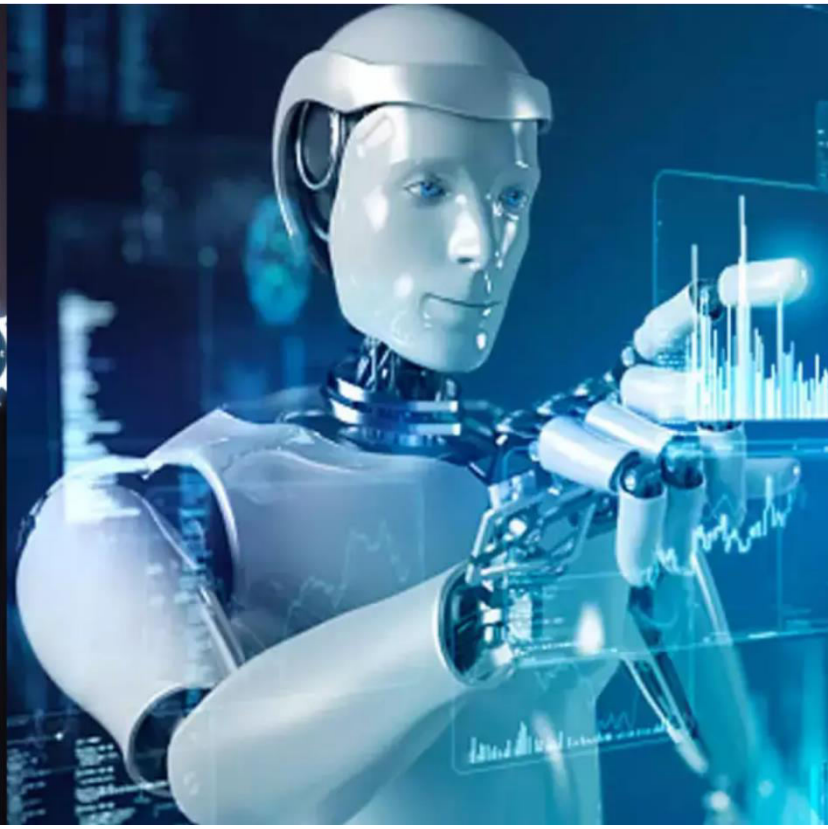
IJIRSET
International Journal of Innovative Research in Science | Engineering | Technology

ISSN(O): 2319-8753
ISSN(P): 2347-6710

International Journal of Innovative Research in Science Engineering and Technology (IJIRSET)

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



Impact Factor: 8.699

Volume 14, Issue 6, June 2025

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 14, Issue 6, June 2025

|DOI: 10.15680/IJIRSET.2025.1406081|

A Review on Phishing Detection using AI & Machine Learning

Prof. Pratiksha Prakash Pansare[1], Zolekar Avishkar Bharat [2], Thakre Prathmesh Sanjay[2], Lokhande Ritesh Motiram[2]

Assistant Professor, Department of Computer Engineering, Samarth College of Engineering and Management, Belhe, Junnar, Pune, Maharshtra, India[1]

Students, Department of Computer Engineering, Samarth College of Engineering and Management, Belhe, Junnar, Pune, Maharshtra, India[2]

ABSTRACT: Phishing attacks are among the most prevalent cyber threats today, exploiting users through deceptive websites to steal sensitive information. This project presents a client-side browser extension capable of detecting phishing URLs in real-time using a Random Forest machine learning model. A secure login interface and intuitive user dashboard are integrated to enhance user experience and system interaction. The model is trained on the UCI Phishing Websites Dataset, with feature extraction and classification implemented directly in JavaScript for rapid detection and privacy preservation. The system showed a 93% accuracy in identifying phishing sites. This tool offers users a lightweight, offline-capable solution for safer browsing.

KEYWORDS: Phishing Detection, Browser Extension, Random Forest, Client-Side Security, Web UI, Authentication

ABSTRACT

Background

The rise in online services has increased users' exposure to phishing threats. Cybercriminals create spoofed web pages to deceive users and steal sensitive data like login credentials and banking information.

Motivation

Current phishing detection systems are often server-dependent, which causes delays and can compromise privacy. A client-side tool can enhance speed and protect user data locally.

Problem Statement

How can we implement a real-time phishing detection system using only client-side features, while ensuring a simple and secure user experience?

I. INTRODUCTION

Phishing attacks are a growing cybersecurity threat, targeting users and organizations to steal sensitive data through deceptive websites. Traditional defense like blacklists and rule-based systems often fall short against increasingly sophisticated phishing techniques.

