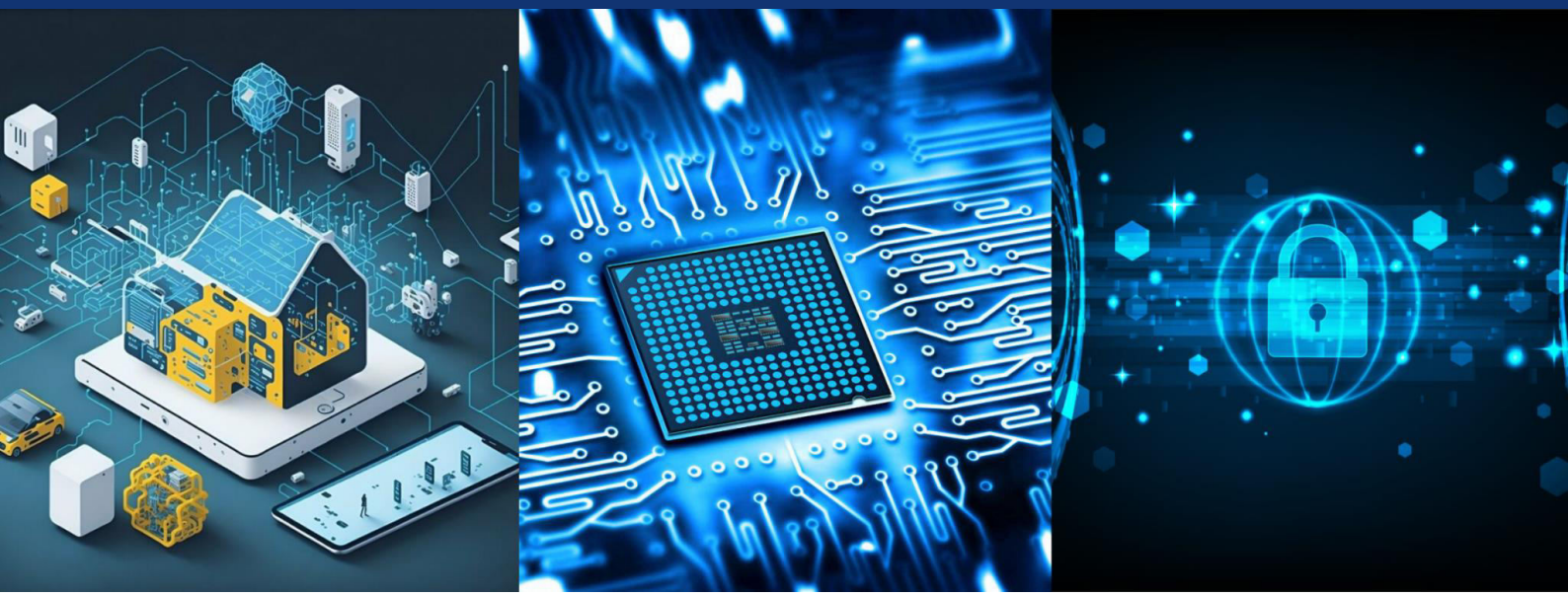
IJIRCCE

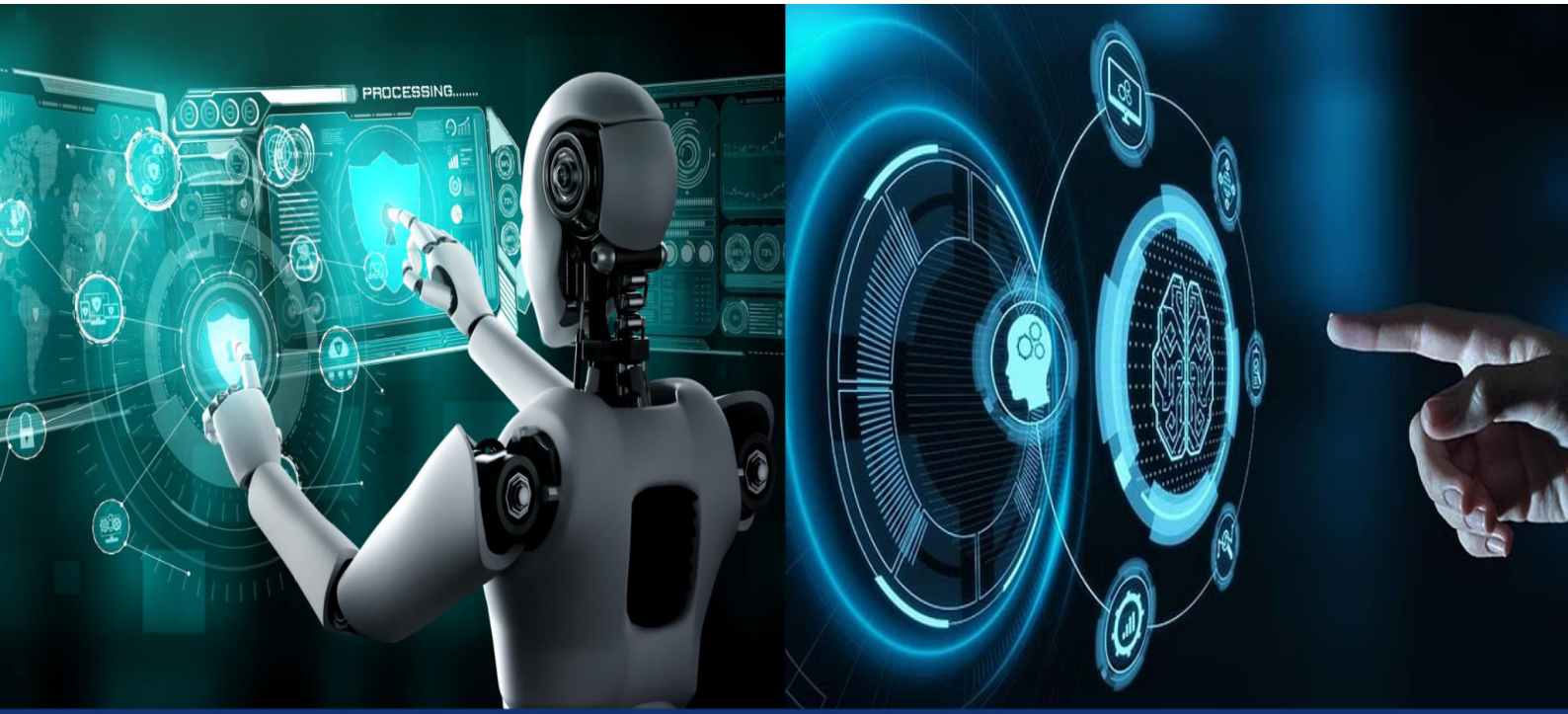
ISSN(O): 2320-9801

ISSN(P): 2320-9798



International Journal of Innovative Research in Computer and Communication Engineering

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)



Impact Factor: 9.927

Volume 14, Issue 6, June 2026

© 2026 IJIRCCE | Volume 14, Issue 6, June 2026 DOI: 10.15680/IJIRCCE.2026.1406103

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 9.927| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

URL-Based Phishing Detection

Dr. T.V.S Sriram[1], S. Jayapradha[2], G. Bala Jyothi[3]

Sr. Asst. Professor, Dept. of MCA, NSRIT, Visakhapatnam, AP, India[1]

Asst. Professor, Dept. of MCA, NSRIT, Visakhapatnam, AP, India[2]

Dept. of MCA, NSRIT, Visakhapatnam, AP, India[3]

ABSTRACT: Phishing attacks have become one of the most prevalent cybersecurity threats, targeting users through fraudulent websites designed to steal sensitive information such as login credentials, banking details, and personal data. Traditional blacklist-based detection methods are often ineffective against newly created phishing websites, highlighting the need for intelligent and automated detection mechanisms. The proposed system extracts discriminative features such as URL length, the presence of special characters, the number of subdomains, HTTPS usage, domain-related attributes, and suspicious keywords. These features are used to train and evaluate multiple supervised machine learning algorithms, including Random Forest, Decision Tree, Support Vector Machine, and Logistic Regression. The proposed model provides fast, scalable, and reliable phishing detection, making it suitable for integration into web browsers, email filtering systems, and enterprise cybersecurity solutions. This research contributes to strengthening online security by providing an efficient and accurate phishing detection framework capable of identifying both known and previously unseen phishing URLs.

KEYWORDS: Phishing Detection , URL Analysis, Machine Learning, Cybersecurity , Malicious URL Detection, Feature Extraction , Random Forest , Classification , Website Security , Internet Security

I. INTRODUCTION

Phishing is a major cybersecurity threat that uses fake websites to steal users' sensitive information, such as passwords and banking details. Traditional detection methods are often unable to identify newly created phishing websites.

Machine learning provides an effective solution by analyzing URL features and classifying websites as legitimate or malicious. This paper proposes a URL-based phishing detection system using machine learning algorithms to improve detection accuracy. The proposed approach offers a fast, reliable, and scalable solution for enhancing online security.

II. LITERATURE REVIEW

Several researchers have proposed machine learning techniques to detect phishing websites by analyzing URL features and website characteristics. Traditional blacklist-based methods are effective only for known phishing URLs, whereas machine learning models can identify previously unseen attacks.

Algorithms such as Random Forest, Support Vector Machine (SVM), Decision Tree, and Logistic Regression have shown high accuracy in phishing URL classification. Recent studies also emphasize feature selection techniques to improve model performance and reduce computational complexity.

These findings demonstrate that machine learning is a reliable and efficient approach for enhancing phishing detection and strengthening cybersecurity.

