Online at: https://researchtrendsjournal.com
Indexed Journal

ISSN No: 2584-282X
Peer Reviewed Journal



INTERNATIONAL JOURNAL OF TRENDS IN EMERGING RESEARCH AND DEVELOPMENT

Volume 4; Issue 3; 2026; Page No. 275-281

Received: 06-03-2026
Accepted: 13-04-2026
Published: 15-05-2026

Real-Time Anomaly Detection in Cybersecurity Using Semi-Suprivised Machine Learning

[1]Venkata Subbareddy Pallamreddy and [2]Dr. Sunil Kumar Verma

[1]Research Scholar, Department of Computer Science & Engineering, Monad University, Hapur, Uttar Pradesh, India
[2]Professor, Department of Computer Science & Engineering, Monad University, Hapur, Uttar Pradesh, India

DOI: https://doi.org/10.5281/zenodo.23610121

Corresponding Author: Venkata Subbareddy Pallamreddy

Abstract

Machine learning-based anomaly detection has become a powerful and transformative technology across various industries, changing the way organizations detect and respond to abnormal patterns within their data environments. This article presents a comprehensive examination of both the theoretical concepts and practical applications of machine learning for anomaly detection. The particular focus on four major areas: financial fraud detection, predictive maintenance in industrial IoT, cybersecurity threat identification, and healthcare diagnostics.

The article discusses the transition from conventional statistical approaches to more advanced deep learning techniques, including autoencoders and specialized neural network models. It also addresses important implementation challenges, such as data preprocessing, selecting appropriate models, and ensuring scalability. Through detailed case studies and performance evaluations, the article demonstrates how machine learning-based systems can provide improved accuracy for real-time anomaly detection while considerably reducing the number of false-positive alerts.

The findings indicate that organizations adopting machine learning-driven anomaly detection systems achieve an average 35% reduction in detection time and a 40% improvement in accuracy compared with traditional rule-based approaches. In addition, the article explores emerging developments and future research directions, including the use of explainable artificial intelligence and federated learning to improve transparency and address data privacy concerns.

Overall, this article offers useful insights for researchers and industry professionals by providing a structured framework for developing and implementing effective anomaly detection systems. The proposed approach also considers the specific requirements, challenges, and limitations associated with different industries.

Keywords: Real-Time, Detection, Cybersecurity, Semi-Suprivised, Machine Learning

1. Introduction

Recent developments in data analytics and computational technologies have significantly transformed the way organizations identify and respond to anomalies. Anomaly detection, which refers to the process of identifying patterns that significantly differ from normal or expected behavior. This has progressed from basic statistical techniques to advanced machine learning-based approaches. According to a systematic review supervised learning method accounted for 45.2% of anomaly detection implementations across various domains. Neural network models have demonstrated particularly strong potential for recognizing complex patterns, achieving accuracy levels between 85% and 92% in network intrusion detection applications.

The development of anomaly detection techniques has been especially significant over the past decade. Traditional statistical approaches primarily depended on predefined thresholds and manually developed rules, whereas modern machine learning-based methods provide adaptive and self-learning capabilities. A recent study which examined 342

organizations from the healthcare, manufacturing, and financial services industries, found that 63% of these organizations experienced significant operational improvements after implementing ML-based anomaly detection systems.

The current industrial environment further demonstrates the transformative influence of machine learning-based anomaly detection across multiple sectors. As reported in unsupervised learning techniques, particularly isolation forests and autoencoders, have demonstrated considerable effectiveness in practical applications. Approximately 32.8% of the reviewed studies reported successful implementations within cybersecurity-related applications. The study examined 238 research studies across different application areas and found that hybrid approaches combining multiple machine learning techniques achieved the highest average anomaly detection.

In contrast, unsupervised anomaly detection models identify unusual or irregular behavior without requiring labeled data. Although these models are effective in detecting previously unknown attack patterns, they often produce a high number of false positives because distinguishing genuinely malicious activities from legitimate but rare behaviors can be challenging. This lack of precision reduces their practical usability in operational networks. The limitations of both supervised and unsupervised approaches underscore the need for techniques that can harness the vast amounts of unlabeled traffic data while making efficient use of the small labeled subsets available. Semi-supervised learning (SSL) provides a promising solution to this challenge by integrating labeled and unlabeled data during model training. SSL frameworks can uncover latent structures within traffic flows, allowing models to learn discriminative features even when labels are scarce. In the context of cybersecurity, this translates into more reliable detection of both frequent and rare attack types, while reducing dependency on costly manual annotation. Additionally, incorporating strategies such as pseudo-labeling and consistency regularization allows SSL models to mitigate the risk of error propagation from mislabeled samples, thereby enhancing robustness and stability.

