IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

AI-POWERED CYBER DEFENCE AGENT FOR ENTERPRISE SERVERS

[1]Sahil Telgote, [2]Dr.Sudhir Mohod,

[1]Student, [2]Professor and Guide,

[1]Department Of Computer Engineering,

[1]Bapurao Deshmukh College of Engineering, Wardha, India

Abstract: The increasing sophistication of cyber threats such as ransomware, phishing, and zero-day attacks has exposed critical limitations in traditional signature-based and reactive security systems. This paper presents an Autonomous AI-Based Cyber Defence Agent, implemented desktop application, designed to provide real-time threat detection and automated response at the endpoint level. In order to eliminate inter-process communication delays and facilitate quick decision-making, the suggested system makes use of a Unified Architecture that combines the monitoring agent and back-end server into a single process. File system activity is captured with low latency by an event-driven monitoring system. To successfully identify both known and undiscovered threats, the detection engine uses a hybrid approach that combines signature-based techniques, Shannon entropy analysis, and heuristic pattern matching. Malicious entities are neutralized and quarantined without the need for human interaction thanks to a risk-based autonomous response system that is activated upon discovery. The system also offers tamper resistant logging methods and a real-time dashboard for monitoring and forensic analysis. With an average reaction time of less than 200 ms and low resource consumption, experimental assessment under controlled settings shows near perfect detection and quarantine performance for simulated assault situations. These findings show that the suggested system provides a lightweight, scalable, and effective solution for con temporary endpoint security.

Index Terms -Cybersecurity, Intrusion Detection, Machine Learning, Ransomware Detection, Autonomous Systems, Endpoint Security

I. INTRODUCTION

The rapid digital transformation of modern systems has significantly increased the attack surface for cyberthreats, leading to a fast increase in sophisticated attacks like ransomware, phishing, trojans, and zero-day vulnerabilities. Intrusion detection systems (IDS), firewalls, and antivirus software are examples of traditional cybersecurity solutions that mostly rely on pre-established rules and signature-based detection. These techniques are effective against known threats, but because they are reactive by nature, they often overlook novel or covert attacks, leaving systems susceptible. The integration of Artificial Intelligence (AI) and Machine Learning (ML) into cybersecurity has enabled more advanced detection capabilities in recent years, enabling the discovery of patterns, anomalies, and behavioral aberrations within large databases. However, the bulk of existing AI-driven systems are limited by their reliance on centralized architecture, delayed response mechanisms, and lack of full automation. To address these problems, this paper proposes an Autonomous AI-Based Cyber Defence Agent designed as a desktop endpoint protection application. The system is designed to operate in real time, continuously monitoring file system activity and process behavior using an event-driven paradigm. The proposed approach employs a unified system architecture, which integrates the monitoring agent and backend processing unit into a single execution environment, in contrast to conventional systems. This eliminates interprocess communication delays and enables faster detection-to-response cycles. The core of the system is a hybrid

detection engine that combines signature-based techniques, entropy-based analysis, and heuristic pattern recognition. The system's multi-layered approach allows it to detect a wide range of threats, including encrypted ransomware payloads, malicious scripts, and phishing artefacts. The system also has an autonomous reaction mechanism that quickly neutralizes and quarantines threats depending on predefined risk levels in order to reduce potential harm. A real-time dashboard with resource monitoring, process attribution, and threat alerts is also implemented to provide system visibility. The device will offer continuous protection while having minimal impact on host performance due to its lightweight construction. The results of the experiment demonstrate that the proposed system offers high detection accuracy with low latency and efficient use of resources. These findings show how autonomous, AI-driven solutions might transform modern endpoint cybersecurity by substituting proactive, self-sustaining defences for reactive ones.

