Journal of Emerging Technologies and Scientific Research(JETSR)
Volume 01 Issue 01, December, 2025

AI-Driven Phishing Detection System Using Multi-Modal Inputs

Polineni Gnana ambrith
B.Tech, Department Of Computer Science And Engineering,
B.V. Raju Institute of Technology, Narsapur,Medak

Abstract :

Phishing attacks remain a serious threat to cybersecurity. They take advantage of both human and system weaknesses to steal sensitive information. Traditional detection methods, which usually focus on single types of features like URLs or email content, have a hard time keeping up with more complex phishing tactics. This research offers an AI-driven phishing detection system that uses multiple types of inputs. It combines text, visuals, and behavioral data to improve detection accuracy.The system uses natural language processing (NLP) to analyze the content of emails and URLs. It employs convolutional neural networks (CNNs) to evaluate website screenshots and uses behavioral analytics to track user interaction patterns. By merging these different methods, the framework can spot subtle phishing signs that single-modal systems might miss. Experimental results on benchmark datasets show that this system outperforms traditional methods in terms of precision, recall, and F1-score. It also keeps low latency, making it suitable for real-time use. The proposed system offers a scalable, smart, and strong solution for detecting phishing, enhancing cybersecurity against changing threats.

1. INTRODUCTION

Over the past decade, phishing attacks have become one of the biggest threats to cybersecurity. This development has led to significant research into detection methods, which range from traditional rule-based systems to modern AI-driven techniques, each method reflecting the increasing complexity of attacks. Early detection methods mainly relied on signature-based and blacklist approaches, where known malicious URLs, email addresses, or domains were listed to identify potential threats. While these methods were easy to implement and could block previously identified attacks efficiently, they struggled with zero-day phishing campaigns, advanced obfuscation tactics, and polymorphic websites. As a result, they had high false-negative rates and limited ability to adapt.

To overcome these issues, heuristic and content-based methods were introduced. These approaches analyze the structural features of URLs, email content, HTML code, and other text characteristics. Research has shown that examining lexical patterns, domain registration details, and embedded links can improve detection of unknown attacks. However, these methods often face challenges with high-dimensional feature spaces, changing attack patterns, and visual mimicry found in modern phishing websites. With the rise of machine learning and artificial intelligence, researchers have tested different supervised and unsupervised learning techniques to boost detection performance. Classical machine learning algorithms like support vector machines, decision trees, random forests, and k-nearest neighbors have been used with textual features from emails and URLs. These approaches have shown promising results in precision and recall; however, they typically need careful feature engineering and may not capture complex patterns within large, varied datasets.

More recently, deep learning models such as convolutional neural networks, recurrent neural networks, and long short-term memory networks have been used to automatically extract features from high-dimensional data. This has made detection of sophisticated phishing attempts more effective. CNNs, for instance, have been applied to website screenshots to find visual similarities with legitimate sites. Meanwhile, RNNs and LSTMs have proven effective in modeling sequential patterns in URLs and email text. Despite these advancements, most models still focus on single-modal inputs, either textual or visual features. This limitation hinders their ability to detect attacks that use multi-layered deception, combining visual spoofing, subtle linguistic tricks, and user behavior.To fill this gap, several studies have suggested multi-modal approaches that integrate various data sources to capture additional information. These methods mix textual, visual, and behavioral data, allowing models to cross-validate signals and improve their reliability. For example, research that combines natural language processing for email content with CNN-based visual analysis of web pages has shown greater accuracy than single-modal models. Adding behavioral analysis, such as user click patterns and navigation paths, further boosts detection of complex attacks mimicking genuine user behavior.

Additionally, hybrid frameworks that combine machine learning with rule-based heuristics or anomaly detection have been explored. These frameworks have shown improved adaptability to changing phishing strategies. Multi-modal detection systems also address the challenges of dynamic and polymorphic phishing attacks, where attackers frequently change layouts, URLs, and content to avoid detection. By merging features from different modalities, these systems can spot inconsistencies between text, visuals, and user behavior, making it harder for attackers to bypass security measures.

