International Journal of Engineering & Extended Technologies Research (IJEETR)

|ISSN: 2322-0163| www.ijeetr.com | A Bimonthly, Peer Reviewed, Scholarly Indexed Journal |

| Volume 8, Issue 2, March - April 2026 |

DOI:10.15662/IJEETR.2026.0802064

NexusSafe: An XAI Powered Cyber Security for Malicious URL Detection

S. Satheeshkumar, P.Keerthika, T.Rakshana, M. Aysha Sugaina, M.Kiruthika

Assistant Professor, Department of Information Technology, Vivekanandha College of Technology for Women, Tiruchengode, Namakkal, Tamil Nadu, India

Department of Information Technology, Vivekanandha College of Technology for Women, Tiruchengode, Namakkal, Tamil Nadu, India

Department of Information Technology, Vivekanandha College of Technology for Women, Tiruchengode, Namakkal, Tamil Nadu, India

Department of Information Technology, Vivekanandha College of Technology for Women, Tiruchengode, Namakkal, Tamil Nadu, India

Department of Information Technology, Vivekanandha College of Technology for Women, Tiruchengode, Namakkal, Tamil Nadu, India

Publication History: Received: 25.02.2026; Revised: 20.03.2026; Accepted: 25.03. 2026; Published: 28.03.2026.

ABSTRACT: NexusSafe is a real-time malicious URL detection system designed to enhance web security using Explainable Artificial Intelligence (XAI). The system integrates a browser-based Chrome Extension with a FastAPI backend to monitor and analyze URLs visited by users. It employs a multi- layered detection approach that includes lexical feature analysis, infrastructure evaluation, behavioral inspection, and threat intelligence integration using external platforms. Unlike traditional blacklist-based systems, which fail to detect newly generated malicious URLs, NexusSafe uses intelligent techniques to identify both known and unknown threats effectively. Additionally, the system generates human-readable explanations for each prediction, allowing users to understand the reasons behind classifying a URL as safe or malicious. This combination of real-time detection, accuracy, and transparency makes NexusSafe a reliable and user-friendly solution for modern cybersecurity challenges.

KEYWORDS: Malicious URL Detection, Explainable Artificial Intelligence (XAI), Phishing Detection, Cybersecurity, Machine Learning, Real-time Analysis, URL Feature Extraction

I. INTRODUCTION

The rapid expansion of internet usage and online services has significantly increased the risk of cyber threats, particularly those involving malicious URLs. These URLs are commonly used in phishing attacks, malware distribution, and fraudulent websites to deceive users and steal sensitive information such as login credentials, banking details, and personal data. Many users are unaware of these threats and often access harmful websites that appear legitimate. Traditional cybersecurity systems mainly rely on blacklist-based approaches, which are limited in detecting newly generated or zero-day malicious URLs. To overcome these challenges, NexusSafe introduces an advanced malicious URL detection system powered by Explainable Artificial Intelligence (XAI). The system performs real-time analysis of URLs using a multi-layered approach that includes lexical analysis, infrastructure profiling, threat intelligence, and behavioral analysis. Additionally, it integrates a Chrome Extension with a FastAPI backend to ensure continuous monitoring and quick response. One of the key features of NexusSafe is its ability to provide clear and understandable explanations for its predictions, helping users make informed decisions. By combining accurate detection with transparency, the system enhances both cybersecurity protection and user awareness.

1.1 MALICIOUS URL ATTACKS

Malicious URL attacks are one of the most widely used techniques by cybercriminals to compromise user security and steal sensitive information. These URLs are crafted to look legitimate but redirect usersto harmful websites that can

International Journal of Engineering & Extended Technologies Research (IJEETR)

|ISSN: 2322-0163| www.ijeetr.com | A Bimonthly, Peer Reviewed, Scholarly Indexed Journal |

| Volume 8, Issue 2, March - April 2026 |

DOI:10.15662/IJEETR.2026.0802064

perform actions such as phishing, malware installation, or data theft. Attackers often use techniques like URL obfuscation, fake domain names, shortened links, and redirection chains to deceive users and bypass traditional security systems. Since new malicious URLs are continuously generated, blacklist-based detection methods become ineffective in identifying , blacklist-based detection methods become ineffective in identifying these evolving threats. Therefore, detecting malicious URLs in real time using intelligent and adaptive techniques is essential to protect users and ensure a secure browsing experience.

