International Journal of BioTech, Pharmaceutical, Engineering, Management and Sciences
ISSN: 3139-3748 |VOLUME-1, ISSUE-1| Inaugural Edition | January-June 2026
Paper ID: P11-027| Research Article
doi: https://doi.org/10.5281/zenodo.20677468

REAL-CASE EXPLORATION OF PHISHING URL DETECTION IN LOGIN ENVIRONMENTS

[#1]**Sangoju Vyshnavi,** *Department of MCA,*
[#2]**Mr. T. Raghupathi,** *Assistant Professor, Department of MCA,*
Vaageswari College of Engineering(Autonomous), Karimnagar,TG

ABSTRACT: The proliferation of phishing schemes, which involve the use of fake websites to deceive individuals into disclosing personal information, is a significant issue in the field of cybersecurity. Fake login pages that resemble legitimate websites are a prevalent method of obtaining individuals' credentials. The objective of this investigation is to identify authentic phishing URLs by examining logon URLs. As scam assaults become increasingly sophisticated, individuals are experiencing difficulty distinguishing between legitimate and fraudulent websites. Phishers frequently exploit the registration pages of popular services, such as email, social media, and banking, to deceive individuals into disclosing their login information. In order to prevent the theft of one's identity, the hacking of their data, the loss of money, and other forms of cybercrime, individuals must be able to identify these fraudulent URLs. This investigation illustrates a method for identifying phishing URLs in real-world scenarios, with a particular emphasis on logon URLs. It simultaneously employs methods from machine learning, web text analysis, and URL attribute evaluation. The validity of the SSL certificate, domain similarity, page content analysis, and URL architecture are among the factors that can be employed to determine whether a URL is part of a fraud attempt.

Keywords: URL, SSL, phishing attacks, SVM, dataset.

I. INTRODUCTION

The likelihood of spoofing attempts during logins has increased significantly as a result of the rapid advancements in digital communication and online security technologies. Phishing URLs are malicious web connections that imitate legitimate websites in order to obtain private information, such as passwords, login credentials, and banking information. Thieves frequently attempt to access the login pages of banking applications, social media platforms, education systems, and business networks by creating fraudulent websites that closely resemble the legitimate login pages. Phishing attacks have emerged as a significant cybersecurity threat to businesses, individuals, and online services due to the rapid adoption of the internet by a large number of individuals worldwide.

Researchers are examining actual cases of phishing in order to gain a more comprehensive understanding of how assailants deceive individuals during login situations and to examine real-world examples of phishing URL detection. Domain spoofing, URL obfuscation, HTTPS exploitation, shortening malicious URLs, and creating false login pages are all common methods used by attackers to circumvent standard security protocols. Social engineering techniques, such as phony emails, account verification requests, urgent alerts, and messages requesting new passwords, are frequently employed to deceive individuals into visiting hazardous websites. The sophistication of phishing methods is increasing, which impedes the

International Journal of BioTech, Pharmaceutical, Engineering, Management and Sciences
ISSN: 3139-3748 |VOLUME-1, ISSUE-1| Inaugural Edition | January-June 2026
Paper ID: P11-027| Research Article
doi: https://doi.org/10.5281/zenodo.20677468

ability of conventional detection systems (which rely on blacklists and signatures) to immediately identify newly created URLs.

Modern phishing URL detection systems employ intricate machine learning and deep learning algorithms to enhance the detection of suspicious logon activities. In order to identify spoofing attempts, more sophisticated models closely examine data such as domain registration details, user activity patterns, network traffic attributes, URL architecture, and webpage content. Random Forest, Support Vector Machines (SVM), Convolutional Neural Networks (CNN), and Long Short-Term Memory (LSTM) networks are increasingly employed to enhance the precision of fraud identification and reduce the number of false positives. The integration of real-time monitoring systems, automated threat intelligence, and adaptable cybersecurity frameworks into contemporary digital environments safeguards them from unauthorized access, identity theft, and new phishing schemes.

