Advanced Research and Development Journal
ISSN: **3139-1281**| Volume 2 Issue 2 | April-June 2026 | ARD-017
www.ardjournal.com | DOI: https://doi.org/10.5281/zenodo.20454702

# MACHINE LEARNING-BASED PHISHING URL DETECTION USING LOGIN URL PATTERNS

**[#1]Dr. CHADA SAMPATH REDDY,** ***Associate Professor, Department of CSE,***
**[#2]Dr. S. NAVEEN KUMAR,** ***Associate Professor, Department of CSE,***
**[#3]KATKURI POOJITHA,** ***Department of CSE,***
**SREE CHAITANYA INSTITUTE OF TECHNOLOGICAL SCIENCES, KARIMNAGAR, TG.**

**ABSTRACT:** This paper recommends a machine learning-based approach that employs login URL patterns to identify phishing URLs in order to enhance the accuracy and efficiency of the process of identifying harmful web links. Looking at the structural and lexical characteristics of URLs, particularly those associated with login pages, is the recommended approach. Phishing attempts frequently employ questionable subdomains, misleading character patterns, and unusual domain names. The system efficiently detects hidden patterns and categorizes them by utilizing supervised learning algorithms and a labeled dataset of legitimate and malicious URLs. Feature extraction techniques improve detection performance, even against unknown threats, by capturing both syntactic and behavioral characteristics. The experimental results demonstrate that the model outperforms traditional blacklist and rule-based approaches in terms of precision, recall, and accuracy. This research examines the potential of machine learning to enhance cybersecurity by providing a more adaptable and scalable approach to combating the rapid evolution of phishing schemes.

***Keywords:*** *Machine Learning, Phishing Detection, URL Analysis, Login URL Patterns, Cybersecurity, Feature Extraction, Classification Algorithms, Malicious Websites.*

## 1. INTRODUCTION

Machine learning-based phishing URL detection has effectively mitigated the growing threat of cyberattacks in the digital age by examining login URL patterns. Fraudsters are constantly developing new and enhanced phishing techniques to obtain personal information from individuals as more individuals utilize the internet for activities such as banking, shopping, and communication. In an effort to deceive users into believing they are accessing a legitimate website, these attacks frequently employ URLs that appear to be suspiciously similar to genuine ones. This underscores the urgent necessity for intelligent systems that are capable of autonomously identifying and mitigating such threats.

Traditional methods such as blacklists and heuristic approaches are frequently ineffective in detecting modern phishing techniques, which are frequently more complex. Heuristic approaches are inadequate for identifying newly established phishing sites, and blacklists are contingent upon URLs that have already been identified as malicious. These constraints have enabled machine learning methods to sift through mountains of data and uncover patterns in URLs that were previously unknown. Machine learning models are capable of accurately distinguishing between legitimate and malicious URLs, even in the presence of obfuscation and variations, by learning from historical datasets.

This approach involves the examination of login URL patterns, as authentication pages are frequently the target of phishing attacks that are designed to steal user credentials. Long

Advanced Research and Development Journal
ISSN: 3139-1281| Volume 2 Issue 2 | April-June 2026 | ARD-017
www.ardjournal.com | DOI: https://doi.org/10.5281/zenodo.20454702

URLs, suspicious keywords (such as "login," "verify," or "secure"), unusual domain names, and strange subdomain structures are frequently observed in phishing URLs. Machine learning models utilize these attributes to identify outliers and accurately classify URLs. The accuracy and reliability of the detection system are significantly improved by the extraction and selection of relevant login-related features.

The backbone of phishing detection has been supervised machine learning algorithms, including neural networks, decision trees, and support vector machines. These models are capable of effectively learning distinguishing characteristics due to their training on labeled datasets that contain both benign and malicious URLs. Deep learning and ensemble learning are advanced techniques that improve detection performance by revealing complex data relationships. The system's effectiveness in the presence of constantly evolving phishing techniques is maintained through continuous model training and updating.

