IJIRCCE

e-ISSN: 2320-9801 | p-ISSN: 2320-9798

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

Volume 12, Issue 7, July 2024

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

Impact Factor: 8.379

9940 572 462 | 6381 907 438 | ijircce@gmail.com | www.ijircce.com

International Journal of Innovative Research in Computer and Communication Engineering

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 |

|| Volume 12, Issue 7, July 2024 ||

| DOI: 10.15680/IJIRCCE.2023.1106112 |

AI-Driven Detection of Phishing Attacks through Multimodal Analysis of Content and Design

Shoeb Ali Syed

University of the Cumberlands, Kentucky, United States of America

ABSTRACT: Machine operators continue facing dangerous phishing attacks through text manipulation and deceptive visual design schemes that trick users. The prevalent detection procedures of traditional systems depend on static rule-based filters and content-only analysis because these methods fail to keep pace with contemporary phishing methods. The suggested AI-based solution for phishing detection utilizes a dual method that examines written content while assessing webpage designs to achieve better results and flexible system operations.

The research employs a mix of legitimate and phishing emails and webpages to implement NLP text features alongside CNN network analyses of visual design elements. The system evaluates linguistic urgency signals, deception markers, and design features, including layout inconsistency, corrupted logos, and color slips.

Multimodal analysis outclasses single-modality assessment in experimental tests because it produces superior detection results with higher precision and recall metrics, and F1 scores against phishing attempts. The model can detect new phishing samples that it has not encountered before testing.

The study demonstrates the value of semantic and visual data in cybersecurity systems because dynamic and intelligent models are required to fight advanced phishing attacks. This proposed framework provides operational benefits for current phishing defense systems operating in email platforms, web browsers, and enterprise security systems.

KEYWORDS: Phishing Detection, Artificial Intelligence, Multimodal Analysis, Cybersecurity, Machine Learning, Content Analysis, UI/UX Detection

I. INTRODUCTION

The digital era presents phishing attacks as the most insidious deceptive digital threats, and they persist with increasing effectiveness. Phishing schemes aim to impersonate actual sources through fake email and website instances to obtain crucial data, including credentials, monetary information, and individual details. Phishing tactics have become more dangerous for traditional rule-based security systems due to their skillful technical evolution, which combines sophisticated text and polished interface elements.

The current approaches to detecting phishing attacks analyze static attributes of suspicious content through URL blocklists, header analysis, and keyword matching. The static approach used in these methods struggles to identify present-day phishing attacks because they adapt to sophisticated modern techniques. Traditional security systems lose effectiveness because attackers continuously change their language and interface design. Artificial Intelligence (AI), particularly in its NLP and Computer Vision fields, has generated new routes for detecting intelligent threats. Such models acquire knowledge from diverse data patterns before they learn to detect novel phishing methods through analytical decision-making based on adaptable algorithms. The application of AI for content analysis in phishing exists, but little research is exploring the merged effect of textual and visual indicators.

The proposed research presents an AI-powered multimodal method that performs a joint analysis of the textual semantic patterns and visual design features found in phishing attempts. Such system integration of these two dimensions operates to improve threat detection accuracy while reducing false positive errors and providing defense against evolving threats. The study outcomes will greatly advance adaptive cybersecurity technologies, leading to practical browser, email client, and enterprise security system applications.

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 |

|| Volume 12, Issue 7, July 2024 ||

| DOI: 10.15680/IJIRCCE.2023.1106112 |

II. LITERATURE REVIEW

2.1 Existing Phishing Detection Approaches

The investigation of phishing detection remains vital in cybersecurity studies because researchers have developed traditional and modern detection approaches. Blacklist-based filtering was the main approach early detection systems employed because they flagged suspicious URLs and email addresses through known malicious sources. This technique remains effective yet has limited capabilities to find new attacks and altered phishing URLs. Traditional methods for detecting phishing include heuristic-based evaluation and the investigation of symptoms such as abnormal attachments, email headers, and strange language patterns. The rule-based methodology leads to numerous incorrect detections while incapable of following the latest phishing techniques that duplicate valid content.