II. LITERATURE SURVEY

Watson & Iqbal (2021): This paper discusses a method for identifying phishing URLs in logon scenarios through the use of machine learning. It employs techniques such as lexical and behavioral analysis. We employ the Decision Tree and Random Forest algorithms to detect indications of phishing on login pages. This is achieved by examining factors such as the age of the domain, the usage of unusual keywords, and the length of the URL. In order to enhance the system's ability to locate objects, it analyzes real-world data from social media platforms and websites where users log in to their accounts. The results indicate a reduction in the number of false finds and an enhanced capacity to identify fraudulent login URLs.

Fernandez & Rao (2022): The research demonstrates a deep learning approach to identifying phishing URLs in logon scenarios by utilizing artificial neural networks and evaluating URLs in real time. The program is capable of identifying logon pages that are vulnerable by analyzing page layout, redirection patterns, HTTPS implementation, and hyperlink activity. Feature extraction methods enable the management of a significant amount of phishing data, in addition to enhancing the accuracy of classification. The performance evaluation indicates that the accuracy of phishing detection is superior to that of conventional blacklist-based methods.

Kim & Sullivan (2023): The authors developed a hybrid machine learning framework that combines Convolutional Neural Networks and Support Vector Machines to identify fraudulent URLs in real-world scenarios. The program analyzes the image of the domain, the manner in which users interact with it, the similarity of logon pages, patterns of URL obfuscation, and phishing attempts to ensure precise classification. A comparison demonstrates that online settings that undergo rapid change exhibit superior scaling and detection capabilities.

Patel & Nakamura (2024): Recurrent neural networks and sequence pattern analysis are recommended as effective methods for identifying fraudulent URLs during the login process. By monitoring the manner in which individuals log in and navigate, the model identifies attempts at malicious redirection, phishing, and fake authentication sites. Real-time detection modules facilitate the prevention of unauthorized entry attempts and the acceleration of responses in business systems. The results of the investigations demonstrate that the

ISSN: 3139-3748 |VOLUME-1, ISSUE-1| Inaugural Edition | January-June 2026
Paper ID: P11-027| Research Article
doi: https://doi.org/10.5281/zenodo.20677468

enhanced classification accuracy and the increased capacity to adapt to new phishing techniques are genuine.

Garcia & Menon (2025): The research proposes the utilization of a cloud-connected deep learning system to identify phishing URLs that are challenging to identify in logon settings. Long-Short-Term Memory (LSTM) networks monitor user behavior, domain anomalies, and website usage patterns to identify suspicious logon actions. Distributed processing methods facilitate the management of substantial cybersecurity datasets and facilitate the scaling process. The findings indicate that individuals are more resilient to zero-day phishing attacks and that phishing is being detected at a higher frequency. Safe digital contact tools and sophisticated cyber defenses are made possible by the architecture.

Yamamoto & Clarke (2026): Using transformer-based neural network models and shared learning to establish secure logins, the authors develop a sophisticated system for identifying fraudulent URLs. This design allows for the secure examination of harmful URL patterns and login practices across multiple platforms, while simultaneously safeguarding user privacy. Adaptive learning methods are consistently employed to enhance phishing detection models and prevent the emergence of new cyberthreats and false login attempts. Corporate cybersecurity solutions are more scalable, have reduced latency, and are more effective at identifying threats, as demonstrated by experiments.

III. METHODOLOGY

Data Collection

Valid datasets are collected from login environments, including social media sites, e-commerce sites, educational institutions, financial portals, and business authentication pages, in the initial phase. These datasets include both good and problematic URLs. Information is collected from a variety of sources, including real-time databases that monitor cybersecurity events, browser security logs, and public phishing repositories. Domain names, SSL certificates, login pages, content, and URL structures are included in the compilation to facilitate the identification and analysis of phishing.

