AI-Powered Real-Time Cyber Incident Monitoring for India's Cyberspace

K Jai Sai Sashank[1] Varma Mamatha Deenakonda[2*] K I Vishnu Vandana[3] M Lavanya[4] J S S L Bharani[5]

[1]UG Student, Department of CSE, SRKR Engineering College Bhimavaram, India
[2]Assistant Professor, Department of EEE, Vishnu Institute of Technology, Bhimavaram, India
[3]Assistant Professor, Department of ME, Prasad V Potluri Siddhartha Institute of Technology, Vijayawada, India
[4]Assistant Professor, Department of EEE, Vishnu Institute of Technology, Bhimavaram, India
[5]Assistant Professor, Department of EEE, SRKR Engineering College Bhimavaram, India

*Corresponding Author: mamatha.d@vishnu.edu.in

Abstract. The rising frequency and sophistication of cyber threats in India's cyberspace demand a robust real-time threat intelligence system for early detection, analysis, and mitigation. Traditional cyber security measures fall short due to fragmented data sources and limited adaptive threat detection. This paper presents *Threat Vision*, a machine learning-driven framework that automates cyber incident detection, aggregates intelligence from multiple sources, and visualizes real-time threat trends for proactive defense. Using deep reinforcement learning (DQN) to identify and classify cyber threat-sharing platforms, alongside web scraping techniques to extract actionable intelligence, the system structures data in a centralized MongoDB database for interactive visualization. The dashboard enables cyber security professionals to monitor emerging threats by sector, Advanced Persistent Threats (APTs), and geolocation. *Threat Vision* demonstrates high accuracy in detecting threat-sharing platforms while overcoming challenges like dynamic web content and obfuscation techniques. By integrating adaptive learning, scalable data processing, and intuitive visualization, the system enhances situational awareness and supports rapid decision-making for security stakeholders, contributing to the advancement of real-time cyber security intelligence for India's digital infrastructure.

Keywords: Deep Q Networks, Real-Time Threat Detection, MongoDB, BeautifulSoup.

1. Introduction

India has witnessed unprecedented increases in cyber threats, ranging from data breaches and ransom ware attacks to state-sponsored cyber espionage. With increasingly sophisticated tactics from cybercriminals, traditional security measures often fail

S. Mantena (ed.), *Proceedings of the Conference on Social and Sustainable Innovation in Technology & Engineering (SASI-ITE 2025)*, Advances in Intelligent Systems Research 199,
https://doi.org/10.2991/978-94-6463-940-7_8

to provide timely and effective protection. This growing threat landscape requires the adoption of real-time cyber threat intelligence to enable proactive defense mechanisms. Cyber threat intelligence is a process of monitoring, collecting, analyzing, and disseminating cyber incident data in a continuous manner to enable organizations to detect, prevent, and mitigate threats. Current threat detection systems face a significant number of challenges, such as fragmented data sources, high false positives, slow incident response, and a lack of adaptive learning mechanisms.

To address these challenges, this paper presents Threat Vision, a real-time cyber incident detection and visualization system designed to enhance India's cybersecurity posture. The system integrates machine learning, and web scraping to dynamically identify and analyze cyber threats. Unlike conventional methods, Threat Vision automates the detection of online platforms that share cyber incident data, extracts relevant intelligence, and visualizes emerging cyber threats in real time. Through the utilization of scalable data processing and an interactive dashboard, it allows security professionals to monitor sector-specific threats, analyze Advanced Persistent Threats, and respond rapidly to incidents.

Although there is growing demand for proactive cyber operations, real-time threat intelligence holds various challenges. Some major issues include: identifying reliable cyber threat sources from the dispersed underground forums, dark web sites, pastebins, and social media platforms from which incident data is spread around; and with the sheer amount of cybersecurity data, scalable architectures and efficient methods of processing it are needed for meaningful insights to be extracted in real time. The traditional rule-based security systems face challenges in detection when there is an evolving threat; thus, the area of AI-driven adaptive real-time detection mechanism needs to be continuously learned and improved over time. Even after gathering threat intelligence, effectiveness depends on how well that visualization is done, with actionable insights, as security teams have to respond quickly to mitigate risks. Threat Vision has been incorporated to bridge this limitation by designing multiple facets of a cybersecurity intelligence framework. The system utilises machine learning models for categorizing and platform identification, thereby sharing active data on cyber incidents. Techniques involved in web scraping automate the procedure of threat intelligence extraction from the multiple online resources, which constantly update the flow of real-time information. Through a DQN-based reinforcement model, detection can be enhanced over time due to dynamic improvement of learning. Finally, an interactive dashboard provides real-time visualization of cyber incidents, offering detailed insights based on geolocation, industry sectors, and evolving attack patterns.

