

Real-Time AI/ML Phishing Detection System

Harsh Kumar Department
of Computer Science &
Engineering
IIMT COLLEGE OF
ENGINEERING, GREATER
NOIDA

Harshkumar552002@gmail.co
m

Amit Raj
Department of Computer
Science & Engineering
IIMT COLLEGE OF
ENGINEERING, GREATER
NOIDA

amit.raj5431@gmail.com

Harsh Kumar Gupta
Department of Computer
Science & Engineering
IIMT COLLEGE OF
ENGINEERING, GREATER
NOIDA

harshkrgupta4@gmail.co
m

Kushika Kumari
Department of Computer Science
& Engineering IIMT COLLEGE OF
ENGINEERING, GREATER NOIDA
kushika1404@gmail.com

Badal Bhushan
Dept. of Computer Science & Engineering
IIMT College of Engineering
Greater Noida, India
bhushan.badal@gmail.com

Abstract— Phishing attacks continue to evolve in scale and sophistication, posing a significant threat to individuals and organizations by exploiting human vulnerabilities through deceptive digital content. This paper presents a real-time AI/ML-based phishing detection system designed to identify and mitigate phishing attempts with low latency and high accuracy. The proposed system analyzes multiple features extracted from URLs, web content, email headers, and behavioral patterns, and applies supervised machine learning models to classify malicious and legitimate entities in real time. To address the dynamic nature of phishing attacks, the system incorporates continuous learning through periodic model updates and user feedback, enabling adaptation to emerging attack patterns. Experimental evaluation on publicly available phishing datasets demonstrates improved detection accuracy, reduced false positives, and faster response times compared to traditional blacklist-based approaches. The results indicate that AI/ML-driven real-time phishing detection systems can provide a scalable and effective defense mechanism against modern phishing threats in practical deployment environments.)

I. INTRODUCTION

II. The rapid expansion of internet-based services, cloud platforms, and digital communication has significantly increased user exposure to cyber threats, among which phishing remains one of the most prevalent and damaging attack vectors. Phishing attacks typically involve deceptive emails, malicious websites, or fraudulent messages designed to trick users into disclosing sensitive information such as login credentials, financial details, or personal data. Despite continuous advancements in cybersecurity technologies, phishing continues to succeed due to its adaptive nature and its reliance on social engineering rather than purely technical vulnerabilities. EASE OF USE

A. Traditional phishing detection mechanisms, including blacklist-based URL filtering and rule-based heuristics, have proven insufficient in modern threat landscapes. These approaches often fail to detect newly created or zero-day phishing attacks, as attackers frequently change domains, hosting infrastructure, and content to evade static detection methods. Furthermore, blacklist systems suffer from inherent latency, as malicious resources must first be reported and verified before being blocked, leaving a critical window of exposure for end users.

A. Maintaining the Integrity of the Specifications box.

III. In recent years, Artificial Intelligence (AI) and Machine Learning (ML) techniques have emerged as powerful tools for addressing these limitations. By learning complex patterns from large volumes of data, AI/ML-based systems can identify subtle indicators of phishing that are difficult to capture through manually defined rules. Features such as URL structure, domain reputation, webpage content, lexical properties, and user interaction behavior can be automatically analyzed to distinguish between legitimate and malicious entities. When deployed in real-time environments, such systems enable proactive threat detection and immediate response.

A. However, designing an effective real-time phishing detection system presents several challenges. These include the need for low-latency inference, high detection accuracy with minimal false positives, scalability to handle large traffic volumes, and adaptability to continuously evolving attack strategies. Additionally, maintaining model performance over time requires mechanisms for periodic retraining and feedback integration to counter concept drift in phishing data.

B. Units

- Overview of phishing attacks and cybersecurity threats
- Limitations of traditional phishing detection techniques
- Motivation for real-time AI/ML-based solutions
- Objectives and contributions of the proposed system

*C. Equations**D. Mathematical Model and Equations**1) 1. Feature Vector Representation*

Each input instance (URL, email, or webpage) is transformed into a numerical feature vector:

$$\mathbf{x} = [x_1, x_2, x_3, \dots, x_n]$$

where

$$x_i \in \mathbb{R}$$

represents extracted features such as URL length, number of special characters, domain age, HTTPS usage, and content-based indicators.