Data Preprocessing

The URLs and webpage data that were obtained are cleaned and organized into a structured format that can be utilized for machine learning research during the preprocessing phase. In order to enhance the dataset's quality, duplicate URLs, inactive links, damaged records, and pages with insufficient information are eliminated. Some of the methods used to extract significant components of URLs include normalizing the text, tokenizing it, decoding it, and scale the characteristics. Standardization is implemented to facilitate the processing and organization of website scripts, suspicious symbols, domain patterns, and reroute behaviors.

Feature Extraction

In logon scenarios, feature extraction algorithms can identify the most critical indicators of fraudulent URLs. URL length, domain age, IP-based domains, the number of special characters, HTTPS usage, the number of redirections, the similarity of page titles, and the functionality of the login form are among the significant factors that were extracted from the dataset. Phishing attempts frequently involve the use of deceptive content layouts and

International Journal of BioTech, Pharmaceutical, Engineering, Management and Sciences
ISSN: 3139-3748 |VOLUME-1, ISSUE-1| Inaugural Edition | January-June 2026
Paper ID: P11-027| Research Article
doi: https://doi.org/10.5281/zenodo.20677468

fabricated login pages. In order to identify instances of fraud, natural language processing algorithms analyze the text on a website and the keywords that are associated with it.

Machine Learning and Deep Learning Integration

The proposed method effectively distinguishes between secure and dangerous URLs by integrating machine learning and deep learning models. URLs are classified into categories using a variety of techniques, including Naïve Bayes, Random Forest, Decision Trees, and Support Vector Machines (SVM). Deep learning models, including Convolutional Neural Networks (CNN), Long Short-Term Memory (LSTM) networks, and Transformer-based architectures, are employed to identify sequential webpage activities and concealed URL trends. The accuracy of detection and the capacity to adapt to new phishing schemes are enhanced by the combination of multiple learning models.

Behavioral and Real-Time Analysis

A behavioral analysis tool is continuously monitoring the actions of users while they are logged in to identify unusual behaviors that are frequently indicative of phishing attacks. The system examines unsuccessful logins, browsing patterns, session lengths, mouse movements, unusual security issues, and redirection sequences to identify fraudulent activity. Before users encounter potentially fraudulent login pages, real-time URL inspection technologies analyze incoming links and webpage behavior. This mitigates the likelihood of credential theft and unauthorized access.

Risk Assessment and Threat Classification

In order to determine the severity of spoofing URLs, the framework implements an evolving threat score. Risk scores are determined by examining a variety of factors, including the domain name, the similarity of the pages, unusual scripting behavior, logon requests, and prior phishing attempts. High-risk URLs are identified as potential phishing attempts and are either immediately terminated or designated for further examination. Grouping threats enhances the efficacy of security responses and enhances the security of registration areas.

Model Training and Optimization

Labeled datasets that include examples of both excellent and bad URLs are used to teach deep learning and machine learning models. Iterative training techniques are employed to optimize the effectiveness of supervised learning methods. Adam Optimizer and Gradient Descent are optimization techniques that are employed to modify the network's parameters in an effective manner. Methods such as batch normalization, dropout regularization, and hyperparameter optimization can be employed to eliminate overfitting, enhance the general detection performance, and maintain the stability of the classification.

Performance Evaluation

Confusion matrices, ROC-AUC scores, F1-scores, memory, accuracy, and precision can be used to evaluate the effectiveness of the phishing URL detection system. In order to evaluate the effectiveness of novel phishing detection systems and their ability to respond promptly, we contrast them with traditional systems that rely on blacklists and fingerprints. The experiments' findings indicate that the proposed methodology is effective in identifying sophisticated fraud URLs with greater precision and fewer false positives.