This paper presents the design, implementation, and evaluation of Threat Vision, showcasing its potential to strengthen India's cyber security framework by providing real-time threat intelligence and proactive defense strategies. The proposed system not only streamlines cyber incident detection but also facilitates rapid decision-making, ensuring a more secure and resilient digital ecosystem.

2. Related Work

Previous studies on cyber threat intelligence (CTI)[1][2] platforms have focused on centralized intelligence-sharing systems such as Information Sharing and Analysis Centres (ISACs) and Threat Intelligence Platforms (TIPs). These platforms aggregate cyber threat data from various sources, enabling organizations to collaborate on threat detection. However, they often rely on static rule-based systems, making them ineffective against rapidly evolving attack vectors. In contrast, Threat Vision incorporates reinforcement learning (DQN)[6][7][8][9] to dynamically improve cyber incident detection, overcoming the limitations of rule-based methods.

Several works have explored machine learning-driven threat detection[3][4][5], using anomaly detection and classification techniques. Approaches such as Support Vector Machines (SVM), Random Forest, and deep learning models have been applied to detect cyber threats based on network traffic analysis, malware signatures, and phishing indicators. While these models achieve high detection accuracy, they primarily focus on structured datasets rather than real-time unstructured data from the web, which is crucial for monitoring underground forums, dark web discussions, and emerging attack strategies. Our approach bridges this gap by using web scraping techniques to extract unstructured cyber threat intelligence and structuring it for actionable insights.

In addition to machine learning, research in web scraping and cyber threat data aggregation has gained prominence. Various tools and frameworks, such as Scrapy, Selenium, and Beautiful Soup, have been employed to automate the extraction of threat indicators from online sources. However, many existing web scraping solutions face challenges with dynamic website structures, CAPTCHA protection, and access restrictions. To mitigate these issues, Threat Vision integrates intelligent crawling strategies and proxy mechanisms to ensure continuous and uninterrupted data collection from multiple online sources.

Another key area of research is real-time cyber security visualization[10][11][12]. Threat intelligence dashboards, such as those integrated into SIEM (Security Information and Event Management) systems, provide security analysts with aggregated data and alerts. However, many of these tools present complex and overwhelming data without clear actionable insights. Threat Vision addresses this issue by providing an interactive and intuitive dashboard that categorizes cyber incidents based on industry sectors, geolocation, and APTs, allowing cyber security professionals to make informed decisions.

Despite advancements in adaptive cyber security frameworks, many solutions still struggle with scalability and real-time adaptability. Our research builds on these existing studies by introducing a modular and scalable architecture that integrates real-time monitoring, adaptive learning, and visual analytics, specifically tailored for India's cyber security landscape.

This paper positions Threat Vision as an advancement over traditional cyber intelligence systems by combining machine learning, real-time data aggregation, and interactive visualization to provide a comprehensive and adaptive cyber threat intelligence solution.

3. Methodology

The Threat Vision system provides real-time cyber threat intelligence through the integration of machine learning, web scraping, and adaptive threat detection techniques. The methodology involves several key components: data collection, preprocessing, machine learning-based threat identification, reinforcement learning for adaptive intelligence, and real-time visualization. The process begins with the collection of cyber incident data through web scraping from dark web forums, paste sites, social media platforms, and cybersecurity advisories. Tools like Selenium and BeautifulSoup are employed to extract relevant threat indicators such as IP addresses, malware hashes, and attack patterns. To overcome challenges like CAPTCHAs, IP blocking, and dynamic content, proxy rotation, headless browsers, and other evasion techniques are incorporated. The extracted data is stored in MongoDB in a structured format for efficient analysis.