Machine learning offers a powerful solution by enabling systems to detect phishing through pattern recognition and feature analysis. By training models on datasets of legitimate and phishing sites, these systems can identify threats with greater accuracy and adaptability.

This project focuses on developing an intelligent phishing detection system using machine learning, aiming to improve

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 14, Issue 6, June 2025

|DOI: 10.15680/IJIRSET.2025.1406081|

detection rates, reduce false positives, and adapt to evolving attack strategies. By leveraging features like URL structure, content behavior, and code analysis, the model enhances threat identification and user protection.
Ultimately, this approach contributes to a smarter, more proactive cybersecurity framework—safeguarding both individual users and organizations in the digital age.

II. LITERATURE SURVEY

Title	Authors	Years	Methodology
Phishing URL detection using machine learning methods	SK Hasane Ahammad a , Sunil D. Kale b , Gopal D. Upadhye b , Sandeep Dwarkanath Pande c,* , E Venkatesh Babu a , Amol V. Dhumane b , Mr. Dilip Kumar Jang Bahadur d	2022	involves applying machine learning techniques to analyze and classify URLs to detect phishing attempts.
Phishing Detection and Prevention using Chrome Extension	M. Amir Syafiq Rohmat Rose; Nurlida Basir; Nur Fatin Nabila Rafie Heng; Nurzi Juana Mohd Zaizi; Madihah Mohd Saudi	2022	The methodology used in the proposed model involves the implementation of a self-destruct detection algorithm that employs supervised machine learning techniques, specifically focusing on URL-based web characteristics for phishing detection and prevention.
Phishing Site Detection Using Similarity of Website Structure.	Suleiman Y. Yerima, Mohammed K. Alzaylaee.	2021	It Proposes feature selection method are also used to increase the accuracy of classification model by selecting best feature & result .
An Intelligent System for Phishing Attack Detection and Prevention	N Megha, K R Remesh Babu Elizabeth Sherly	2020	Implementing a multi agent based architecture and ML classifier for detecting and rectifying web phishing attacks

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 14, Issue 6, June 2025

|DOI: 10.15680/IJIRSET.2025.1406081|

III. PROPOSED SYSTEM

The proposed system is a client-side phishing detection browser extension that provides real-time protection against phishing websites. Unlike traditional systems that rely on server-side processing or external APIs, this extension functions entirely within the browser, enhancing both speed and user privacy.

Key components of the proposed system include:

1. Client-Side Feature Extraction
The extension analyzes webpage characteristics such as URL patterns, presence of special characters, number of subdomains, HTTPS usage, and more — all without sending any data to external servers.

2. Machine Learning-Based Classifier (Random Forest)
A pre-trained Random Forest model (initially trained using Python and the UCI Phishing Dataset) is translated into JavaScript logic for integration into the browser. This model predicts whether the current site is legitimate or phishing in real time.

3. Integrated Login System
A login page restricts access to authenticated users, ensuring that only verified individuals can manage extension settings and logs. User authentication can be implemented using local storage or Firebase for real-time syncing.

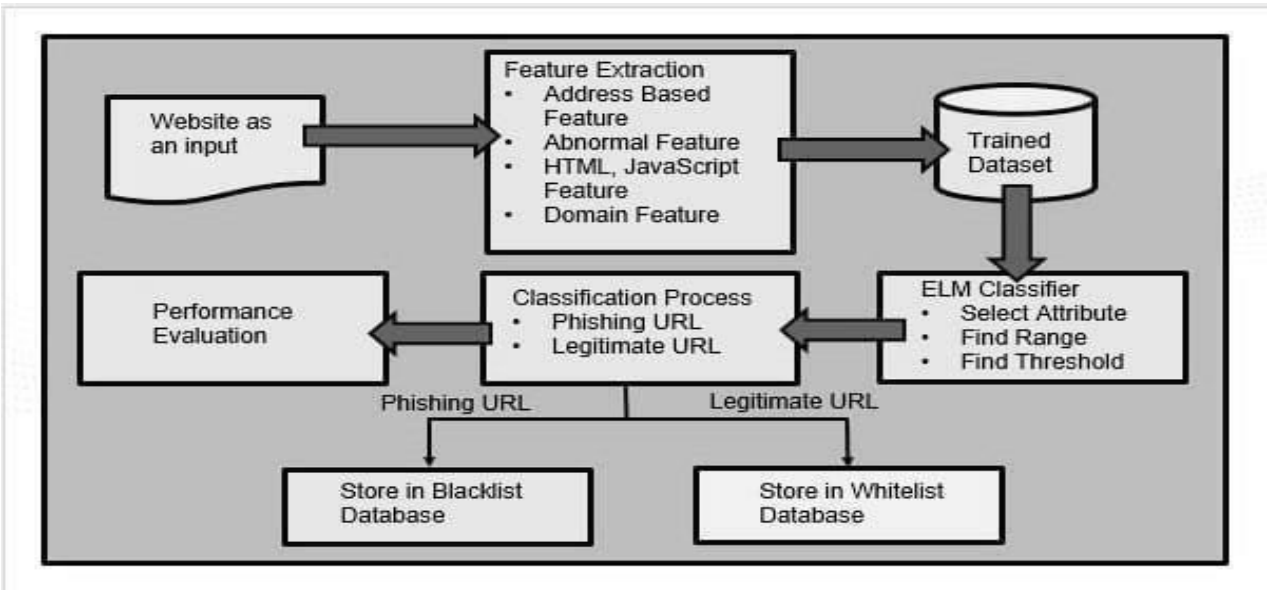
4. User Interface (UI)
A clean and simple user interface is provided as a popup in the browser. It displays:

Current website status (Phishing/Legit)

Option to report a suspicious URL

User logs and session data

5. Real-Time Alerting
As soon as a user navigates to a new page, the extension evaluates it and shows a warning if phishing is detected — before the user enters any credentials.III. SYSTEM REQUIREMENTS



International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 14, Issue 6, June 2025

|DOI: 10.15680/IJIRSET.2025.1406081|

Login Module: Authenticates users before enabling extension features

Feature Extractor: Pulls real-time website data (e.g., URL structure, HTTPS presence)

ML Classifier: JavaScript port of Random Forest model

UI Module: Displays status (Phishing/Legit), logs, and reports

Requirements

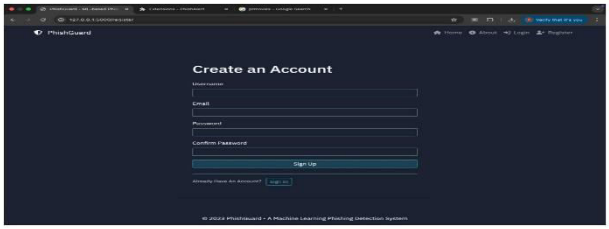
Hardware: Any device with a Chromium-based browser

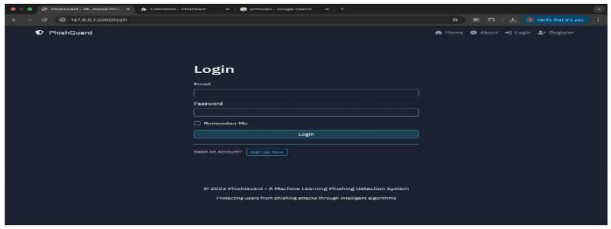
Software: HTML, CSS, JS, VS Code, Chrome, Python (for training)

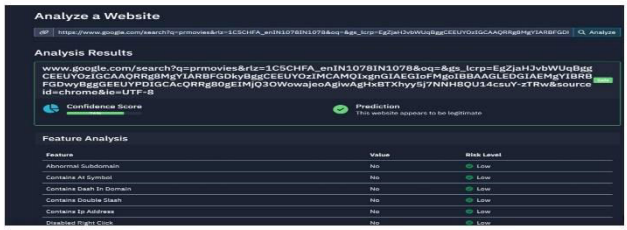
IV. SCOPE

- URL & Website Analysis: Detecting phishing through lexical, domain, and content-based features.
- Email Inspection with NLP: Leveraging email headers, metadata, and text analysis to spot phishing attempts.
- AI-Powered Classification: Using machine learning and deep learning to distinguish phishing from legitimate sources.
- Zero-Day Threat Detection: Real-time systems designed to catch previously unseen attacks.
- Behavioral & Intelligence-Based Defense: Analyzing user behavior and integrating threat intelligence for proactive protection.