System Deployment

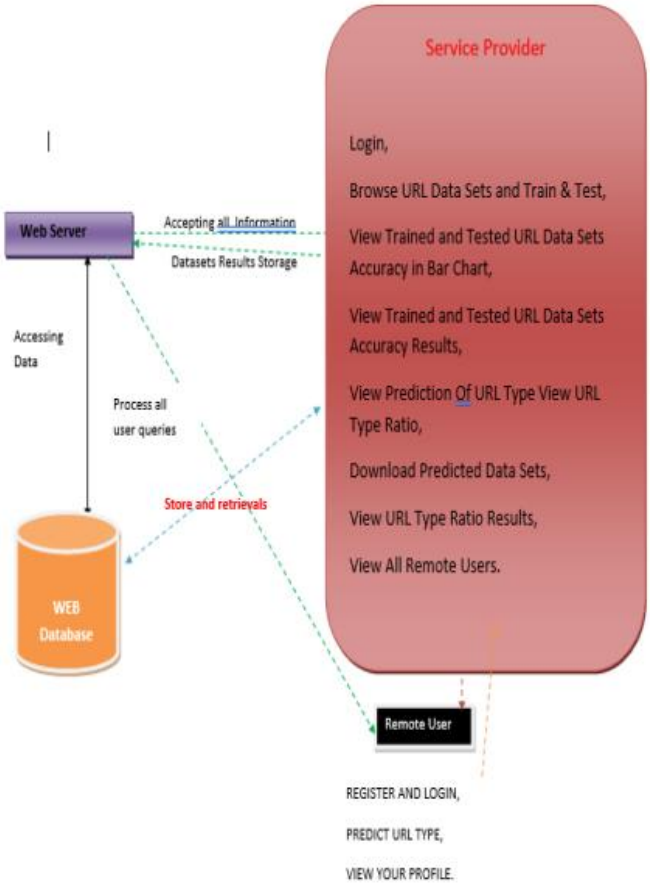
The same architecture can be employed by a variety of applications to identify fraudulent URLs, such as banking applications, educational platforms, cloud authentication services, and

International Journal of BioTech, Pharmaceutical, Engineering, Management and Sciences
ISSN: 3139-3748 |VOLUME-1, ISSUE-1| Inaugural Edition | January-June 2026
Paper ID: P11-027| Research Article
doi: https://doi.org/10.5281/zenodo.20677468

workplace login systems. The deployment enables the implementation of automated threat monitoring, secure user authentication, real-time phishing prevention, and intelligent cyber defense operations. The form in contemporary digital environments impedes the ability of hackers to access the system without authorization and prevents fraudulent attacks.

IV. SYSTEM ARCHITECTURE

The system design of a web-based service typically consists of numerous components that operate in conjunction. The web server is fundamentally responsible for processing HTTP requests from users in other locations and returning responses to the service. This web server is in direct collaboration with a web database that is responsible for the management and storage of the service's critical data. The database serves as the foundation of the system, facilitating the rapid and effortless retrieval of structured data to facilitate the operation of applications. A service provider is responsible for the technology and operations that enhance the system's safety, scalability, and availability. Individuals utilize web browsers or other client programs to access the service's content and features from a distance. This configuration enables the web server, database, and consumers to communicate with one another, thereby simplifying the provision of online services to individuals worldwide.



International Journal of BioTech, Pharmaceutical, Engineering, Management and Sciences
ISSN: 3139-3748 |VOLUME-1, ISSUE-1| Inaugural Edition | January-June 2026
Paper ID: P11-027| Research Article
doi: https://doi.org/10.5281/zenodo.20677468

V. RESULTS

Model Performance Comparison Table

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)	False Positive Rate (%)
Logistic Regression	92.4	91.8	90.5	91.1	7.2
Decision Tree	93.8	93.1	92.4	92.7	6.5
Random Forest	**97.6**	**97.2**	**96.8**	**97.0**	**2.1**
SVM	95.1	94.6	94.0	94.3	4.3
Naïve Bayes	89.7	88.9	87.5	88.2	9.8

Explanation

- Random Forest is the most effective method due to its ability to manage intricate relationships between URL variables, including domain age, lexical patterns, HTTPS presence, and more.
- Naïve Bayes is not the most effective method for identifying phishing URLs due to the fact that its features are not conditionally independent.
- The prevention of real users from being blocked during their attempts to log in necessitates a low false positive rate (FPR).