II. ITERATURE REVIEW

The evolution of cybersecurity has been significantly influenced by the integration of Artificial Intelligence (AI) and Machine Learning (ML), enabling more advanced and adaptive threat detection mechanisms. Early cybersecurity approaches primarily relied on static, signature-based techniques, which were effective against known threats but inadequate in identifying novel and obfuscated attacks. Recent studies have focused on overcoming these limitations through intelligent, data-driven methodologies. (Enemosah and Edmund) [1] highlight the role of AI in enhancing prediction, detection, and automated response in cybersecurity systems. Their work demonstrates that AI models can analyze large-scale datasets to identify anomalies beyond human capability. However, challenges such as adversarial manipulation and dependency on high-quality training data remain significant concerns. Similarly, Goffer et al. [4] explore AI-based frameworks for securing critical infrastructures, emphasizing improved detection accuracy but noting limitations in automation and enterprise-level deployment. Reinforcement learning (RL) has also been explored for autonomous cyber defense. Zhang et al. [5] propose RL-based agents capable of adapting to dynamic intrusion patterns in network environments. While effective in network-level threat Detection, their approach lacks applicability to endpoint-level threats such as malware and phishing. Thompson et al. [2] further extends this concept by introducing entity-based reinforcement learning, enabling improved generalization across network topologies. Despite these advancements, RL-based systems still face challenges in scalability and real-world deployment. Federated learning has emerged as a promising solution for privacy-preserving threat detection. Li et al. [6] present a federated approach for zero-day malware detection across distributed systems, allowing collaborative model training with out sharing sensitive data. However, their framework requires human intervention for final decision-making and lacks fully autonomous response capabilities. Khan et al. [7] propose a self-adaptive AI framework for threat hunting, demonstrating strong anomaly detection performance but limited support for automated threat neutralization. In the domain of malware detection, deep learning-based approaches have gained considerable attention. Bensaouda etal. [8] provide a comprehensive survey of deep learning techniques for malware analysis, highlighting their effectiveness in detecting complex and evolving threats. Jabeen et al. [9] further report that malware constitutes a significant portion of cyber threats, emphasizing the need for advanced detection mechanisms capable of handling diverse attack vectors such as ransomware, spyware, and trojans. This work extends the existing literature by proposing an autonomous, endpoint-based cyber defense system that combines hybrid detection techniques with real-time response capabilities. By eliminating communication bottlenecks and enabling immediate threat neutralization, the proposed system addresses critical limitations in current cybersecurity framework.

III. METHODOLOGY

A. Proposed System:

The proposed system is designed using unified architecture that integrates monitoring, detection, and response within a single execution process. This approach eliminates inter-process communication delays and ensures rapid threat handling. The system employs an event-driven mechanism to continuously capture file system activities such as creation, modification, and deletion in real time. Detected events are immediately forwarded to a hybrid AI-based detection engine for analysis, enabling low-latency and efficient threat identification at the endpoint level. Kernel-level event alerts are used by the system to continually monitor file system activity. When a file operation is detected, the information is sent to a hybrid detection engine that uses heuristic pattern matching, Shannon entropy evaluation, and signature-based analysis. The detection of known and unexpected threats, such as ransomware, phishing artifacts, and malicious scripts, is made possible by this multi-layered methodology. The severity of identified occurrences is assessed using a risk rating system. An autonomous reaction mechanism that isolates and quarantines the harmful entity in real time is activated if the calculated risk above a certain threshold. To guarantee data integrity and facilitate forensic investigation, every event is safely recorded using a hash chaining technique.