V. RESULT







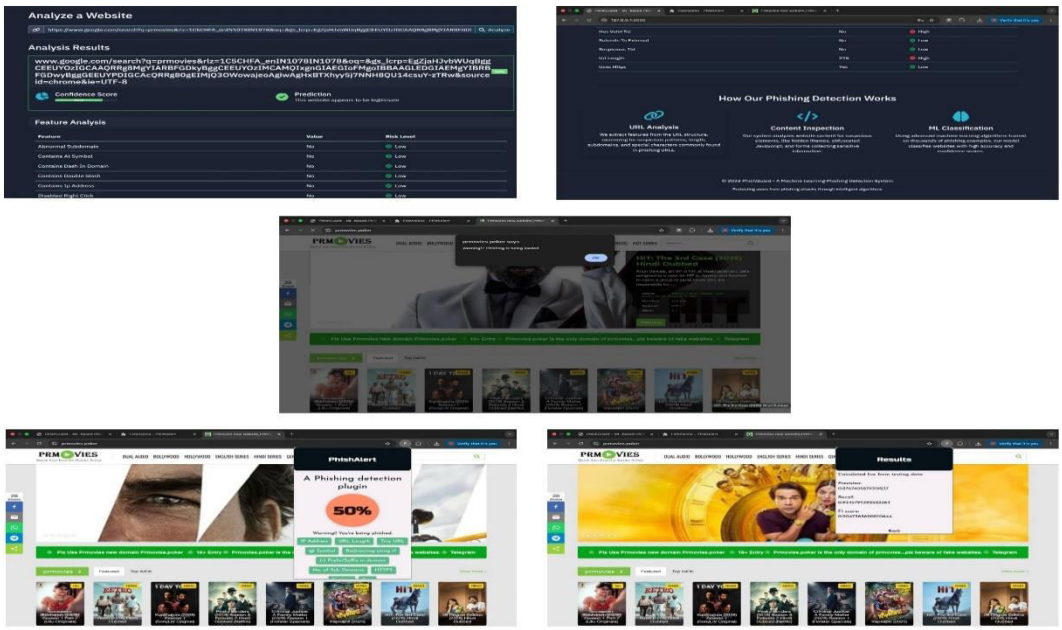
International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 14, Issue 6, June 2025

|DOI: 10.15680/IJIRSET.2025.1406081|



TESTING & EVALUATION

Testing Methods
Unit tests for feature extractor
Integration tests for UI-model connection
Manual testing with real phishing and legitimate URL
Test Cases
Valid phishing site → Alert
Legitimate site → No alert
Invalid login → Blocked access
Performance
Accuracy: ~93%
Speed: Instant feedback (<1 sec)
Lightweight: No internet/server dependency
Challenges
Fewer features available client-side
Managing model complexity in JS.

VI. CONCLUSION AND FUTURE IMPROVEMENT

Client-Side Phishing Detection Using Random Forest With the growing reliance on websites across domains like healthcare, business, and education, cyber threats—especially phishing—are increasingly prevalent. Phishing attacks trick users into revealing sensitive information through malicious websites. Traditional server-based detection systems often rely on features that require network access, compromising both speed and user privacy.

This work introduces a client-side phishing detection system using a Random Forest classifier implemented in JavaScript. It leverages only those webpage features that can be extracted locally on the user's browser, enabling real-time, private, and network-independent detection. While using fewer features may slightly reduce accuracy, it significantly improves speed and usability. The system uses the UCI Phishing Website dataset and demonstrates that lightweight, local analysis can still offer effective protection against phishing attacks.

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 14, Issue 6, June 2025

|DOI: 10.15680/IJIRSET.2025.1406081|

Future Work

Add server-sync for flagged websites
Explore deep learning alternatives (e.g., ELM, CNN)
Multi-browser support (Firefox, Edge

REFERENCES

1. Ahammad, S.K.H., Kale, S.D., Upadhye, G.D., Pande, S.D., Babu, E.V., Dhumane, A.V., Bahadur, M.D.K.J. (2022). "Phishing URL detection using machine learning methods."
2. Rohmat Rose, M.A.S., Basir, N., Heng, N.F.N.R., Zaizi, N.J.M., Saudi, M.M. (2022)."Phishing Detection and Prevention using Chrome Extension."
3. Yerima, S.Y., Alzaylaee, M.K. (2021). "Phishing Site Detection Using Similarity of WebsiteStructure."
4. Megha, N., Remesh, K.R., Babu, E.S. (2020). "An Intelligent System for Phishing AttackDetection and Preven
5. Alkawaz, M.H., Steven, S.J., Hajamydeen, A.I. (2020). "Detecting Phishing Website UsingMachine Learning."
6. Yerima, S.Y., Alzaylaee, M.K. (2020). "High Accuracy Phishing Detection Based onConvolutional Neural Network."
7. Satapathy, S.K., Mishra, S., Mallick, P.K., Badiginchala, L., Gudur, R.R., Guttha, S.C.(2019). "Phishing Detection Using Machine Learning."
8. Jain, A.K., Gupta, B.B. (2016). "A novel approach to protect against phishing
9. attacks atclient-side using auto-updated white-list."

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

INNO SPACE SJIF Scientific Journal Impact Factor

doi cross ref

निस्केयर NISCAIR

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

9940 572 462 6381 907 438 ijirset@gmail.com

www.ijirset.com

Scan to save the contact details