EXPLAINABLE ARTIFICIAL INTELLIGENCE (XAI)

Explainable Artificial Intelligence (XAI) is a concept that focuses on making the decisions of machine learning models clear, transparent, and understandable to humans. In many traditional AI- based cybersecurity systems, models act as black boxes, providing predictions without explaining how those decisions were made. This lack of transparency reduces user trust and makes it difficult for security analysts to take appropriate actions. XAI addresses this issue by using techniques such as SHAP and LIME to highlight the important features that influence a model's prediction. In the NexusSafe system, XAI plays a crucial role by providing detailed explanations for why a URL is classified as safe, suspicious, or malicious, thereby improving trust, interpretability, and effective decision-making in cybersecurity.

URL FEATURE ANALYSIS

URL feature analysis is a key technique used to identify whether a website is safe or malicious by examining the structure and properties of its URL. This process involves analyzing lexical features such as URL length, presence of special characters, use of symbols like "@" or "//", number of subdomains, and suspicious keywords. In addition, host-based features such as domain age, IP address usage, DNS records, and HTTPS security are also considered. By extracting and evaluating these features, the system can detect hidden patterns and anomalies associated with malicious behavior. This allows NexusSafe to classify URLs effectively without fully loading the website, ensuring faster detection and improved user safety.

CYBERSECURITY DEFENSE

Cybersecurity defense refers to the set of technologies, strategies, and practices used to protect systems, networks, and data from cyber threats and unauthorized access. With the increasing complexity of cyberattacks, modern defense mechanisms require a multi-layered approach that includes prevention, detection, and response. NexusSafe enhances cybersecurity defense by continuously monitoring URLs in real time, analyzing potential threats using advanced techniques, and integrating threat intelligence from external sources. By combining these capabilities with Explainable AI, the system not only detects malicious activities accurately but also provides clear explanations, enabling users and organizations to understand threats and take appropriate actions to strengthen their overall security.

II. LITERATURE REVIEW

Recent advancements in cybersecurity have led to the development of various techniques for detecting malicious URLs, ranging from traditional blacklist- based approaches to advanced machine learning and deep learning models. Initially, most systems relied on blacklist databases, which compare URLs against known malicious entries. However, these methods are limited because they cannot detect newly generated or zero-day attacks and require constant updates.

To overcome these limitations, researchers introduced machine learning techniques such as Decision Trees, Support Vector Machines (SVM), Random Forest, and ensemble models. These approaches analyze features like URL structure, domain information, and website behavior to identify hidden patterns associated with malicious activities. Studies have shown that machine learning significantly improves detection accuracy and adaptability compared to traditional methods.

In recent years, deep learning techniques such as Convolutional Neural Networks (CNN) and Recurrent Neural Networks (RNN) have been applied to phishing and malicious URL detection. These models can automatically extract complex features from large datasets and detect sophisticated attack patterns, making them highly effective against evolving threats.

Despite these advancements, a major challenge in existing systems is the lack of interpretability. Most models act as black boxes, making it difficult for users and analysts to understand the reasoning behind predictions. To address this issue, Explainable Artificial Intelligence (XAI) techniques such as

International Journal of Engineering & Extended Technologies Research (IJEETR)

|ISSN: 2322-0163| www.ijeetr.com | A Bimonthly, Peer Reviewed, Scholarly Indexed Journal |

| Volume 8, Issue 2, March - April 2026 |

DOI:10.15662/IJEETR.2026.0802064

SHAP and LIME have been introduced, enabling models to provide clear explanations of their decisions.
The proposed NexusSafe system builds upon these research developments by integrating multi-layered detection techniques with XAI, ensuring both high accuracy and transparency in malicious URL detection.