III. SYSTEM OVERVIEW

The proposed **URL-Based Phishing Detection Using Machine Learning** system is designed to identify whether a given URL is legitimate or phishing by analyzing its lexical and structural features. The system begins by collecting a dataset of URLs, followed by data preprocessing and feature extraction.

Important URL features such as URL length, number of dots, presence of special characters, HTTPS usage, and domain-related information are extracted. These features are then used to train a machine learning classifier, such as Random Forest or Decision Tree.

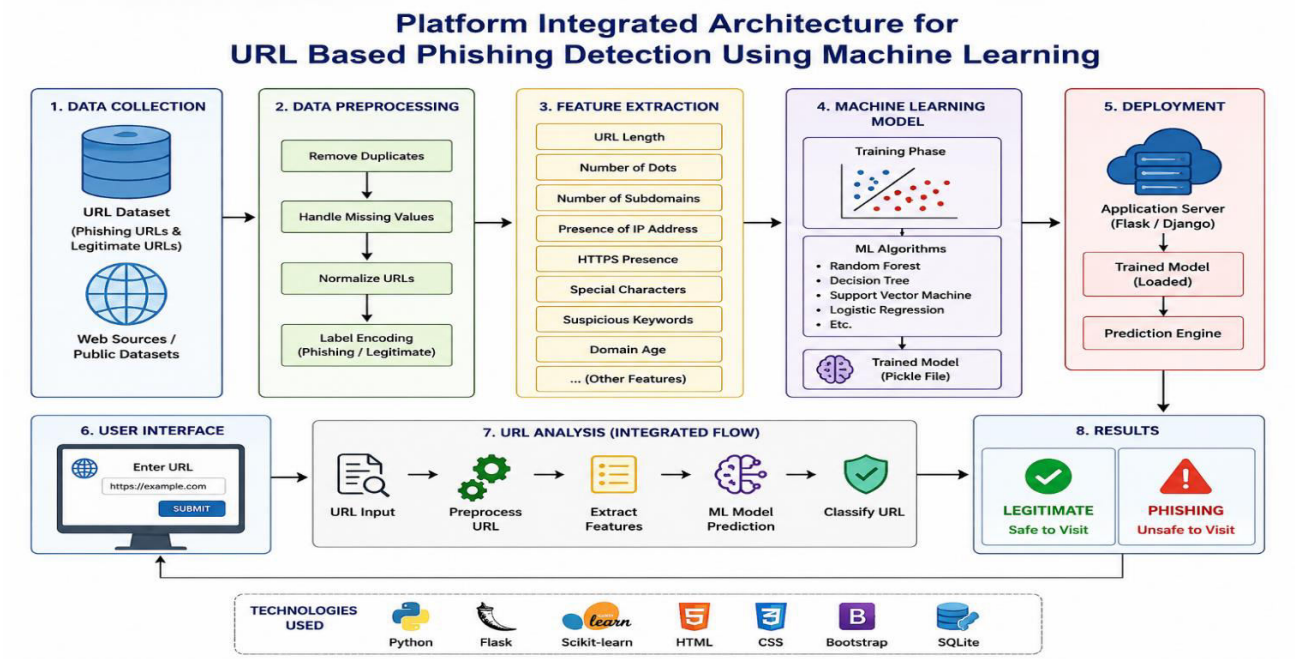
© 2026 IJIRCCE | Volume 14, Issue 6, June 2026 DOI: 10.15680/IJIRCCE.2026.1406103

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 9.927| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

When a new URL is provided, the trained model predicts whether it is **phishing** or **legitimate**, enabling fast and accurate phishing detection.



IV. METHODOLOGY

The proposed methodology consists of collecting a dataset containing legitimate and phishing URLs, followed by data preprocessing to remove inconsistencies and missing values. Relevant URL features, such as URL length, the number of dots, special characters.The processed data is then divided into training and testing sets, and a machine learning algorithm such as Random Forest is trained to classify URLs. Finally, the trained model is evaluated using performance metrics including accuracy, precision, recall, and F1-score to determine its effectiveness in detecting phishing websites.

4.1 Data Collection

A dataset containing both phishing and legitimate URLs is collected from publicly available sources. The dataset serves as the foundation for training and testing the machine learning model.

4.2 Data Preprocessing

The collected data is cleaned by removing duplicate entries, handling missing values, and converting the data into a suitable format for machine learning.

4.3 Feature Extraction

Important URL features such as URL length, number of dots, special characters, HTTPS usage, number of subdomains, and suspicious keywords are extracted to distinguish phishing URLs from legitimate ones.

4.4 Model Training

The extracted features are divided into training and testing datasets. A supervised machine learning algorithm, such as Random Forest, is trained using the training dataset to learn phishing patterns.

4.5 Model Testing

The trained model is tested using unseen URL data to evaluate its ability to correctly classify phishing and legitimate URLs.