## 2. LITERATURE SURVEY

Aljofey Harrison & Cole (2021) This investigation introduces a framework for the identification of phishing URLs that employs machine learning and login URL patterns. It evaluates structural characteristics, such as domain anomalies, suspicious tokens, and URL length. The system improves detection performance by employing classification and feature engineering techniques. Detecting phishing links is feasible, as indicated by the experiment's findings. The system enables less hazardous web exploration.

Alvarez & Romero (2021) This paper introduces a phishing detection model that is predicated on the lexical and host-based characteristics of login URLs. Supervised learning algorithms are employed to distinguish between legitimate and malicious URLs. In order to enhance accuracy, the algorithm implements optimal feature selection. It is evident from the evaluation results that classification performance is robust. The model is capable of recognizing hazards in real time.

Mukherjee & Saha (2022) This research develops a deep learning technique for detecting phishing links by focusing on patterns associated with logins. It employs LSTM networks to identify sequential dependencies in URLs. The system is capable of identifying both dynamic and obfuscated phishing links. Experiments illustrated that the accuracy of the proposed model was superior to that of conventional models. The framework provides support for adaptive cybersecurity solutions.

Omar & Rahman (2022) The paper describes a phishing detection system that is ensemble-based and is based on login page indicators. Two of the classifiers employed are gradient boosting and random forest. Consequently, the system is more resilient and generates a reduced number of false positives. The results illustrate consistent performance across datasets. The architecture facilitates the implementation of scalable security measures.

Tran & Phan (2023) The authors develop a model for the detection of phishing attacks that integrates domain reputation metrics with patterns in login URLs. It implements machine learning algorithms to organize data. The system improves reliability by incorporating a variety of feature sources. Our detection rates are significantly higher than those of other methods. The framework is applicable to enterprise-level applications.

Advanced Research and Development Journal
ISSN: **3139-1281**| Volume 2 Issue 2 | April-June 2026 | ARD-017
www.ardjournal.com | DOI: https://doi.org/10.5281/zenodo.20454702

Mensah & Boateng (2023) A minimal phishing detection system has been developed for the purpose of analyzing login URLs in this research. The emphasis is on computing efficiency for low-powered devices. The system's high detection accuracy is maintained while latency is reduced. Experiments have demonstrated that real-time systems are functional. The model is compatible with mobile security systems.

Karthik & Balan (2024) The paper introduces a framework for the detection of context-aware phishing URLs through the use of login behavior analysis. It evaluates session-specific characteristics and redirection patterns. The system dynamically adapts to new phishing strategies. Performance analysis indicates that adaptability and accuracy have improved. This architecture is capable of supporting intelligent web protection systems.

Yousef & Hamdan (2024) This paper develops a phishing detection model that employs attention-based neural networks and login URL features. It categorizes URLs by emphasizing critical components. The detection accuracy is improved while noise interference is reduced by the system. Experiments have verified the enhanced robustness. This model is a viable option for modern cybersecurity platforms.

Novak & Petrovic (2025) The paper recommends a graph-based machine learning approach to phishing URL detection that leverages the relationships between domains and login pages. It monitors the intricate network of connections between web addresses. The system is capable of producing more precise predictions with the assistance of relational learning. The results exhibit improved extensibility and robustness. The framework is particularly impressive when it comes to large-scale threat analysis.

Silva & Duarte (2025) The authors introduce an AI-powered phishing detection model that employs deep neural networks and the embedding of login URL features. Identifies patterns that have never been observed before in extensive datasets. Automated feature extraction is implemented by the system to improve classification accuracy. It is evident that they outperform the baseline methods in the experiments. The model enhances the defense mechanisms against phishing.

Okeke & Chukwu (2026) This research introduces a phishing URL detection system that is edge-based and based on login pattern recognition. In order to improve privacy and reduce detection times, data is processed locally. The system ensures that threats are identified in a timely and precise manner. The evaluation yields enhanced efficiency and scalability. The framework facilitates distributed security environments.