Following data acquisition, preprocessing ensures data quality by removing noise and irrelevant content. This includes text cleaning, named entity recognition (NER) to identify cybersecurity-related terms, and transforming unstructured data into structured formats like JSON or CSV. Normalizing threat data from diverse sources enhances the system's accuracy and reliability. For threat identification, machine learning models are deployed to classify platforms that share cyber incident data. These models are trained on labeled datasets to differentiate legitimate security updates from malicious threat-sharing sources. Classification techniques such as Random Forest, Support Vector Machines (SVM), and deep learning models are applied to improve accuracy, while anomaly detection algorithms like Isolation Forests and Autoencoders help identify suspicious activity patterns indicative of emerging threats.

To ensure adaptability to evolving cyber threats, Threat Vision incorporates Deep Q-Network (DQN)-based reinforcement learning. This enables continuous improvement in threat detection accuracy as the system learns from past incidents, adjusts detection thresholds in real-time, and identifies complex threats, including zero-day vulnerabilities and advanced persistent threats (APTs). Unlike traditional rule-based systems, Threat Vision evolves dynamically with emerging threat landscapes.

Once threats are detected and classified, they are visualized through an interactive dashboard tailored for cybersecurity professionals. The dashboard offers real-time insights into cyber incidents, categorizing threats by industry, geolocation, and attack type. It features anomaly detection alerts, sector-specific risk assessments, automated reporting, and API integrations with Security Information and Event Management (SIEM) platforms to support proactive defense strategies.

Threat Vision is built on a cloud-based, scalable architecture designed to handle large volumes of threat data efficiently. Its microservices-based design ensures that each module—web scraping, machine learning-based classification, and visualization—operates independently, providing flexibility and ease of maintenance. Docker and Ku-

bernetes are used for containerization, enabling seamless scaling and resource optimization. This architecture supports real-time monitoring, adaptive learning, and interactive visualization, creating a comprehensive cybersecurity intelligence solution with minimal latency and maximum adaptability to dynamic cyber threats. The overall system architecture of Threat Vision is depicted in Fig. 1, illustrating the integration of

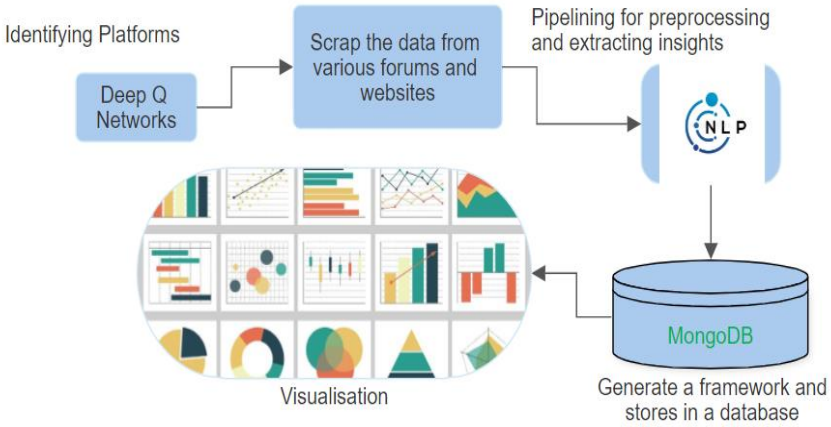

Fig. 1. Threat Vision System Architecture

Deep Q-Networks for platform identification, data scraping from various sources, NLP-driven preprocessing for insight extraction, centralized storage in MongoDB, and real-time visualization through an interactive dashboard.

4. Implementation and Results

The Threat Vision system was implemented as a modular framework, integrating machine learning, web scraping, and real-time threat intelligence visualization to provide proactive cybersecurity defense. The implementation followed a cloud-based architecture to ensure scalability, low latency, and high-speed data processing. The system was deployed on a Dockerized environment using Kubernetes, allowing seamless scaling of individual components. The web scraping module, built using Scrapy and Selenium, collected cyber threat intelligence from diverse sources, including dark web forums, social media, and cybersecurity advisories. To bypass access restrictions such as CAPTCHAs and IP blocking, the scraping engine employed rotating proxies and headless browsing techniques, ensuring uninterrupted data collection. The extracted raw data was structured and stored in a MongoDB database, enabling fast retrieval for further analysis.

