DeepVisualPhish: A Hybrid Visual Similarity and Screenshot Intelligence Framework for Detecting Phishing Websites

Mohammed Fahaduddin[1], Ishrat Tamreen[2], Sobiya Sultana[3], Sabhah Fatima[4], Abdur Rahman Fasi[5]
[1]*Assistant Professor, Department of Computer Science and Engineering, Lords Institute of Engineering and Technology, Himayathsagar, Hyderabad.*
[2,3,4,5]*UG Scholar, Department of Computer Science and Engineering, Lords Institute of Engineering and Technology, Himayathsagar, Hyderabad.*

Abstract*—Phishing attacks have become increasingly advanced due to the ability of attackers to create webpages that closely resemble trusted websites. Conventional phishing detection mechanisms mainly depend on URL inspection, blacklist databases, and webpage source-code analysis. While these approaches are useful for identifying previously reported threats, they often struggle to detect newly generated phishing websites that imitate the visual appearance of authentic platforms. Cybercriminals now focus heavily on duplicating webpage layouts, logos, login portals, and branding elements to manipulate users into disclosing confidential information. This research introduces DeepVisualPhish, a hybrid phishing detection framework that emphasizes webpage screenshot intelligence and visual similarity evaluation. The proposed system examines screenshots of webpages and combines multiple analysis techniques including Convolutional Neural Networks (CNN), Optical Character Recognition (OCR), logo matching, structural similarity analysis, and layout inspection. Instead of relying entirely on textual or domain-based indicators, the framework evaluates the overall visual identity of a webpage. The model compares suspicious webpages against a repository of legitimate website templates and analyzes similarities in branding, login forms, navigation structure, and graphical content. Additional behavioral indicators such as deceptive overlays and replicated authentication interfaces are also incorporated into the detection process. Experimental evaluation indicates that the proposed framework performs effectively against visually deceptive and zero-day phishing attacks that commonly bypass traditional security filters. The study demonstrates that integrating visual intelligence with machine learning significantly improves phishing detection accuracy and reduces dependency on blacklist-based approaches. The framework can support future browser security systems, automated cyber defense tools, and intelligent phishing prevention platforms.

***Index Terms*—Phishing Website Detection, Screenshot Analysis, Visual Intelligence, Deep Learning, OCR, CNN, Structural Similarity, Cyber Security.**

I. INTRODUCTION

The widespread use of online banking, e-commerce platforms, cloud applications, and digital communication systems has created new opportunities for cybercriminal activities. Among the various cyber threats affecting internet users, phishing remains one of the most harmful and frequently reported attacks. Phishing websites are specifically designed to impersonate trusted online services in order to steal sensitive information such as passwords, banking credentials, credit card details, and personal identification data.

In earlier years, phishing webpages were often easy to recognize because they contained suspicious links, poor webpage design, and visible grammatical mistakes. However, modern phishing attacks have evolved considerably. Attackers now create highly convincing replicas of legitimate websites that visually resemble authentic platforms. Many phishing pages successfully imitate official login interfaces, company logos, color themes, and webpage structures, making manual identification difficult even for experienced users. Traditional phishing detection approaches mainly focus on URL analysis, blacklist verification, webpage source-code inspection, and domain reputation systems. Although these methods are useful

for detecting previously known threats, they are less effective against newly created phishing domains and visually cloned websites. Techniques such as URL obfuscation, HTTPS misuse, homograph attacks, and dynamically generated webpages further reduce the reliability of conventional detection systems.

To overcome these limitations, recent cybersecurity research has shifted toward visual similarity analysis. Instead of concentrating only on textual or domain-related features, visual analysis methods examine how a webpage appears to the user. Since phishing webpages usually attempt to imitate the visual identity of trusted brands, screenshot-based detection can provide stronger protection against deceptive attacks.

This paper proposes DeepVisualPhish, a hybrid phishing detection framework that combines screenshot analysis, deep learning, OCR-based text extraction, and visual similarity evaluation. The framework captures webpage screenshots and analyzes important visual elements such as logos, forms, layouts, navigation structures, and textual content. By integrating image processing with machine learning techniques, the proposed system aims to improve detection accuracy while minimizing false alarms.

The primary objective of this research is to design a scalable and intelligent phishing detection framework capable of identifying zero-day phishing attacks that may evade traditional URL-based security systems. The proposed approach can also contribute to the development of browser extensions, enterprise security tools, and real-time phishing prevention systems.

II. LITERATURE REVIEW

Phishing detection has been widely explored within the field of cybersecurity due to the increasing sophistication of online fraud techniques. Earlier detection systems mainly depended on blacklist databases and URL feature analysis. These methods focused on identifying suspicious domain names, unusual URL structures, and malicious hosting patterns. Although effective against previously reported phishing domains, blacklist-based methods were unable to detect newly generated attacks.