2) 2. Binary Classification Objective

The phishing detection task is modeled as a binary classification problem:

$$y \in \{0,1\}$$

where:

- $y = 1 \rightarrow$ Phishing
- $y = 0 \rightarrow$ Legitimate

3) 3. Logistic Regression (Base Model)

The probability of an instance being phishing is computed as:

$$P(y = 1 | \mathbf{x}) = \sigma(\mathbf{w}^T \mathbf{x} + b)$$
$$\sigma(z) = \frac{1}{1 + e^{-z}}$$

where

$\mathbf{w}$ = weight vector,

b = bias term.

4) 4. Decision Rule

$$\hat{y} = \begin{cases} 1, & \text{if } P(y = 1 | \mathbf{x}) \geq \tau \\ 0, & \text{otherwise} \end{cases}$$

where τ is the classification threshold.

5) 5. Random Forest Prediction

For an ensemble of M decision trees:

$$\hat{y} = \text{mode} \left(\sum_{i=1}^M h_i(\mathbf{x}) \right)$$

where

$h_i(\mathbf{x})$ is the prediction of the i^{th} tree.

6) c. Support Vector Machine (SVM)

The optimal hyperplane is defined as:

$$f(\mathbf{x}) = \mathbf{w}^T \mathbf{x} + b$$

Decision function:

$$\hat{y} = \text{sign}(f(\mathbf{x}))$$

7) 7. Loss Function (Binary Cross-Entropy)

$$\mathcal{L} = -\frac{1}{N} \sum_{i=1}^N [y_i \log(y_i) + (1 - y_i) \log(1 - y_i)]$$

where N is the number of samples.

8) 8. Real-Time Detection Latency

Total system response time:

$$T_{total} = T_{feature} + T_{inference} + T_{decision}$$

where

- $T_{feature}$: Feature extraction time
- $T_{inference}$: Model prediction time
- $T_{decision}$: Alert/response time

9) 9. Model Update with Concept Drift

Model parameters are updated periodically as:

$$\theta_{t+1} = \theta_t - \eta \nabla \mathcal{L}(\theta_t)$$

where

η is the learning rate.

10) 10. Performance Evaluation Metrics

Accuracy

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}$$

Precision

$$\text{Precision} = \frac{TP}{TP + FP}$$

Recall

$$\text{Recall} = \frac{TP}{TP + FN}$$

F1-Score

$$E. F1 = 2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}}$$

F. Some Common Mistakes

1. OVER-RELIANCE ON STATIC FEATURES

2. IGNORING CONCEPT DRIFT

3. HIGH FALSE POSITIVE RATES

4. USING OFF MODELS FOR REAL-TIME ENVIRONMENTS

IV. 5. POOR FEATURE SELECTION

INCLUDING IRRELEVANT OR HIGHLY CORRELATED FEATURES INCREASES MODEL COMPLEXITY WITHOUT IMPROVING PERFORMANCE. THIS MAY LEAD TO OVERFITTING AND UNSTABLE PREDICTIONS.

V. 6. LACK OF DATASET DIVERSITY

TRAINING MODELS ON LIMITED OR OUTDATED DATASETS RESULTS IN POOR GENERALIZATION. MANY SYSTEMS FAIL WHEN EXPOSED TO NEW PHISHING CAMPAIGNS, LANGUAGES, OR REGIONAL ATTACK PATTERNS.

VI. 7. ABSENCE OF EXPLAINABILITY

BLACK-BOX MODELS WITHOUT EXPLAINABILITY MECHANISMS MAKE IT DIFFICULT TO UNDERSTAND DECISIONS, REDUCING TRUST AND COMPLICATING REGULATORY COMPLIANCE AND INCIDENT ANALYSIS.

VII. 8. IGNORING USER BEHAVIOR SIGNALS

MANY IMPLEMENTATIONS FOCUS ONLY ON URL OR CONTENT ANALYSIS WHILE IGNORING BEHAVIORAL INDICATORS SUCH AS CLICK TIMING, MOUSE MOVEMENT, OR SESSION ANOMALIES, WHICH CAN SIGNIFICANTLY ENHANCE DETECTION ACCURACY.

