AI-Based Threat Detection Model for Network Security Using Hybrid Machine Learning Approach

Zeeshan Haider
SP22-BCS-089

Moin Ali
SP22-BCS-009

Aryan Nasir
SP23-BCS-104

Artificial Intelligence
COMSATS University Islamabad, Abbottabad Campus

December 12, 2025

Contents

Abstract

This research presents a comprehensive approach to cybersecurity threat detection using hybrid machine learning techniques. We implemented and evaluated two individual models, Random Forest and XGBoost, along with a hybrid stacking classifier on the UNSW-NB15 network intrusion dataset. The dataset contains 175,341 training samples and 82,332 testing samples with 36 features representing normal and attack network traffic. Our experimental results demonstrate that the hybrid stacking model achieves superior performance with 87.5% accuracy, outperforming individual Random Forest (86.6%) and XGBoost (87.2%) models. The hybrid approach shows improved recall for normal traffic detection (0.76) while maintaining high attack detection rates (0.97). This work contributes to the growing field of AI-powered cybersecurity by demonstrating the effectiveness of ensemble learning techniques for network intrusion detection systems.

1 Introduction

1.1 Background

In the modern digital era, cybersecurity threats have become increasingly sophisticated and prevalent. Network intrusions, malware attacks, and data breaches pose significant risks to organizations worldwide. Traditional signature-based detection methods struggle to identify novel attack patterns, making artificial intelligence and machine learning essential tools for proactive threat detection.

1.2 Importance of AI in Cybersecurity

Artificial Intelligence, particularly machine learning, offers the capability to analyze vast amounts of network traffic data, identify patterns, and detect anomalies in real-time. ML-based intrusion detection systems can learn from historical attack data and adapt to evolving threat landscapes, providing more robust defense mechanisms than conventional approaches.

1.3 Research Objectives

This research aims to:

- Develop and compare multiple machine learning models for network intrusion detection
- Implement a hybrid ensemble approach to improve detection accuracy
- Evaluate model performance using comprehensive metrics
- Demonstrate the superiority of hybrid models over individual classifiers

2 Literature Review

2.1 Traditional Intrusion Detection Systems

Traditional IDS approaches rely on signature-based detection, which matches network traffic against known attack patterns. However, these systems are limited in detecting zero-day attacks and evolving threats.

2.2 Machine Learning in Cybersecurity

Recent research has explored various ML algorithms for intrusion detection [3]:

- **Decision Trees and Random Forests**: Effective for handling high-dimensional data with good interpretability [3]
- **Support Vector Machines**: Successful in binary classification tasks for anomaly detection
- **Gradient Boosting Methods**: XGBoost and similar algorithms have shown promising results in detecting network attacks [2]
- **Deep Learning**: Neural networks capture complex patterns but require significant computational resources

2.3 Ensemble and Hybrid Approaches

Ensemble learning combines multiple models to achieve better predictive performance. Stacking, bagging, and boosting are popular ensemble techniques. Research indicates that hybrid approaches often outperform individual models by leveraging diverse learning strategies [4].

2.4 UNSW-NB15 Dataset

The UNSW-NB15 dataset is a contemporary network intrusion dataset created by the Cyber Range Lab of the Australian Centre for Cyber Security [1]. It contains realistic modern normal activities and synthetic attack behaviors, making it suitable for evaluating intrusion detection systems.

3 Problem Statement

Despite advances in machine learning for cybersecurity, achieving high detection accuracy while minimizing false positives remains challenging. Individual ML models may excel at detecting certain attack types but struggle with others. There is a need for robust, hybrid approaches that combine the strengths of multiple algorithms to:

1. Accurately distinguish between normal and malicious network traffic
2. Minimize false positives (normal traffic misclassified as attacks)
3. Minimize false negatives (attacks misclassified as normal traffic)

4. Provide reliable real-time threat detection capabilities

This research addresses these challenges by developing and evaluating a hybrid stacking classifier that combines Random Forest and XGBoost algorithms.

