Title: Intrusion Detection Prediction Using Random Forest and Support Vector Machine Learning Algorithms

Contents

1. Abstract

Intrusion Detection Systems (IDS) play a critical role in contemporary cyber security, assisting in the detection and prevention of malicious behaviour. This research suggests IDS based on machine learning using Random Forest (RF) and Support Vector Machine (SVM) to improve detection rates, flexibility, and minimize false positives. We anticipate that this method will exhibit better performance than baseline IDS models, which will be empirically tested. Moreover, this research seeks to enhance model transparency by integrating Explainable AI (ExAI) methods to increase interpretability. The results of this research will help enhance IDS efficiency, making them more trustworthy for real-world applications in cyber security.

2. Introduction

2.1 Background

This research suggests an Intrusion Detection System (IDS) based on machine learning techniques employing Random Forest (RF) and Support Vector Machine (SVM) in order to upgrade cyber threat identification with enhanced precision and interpretability. Through combining feature selection mechanisms (RFE & PCA) and Explainable AI (SHAP & LIME), the model provides efficient detection and interpretability in the decision-making process (Band, 2023).

In the fast-changing field of information technology, protecting digital assets and sensitive data is now a top priority. The growing complexity and frequency of cyber threats have made network security an essential focus of research. The main importance of Intrusion Detection Systems (IDS) is its proactive approach. IDS proactively monitors networks and systems in real time instead of only reacting to security events after they happen. IDS are crucial for identifying and mitigating unauthorized access attempts and malicious activities within a network(Bejtlich, 2004).

While conventional ML-based IDS models prioritize accuracy, they tend to neglect adversarial vulnerability and real-time adaptability. Existing research has demonstrated that adversarial attacks can lead to serious performance deterioration of IDS, making them vulnerable to evasion attack (Alhajjar, Maxwell and Bastian, 2021). Explainable AI methods, including SHAP and LIME, have been investigated to enhance interpretability in IDS models, while adversarial learning approaches have been proposed to enhance model resilience against sophisticated threats. Traditional IDS typically depend on predefined rules and signature-based detection methods, which often struggle to adapt to new and evolving threats.

2.2 Problem Statement

The integrity and security of modern computer networks face significant challenges due to the rapid rise of cyber threats. Traditional signature-based Intrusion Detection Systems (IDS) often fall short against sophisticated attacks, resulting in a high number of false negatives and slow response times. Moreover, the sheer volume of network data complicates real-time anomaly detection(García-Teodoro et al., 2009).

Machine Learning (ML) presents a promising solution to enhance the effectiveness of IDS, but selecting the right algorithm can be difficult. This study suggests an advanced Intrusion Detection System (IDS) by incorporating feature selection methods with Random Forest (RF) and Support Vector Machine (SVM) for enhanced detection precision and minimized false alarms. The research also examines the robustness of these models against adversarial attacks, testing their flexibility in adapting to novel threats.

2.3 Objectives

- Improve IDS performance using sophisticated feature selection methods like Recursive Feature. Elimination (RFE) and Principal Component Analysis (PCA) to enhance the accuracy of detection and decrease the computational complexity.
- Improve model interpretability by incorporating Explainable AI (ExAI) methods to offer transparency in IDS decision-making.
- Compare and assess the performance of Random Forest (RF) and Support Vector Machine (SVM) with baseline IDS models on common intrusion datasets like NSL-KDD and CIC-IDS 2017

2.4 Significance

This study contributes to Intrusion Detection Systems (IDS) by addressing major limitations of conventional machine learning models. In contrast to standard studies concerned with detection accuracy only, this study brings in feature selection techniques to increase computational efficiency and enhance classification performance(Mwadulo, 2016).

In addition, the real-time flexibility of IDS is tested, such that the models can efficiently identify new and changing threats within dynamic network systems. This field application makes the research more viable for real-world security solutions. To support trust and transparency, Explainable AI (EXAI) methods like SHAP and LIME are incorporated, rendering IDS's decision-making process more explainable to security analysts(Mohale &Obagbuwa, 2025).

