International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 08 Issue: 06 | June - 2024 SJIF Rating: 8.448 ISSN: 2582-3930

Reference-Based AI Decision Support for Cyber Security

Mr. Sunad Kumar AN, Asst. Professor,

Dept of computer science and engineering, BGS Institute of Technology,

Adhichunchanagiri university, BG Nagara, Karnataka

Mr. Keerthi Raj J ,4th year ,8th sem ,

Dept of computer science and engineering , BGS Institute of Technology , Adhichunchanagiri university , BG Nagara , Karnataka

ABSTRACT

The cyber environment, massive amounts of data are generated daily. Artificial Intelligence (AI) technologies can effectively manage this vast data to support efficient operations in the cyber environment. Thanks to active research, AI has advanced significantly in this regard. However, as AI achieves higher performance, it becomes increasingly complex, which results in the low interpretability of AI outputs. This black-box nature of AI technology makes AI challenging to apply in fields like cybersecurity, where the risk of false positives is significant. To address this issue, researchers have been working one explainable Artificial Intelligence (XAI) technology, with the intention to enhance the utility of AI by providing interpretations of AI predictions. Most previous research has focused on understanding how models function in terms of feature importance to interpret AI results. However, this approach fails to provide clear interpretations in fields where interpretability is crucial, such as security. Therefore, this paper proposes a framework that offers interpretations of AI results, even in unsupervised environments that are suitable for security scenarios. Additionally, we have improved the logic of calculation Reference and have enhanced the function and performance compared with previous research. We provide additional information that supports interpretation, such as PValues and References, to offer more effective decision support to security analysts and to ultimately reduce false alarms and enhance model performance.

INTRODUCTION

Due to its ability to effectively process and utilize big data, Artificial Intelligence (AI) technology has been a field of active research and development since its early stages. Thanks to these research efforts, AI technology has evolved and has become applicable to various fields. However, in pursuit of such performance improvements, AI technology has adopted a more complex output logic, which has resulted in the decreased interpretability of its output. In other words, although it has demonstrated excellent performance, AI technology has acquired a black-box nature that makes it difficult to identify the mechanism behind its output. This characteristic of AI technology has become an obstacle to its adoption in fields that have a high risk of false positives. To address these shortcomings and make AI effective even in fields with a high risk of false positives, explainable Artificial Intelligence (XAI) technology is being develop. In an era dominated by digitalization, the significance of cybersecurity cannot be overstated. As organizations increasingly rely on interconnected systems and data, the vulnerability to cyber threats escalates proportionately. Traditional approaches to cybersecurity, while effective to some extent, are often inadequate in addressing the dynamic and sophisticated nature of modern cyber threats. In response to this pressing need, Reference-Based AI Decision Support .

RBADS leverages the power of artificial intelligence (AI) to enhance decision-making processes within cybersecurity frameworks. Unlike conventional methods that rely heavily on predefined rules and signatures, RBADS operates on the principle of continuous learning and adaptation. At its core, RBADS analyzes vast amounts of data, drawing insights from both historical incidents and real-time inputs to identify patterns, anomalies, and potential threats.

One of the distinguishing features of RBADS is its reliance on references. These references encompass a broad spectrum of sources, including past cyber incidents, industry best practices, regulatory standards, and threat intelligence feeds. By incorporating diverse references, RBADS ensures a comprehensive and contextually rich understanding of thecybersecurity landscape. This multidimensional perspective enables RBADS to not only detect known threats but also anticipate emerging risks and vulnerabilities.

Central to RBADS is its ability to provide decision support to cybersecurity professionals. Rather than replacing human expertise, RBADS functions as a force multiplier, augmenting the analytical capabilities of security teams. Through advanced analytics and machine learning algorithms, RBADS assists in prioritizing alerts, validating potential threats, and recommending appropriate response strategies. This collaborative approach enables organizations to effectively allocate resources and mitigate risks in a timely manner.

In an era dominated by digitalization, the significance of cybersecurity cannot be overstated. As organizations increasingly rely on interconnected systems and data, the vulnerability to cyber threats escalates proportionately. Traditional approaches to cybersecurity, while effective to some extent, are often inadequate in addressing the dynamic and sophisticated nature of modern cyber threats. In response to this pressing need, Reference-Based AI Decision Support.