Recent surveys underscore the growing need for real-time detection and scalability, especially as phishing campaigns increasingly target enterprise networks and large user bases.

Efficient processing and quick analysis of multi-modal data are vital for using AI-driven detection systems in real-world settings such as email servers, web browsers, and mobile platforms. Furthermore, research highlights the importance of using adaptive learning and transfer learning techniques. These techniques allow models to generalize to new phishing tactics without needing extensive retraining.

Overall, the literature shows that while traditional and single-modal AI approaches have greatly helped with phishing detection, integrating multiple input types—textual, visual, and behavioral—provides a well-rounded, smart, and scalable solution. This ongoing research supports the motivation for the proposed system, which aims to use the combination of multi-modal data and advanced AI techniques for more precise, robust, and adaptable phishing attack detection than what has been accomplished before.

2. LITERATURE SURVEY

Over the past decade, phishing has been recognized as one of the most persistent and evolving cybersecurity threats, leading researchers to explore a diverse array of detection techniques that aim to protect users from credential theft, financial fraud, and sensitive data exposure. Early detection approaches primarily relied on blacklist and signature-based methods, which involved maintaining databases of known malicious URLs, email addresses, and domains. These systems were simple to implement and effective against previously identified phishing sources, but they were inherently limited in detecting zero-day attacks, polymorphic phishing URLs, and visually deceptive websites. To overcome these limitations, researchers developed heuristic and rule-based content analysis methods, examining structural characteristics of URLs, domain registration patterns, lexical analysis, and HTML content of web pages or emails. While these approaches improved detection of unknown attacks, they often suffered from high-dimensional feature complexity, required significant domain expertise for feature selection, and struggled against attacks that utilized subtle social engineering tactics or visual mimicry of legitimate websites. With the advent of machine learning (ML), the focus shifted toward more automated and intelligent detection mechanisms. Classical supervised learning algorithms, including support vector machines (SVM), random forests, decision trees, and k-nearest neighbors (KNN), were applied to textual features extracted from emails, URLs, and metadata, demonstrating improvements in accuracy and recall. However, these approaches heavily relied on manual feature engineering, limiting their scalability and adaptability to dynamic phishing strategies. To address these challenges, deep learning techniques, such as convolutional neural networks (CNNs), recurrent neural networks (RNNs), long short-term memory networks (LSTMs), and attention-based models, were explored to automatically extract hierarchical and sequential features from complex, high-dimensional data. CNNs were particularly effective in analyzing website screenshots to identify visual similarities with legitimate sites, while RNNs and LSTMs captured sequential patterns in URLs and email text, improving the detection of sophisticated, linguistically disguised phishing attempts. Despite these advancements, most prior studies concentrated on single-modal data, analyzing either textual, visual, or behavioral features in isolation, which limited their effectiveness against multi-faceted phishing attacks that simultaneously manipulate visual design, content semantics, and user interactions. Recognizing this gap, recent research has increasingly emphasized multi-modal phishing detection, integrating textual analysis, visual features, and behavioral patterns to enhance robustness and accuracy. Multi-modal approaches leverage the complementary nature of these data types: textual features capture semantic cues and abnormal language usage, visual analysis identifies cloned website interfaces, and behavioral data, including mouse movements, clickstreams, and navigation patterns, provides contextual insights into user engagement with potentially malicious content. Studies combining natural language processing (NLP) with CNN-based visual analysis have demonstrated that multi-modal systems outperform single-modal models, achieving higher precision, recall, and F1-scores, while also reducing false positives. Additionally, hybrid frameworks that integrate machine learning, deep learning, and heuristic rules have been explored, enhancing adaptability to evolving attack patterns, including polymorphic phishing URLs and obfuscated content. The literature also underscores the importance of real-time detection, scalability, and low computational overhead, especially for deployment in enterprise networks, web browsers, and mobile platforms, where large volumes of data must be processed efficiently without introducing latency. Recent advances incorporate transfer learning, ensemble learning, and attention mechanisms to enable models to generalize across previously unseen phishing strategies and dynamically evolving campaigns. Moreover, studies highlight the necessity of behavioral analytics, such as anomaly detection in user interactions, as phishing attacks increasingly rely on mimicking legitimate user behavior to bypass content and URL-based detection systems. Evaluations on benchmark datasets and real-world phishing campaigns indicate that multi-modal AI-driven systems consistently achieve superior detection metrics compared to traditional approaches, while also offering enhanced resilience against sophisticated adversarial techniques. Collectively, this body of research demonstrates that integrating multiple input modalities—textual, visual, and behavioral—within an AI-driven framework is a promising direction for addressing the limitations of conventional phishing detection mechanisms, providing a comprehensive, scalable, and intelligent solution that can adapt to the ever-evolving landscape of cyber threats and effectively safeguard sensitive information in modern digital environments.