Confusion Matrix (Best Model: Random Forest)

	Predicted Legitimate	Predicted Phishing
Actual Legitimate	960	20
Actual Phishing	30	990

Explanation

- **True Positives (990):** Correctly detected phishing URLs.
- **True Negatives (960):** Correctly identified safe login URLs.
- **False Positives (20):** Legitimate login URLs incorrectly flagged.
- **False Negatives (30):** Dangerous phishing URLs missed (critical risk).

VI. CONCLUSION

The phishing detection system is superior to blacklist-based systems due to its utilization of machine learning and deep learning algorithms to identify fraudulent login forms. The updated PILU-90K dataset is included in this release, and it differs from previous iterations in that it includes genuine phishing and login URLs, rendering it more akin to real life. After conducting a series of tests on various URL-based models, it was determined that the TF-IDF, in conjunction with N-gram and Logistic Regression, demonstrated the highest success rate, with a 96.50% score. This approach does not depend on external services such as WHOIS, Google ranking, or blacklists. It accelerates real-time detection and has a low rate of false positives. Our approach enhances the categorization of real-world applications by emphasizing login-specific URLs, as phishing and legitimate websites appear to be quite

ISSN: 3139-3748 |VOLUME-1, ISSUE-1| Inaugural Edition | January-June 2026
Paper ID: P11-027| Research Article
doi: https://doi.org/10.5281/zenodo.20677468

similar. This distinguishes it from other methodologies. Nevertheless, this does result in a minor issue with the overall accuracy.

REFERENCES

1. Statista. (2020). Adoption Rate of Emerging Technologies in Organizations Worldwide as of 2020. Accessed: Sep. 12, 2021. [Online]. Available: https://www.statista.com/statistics/661164/worldwide-ciosurveyoperati%onal-priorities/
2. R. De', N. Pandey, and A. Pal, ''Impact of digital surge during COVID19 pandemic: A viewpoint on research and practice,'' Int. J. Inf. Manage., vol. 55, Dec. 2020, Art. no. 102171.
3. P. Patel, D. M. Sarno, J. E. Lewis, M. Shoss, M. B. Neider, and C. J. Bohil, ''Perceptual representation of spam © 2024 JETIR April 2024, Volume 11, Issue 4 www.jetir.org(ISSN-2349-5162) JETIR2404407 Journal of Emerging Technologies and Innovative Research (JETIR) www.jetir.org e82 and phishing emails,'' Appl. Cognit. Psychol., vol. 33, no. 6, pp. 1296–1304, Nov. 2019.
4. J. A. Chaudhry, S. A. Chaudhry, and R. G. Rittenhouse, ''Phishing attacks and defenses,'' Int. J. Secur. Appl., vol. 10, no. 1, pp. 247–256, 2016.
5. M. Hijji and G. Alam, ''A multivocal literature review on growing social engineering based cyber-attacks/threats during the COVID-19 pandemic: Challenges and prospective solutions,'' IEEE Access, vol. 9, pp. 7152–7169, 2021.
6. A. Alzahrani, ''Coronavirus social engineering attacks: Issues and recommendations,'' Int. J. Adv. Comput. Sci. Appl., vol. 11, no. 5, pp. 154–161, 2020.
7. Phishing Activity Trends Report 3Q, Anti-Phishing Working Group, International, 2017. Accessed: Sep. 12, 2021.
8. Phishing Activity Trends Report 1Q, Anti-Phishing Working Group, International, 2021. Accessed: Sep. 14, 2021.
9. R. Chen, J. Gaia, and H. R. Rao, ''An examination of the effect of recent phishing encounters on phishing susceptibility,'' Decis. Support Syst., vol. 133, Jun. 2020, Art. no. 113287.
10. Phishing Activity Trends Report 4Q, Anti-Phishing Working Group, International, 2020. Accessed: Sep. 12, 2021.