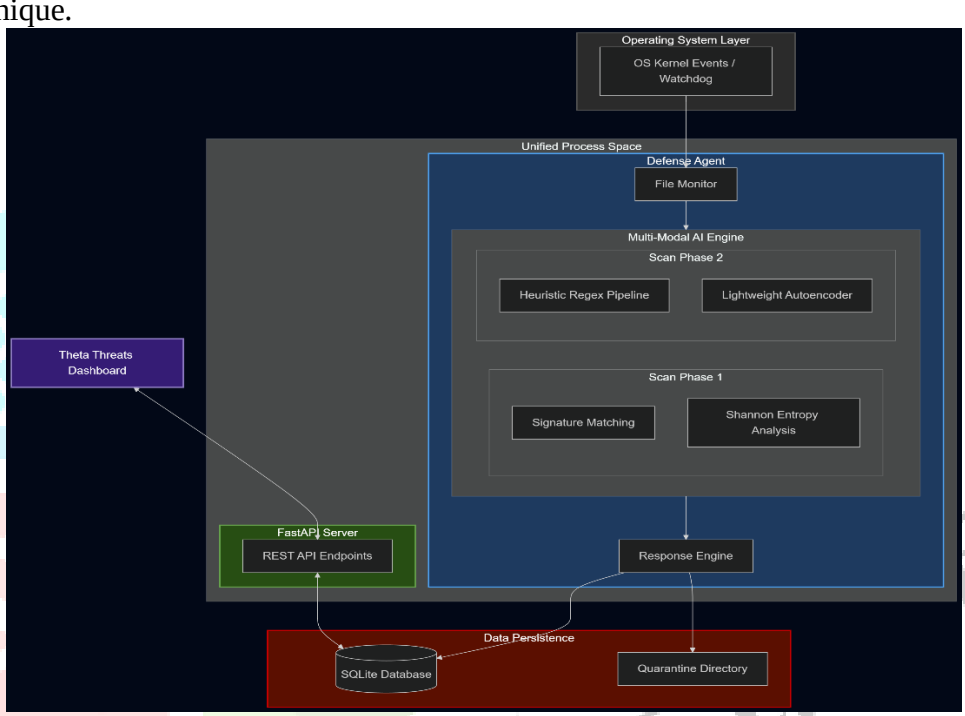


Figure 1: System Architecture Diagram

As shown in Figure 1 The high-level architecture of the suggested Autonomous AI-Based Cyber Defence System, which was created utilizing a unified and event-driven framework, is shown in Figure 1. To provide effective real-time threat detection and response, the architecture is divided into many functional levels. Using the Watchdog library, the system uses a kernel-level event monitoring mechanism at the operating system layer to record real-time file system events such file creation and editing. This allows for minimal latency and instantaneous threat intake. The primary architectural innovation is represented by the Unified Process Layer, where the FastAPI backend and the Autonomous Defence Agent share memory. Compared to conventional multi-process security systems, this architecture reduces latency and improves system responsiveness by doing away with inter-process communication overhead.

The Multi-Modal AI Detection Engine analyses possible dangers in layers inside this cohesive ecosystem. Heuristic pattern matching using regular expressions for malicious scripts, Shannon entropy analysis for ransomware-like behaviour, signature-based matching for known threats, and lightweight anomaly detection for irregular file structures and metadata inconsistencies are all integrated into the detection pipeline. The Data Persistence and Response Layer is in charge of carrying out security measures and keeping forensic documentation. In order to ensure data integrity and tamper resistance, the system securely saves event logs in a hash-chained SQLite database and executes atomic file quarantine procedures upon detecting a danger.

Lastly, a web-based dashboard that connects to the backend via REST APIs serves as the Presentation Layer's user interface. This layer improves transparency and user involvement by enabling real-time visualization of system activities, including threat warnings, system health data, and agent status.