2.5 Scope and Limitation

This research improves Intrusion Detection System (IDS) effectiveness through the incorporation of feature selection methods to maximize performance. The study tests the adversarial robustness of Random Forest (RF) and Support Vector Machine (SVM) to attacks to make them resistant to changing threats. The study further analyses the real-time flexibility of these models within changing network scenarios.

The research suffers from computational complexity because feature selection algorithms can make the processing time larger, which could affect real-time performance. The second limitation is that of model generalizability since the results obtained are specific to Random Forest (RF) and Support Vector Machine (SVM) and might not generalize to any other machine learning algorithm. Although SHAP and LIME promote explain ability, there are trade-offs since they fail to represent intricate patterns in IDS classification. Finally, real-world deployment limitations create challenges, especially for high-traffic networks where scalability and efficiency become paramount (Thakkar and Lohiya, 2022).

2.6 Research Questions

- How does Explainable AI (ExAI) enhance the interpretability and transparency of IDS decision-making with the Random Forest (RF) and Support Vector Machine (SVM) algorithms?
- Which feature selection methods, including Recursive Feature Elimination (RFE) and Principal Component Analysis (PCA), will make the highest contribution to increasing IDS accuracy without sacrificing computational complexity?

- What are the advantages and limitations of RF and SVM-based IDS models compared to other conventional intrusion detection methods?

2.7 Explanation

1. How does Explainable AI (ExAI) enhance the interpretability and transparency of IDS decision-making with the Random Forest (RF) and Support Vector Machine (SVM) algorithms?

Explainable AI (ExAI) improves Intrusion Detection System (IDS) decision-making interpretability and transparency using Random Forest (RF) and Support Vector Machine (SVM) algorithms in the following manners:

- Feature Contribution Analysis – SHAP (SHapley Additive Explanations) measures the contribution of each feature to RF and SVM predictions so that security analysts can see why an alert was generated (Ahmed, 2024).

- Local Interpretability – LIME (Local Interpretable Model-Agnostic Explanations) produces interpretable approximations of difficult model decisions, assisting in case-by-case attack analysis (Ahmed, 2024).

- Identifying Decision Boundaries – In SVM, ExAI points out how support vectors affect classification, explaining how the model discriminates between normal and malicious traffic (Trivedi, 2024).

- Rule-Based Interpretability – For RF, ExAI can pull out decision paths from trees, which simplifies the understanding of why particular rules identify traffic as an attack.

- Trust and Usability – By explaining RF and SVM decisions more clearly, ExAI builds trust in the IDS so that cyber security experts can verify and refine detection methods.

- Minimizing False Positives – ExAI insights can be leveraged by analysts to optimize feature selection and model training, enhancing IDS accuracy and reducing false positives (Saraswat, 2022).

2.Which feature selection methods, including Recursive Feature Elimination (RFE) and Principal Component Analysis (PCA), will make the highest contribution to increasing IDS accuracy without sacrificing computational complexity?

Feature selection techniques, such as Recursive Feature Elimination (RFE) and Principal Component Analysis (PCA), help to improve IDS accuracy while maintaining computational efficiency in the following manners:

1. Recursive Feature Elimination (RFE):
   - Gradually eliminates less important features, keeping only the most important ones.
   - Functions nicely with Support Vector Machine (SVM) and Random Forest (RF) as it is used to optimize subsets of features for improved classification (Boateng, 2020).
   - Tackles over fitting, ensuring more accurate results at the cost of interpretability.
   - Ideally works when computer efficiency is not a top constraint but there is an emphasis on accuracy.

2. Principal Component Analysis (PCA):
   - Minimizes dimensionality by converting correlated features to a lower set of uncorrelated principal components.
   - Enhances IDS performance by minimizing computational overhead without appreciable loss in accuracy.
   - Suits best high-dimensional data such as NSL-KDD and CIC-IDS 2017, where there is high redundancy of features (Bakro, 2024).

3. What are the advantages and limitations of RF and SVM-based IDS models compared to other conventional intrusion detection methods?

