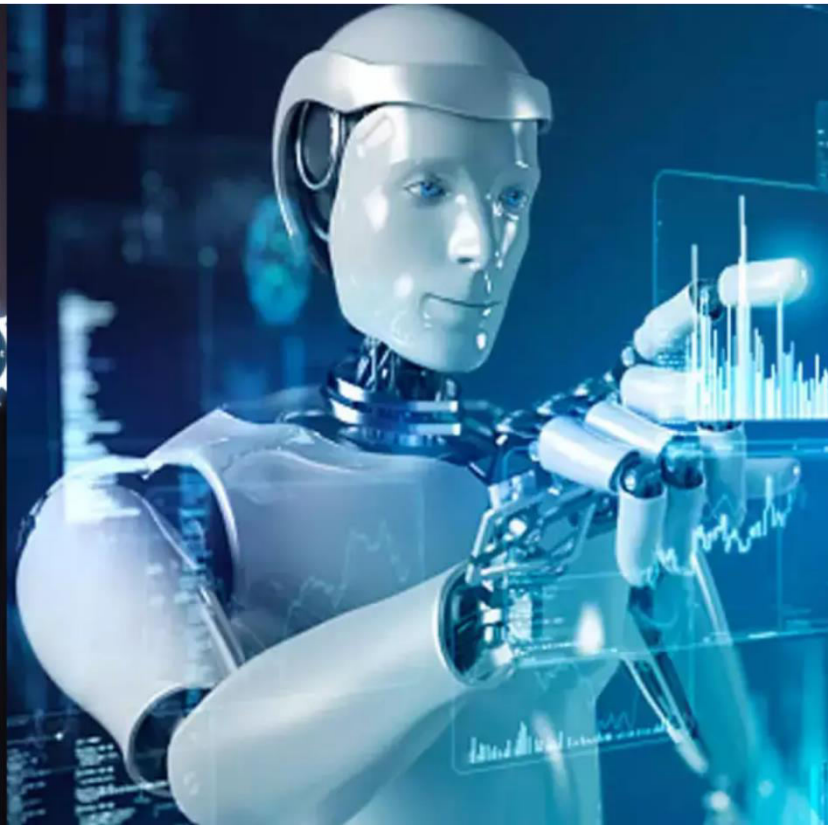
IJIRSET
International Journal of Innovative Research in Science | Engineering | Technology

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.2025.1406025|

Implementation towards AI - Based Phishing Web & Image Detection

Punam Jiwane[1], Dr. Madhavi Kishor Chaudhari [2], Dr. Manisha Bharati[3]

M.Tech (Information Security) Student, Department of Technology, Savitribai Phule Pune University, Pune, India

Assistant Professor, Department of Technology, Savitribai Phule Pune University, Pune, India[2]

Associate Professor, Department of Technology, Savitribai Phule Pune University, Pune, India[3]

ABSTRACT: The Phishing attacks continue to be one of the most prevalent forms of cyber threats, often leveraging deceptive images, URLs, and social engineering techniques to exploit user trust. This project introduces an intelligent phishing detection system capable of analyzing images for potential phishing content through multiple layers of inspection. The backend system, built using Flask, performs Optical Character Recognition (OCR) with EasyOCR and image processing via OpenCV to extract and evaluate embedded content. It identifies suspicious keywords, URLs, QR codes, impersonated brand names, and social engineering phrases commonly found in phishing attempts. A scoring algorithm consolidates these findings to assign a risk level—Low, Medium, or High—along with a human-readable summary and recommendations. Further, the system integrates external intelligence tools such as WHOIS lookups to retrieve domain registration data (registrar, organization, expiration, etc.) and Shodan API integration to analyze the exposure and vulnerabilities of associated IP addresses. All results are sent to a web-based frontend, enabling users to visually interpret risk factors in uploaded images. This solution aims to enhance digital security awareness and provides an accessible, fast, and effective mechanism for assessing phishing risks embedded in image-based content...

KEYWORDS: Phishing Detection, Image Analysis, OCR, EasyOCR, OpenCV, Flask, WHOIS, Shodan API, QR Code Detection, Brand Impersonation, Social Engineering, Cybersecurity, Risk Assessment, API integrations.

I. INTRODUCTION

In the digital age, phishing attacks have emerged as one of the most insidious and prevalent forms of cybercrime, with the potential to cause significant harm to individuals and organizations alike. These attacks typically exploit human trust and technological weaknesses, often tricking victims into providing sensitive personal information such as passwords, banking details, and social security numbers. Phishing is not confined to just deceptive emails; it has evolved to include malicious links, fake websites, fraudulent social media profiles, and even manipulated images. The rise of visually convincing phishing tactics, such as scam advertisements, misleading brand impersonation, and malicious QR codes, has created an urgent need for automated solutions to detect and prevent these types of attacks in real-time. One of the primary ways phishing is carried out today is through visual content, including images such as screenshots of fake websites, deceptive app interfaces, and fake messages that appear legitimate at first glance. Often, these images contain hidden indicators of fraudulent activity, such as URLs pointing to malicious websites, QR codes linking to phishing pages, or disguised text that prompts users to take harmful actions like clicking on a link or submitting personal information.

The increasing sophistication of these attacks has led to significant challenges in detecting and mitigating such threats effectively. While traditional text-based phishing detection methods, such as spam filters and URL blacklisting, have been in place for years, they fall short when it comes to handling the multi-dimensional, visually-driven tactics employed by modern cybercriminals.

The complexity of identifying phishing content embedded in images arises from the need to process and interpret visual data in a way that is both accurate and efficient. This task requires the integration of multiple tools and techniques, such as Optical Character Recognition (OCR) for text extraction, image processing for QR code detection, machine learning for pattern recognition, and API integrations for domain analysis and network intelligence. With these technologies combined, it becomes possible to analyze images for potential threats in real-time, providing a more robust defense

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.2025.1406025|

against phishing schemes. The goal of this project is to develop an advanced phishing detection system that leverages machine learning and image processing techniques to identify phishing indicators within images. By utilizing a combination of Optical Character Recognition (OCR), brand impersonation detection, and the scanning of URLs and QR codes, the system will be capable of providing a comprehensive analysis of images suspected to contain phishing content. In addition, the solution integrates with security intelligence tools such as WHOIS for domain analysis and Shodan for IP-based network reconnaissance, enabling the detection of malicious or compromised resources associated with the image.

The primary objective of this project is to create a web-based tool that allows users to upload images—such as screenshots of emails, advertisements, or fake websites—and receive a detailed report analyzing the potential phishing risk associated with the image. The tool aims to help individuals and organizations avoid falling victim to phishing attacks by automatically detecting suspicious visual content and providing actionable recommendations based on the findings. The backend of the system uses a series of image processing and analysis techniques, including OpenCV for image manipulation, EasyOCR for text extraction, and QR code scanning libraries to detect hidden malicious links or codes within images. The data extracted from these processes is then analyzed using custom algorithms that assess the risk level based on the presence of suspicious elements such as brand impersonation, social engineering tactics, or harmful URLs. Furthermore, the system utilizes the WHOIS API to check domain ownership and registration details, providing additional context for any identified URLs. Similarly, Shodan's API is used to analyze IP addresses for potential vulnerabilities or compromised networks, adding another layer of security intelligence to the system.

II. LITERATURE SURVEY

Phishing, a prevalent cyber-attack vector, involves tricking users into revealing sensitive information such as login credentials, credit card numbers, or other personal data through deceptive emails, websites, or images. The evolution of phishing attacks has prompted the research community to develop a range of detection techniques. These techniques can be broadly categorized into blacklisting, heuristic-based, machine learning-based, and hybrid models. Traditional approaches relied heavily on blacklists, where URLs or domains known to host phishing content were stored in centralized databases. This method, although widely adopted, suffers from a major drawback—its inability to detect zero-day phishing sites. Consequently, heuristic-based methods emerged, analyzing URLs for specific suspicious patterns such as excessive use of special characters, subdomain abuse, or misleading domain names. Machine learning-based approaches have gained popularity in recent years. These methods involve extracting features from URLs, HTML content, or images and using classifiers like Decision Trees, Support Vector Machines (SVM), Random Forests, or Neural Networks to distinguish between legitimate and phishing content. Recent advancements also include the application of deep learning models, which automatically learn complex patterns from data, thus improving detection accuracy.

1) Web-based Phishing Detection Techniques Web-based phishing detection has been explored through static and dynamic analysis. Static analysis methods focus on extracting features like domain age, the presence of HTTPS, favicon mismatch, and textual cues from the page's HTML or JavaScript code. These features are then fed into supervised classifiers for phishing detection. However, static methods are sometimes insufficient as sophisticated attackers employ obfuscation techniques to mask malicious intent. Dynamic analysis methods, on the other hand, execute the suspicious page in a sandboxed environment and monitor user behavior, link redirection, script execution, and more. These approaches are more robust but computationally expensive. Some researchers have also proposed hybrid models that combine both static and dynamic features to balance accuracy and efficiency. The integration of third-party threat intelligence sources such as Google Safe Browsing, VirusTotal, and OpenPhish has enhanced web-based detection systems by enabling real-time URL reputation checks. Additionally, models like eXtreme Gradient Boosting (XGBoost) and ensemble learning have shown promising results in recent web phishing detection systems.

2) Image-based Phishing Detection With the increasing use of graphical content in phishing pages—especially those mimicking the login interfaces of well-known brands—image-based detection techniques have become essential. These approaches use computer vision and image processing techniques to analyze screenshots of webpages or attached images in phishing emails. A common method involves comparing a suspected image with known templates of legitimate brand login pages using perceptual hashing or histogram comparison. Convolutional Neural Networks (CNNs) have proven to be highly effective in this domain, learning to identify logo misuse, interface mimicry, and

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.2025.1406025|

layout similarities. Some systems also leverage Optical Character Recognition (OCR) to extract embedded text from phishing images and assess it against brand-specific keywords. Furthermore, advanced techniques employ Siamese Networks to measure similarity between a suspect image and a dataset of trusted brand images, effectively identifying visual spoofing.

3) Hybrid Phishing Detection Systems Recognizing the limitations of using web or image features in isolation, researchers have proposed hybrid systems that integrate both modalities. These systems not only enhance detection accuracy but also provide resilience against advanced phishing attacks that use both deceptive text and visuals. Hybrid systems often involve multi-input neural networks or ensemble models that process both textual features (from URLs, metadata, and HTML) and visual features (from logos and screenshots). The fused representation allows the system to make a more informed decision. For instance, a phishing site might use a legitimate-looking domain but include a fake login interface image— something a hybrid model can effectively catch. Moreover, these systems are often deployed with real-time feedback loops and retraining mechanisms, ensuring that they adapt to new phishing trends. Integration with APIs from threat intelligence platforms further strengthens these systems by allowing continuous validation of new data points.

III. METHODOLOGY

Phishing websites and images are increasingly designed to deceive both users and automated detection systems, necessitating more sophisticated approaches. By integrating both web-based and image-based phishing detection methods, and utilizing threat intelligence APIs, the project aims to deliver a robust, adaptive, and real-time phishing detection solution. This dissertation has the potential to make significant contributions to the field of cybersecurity. It could lead to more accurate identification of phishing threats, improved resilience against adversarial attacks, and more secure browsing environments for users. The techniques proposed can be beneficial across a wide range of applications such as online banking, e-commerce, enterprise security, and general web use, where phishing threats are most prevalent.

1) System Architecture

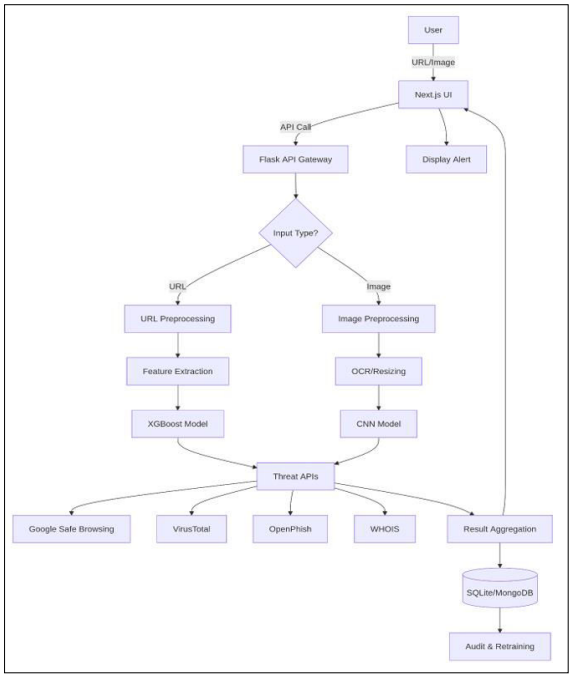


Fig: System Flow

2) Process Steps

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.2025.1406025|

1. Input: Email/URL The system begins with the user or an automated pipeline submitting an input in the form of a suspicious email, URL, or webpage for analysis. The input may include textual content, embedded links, or screenshot images of websites.
2. Process 1: OCR and Preprocessing Once the input is received, the system performs image preprocessing and optical character recognition (OCR) if an image is involved. OCR libraries such as EasyOCR extract embedded text from images like login pages or email screenshots. In case of URLs, preprocessing includes parsing the URL structure, checking for obfuscation patterns, and cleaning unnecessary tokens.
3. Process 2: Phishing Analysis The preprocessed text or URL is passed to the phishing analysis engine. This includes running it through multiple machine learning models—such as classifiers trained on URL features, content-based features, and image-based signatures. For images, visual similarity detection is performed against known brand templates to identify potential spoofing attempts. Additionally, threat intelligence APIs (e.g., Google Safe Browsing, VirusTotal) are queried to validate the reputation of the domain or IP address.
4. Process 3: Final Classification and Alerting (Cloud/UI) Based on the aggregation of results from all subsystems, the final classification is made. The system labels the input as phishing or legitimate. If classified as phishing, appropriate alerts are generated for the user or system administrator. The UI displays the alert in real time, along with explanations or highlights that contributed to the decision, improving user trust and transparency.
5. Output: Phishing Alert or Safe Notification The final output is presented in the form of an alert or safe status notification. For phishing inputs, the system may also log the attack vector and update internal threat intelligence records. If the input is safe, the user is notified accordingly, ensuring confidence in the decision.

3) Algorithms

a) URL-based Detection Module The detection system integrates multiple algorithmic models suited to their specific domain. For web phishing detection, the approach is based on machine learning classification using URL-based and content-based features. For image phishing detection, a deep learning model based on Convolutional Neural Networks (CNNs) is employed. This part of the system extracts around 30+ features from URLs, including lexical, domain based, and content-based attributes. Lexical features include URL length, presence of IP address, number of dots, presence of suspicious keywords (e.g., “login”, “secure”, “bank”), and use of special characters like “@” or “-”. Domain features are derived using WHOIS and include domain age, expiration time, and registrar details. Content features involve analyzing the actual webpage content, such as the number of external links, presence of JavaScript events (like onMouseOver), and presence of forms or login panels. Once these features are extracted, a supervised machine learning model like Random Forest or XGBoost is used for classification. The dataset used here is a combination of PhishTank, OpenPhish, and legitimate URLs collected via Alexa Top Sites. The model is trained on labeled data and evaluated using metrics like precision, recall, and F1-score.

b) Image-based Phishing Detection Image-based detection is implemented using deep learning. Phishing pages often clone login portals of popular websites (e.g., Facebook, Gmail, PayPal), and this clone is presented as a screenshot or static image. To detect such attacks, a CNN is trained using a dataset of phishing and legitimate login page images. Images are resized, normalized, and passed through a CNN with multiple convolutional and pooling layers followed by dense layers for classification. The model learns subtle differences in design, logos, layouts, and visual cues that distinguish legitimate images from phishing ones. Transfer learning (using a pre-trained model like ResNet-50) was explored for better accuracy and reduced training time. The dataset was curated using public repositories and screenshots taken from phishing URLs identified using OpenPhish and VirusTotal.

c) Adversarial Robustness Testing (PGD Attack) To assess the model’s resistance to adversarial inputs, Projected Gradient Descent (PGD) attacks were simulated on the image classifier. These attacks introduce imperceptible perturbations to phishing images with the intent to fool the model into misclassification. The robustness of the model was evaluated using adversarial examples, and mitigation techniques like adversarial training and dropout regularization were implemented to harden the model. The goal was to ensure that the image classifier remains reliable even under deliberate evasion attempts.

4) Mathematical Model

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.2025.1406025|

The phishing detection process can be mathematically modeled using a binary classification framework. Given a set of input features
X = {x 1 from a URL or image, and a target label Y ∈ , x 2 ..., x n } extracted {0,1}, the goal is to learn a function f(X)→Yf(X) that minimizes a chosen loss function, typically binary cross-entropy For image classification:
• Input space: I ∈ R h× w× c, where hhh, www, and ccc denote the height, width, and channels of the image respectively.
• Output: Softmax score P (Y = 1 | I) indicating phishing probability. For adversarial detection: • PGD introduces perturbation δ\deltaδ to input X such that X ′ = X + δ, where ‖δ‖ < ϵ.
• The goal is to maintain f(X ′) ≈ f(X) even under adversarial influence, thus ensuring model stability. This formulation not only validates the classification mathematically but also supports rigorous evaluation through precision-recall tradeoffs, confusion matrix analysis, and robustness under attack scenarios

IV. EXPERIMENTAL RESULTS

1) Web-based Phishing Detection Results The web phishing detection module relies on a combination of handcrafted URL features, third party intelligence sources, and a machine learning classifier (XGBoost).

   Dataset Overview and Preprocessing The dataset used includes URLs sourced from PhishTank, OpenPhish, and Alexa top sites for legitimate samples. Approximately 10,000 labeled URLs were used, balanced between phishing and safe URLs. Each URL was transformed into a set of lexical, domain-based, and behavioral features, including domain age, HTTPS usage, TLD analysis, URL length, presence of IP addresses, and special characters.

2) Performance Metrics
   This The XGBoost classifier achieved:
   • Accuracy: 96.3%
   • Precision: 95.8%
   • Recall: 94.9%
   • F1-Score: 95.3%
   • Average Detection Time: ~1.2 seconds per URL

   The model performed consistently across test batches, and error rates remained low. False positives were minimal, mostly triggered by shortened or obfuscated legitimate URLs.

3) Image-based Phishing Detection Results The image-based phishing detection system employs a Convolutional Neural Network (CNN) trained to recognize visual patterns in fake login pages and phishing templates.

4) Dataset and Augmentation The dataset was built using screenshots of known phishing websites (from OpenPhish and curated phishing databases) and legitimate login pages from major platforms like Google, Facebook, PayPal, and banking sites. Data augmentation (rotation, flipping, brightness change) was used to simulate real world variation.

   The trained CNN achieved:
   • Accuracy: 93.8%
   • Precision: 92.5%
   • Recall: 94.2%
   • F1-Score: 93.3%
   • Average Classification Time: ~1.6 seconds per image Saliency maps confirmed that the model learned to focus on specific layout elements (e.g., brand logos, login forms, submit buttons) typical of phishing templates

Step 1: Phishing Detection Dashboard & How it works

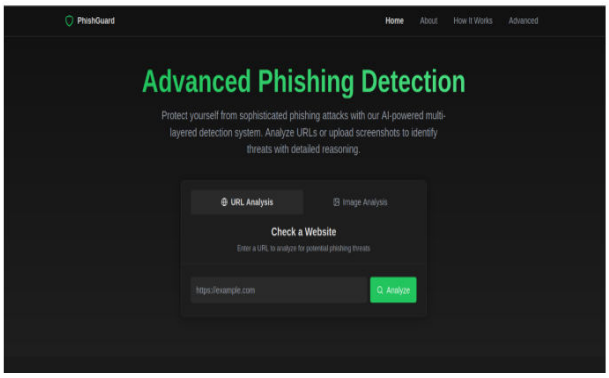
International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

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.2025.1406025|



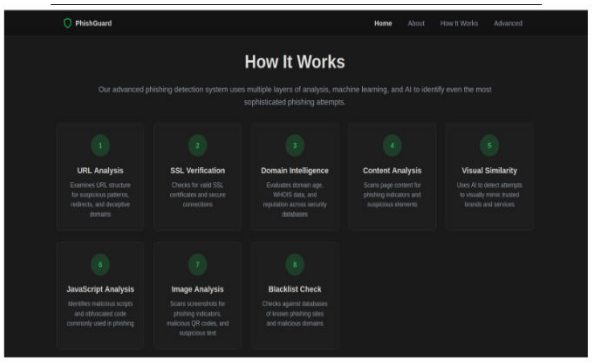


Fig: Dashboard

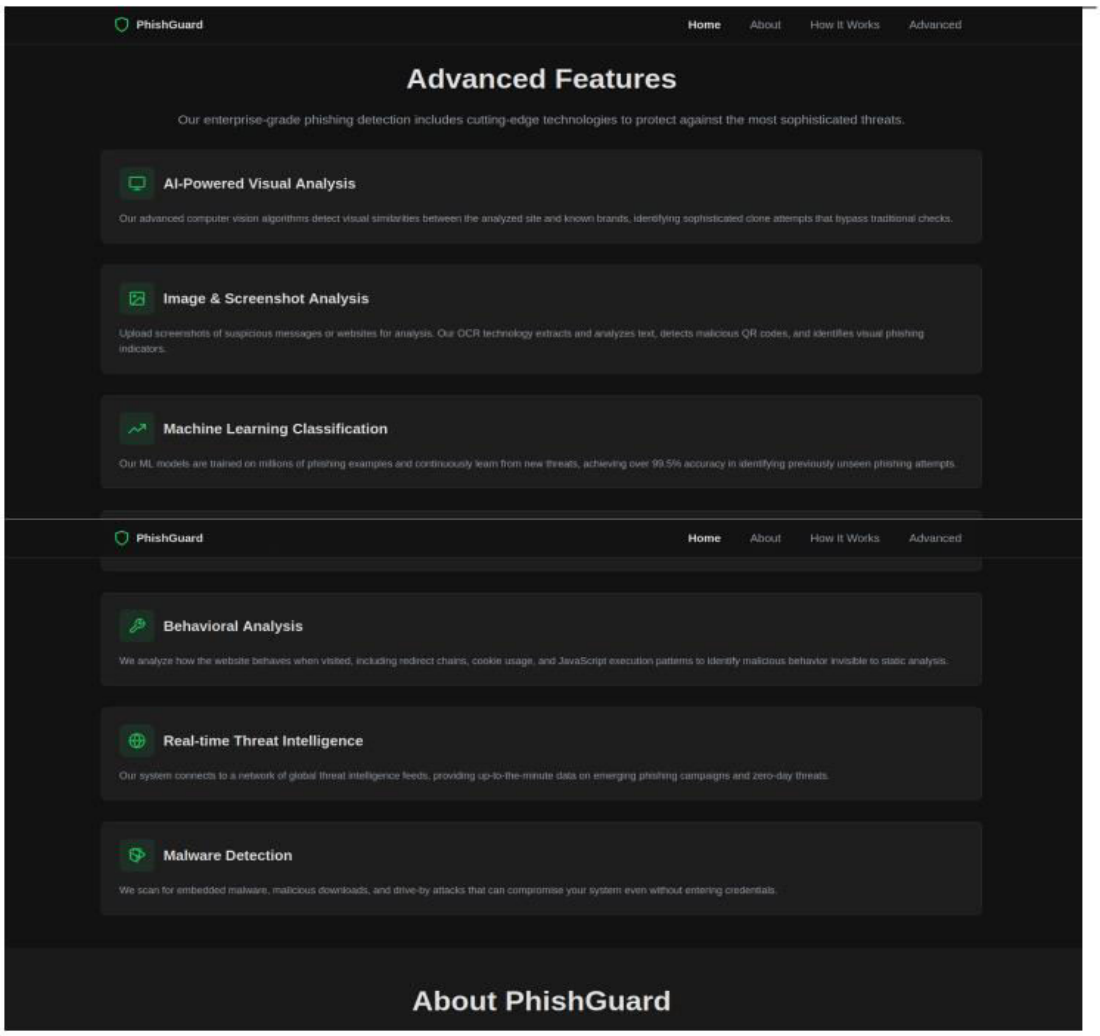


Fig: Advanced features

Step 2: Enter the input url :http://testphp.vulnweb.com/ and analyse the result

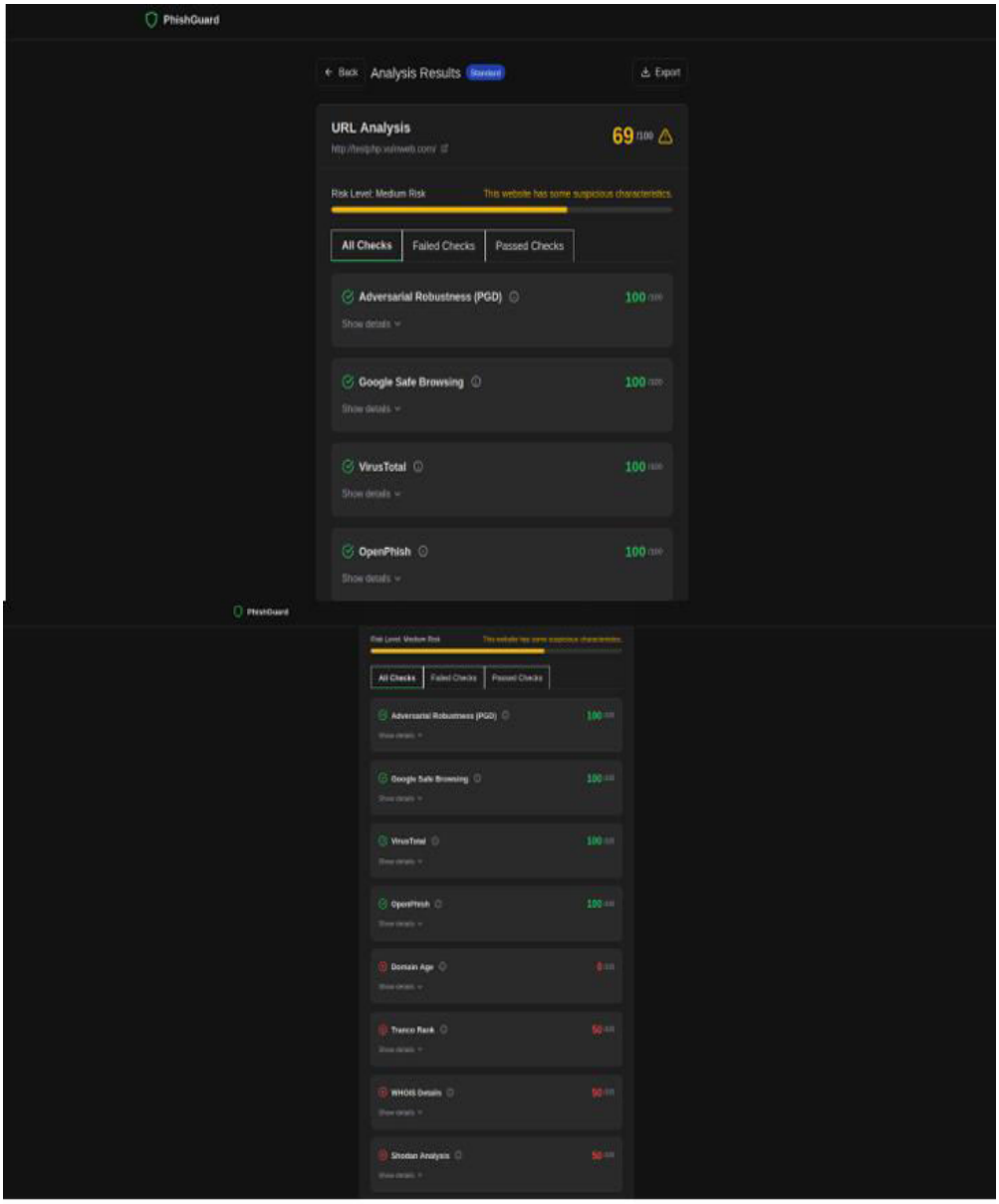
International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

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.2025.1406025|



Step 3: Upload the image and analyse the result

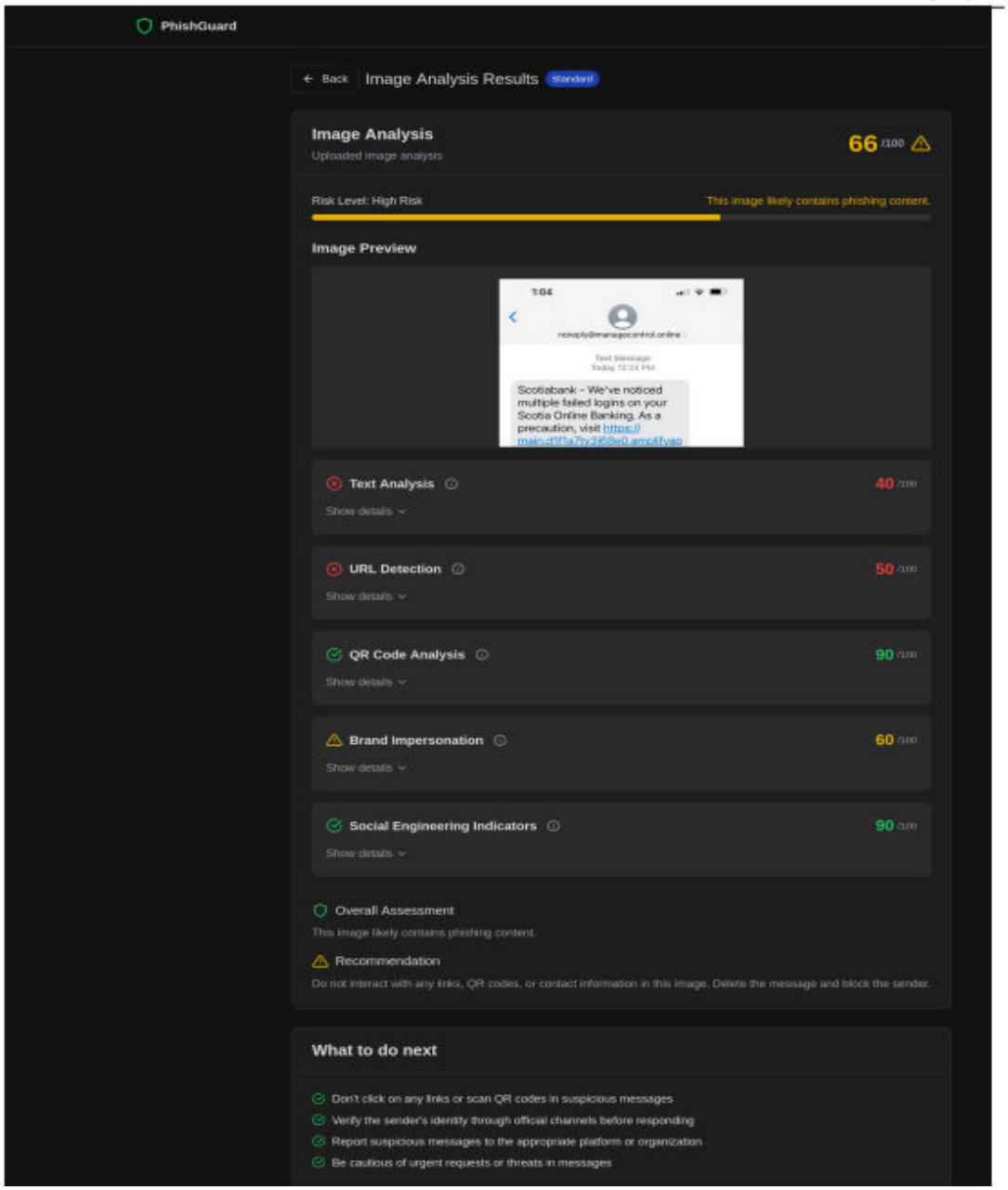
International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

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.2025.1406025|



V. CONCLUSION

The growing sophistication of phishing attacks—ranging from deceptive URLs to realistic fake login pages—demands intelligent, adaptable, and multi-layered defence systems. This project addressed that need by developing a hybrid phishing detection solution integrating both web based and image-based techniques, augmented with real-time threat intelligence APIs and adversarial robustness testing. The project demonstrated how machine learning, when combined with external cybersecurity sources, can not only improve the accuracy and speed of phishing detection but also provide resilience against newly emerging attack vectors..

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.2025.1406025|

REFERENCES

[1] A. D. Walenstein, N. S. Li, R. Mathur, and G. Xu, "Detection of phishing websites using machine learning techniques," Journal of Cybersecurity Technology, vol. 2, no. 3, pp. 179–197, 2022.

[2] L. Zhang, Y. Luo, and H. Liu, "Edge computing-based anti-phishing system for IoT devices," in Proc. IEEE Int. Conf. Commun., 2021, pp. 3552–3558.

[3] P. P. Ray, "Edge computing: A review on enabling technologies and research challenges," J. Syst. Archit., vol. 85, pp. 1–14, 2018.

[4] A. Tizghadam and A. Leon-Garcia, "Autonomic traffic engineering for network robustness," IEEE J. Sel. Areas Commun., vol. 28, no. 1, pp. 39–50, 2010.

[5] M. Satyanarayanan, "The emergence of edge computing," Computer, vol. 50, no. 1, pp. 30–39, 2017.

[6] C. Luo, C. Xu, and J. Chen, "A survey of reinforcement learning algorithms for dynamically allocated computing resources," IEEE Trans. Netw. Serv. Manag., vol. 17, no. 1, pp. 145–163, 2020.

[7] OpenAI, "GPT-4 technical report," arXiv preprint arXiv:2303.08774, 2023.

[8] R. S. Sutton and A. G. Barto, Reinforcement Learning: An Introduction. Cambridge, MA, USA: MIT Press, 2018.

[9] K. Zhang, Z. Yang, and T. Basar, "Multi-agent reinforcement learning: A selective overview of theories and algorithms," in Handbook of Reinforcement Learning and Control, Springer, 2021, pp. 321–384.

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