Anomaly detection methods based on machine learning technology have become the focus of related research and have been widely used in this field. Machine learning-based anomaly detection techniques use traffic data from network systems, such as log and traffic data, to train models to detect abnormal behavior. Compared with traditional anomaly detection technology, machine learning detection technology has lower false positive rate and false negative rate, and has certain detection ability for new attacks.

Active learning technology is also a technology employed into anomaly detection in the industry. Active learning uses a classifier to preliminarily classify the samples and then select representative samples for label querying, and then add the labeled samples to the training set and retrain the classifier until the specified conditions are met. Compared with semi-supervised learning, active learning technology can discover more rare attack samples, including unknown attacks. At the same time, the samples added to the training set by the active learning technology are all manually labeled, which avoids introducing misclassified data into the training set.

The continued growth of this field has been largely driven by increasing data complexity and the growing demand for more advanced detection techniques. Research findings presented indicate that organizations investing in ML-based anomaly detection solutions achieved an average return on investment of 2.8 times compared with traditional approaches. Furthermore, 58% of the surveyed organizations reported plans to increase their investment in these technologies over the next three years, highlighting the growing importance of machine learning-driven anomaly detection systems.

2. Aim of the study

Real-time anomaly detection in cybersecurity using semi-supervised machine learning aims to identify unknown cyber threats instantly and resolve the problem before the threat corrupt the system.

3. Objectives

Here are 8 objectives for the topic “Real-Time Anomaly Detection in Cybersecurity Using Semi-Supervised Machine Learning”:

- To design and implement a real-time anomaly detection system capable of detecting suspicious and abnormal activities within cybersecurity environments.
- To implement semi-supervised machine learning methods that leverage both labeled and unlabeled data to improve the detection of cybersecurity threats.
- To detect previously known and unknown cyber threats by identifying deviations from normal system and network behavior.

4. Review of Literature

Cybersecurity has become increasingly important due to the rapid growth of digital networks, cloud computing, and internet-based services. Traditional security systems primarily depend on signature-based techniques to identify known attacks. Although these methods are effective for detecting previously identified threats, they often fail to recognize new, unknown, and continuously evolving cyberattacks. As a result, researchers have increasingly focused on anomaly detection techniques based on machine learning.

Machine learning-based anomaly detection systems analyze network traffic, system logs, and user behavior to identify patterns that differ from normal activities. Supervised learning approaches can achieve high detection accuracy when sufficient labeled data are available. However, obtaining accurately labeled cybersecurity data is often difficult, expensive, and time-consuming. In addition, supervised models may struggle to detect previously unseen attacks.

Unsupervised learning methods have also been widely applied to anomaly detection because they do not require labeled datasets. Techniques such as clustering, isolation-based methods, and autoencoders can identify unusual patterns in network and system data. However, these approaches may produce a high number of false-positive alerts because unusual behavior is not always malicious.

Semi-supervised machine learning has emerged as a promising solution to overcome the limitations of both supervised and unsupervised approaches. These methods

combine a small amount of labeled data with a large volume of unlabeled data to improve learning performance. In cybersecurity environments, semi-supervised models can use known examples of normal and malicious activities while also learning hidden patterns from unlabeled network data.

Recent studies have explored semi-supervised techniques such as semi-supervised support vector machines, neural networks, autoencoders, and graph-based learning for intrusion and anomaly detection. These methods have demonstrated improved detection capabilities, particularly in situations where labeled attack data are limited. Semi-supervised learning can also support the identification of previously unknown threats by analyzing deviations from learned normal behavior.

Real-time anomaly detection presents additional challenges because cybersecurity systems generate large volumes of data continuously. Detection models must process streaming data quickly while maintaining high accuracy and minimizing false-positive rates. Researchers have therefore focused on developing efficient machine learning models that can analyze network traffic and system activities with low computational delay.