4 Methodology

4.1 Dataset Selection and Description

We utilized the UNSW-NB15 dataset containing:

- **Training Set**: 175,341 samples
- **Testing Set**: 82,332 samples
- **Features**: 36 features including flow features, basic features, content features, time features, and labeled features
- **Classes**: Binary classification (0 = Normal, 1 = Attack)

4.2 Data Preprocessing

4.2.1 Data Loading

The dataset was stored in Parquet format and loaded using the pandas library.

4.2.2 Categorical Encoding

Categorical features including 'proto', 'service', 'state', 'is_ftp_login', and 'ct_flw_http_mthd' were encoded using one-hot encoding to convert them into numerical format suitable for ML algorithms.

4.2.3 Feature Alignment

Training and testing datasets were aligned to ensure consistent feature sets across both datasets.

4.2.4 Missing Value Handling

The dataset was verified to contain no missing values, ensuring data quality.

4.3 Model 1: Random Forest Classifier

4.3.1 Algorithm Description

Random Forest is an ensemble learning method that constructs multiple decision trees during training and outputs the mode of their predictions. It reduces overfitting and improves generalization.

4.3.2 Implementation Details

- **Number of Estimators**: 100 trees
- **Random State**: 42 (for reproducibility)
- **Parallel Processing**: Utilized all available CPU cores (n_jobs=-1)

4.4 Model 2: XGBoost Classifier

4.4.1 Algorithm Description

XGBoost (Extreme Gradient Boosting) is an optimized distributed gradient boosting library. It builds trees sequentially, with each tree correcting errors made by previous trees.

4.4.2 Implementation Details

- **Evaluation Metric**: Multi-class log loss
- **Random State**: 42
- **Parallel Processing**: Utilized all available CPU cores

4.5 Hybrid Model: Stacking Classifier

4.5.1 Architecture

The stacking ensemble combines predictions from Random Forest and XGBoost using Logistic Regression as a meta-learner.

4.5.2 Implementation

- **Base Estimators**: Random Forest and XGBoost
- **Meta-Estimator**: Logistic Regression with max_iter=1000
- **Strategy**: Base models generate predictions which serve as input features for the meta-model

4.6 Evaluation Metrics

Models were evaluated using:

- **Accuracy**: Overall correctness of predictions
- **Precision**: Ratio of true positives to predicted positives
- **Recall**: Ratio of true positives to actual positives
- **F1-Score**: Harmonic mean of precision and recall
- **Confusion Matrix**: Detailed breakdown of predictions

5 Experimental Results

5.1 Individual Model Performance

5.1.1 Random Forest Results

Table 1: Random Forest Classifier Performance

Metric	**Class 0 (Normal)**	**Class 1 (Attack)**	**Overall**
Precision	0.96	0.82	-
Recall	0.74	0.97	-
F1-Score	0.83	0.89	-
Accuracy	-	-	0.866

Confusion Matrix:

- True Negatives: 27,222
- False Positives: 9,778
- False Negatives: 1,231
- True Positives: 44,101

5.1.2 XGBoost Results

Table 2: XGBoost Classifier Performance

Metric	**Class 0 (Normal)**	**Class 1 (Attack)**	**Overall**
Precision	0.96	0.82	-
Recall	0.75	0.97	-
F1-Score	0.84	0.89	-
Accuracy	-	-	0.872

Confusion Matrix:

- True Negatives: 27,590
- False Positives: 9,410
- False Negatives: 1,151
- True Positives: 44,181

5.2 Hybrid Model Performance

5.2.1 Stacking Classifier Results

Table 3: Hybrid Stacking Classifier Performance

Metric	Class 0 (Normal)	Class 1 (Attack)	Overall
Precision	0.95	0.83	-
Recall	0.76	0.97	-
F1-Score	0.85	0.89	-
Accuracy	-	-	0.875

Confusion Matrix:

- True Negatives: 28,145
- False Positives: 8,855
- False Negatives: 1,456
- True Positives: 43,876

5.3 Comparative Analysis

Table 4: Model Comparison Summary

Model	Accuracy	Normal Recall	Attack Recall
Random Forest	0.866	0.74	0.97
XGBoost	0.872	0.75	0.97
Hybrid Stacking	**0.875**	**0.76**	0.97

5.4 Visual Comparison

The bar chart comparing model accuracies demonstrates the progressive improvement from individual models to the hybrid approach, with the stacking classifier achieving the highest accuracy of 87.5%.