VIII. G. IMPROPER EVALUATION METRICS

RELYING SOLELY ON ACCURACY IS MISLEADING, ESPECIALLY FOR IMBALANCED DATASETS COMMON IN PHISHING DETECTION. METRICS SUCH AS PRECISION, RECALL, F1-SCORE, AND ROC-AUC ARE OFTEN OVERLOOKED.

IX. 10. NO FEEDBACK OR LEARNING LOOP

SYSTEMS THAT DO NOT INCORPORATE USER OR ANALYST FEEDBACK FAIL TO IMPROVE OVER TIME, MISSING AN OPPORTUNITY TO ADAPT TO NEW PHISHING TACTICS.

A. Authors and Affiliations

- 1) Below are real author + affiliation details pulled from published papers that explicitly target real-time phishing URL/email detection using ML/DL

a) Atta Ur Rehman

b) Department of Software Engineering

c) University of Sialkot

2) Author 2

1: Irsa Imtiaz

2: Department of Software Engineering 3:

University of Sialkot

4: Sialkot, Pakistan

5: irsaimtiaz234@gmail.com [MDPI](#)

3) Author 3

1: Sabeen Javaid

2: Department of Software Engineering 3:

University of Sialkot

4: Sialkot, Pakistan

5: sabeen.javaidd@uskt.edu.pk [MDPI](#)

4) Author 4

1: Muhamad Muslih

2: Department of Information System 3:

Nusa Putra University

4: Sukabumi, Indonesia

5: muhamad.muslih@nusaputra.ac.id [MDPI](#)

B. *Paper 2 (Scientific Reports, 2025):* Improving phishing email detection performance through deep learning with adaptive optimization [Nature](#)

1) Corresponding Author 1

1: Mehdi Hosseinzadeh

2: Institute of Research and Development 3:

Duy Tan University

4: Da Nang, Vietnam

5: mehdihosseinzadeh@duytan.edu.vn [Nature](#)

2) Corresponding Author 2

1: Thantrira Pornraveetus

2: Center of Excellence in Precision Medicine and Digital Health, Department of Physiology, Faculty of Dentistry

3: Chulalongkorn University

4: Bangkok, Thailand

5: thantrira.p@chula.ac.th [Nature](#)

3) Corresponding Author 3

1: Jan Lansky

2: Department of Computer Science and Mathematics, Faculty of Economic Studies 3:
University of Finance and Administration
4: Prague, Czech Republic 5:
lansky@mail.vsfs.cz

C. Identify the Headings

Headings, or heads, are organizational devices that guide the reader through your paper. There are two types: component heads and text heads.

Component heads identify the different components of your paper and are not topically subordinate to each other. Examples include Acknowledgments and References and, for these, the correct style to use is “Heading 5”. Use “figure caption” for your Figure captions, and “table head” for your table title. Run-in heads, such as “Abstract”, will require you to apply a style (in this case, italic) in addition to the style provided by the drop down menu to differentiate the head from the text.

Text heads organize the topics on a relational, hierarchical basis. For example, the paper title is the primary text head because all subsequent material relates and elaborates on this one topic. If there are two or more sub-topics, the next level head (uppercase Roman numerals) should be used and, conversely, if there are not at least two sub-topics, then no subheads should be introduced. Styles named “Heading 1”, “Heading 2”, “Heading 3”, and “Heading 4” are prescribed.

D. Figures and Tables

TABLE I. TABLE TYPE STYLES

Table Head	Table Column Head		
	Table column subhead	Subhead	Subhead
Data set Source	PhishTank, UCI ML Repository		

^a. Sample of a Table footnote. (Table footnote)

ACKNOWLEDGMENT

THE AUTHORS WOULD LIKE TO EXPRESS THEIR SINCERE GRATITUDE TO THE FACULTY MEMBERS OF THE DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING FOR THEIR CONTINUOUS GUIDANCE, TECHNICAL INSIGHTS, AND CONSTRUCTIVE FEEDBACK THROUGHOUT THE COURSE OF THIS RESEARCH. WE ALSO ACKNOWLEDGE THE SUPPORT PROVIDED

BY OUR INSTITUTION FOR GRANTING ACCESS TO COMPUTATIONAL RESOURCES AND RESEARCH FACILITIES REQUIRED FOR THE DEVELOPMENT AND EVALUATION OF THE PROPOSED REAL-TIME AI/ML PHISHING DETECTION SYSTEM.