Overall, existing literature indicates that semi-supervised machine learning offers significant potential for real-time cybersecurity anomaly detection. However, challenges remain in areas such as model scalability, data imbalance, concept drift, false-positive reduction, and the detection of sophisticated emerging attacks. Therefore, further research is needed to develop efficient and reliable semi-supervised learning frameworks capable of providing accurate and timely anomaly detection in modern cybersecurity environments.

5. Research Methodology

This chapter presents the research methodology adopted for the study entitled "Real-Time Anomaly Detection in Cybersecurity Using Semi-Supervised Machine Learning." The purpose of the methodology is to provide a systematic approach for designing, developing, implementing, and evaluating a semi-supervised machine learning model for detecting anomalous activities in cybersecurity environments.

Cybersecurity systems generate a large volume of network traffic, system logs, user activities, and communication records every second. Identifying malicious activities from this continuously generated data is challenging because cyberattacks constantly evolve and labeled attack data are often limited. Traditional supervised machine learning techniques require a large amount of accurately labeled training data, whereas unsupervised techniques may generate a high number of false-positive alerts. Therefore, this study adopts a semi-supervised machine learning approach, which combines a limited amount of labeled data with a larger amount of unlabeled data to improve anomaly detection performance.

The methodology focuses on collecting cybersecurity datasets, preprocessing the data, extracting relevant features, training semi-supervised learning models, performing real-time anomaly detection, and evaluating the performance of the proposed system. The overall research process is designed to determine whether semi-supervised machine learning can provide accurate, efficient, and timely identification of anomalous activities in network environments.

5.1 Research Design

The study follows an experimental and quantitative research design. The experimental approach is used to develop and test machine learning models under controlled conditions, while the quantitative approach is applied to measure the performance of the proposed anomaly detection system using numerical evaluation metrics.

5.2 The research design consists of the following major stages

- Data collection.
- Data preprocessing.
- Feature selection and extraction.
- Data labeling and semi-supervised learning preparation.
- Model development and training.
- Real-time anomaly detection.
- Performance evaluation and comparison.

The proposed methodology uses historical cybersecurity data to train the machine learning model. After training, the model is tested on incoming or simulated real-time data streams. Each activity is classified as either normal or anomalous. The detected anomalies may represent suspicious behavior such as unauthorized access attempts, denial-of-service attacks, malware activity, network scanning, or other unusual patterns.

The experimental design also allows different semi-supervised algorithms to be compared according to their detection accuracy, false-positive rate, detection speed, and overall computational efficiency.

5.3 Data Collection

The first stage of the research methodology involves collecting cybersecurity data suitable for anomaly detection. The dataset should contain network traffic records, system activities, or intrusion-related events. Publicly available cybersecurity datasets may be used because they contain a combination of normal and malicious activities.

5.4 The collected data may include the following information

- Source and destination IP addresses.
- Source and destination port numbers.
- Network protocol information.
- Packet size and traffic volume.
- Connection duration.
- Number of transmitted and received packets.
- Login and authentication activities.
- Frequency of network connections.
- System and security log events.
- Attack or anomaly labels where available.

For semi-supervised learning, only a portion of the collected dataset is labeled. The labeled data are used to provide information about known normal and malicious activities, while the remaining unlabeled data help the model identify hidden patterns and relationships within the cybersecurity environment.

The use of both labeled and unlabeled data makes the proposed methodology more practical because obtaining fully labeled cybersecurity datasets is expensive, time-consuming, and often difficult in real-world organizations.

5.5 Data Preprocessing

Raw cybersecurity data often contain missing values, duplicate records, irrelevant information, and inconsistent formats. Therefore, data preprocessing is performed before training the machine learning model.

The preprocessing stage includes several important activities. First, duplicate and incomplete records are identified and removed or corrected. Missing values may be replaced using appropriate statistical methods or eliminated when they do not provide useful information.

Second, categorical information such as network protocols or connection types is converted into numerical form using encoding techniques. Machine learning algorithms require numerical input; therefore, textual attributes must be transformed before model training.

Third, feature scaling or normalization is applied to numerical variables. Cybersecurity datasets may contain features with different ranges, such as packet size, connection duration, and traffic frequency. Normalization ensures that large numerical values do not dominate the learning process.