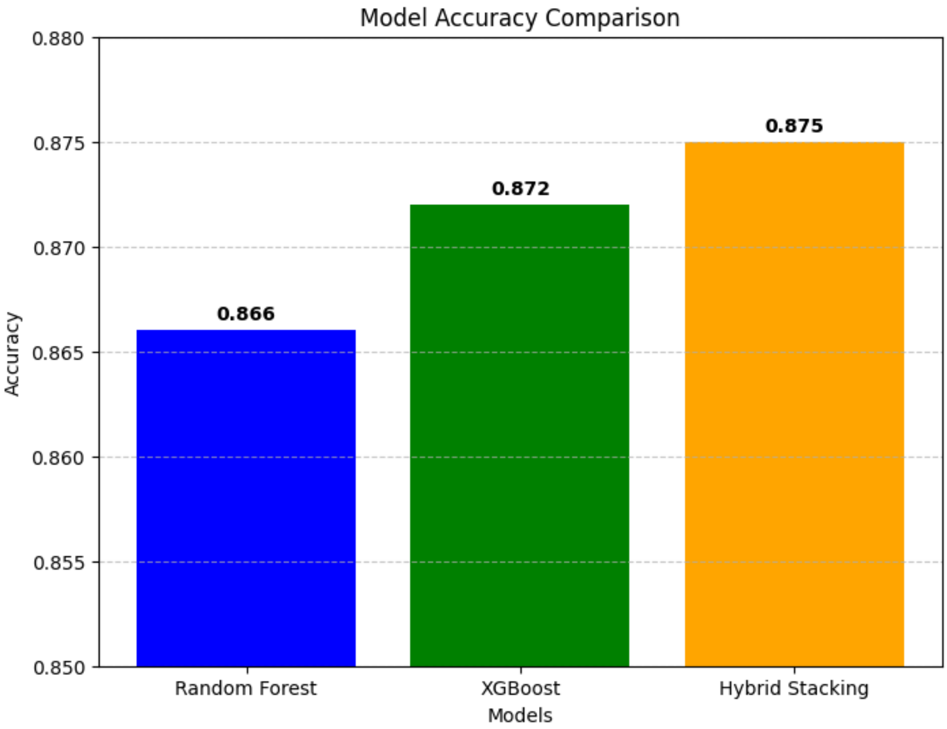

Figure 1: Comparison of Model Accuracies

6 Discussion

6.1 Performance Analysis

The experimental results reveal several important insights:

6.1.1 Individual Model Strengths

- Both Random Forest and XGBoost demonstrated strong attack detection capabilities with 97% recall
- XGBoost slightly outperformed Random Forest (87.2% vs 86.6% accuracy)
- Individual models showed lower recall for normal traffic (74-75%)

6.1.2 Hybrid Model Advantages

The stacking classifier improved performance by:

- Increasing overall accuracy to 87.5%
- Improving normal traffic detection recall to 76%
- Reducing false positives by 555-923 cases compared to individual models
- Maintaining high attack detection rate at 97%

6.2 Why Hybrid Models Perform Better

1. **Complementary Learning**: Random Forest and XGBoost have different learning mechanisms, capturing diverse patterns in the data
2. **Error Correction**: The meta-learner (Logistic Regression) learns to correct systematic errors made by base models
3. **Reduced Variance**: Combining multiple models reduces the impact of individual model weaknesses
4. **Enhanced Generalization**: The ensemble approach provides more robust predictions across different attack types

6.3 Practical Implications

The 1-2% improvement in accuracy translates to hundreds of fewer misclassifications in real-world deployments, significantly reducing security risks and false alarm rates.