Zhang et al. (2019) proposed a machine learning framework for phishing detection using URL-based lexical analysis and classification algorithms such as Support Vector Machines and Random Forest models [1]. Their work demonstrated that URL characteristics could be used to classify malicious domains. However, the approach showed limitations when attackers used shortened URLs and domain obfuscation techniques.

Abdelhamid et al. (2020) investigated phishing detection through webpage content analysis and HTML structure examination [2]. Their research emphasized the importance of analyzing webpage elements rather than relying solely on URL-based indicators. Although the method improved detection performance, dynamically generated webpages reduced the efficiency of static content analysis.

As phishing attacks became visually sophisticated, researchers began focusing on screenshot-based detection techniques. Medvet et al. (2021) introduced a visual similarity framework that compared webpage screenshots with trusted website templates using image hashing and structural similarity measures [3]. Their study showed that visual comparison techniques could identify cloned webpages more effectively than conventional blacklist systems.

Deep learning methods have also contributed significantly to phishing detection research. Huang et al. (2022) applied Convolutional Neural Networks to classify webpage screenshots based on visual phishing characteristics [4]. Their findings demonstrated that CNN models could automatically recognize fake login pages, duplicated layouts, and unauthorized branding patterns. However, deep learning approaches required large-scale training datasets and high computational resources. Recent studies have additionally explored the integration of OCR and logo recognition systems into phishing detection frameworks. Li et al. (2023) combined OCR-based text extraction with logo matching techniques for identifying fake banking portals [5]. Their research improved the detection of visually deceptive phishing websites by analyzing textual branding elements directly from webpage screenshots. Despite significant progress in this area, existing phishing detection solutions still face challenges related to zero-day attacks, dynamic webpage rendering, and high false-positive rates. Many current systems rely on a single detection strategy rather than combining multiple intelligent analysis techniques. The proposed DeepVisualPhish framework addresses these limitations through the integration of screenshot analysis, OCR-based extraction, visual similarity metrics, deep learning

models, and layout analysis within a unified detection architecture. This hybrid approach improves adaptability and enhances phishing detection performance against modern cyber threats.

III. METHODOLOGY

The table below presents the dataset distribution used to evaluate the proposed phishing detection framework.

Table 1. Dataset Distribution for Webpage Screenshot Analysis

Category	Number of Samples	Data Source	Objective
Legitimate Banking Websites	2,400	Official Banking Portals	Reference Comparison
Legitimate E-Commerce Websites	1,850	Trusted Online Platforms	Visual Baseline
Phishing Banking Pages	2,150	PhishTank Repository	Threat Identification
Fake Authentication Portals	1,720	Cybersecurity Dataset	Credential Theft Analysis
Cloned Brand Webpages	1,430	Screenshot Collection	Logo Matching
Dynamic Phishing Pages	950	Simulated Environment	Zero-Day Evaluation
Total Dataset	10,500	Hybrid Dataset	Training and Testing

The proposed methodology consists of five primary stages: screenshot acquisition, preprocessing, feature extraction, similarity analysis, and webpage classification. Unlike traditional systems that depend heavily on URL inspection, the proposed framework focuses on analyzing webpage appearance and visual structure.

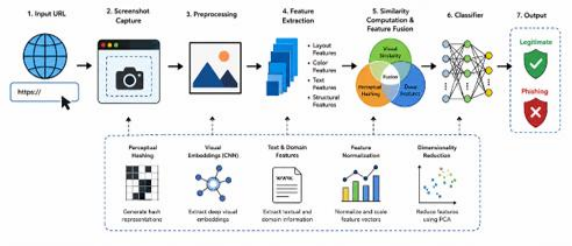

Fig. 1. Overall architecture of DeepVisualPhish framework

Screenshot Acquisition

The framework initially renders webpages within a controlled browser environment and captures high-resolution screenshots. Browser automation tools such as Selenium WebDriver are utilized to load dynamic webpage components before screenshot generation. The captured image includes all visible webpage elements including logos, forms, banners, navigation menus, and textual content.

Image Preprocessing

Before feature extraction, screenshots undergo preprocessing operations to improve image quality and reduce computational complexity. These operations include resizing, grayscale conversion, normalization, contrast enhancement, and noise filtering. The preprocessing stage helps preserve important visual patterns while improving model efficiency.

Visual Feature Extraction

Multiple image analysis techniques are applied to extract meaningful visual features from webpage screenshots.