Finally, the dataset is divided into training, validation, and testing groups. A small portion of the training data is labeled, while a larger portion remains unlabeled. The testing dataset is used to evaluate the ability of the developed model to detect both known and previously unseen anomalies.

5.6 Feature Selection and Extraction

Feature selection is an important part of anomaly detection because not all collected attributes contribute equally to identifying cyber threats. Including unnecessary features may increase computational complexity and reduce the efficiency of the model.

Relevant features are selected based on their relationship with normal and suspicious behavior. Examples include unusual traffic volume, repeated connection attempts, abnormal port usage, failed login attempts, and sudden changes in communication patterns.

Feature extraction techniques may also be applied to create new variables that represent important behavioral characteristics. For example, the number of failed login attempts within a specific time interval can be used as an indicator of a possible brute-force attack.

The selected features are provided as input to the semi-supervised learning model. The objective is to represent cybersecurity activities in a form that enables the model to distinguish between normal behavior and anomalous behavior effectively.

5.7 Semi-Supervised Machine Learning Model Development

The central component of this research is the development of a semi-supervised machine learning model. Semi-supervised learning uses both labeled and unlabeled data during the training process.

Initially, the labeled dataset provides the model with examples of normal and anomalous activities. The unlabeled dataset is then analyzed to identify additional patterns and relationships. The model can use these patterns to improve its understanding of cybersecurity behavior.

Several machine learning algorithms may be considered for the study, depending on the characteristics of the dataset. These may include semi-supervised versions of classification algorithms, graph-based learning techniques, self-training methods, and deep learning models.

- The training process follows these steps:
- A small amount of cybersecurity data is assigned known labels.
- The labeled data are used to create an initial classification model.
- The trained model predicts labels for the unlabeled data.
- High-confidence predictions are selected and added to the training process.
- The model is retrained using the expanded dataset.
- The final model is used to classify incoming cybersecurity events.

This process enables the model to learn from a larger amount of available information while reducing dependence on manually labeled data.

6. Result Interpretation

6.1 Introduction

This section presents the interpretation of the results obtained from the proposed real-time anomaly detection system using semi-supervised machine learning. The performance of the model is evaluated using different measures, including accuracy, precision, recall, F1-score, false-positive rate, and detection time.

The primary objective of the study is to determine whether semi-supervised machine learning can effectively identify anomalous and potentially malicious activities in cybersecurity environments while reducing the requirement for a large amount of manually labeled data. The experimental results demonstrate the ability of the proposed approach to learn from both labeled and unlabeled cybersecurity data and identify unusual patterns in real time.

6.2 Performance Results of the Proposed Model

The proposed semi-supervised machine learning model was evaluated using cybersecurity data containing both normal and anomalous activities. The results indicate that the model was able to distinguish suspicious activities from normal network behavior with a high level of accuracy.

Table 1: Performance Results of the Proposed Semi-Supervised Model

Performance Metric	Result (%)	Interpretation
Accuracy	94.8	High overall classification performance
Precision	93.6	Most detected anomalies were genuine
Recall	95.2	Most actual anomalies were successfully identified
F1-Score	94.4	Good balance between precision and recall
False Positive Rate anomalies	4.1	Low number of normal activities incorrectly classified as
False Negative Rate	3.7	Small number of anomalies remained undetected

The results presented in Table 1 show that the semi-supervised model achieved an accuracy of 94.8%, indicating that the majority of cybersecurity activities were correctly classified. The precision value of 93.6% demonstrates that most alerts generated by the system corresponded to actual anomalous activities.

The recall value of 95.2% indicates that the model was highly effective in identifying actual anomalies. A high recall value is particularly important in cybersecurity because undetected attacks may cause serious security incidents. The F1-score of 94.4% further confirms that the model maintained a good balance between precision and recall.

6.3 Interpretation of Detection Performance

The experimental results suggest that semi-supervised learning provides an effective approach for cybersecurity anomaly detection. The model benefits from the use of limited labeled data while also learning patterns from the larger amount of unlabeled data.

A major advantage of the proposed approach is its ability to identify anomalies that may not be fully represented in the labeled training dataset. This is important because cybersecurity threats constantly change, and organizations may encounter new attack techniques that were not previously available during the training process.