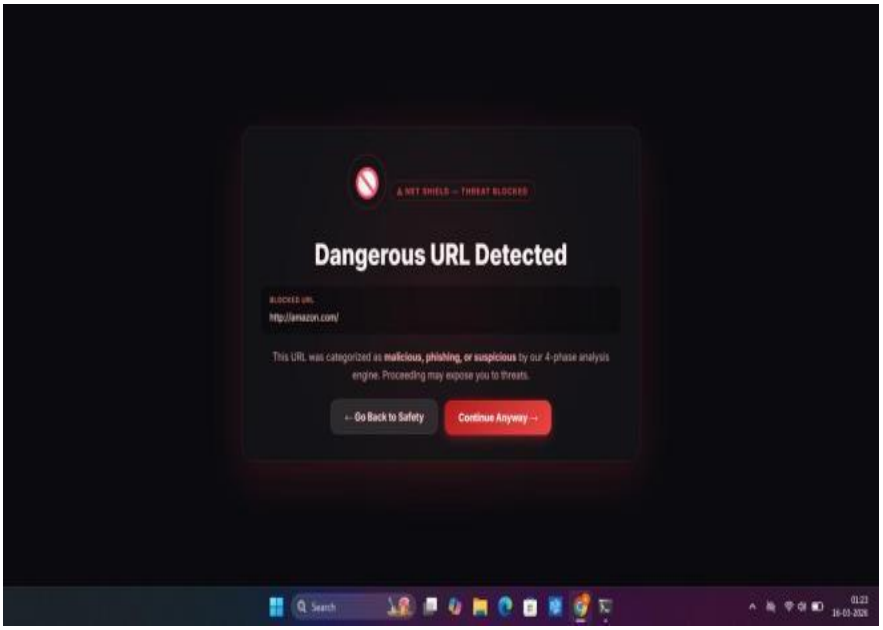
EXISTING SYSTEM

Existing systems for malicious URL detection primarily rely on traditional security mechanisms such as blacklist-based filtering, antivirus software, and browser security tools. These systems maintain databases of known malicious websites and block access when a match is found. While they provide basic protection against previously identified threats, they are ineffective in detecting newly created or zero-day malicious URLs, as such URLs are not yet included in the database.

In addition to blacklist approaches, some modern systems use machine learning models like Decision Trees, Random Forest, and Support Vector Machines to classify URLs based on extracted features. Although these methods improve detection accuracy compared to traditional techniques, they still face limitations such as high false positive rates and lack of real-time adaptability. Moreover, most existing systems do not provide clear explanations for their predictions, making them less transparent and difficult for users to trust.

Therefore, the limitations of existing systems highlight the need for a more advanced solution that can detect unknown threats in real time while also providing understandable insights into its decision- making process.

III. PROPOSED SYSTEM



The proposed system, NexusSafe, is an advanced malicious URL detection framework that leverages Explainable Artificial Intelligence (XAI) to provide accurate and transparent cybersecurity protection.

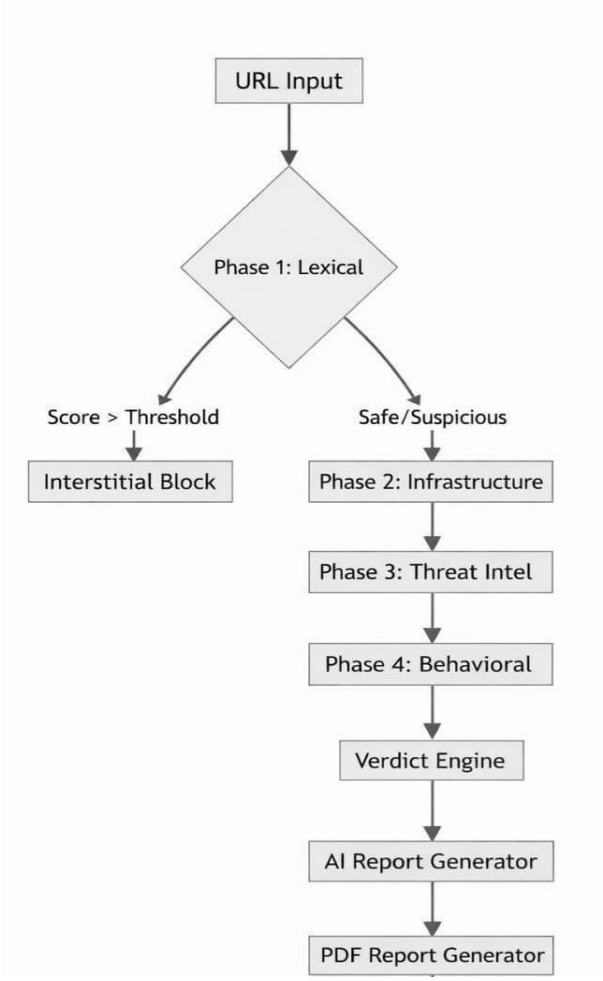
International Journal of Engineering & Extended Technologies Research (IJEETR)