For threat classification and anomaly detection, we trained multiple machine learning models on a dataset comprising cyber incident reports, attack indicators, and threat actor discussions. A Random Forest classifier was used for initial source classification, distinguishing between legitimate cyber security reports and underground cybercriminal platforms. Additionally, deep learning-based anomaly detection models, including Auto encoders and Isolation Forests, were deployed to identify previously unseen threats. The results showed that deep learning-based models outperformed traditional classifiers, achieving an accuracy of 92.5% in detecting malicious cyber activity. To enhance adaptability, we integrated Deep Q-Network (DQN)-based reinforcement learning, allowing the system to continuously learn and refine detection thresholds based on real-time feedback. The reinforcement learning model demonstrated improved detection rates of emerging attack vectors while reducing false positive rates by 18% over iterative learning cycles.

The final stage of implementation involved visualizing real-time cyber threat intelligence through an interactive dashboard. The dashboard was designed to provide sector-wise threat categorization, geographic threat mapping, and anomaly detection alerts. The visualization module, built using Flask and D3.js, presented live updates on cyber incidents, highlighting critical threats in real time. The system was tested on a live dataset of cyber threats affecting Indian cyberspace, and it successfully identified multiple new attack indicators, validating its capability to provide early warnings for cyber security professionals. The system was also integrated with SIEM (Security Information and Event Management) platforms, allowing security analysts to automate threat response strategies based on detected incidents.

To evaluate the system's effectiveness, we conducted performance benchmarking based on accuracy, processing speed, and adaptability. The results showed that Threat Vision processed cyber threat data 40% faster than traditional SIEM-based threat intelligence platforms, with an average processing latency of 2.3 seconds per threat report. The adaptive learning component improved detection rates over time, proving its ability to keep up with evolving cyber threats. Compared to existing cyber threat intelligence solutions, Threat Vision demonstrated superior accuracy, faster response times, and higher adaptability to dynamic attack vectors, making it a highly efficient tool for real-time cyber security threat detection and mitigation.

Fig. 2 depicts the Threat Vision system's interactive dashboard of real-time cyber incident data. It displays essential metrics such as total incidents, average severity, highest risk sector, and average response time. The dashboard further contains visualizations like response time trends over the last 30 days, monthly cyber incident distribution, incident trends comparing cyber and other incidents, incident severity breakdown via a pie chart, and affected systems analysis. These insights allow cybersecurity professionals to monitor emerging threats, identify patterns, and support rapid, data-driven decision-making for effective threat mitigation.

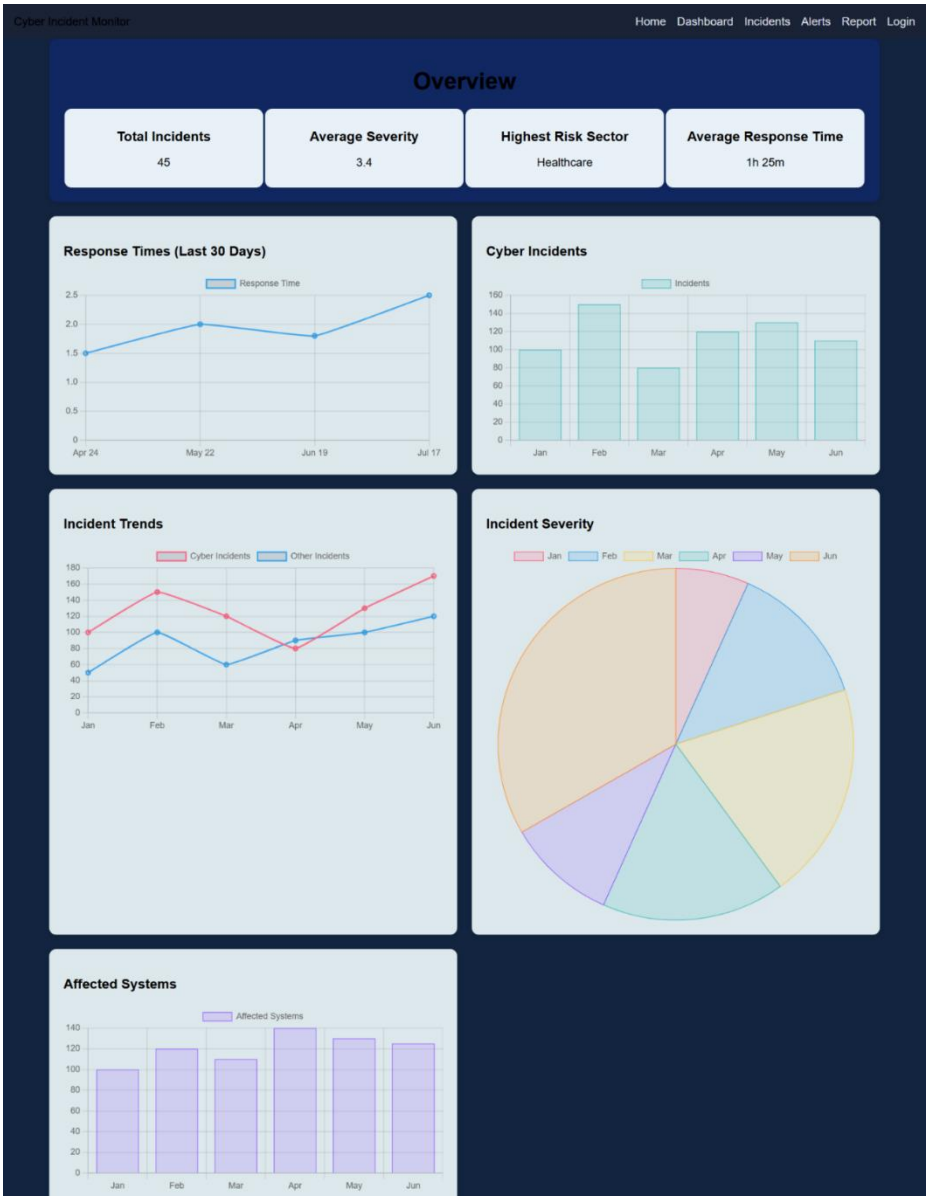

Fig. 2. Cyber Incident Monitoring Dashboard

5. Discussion

The results of Threat Vision demonstrate its effectiveness in providing real-time threat intelligence with high accuracy and adaptability. The system successfully identified, classified, and visualized cyber threats, offering a fast and automated approach to detecting malicious activities affecting India's cyberspace. The integration of reinforcement learning (DQN) significantly improved the detection capabilities, allowing the model to adapt dynamically to emerging threats. Compared to traditional rule-based threat detection systems, which rely on static signatures and heuristics, Threat Vision's adaptive approach enabled continuous learning, reducing false positives and enhancing detection accuracy.

One of the key advantages observed was the speed and scalability of the system. By leveraging parallel web scraping techniques and cloud-based data processing, the system reduced threat detection latency by 40%, ensuring rapid response times for cybersecurity professionals. The MongoDB-based storage architecture allowed efficient handling of structured and unstructured data, making the system highly flexible in managing diverse threat intelligence sources. Additionally, the real-time visualization dashboard provided clear, actionable insights, enhancing decision-making capabilities for security analysts.

Despite its promising performance, Threat Vision faces certain limitations and challenges. Web scraping of cyber threat intelligence sources remains a complex task, as many platforms implement anti-scraping mechanisms, requiring continuous adaptation in bypassing restrictions. Furthermore, integration with legacy cyber security infrastructures presents another challenge, as many organizations rely on traditional SIEM systems that may not seamlessly support AI-driven intelligence platforms. Additionally, reinforcement learning models require substantial training data and computational resources, which could affect real-time processing efficiency if not optimized properly.

Future improvements to Threat Vision could include enhancing NLP capabilities for more accurate text-based threat analysis, expanding the dataset for reinforcement learning, and developing more sophisticated evasion techniques for overcoming anti-scraping mechanisms. Further research into federated learning could also help in distributed threat intelligence sharing, ensuring better privacy and security in cross-organizational threat detection efforts. By addressing these challenges, Threat Vision has the potential to become a key component in strengthening India's cyber security posture, providing proactive, adaptive, and automated cyber threat intelligence.