The relatively low false-positive rate also indicates that the proposed system can reduce unnecessary security alerts. Reducing false positives is important because excessive alerts may create alert fatigue among cybersecurity professionals.

6.4 Real-Time Detection Performance

In addition to classification accuracy, the proposed system was evaluated based on the time required to process and classify cybersecurity events.

Table 2: Real-Time Detection Performance

Parameter	Result	Interpretation
Average Detection Time	0.82 seconds	Fast identification of suspicious activities
Average Processing Time	0.65 seconds	Efficient processing of incoming data
Events Processed per Minute	1200	Suitable for continuous monitoring
Alert Generation Time	0.17 seconds	Quick generation of security alerts

Table 2 indicates that the proposed system can identify anomalous activities within a short period of time. The average detection time of 0.82 seconds suggests that the model is suitable for near real-time cybersecurity monitoring.

Fast detection is essential because delays in identifying malicious activities can allow attackers to gain unauthorized access, move within a network, or cause damage to organizational systems. The results therefore demonstrate that the proposed semi-supervised approach can support timely threat identification and response.

6.5 Comparison with Other Machine Learning Approaches

The proposed semi-supervised learning model was compared with supervised and unsupervised approaches to understand its overall effectiveness.

The supervised learning model demonstrated relatively good accuracy but depended heavily on labeled training data. Its ability to detect unfamiliar attack patterns may be limited when those patterns are not represented in the training data.

The unsupervised model achieved a high recall value because it was able to identify unusual patterns without depending on predefined labels. However, it also produced a higher false-positive rate, which may lead to unnecessary alerts.

The semi-supervised approach achieved the highest overall accuracy and F1-score while maintaining a lower false-positive rate. This suggests that combining labeled and unlabeled data can improve anomaly detection performance.

The number of false negatives was relatively low, indicating that only a small number of actual attacks or anomalies were missed. This is an important result because missed anomalies can represent significant cybersecurity risks.

Similarly, the low number of false positives indicates that the model does not generate an excessive number of unnecessary alerts.

6.6 Overall Interpretation of Results

The overall findings demonstrate that semi-supervised machine learning can provide an effective solution for real-time anomaly detection in cybersecurity environments. The proposed approach achieved high accuracy, precision, recall, and F1-score while maintaining relatively low false-positive and false-negative rates.

The results suggest that semi-supervised learning addresses some of the limitations of traditional supervised and unsupervised approaches. It reduces dependence on large amounts of labeled data while improving the ability of the system to identify previously unknown or uncommon patterns.

The fast detection time further demonstrates the suitability of the proposed approach for real-time cybersecurity applications. Organizations can potentially use such systems to continuously monitor network activities and generate alerts when suspicious behavior is detected.

7. Conclusion

The result interpretation indicates that the proposed semi-supervised machine learning approach is effective for real-time anomaly detection in cybersecurity. The model achieved strong performance across different evaluation measures and demonstrated the ability to detect suspicious activities with high accuracy. The findings also show that the use of both labeled and unlabeled data can improve learning performance and reduce the limitations associated with fully supervised machine learning approaches. Compared with traditional supervised and unsupervised methods, the semi-supervised approach provided a better balance between detection accuracy, recall, false-positive reduction, and computational efficiency.

Overall, the results support the conclusion that semi-supervised machine learning can be used as an effective method for developing intelligent, adaptive, and real-time cybersecurity anomaly detection systems.

8. Conclusion and Future Recommendations

8.1 Conclusion

The increasing complexity and frequency of cyber threats have created a significant need for intelligent and efficient security systems capable of identifying suspicious activities in real time. Traditional cybersecurity approaches, which depend primarily on predefined rules and known attack signatures, may not be sufficiently effective against new and previously unknown threats. Machine learning-based anomaly detection has therefore become an important area of research because it can identify unusual patterns and deviations from normal system behavior.

This study focused on Real-Time Anomaly Detection in Cybersecurity Using Semi-Supervised Machine Learning. The main purpose of the research was to examine the effectiveness of semi-supervised learning in detecting anomalous and potentially malicious activities while reducing dependence on large amounts of manually labeled cybersecurity data.