Central to RBADS is its ability to provide decision support to cybersecurity professionals. Rather than replacing human expertise, RBADS functions as a force multiplier, augmenting the analytical capabilities of security teams. Through advanced analytics and machine learningalgorithms, RBADS assists in prioritizing alerts, validating potential threats, and recommending appropriate response strategies. This collaborative approach enables organizations to effectively allocate resources and mitigate risks in a timely manner.

Furthermore, RBADS facilitates adaptive defense mechanisms by continuously refining its models based on feedback loops and evolving threat landscapes. This dynamic adaptation is particularly crucial in combating zero-day exploits and novel attack vectors, where conventional defenses often fall short. By iteratively learning from both successes and failures, RBADS enhances its predictive capabilities and resilience against emerging threats.

The implementation of RBADS encompasses a range of technical components, including data collection, preprocessing, feature extraction, modeling, and decision support modules. Leveraging advancements in AI, machine learning, and big data technologies, RBADS integrates seamlessly into existing cybersecurity architectures, providing scalability and interoperability across diverse environments.

However, the adoption of RBADS is not without challenges. Privacy concerns, data governance issues, and algorithmic biases pose significant hurdles that must be addressed to ensure ethical and responsible deployment. Additionally, the complexity of cyber threats necessitates ongoing research and collaboration to enhance the effectiveness and robustness ofRBADS solutions.

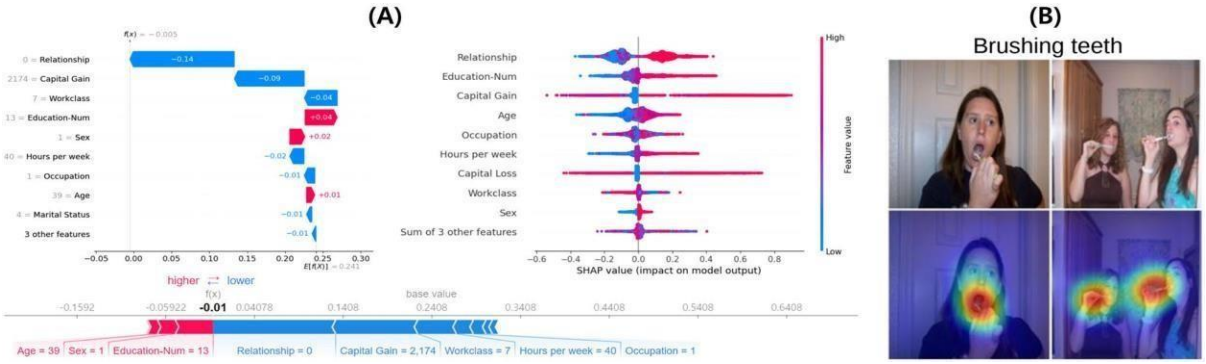

Fig 1.1 Examples of interpretations provided by feature importance-based XAI [1], [2].

Previous well-known XAI technologies, such as Shap- ley Additive exPlanation(SHAP)

[1] and Class Activation Map(CAM) [2], have primarily provided visual interpreta- tions based on feature importance. These research approaches aim to explain how each feature contributes to AI's decision- making by calculating the importance of each feature when AI makes specific decisions.

Such previous research mainly focused on providing inter- pretations through anunderstanding of AI's operational logic and was developed to target supervised learning models to compute feature importance for each label. However, visual interpretations based on feature importance in cybersecurity may not always provide clear explanations. This is because cybersecurity relies on attack detection based on the differ- ences between the original feature values in normal scenarios and those in attack scenarios. Therefore, in cybersecurity, interpreting attack decisions based on differences in terms of feature values in the actual data, rather than relying solely on feature importance, can offer clearer interpretability.

However, as mentioned in previous studies [3], [4], [5], [6], the cybersecurity environment often needs greater resources to label all data, and most of the collected data consists of nor- mal data, which leads to class imbalance issues. Additionally, supervised learning relies on pre-labeled data for detection while achieving high discrimination accuracy, which makes it challenging to respond to unknown threats and attacks that have yet to be detected. Unsupervised learning models can effectively operate in cybersecurity environments, which do not require label information during AI training and the result-generation processes.

