iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 14 **Issue:** IV **Month of publication:** April 2026

DOI: https://doi.org/10.22214/ijraset.2026.80606

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 14 Issue IV Apr 2026- Available at www.ijraset.com

AI-Based Phishing Detection System

Shaik Dharvesh Abbas[1], Chintha Karthikeya[2], Pragada Dyva Sagar[3], Saka Sri Praveen[4]
Department of Computer Applications, Aditya University, Surampalem, India,

Abstract: ***Phishing attacks remain one of the most critical cybersecurity threats, exploiting user trust through deceptive emails, malicious URLs, and fake web interfaces. Traditional detection approaches such as blacklist-based systems are ineffective against newly emerging and dynamically generated phishing attacks. This paper proposes an intelligent phishing detection framework that integrates machine learning and natural language processing techniques for improved accuracy and adaptability. The system analyzes a combination of URL-based, content-based, and domain-related features to identify malicious patterns. Multiple supervised learning models, including Random Forest, Support Vector Machine, and Logistic Regression, are evaluated using standard performance metrics. Experimental results demonstrate that the proposed hybrid approach achieves high detection accuracy while reducing false positives, making it suitable for real-time cybersecurity applications***

I. INTRODUCTION

The rapid expansion of internet services and digital communication platforms has significantly increased the exchange of sensitive information across online environments. While this digital transformation has improved accessibility and convenience, it has also created opportunities for various cyber threats. Among these, phishing attacks have emerged as one of the most prevalent and damaging forms of cybercrime. In such attacks, malicious actors impersonate legitimate organizations or individuals to deceive users into revealing confidential information such as login credentials, banking details, and personal data.

Phishing attacks are typically carried out through deceptive emails, fraudulent websites, and manipulated URLs that closely resemble trusted sources. Due to their realistic appearance and evolving strategies, these attacks are often difficult for users to identify. Traditional security mechanisms, including blacklist-based filtering and rule-based detection systems, rely heavily on previously identified threats. Although effective against known attacks, these methods are inadequate in detecting newly generated or obfuscated phishing attempts, which limits their effectiveness in dynamic environments.

To overcome these limitations, intelligent detection techniques based on machine learning have gained considerable attention in recent years. Machine learning models can analyze large volumes of data, learn hidden patterns, and classify malicious activities with improved accuracy. By leveraging features such as URL structure, domain information, and webpage content, these models can identify suspicious behavior even in previously unseen data. In addition, Natural Language Processing (NLP) techniques enhance detection capabilities by examining textual content, enabling the identification of deceptive language patterns commonly used in phishing messages.

This paper proposes a hybrid phishing detection system that integrates machine learning algorithms with both structural and content-based analysis. The system extracts multiple features from URLs and textual data, which are then used to train supervised classification models. By combining different feature types, the proposed approach improves detection performance while reducing false positives.

The main contribution of this work lies in the development of a unified framework that incorporates URL-based feature extraction and NLP-driven content analysis to enhance phishing detection accuracy. The system is designed to be adaptive, scalable, and suitable for real-time deployment, addressing the challenges posed by continuously evolving phishing techniques.

II. LITERATURE REVIEW

Over the years, multiple approaches have been proposed to address phishing detection challenges. Early solutions primarily relied on blacklist-based filtering and heuristic rules, where URLs and email sources were compared against known malicious entries. Although simple to implement, these approaches required frequent updates and were ineffective against newly generated phishing attacks.

With the introduction of machine learning techniques, researchers began focusing on feature-based detection methods. Studies have shown that analyzing URL characteristics, such as length, presence of special symbols, and domain-related attributes, can significantly improve classification performance. Algorithms such as Decision Trees, Support Vector Machines (SVM), and Naive Bayes have been widely used for this purpose.

Among various models, ensemble techniques like Random Forest have demonstrated strong performance due to their ability to combine multiple decision trees and reduce overfitting. In addition to structural analysis, Natural Language Processing (NLP) has been applied to examine email content, including vocabulary patterns, writing style, and suspicious keywords.

Recent advancements have also explored deep learning models, including Convolutional Neural Networks (CNNs) and Long Short-Term Memory (LSTM) networks, to capture complex relationships within data. Despite these improvements, challenges such as dataset imbalance, feature selection, and computational complexity remain significant. The system proposed in this work builds upon these existing methods by integrating effective feature extraction with supervised learning techniques to enhance detection reliability.