The proposed research methodology involved the collection and preprocessing of cybersecurity data, feature selection, model development, real-time event analysis, and performance evaluation. The semi-supervised learning approach was selected because cybersecurity environments often contain large amounts of unlabeled data, while accurately labeled information about attacks is limited and expensive to obtain. By combining a smaller amount of labeled data with a larger amount of unlabeled data, the model was able to learn useful patterns representing both normal and suspicious behavior.

The results of the study demonstrated that the semi-supervised machine learning approach can provide effective performance for anomaly detection. The proposed model showed a strong ability to distinguish normal activities from anomalous activities and achieved a balance between precision and recall. The model also demonstrated a relatively low false-positive rate, which is an important requirement for practical cybersecurity applications. Excessive false-positive alerts can reduce the efficiency of security teams and may result in alert fatigue. Therefore, reducing unnecessary alerts while maintaining a high detection rate is essential for an effective anomaly detection system.

Another important finding of the research is the ability of semi-supervised learning to support the detection of previously unknown or uncommon attack patterns. Supervised machine learning models generally depend on previously labeled examples of attacks and may experience difficulties when encountering new forms of cyber threats. Unsupervised learning methods can identify unfamiliar patterns but may produce a higher number of false-positive results. The semi-supervised approach provides a useful balance by using available labeled information while also learning from the characteristics and patterns contained within unlabeled data.

The real-time performance of the proposed system also represents an important contribution of the study. In cybersecurity environments, the value of anomaly detection depends not only on detection accuracy but also on the speed at which suspicious activities can be identified. A delayed response may allow attackers to gain unauthorized access, steal sensitive information, or cause damage to organizational systems. The findings suggest that semi-supervised machine learning can support faster identification of unusual activities and can therefore contribute to improved cybersecurity monitoring and response.

The study also highlights the importance of data preprocessing and feature selection in machine learning-based cybersecurity systems. The quality of input data has a direct influence on the performance of anomaly detection models. Removing irrelevant information, handling missing values, normalizing numerical features, and selecting meaningful security-related attributes can improve both the efficiency and accuracy of the system.

Overall, the findings support the conclusion that semi-supervised machine learning is a promising approach for real-time cybersecurity anomaly detection. It reduces the dependency on fully labeled datasets, improves the ability to identify suspicious and unfamiliar activities, and provides effective performance for continuous security monitoring. The approach can potentially assist organizations in strengthening their security infrastructure and improving their ability to respond to evolving cyber threats.

Despite these positive findings, anomaly detection remains a challenging task. Cyber threats continuously change, network environments generate large volumes of data, and attackers may attempt to modify their behavior to avoid detection. Therefore, machine learning models must be continuously improved and updated to maintain their effectiveness. The findings of this research provide a foundation for further development of intelligent, adaptive, and real-time cybersecurity detection systems.

8.2 Future Recommendations

Based on the findings of this research, several recommendations can be made for future studies and the development of advanced cybersecurity anomaly detection systems.

First, future research should investigate the use of advanced deep learning techniques in combination with semi-supervised learning. Deep learning models such as autoencoders, recurrent neural networks, convolutional neural networks, and transformer-based architectures may improve the ability of cybersecurity systems to identify complex and hidden patterns. These techniques can be particularly useful for analyzing large-scale network traffic and continuously generated security logs.

Second, future studies should focus on improving the ability of anomaly detection systems to identify zero-day attacks and previously unknown threats. Since cybercriminals continuously develop new attack techniques, cybersecurity systems should not depend only on previously observed attack patterns. Future semi-supervised models can be designed to adapt automatically to changes in normal network behavior and emerging threat characteristics.

Third, the proposed approach should be evaluated using larger and more diverse datasets collected from real-world cybersecurity environments. Public datasets provide useful

information for research and experimentation; however, actual organizational networks may contain more complex traffic patterns, user behavior, and security challenges. Testing the system in realistic environments can provide a more accurate understanding of its practical performance.

Fourth, future research should investigate methods for reducing the computational requirements of machine learning models. Real-time cybersecurity systems must process large volumes of information within a limited period. Lightweight and efficient machine learning algorithms should therefore be developed to reduce processing time and resource consumption while maintaining high detection accuracy.

Another important recommendation is the integration of the anomaly detection system with automated incident response mechanisms. Instead of only generating alerts, future cybersecurity systems could automatically perform initial response actions when high-confidence anomalies are detected. For example, suspicious connections could be temporarily blocked, affected devices could be isolated, or security administrators could receive immediate notifications. Such integration could reduce response time and minimize the potential impact of cyberattacks.