Advantages and Disadvantages of RF and SVM-Based IDS Models over Traditional IDS

Advantages:

1. More Accuracy & Detection Rates
   - Random Forest (RF): Efficient with large datasets, minimizes overfitting, and offers excellent detection accuracy.
   - Support Vector Machine (SVM): Well-suited in differentiating between normal and suspicious traffic, particularly in binary classifying tasks (Odera, 2023).

2. Feature Selection & Interpretability
   - RF: Offers feature importance scores, thereby simplifying interpretation of the most important indicators of attacks (Maldonado, 2022).
   - SVM: Exemplifies great compatibility with feature selection methods like Recursive Feature Elimination (RFE) for best results.

3. Better Generalization

RF and SVM both are good across various IDS datasets (NSL-KDD, CIC-IDS 2017), varying with multiple attack patterns.

4. Overfitting Resistant

- RF: Comprises multiple decision trees and thus is not so overfitting prone.
- SVM: Is effective with kernel functions, leading to improved decision boundaries.

5. Explain ability with ExAI Integration
SHAP and LIME can be used in RF and SVM to improve IDS transparency and decision-making for security analysts (Hassan, 2023).

Limitations:
1. Computational Complexity:

- RF: Slow when handling very large datasets because of the use of many decision trees.
- SVM: Computationally intensive for high-dimensional data, particularly in multi-class classification ({Pappu, 2014).

2. Sensitivity to Hyper parameters:
RF and SVM need parameter tuning (Grid Search, Cross-Validation) to perform the best, hindering deployment.
3. Limited against Changing Attacks:
Deep learning-based IDS are different in the sense that RF and SVM could have trouble handling *zero-day attacks necessitating adaptive learning.
4. Heavy Reliance on Feature Engineering:
Performance highly relies on feature selection techniques (RFE, PCA) to achieve best accuracy without unwarranted complexity.

3. Literature Review

Intrusion Detection Systems (IDS) have undergone a tremendous change with the adoption of machine learning (ML) methods for the detection and prevention of cyber-attacks. Recent research emphasizes that deep learning-based IDS models have exhibited high accuracy in attack detection; however, they are plagued by issues like high computational expenses, lack of interpretability, and real-time deployment difficulties (Moustafa, 2023).

3.1 Machine Learning in IDS

Conventional IDS models are based on rule-based detection techniques, which usually fail to deal with new or emerging threats. ML-based IDS, such as Random Forest (RF) and Support Vector Machine (SVM), have been extensively researched for their effectiveness in anomaly detection (Azam, 2023). These approaches provide strong classification performance, but it is vulnerable to bad feature selection and poor-quality datasets.

3.2 Feature Selection Methods for IDS

High-dimensional data in IDS can lead to increased computational costs and model inefficiency. Techniques such as Recursive Feature Elimination (RFE) and Principal Component Analysis (PCA) help reduce dimensionality while preserving critical information for intrusion detection (Tufail, 2025). This study focuses on evaluating these feature selection methods to enhance IDS performance.

3.3 Explainable AI (ExAI) in Cyber security

One major disadvantage of most ML-based IDS models is their black-box nature, which does not allow security analysts to understand decisions. Explainable AI (ExAI) methods, including SHAP (SHapley Additive Explanations) and LIME (Local Interpretable Model-agnostic Explanations), seek to enhance transparency in IDS decision-making (Saraswat, 2022). This work incorporates ExAI to promote the interpretability of RF and SVM-based IDS.

3.4 Gap Analysis: Bridging the Shortcomings of Deep Learning-Based IDS

Whereas deep learning approaches like Convolution Neural Networks (CNNs) and Recurrent Neural Networks (RNNs) proved to be highly effective in IDS, they do so at the cost of large datasets, high computational capabilities, and absence of interpretability. This work fills the gap by utilizing RF and SVM using ExAI for an efficient, interpretable, and computationally viable IDS solution.

---

4. Research Methodology

This paper introduces a machine learning-based Intrusion Detection System (IDS) based on Random Forest (RF) and Support Vector Machine (SVM) for detecting intrusions. The approach is designed into five fundamental phases: data collection and pre-processing, feature selection, model development, performance evaluation, and integration of Explainable AI (ExAI).