III. SYSTEM ARCHITECTURE

The proposed phishing detection system is structured as a modular pipeline consisting of multiple interconnected components. Each module performs a specific task to ensure accurate and efficient classification of inputs.

The process begins with the data collection module, which gathers labeled datasets containing both phishing and legitimate samples from publicly available sources. These datasets serve as the foundation for training and evaluating the model.

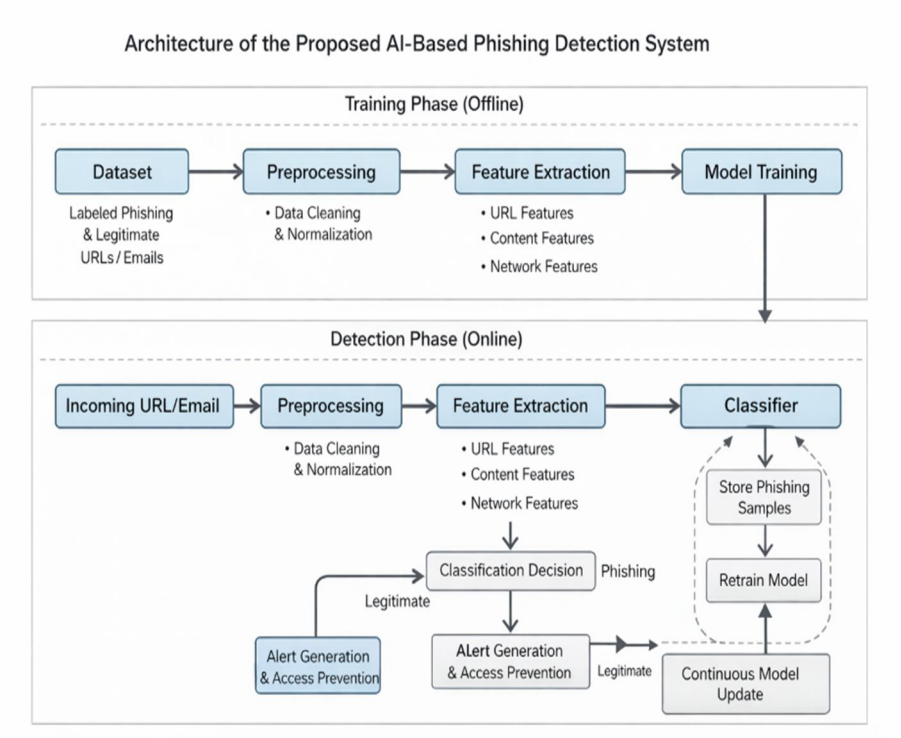
The preprocessing stage focuses on improving data quality by removing inconsistencies such as missing values, duplicate records, and irrelevant information. Text normalization techniques are applied to standardize the input data, ensuring consistency across samples.

Feature extraction plays a critical role in the system by identifying meaningful attributes from the data. These features are categorized into URL-based features (e.g., URL length, presence of IP address), content-based features (e.g., keyword frequency, HTML structure), and network-based features (e.g., domain information and DNS records).

The classification module applies supervised machine learning algorithms to analyze the extracted features and predict whether a given input is malicious or legitimate. Once trained, the model can process new inputs in real time. If a phishing attempt is detected, the system generates an alert and restricts access to the suspicious content.

An additional feedback mechanism is incorporated to continuously update the model using newly identified phishing samples, ensuring that the system remains effective against evolving threats.

Fig. 1. Proposed System Architecture



ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 14 Issue IV Apr 2026- Available at www.ijraset.com

IV. PROPOSED METHODOLOGY

The development of the proposed system follows a structured workflow consisting of several key stages.

1) Initially, a dataset containing both phishing and legitimate samples is collected and prepared for analysis. The preprocessing stage involves cleaning the data, handling missing values, and converting textual information into a standardized format.