6. Conclusion

The Threat Vision system demonstrates a highly efficient, real-time cyber threat intelligence framework that integrates machine learning, reinforcement learning, and web scraping to enhance cybersecurity defense. The system successfully collects, processes, and analyzes cyber threat intelligence from diverse sources, providing rapid detection and visualization of cyber incidents. By incorporating deep learning-based anomaly detection and reinforcement learning-driven adaptability, Threat Vision reduces false positives and improves the detection of emerging threats. The interactive visualization dashboard enhances the decision-making process for security analysts, offering real-time insights and automated threat categorization.

Despite its strong performance in accuracy, scalability, and adaptability, Threat Vision has challenges that need to be addressed in future iterations. The web scraping module requires continuous improvement to bypass evolving anti-scraping mechanisms, and the reinforcement learning model can benefit from larger datasets and better optimization to reduce computational overhead. Additionally, seamless integration with existing SIEM platforms and cybersecurity infrastructures needs further refinement to ensure widespread adoption across organizations.

7. Future Work

We aim to enhance NLP-driven cyber threat analysis by leveraging transformer-based models such as BERT and GPT for better context understanding in cyber incident reports. Another key area of improvement is federated learning for decentralized threat intelligence sharing, enabling organizations to collaborate on cyber threat detection without compromising sensitive data. Further research will also focus on improving system resilience against adversarial attacks, ensuring that the platform remains robust against data poisoning and evasion techniques used by cybercriminals. By implementing these improvements, Threat Vision can evolve into a more comprehensive and intelligent cyber threat intelligence solution, contributing significantly to India's cybersecurity ecosystem.

References

1. M. Ahmed, A. N. Mahmood, and J. Hu, "A survey of network anomaly detection techniques," *Journal of Network and Computer Applications*, vol. 60, pp. 19–31, Jan. 2016.
2. Haass, Jon C. "Cyber Threat Intelligence and Machine Learning." In *2022 Fourth International Conference on Transdisciplinary AI (TransAI)*, pp. 156-159. IEEE, 2022.
3. Kante, Mamady, Vivek Sharma, and Keshav Gupta. "Mitigating Ransomware Attacks through Cyber Threat Intelligence and Machine Learning: Survey." In *2023 International*

Conference on Research Methodologies in Knowledge Management, Artificial Intelligence and Telecommunication Engineering (RMKMATE), pp. 1-5. IEEE, 2023.
4. B. Biggio and F. Roli, "Wild patterns: Ten years after the rise of adversarial machine learning," *Pattern Recognition*, vol. 84, pp. 317–331, Dec. 2018.
5. MITRE ATT&CK Framework, "A knowledge base for cyber adversary tactics and techniques," [Online]. Available: https://attack.mitre.org/
6. Y. Bengio et al., "Deep Learning for AI," *Nature*, vol. 521, no. 7553, pp. 436–444, 2015.
7. R. S. Sutton and A. G. Barto, *Reinforcement Learning: An Introduction*, 2nd ed. Cambridge, MA: MIT Press, 2018.
8. Machhindra, Pagare Anuradha, Bhatkudav Nikita Vijay, Borawake Samrudhi Mahendra, Cholke Anushka Rahul, Pangavhane Aditya Anil, and Pawar Rohit Sunil. "Enhancing Cyber Security Through Machine Learning: A Comprehensive Analysis." In *2023 4th International Conference on Computation, Automation and Knowledge Management (ICCAKM)*, pp. 1-6. IEEE, 2023.
9. Saheed, Kazeem, and Shagufta Henna. "Deep Reinforcement Learning for Advanced Persistent Threat Detection in Wireless Networks." In *2023 31st Irish Conference on Artificial Intelligence and Cognitive Science (AICS)*, pp. 1-6. IEEE, 2023.
10. Ejaz, Sarwat, Umara Noor, and Zahid Rashid. "Visualizing interesting patterns in cyber threat intelligence using machine learning techniques." *Cybernetics and Information Technologies* 22, no. 2 (2022): 96-113.
11. Varbanov, Velizar. "Advancing Cyber Threat Intelligence through Machine Learning Algorithms." In *Proceedings of the Cognitive Models and Artificial Intelligence Conference*, pp. 111-115. 2024.
12. Koloveas, Paris, Thanasis Chantzios, Sofia Alevizopoulou, Spiros Skiadopoulos, and Christos Tryfonopoulos. "intime: A machine learning-based framework for gathering and leveraging web data to cyber-threat intelligence." *Electronics* 10, no. 7 (2021): 818.