STM JOURNALS

Journal of
Network Security

https://journals.stmjournals.com/jons

ISSN: 2395-6739 (Online)
ISSN: 2321-8517 (Print)
Volume 13, Issue 3, 2025
September–December
DOI (Journal): 10.37591/JoNS

Review JoNS

A Survey on Ensemble Technique for Enhanced Cyberattack Detection

Devashish M. Patil[1,*], Darshan B. Khairnar[1], Atharva B. Jagzap[1], Anand N. Gharu[2]

Abstract

It is now more difficult than ever to safeguard enterprises against cyberattacks due to their fast growth and growing sophistication. Stronger cyberattack detection systems are becoming more and more necessary as hostile strategies continue to evolve in order to safeguard information, preserve corporate trust, and protect sensitive data. An overview of contemporary detection techniques is given in this study, with a focus on integrating machine learning (ML) to increase efficacy. A thorough analysis of conventional methods is conducted, including hybrid anomaly-based detection and signature-based detection, pointing out both their benefits and drawbacks. Although these techniques have been useful in recognizing known dangers, they frequently fail to identify new or developing assaults, exposing businesses to serious risks. A useful remedy for these drawbacks is offered by machine learning approaches, which provide the capacity to continuously learn, adapt, and recognize intricate attack patterns that traditional systems could miss. Additionally, ML-based models increase detection speed and accuracy while reducing human error. The operational issues with conventional systems are also covered in the study, along with cutting-edge solutions to these problems. Organizations may greatly improve cybersecurity resilience and adjust to the ever-changing threat landscape by utilizing machine learning.

Keywords: Ensemble learning, classification, machine learning, feature extraction, adaptive learning, random forest

INTRODUCTION

Ensemble Machine Learning (ML) approaches are becoming increasingly essential for enhancing cyberattack detection, given the growing complexity and variety of cyber threats. As cyber adversaries adopt more advanced and evasive tactics, traditional detection methods often struggle to detect a broad range of attacks, especially novel ones. This study explores the potential of ensemble ML methodologies to address these challenges by combining multiple models, resulting in a more accurate and resilient detection system.

***Author for Correspondence**
Devashish M. Patil
E-mail: devashishp.comp_ioe@bkc.met.edu

[1]Student, Department. of Computer Engineering, MET Bhujbal Knowledge City Institute of Engineering, Savitribai Phule Pune University, Nashik, Maharashtra, India
[2]Assistant Professor, Department. of Computer Engineering, MET Bhujbal Knowledge City Institute of Engineering, Savitribai Phule Pune University, Nashik, Maharashtra, India

Received Date: June 12, 2025
Accepted Date: September 05, 2025
Published Date: September 17, 2025

Citation: Devashish M. Patil, Darshan B. Khairnar, Atharva B. Jagzap, Anand N. Gharu. A Survey on Ensemble Technique for Enhanced Cyberattack Detection. Journal of Network Security. 2025; 13(3): 50–54p.

Traditional detection systems, such as signature-based and anomaly-based methods, typically rely on predefined patterns, making them less effective against unknown or rapidly evolving threats. Ensemble methods, such as Support Vector Machines (SVM), Naïve Bayes, and Random Forest, can integrate these diverse models to improve detection accuracy by leveraging the strengths of each algorithm. This integrated approach is typically achieved through techniques such as majority voting, where the final classification decision is based on the majority output of individual models. By combining predictions, ensemble methods compensate for the limitations of individual models and improve overall performance.

Ensemble methods also offer the advantage of reducing false positives, a common issue in traditional detection systems that can lead to alert fatigue. By cross-validating results from multiple models, ensemble methods enhance the precision of threat detection, improving efficiency. Moreover, ensemble systems are more adaptive, enabling them to detect both known and previously unseen attack patterns.

Feature engineering plays a vital role in ML-based detection systems. While traditional ML models rely on manual feature extraction, deep learning (DL) models can automatically uncover complex patterns from raw data, making them more efficient at recognizing subtle relationships. Emphasizing both manual and automated feature extraction, we propose an integrated framework that leverages domain-specific knowledge along with automated feature learning to improve detection capabilities.