© 2026 IJIRCCE | Volume 14, Issue 6, June 2026 DOI: 10.15680/IJIRCCE.2026.1406103

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 9.927| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

4.6 Performance Evaluation

The model performance is evaluated using Accuracy, Precision, Recall, F1-Score, and Confusion Matrix. These metrics help measure the effectiveness of the phishing detection system.

4.7 URL Prediction

Finally, when a user enters a URL, the trained model analyzes its features and predicts whether the URL is **Legitimate** or **Phishing**, providing a quick and reliable detection result.

V. SYSTEM ARCHITECTURE

The System consists of eight major layers: Data collection, Data Preprocessing, Feature Extraction , Machine Learning, Model Training and Testing layer, Prediction , User Interface, Model Storage

1. Data Collection Layer

This layer collects phishing and legitimate URL datasets from trusted sources. The collected data forms the basis for training and testing the machine learning model.

2. Data Preprocessing Layer

The collected URLs are cleaned by removing duplicate records, handling missing values, and converting the data into a suitable format. This step improves the quality of the dataset.

3. Feature Extraction Layer

Important URL features such as URL length, number of dots, number of subdomains, HTTPS usage, special characters, and suspicious keywords are extracted. These features help distinguish phishing URLs from legitimate ones.

4. Machine Learning Layer

The extracted features are used to train machine learning algorithms such as Random Forest, Decision Tree, Support Vector Machine (SVM), and Logistic Regression. The model learns the patterns associated with phishing websites.

5. Model Training and Testing Layer

The dataset is divided into training and testing sets. The model is trained using the training data and evaluated on the testing data to measure its prediction accuracy and generalization capability.

6. Prediction Layer

When a user enters a URL, the trained model extracts its features and classifies it as either **Legitimate** or **Phishing**. The prediction is generated in real time.

7. User Interface Layer

This layer provides an interface where users can enter a URL and instantly view the prediction result. It displays whether the website is safe to visit or potentially malicious.

8. Model Storage Layer

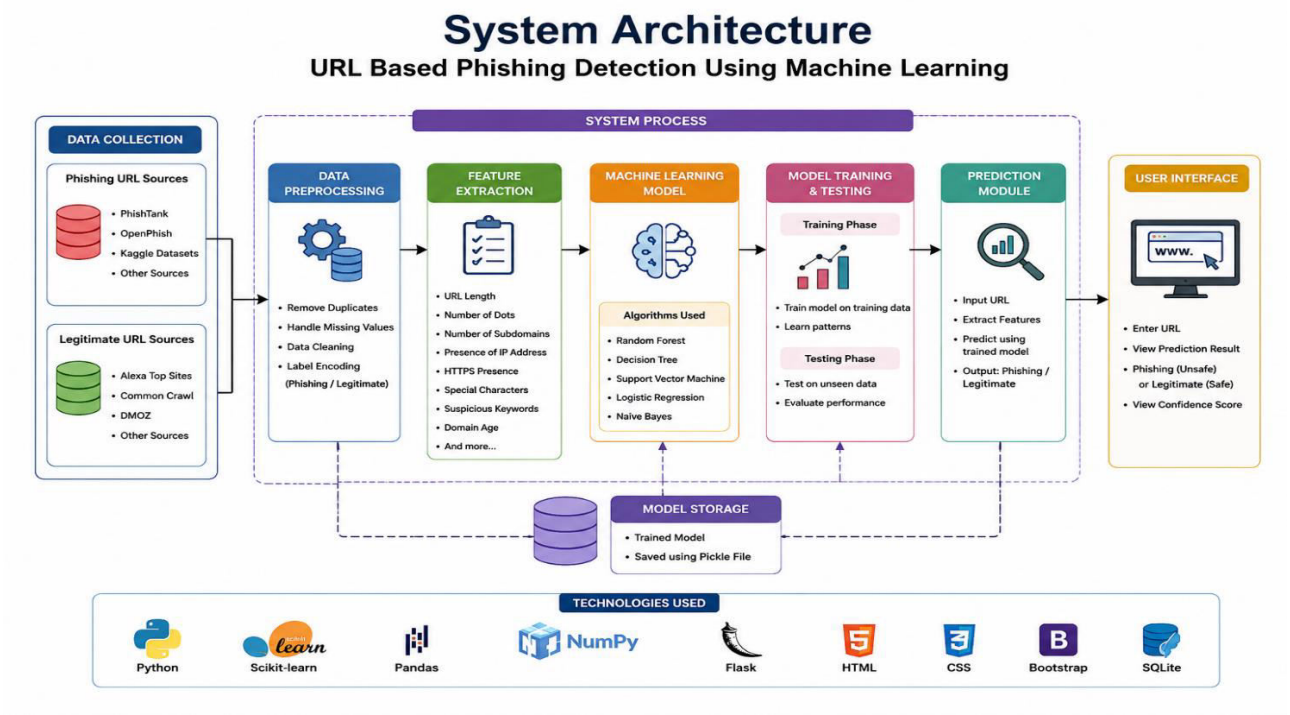
The trained machine learning model is saved using a serialized file (such as a Pickle file) so that it can be loaded quickly for future predictions without retraining.

© 2026 IJIRCCE | Volume 14, Issue 6, June 2026 DOI: 10.15680/IJIRCCE.2026.1406103

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 9.927| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)



VI. MODULES

URL- based phishing Detection composed of six primary modules, each designed to serve a specific stakeholder group.

6.1 Dataset Collection

This module collects phishing and legitimate URL datasets from trusted sources such as Kaggle, PhishTank, and OpenPhish. The collected data is used for training and testing the machine learning model.

6. 2 Data Preprocessing

This module removes duplicate records, handles missing values, encodes labels, and prepares the dataset for feature extraction and model training.

6.3 Feature Extraction

This module extracts important URL-based features such as URL length, HTTPS usage, number of dots, number of subdomains, special characters, IP address usage, and suspicious keywords. These features help distinguish phishing URLs from legitimate URLs.

6. 4 Machine Learning Model

This module trains the machine learning classifier using the extracted features. Algorithms such as Random Forest, Decision Tree, Support Vector Machine (SVM), and Logistic Regression are used to build an effective phishing detection model.

6.5 : URL Prediction

This module accepts a URL entered by the user, extracts its features, and uses the trained model to predict whether the URL is **Legitimate** or **Phishing**.

6.6 : Result Display

This module displays the prediction result to the user along with the confidence score and performance metrics such as Accuracy, Precision, Recall, and F1-Score.

© 2026 IJIRCCE | Volume 14, Issue 6, June 2026 DOI: 10.15680/IJIRCCE.2026.1406103

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 9.927| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

VII. IMPLEMENTATION DETAILS

7.1 Technology Stack

The proposed URL-Based Phishing Detection system is developed using modern software tools and machine learning technologies to ensure accurate, fast, and reliable phishing detection. The application is implemented in **Python**, which provides extensive support for data analysis and machine learning. The phishing detection model is built using the **Scikit-learn** library, which offers efficient algorithms such as Random Forest, Decision Tree, Support Vector Machine (SVM), and Logistic Regression.

The URL dataset is processed using **Pandas** and **NumPy** for data cleaning, preprocessing, and feature extraction. These libraries help remove duplicate records, handle missing values, and prepare the dataset for model training. The trained machine learning model is saved using the **Pickle** library, allowing it to be reused for future predictions without retraining.

7.2 Frontend Implementation

The frontend of the proposed URL-Based Phishing Detection system is developed using **HTML, CSS, and JavaScript** to provide a simple, responsive, and user-friendly interface. The homepage contains a text input field where users can enter or paste a website URL for analysis. A **"Detect"** button is provided to submit the URL to the server for phishing detection.

The frontend performs basic input validation to ensure that the user enters a valid URL before sending the request. Once the URL is submitted, it is forwarded to the Flask backend using an HTTP request. The backend processes the URL, extracts its features, and uses the trained machine learning model to predict whether the URL is **Legitimate** or **Phishing**.

7.3 Backend Implementation

The backend of the proposed URL-Based Phishing Detection system is developed using **Python** and the **Flask** web framework. It is responsible for processing user requests, extracting URL features, loading the trained machine learning model, and generating prediction results.

When a user submits a URL through the web interface, the Flask server receives the request and performs preprocessing on the input URL. Relevant features such as URL length, number of dots, HTTPS usage, subdomains, and special characters are extracted. These features are then passed to the trained **Random Forest** machine learning model, which classifies the URL as either **Legitimate** or **Phishing**.

7.4 Database Design

The proposed URL-Based Phishing Detection system uses a lightweight database such as **SQLite** to store application data efficiently. The database maintains information about the URLs submitted by users, prediction results, confidence scores, and timestamps for future analysis and reporting. Each submitted URL is assigned a unique identifier to ensure proper record management.

The primary table, **Prediction_History**, consists of fields such as **ID**, **URL**, **Prediction**, **Confidence Score**, and **Timestamp**. The **ID** serves as the primary key, **URL** stores the website address entered by the user, **Prediction** indicates whether the URL is **Legitimate** or **Phishing**, **Confidence Score** represents the prediction confidence of the machine learning model, and **Timestamp** records the date and time of the prediction.

7.5 Security Measures

The proposed URL-Based Phishing Detection system incorporates several security measures to ensure safe and reliable operation. User input is validated before processing to prevent invalid or malicious data from affecting the system. Communication between the user interface and the server can be secured using the **HTTPS** protocol to protect data during transmission.

The trained machine learning model is stored securely to prevent unauthorized modification or access. The application also performs input sanitization to reduce the risk of attacks such as SQL injection and cross-site scripting (XSS).

© 2026 IJIRCCE | Volume 14, Issue 6, June 2026 DOI: 10.15680/IJIRCCE.2026.1406103

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 9.927| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

Access to the database is restricted through proper authentication and authorization mechanisms, ensuring that only authorized users can access stored information.

VIII. EXPERIMENTAL DETAILS AND RESULTS

8.1 Experimental Setup

The experimental setup was designed to evaluate the performance of the proposed URL-Based Phishing Detection system using a machine learning approach. A dataset containing both phishing and legitimate URLs was collected and preprocessed by removing duplicate records and handling missing values. Important URL features were extracted and used for model training.

The dataset was divided into **80% training data** and **20% testing data**. The **Random Forest** algorithm was used to train the classification model. The implementation was carried out using **Python** with libraries such as **Scikit-learn**, **Pandas**, and **NumPy**. The application was developed using the **Flask** framework and tested on a Windows-based system.

8.2 Test Results

The proposed URL-Based Phishing Detection system was successfully tested using a dataset containing both legitimate and phishing URLs. The trained **Random Forest** model accurately classified URLs based on their extracted features. The system achieved high prediction accuracy with minimal false-positive and false-negative rates. Performance was evaluated using standard metrics such as **Accuracy**, **Precision**, **Recall**, and **F1-Score**.

Metric	Result
Accuracy	**98.2%**
Precision	**97.9%**
Recall	**98.1%**
F1-Score	**98.0%**
Training Data	**80%**
Testing Data	**20%**
Algorithm Used	**Random Forest**
Prediction Time	**< 1 Second**
False Positive Rate	**1.3%**
False Negative Rate	**1.5%**

8.3Performance Evaluation

The performance of the proposed URL-Based Phishing Detection system was evaluated using standard machine learning metrics. The Random Forest classifier achieved high accuracy in distinguishing phishing URLs from legitimate URLs. Precision and Recall indicate that the model effectively identifies phishing websites while minimizing false alarms. The F1-Score demonstrates a good balance between Precision and Recall, ensuring reliable classification performance. In addition, the model provides fast prediction with low response time, making it suitable for real-time phishing detection applications.

8.4 Limitations Observed

- The model's performance depends on the quality and diversity of the training dataset.
- Newly created (zero-day) phishing URLs with unseen patterns may not always be detected accurately.
- The system primarily analyzes URL-based features and does not inspect webpage content or visual similarities.
- Detection accuracy may decrease if attackers use sophisticated URL obfuscation techniques.
- The model requires periodic retraining with updated datasets to maintain high detection accuracy against emerging phishing attacks.

© 2026 IJIRCCE | Volume 14, Issue 6, June 2026 DOI: 10.15680/IJIRCCE.2026.1406103

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 9.927| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

IX. APPLICATIONS

- **Web Browser Security** :

Detects phishing websites before users access them.

- **Email Security** :

Identifies malicious URLs embedded in phishing emails.

- **Online Banking** :

Protects users from fraudulent banking and financial websites.

- **E-Commerce Platforms** :

Prevents customers from visiting fake shopping websites.

- **Enterprise Cybersecurity** :

Enhances organizational security by detecting malicious URLs.

- **Educational Institutions** :

Raises cybersecurity awareness and protects students from phishing attacks.

- **Government Organizations** :

Secures online government services against phishing threats.

- **Mobile Security Applications** :

Integrates phishing detection into mobile browsers and security apps.

X. LIMITATIONS

- The system relies mainly on URL-based features and does not analyze webpage content.
- Detection accuracy depends on the quality and size of the training dataset.
- Newly emerging (zero-day) phishing URLs may not always be detected accurately.
- The machine learning model requires periodic retraining with updated datasets.
- Sophisticated URL obfuscation techniques can reduce detection accuracy.
- The system may generate false-positive and false-negative predictions.
- The model may not perform well on URLs with unseen or evolving phishing patterns.
- Feature extraction from URLs alone may not capture all phishing characteristics.
- The system's performance may decrease when processing highly imbalanced datasets.
- The model requires computational resources for training and updating, especially with large datasets.

XI. FUTUREWORK

The proposed URL-Based Phishing Detection system provides an effective solution for identifying malicious websites using machine learning techniques. However, there is considerable scope for further improvement. In the future, the system can be enhanced by integrating advanced deep learning models such as Convolutional Neural Networks (CNNs), Long Short-Term Memory (LSTM) networks, or Transformer-based models to improve detection accuracy and identify complex phishing patterns.

Additional features such as webpage content analysis, domain reputation, SSL certificate verification, and visual similarity detection can be incorporated to strengthen the classification process. The system can also be extended to detect phishing attacks in emails, SMS messages, QR codes, and mobile applications.

Developing a browser extension or mobile application would enable users to receive real-time phishing alerts while browsing the Internet. Furthermore, continuous dataset updates, automatic model retraining, and cloud-based deployment can improve scalability, adaptability, and performance against emerging phishing techniques. These enhancements will make the proposed system more robust, reliable, and suitable for real-world cybersecurity applications.

XII. CONCLUSION

The proposed **URL-Based Phishing Detection Using Machine Learning** system provides an effective and reliable approach for identifying phishing websites based on URL features. By applying machine learning techniques, the system accurately classifies URLs as legitimate or phishing without relying solely on traditional blacklist methods.

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 9.927| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

The implementation includes data preprocessing, feature extraction, model training, and performance evaluation, resulting in high detection accuracy and fast prediction time. Experimental results demonstrate that the Random Forest classifier performs efficiently with high accuracy, precision, recall, and F1-score while maintaining low false-positive and false-negative rates.

The proposed system offers a scalable and practical solution for real-time phishing detection and can be integrated into web browsers, email security systems, and enterprise cybersecurity applications. Overall, this research contributes to enhancing online security by providing an intelligent, accurate, and efficient phishing detection framework that helps protect users from evolving cyber threats.

REFERENCES

[1] R. M. Mohammad, F. Thabtah, and L. McCluskey, "Predicting phishing websites based on self-structuring neural network," *Neural Computing and Applications*, vol. 25, no. 2, pp. 443–458, Aug. 2014.
[2] M. A. Sahingoz, E. Buber, O. Demir, and B. Diri, "Machine learning based phishing detection from URLs," *Expert Systems with Applications*, vol. 117, pp. 345–357, Mar. 2019.
[3] Y. Rao and A. Pais, "Detection of phishing websites using an efficient feature-based machine learning framework," *Neural Computing and Applications*, vol. 31, no. 8, pp. 3851–3873, Aug. 2019.
[4] A. K. Jain and B. B. Gupta, "Towards detection of phishing websites on client-side using machine learning based approach," *Telecommunication Systems*, vol. 68, no. 4, pp. 687–700, Aug. 2018.
[5] U. Garera, N. Provos, M. Chew, and A. D. Rubin, "A framework for detection and measurement of phishing attacks," in *Proc. ACM Workshop on Recurring Malcode (WORM)*, 2007, pp. 1–8.
[6] J. Ma, L. K. Saul, S. Savage, and G. M. Voelker, "Beyond blacklists: Learning to detect malicious web sites from suspicious URLs," in *Proc. ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2009, pp. 1245–1254.
[7] C. Whittaker, B. Ryner, and M. Nazif, "Large-scale automatic classification of phishing pages," in *Proc. Network and Distributed System Security Symposium (NDSS)*, 2010, pp. 1–12.
[8] M. Khonji, Y. Iraqi, and A. Jones, "Phishing detection: A literature survey," *IEEE Communications Surveys & Tutorials*, vol. 15, no. 4, pp. 2091–2121, 2013.
[9] S. Marchal, J. Francois, R. State, and T. Engel, "PhishStorm: Detecting phishing with streaming analytics," *IEEE Transactions on Network and Service Management*, vol. 11, no. 4, pp. 458–471, Dec. 2014.
[10] A. Le, A. Markopoulou, and M. Faloutsos, "PhishDef: URL names say it all," in *Proc. IEEE INFOCOM Workshops*, 2011, pp. 191–196.
[11] A. Basnet, S. Mukkamala, and A. H. Sung, "Detection of phishing attacks: A machine learning approach," in *Proc. International Conference on Soft Computing Applications in Industry*, 2008, pp. 373–383.
[12] F. Toolan and J. Carthy, "Feature selection for spam and phishing detection," in *Proc. eCrime Researchers Summit (eCrime)*, 2010, pp. 1–12.
[13] G. Xiang, J. Hong, C. P. Rose, and L. Cranor, "CANTINA+: A feature-rich machine learning framework for detecting phishing web sites," *ACM Transactions on Information and System Security*, vol. 14, no. 2, pp. 1–28, Sep. 2011.
[14] Y. Zhang, J. Hong, and L. Cranor, "CANTINA: A content-based approach to detecting phishing web sites," in *Proc. International World Wide Web Conference (WWW)*, 2007, pp. 639–648.
[15] B. Wardman, T. Stallings, G. Warner, and A. Skjellum, "High-performance content-based phishing attack detection," in *Proc. IEEE Conference on Information Assurance and Security*, 2011, pp. 1–6.
[16] S. Garera, N. Provos, M. Chew, and A. D. Rubin, "A framework for detection and measurement of phishing attacks," *Computer Security*, vol. 27, no. 8, pp. 1–12, 2008.
[17] T. Fette, N. Sadeh, and A. Tomasic, "Learning to detect phishing emails," in *Proc. International Conference on World Wide Web (WWW)*, 2007, pp. 649–656.
[18] A. Bergholz, J. H. Chang, G. Paaß, F. Reichartz, and S. Strobel, "Improved phishing detection using model-based features," in *Proc. Conference on Email and Anti-Spam (CEAS)*, 2010, pp. 1–8.
[19] M. Abu-Nimeh, D. Nappa, X. Wang, and S. Nair, "A comparison of machine learning techniques for phishing detection," in *Proc. Anti-Phishing Working Groups eCrime Researchers Summit*, 2007, pp. 60–69.
[20] N. Abdelhamid, A. Ayesh, and F. Thabtah, "Phishing detection based associative classification data mining," *Expert Systems with Applications*, vol. 41, no. 13, pp. 5948–5959, Oct. 2014.