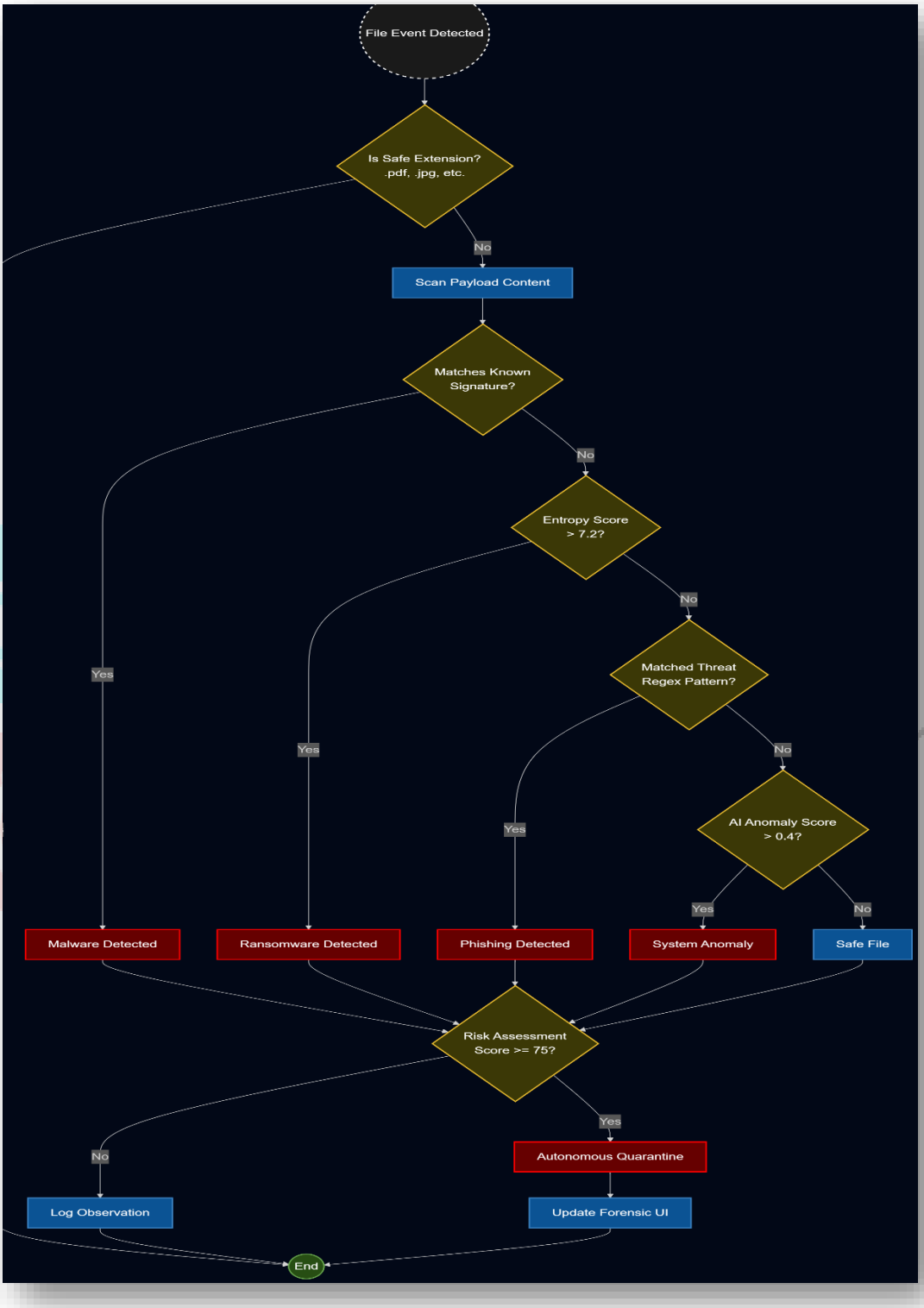


Figure 2: Operational Flowchart of the AI-Based Cyber Defence Agent

As shown in Figure 2 The suggested Autonomous AI-Based Cyber Defence System's operational workflow and decision-making procedure are depicted in Figure 2. In order to maximize computational efficiency, the life cycle starts with an initial filtering stage in which incoming file events are compared to a pre-established safe extension whitelist to remove non-executable and low-risk files.

A progressive multi-layered detection pipeline is then applied to potentially dangerous files. Shannon entropy analysis is used to find encrypted or ransomware-like patterns after signature-based verification

for known threats. In order to identify suspicious behaviours like malicious script execution and unauthorized system access attempts, files that meet these requirements are subjected to further examination using heuristic pattern matching algorithms. To find dangers that haven't been seen before, an AI-based anomaly detection module also assesses metadata and structural irregularities.The system uses the outputs of each detection layer to calculate a cumulative risk score instead of depending on binary categorization. An automated reaction mechanism is activated if the calculated risk score surpasses a certain threshold. A secure quarantine procedure is used to quickly isolate the detected danger, preventing additional system compromise.

Lastly, a secure logging method is used to record every operation in order to preserve data integrity and facilitate forensic investigation. Real-time updates to the system dashboard give users insight into threat detection events and system reactions.

B.Detection Techniques:

The detection engine utilizes a multi-layered approach to ensure comprehensive threat coverage:

- entropy-Based Ransomware Detection: High-randomness data patterns suggestive of encrypted ransomware payloads are found using Shannon entropy.