To handle evolving threats, it is crucial to periodically retrain models or use adaptive learning techniques. Retraining the models with new data ensures that the detection system remains effective against newly emerging attack patterns.

LITERATURE REVIEW

1. *Cybersecurity in the Digital Sector:* Discuss the growing importance of cybersecurity in protecting sensitive data, highlighting its relevance across various industries. Effective cybersecurity practices are essential to prevent unauthorized access and safeguard information from malicious entities [1].
2. *Social Engineering in Cybersecurity: Mechanisms, Vulnerabilities, and Attacks:* Analyze the dynamics of social engineering attacks, examining how manipulation and deception exploit human vulnerabilities to bypass security measures. Their work identifies critical human factors and provides valuable insights for enhancing security awareness [2].
3. *The Concept of Social Engineering in Cybersecurity:* Revisits the conceptual definition of social engineering in the context of cybersecurity, providing a refined framework for understanding these attacks. It also addresses the challenges of inconsistent definitions and overgeneralization in current literature [3].
4. *Legal Framework for Cybersecurity in AI Technology:* Examine the regulatory and legal challenges in cybersecurity, particularly in the application of artificial intelligence (AI). They highlight the need for updated legal frameworks to counter the rising volume and sophistication of cyber threats [4].
5. *Cybersecurity for Embedded Systems:* The challenges of securing embedded systems, which are widely used in critical infrastructures. Their work identifies key factors affecting the security of these systems, emphasizing the need for lightweight yet robust security mechanisms [5].
6. *Big Data in Cybersecurity: Applications and Trends:* Explores the intersection of big data and cybersecurity, showcasing how large-scale data analysis can enhance threat detection. The paper discusses various applications, including intrusion detection and malware identification, and highlights future research directions [6].
7. Cyberattacks discussed in earlier studies need to be considered [7–9].

PROPOSED SYSTEM

The proposed system integrates several key features to improve security and usability (Figure 1). It incorporates user authentication for secure access and automated data preprocessing to ensure high-quality data for analysis. The system utilizes multiple classifiers (SVM, Naïve Bayes, and Random Forest) to enhance detection accuracy and provides real-time threat detection to minimize response times. Furthermore, the user interface offers intuitive visualizations and actionable insights [10–13].

1. *User interaction (registration and login):* Users must register and log in to access the system, ensuring secure and authorized access. After logging in, users can upload network traffic data in CSV format for analysis.

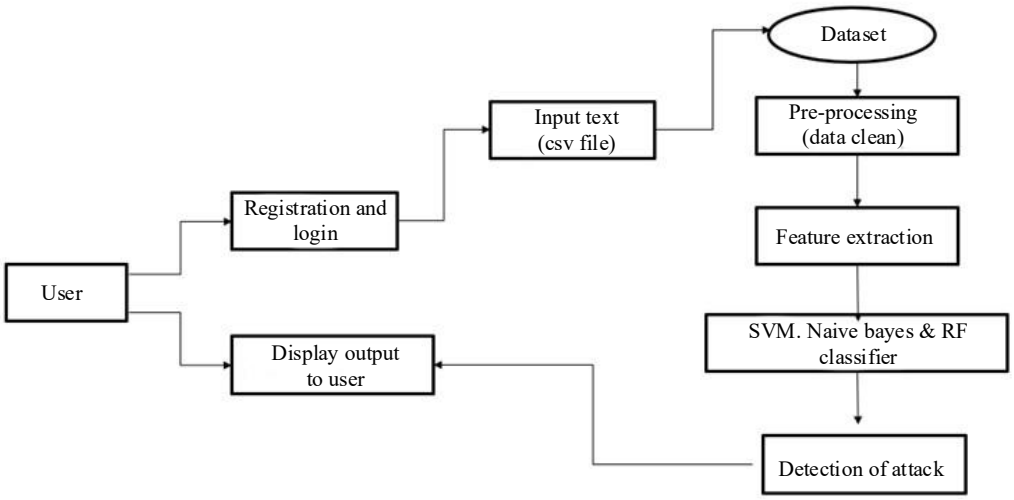

Figure 1. Proposed system.