|ISSN: 2322-0163| www.ijeetr.com | A Bimonthly, Peer Reviewed, Scholarly Indexed Journal |

| Volume 8, Issue 2, March - April 2026 |

DOI:10.15662/IJEETR.2026.0802064

The system integrates a Chrome Extension with a FastAPI backend server to monitor and analyze URLs in real time. When a user visits a website, the extension captures the URL and sends it backend for detailed analysis. The system employs a multi- layered detection approach that includes lexical feature analysis, infrastructure profiling, behavioral analysis, and threat intelligence integration using external platforms such as VirusTotal and urlscan.io. Based on these analyses, the system calculates a risk score and classifies the URL as safe, suspicious, or malicious.



A. Lexical Analysis

The Lexical Analysis phase is the first and one of the most important stages in the NexusSafe detection engine, as it performs a quick and efficient evaluation of a URL based solely on its structure and textual characteristics. This phase does not require accessing the actual website, making it faster and safer for initial screening. The system analyzes various lexical features of the URL to identify patterns commonly associated with malicious activity. One of the key checks performed in this phase is the detection of unusual URL length and the presence of excessive special characters such as "@", "-", "//", or multiple dots, which are often used to confuse users. It also identifies typo squatting, where attackers create fake domains that closely resemble legitimate websites (for example, slight spelling changes). Another important technique is detecting homoglyph attacks, where visually similar characters (like replacing "o" with "0") are used to trick users.

Additionally, the system examines encoded content within the URL, such as Base64 encoding, which may hide malicious payloads. It also checks for the use of suspicious keywords and patterns that indicate phishing attempts. A specialized feature of this phase is the detection of tunnel or proxy domains like ngrok and TryCloudflare, which are frequently used by attackers to host temporary malicious sites and bypass traditional security measures.
By analyzing these features, the Lexical Analysis phase provides a fast and effective way to flag potentially harmful

|ISSN: 2322-0163| www.ijeetr.com | A Bimonthly, Peer Reviewed, Scholarly Indexed Journal |

| Volume 8, Issue 2, March - April 2026 |

DOI:10.15662/IJEETR.2026.0802064

URLs at an early stage. This reduces the need for deeper analysis on clearly safe URLs and ensures that suspicious links are immediately passed to the next phases for further investigation, improving the overall efficiency and accuracy of the NexusSafe system.

B. Infrastructure Profiling

The Infrastructure Profiling phase focuses on analyzing the backend and hosting-related details of a URL to evaluate its trustworthiness. Unlike lexical analysis, which examines the URL structure, this phase investigates where and how the website is hosted. It provides deeper insight into the credibility of the domain by examining multiple infrastructure-level attributes.

One of the key components of this phase is analyzing the hosting provider and the Autonomous System Number (ASN) associated with the domain. Some hosting providers or ASNs may have a poor\ reputation due to frequent involvement in malicious activities. By checking ASN reputation, the system can identify whether the domain is hosted in a suspicious or high-risk network environment.

Another important aspect is the validation of the TLS/SSL certificate, which ensures secure communication between the user and the website. The system checks whether the certificate is valid, expired, self-signed, or improperly configured. Malicious websites often use invalid or weak certificates, which can be a strong indicator of risk.

Additionally, the system performs GeoIP mapping to determine the geographical location of the server hosting the website. If the domain is hosted in a location known for cybercrime activities or differs significantly from the expected region of the service, it may raise suspicion.

By combining all these factors, the Infrastructure Profiling phase helps identify hidden risks that are not visible from the URL alone. This phase strengthens the detection process by evaluating the reliability of the hosting environment, ensuring that suspicious domains are accurately flagged for further analysis.

C. Threat Intelligence

The Threat Intelligence phase enhances the detection capability of the NexusSafe system by integrating real-time data from external cybersecurity platforms. This phase plays a crucial role in identifying known threats as well as detecting newly emerging malicious URLs using global threat databases and scanning services.

In this phase, the system connects to platforms such as VirusTotal and urlscan.io, which aggregate threat data from multiple security vendors and researchers worldwide. These platforms provide detailed information about URLs, including whether they have been previously reported as malicious, associated with malware, phishing activity, or suspicious behavior. By leveraging this data, the system can quickly determine if a URL is already known to be harmful.

A key advantage of this phase is its ability to handle zero-day threats, which are newly created URLs that have not yet been widely reported. The system can trigger real-time scans on these platforms to analyze unknown URLs instantly. This ensures that even previously unseen threats can be detected effectively. The phase also includes mechanisms to handle API rate limits and ensures efficient use of external services without delays. By combining data from multiple sources, the system increases its detection accuracy and reduces false positives.
Overall, the Threat Intelligence phase acts as a global knowledge layer, enabling the NexusSafe system to stay updated with the latest cyber threats and provide reliable, real-time protection against both known and emerging malicious URLs.

D. Behavioral Analysis

The Behavioral Analysis phase focuses on understanding how a website behaves when accessed, rather than just analyzing its structure or infrastructure. This phase is essential for detecting advanced threats that may appear safe initially but perform malicious actions during interaction. By observing the runtime behavior of the URL, the system can uncover hidden risks that are not detectable through static analysis.

One of the primary techniques used in this phase is redirection chain analysis. Many malicious URLs do not directly host harmful content but instead redirect users through multiple intermediate links before reaching the final destination. The system tracks these redirections to identify suspicious or hidden paths that may lead to phishing pages or malware downloads.

International Journal of Engineering & Extended Technologies Research (IJEETR)

|ISSN: 2322-0163| www.ijeetr.com | A Bimonthly, Peer Reviewed, Scholarly Indexed Journal |

| Volume 8, Issue 2, March - April 2026 |

DOI:10.15662/IJEETR.2026.0802064

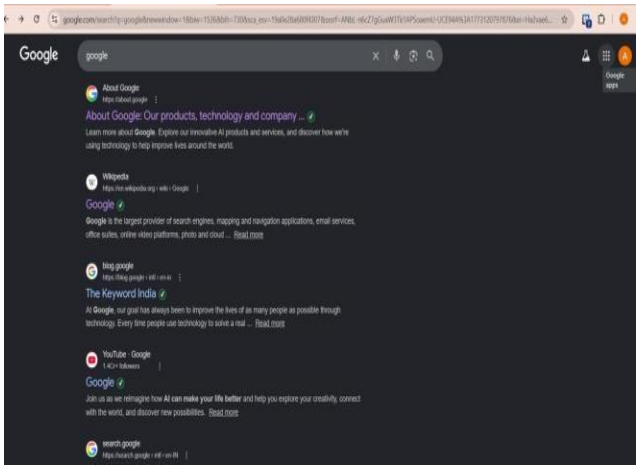
In addition, the system evaluates important security headers such as HTTP Strict Transport Security (HSTS), Content Security Policy (CSP), and X- Frame-Options. These headers play a crucial role in protecting websites from common attacks like clickjacking and data injection. The absence or misconfiguration of these headers may indicate poor security practices or malicious intent.