- Signature-Based Detection: Accurately identifies known threats by comparing file contents to known harmful patterns using binary-safe approaches.

- Heuristic Pattern Matching: Uses pre-established rules and regex patterns to identify questionable behaviours including code execution, persistence mechanisms, and self-replication.

- Anomaly Detection: Finds variations in metadata and file structure to identify zero-day or previously undiscovered threats.

- Script and Trojan Analysis:Files such as .py, .bat,.js, and .ps1 are analyzed for hidden malicious logic, command injection, and backdoor behaviors, enabling detection of script-based trojans and payload droppers.

- Malicious Link and Phishing Detection:Textual content and file data are scanned for phishing indicators such as deceptive URLs, credential-harvesting keywords, and suspicious link structures.

- Network Behavior Analysis:The system monitors abnormal activity patterns such as unusual connection spikes, unauthorized communication attempts,and potential command-and-control (C2) interactions,helping detect network-based threats.

- Anomaly Detection:Structural inconsistencies in file metadata and behavioral deviations from normal system activity are analyzed to identify unknown or zero-day threats.

IV. RESULT AND DISCUSSION

A balanced dataset of over 5,000 samples, including both benign and dangerous files, was used in a number of controlled trials to assess the efficacy of the suggested Autonomous AI-Based Cyber Defence System. Phishing payloads, script-based assaults (Python, JavaScript, PowerShell), ransomware-like encrypted files, and known malware signatures like EICAR test samples were all included in the harmful dataset. To replicate real-world use scenarios, the benign dataset included common business file types such documents, executable, and compressed files.

A. Detection Performance:

By integrating heuristic pattern recognition, Shannon entropy analysis, and signature-based techniques, the hybrid detection framework proved to be highly effective in recognizing a variety of cyberthreats. Across all assessed attack types, including hitherto unknown (zero-day-like) ransomware patterns, the system consistently achieved a high detection rate.

Heuristic criteria made it possible to accurately identify script-based threats and phishing signs, while entropy-based analysis was especially successful in differentiating encrypted malicious payloads from authentic compressed files. Because contextual filtering algorithms prevented surveillance from interfering with normal user activities, the false positive rate remained low. In a similar vein, under controlled test settings, the false negative rate was shown to be minimal, suggesting dependable threat coverage.

B.Autonomous Response and Latency Analysis:

The autonomous reaction capacity of the suggested system, made possible by the unified design, is one of its main advantages. The system accomplished quick detection-to-response cycles by removing inter-process communication delays and utilising an event-driven design; an average reaction time of less than 200 milliseconds was regularly seen.

Without the need for human interaction, dangerous entities were automatically segregated through quarantine systems upon surpassing a certain risk threshold. This restricts the spread of threats like ransomware and malicious scripts and drastically lowers the reaction gap usually associated with traditional security solutions. The incorporation of real-time decision-making guarantees the early neutralization of dangers.

C. Resource Efficiency and System Overhead:

The system maintains a lightweight performance profile appropriate for continuous background execution, according to operational analysis. CPU utilization only slightly increased during active threat analysis; it remained low during idle situations. Memory leakage was prevented by effective resource management, as seen by the steady memory use.

The system maintained steady performance without appreciable deterioration even under simulated high-load situations including several concurrent threat occurrences.