2. *Dataset Preprocessing:* The uploaded dataset undergoes cleaning to remove inconsistencies, missing values, and irrelevant information. Techniques like normalization and scaling are applied to prepare the data for analysis.
3. *Feature extraction*: Feature extraction methods such as Principal Component Analysis (PCA), Chi-Square tests, and correlation-based methods are employed to identify key attributes from the dataset. These features are then used as inputs for classification models.
4. *Model training and testing (SVM, Naïve Bayes, and Random Forest):* The cleaned and processed data is divided into training and testing sets. Machine learning classifiers such as SVM, Naïve Bayes, and Random Forest are used to train models for detecting both normal and malicious network activities. The models are integrated using ensemble learning techniques such as majority voting to improve detection accuracy.
5. *Attack detection*: After training, the models are applied to incoming data to classify it as either normal or malicious. If an attack is detected, it is flagged for further action by the network administrator.
6. *User output display*: The system displays detection results, including attack types, confidence scores, and visual representations of network security.

ALGORITHMS

Support Vector Machine (SVM)

- *Import libraries*: Use Numpy, Pandas, and Matplotlib for data manipulation and visualization. The scikit-learn library offers SVM functionality for classification.
- *Dataset preparation*: Load the dataset using pandas.read_csv() into a DataFrame. Split the data into features (X) and target (y). Remove missing values using dropna() to ensure clean data.
- *Model training*: Initialize an SVM classifier with a linear kernel and train using fit() on the training set. This model finds the best hyperplane to separate classes.
- *Prediction and evaluation*: Use the trained model to predict target values for the test data and evaluate performance [14–17].

Random Forest (RF)

- *Import libraries*: Similar to SVM, use Pandas, Numpy, and Matplotlib. The Sklearn library is used for Random Forest.
- *Dataset preparation*: Load the dataset into a DataFrame and clean it by removing missing values. Split the data into features (X) and target (y).
- *Model training*: Train a Random Forest classifier with a specified number of trees (n_estimators). The classifier creates multiple decision trees to learn data patterns.
- *Prediction and evaluation*: After training, predict target labels for the test set and assess accuracy.

Naïve Bayes

- *Import libraries*: Use Pandas, Numpy, Matplotlib, and Sklearn for loading the dataset, data manipulation, and modeling.
- *Dataset preparation*: Load the data from a CSV file, clean it by removing rows with missing values, and split it into features (X) and target (y).
- *Model training*: Train a Naïve Bayes classifier (GaussianNB, MultinomialNB, or BernoulliNB depending on the data type) to learn the probability distributions between features and the target.
- *Prediction and evaluation*: Predict target values for the test data and evaluate performance based on accuracy.

ADVANTAGES

1. *Improved performance*: By combining multiple models, ensemble methods can significantly improve detection accuracy and reduce false positives.
2. *Robustness to noise*: Ensemble methods are less sensitive to data anomalies, making them more resilient in varied conditions.
3. *Adaptability to new threats*: Periodically retraining models or using adaptive learning techniques ensures that ensemble methods can adapt to new attack patterns.
4. *Scalability*: Ensemble systems can handle large volumes of data, making them suitable for deployment in large-scale organizations.

LIMITATIONS

1. *Complexity and resource demands*: Implementing and fine-tuning ensemble models can be computationally intensive and resource-demanding.
2. *Overfitting risk*: If not properly managed, ensemble models may overfit the training data, reducing their generalization ability.
3. *Higher computational cost*: Ensemble methods require more memory and processing power than single models, which can be a limitation in resource-constrained environments.
4. *Interpretability challenges*: Ensemble models can be difficult to interpret, posing a challenge in security contexts where transparency is crucial.

CONCLUSION

In conclusion, the integration of machine learning (ML) models, particularly ensemble techniques, significantly enhances cybersecurity systems by leveraging predictive modeling and advanced data analysis. These models effectively identify subtle patterns and anomalies that traditional systems may overlook. By continuously adapting to evolving threats through retraining or adaptive learning techniques, ML-powered systems can detect both known and unknown cyberattacks in real-time, allowing for faster response and mitigation. Machine learning automation helps large organizations manage the growing volume of data while enabling security professionals to focus on strategic decision-making. The integration of ML models into cybersecurity frameworks represents a major advancement in strengthening organizational security by proactively addressing both current and future threats.

REFERENCES

1. Wang W, Lu Z. Cyber security in the smart grid: Survey and challenges. Comput Netw. 2013 Apr 7; 57(5): 1344–71.
2. Datta P, Panda SN, Bajaj S. Data analysis of cyber security for women in Haryana. In 2020 IEEE 8th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions) (ICRITO). 2020 Jun 4; 763–767.
3. Anand A, Chirputkar A, Ashok P. Mitigating Cyber-Security Risks using Cyber-Analytics. In 2023 IEEE 7th International Conference on Trends in Electronics and Informatics (ICOEI). 2023 Apr 11; 630–635.
4. Ulven JB, Wangen G. A systematic review of cybersecurity risks in higher education. Future Internet. 2021 Feb 2; 13(2): 39.

5. Liao Y, Vemuri VR. Use of k-nearest neighbor classifier for intrusion detection. Comput Secur. 2002 Oct 1; 21(5): 439–48.
6. Givargis T. Embedded system design: a unified hardware/software introduction. Wiley; New Jersey, USA. 2002.
7. Dabhade V, Alvi AS. Malicious Node Detection and Prevention for Secured Communication in WSN. In Computer Networks, Big Data and IoT: Proceedings of ICCBI 2021. Singapore: Springer Nature Singapore; 2022 May 22; 121–136.
8. Pabale AR, Kolhe RV, William P, Deshpande N, Paithankar DN, Yawalkar PM. Smart crack detection system using nanostructured materials with integrated optimization technology. Measurement: Sensors. 2023; 27: 100789.
9. Panda M. Security in wireless sensor networks using cryptographic techniques. Am J Eng Res. 2014 Oct; 3(01): 50–6.
10. Papp D, Ma Z, Buttyan L. Embedded systems security: Threats, vulnerabilities, and attack taxonomy. In 2015 IEEE 13th Annual Conference on Privacy, Security and Trust (PST). 2015 Jul 21; 145–152.
11. Van Huong P, Binh NN. Embedded system architecture design and optimization at the model level. International Journal of Computer and Communication Engineering (IJCCE). 2012 Nov 1; 1(4): 345–349.
12. Lo CH, Ansari N. CONSUMER: A novel hybrid intrusion detection system for distribution networks in smart grid. IEEE Trans Emerg Topics Comput. 2013 Jul 18; 1(1): 33–44.
13. Anwar A, Mahmood AN. Cyber security of smart grid infrastructure. arXiv preprint arXiv:1401.3936. 2014 Jan 16.
14. Cobilean V, Mavikumbure HS, Mcbride BJ, Vaagensmith B, Singh VK, Li R, Rieger C, Manic M. A review of visualization methods for cyber-physical security: Smart grid case study. IEEE Access. 2023 Jun 14; 11: 59788–803.
15. Sridhar S, Hahn A, Govindarasu M. Cyber–physical system security for the electric power grid. Proc IEEE. 2011 Oct 3; 100(1): 210–24.
16. Cardenas AA, Amin S, Sastry S. Secure control: Towards survivable cyber-physical systems. In 2008 IEEE 28th International Conference on Distributed Computing Systems Workshops. 2008 Jun 17; 495–500.
17. Mohammed A, George G. Vulnerabilities and strategies of cybersecurity in smart grid-evaluation and review. In 2022 IEEE 3rd International Conference on Smart Grid and Renewable Energy (SGRE). 2022 Mar 20; 1–6.