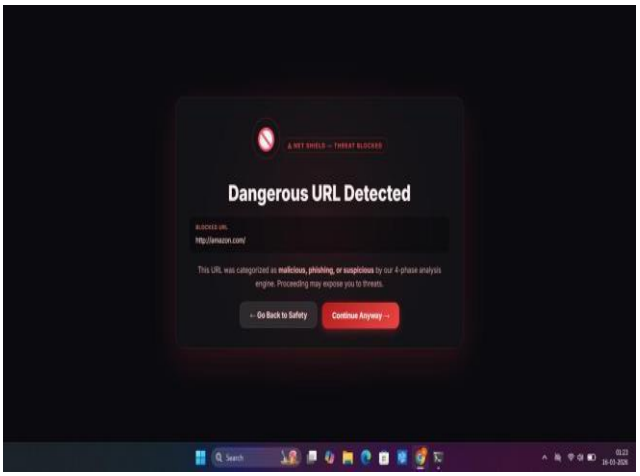
The system may also analyze other behavioral indicators such as unusual response patterns, unexpected scripts, or abnormal interactions that could signal phishing or exploitation attempts.

By combining these observations, the Behavioral Analysis phase provides a dynamic view of the website's activity, ensuring that even sophisticated and hidden threats are detected. This phase significantly strengthens the overall security of the NexusSafe system by identifying risks that are not visible through earlier stages of analysis.

IV. RESULT ANALYSIS

Experimental Results



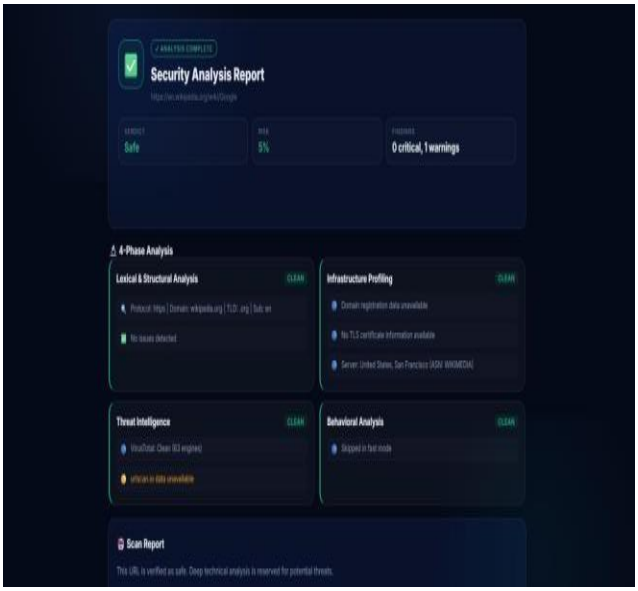


International Journal of Engineering & Extended Technologies Research (IJEETR)

|ISSN: 2322-0163| www.ijeetr.com | A Bimonthly, Peer Reviewed, Scholarly Indexed Journal |

| Volume 8, Issue 2, March - April 2026 |

DOI:10.15662/IJEETR.2026.0802064



The performance of the NexusSafe system is evaluated based on its ability to accurately detect malicious URLs in real time while providing clear and understandable explanations. The system was tested using a dataset containing both benign and malicious URLs collected from various sources. The evaluation of metrics used include accuracy, precision, recall, and F1-score to measure the effectiveness of the detection model.

The results show that the multi-layered detection approach significantly improves the overall accuracy of the system, achieving high detection rates for both known and unknown threats. The integration of lexical analysis, infrastructure profiling, threat intelligence, and behavioral analysis ensures that even complex and obfuscated malicious URLs are identified effectively. The use of external threat intelligence platforms further enhances detection by providing real-time updates on emerging threats.

In addition to accuracy, the system demonstrates a low false positive rate, meaning that legitimate websites are rarely misclassified as malicious. This improves user trust and reduces unnecessary interruptions during browsing. The Explainable AI component also plays a crucial role by generating clear explanations for each prediction, helping users understand the reasons behind the classification and making the system more transparent and user- friendly.

Overall, the results indicate that NexusSafe provides a reliable and efficient solution for malicious URL detection. The combination of high accuracy, real-time performance, low false positives, and explainability makes it a strong and effective cybersecurity system suitable for practical applications.