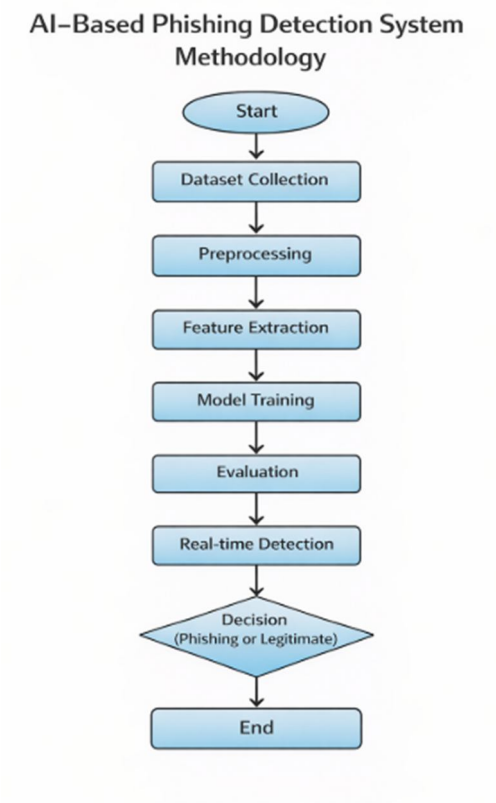
2) In the feature extraction phase, relevant attributes are derived from the dataset. URL-based features include characteristics such as the use of HTTPS, number of subdomains, and presence of special characters. Content-based features focus on identifying suspicious keywords and analyzing text patterns within emails or web pages.

3) The dataset is then divided into training and testing subsets to evaluate model performance. Multiple machine learning algorithms, including Random Forest, Logistic Regression, and Support Vector Machine, are applied to identify the most effective model.

4) Model evaluation is carried out using performance metrics such as accuracy, precision, recall, and F1-score. These metrics provide insight into the system's ability to correctly identify phishing attempts while minimizing errors.

5) Finally, the selected model is deployed in a real-time detection environment, where incoming data is processed and classified automatically. The system responds immediately to detected threats by generating alerts and preventing user interaction with malicious content.

Fig. 2. Proposed Methodology Flow Diagram



V. EXPERIMENTAL RESULTS AND ANALYSIS

The proposed system was evaluated using the publicly available UCI Phishing Websites Dataset, which contains both legitimate and phishing samples. The dataset was preprocessed and divided into training (80%) and testing (20%) subsets.

Three machine learning algorithms were implemented and compared:

- Random Forest
- Support Vector Machine (SVM)
- Logistic Regression

The performance of each model was evaluated using standard metrics such as Accuracy, Precision, Recall, and F1-score.

Algorithm	Accuracy	Precision	Recall	F1-Score
Random Forest	96.2%	95.8%	95.1%	95.4%
SVM	93.5%	92.7%	91.9%	92.3%
Logistic Regression	91.8%	90.5%	89.6%	90.0%

The results indicate that the Random Forest algorithm outperforms other models due to its ability to handle complex feature interactions and reduce overfitting. The integration of both structural and textual features significantly improves detection performance compared to single-feature approaches.