4.1 Data Collection and Preprocessing

Datasets:
This research will use some publicly accessible datasets of IDS, such as:

- NSL-KDD: A cleaned version of the KDD'99 dataset, tailored to remove redundancy problems.

- CIC-IDS 2017: A modern intrusion detection dataset incorporating real-world attack scenarios (Hewapathirana, 2025).

Data Preprocessing:

To facilitate quality input for model training, the following preprocessing will be done:

- Data Cleaning: Missing values will be handled via mean/mode imputation.
- Categorical Encoding: Transforming categorical variables into a numerical representation.
- Feature Scaling: Normalizing or standardizing numerical features to improve model performance.

4.2 Feature Selection Approaches

In order to improve IDS effectiveness, two feature selection methods will be used:

- Recursive Feature Elimination (RFE): An iterative algorithm which orders features in terms of importance and eliminates the least important ones.
- Principal Component Analysis (PCA): A feature reduction method that converts features into principal components holding key variance.
- These approaches tend to minimize computational complexity while enhancing detection accuracy.

4.3 Model Building and Training

Algorithm Selection:

Two supervised machine learning algorithms shall be used:

- Random Forest (RF): Ensemble learning algorithm based on decision trees that is highly accurate and robust.
- Support Vector Machine (SVM): A strong classifier that performs well in high-dimensional spaces.

Hyper parameter Tuning:

Grid Search and Cross-Validation algorithms will be used to maximize model performance. These algorithms will determine the optimal set of hyperparameters for RF and SVM.

Training Process:

- The dataset will be split into 80% training and 20% testing sets.
- The models will be trained on the training set and tested on the test set.

4.4 Performance Measurement Metrics

The performance of the IDS models will be evaluated with several evaluation measures:

- Accuracy: Measures of overall correctness.
- Precision, Recall, and F1-Score: Measures how well the model correctly classifies attacks, especially for imbalanced datasets.
- ROC-AUC Score: Evaluates how well the classifier can distinguish normal traffic from malicious traffic.

These metrics will give a complete assessment of the strengths and weaknesses of the models.

4.5 Explainable AI (ExAI) Integration

The main problem with ML-based IDS is the non-explainability. To overcome this, Explainable AI (ExAI) methods will be integrated:

- SHAP (SHapley Additive Explanations): Gives insight into feature contributions to model predictions.
- LIME (Local Interpretable Model-agnostic Explanations): Produces interpretable explanations for predictions.

These methods will increase transparency, allowing security analysts to comprehend and believe in model decisions (Kanti, 2024).

6. Comparison to Baseline Models

In order to determine the effectiveness of the proposed IDS, its performance will be compared to:

- Conventional rule-based IDS (e.g., Suricata, Snort).
- Other machine learning-driven IDS include Decision Trees and K-Nearest Neighbours (KNN) (Hussein, 2024).

This comparison will identify the strengths and possible weaknesses of RF and SVM-based IDS models.

---

5. Cyber Security Body of Knowledge (CYBOK) Alignment

The suggested Intrusion Detection System (IDS) based on Random Forest (RF) and Support Vector Machine (SVM) has alignment with various domains of the Cybersecurity Body of Knowledge (CyBOK). This part of the section identifies significant areas of alignment as per the research methodology and objectives.

5.1 Network Security

The research uses machine learning-based IDS to improve the detection of cyber threats within network traffic.

NSL-KDD and CIC-IDS 2017 datasets are utilized to simulate real-world attacks like DoS, malware, and intrusion attempts.

The system enhances real-time threat detection and response strategies against dynamic cyber threats.

5.2 Security Operations & Incident Management

The projected IDS assist security analysts with automated threat detection and actionable intelligence.

Explainable AI (ExAI) methods (SHAP and LIME) improve interpretability, allowing security teams to have faith and authenticate ML-driven decisions (Sommer, 2010).

The system is an input to threat intelligence and proactive incident response plans.

5.3 Artificial Intelligence & Cybersecurity

This research investigates ML-based intrusion detection, making use of feature selection algorithms (RFE & PCA) to improve the accuracy of detection (Verma, 2020).

Hyperparameter tuning (Grid Search, Cross-Validation) increases model strength.