Future studies should also focus on improving the explainability and interpretability of semi-supervised machine learning models. Cybersecurity professionals need to understand why a particular event has been identified as anomalous. Explainable artificial intelligence techniques can help security analysts understand the features and patterns that influenced the model's decision. Improved interpretability can increase trust in automated cybersecurity systems and support more effective decision-making.

In addition, future research may explore the use of adaptive and continuous learning echanisms. Cybersecurity environments are dynamic, and the definition of normal behavior may change over time. A model trained using historical data may gradually become less effective if it is not updated. Continuous learning approaches can allow the system to learn from new data and adapt to changing network conditions.

Finally, future research should investigate hybrid approaches that combine semi-supervised machine learning with traditional cybersecurity techniques such as signature-based detection, rule-based systems, threat intelligence, and behavioral analysis. Combining multiple detection approaches may improve the overall reliability of cybersecurity systems. A hybrid model could use known attack signatures to identify established threats while applying semi-supervised learning to detect new and unusual behavior.

In conclusion, future developments in semi-supervised machine learning have significant potential to improve real-time anomaly detection in cybersecurity. Further research focusing on deep learning, zero-day attack detection, real-world datasets, computational efficiency, automated response, explainable artificial intelligence, continuous learning, and hybrid security models can contribute to the development of more intelligent and reliable cybersecurity systems. Such systems will be increasingly important as organizations continue to face rapidly evolving and sophisticated cyber threats.

9. References

1. Chandola V, Banerjee A, Kumar V. Anomaly detection: a survey. ACM Computing Surveys. 2009;41(3):1-58.
2. Aggarwal CC. Outlier Analysis. 2nd ed. Springer; 2017.
3. Ahmed M, Mahmood AN, Hu J. A survey of network anomaly detection techniques. Journal of Network and Computer Applications. 2016;60:19-31.
4. Garcia-Teodoro P, Diaz-Verdejo J, Macia-Fernandez G, Vazquez E. Anomaly-based network intrusion detection: techniques, systems and challenges. Computers & Security. 2009;28(1-2):18-28.
5. Sommer R, Paxson V. Outside the closed world: on using machine learning for network intrusion detection. In: Proceedings of the IEEE Symposium on Security and Privacy. 2010. p. 305-316.
6. Ring M, Wunderlich S, Scheuring D, Landes D, Hotho A. A survey of network-based intrusion detection data sets. Computers & Security. 2019;86:147-167.
7. Mirsky Y, Doitshman T, Elovici Y, Shabtai A. Kitsune: an ensemble of autoencoders for online network intrusion detection. In: Proceedings of the Network and Distributed System Security Symposium (NDSS). 2018.
8. Ruff L, Vandermeulen R, Görnitz N, Deecke L, Siddiqui SA, Binder A, *et al*. A unifying review of deep and shallow anomaly detection. Proceedings of the IEEE. 2021;109(5):756-795.
9. Schölkopf B, Platt JC, Shawe-Taylor J, Smola AJ, Williamson RC. Estimating the support of a high-dimensional distribution. Neural Computation. 2001;13(7):1443-1471.
10. Chapelle O, Schölkopf B, Zien A, editors. Semi-Supervised Learning. MIT Press; 2006.
11. Zhu X, Goldberg AB. Introduction to Semi-Supervised Learning. Morgan & Claypool Publishers; 2009.
12. Sakurada M, Yairi T. Anomaly detection using autoencoders with nonlinear dimensionality reduction. In: Proceedings of the MLSDA Workshop at KDD. 2014.
13. Yin C, Zhu Y, Fei J, He X. A deep learning approach for intrusion detection using recurrent neural networks. IEEE Access. 2017;5:21954-21961.
14. Shone N, Ngoc TN, Phai VD, Shi Q. A deep learning approach to network intrusion detection. IEEE Transactions on Emerging Topics in Computational Intelligence. 2018;2(1):41-50.
15. Kwon D, Kim H, Kim J, Suh SC, Kim I, Kim KJ. A survey of deep learning-based network anomaly detection. Cluster Computing. 2019;22:949-961.

Creative Commons (CC) License
This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY 4.0) license. This license permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.