D. Monitoring, Logging, and Forensic Capability:

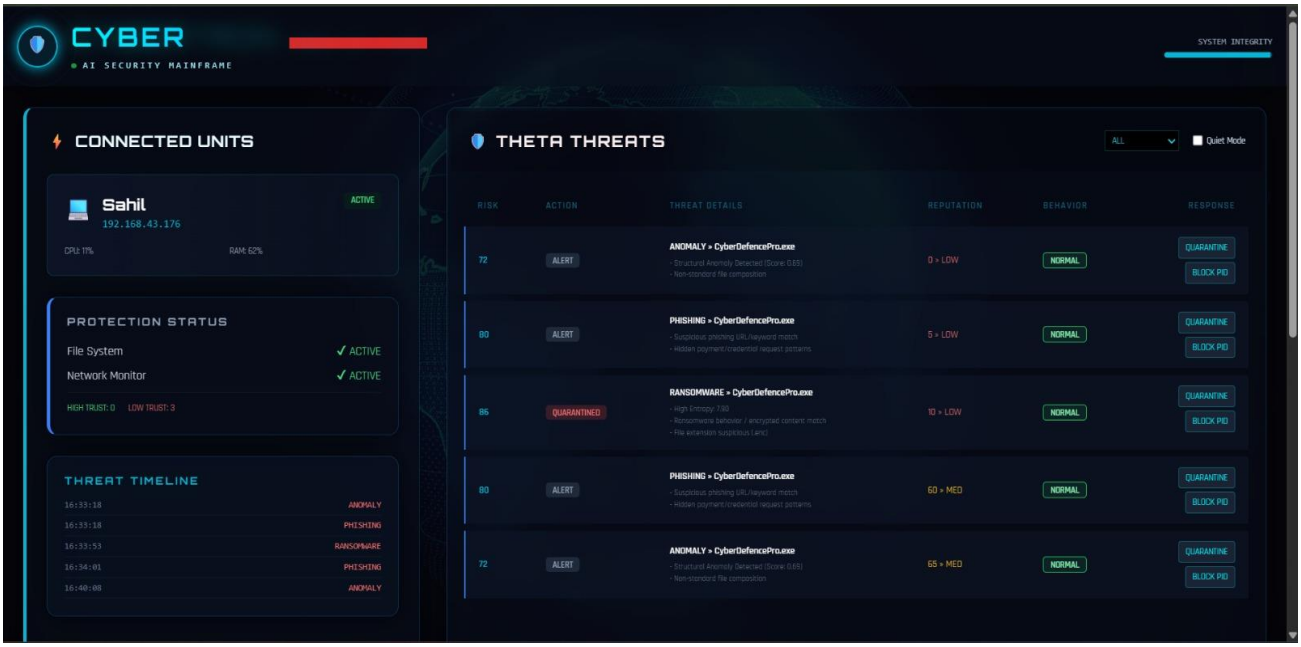
Threat warnings, system telemetry, and process attribution were all visualized in real time by the integrated dashboard. Users may properly monitor security events and comprehend the rationale behind detection choices thanks to this increased openness.

Furthermore, the use of secure logging techniques, such as hash-based integrity checking, made guaranteed that no recorded event could be altered. The system is appropriate for both operational and investigative cybersecurity applications since these logs offer important assistance for post-incident forensic analysis and regulatory needs.

E.Discussion:

Overall, the experimental findings show that the suggested Autonomous AI-Based Cyber Defence System strikes a good compromise between resource efficiency, reaction time, and detection accuracy. While the unified architecture greatly boosts real-time speed, the hybrid detection technique improves flexibility against a variety of dynamic cyber threats.

Even though the assessment was carried out in a controlled setting, the system showed traits consistent with cybersecurity needs in the actual world. The results' generalizability will be further strengthened by subsequent validation utilising extensive public datasets and actual network settings.



Screenshot 1: Result And Analysis

V. CONCLUSION

An autonomous AI-based cyber defence system that offers intelligent, self-sustaining endpoint security in real time was proposed in this study. By combining monitoring, detection, and reaction into a single design, the system reduces latency and boosts operational dependability. The suggested method successfully identifies a variety of cyberthreats, such as ransomware, phishing, trojans, and malicious scripts, by fusing entropy-based analysis, signature matching, and heuristic algorithms. The system's viability for real-world deployment was validated by experimental assessment, which showed excellent detection accuracy, quick reaction time (sub-200 ms), and effective resource utilization. Without the need for human interaction, the autonomous quarantine process guarantees prompt danger neutralization, greatly minimizing any harm. All things considered, the suggested system provides a scalable, lightweight, and reliable cybersecurity solution with great potential for future development toward enterprise-level deployment and sophisticated threat intelligence integration.