Comparison with conventional IDS and ML-based IDS guarantees the performance of AI in cybersecurity deployments.

5.4 Risk Management & Security Assurance

The suggested IDS counteracts cybersecurity threats by detecting threats with high precision and recall.

ROC-AUC analysis assesses detection effectiveness, minimizing false positives and negatives.

Model efficiency is promoted by feature selection methods (*RFE & PCA*) to bring about a harmony between system performance and security.

5.5 Human Factors & Explainable AI in Cybersecurity

Integration of Explainable AI (ExAI) provides transparency in ML-based IDS.

SHAP and LIME give us an understanding of why the model is classifying some network behaviours as attacks (Samek, 2021).

Increasing IDS explainability enhances trust, usability, and decision-making among cybersecurity experts.

6. Expected Outcomes

1. Improved Intrusion Detection Accuracy – Hyperparameter optimization (RFE & PCA) and feature selection will improve detection accuracy while lowering computational complexity.

2. Better Interpretability using Explainable AI (ExAI) – Combination of SHAP and LIME will add transparency, allowing security analysts to comprehend model choices.

3. Comparison with Conventional IDS – The new IDS will be more accurate and efficient compared to rule-based and other ML-based IDS (e.g., Decision Trees, KNN).

4. Cybersecurity Body of Knowledge (CyBOK) Contribution – The study contributes to major CyBOK domains, specifically in AI-based security, risk management, and network security.

5. Real-World Relevance – The NSL-KDD and CIC-IDS 2017 datasets will be tested to assure effectiveness against new cyber threats.

6. Deployable IDS Solution – Real-time security monitoring and anticipatory cyber defense approaches will be supported by the model.

Conclusion

The work suggests an effective, understandable, and high-performance Intrusion Detection System (IDS) incorporating Random Forest (RF) and Support Vector Machine (SVM) for improved cyber threat detection. Implementing Recursive Feature Elimination (RFE) and Principal Component Analysis (PCA) helps in feature reduction, optimizing the accuracy of the model with reduced computational complexity. Hyperparameter tuning (Grid Search & Cross-Validation) also improves the detection performance.

To overcome the model interpretability challenge, Explainable AI (ExAI) methods, such as SHAP and LIME, are incorporated, enabling security analysts to comprehend and rely on the system's outputs. The envisioned IDS is compared with conventional rule-based systems (e.g., Snort, Suricata) and other ML-based IDS models (e.g., Decision Trees, KNN) to ensure an extensive performance analysis*.

Timeline and Project Plan

	February	March	April	May	June	July	August
Proposal Preparation & Submission	🟩	🟩					
	Meeting 1	Meeting 2					
Conduct Literature Review			🟩				
			Meeting 3				
Data Acquisition and Preprocessing				🟩			
				Meeting 4			
Model Implementation and Training					🟩		
					Meeting 5		
Hyperparameter Tuning and Evaluation						🟩	
						Meeting 6	
Thesis Writing and Revision							🟩
							Meeting 7

Meeting 1	18-Feb-2025
Meeting 2	5-Mar-2025
Meeting 3	Apr-2025
Meeting 4	May-2025
Meeting 5	June-2025
Meeting 6	July-2025
Meeting 7	Aug-2025

References

{Pappu, V. a. P. P. M., 2014. High-dimensional data classification. *Clusters, Orders, and Trees: Methods and Applications: In Honor of Boris Mirkin's 70th Birthday},* pp. 119--150.

Ahmed, S. a. K. M. S. a. H. M. S. a. A. K., 2024. A comparative analysis of lime and shap interpreters with explainable ml-based diabetes predictions. *IEEE Access.*

Ahmed, U. a. J. Z. a. A. A. a. S. M. a. R. A. U. a. S. M. a. C. J., 2024. Hybrid bagging and boosting with SHAP based feature selection for enhanced predictive modeling in intrusion detection systems. *Scientific Reports,* p. 30532.

Azam, Z. a. I. M. M. a. H. M. N., 2023. Comparative analysis of intrusion detection systems and machine learning-based model analysis through decision tree. *IEEE Access,* pp. 80348--80391.