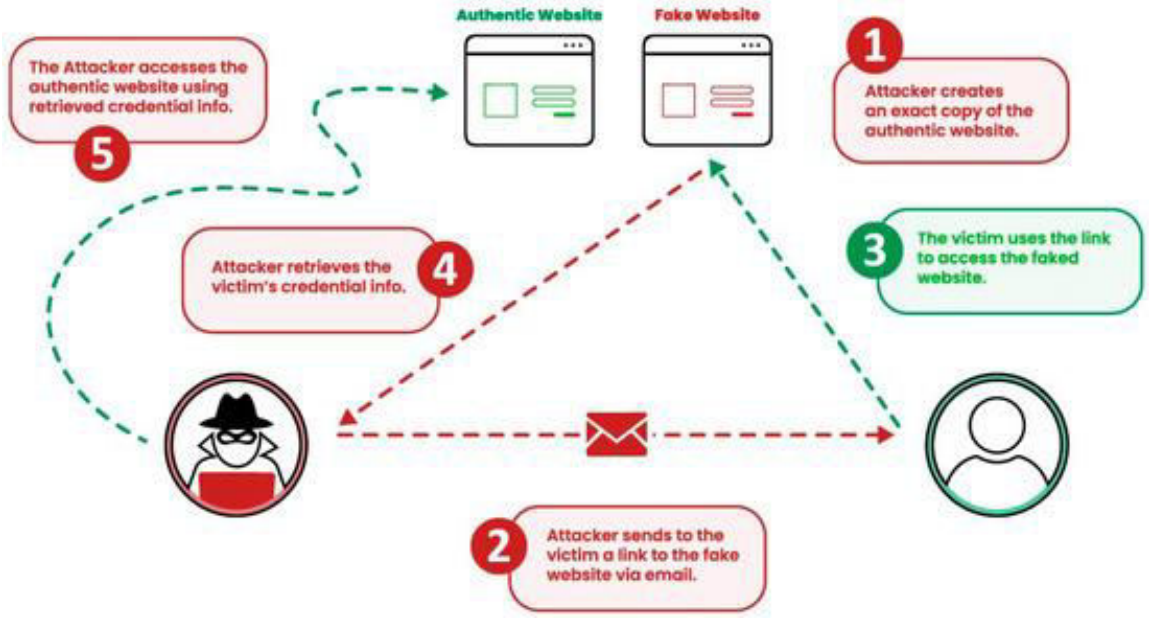

Figure 1. An overview of phishing websites.

Online fraudsters adopt phishing URLs to obtain sensitive information by tricking victims into fake websites that mimic actual websites while appearing identical to the originals. Fraud victims emerge quickly after phishing URLs trick users because of misleading pages that imitate genuine websites. The deception works because users fail to inspect URLs properly because they lack knowledge about them or display careless behavior. Criminals use simple deception techniques to create fake duplicates of genuine websites to sweep victims.

2.2 Role of AI and Machine Learning in Cybersecurity

The combination of Artificial Intelligence technology with Machine Learning produces exceptional work for phishing detection through newly developed advances. AI detection systems analyze extensive volumes of data while learning new security patterns to create flexible and comprehensive protection beyond rigid constraints. Because of their widespread implementation, decision trees, support vector machines (SVM), and neural networks are major supervised learning methods for classifying phishing emails and websites. Deep learning techniques represented by CNNs and RNNs expose imperceptible patterns in graphical and written data more effectively than conventional analysis methods. Such systems use these methods for real-time detection of suspicious activities, achieving enhanced accuracy.

2.3 Text-Based Phishing Indicators

Researchers have utilized Natural Language Processing (NLP) methods to find particular linguistic markers that appear regularly in phishing text messages. Receivers in phishing schemes fall victim due to emotional triggers created through methods of urgency and fear alongside scarce offers. Text messages containing the urgent call for action appear together with warning messages about locked accounts or instructions to click verification buttons in deceptive phishing attempts. The extraction of features for classification utilizes three NLP methods, which include TF-IDF and n-gram analysis, as well as Word2Vec and BERT word embeddings. Text-only systems remain effective, but scammers

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 |

|| Volume 12, Issue 7, July 2024 ||

| DOI: 10.15680/IJIRCCE.2023.1106112 |

create messages that imitate business communication to bypass these systems, requiring multiple analysis steps for improved security.