Signal Analysis Overview of URL Threat Detection in NexusSafe

The signal analysis overview illustrates the internal feature evaluation performed by the proposed system. Various signals such as lexical patterns, URL structure, and behavioral indicators are analyzed to determine whether a given URL is safe or suspicious. The visualization highlights how multiple detection layers contribute to the final classification, thereby improving the accuracy and reliability of the system.

|ISSN: 2322-0163| www.ijeetr.com | A Bimonthly, Peer Reviewed, Scholarly Indexed Journal |

| Volume 8, Issue 2, March - April 2026 |

DOI:10.15662/IJEETR.2026.0802064

V. CONCLUSION

NexusSafe presents an effective and intelligent solution for detecting malicious URLs using Explainable Artificial Intelligence (XAI). By combining a multi-layered detection approach that includes lexical analysis, infrastructure profiling, threat intelligence, and behavioral analysis, the system is capable of identifying both known and unknown cyber threats with high accuracy. Unlike traditional systems that rely only on blacklists, NexusSafe provides real-time analysis and adapts to emerging threats, making it more reliable and efficient.

A key strength of the system is its ability to generate clear and human-readable explanations for its predictions, which enhances transparency, user trust, and decision-making. The integration of a Chrome Extension with a FastAPI backend ensures seamless operation and continuous monitoring of user activity without affecting performance.
Overall, NexusSafe not only improves cybersecurity protection but also increases user awareness by explaining the risks associated with malicious websites. This makes it a valuable tool for both individual users and organizations, contributing to a safer and more secure online environment

VI. FUTURE WORK

The NexusSafe system can be further enhanced by incorporating advanced machine learning and deep learning models to improve detection accuracy and adapt to more sophisticated cyber threats. Future work may include the integration of neural network- based approaches such as deep learning models for better pattern recognition and handling of complex malicious behaviors. Additionally, the system can be expanded to support multiple browsers like Firefox and Microsoft Edge, as well as mobile platforms, to provide wider accessibility and protection across devices.

Another important improvement is the integration of more comprehensive global threat intelligence feeds to keep the system continuously updated with the latest cyber threats. Enhancing the Explainable AI component with more interactive and visual explanations can further improve user understanding and trust. The system can also include automated learning mechanisms that update the model based on new threats and user feedback, ensuring continuous improvement.

Furthermore, future developments may focus on real- time large-scale deployment, cloud integration, and improved performance optimization for handling high volumes of web traffic. These enhancements will make NexusSafe more scalable, efficient, and capable of providing stronger cybersecurity protection in dynamic and evolving digital environment

REFERENCES

1) H. Kibriya, R. Amin, S. S. Alshamrani, et al., “Lightweight Malicious URL Detection Using Deep Learning and Large Language Models,” Scientific Reports, vol. 15, 2025.
2) Türk and M. Kılıçaslan, “Malicious URL Detection with Advanced Machine Learning and Optimization-Supported Deep Learning Models,” Applied Sciences, vol. 15, no. 18, 2025.
3) A. Nair et al., “Securing Healthcare Systems: A Random Forest Approach to Malicious URL Detection,” Journal of Computer Virology and Hacking
4) Techniques, 2025.
5) T. Kehkashan et al., “Explainable Phishing Website Detection for Secure and Sustainable Cyber Infrastructure,” Scientific Reports, 2025.
6) D. Karapiperis et al., “Explainable AI for URL Threat Detection Using LIME and SHAP,” Studies in Health Technology and Informatics, 2025.
7) L. Chen and L. Meng, “Metadata Driven Malicious URL Detection Using RoBERTa and Multi-Source Threat Intelligence,” Scientific Reports, 2026.
8) C.Nagarajan and M.Madheswaran - ‘Stability Analysis of Series Parallel Resonant Converter with Fuzzy Logic Controller Using State Space Techniques’- Taylor &Francis, Electric Power Components and Systems, Vol.39 (8), pp.780-793, May 2011. DOI: 10.1080/15325008.2010.541746
9) C.Nagarajan and M.Madheswaran - ‘Experimental verification and stability state space analysis of CLL-T Series Parallel Resonant Converter’ - Journal of Electrical Engineering, Vol.63 (6), pp.365-372, Dec.2012. DOI: 10.2478/v10187-012-0054-2
10) C.Nagarajan and M.Madheswaran - ‘Performance Analysis of LCL-T Resonant Converter with Fuzzy/PID Using State Space Analysis’- Springer, Electrical Engineering, Vol.93 (3), pp.167-178, September 2011. DOI