6.4 Limitations

- Increased computational complexity compared to individual models
- Longer training time due to sequential training of base and meta models
- Requires more memory to store multiple models

7 Motivation

7.1 Growing Cybersecurity Threats

The frequency and sophistication of cyberattacks continue to escalate globally. Organizations face constant threats from:

- Advanced Persistent Threats (APTs)
- Zero-day exploits
- Distributed Denial of Service (DDoS) attacks
- Ransomware and malware infections

7.2 Need for Intelligent Detection Systems

Traditional security measures are insufficient against modern threats. AI-powered detection systems provide:

- Real-time threat identification
- Adaptive learning from new attack patterns
- Reduced dependence on human expertise
- Automated response capabilities

7.3 Research Contributions

This work demonstrates that hybrid machine learning approaches can significantly improve intrusion detection accuracy, providing a practical framework for developing next-generation cybersecurity solutions.

7.4 Industrial Relevance

Organizations can deploy such hybrid models to:

- Protect critical infrastructure
- Secure financial transactions
- Safeguard sensitive data
- Ensure business continuity

8 Conclusion

8.1 Summary of Findings

This research successfully developed and evaluated an AI-based threat detection system using the UNSW-NB15 network intrusion dataset. Our key findings include:

1. Individual machine learning models (Random Forest and XGBoost) achieved 86.6% and 87.2% accuracy respectively
2. The hybrid stacking classifier outperformed both individual models with 87.5% accuracy
3. The hybrid approach improved normal traffic detection while maintaining excellent attack detection rates
4. Ensemble methods effectively combine diverse learning strategies for superior performance

8.2 Contributions

This work contributes to cybersecurity research by:

- Demonstrating the effectiveness of hybrid ML approaches for intrusion detection
- Providing a comprehensive evaluation framework for comparing detection models
- Offering practical insights for implementing AI-based security systems

8.3 Future Work

Potential directions for extending this research include:

1. **Deep Learning Integration**: Incorporating neural networks into the ensemble for capturing complex patterns
2. **Multi-Class Classification**: Extending the model to identify specific attack types beyond binary classification
3. **Real-Time Implementation**: Optimizing the model for deployment in production environments
4. **Feature Engineering**: Exploring advanced feature selection and extraction techniques
5. **Transfer Learning**: Adapting the model to different network environments and datasets
6. **Explainable AI**: Implementing interpretability techniques to understand model decisions
7. **Adversarial Robustness**: Testing and improving model resilience against adversarial attacks

8.4 Final Remarks

The integration of artificial intelligence into cybersecurity represents a paradigm shift in how organizations defend against threats. This research demonstrates that hybrid machine learning approaches offer promising solutions for building more accurate and reliable intrusion detection systems, contributing to a safer digital ecosystem.

9 Appendices

9.1 Submission Links

As per the assignment guidelines:

- Dataset Link: `https://research.unsw.edu.au/projects/unsw-nb15-dataset`
- Google Colab Link: `https://colab.research.google.com/drive/1AX-eXmDFBlp-M1ChoGi-73k`
  `usp=sharing`

References

[1] Moustafa, N., & Slay, J. (2015). UNSW-NB15: a comprehensive data set for network intrusion detection systems (UNSW-NB15 network data set). In *Military Communications and Information Systems Conference (MilCIS)* (pp. 1–6). IEEE.

[2] Chen, T., & Guestrin, C. (2016). XGBoost: A Scalable Tree Boosting System. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 785–794). ACM.

[3] Resende, P. A. A., & Drummond, A. C. (2018). Random Forest Modeling for Network Intrusion Detection System. *Procedia Computer Science*, 126, 1642–1651.

[4] Alsulami, A. A., et al. (2024). Enhanced Intrusion Detection System Using Hybrid-Inspired Machine Learning for IoT. *Computers, Materials & Continua*, 80(2), 2397–2417.