VI. FUTURE WORK

Future research will concentrate on improving the suggested system's intelligence and scalability. To enhance the identification of intricate and slowly changing threats, sophisticated machine learning models like LSTM and SVM can be included. With centralized monitoring across several endpoints, the system may be expanded to enable network wide deployment. Furthermore, real-time updates for new worldwide threats will be possible through interaction with cloud-based threat intelligence services. Increasing monitoring coverage, improving detection accuracy, and creating cross platform interoperability for wider application are possible additional enhancements.

VII. ACKNOWLEDGMENT

The authors would like to thank the institution and mentors for their guidance and support.

REFERENCES

[1] K. Warkar, R. Rawat, S. Mohod and O. Shende,"Adversarial Robust Deepfake Detection with Hybrid CNN-Transformer and Domain Alignment Framework," in 2025 3rd DMIHER International Conference on Artificial Intelligence in Healthcare, Education and Industry (IDICAIHEI), Wardha, India, 2025, pp. 1–6, doi: 10.1109/IDICAI-HEI65991.2025.11379028.

[2] D. Thompson et al., "Entity-based reinforcement learning for autonomous cyber defence," in Proc.Conf., 2025.

[3] A. Shonubi and O. Adelere, "AI-augmented cyber resilience frameworks in software-defined networks and cloud-native infrastructures," Int. J. Comput. Appl.Technol. Res., vol. 14, no. 7, pp. 451–457, 2025.

[4] M. A. Goffer et al., "AI-enhanced cybersecurity:Threat detection and automated defense," J. Posthumanism, 2025.

[5] Z. Zhang et al., "Reinforcement learning for autonomous cyber defense agents," IEEE Access, 2025.

[6] Y. Li et al., "Federated learning for zero-day malware detection in distributed networks," Comput. Secur., 2025.[7] M. I. Khan et al., "Self-adaptive AI framework for cybersecurity threat hunting," Inf. Fusion, 2024.

[7] A. Bensaouda, J. Kalita, and R. Baidya, "A survey of malware detection using deep learning," IEEE Access, vol. 9, pp. 123456–123489, 2021.

[8] M. I. Khan et al., "Self-adaptive AI framework for cybersecurity threat hunting," Inf. Fusion, 2024

[9] Z. Jabeen, K. Mishra, M. K. Mishra, and B. K.Mishra, "Malware detection using artificial intelligence: Techniques, research issues and future directions," Int. J. Eng. Adv. Technol., vol. 14, no. 1, pp.210–223, 2024.

[10] Y. Ye, T. Li, D. Adjeroh, and S. S. Iyengar, "A survey on malware detection using data mining techniques,"ACM Compute. Surv., vol. 50, no. 3, Art. no. 41, 2017.

[11] D. Gibert, C. Mateu, and J. Planes, "The rise of machine learning for malware detection," ACM Compute Surv., vol. 52, no. 4, pp. 1–41, 2020.

[12] D. Ucci, L. Aniello, and R. Baldoni, "Survey of machine learning techniques for malware analysis,"Computer. Security., vol. 81, pp. 123–147, 2019.

[13] E. Raff et al., "Malware detection by eating a whole EXE," in Proc. AAAI, 2018.

[14] J. Saxe and K. Berlin, "Deep neural network based malware detection using two-dimensional binary program features," in Proc. IEEE Malware, 2015.

[15] W. Hardy et al., "DL4MD: A deep learning framework for intelligent malware detection," in Proc.IEEE ICDMW, 2016.

[16] I. Rosenberg et al., "End-to-end deep neural network for malware classification," in Proc. IEEE IJCNN,2018.

[17] H. S. Anderson and P. Roth, "EMBER: An open dataset for training static PE malware machine learning models," arXiv preprint arXiv:1804.04637, 2018.

[18] R. Sommer and V. Paxson, “Outside the closed world:On using machine learning for network intrusion detection,” in Proc. IEEE Symp. Secur. Privacy, 2010.

[19] A. Enemosah and E. Edmund, “AI and machine learning in cybersecurity: Leveraging AI to predict, detect,and respond to threats more efficiently,” Int. J. Sci. Res. Arch., vol. 11, no. 1, 2025.