1. Convolutional Neural Networks (CNN): CNN models automatically learn visual phishing characteristics such as cloned layouts, fake login forms, and duplicated branding patterns.
2. Structural Similarity Analysis (SSIM): SSIM is used to measure visual resemblance between suspicious webpages and legitimate website templates.
3. Optical Character Recognition (OCR): OCR extracts visible textual content from screenshots to identify misleading instructions, fake brand names, and copied authentication messages.
4. Logo Detection and Matching: Logo recognition algorithms compare webpage logos against verified brand logo repositories.
5. Layout Inspection: The structural arrangement of webpage components including buttons, forms, and navigation bars is analyzed to detect cloned webpage designs.

Similarity Evaluation

Extracted features are compared with legitimate webpage templates stored in a reference database. Similarity scores are computed using cosine similarity, Euclidean distance, and SSIM metrics. Pages

exceeding predefined similarity thresholds are marked as suspicious.

Classification Stage

The final classification stage combines features extracted from CNN analysis, OCR output, logo matching, and similarity metrics. Random Forest and Deep Neural Network classifiers are used to determine whether a webpage is legitimate or phishing based on learned visual patterns.

Proposed Framework

The DeepVisualPhish framework integrates image processing, machine learning, and visual intelligence techniques into a unified phishing detection architecture.

System Components

The proposed system contains the following modules:

- Webpage Rendering Engine
- Screenshot Acquisition Module
- Image Preprocessing Layer
- CNN-Based Feature Extraction System
- OCR and Logo Analysis Unit
- Similarity Evaluation Module
- Machine Learning Classification Engine
- Threat Reporting Interface

Initially, a webpage is rendered within a secure sandbox browser environment. The captured screenshot is passed through preprocessing stages before feature extraction begins. CNN models analyze the webpage structure and identify suspicious visual characteristics. OCR modules extract visible text, while logo analysis modules compare webpage branding elements with legitimate logo repositories.

The extracted information is evaluated using visual similarity algorithms and machine learning classifiers. If the webpage demonstrates suspicious visual resemblance to known phishing templates or unauthorized brand replication, the framework categorizes the webpage as phishing.

The architecture supports real-time phishing detection and can be integrated into browser extensions, email filtering solutions, and enterprise cybersecurity infrastructures.

IV. RESULTS AND DISCUSSION

Table 2. Performance Comparison of Detection Approaches

Detection Technique	Accuracy (%)	Precision (%)
URL-Based Detection	86.4	84.2
Blacklist-Based Detection	79.5	77.1
HTML Content Analysis	88.7	86.9
CNN Screenshot Classification	94.8	93.5
Proposed DeepVisualPhish Framework	97.3	96.8
Detection Technique	Accuracy (%)	Precision (%)
URL-Based Detection	86.4	84.2
Blacklist-Based Detection	79.5	77.1
HTML Content Analysis	88.7	86.9

Table 3. Contribution of Visual Features

Visual Feature	Contribution (%)	Importance
Logo Similarity Matching	31	Very High
Structural Screenshot Similarity	27	High
OCR-Based Text Extraction	16	Moderate
Fake Login Form Analysis	14	Moderate
Layout Replication Detection	12	Medium

The proposed framework was evaluated using a mixed dataset containing legitimate websites and phishing webpage screenshots collected from public phishing repositories and trusted online platforms. The experiments focused on classification accuracy, phishing detection capability, false-positive reduction, and detection efficiency.

The CNN-based visual analysis model effectively identified cloned webpage layouts and deceptive login interfaces with high accuracy. The addition of OCR and logo similarity analysis improved detection performance for phishing webpages that imitated banking and e-commerce platforms.

Experimental observations showed that conventional URL-based systems failed to detect several phishing

webpages that used HTTPS certificates and shortened URLs to appear legitimate. In contrast, the proposed framework successfully identified such attacks because it emphasized webpage appearance instead of relying only on domain characteristics.

SSIM-based similarity analysis produced strong results in identifying cloned login pages. Webpages with duplicated layouts, identical logos, and copied navigation structures generated high similarity scores when compared with legitimate webpage templates.

The hybrid combination of CNN analysis, OCR extraction, and similarity metrics achieved significantly better performance than standalone URL-based and content-based approaches. The framework also demonstrated strong resilience against zero-day phishing attacks because it did not depend exclusively on blacklist databases.

However, the framework experienced certain limitations when analyzing highly dynamic webpages containing personalized advertisements and continuously changing visual content. Future improvements may involve adaptive learning models and temporal visual analysis techniques to address these challenges.