VI. CONCLUSION

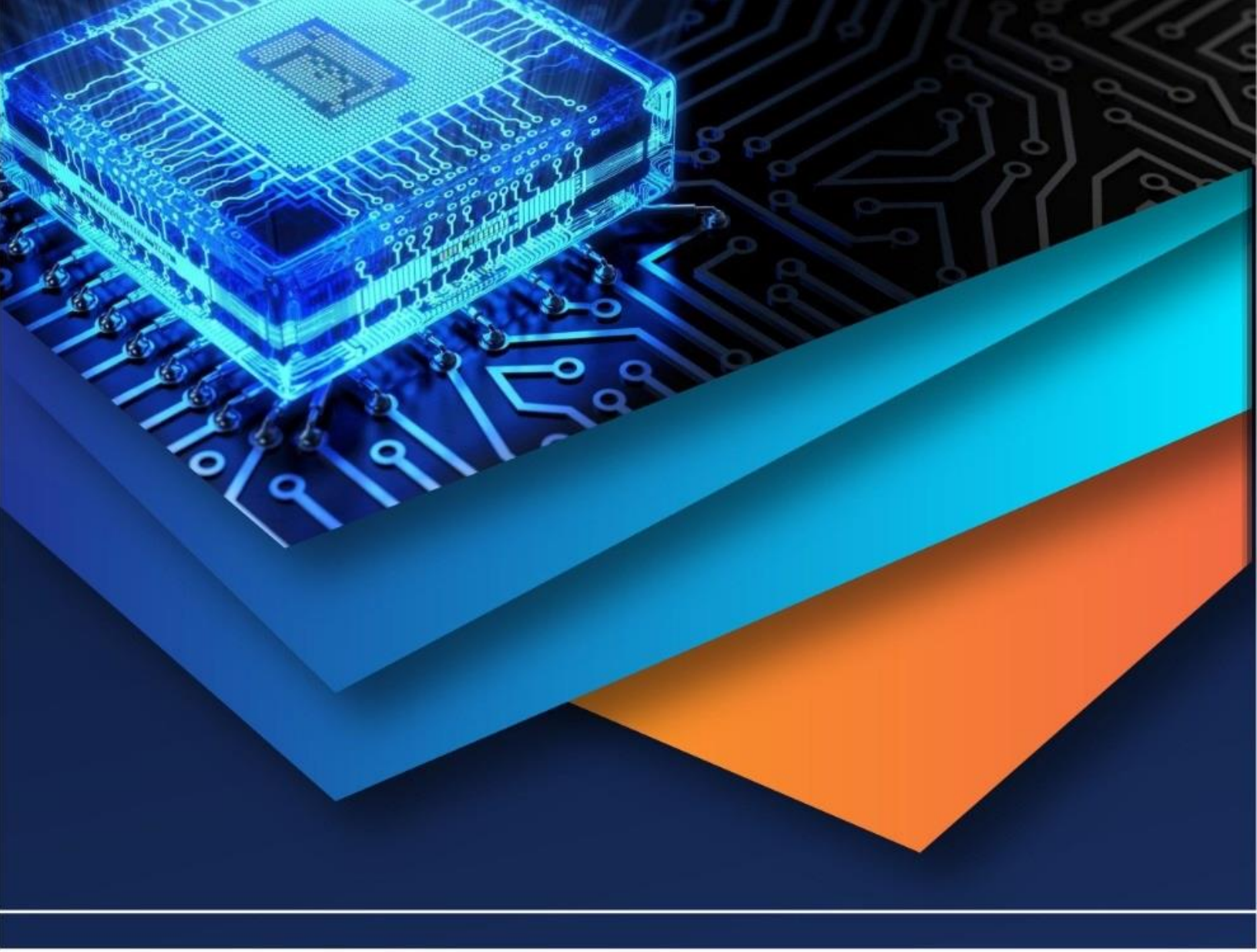
This paper presented a hybrid machine learning-based phishing detection system that combines URL analysis with content-based feature extraction using NLP techniques. The proposed model addresses the limitations of traditional detection approaches by improving adaptability and detection accuracy. Experimental results demonstrate that the system achieves high performance across multiple evaluation metrics, making it suitable for real-time deployment. Future work may focus on incorporating deep learning techniques and real-time browser integration to further enhance system effectiveness and scalability.

VII. ACKNOWLEDGMENT

The authors would like to express their sincere gratitude to the Department of Computer Applications, Aditya University, Surampalem, for providing the necessary resources, infrastructure, and academic support required to carry out this research. The authors are also deeply thankful to the faculty members and mentors for their continuous encouragement, invaluable guidance, and constructive feedback throughout the course of this work.

REFERENCES

[1] M. Jakobsson and S. Myers, Phishing and Countermeasures: Understanding the Increasing Problem of Electronic Identity Theft, Wiley, 2006.
[2] A. Bergholz et al., "New filtering approaches for phishing email," Journal of Computer Security, vol. 18, no. 1, pp. 7–35, 2010.
[3] J. Ma et al., "Beyond Blacklists: Learning to Detect Malicious Web Sites from Suspicious URLs," Proc. ACM SIGKDD, 2009.
[4] R. Verma and A. Das, "What's in a URL: Fast Feature Extraction and Malicious URL Detection," Proc. ACM BADGERS, 2017.
[5] S. Garera et al., "A Framework for Detection and Measurement of Phishing Attacks," Proc. ACM Workshop on Recurring Malcode, 2007.
[6] F. Salahdine, Z. El Mrabet, and N. Kaabouch, "Phishing Attacks Detection Using Machine Learning," 2022.
[7] D. Sahoo et al., "Malicious URL Detection Using Machine Learning: A Survey," 2017.
[8] V. Shahrivari et al., "Phishing Detection Using Machine Learning Techniques," 2020.
[9] R. Jayaraj et al., "Intrusion Detection Based on Phishing Detection Using Machine Learning," 2024.
[10] A. Daud et al., "Phishing Website Detection Using Deep Learning Models," 2023.



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY