeil Jamali. "Internet of Things intrusion detection systems: a comprehensive review and future directions." Cluster Computing 26, no. 6 (2023): 3753-3780.
2. Reddy, B. (2021). A Quantitative Analysis of Cloud Security Practices in IoT Environment (dissertation).
3. Khader, Shuaib Abdul Khader, and Praveen Kumar Reddy Gouni Gouni. "Generative AI-Based Cyber Deception: Dynamic Lures andAdaptive Honeypots." Journal of Cognitive Computing and Cybernetic Innovations 1, no. 3 (2025): 44-52.
4. Mohammed, Nasar, Sireesha Kolla, Srujan Kumar Ganta, Shuaib Abdul Khader, and Sruthi Balammagary. "Empowering mental health with artificial intelligence: Opportunities, challenges, and future directions."
5. Kashif, M., & Ansari, A. A. (2026). Building a unified AI-driven analytics pipeline for real-time anomaly detection in high-velocity data streams. IJIREEICE, 14(1), 66–75. https://doi.org/10.17148/ijireeice.2026.14111
6. Ansari, Meraj Farheen, and Syed Sharik Ali. "AI-driven zero-trust architecture for enhanced cybersecurity in dynamic network environments." International Journal of

  

Volume: 09, Issue: 04 July - August 2026

ISSN 2582-0176

Innovative Research in Electrical, Electronics, Instrumentation and Control Engineering 13 (2025): 12.

7. Chittoju, Siva Sai Ram, Sireesha Kolla, Mubashir Ali Ahmed, and Abdul Raheman Mohammed. "Synergistic Integration of Blockchain and Artificial Intelligence for Robust IoT and Critical Infrastructure Security."
8. Khader, Shuaib Abdul, Amir Ahmed Ansari, and Syed Sharik Ali. "Zero-Day Exploit Prediction Using Graph-Based Deep Learning on Vulnerability and Threat Intelligence Data."
9. Ansari, Mohammed Azmath, Shaik Aqheel Pasha, and Narendar Kandula. "Reinforcement Learning for Adaptive Network Defense in Financial Institutions."
10. Chittoju, S. R., and Siraj Farheen Ansari. "Blockchain's Evolution in Financial Services: Enhancing Security, Transparency, and Operational Efficiency." International Journal of Advanced Research in Computer and Communication Engineering 13, no. 12 (2024): 1-5.
11. Kashif, Mohammed, Mohammed Aasimuddin, Mubashir Ali Ahmed, Laxmi Bhavani Cheekatimalla, Eraj Farheen Ansari, and Ahwan Mishra. "AI-DRIVEN CTI FOR BUSINESS: EMERGING THREATS, ATTACK STRATEGIES, AND DEFENSIVE MEASURES."
12. RAHEEM, MOHD ABDUL, and MOHAMMED AZMATH ANSARI. "INTELLIGENT AND TRUSTWORTHY 6G: AI-DRIVEN ARCHITECTURES, APPLICATIONS, AND SECURITY FRAMEWORKS.
13. Mohammed, Naveed Uddin, Zubair Ahmed Mohammed, Shravan Kumar Reddy Gunda, Akheel Mohammed, and Moin Uddin Khaja. "Networking with AI: Optimizing Network Planning, Management, and Security through the medium of Artificial Intelligence.
14. Sultana, Ghousia, Siraj Farheen Ansari, Mohammed Imran Ahmed, Abdul Faiyaz Shaik, Moin Uddin Khaja, and Bibhu Dash. "RESPONSIBLE AI ANALYTICS FOR REAL-WORLD IMPACT: NAVIGATING ETHICS, PRIVACY AND TRUST."
15. Integrating Machine Learning and Engineering Management to Optimize Construction Scheduling and Resource Allocation
16. Mohammed, Abdul Faisal, and Mohammed Akifuddin Ghori. "AI-Enhanced Safety for Heavy Load Construction Vehicles: An Integrated Embedded C++ Software Approach."
17. KASHIF, MOHAMMED, ABDUL RAHMAN JIBRAN SYED, and MUBASHIR ALI AHMED. "ADVANCING ANOMALY AND FRAUD DETECTION IN BIG DATA WITH ARTIFICIAL INTELLIGENCE."
18. Wang, Zheng-Xin, Dan-Dan Li, and Hong-Hao Zheng. "Model comparison of GM (1, 1) and DGM (1, 1) based on Monte-Carlo simulation." Physica A: Statistical Mechanics and Its Applications 542 (2020): 123341.
19. Mohammed, Akheel, Zubair Ahmed Mohammed, Naveed Uddin Mohammed, Shravan Kumar Gunda, Mohammed Azmath Ansari, and Mohd Abdul Raheem. "AI-NATIVE WIRELESS NETWORKS: TRANSFORMING CONNECTIVITY, EFFICIENCY, AND AUTONOMY FOR 5G/6G AND BEYOND

  