Bakro, M. a. K. R. R. a. H. M. a. A. Z. a. A. A. a. Y. S. I. a. A. M. N. a. P. N., 2024. Building a cloud-IDS by hybrid bio-inspired feature selection algorithms along with random forest model. *IEEE Access,* pp. 8846--8874.

Band, S. S. a. Y. A. a. H. C.-C. a. B. M. a. S. M. a. A. R. a. D. I. a. C. A. T. a. L. H.-W., 2023. Application of explainable artificial intelligence in medical health: A systematic review of interpretability methods. *Informatics in Medicine Unlocked,* p. 101286.

Boateng, E. Y. a. O. J. a. A. D. A., 2020. Basic tenets of classification algorithms K-nearest-neighbor, support vector machine, random forest and neural network: A review. *Journal of Data Analysis and Information Processing,* pp. 341--357.

Hassan, F. a. Y. J. a. S. Z. S. a. M. A. H. a. A. N., 2023. Developing Transparent IDS for VANETs Using LIME and SHAP: An Empirical Study.},. *Computers, Materials \& Continua}.*

Hewapathirana, I. U., 2025. A Comparative Study of Two-Stage Intrusion Detection Using Modern Machine Learning Approaches on the CSE-CIC-IDS2018 Dataset. *Knowledge,* p. 6.

Hussein, S. M. a. A. A. M., 2024. Machine Learning-Driven Intrusion Detection Systems: Reducing False Alarms and Enhancing Accuracy. *{EURASIAN JOURNAL OF SCIENCE AND ENGINEERING,* pp. 85--96.

Islam, M. T. a. S. M. K. a. R. M. G. a. D. D., 2024. Bridging the gap: advancing the transparency and trustworthiness of network intrusion detection with explainable AI}. *International Journal of Machine Learning and Cybernetics,* pp. 5337--5360.

Kanti, P. K. a. S. P. a. W. V. V. a. S. N. M., 2024. Explainable machine learning techniques for hybrid nanofluids transport characteristics: an evaluation of shapley additive and local interpretable model-agnostic explanations. *Journal of Thermal Analysis and Calorimetry,* pp. 11599--11618.

Maldonado, J. a. R. M. C. a. N. B., 2022. A review of recent approaches on wrapper feature selection for intrusion detection. *Expert Systems with Applications,* p. 116822.

Moustafa, N. a. K. N. a. K. M. a. Z. A. Y. a. T. Z., 2023. Explainable intrusion detection for cyber defences in the internet of things: Opportunities and solutions. *IEEE Communications Surveys \& Tutorials,* pp. 1775--1807.

Odera, D. a. O. G., 2023. A comparative analysis of recurrent neural network and support vector machine for binary classification of spam short message service}. *World Journal of Advanced Engineering Technology and Sciences,* pp. 127--152.

Samek, W. a. M. G. a. L. S. a. A. C. J. a. M. K.-R., 2021. Explaining deep neural networks and beyond: A review of methods and applications. *Proceedings of the IEEE,* pp. 247--278.

Saraswat, D. a. B. P. a. V. A. a. P. V. K. a. T. S. a. S. G. a. B. P. N. a. S. R., 2022. Explainable AI for healthcare 5.0: opportunities and challenges. *IEEe Access*, pp. 84486--84517.

Sommer, R. a. P. V., 2010. *2010 IEEE symposium on security and privacy.* s.l.:IEEE.

Trivedi, C. a. B. P. a. P. V. K. a. P. V. a. S. A. a. T. S. a. S. R. a. A. S. a. P. G. a. S. G., 2024. Explainable AI for Industry 5.0: vision, architecture, and potential directions. *IEEE Open Journal of Industry Applications.*

Tufail, S. a. I. H. a. T. M. a. S. A., 2025. A Hybrid Machine Learning-Based Framework for Data Injection Attack Detection in Smart Grids Using PCA and Stacked Autoencoders. *IEEE Access.*

Verma, A. a. R. V., 2020. Machine learning based intrusion detection systems for IoT applications}. *Wireless Personal Communications*, pp. 2287--2310.