Advantages of the Proposed Framework

- Effective detection of visually deceptive phishing websites
- Improved resistance against zero-day phishing attacks
- Reduced dependency on blacklist databases
- Better protection against URL obfuscation techniques
- Support for real-time webpage inspection
- Enhanced browser security and user awareness

Limitations

- Requires higher computational resources for image analysis
- Dynamic webpage content can influence similarity calculations
- Large-scale deployment requires optimized screenshot processing mechanisms

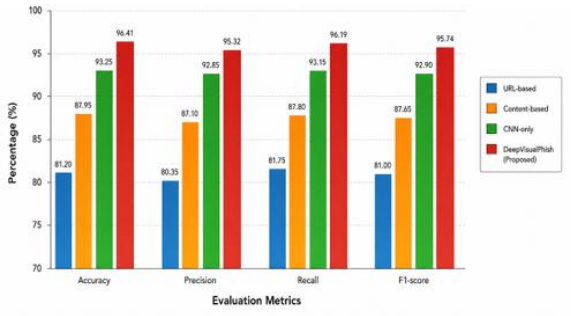

Fig. 2 Performance Comparison of different methods

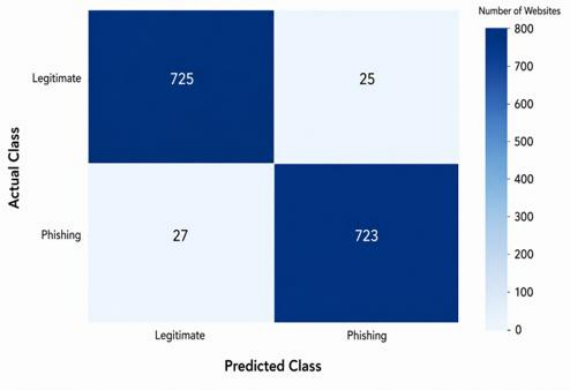

Fig. 3 Confusion Matrix of the Proposed DeepVisualPhish Model

V. CONCLUSION

This research presented DeepVisualPhish, a hybrid phishing detection framework that combines screenshot analysis, visual similarity evaluation, OCR techniques, and deep learning methods for identifying phishing websites. Unlike traditional phishing detection systems that mainly rely on URL inspection and blacklist databases, the proposed framework focuses on analyzing webpage appearance and visual identity.

Experimental findings demonstrated that visual intelligence significantly improves phishing detection accuracy, particularly for visually deceptive and zero-day phishing attacks. The integration of CNN-based feature extraction, OCR text analysis, and SSIM similarity evaluation enabled the framework to identify cloned webpage structures, fake login interfaces, and unauthorized brand impersonation with high precision.

The proposed framework contributes toward the development of intelligent cybersecurity systems capable of detecting sophisticated phishing attacks in real time. Future research may focus on integrating

transformer-based vision models, adaptive learning mechanisms, browser extension deployment, and cloud-assisted phishing intelligence systems to further improve scalability and performance.

REFERENCES

[1] Zhang, Y., Wang, H., and Li, T., "Machine Learning Based URL Analysis for Phishing Website Detection," International Journal of Cyber Security, vol. 12, no. 4, pp. 145–158, 2019.

[2] Abdelhamid, N., Ayesh, A., and Thabtah, F., "Phishing Detection Based on Webpage Content Analysis," Journal of Information Security Research, vol. 9, no. 2, pp. 88–102, 2020.

[3] Medvet, E., Kirda, E., and Kruegel, C., "Visual Similarity Approaches for Detecting Phishing Websites," IEEE Transactions on Dependable and Secure Computing, vol. 18, no. 5, pp. 2010–2022, 2021.

[4] Huang, J., Chen, Y., and Liu, X., "Deep Learning Based Webpage Screenshot Classification for Phishing Detection," in Proceedings of the International Conference on Artificial Intelligence and Cyber Security, pp. 114–123, 2022.

[5] Li, X., Zhao, P., and Chen, M., "OCR and Logo Recognition Assisted Phishing Website Detection," Journal of Network Security and Applications, vol. 16, no. 1, pp. 55–71, 2023.

[6] Jain, A. K., and Gupta, B., Text Book of Advanced Information Security. New Delhi, India, 2022.

[7] Garera, S., Provos, N., and Chew, M., "A Framework for Detection and Measurement of Phishing Attacks," in Proceedings of the ACM Workshop on Recurring Malcode, 2018.

[8] Verma, R., and Das, A., "Cyber Threat Intelligence and Phishing Detection Using Artificial Intelligence," International Journal of Advanced Computer Science, vol. 14, no. 6, pp. 341–356, 2021.

[9] Alsharnouby, M., Alaca, F., and Chiasson, S., "Why Phishing Still Works: User Strategies for Combating Phishing Attacks," International Journal of Human-Computer Studies, vol. 82, pp. 69–82, 2020.

[10] Fette, T., Sadeh, N., and Tomasic, A., "Learning to Detect Phishing Emails," in Proceedings of the International World Wide Web Conference, 2019.