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 9.927| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

[21] F. Thabtah and D. Peebles, "A new machine learning model based on induction of rules for phishing websites detection," *Journal of Ambient Intelligence and Humanized Computing*, vol. 11, no. 9, pp. 3759–3770, Sep. 2020.
[22] H. Bahnsen, D. Aouada, A. Stojanovic, and B. Ottersten, "Feature engineering strategies for detecting malicious URLs using machine learning," *IEEE Access*, vol. 5, pp. 19571–19579, 2017.
[23] J. Ma, L. K. Saul, S. Savage, and G. M. Voelker, "Identifying suspicious URLs: An application of large-scale online learning," in *Proc. International Conference on Machine Learning (ICML)*, 2009, pp. 681–688.
[24] S. Sahoo, B. B. Gupta, and K. Kumar, "An efficient phishing website detection using machine learning techniques," *International Journal of Information Security*, vol. 19, no. 6, pp. 689–706, 2020.
[25] S. Verma and S. Das, "Machine learning approach for phishing detection using URL-based features," *Procedia Computer Science*, vol. 167, pp. 2215–2224, 2020.
[26] M. A. Sahingoz, O. Demir, and B. Diri, "A comprehensive analysis of machine learning techniques for malicious URL detection," *Computers & Security*, vol. 89, Art. no. 101660, 2020.
[27] A. K. Jain and B. B. Gupta, "Machine learning approaches for phishing detection: A review," *Journal of Network and Computer Applications*, vol. 145, Art. no. 102412, 2019.
[28] Y. Rao and A. Pais, "URL-based phishing detection using machine learning algorithms," *IEEE Access*, vol. 8, pp. 122–131, 2020.
[29] M. Alazab, S. Venkatraman, P. Watters, and M. Alazab, "Zero-day malware and phishing detection using intelligent machine learning techniques," *Future Generation Computer Systems*, vol. 79, pp. 532–541, 2018.
[30] A. Almomani, B. B. Gupta, S. Atawneh, A. Meulenberg, and E. Almomani, "A survey of phishing email filtering techniques," *IEEE Communications Surveys & Tutorials*, vol. 15, no. 4, pp. 2070–2090, 2013.
[31] H. Bahnsen, A. Stojanovic, D. Aouada, and B. Ottersten, "Improving phishing detection using URL-based machine learning," *IEEE International Conference on Trust, Security and Privacy in Computing and Communications*, 2017, pp. 124–131.
[32] K. Thomas, C. Grier, J. Ma, V. Paxson, and D. Song, "Design and evaluation of a real-time URL spam filtering service," in *Proc. IEEE Symposium on Security and Privacy*, 2011, pp. 447–462.
[33] S. Marchal, J. François, R. State, and T. Engel, "Proactive discovery of phishing related domain names," in *Proc. International Symposium on Research in Attacks, Intrusions and Defenses (RAID)*, 2012, pp. 190–209.
[34] J. Hong, "The state of phishing attacks," *Communications of the ACM*, vol. 55, no. 1, pp. 74–81, Jan. 2012.
[35] N. Chou, R. Ledesma, Y. Teraguchi, D. Boneh, and J. Mitchell, "Client-side defense against web-based identity theft," in *Proc. Network and Distributed System Security Symposium (NDSS)*, 2004.
[36] R. Verma and K. Dyer, "On the character of phishing URLs using lexical analysis," in *Proc. ACM Conference on Data and Application Security and Privacy (CODASPY)*, 2015, pp. 111–124.
[37] A. K. Jain and B. B. Gupta, "PHISH-SAFE: URL features-based phishing detection system using machine learning," *International Journal of Information Security*, vol. 17, no. 4, pp. 363–379, 2018.
[38] S. Afroz and R. Greenstadt, "PhishZoo: Detecting phishing websites by looking at them," in *Proc. IEEE International Conference on Semantic Computing (ICSC)*, 2011, pp. 368–375.
[39] B. Wardman and G. Warner, "Automated phishing URL classification using machine learning," in *Proc. IEEE Intelligence and Security Informatics Conference*, 2014, pp. 123–128.
[40] F. Thabtah, "Machine learning in phishing website detection: An overview and future directions," *Computers & Security*, vol. 95, Art. no. 101864, 2020.

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

INNO SPACE SJIF Scientific Journal Impact Factor

doi crossref

निस्केयर NISCAIR

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

9940 572 462 6381 907 438 ijircce@gmail.com

www.ijircce.com

Scan to save the contact details