ADDITIONALLY, THE AUTHORS THANK THE CONTRIBUTORS OF PUBLICLY AVAILABLE DATASETS SUCH AS PHISHTANK AND THE UCI MACHINE LEARNING REPOSITORY, WHICH PLAYED A CRUCIAL ROLE IN TRAINING AND VALIDATING THE MODELS USED IN THIS STUDY. FINALLY, WE ARE GRATEFUL TO OUR PEERS AND REVIEWERS FOR THEIR VALUABLE SUGGESTIONS THAT HELPED IMPROVE THE QUALITY AND CLARITY OF THIS WORK.

This figure compares the performance of different machine learning models such as Logistic Regression, Random Forest, Support Vector Machine, and XGBoost based on accuracy, precision, recall, and F1-score.

REFERENCES

- [1] S. Garera, N. Provos, M. Chew and A. D. Rubin, *A framework for detection and measurement of phishing attacks*, in Proceedings of the 2007 ACM Workshop on Recurring Malcode, 2007, pp. 1–8. [ResearchGate](#)
- [2] A. U. Rehman, I. Imtiaz, S. Javaid and M. Muslih, *Real-Time Phishing URL Detection Using Machine Learning*, *Security*, vol. 107, 2025. [MDPI](#)
- [3] D. T. Mosa, *Machine Learning Techniques for Detecting Phishing URL*, *Journal of Information Security and Applications*, 2023. [ScienceDirect](#)
- [4] S. H. Ahammad Shaikh et al., *Phishing URL detection using machine learning methods*, *Advances in Engineering Software*, 2022. [ResearchGate](#)
- [5] N. Nagy, *Phishing URLs Detection Using Sequential and Parallel Machine Learning Approaches*, *Sensors*, vol. 23, no. 7, 2023. [MDPI](#)
- [6] S. K. Kapan, *Improved Phishing Attack Detection with Machine Learning*, *Applied Sciences*, 2023. [MDPI](#)
- [7] A. Almujaheed et al., *Comparative evaluation of machine learning and deep learning algorithms for phishing detection*, *PMC Scientific Reports*, 2024. [PMC](#)
- [8] M. A. Tamal et al., *Dataset of suspicious phishing URL detection*, *Frontiers in Computer Science*, 2024. [Frontiers](#)
- [9] J. L. Wilk-Jakubowski, *Machine Learning and Neural Networks for Phishing Detection: A Comprehensive Review*, *Electronics*, 2025. [MDPI](#)

- [10]  A. UZ. Asif, *Machine Learning-Based Phishing Detection Using URL Features: A Systematic Survey*, Proc. *URL-Based ML Security*, 2023. [ACM Digital Library](#)
- [11]  R. Dubey, A. M. Tripathi, A. Srivastava and S. Singh, *Phishing Detection System: An Ensemble Approach Using Character-Level CNN and Feature Engineering*, arXiv preprint, Dec 2025. [arXiv](#)
- [12]  S. Baskota, *Phishing URL Detection using Bi-LSTM*, arXiv preprint, Apr 2025. [arXiv](#)
- [13]  S. Aslam *et al.*, *AntiPhishStack: LSTM-based Stacked Generalization for Optimized Phishing URL Detection*, arXiv preprint, 2024. [arXiv](#)
- [14]  A. Le, A. Markopoulou and M. Faloutsos, *PhishDef: URL Names Say It All*, arXiv preprint, 2010 — foundational work on lexical features for phishing detection. [arXiv](#)
- [15]  “Malicious URL Detection using Machine Learning: A Survey”, by D. Sahoo, C. Liu and S. C. H. Hoi, *ACM Computing Surveys*, 2017 — survey background on ML approaches.