1.1 Objectives

- RBADS aims to enhance the capability of cybersecurity systems to detect and identify awide range of cyber threats, including known vulnerabilities, zero-day exploits, and novel attack vectors. By leveraging diverse references such as historical incidents, threat intelligence feeds, and industry best practices, RBADS can provide comprehensive threat detection capabilities, reducing false positives and increasing the accuracy of threat identification.
- RBADS seeks to provide cybersecurity professionals with a holistic view of the cybersecurity landscape, enabling them to gain insights into emerging risks, vulnerabilities, and trends. By continuously analyzing and synthesizing information from various sources, RBADS enhances situational awareness, empowering organizations to make informed decisions and proactively respond to evolving threats.
- RBADS aims to facilitate dynamic risk assessment and prioritization by analyzing the

SYSTEM ANALYSIS AND DESIGN

2.1 Existing System

Reference-Based AI Decision Support for Cybersecurity is a sophisticated system designedto enhance security operations by leveraging artificial intelligence (AI) algorithms and referencing a comprehensive database of known cyber threats and attack patterns. At its core,this system aims to empower cybersecurity professionals with timely and accurate decision- making support, enabling them to proactively identify, mitigate, and respond to potential security incidents effectively. The system operates by continuously collecting and analyzing vast amounts of data from various sources, including network traffic, system logs, and threat intelligence feeds. Using advanced machine learning techniques, it identifies patterns, anomalies, and indicators of compromise that may indicate malicious activity.

One of the key features of this system is its reference-based approach. By comparing current data against a repository of known threats and attack signatures, it can quickly identify similarities and deviations from established norms. This allows it to detect both known and emerging threats, providing security teams with valuable insights into potential risks. Moreover, the AI decision support component of the system provides recommendations and alerts based on the analysis of incoming data. These recommendations can range fromsuggesting immediate actions to mitigate an ongoing attack to providing guidance on strengthening overall security posture. By automating routine tasks and providing actionable insights, the system enables security analysts to focus their efforts on more strategic activities, such as threat hunting and incident response planning.

Additionally, the system's reference-based approach enables it to adapt and evolve over time. As new threats emerge and attack techniques evolve, the system continuously updates its reference database to ensure it remains effective against the latest cyber threats. This dynamic nature is crucial in the ever-changing landscape of cybersecurity, where staying ahead of adversaries requires constant vigilance and adaptation Furthermore, the system incorporates advanced visualization techniques to present complex data in a clear and intuitive manner. Interactive dashboards and reports allow security analysts to explore trends,correlations, and dependencies within the data, facilitating informed decision-making and
response to security incide.

In summary, Reference-Based AI Decision Support for Cybersecurity represents a cutting- edge approach to enhancing security operations in today's digital landscape. By leveragingAI algorithms and a comprehensive reference database of known threats, the system empowers security professionals with timely and accurate decision-making support. Through its dynamic nature, advanced visualization capabilities, and emphasis on automation, the system enables organizations to strengthen their security posture and effectively defendagainst a wide range of cyber threats.

2.2 Proposed System

The proposed framework introduces a novel approach to interpreting and enhancing the decision support provided by an Autoencoder model for anomaly detection, specificallytailored to raw log data collected in Endpoint Detection and Response (EDR) systems. By integrating advanced techniques for unsupervised learning interpretation and leveraging insights from existing reference generation algorithms, such as those proposed by Han and others, the framework aims to provide clear and actionable interpretations in the cybersecurity domain. Central to the framework is the improved reference generation logic, which focuses on calculating an Optimum Reference that

better captures the nuances of the raw log data. Through rigorous validation work, it has been demonstrated that this enhanced reference generation process significantly improves functional and performance metrics, leading to more accurate anomaly detection outcomes.

Building upon this Optimum Reference, the framework generates a range of interpretive metrics, including Nearest Real Data, P-Value, and others. These metrics serve to enhance theclarity of interpretation by providing analysts with deeper insights into the context andsignificance of detected anomalies. Crucially, the framework employs AI decision support mechanisms to guide analysts in assessing the reliability of each AI prediction. By analyzing the comprehensive set of interpretive metrics, the system offers guidance on whether to accept or question the conclusions drawn by the AI model. This empowers analysts to make informed decisions based on a holistic understanding of the data and the model's predictions.

In instances where false alarms occur due to incorrect AI predictions, analysts play a pivotal role in contributing to false alarm reduction and enhancing the performance of the AI system. By refraining from citing erroneous predictions and providing feedback on misclassifications,

analysts facilitate continuous improvement and refinement of the AI model.

Overall, the proposed framework represents a significant advancement in the field of AI- driven anomaly detection in cybersecurity. By prioritizing interpretability, leveraging sophisticated reference generation techniques, and integrating AI decision support mechanisms, the framework enables analysts to effectively navigate the complexities of raw log data analysis and make informed decisions that enhance both security posture and AI performance.

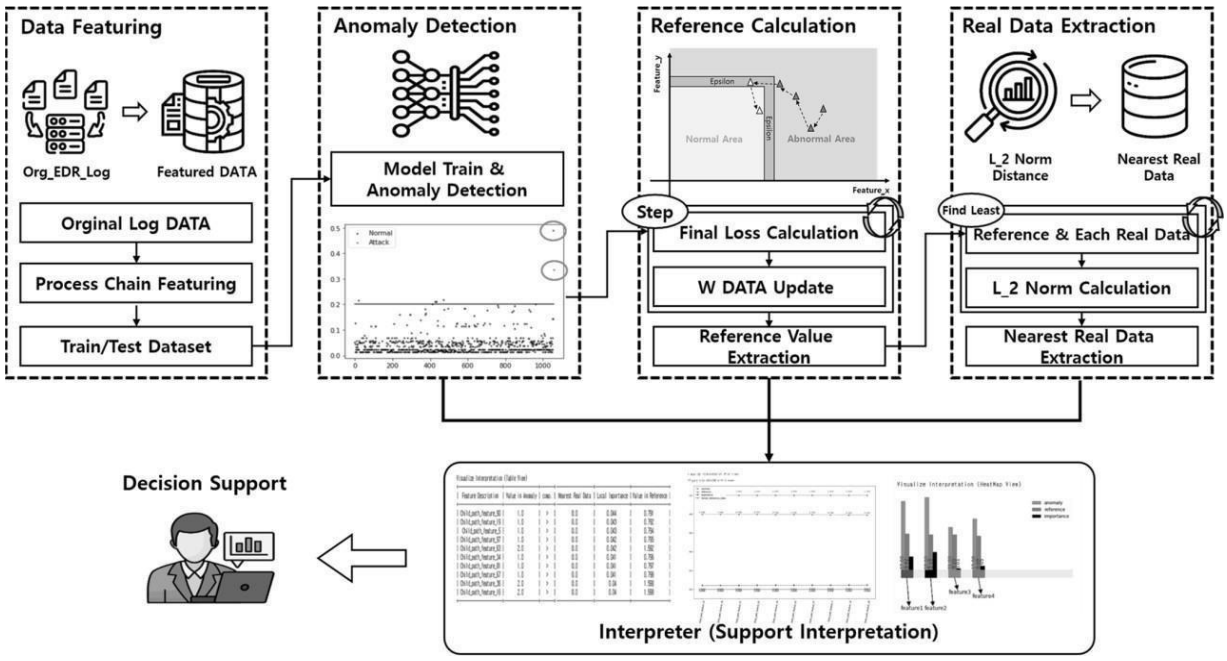

FIGURE 2.1 Proposed framework: Providing clear interpretability for anomaly data.

2.3 Advantages Of Proposed System

1. Enhanced Decision Making: By leveraging user references, the AI system gains insights into specific threats, attack patterns, and vulnerabilities relevant to the organization. This enhances decision-making capabilities by providing tailored recommendations that align with the organization's unique security needs and concerns.

2. Contextual Understanding: User references offer context to the AI system, allowing it to better understand the organization's network architecture, infrastructure, and security policies. This contextual understanding enables more accurate threat detection and mitigation, as the system can differentiate between benign activities and genuine security threats based on the organization's specific context.

3. Improved Accuracy: User references help refine the AI algorithms, making them more accurate in identifying and classifying security incidents. By learning from user-provided dataand feedback, the AI system continuously improves its detection capabilities, reducing false positives and false negatives and enhancing overall accuracy.

4. Customization and Flexibility: User reference-based AI decision support allows organizations to customize the system according to their unique requirements and preferences.Users can provide feedback, adjust parameters, and update reference databases to ensure that the AI system reflects the latest threats and security trends relevant to their environment.

5. Proactive Threat Intelligence: By incorporating user references, the AI system gains access to real-world threat intelligence curated by security experts within the organization. This proactive approach enables the system to anticipate emerging threats and vulnerabilities, empowering security teams to take preemptive measures to protect against potential attacks.

6. Efficient Resource Allocation: User reference-based AI decision support helps optimize resource allocation by prioritizing security incidents based on their relevance and potential impact on the organization. By focusing on threats identified through user references, security teams can allocate resources more efficiently, addressing the most critical issues first.

7. Collaborative Defense: User reference-based AI decision support promotes collaboration and information sharing among security professionals within and across organizations. By pooling together user references and sharing insights, organizations can collectively strengthen their defenses and respond more effectively to evolving cyber threats.

8. Compliance and Governance: User reference-based AI decision support aids organizations in meeting regulatory compliance requirements and industry standards by providing a transparent and auditable process for threat detection and response. By incorporating user references into the decision-making process, organizations can demonstrate due diligence and adherence to security best practices.

9. Continuous Learning and Improvement: User reference-based AI decision support fosters a culture of continuous learning and improvement within the organization. By actively engaging users in the threat detection and mitigation process, the AI system can adapt and evolve in real- time, staying ahead of evolving threats and

vulnerabilities.

10. Cost-Effective Security: Ultimately, user reference-based AI decision support offers a cost-effective approach to cybersecurity by leveraging existing knowledge and expertise within the organization. By harnessing user-generated data and insights, organizations can enhance their security posture without the need for significant additional investments in technology or resources.

APPLICATIONS

1. Enterprise Security: In enterprise security, Reference-Based AI Decision Support enhances threat detection by leveraging user-provided references. This approach refines anomaly detection algorithms, enabling tailored recommendations aligned with theorganization's context. By incorporating user insights, the system improves accuracy, fosters proactive threat intelligence, and empowers efficient decision-making to mitigate cyber risks effectively.

2. Financial Services: Banks, financial institutions, and fintech companies can benefit from User Reference-Based AI Decision Support to safeguard sensitive financial data, prevent fraud, and ensure compliance with regulatory requirements such as PCI DSS and GDPR. In financial services, Reference-Based AI Decision Support bolsters cybersecurity byintegrating user-provided references. This strengthens fraud detection, regulatory compliance, and data protection measures tailored to the industry's specific risks. By leveraging user insights, the system enhances threat intelligence, mitigates vulnerabilities, and ensures the security of sensitive financial information.

3. Healthcare: Healthcare providers and organizations handling sensitive patient information can use this technology to protect electronic health records (EHRs), medical devices, and other critical infrastructure from cyberattacks and data breaches. In healthcare, Reference-Based AI Decision Support fortifies cybersecurity by incorporating user-providedreferences. This enhances the protection of patient records, medical devices, and critical infrastructure from cyber threats. By leveraging user insights, the system improves threat detection, compliance with regulations like HIPAA, and safeguards against data breaches and unauthorized access.

4. Government and Defense: Government agencies, military organizations, and defense contractors can leverage User Reference-Based AI Decision Support to defend against sophisticated cyber threats targeting national security interests, critical infrastructure, and classified information. In Government and Defense, Reference-Based AI Decision Support elevates cybersecurity by integrating user-provided references. This strengthens nation

METHODOLOGY

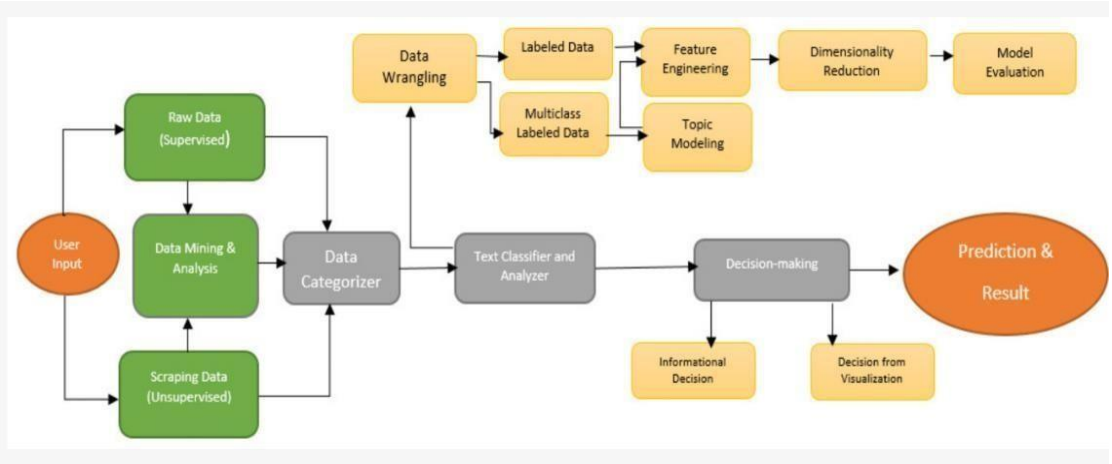

Figure 4.1 The RAIDSS text classifier model.

The RAIDSS classifier model is seen in Figure 1, where the user gives a specific keyword ortopic to extract information from the Web or to specifically label the dataset to get results. After Web scraping or mining, the data need to be categorized for the classifier. Thecategorization process identifies which information is the user's given and mining data. A text classifier formats these data for further analysis such as data preparation, model evaluation, builds an application and evaluates performance prediction and the results.

The objective of a text classifier is to send information to either supervised or unsupervised learning, where a given sample of data gets the desired output. It shows the relationships between input and output as visual information. After mining information that contains raw data, classifiers receive information as a named given dataset or as mined data. Therefore, the RAIDSS model does the assessment and provides the prediction or output. The most significant undertakings inside unsupervised learning are clustering, portrayal learning, and density estimation [25]. However, the dataset is prepared by topic modeling with multiclass text classifications, where the data-wrangling classifier first applies labeling and then goes into model evaluation. If the user input contains a labelling corpus, then the classifier sends it for supervised processing. Models have earlier information on what the output determines our samples ought to be. Thus, it is learning conceivable text and needs to apply binary or multiclass text classifications. Classification's goal is to infer the natural structure or hierarchical structures that present data points [19]. After model evaluation, users get their desired answers through several decision-making graph visualizations and informational output by the chatbot application.

CONCLUSION AND FUTURE WORK

Conclusion

In this paper, to provide a conclusion based on reference-based AI decision support for cybersecurity, it's essential to highlight the significance and implications of this approach. By leveraging a repository of prior cybersecurity incidents and solutions, AI systems can offer invaluable support to cybersecurity professionals in making informed decisions, identifying threats, and mitigating risks effectively. This method not only enhances the efficiency of cybersecurity operations but also facilitates continuous learning and adaptation in the face of evolving cyber threats.

Moreover, by incorporating diverse perspectives and experiences from past incidents, AI decision support systems can contribute to more comprehensive and robust cybersecurity strategies. Overall, reference-based AI decision support holds great promise in strengthening cyber defense capabilities and bolstering resilience against emerging cyber threats.

Future Work

1. Enhanced Threat Intelligence Integration: Future work could focus on refining the integration of threat intelligence feeds into the reference-based AI decision support system. This could involve developing more sophisticated algorithms to analyze and prioritize threat data, ensuring that the system can effectively identify and respond to emerging threats inreal- time.

2.Behavioral Analysis Capabilities: Expanding the system's capabilities to include advanced behavioral analysis techniques could significantly enhance its ability to detect and mitigate cybersecurity threats. This might involve integrating machine learning models capable of identifying anomalous behavior patterns within network traffic or user activity.

3. Automated Response Mechanisms: Developing automated response mechanisms within the decision support system could enable it to autonomously implement countermeasures against detected threats. This could involve integrating with existing security infrastructure to automatically block malicious IP addresses, quarantine compromised devices, or initiate incident response procedures.

4. Continuous Learning and Adaptation: Implementing mechanisms for continuouslearning and adaptation will be crucial for ensuring the long-term effectiveness of thedecision support system. This could involve leveraging techniques such as reinforcement learning to enable the system to adapt its decision-making processes based on feedback fromprevious actions and evolving threat landscapes.

5. Interoperability and Integration: Enhancing interoperability and integration with other cybersecurity tools and platforms will be essential for maximizing the utility of the decision support system within complex organizational environments. This could involve developing standardized interfaces and APIs to facilitate seamless communication and data sharing between the decision support system and other security solutions.

6. Human-Machine Collaboration: Exploring ways to facilitate more effective collaboration between human analysts and the AI decision support system will be critical forleveraging the strengths of both humans and machines in cybersecurity operations. This could involve developing user-friendly interfaces that provide analysts with actionable insights and recommendations based on the system's analysis.

7.Scalability and Performance Optimization: As the volume and complexity of cybersecurity threats continue to increase, ensuring scalability and optimizing theperformance of the decision support system will be essential. This could involve leveraging distributed computing techniques, optimizing algorithms for parallel processing, and leveraging cloud-based infrastructure to handle large-scale data analysis tasks efficiently.

8. Regulatory Compliance and Ethical Considerations: Addressing regulatory compliance requirements and ethical considerations will be paramount in the development and deployment of the decision support system. This could involve implementing mechanisms to ensure compliance with data protection regulations, as well as

incorporating ethical principles such as transparency, accountability, and fairness into the system's design and operation.

REFERENCES

[1] W. Li, W. Meng, X. Luo, and L. F. Kwok, ‘‘MVPSys: Toward practical multi-view based false alarm reduction system in network intrusion detec- tion,’’ *Comput. Secur.*, vol. 60, pp. 177–192, Jul. 2016.

[2] M. Eslahi, H. Hashim, and N. M. Tahir, ‘‘An efficient false alarm reduction approach in HTTP-based botnet detection,’’ in *Proc. IEEE Symp. Comput. Informat. (ISCI)*, Apr. 2013,
pp. 201–205.

[3] B. Subba, S. Biswas, and S. Karmakar, ‘‘False alarm reduction in signature- based IDS: Game theory approach,’’ *Secur. Commun. Netw.*, vol. 9, no. 18,
pp. 4863–4881, Dec. 2016.

[4] H. S. Milan and K. Singh, ‘‘Reducing false alarms in intrusion detection systems—Asurvey,’’ *Int. Res. J. Eng. Technol.*, 2018.

[5] R. Dwivedi, D. Dave, H. Naik, S. Singhal, R. Omer, P. Patel, B. Qian,

Z. Wen, T. Shah, G. Morgan, and R. Ranjan, ‘‘Explainable AI (XAI): Core ideas,techniques, and solutions,’’ *ACM Comput. Surv.*, vol. 55, no. 9,
pp. 1–33, Sep. 2023.

[6] M. T. Ribeiro, S. Singh, and C. Guestrin, ‘‘Anchors: High-precision model- agnostic explanations,’’ in *Proc. AAAI Conf. Artif. Intell.*, 2018, vol. 18, no. 1, pp. 1527–1535.

[7] L. Antwarg, R. M. Miller, B. Shapira, and L. Rokach, ‘‘Explaining anoma- lies detected by autoencoders using Shapley additive explanations,’’ *Expert Syst. Appl.*, vol. 186,Dec. 2021, Art. no. 115736.

[8] M. A. Salahuddin, M. F. Bari, H. A. Alameddine, V. Pourahmadi, and R. Boutaba, ‘‘Time-based anomaly detection using autoencoder,’’ in *Proc. 16th Int. Conf. Netw. Service Manage. (CNSM)*, Nov. 2020,
pp. 1–9.

[9] A. Takiddin, M. Ismail, U. Zafar, and E. Serpedin, ‘‘Deep autoencoder- based anomaly detection of electricity theft cyberattacks in smart grids,’’ *IEEE Syst. J.*, vol. 16, no.3, pp. 4106–4117, Sep. 2022.

[10] W. Xu, J. Jang-Jaccard, A. Singh, Y. Wei, and F. Sabrina, ‘‘Improv- ing performance of autoencoder-based network anomaly detection on NSL-KDD dataset,’’