|ISSN: 2322-0163| www.ijeetr.com | A Bimonthly, Peer Reviewed, Scholarly Indexed Journal |

| Volume 8, Issue 2, March - April 2026 |

DOI:10.15662/IJEETR.2026.0802064

10.1007/s00202-011-0203-9

11) S.Tamilselvi, R.Prakash, C.Nagarajan,"Solar System Integrated Smart Grid Utilizing Hybrid Coot-Genetic Algorithm Optimized ANN Controller" Iranian Journal Of Science And Technology-Transactions Of Electrical Engineering, DOI10.1007/s40998-025-00917-z,2025
12) S.Tamilselvi, R.Prakash, C.Nagarajan," Adaptive sliding mode control of multilevel grid-connected inverters using reinforcement learning for enhanced LVRT performance" Electric Power Systems Research 253 (2026) 112428, doi.org/10.1016/j.epsr.2025.112428
13) S.Thirunavukkarasu, C. Nagarajan, 2024, "Performance Investigation on OCF and SCF study in BLDC machine using FTANN Controller," Journal of Electrical Engineering And Technology, Volume 20, pages 2675–2688, (2025), doi.org/10.1007/s42835-024-02126-w
14) C.Nagarajan,M.Madheswaran and D.Ramasubramanian- 'Development of DSP based Robust Control Method for General Resonant Converter Topologies using Transfer Function Model'- Acta Electrotechnica et Informatica Journal , Vol.13 (2), pp.18-31,April-June.2013, DOI: 10.2478/aeei-2013-0025.
15) C.Nagarajan and M.Madheswaran - 'DSP Based Fuzzy Controller for Series Parallel Resonant converter'- Springer, Frontiers of Electrical and Electronic Engineering, Vol. 7(4), pp. 438-446, Dec.12. DOI 10.1007/s11460-012-0212-0.
16) C.Nagarajan and M.Madheswaran - 'Experimental Study and steady state stability analysis of CLL-T Series Parallel Resonant Converter with Fuzzy controller using State Space Analysis'- Iranian Journal of Electrical & Electronic Engineering, Vol.8 (3), pp.259-267, September 2012.
17) C.Nagarajan and M.Madheswaran, "Analysis and Simulation of LCL Series Resonant Full Bridge Converter Using PWM Technique with Load Independent Operation" has been presented in ICTES'08, a IEEE / IET International Conference organized by M.G.R.University, Chennai.Vol.no.1, pp.190-195, Dec.2007
18) Suganthi Mullainathan, Ramesh Natarajan, "An SPSS and CNN modelling based quality assessment using ceramic materials and membrane filtration techniques", Revista Materia (Rio J.) Vol. 30, 2025, DOI: https://doi.org/10.1590/1517-7076-RMAT-2024-0721
19) M Suganthi, N Ramesh, "Treatment of water using natural zeolite as membrane filter", Journal of Environmental Protection and Ecology, Volume 23, Issue 2, pp: 520-530,2022
20) "From Past to Present: A Survey of Malicious URL Detection Techniques, Datasets and Code Repositories," Computer Science Review, 2025.
21) "A Novel Approach for Malicious URL Detection Using RoBERTa and Sparse Autoencoder," Journal of Information Security and Applications, 2025.
22) Y. Tian et al., "URL2Graph++: Unified Semantic- Structural Learning for Malicious URL Detection," arXiv, 2025.
23) Y. Tian et al., "WebGuard++: Interpretable Malicious URL Detection via BERT and HTML Subgraphs," arXiv, 2025.