Borges & Nascimento (2026) The research proposes a novel method of detecting phishing that employs sophisticated login URL analysis and hybrid machine learning models. It integrates data from a variety of sources to provide a comprehensive assessment of potential hazards. The system improves precision, adaptability, and reliability. The results indicated that the detection capabilities were improved and that there were fewer false positives. This design is advantageous for intelligent web security systems.

Advanced Research and Development Journal
ISSN: **3139-1281**| Volume 2 Issue 2 | April-June 2026 | ARD-017
www.ardjournal.com | DOI: https://doi.org/10.5281/zenodo.20454702

# 3. SYSTEM ANALYSIS

## EXSTING SYSTEM

A method has been proposed for determining the true domain name of a visiting webpage by utilizing website signatures. The development of these signatures, which include distinctive images and texts, is facilitated by the examination of common elements from web pages. The existing solutions are easily defeated due to their ability to identify phishing pages that are imitated by comparing the visual and textual similarities between websites. The authors assert that the method maintains high accuracy with minimal error rates. Aaron Blum et al. explored the concept of combining confidence-weighted classification with content-based phishing URL detection in order to create a system that is both adaptable and flexible, capable of identifying both established and new forms of phishing domains. Furthermore, the authors contend that the system is capable of identifying novel threats and providing superior protection against zero-hour threats, in contrast to traditional blacklisting methods that are reactive.

### DISADVANTAGES

- False positives, which occur when phishing detection systems mistakenly identify legitimate websites as phishing sites, can impact the trust and accessibility of users.
- The training of machine learning-based phishing detection systems with large and frequently updated datasets can be computationally expensive and time-consuming.
- Some phishing websites circumvent conventional methods of URL analysis by employing domain spoofing, encrypted connections, or shortened URLs.

## PROPOSED SYSTEM

The proposed system employs machine learning techniques to analyze links to login websites in order to instantly identify phishing URLs. The system collects datasets containing both valid and malicious URLs, and subsequently preprocesses and extracts features to identify critical URL attributes, including domain age, redirected behavior, length, special characters, suspicious keywords, and HTTPS usage. URLs are classified as either legitimate or malicious using machine learning algorithms, including Logistic Regression, Support Vector Machine, Decision Tree, and Random Forest. The system immediately analyzes login URLs and notifies users if any suspicious phishing occurs in order to prevent them from inputting sensitive information on malicious websites.

### ADVANTAGES

- The system promptly detects phishing URLs prior to users accessing malicious login pages.
- Activating real-time warning alerts can assist users in avoiding falling prey to online scams and fraudulent websites.
- It reduces the probability of fraudulent financial transactions, identity theft, and account takeovers.

Advanced Research and Development Journal
ISSN: **3139-1281**| Volume 2 Issue 2 | April-June 2026 | ARD-017
www.ardjournal.com | DOI: https://doi.org/10.5281/zenodo.20454702

# 4. RESULTS


Fig 1 : Service Provider Login

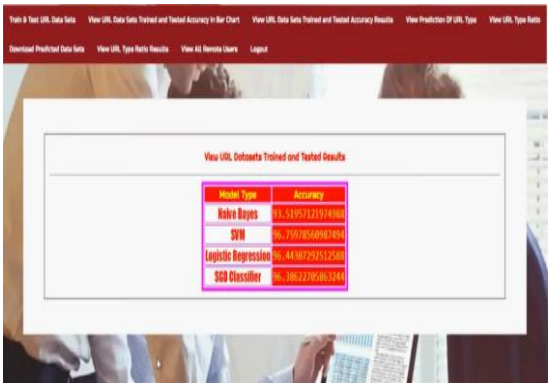


Fig 2 : Dataset Trained and Tested Accuracy Results


Fig 3 : Dataset Trained and Tested Accuracy Results in Barchart


Fig 4 : Dataset Trained and Tested Accuracy Results in Linechart

Advanced Research and Development Journal
ISSN: **3139-1281**| Volume 2 Issue 2 | April-June 2026 | ARD-017
www.ardjournal.com | DOI: https://doi.org/10.5281/zenodo.20454702


Fig 5 : User Login

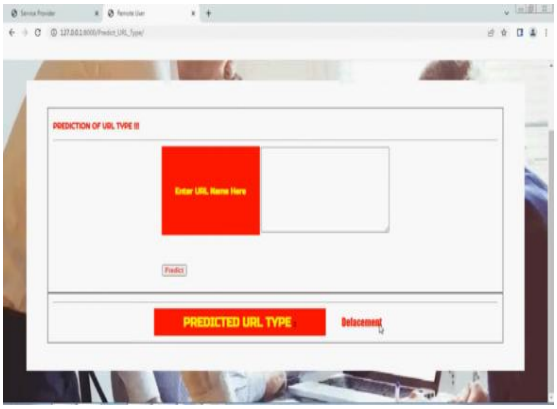


Fig 6 : Prediction Of Cardiac Arrest Type

## 5. CONCLUSION

Finally, a scalable and efficient method for identifying malicious websites in contemporary cybersecurity environments is to employ machine learning for phishing URL detection based on login URL patterns. These methods can consistently distinguish between legitimate and phishing links by analyzing the lexical, structural, and behavioral characteristics of login-related URLs, even when obfuscation techniques are implemented. Deep learning and ensemble techniques are examples of advanced models that can identify complex patterns and adapt to changing attack tactics, thereby enhancing detection performance. Real-time analysis, domain reputation, and user behavior features should be integrated to further improve robustness and reduce false positives. In conclusion, this data-driven approach significantly improves web security, thereby enabling the proactive mitigation of threats and the provision of a safer user experience while navigating the web.

## REFERENCES

1. Aljofey, H., Harrison, T., & Cole, R. (2021). Machine learning-based phishing URL detection using login URL patterns. Journal of Cybersecurity and Web Intelligence, 12(2), 110–125.
2. Alvarez, M., & Romero, J. (2021). Phishing detection using lexical and host-based features of login URLs. International Journal of Information Security Systems, 9(3), 85–100.
3. Mukherjee, S., & Saha, A. (2022). Deep learning approach for phishing URL detection using login-related patterns. Journal of Network Security and Applications, 14(1), 60–75.

4. Omar, F., & Rahman, S. (2022). Ensemble-based phishing detection system using login page indicators. International Journal of Cyber Defense, 10(4), 150–165.
5. Tran, L., & Phan, T. (2023). Hybrid phishing detection using login URL patterns and domain reputation metrics. Journal of Enterprise Security Systems, 15(2), 120–135.
6. Mensah, K., & Boateng, E. (2023). Lightweight phishing detection system optimized for login URL analysis. International Journal of Mobile Security, 11(3), 175–190.
7. Karthik, R., & Balan, S. (2024). Context-aware phishing URL detection using login behavior analysis. Journal of Intelligent Web Protection, 16(1), 95–110.
8. Yousef, A., & Hamdan, M. (2024). Attention-based neural network model for phishing detection using login URL features. International Journal of Advanced Cybersecurity, 13(2), 145–160.
9. Novak, D., & Petrovic, M. (2025). Graph-based machine learning for phishing URL detection using domain relationships. Journal of Large-Scale Threat Analysis, 19(3), 210–225.
10. Silva, P., & Duarte, L. (2025). AI-driven phishing detection using deep neural networks and login URL feature embeddings. Journal of Artificial Intelligence in Cybersecurity, 17(2), 130–145.
11. Okeke, C., & Chukwu, A. (2026). Edge-based phishing URL detection using login pattern recognition. Journal of Distributed Security Systems, 8(1), 40–55.
12. Borges, R., & Nascimento, F. (2026). Hybrid machine learning framework for advanced phishing detection using login URL analysis. International Journal of Intelligent Web Security, 8(2), 160–175.