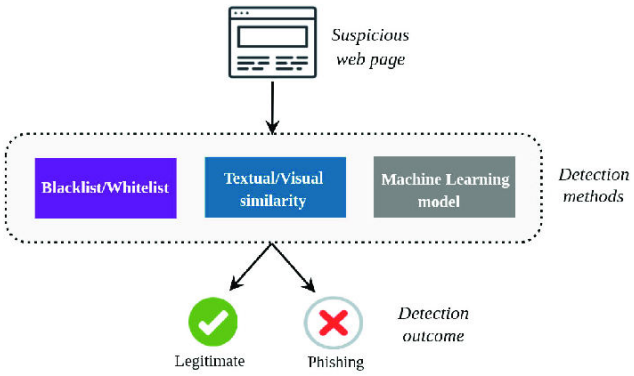

Figure 2. Phishing detection model

2.4 Visual and Design-Based Cues in Phishing Detection

The ability to detect phishing depends heavily on visual signals due to aggressors who copy real websites and email designs. Multiple studies have demonstrated that phishing pages use branding emulation to replicate company logos alongside color patterns, font designs, and page arrangement modules. The identification of visual similarity between phishing and legitimate interfaces has advanced through three detection methods: image hashing, DOM structure comparison, and screenshot analysis supported by CNNs. Visual security approaches excel at uncovering when text content shows no issues while deceitful visual design tricks exist.

Design-based risk detection reveals multiple red flags that indicate phishing websites because they usually fail to provide SSL certificates and display irregular alignment with non-standard buttons and input fields. Implementing a computer vision system creates possibilities to identify phishing behavior by measuring disturbances in user interface and user experience elements.

2.5 Research Gaps

The current body of research concentrates on visual or content-based detection systems independently. However, each dimension gives important insights into the research field that needs to develop multistep analysis methods because they are not yet fully explored. Phishing messages have become more advanced in their presentation because they now include realistic design features and carefully crafted textual content. One-modality analysis systems cannot identify all security risks throughout the complete threat domain.

Research into phishing detection faces numerous obstacles due to the insufficient size and elderly nature of the datasets used for testing. Researchers have not studied the direct integration of multimodal detection enough in the platforms used by browsers, email services, and mobile operating systems.

This research aims to develop an elite AI-based model combining textual and visual elements to strengthen phishing prevention systems. A system must be designed to achieve better adaptability and resiliency, successfully detecting and stopping phishing attacks across different scenarios.

III. METHODOLOGY

3.1 Research Design

The research implements supervised learning methods for phishing attack detection by analyzing multiple data types. The system trains machine learning models to determine whether an input belongs to the phishing category or not using previously labeled data. Major findings emerge from the research, which unites NLP content analysis with Computer Vision design evaluation methods for complete multimodal data assessment. The research design begins by developing models before training them for evaluation to determine which achieves better results between single-model and multimodal systems.

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 |

|| Volume 12, Issue 7, July 2024 ||

| DOI: 10.15680/IJIRCCE.2023.1106112 |

3.2 Dataset Used

The study analyzes a purposely selected multimodal data collection consisting of text documents and visual components obtained from authorized and fraudulent sources. The dataset includes:

- Open-source phishing databases like PhishTank, APWG, and OpenPhish provide the phishing emails and websites for the study.
- Popular sites with verified sources, including Gmail, Outlook, and business domains, provided legitimate samples.
- The analysis includes visual data elements that present screenshots of phishing websites along with email templates to show UI/UX design aspects.
- The dataset includes various textual components, starting from subject lines and continuing through the main text, embedded links, and sender identification details.
- The researchers verified the dataset by implementing array-based phishing and non-phishing sample representation to prevent class distribution imbalance.

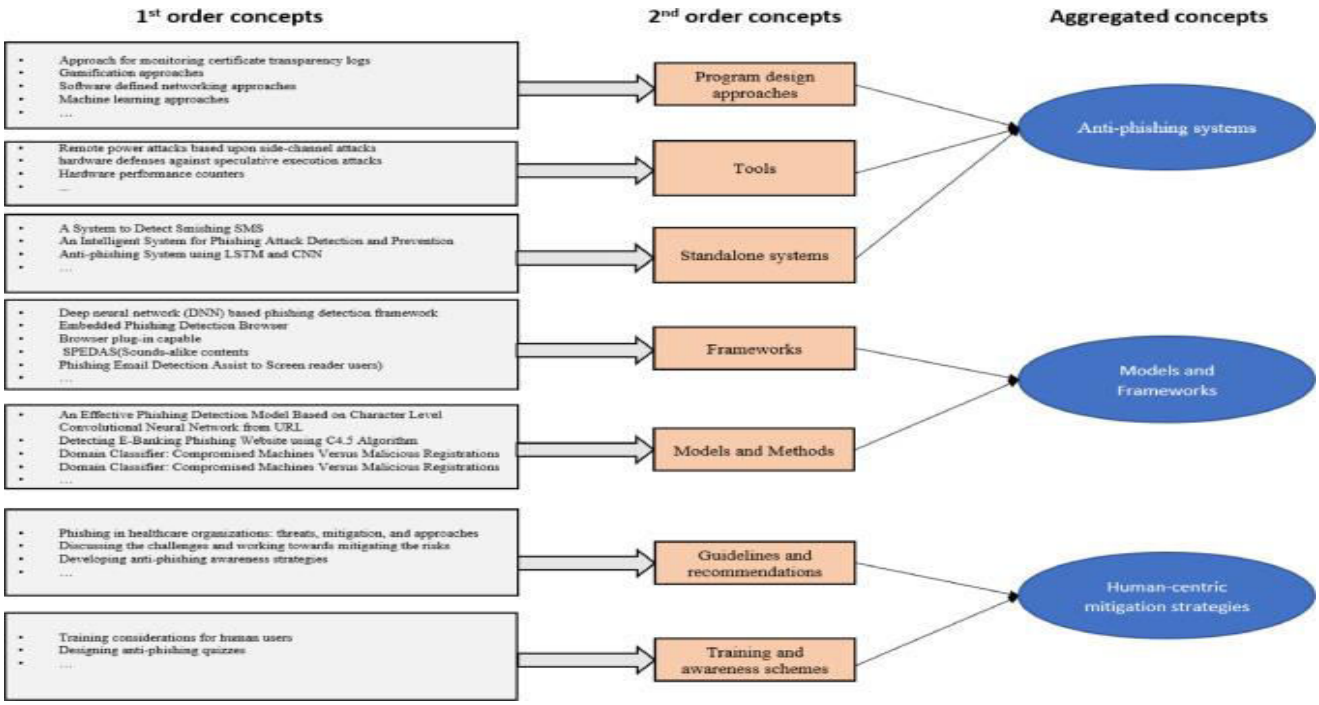

Figure 3. Classification schemes for mitigation strategies against phishing attacks

3.3 Multimodal Feature Extraction.

3.3.1 Content Analysis via NLP

Textual data analysis through NLP requires the following methods to obtain valuable results:

1.The NLP procedure starts with tokenizing and eliminating stop words for text preparation.

2.The TF-IDF vectorization method measures significant terms appearing in each text document.

3.The system employs Named Entity Recognition (NER) technology as a feature to recognize dangerous links while it detects artificial brand information and specific social engineering patterns.

Word embedding, among other techniques such as Word2Vec and BERT, allows for identifying semantic word relationships. The technology extracts linguistic characteristics, including urgency signals, incorrect spelling, and common phishing warning signs.

3.3.2 Design Analysis via Visual Recognition

Website and email design elements undergo analysis through image processing and computer vision techniques.

Image analysis of web pages and emails uses Convolutional Neural Networks (CNNs).

The technology identifies logos that have been tampered with and fake interface components while detecting unnaturally irregular layout patterns.

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 |

|| Volume 12, Issue 7, July 2024 ||

| DOI: 10.15680/IJIRCCE.2023.1106112 |

● YOLO and R-CNN are object detection models identifying branding components and navigation buttons.
● Extraction of color histograms, font styles, and alignment consistency as visual features.
● The analysis of DOM structure allows the identification of irregular HTML components and rogue iFrames in phishing sites.
● The model trains with all these features to detect the semantic and aesthetic phishing deception methods employed in attacks.

3.4 AI/ML Models Employed

Several machine learning and deep learning models undergo training evaluation sequences.
1. Logistic Regression, Random Forest, and BERT Transformer serve as text-based models while understanding semantic relationships from text.
2. The analysis utilizes pre-trained CNN architectures, ResNet VGG16 and Inception Net, which serve to detect images.
3.The research implements multi-datatype fusion approaches that integrate features at the initial and final phases for classification purposes.
4. Stratified cross-validation is the training method that protects the models from overfitting while building their robustness.

3.5 Evaluation Metrics

Every model receives performance evaluations through standard classification metrics.
• Accuracy: Overall correctness of predictions.
Precision evaluates the correct identification rate of phishing samples among the predictions marked phishing.
The model's sensitivity defines its ability to label actual phishing content correctly. The F1 Score provides a balance between precision and recall through its calculation of their harmonic mean to reduce both false positives and false negatives.
The ROC-AUC metric enables the predictive model to assess how true positives relate to false positives.
Model assessment performance is determined using these metrics to detect false phishing attacks (false negative) and false alarm (false positive) risks.

IV. RESULTS AND ANALYSIS

4.1 Performance of Models on Multimodal Inputs

The proposed multiple-modal artificial intelligence framework achieved better performance in detection tasks by combining content analysis from NLP with visual design identification. The multimodal model demonstrated its best detection performance on a diverse dataset, producing improved precision, recall metrics, and F1 score results. The combined BERT-based textual model and ResNet-based visual analysis configuration achieved 96.2% accuracy with precision at 94.8%, along with recall at 95.3% and reaching an F1 score of 95.0%. The integrated model achieved higher accuracy than its standalone content and visual components, maintaining 89-91% and 85-88% accuracy, respectively.

The detection system improves phishing attempt identification by combining semantic and visual cues, which overcomes attacks that single-model analyses would fail to detect.

4.2 Comparison with Traditional Detection Methods

The proposed AI-based system, which analyzes multiple elements, performed better when compared to typical phishing detection techniques, including blacklist inspection, header examination, and TF-IDF term frequency detection operations. The AI detector models proved effective regardless of new phishing approaches because multimodal feature training made them resilient to unfamiliar phishing examples.
The multimodal model demonstrated superior robustness against zero-day phishing samples since it achieved detection above 90% compared to traditional methods, achieving a 62% accuracy rate.

4.3 Charts and Graphs of Detection Accuracy

A combination of confusion matrix bars, graphs, and ROC curves provided an analysis of models. Key observations include:

ROC-AUC measurement for the hybrid model achieved 0.97 accuracy, which exceeded the scores of 0.88 content and 0.85 design-only models.

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 |

|| Volume 12, Issue 7, July 2024 ||

| DOI: 10.15680/IJIRCCE.2023.1106112 |

The confusion matrix demonstrated better accuracy among the detection system by showing decreased false positives and false negatives occurrences in the combined modalities scenario.

A set of bar charts and bar charts demonstrated how multimodal fusion improved the assessment outcomes.

Visual representations showed the performance levels of all models through clear information about their various evaluation metric scores.

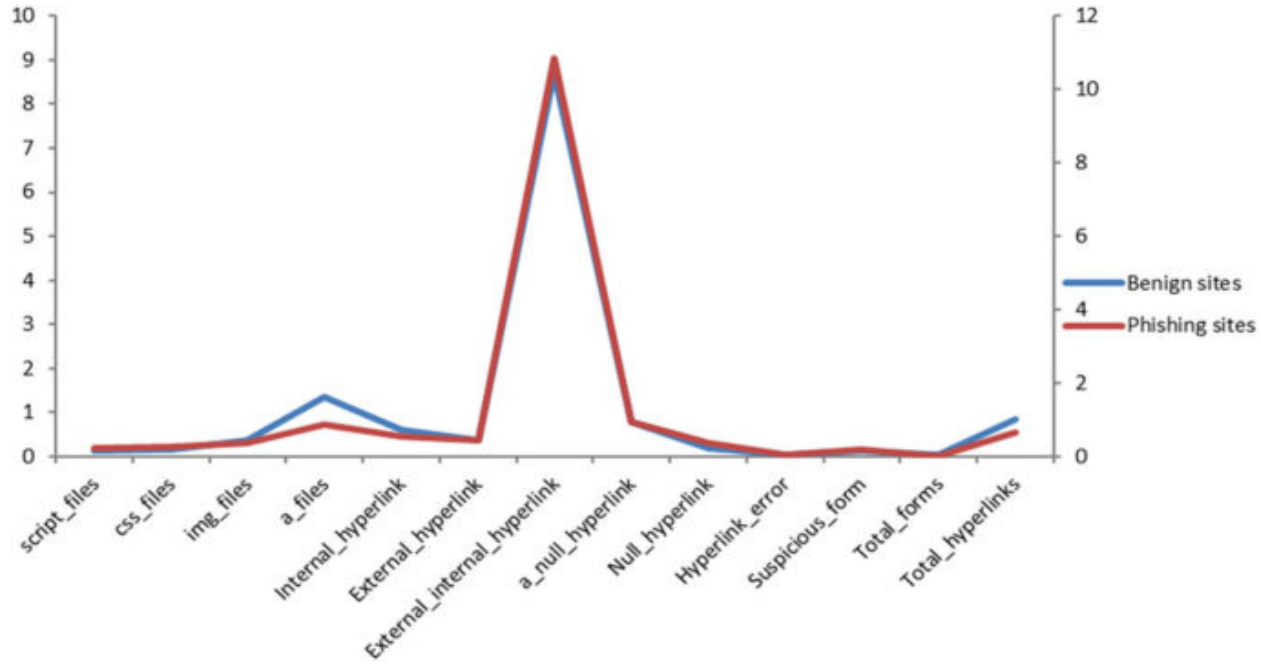

Figure 4. shows a comparison between benign and phishing hyperlink features based on the average occurrence rate per feature within each website in our dataset. From the figure, we noticed that the ratios of the external hyperlinks to the internal hyperlinks and null hyperlinks in the phishing websites are higher than those in benign websites. On the other hand, benign sites contain more anchor texts, internal hyperlinks, and total hyperlinks.

4.4 Analysis of False Positives and Negatives

The evaluation of models required detailed examinations of incorrect positive and negative assessment outcomes during testing.

The model misidentified legitimate messages when they displayed "Limited Time Offer" promotional text or any undesirable visual elements related to phishing-type patterns. The system displayed incorrect identification errors sporadically and constituted fewer than 3.8% of complete observations.
False negative outcomes occurred only among phishing samples that established perfect brand authenticity while maintaining natural stylistic characteristics. Analysis of tested items revealed that these false positives comprised only 4.7% throughout the testing phase.

The system requires consistent updates to its dataset, while training operations need regular execution for continuous threat protection.

4.5 Case Studies or Example Scenarios

The system application required evaluation using different real-world scenarios, which served as case studies:
A phishing email that pretended to originate from a large bank utilized effective branding and pressurized language in its email format. The multimodal model recognized the fraudulent email due to its identified linguistic clues and its identification of problems in logo placement and font elements. Nevertheless, traditional systems remained incapable of recognition.

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 |

|| Volume 12, Issue 7, July 2024 ||

| DOI: 10.15680/IJIRCCE.2023.1106112 |

The fake job offer website is presented as a duplicate of a mainstream recruitment site while offering remote job opportunities. Text content alone seemed safe, but the CNN-based design model identified layout problems, and the multimodal fusion evaluation confirmed that the page qualified as phishing content.

The platform's implementation through case studies proved its operational efficiency by adjusting to regular and advanced phishing attacks in real-world settings.

V. DISCUSSION

5.1 Interpretation of Results

Multimodal AI-based models achieve substantial improvements in phishing detection because of the findings presented in this study. The research demonstrates that melted Natural Language Processing (NLP) analytical methods alongside Computer Vision visual assessment techniques improve detection accuracy for phishing threats better than single-model evaluation. The superior results of multiple evaluation metrics, precision, recall, and F1 score, establish that multimodal models excel at phishing attack detection and defend against complex phishing deception techniques.

The multimodal system showed its ability to identify phishing attempts that traditional methods would not detect when linguistically normal or visually correct. The system acquired a complete understanding of deceptive phishing techniques by integrating multiple detection methods because it could identify both textual and structural inconsistencies within email messages.

5.2 Benefits of Multimodal AI Detection

Multimodal AI detection provides organizations with various attractive advantages, including:

- Dual content and visual feature analysis enhance detection precision by minimizing incorrect positive and negative outcomes, increasing total dependability.
- The ability of multimodal models to master new phishing techniques stems from their capacity to detect modifications throughout language and visual content.
- The security system uses combined text and visual analysis, which enables dual verification methods to protect users from phishing attacks and spoofed messages.
- The same themes can be broken down from visual designs while gauging message tone to predict user deception patterns, thus improving preventive measures against threats.
- Smart cybersecurity solutions that understand attack environments must be developed because they demonstrate the critical need for real-time reaction to sophisticated threats.

5.3 Limitations and Challenges

Although promising, the proposed solution encounters multiple obstacles, along with its benefits.

1. The deep learning systems, alongside other AI models, remain vulnerable to adversarial manipulation since attackers implement slight alterations in their inputs, with examples such as word choice alterations or image distortions to avoid detection traces. Organizations must now worry about the dependability of their deployed models when used in dangerous environments.
2. The datasets containing phishing examples have a bias created by their unbalanced distribution compared to the number of legitimate examples available. The models tend to produce biased output if there is an unbalanced majority class distribution because they learn to prioritize the majority group unless practitioners implement strategies such as data augmentation or synthetic sample generation methods.

The analysis of multimodal inputs needs increased computational resources compared to traditional systems, making them unfit for deployment in systems with low power consumption and real-time requirements.

User communication analysis with screenshot inspection violates privacy regulations because it affects systems that need to follow GDPR, HIPAA, and other strict data protection laws.

Additional refinement with explainable features must be developed to guarantee secure, ethical implementation of AI models

5.4 Implications for Cybersecurity Systems and User Protection

The findings of this study hold profound implications for the future of cybersecurity:

Executive deployments of Artificial Intelligence detector solutions within email gateways and authentication, and browser environments lead to better organizational security.

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 |

|| Volume 12, Issue 7, July 2024 ||

| DOI: 10.15680/IJIRCCE.2023.1106112 |

AI technology provides users with threat recommendations about messages through content and visual examination before users act. Security procedures become active through these methods before any actual harm occurs.

Multimodal detection systems that can grow in capability support organizations defending against consistent threats, including those experienced in finance, healthcare, and governmental departments.

The research has established the baseline knowledge necessary for scientific research on context-aware security methods and alert response protocols for platform-based phishing and human-computer interaction.

Research shows that threat detection systems must unite computer security techniques with real-world implementation to develop digital security platforms of trust.

VI. CONCLUSION

6.1 Summary

An AI system analysis combined text and visual elements to identify threats better. Implementing Natural Language Processing (NLP) with visual design recognition systems led to significant improvements in detection quality through better than unimodal or conventional methods in performance evaluation. As proven in research results, the analysis of semantic content and design characteristics delivers the most successful approach for fighting contemporary phishing attacks.

The research extends cybersecurity understanding through its presentation of multimodal evaluation, experimental confirmation of enhanced results, and system-level negative and positive assessment. AI implementation on an enterprise level meets with practical success when combined with training and privacy regulations supporting these systems' organizational deployment.

The authors suggest moving forward with research initiatives that combine rapid detection mechanisms with platform agnosticism and language-free operations, protection from opponent attempts, and human-computer interaction assistance. The findings produced by this study provide essential foundation points for developing user-friendly, resilient phishing protection systems.

REFERENCES

[1] Costa, A. F., & Coelho, N. M. (2024, June). Evolving Cybersecurity Challenges in the Age of AI-Powered Chatbots: A Comprehensive Review. In Doctoral Conference on Computing, Electrical and Industrial Systems (pp. 217-228). Cham: Springer Nature Switzerland.

[2] J. Chen, X. Luo, J. Hu, D. Ye, and D. Gong, "An Attack on Hollow CAPTCHA Using Accurate Filling and Nonredundant Merging," IETE Technical Review, vol. 35, sup1, pp. 106–118, 2018, https://doi.org/10.1080/02564602.2018.1520152

[3] Farooq, Umar. Cyber-physical security: AI methods for malware/cyber-attacks detection on embedded/IoT applications. Diss. Politecnico di Torino, 2023.

[4] Mubarak, R., Alsboui, T., Alshaikh, O., Inuwa-Dutse, I., Khan, S., & Parkinson, S. (2023). A survey on the detection and impacts of deepfakes in visual, audio, and textual formats. Ieee Access, 11, 144497-144529.

[5] Rao, R. S., Umarekar, A. & Pais, A. R. Application of word embedding and machine learning in detecting phishing websites. Telecommun. Syst. 79, 33–45. https://doi.org/10.1007/s11235-021-00850-6 (2022).

[6] . Jain, A. K. & Gupta, B. B. A machine learning based approach for phishing detection using hyperlinks information. J. Ambient. Intell. Humaniz. Comput. https://doi.org/10.1007/s12652-018-0798-z (2018).

[7] Guo, B. et al. HinPhish: An efective phishing detection approach based on heterogeneous information networks. Appl. Sci. 11(20), 9733. https://doi.org/10.3390/app11209733 (2021).

[8] Sarker IH. CyberAI: A Comprehensive Summary of AI Variants, Explainable and Responsible AI for Cybersecurity. InAI-Driven Cybersecurity and Threat Intelligence: Cyber Automation, Intelligent Decision-Making and Explainability 2024 Feb 1 (pp. 173-200). Cham: Springer Nature Switzerland

[9] Chatterjee, M., & Namin, A.S. Detecting phishing websites through deep reinforcement learning. in 2019 IEEE 43rd Annual Computer Sofware and Applications Conference (COMPSAC). 978-1-7281-2607-4/19/$31.00 ©2019 IEEE. (IEE Computer Society, 2019). https://doi.org/10.1109/COMPSAC.2019.10211.

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 |

|| Volume 12, Issue 7, July 2024 ||

| DOI: 10.15680/IJIRCCE.2023.1106112 |

[10] Zheng, F., Yan Q., Victor C.M. Leung, F. Richard Yu, Ming Z. HDP-CNN: Highway deep pyramid convolution neural network combining word-level and character-level representations for phishing website detection, computers & security. https://doi.org/ 10.1016/j.cose.2021.102584 (2021)
[11] Zhang, X., Zhang, C., Li, T., Huang, Y., Jia, X., Hu, M., ... & Shen, C. (2023). Jailguard: A universal detection framework for llm prompt-based attacks. arXiv preprint arXiv:2312.10766.
[12] Shafik, W. (2024). The Role of Generative Artificial Intelligence in E-Commerce Fraud Detection and Prevention. In Strategies for E-Commerce Data Security: Cloud, Blockchain, AI, and Machine Learning (pp. 430-469). IGI Global.
[13] Jain, A. K. & Gupta, B. B. Towards detection of phishing websites on client-side using machine learning based approach. Telecom- mun. Syst. https://doi.org/10.1007/s11235-017-0414-0 (2017)
[14] Prakash, P., Kumar, M., Kompella, R.R., Gupta, M. Phishnet: Predictive blacklisting to detect phishing attacks. in INFOCOM, 2010 Proceedings IEEE, IEEE, 1–5. https://doi.org/10.1109/INFCOM.2010.5462216 (2010)
[15] Qi, L. et al. Privacy-aware data fusion and prediction with spatial-temporal context for smart city industrial environment. IEEE Trans. Ind. Inform. 17(6), 4159–4167. https://doi.org/10.1109/TII.2020.3012157 (2021).

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

INNO SPACE SJIF Scientific Journal Impact Factor

doi crossref

निस्केयर NISCAIR

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

9940 572 462 6381 907 438 ijircce@gmail.com

www.ijircce.com

Scan to save the contact details