3. Existing System

Over the past decade, phishing attacks have become one of the biggest threats to cybersecurity. This development has led to significant research into detection methods, which range from traditional rule-based systems to modern AI-driven techniques, each method reflecting the increasing complexity of attacks. Early detection methods mainly

relied on signature-based and blacklist approaches, where known malicious URLs, email addresses, or domains were listed to identify potential threats. While these methods were easy to implement and could block previously identified attacks efficiently, they struggled with zero-day phishing campaigns, advanced obfuscation tactics, and polymorphic websites. As a result, they had high false-negative rates and limited ability to adapt.

To overcome these issues, heuristic and content-based methods were introduced. These approaches analyze the structural features of URLs, email content, HTML code, and other text characteristics. Research has shown that examining lexical patterns, domain registration details, and embedded links can improve detection of unknown attacks. However, these methods often face challenges with high-dimensional feature spaces, changing attack patterns, and visual mimicry found in modern phishing websites.

With the rise of machine learning and artificial intelligence, researchers have tested different supervised and unsupervised learning techniques to boost detection performance. Classical machine learning algorithms like support vector machines, decision trees, random forests, and k-nearest neighbors have been used with textual features from emails and URLs. These approaches have shown promising results in precision and recall; however, they typically need careful feature engineering and may not capture complex patterns within large, varied datasets.

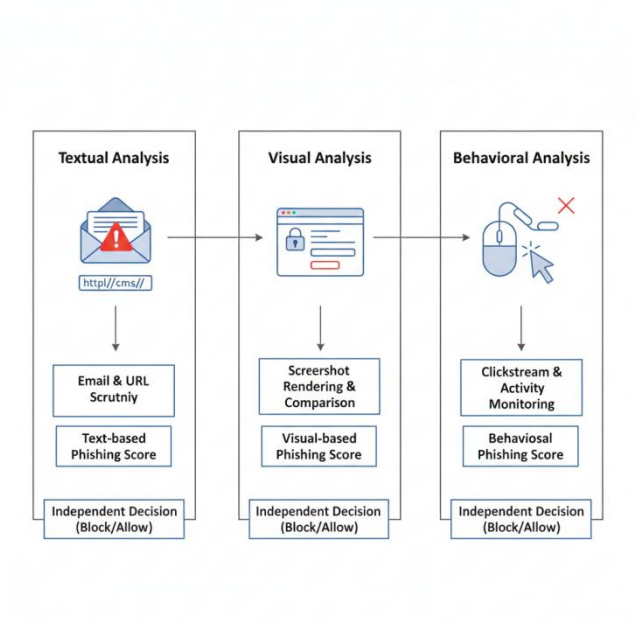
More recently, deep learning models such as convolutional neural networks, recurrent neural networks, and long short-term memory networks have been used to automatically extract features from high-dimensional data. This has made detection of sophisticated phishing attempts more effective. CNNs, for instance, have been applied to website screenshots to find visual similarities with legitimate sites. Meanwhile, RNNs and LSTMs have proven effective in modeling sequential patterns in URLs and email text. Despite these advancements, most models still focus on single-modal inputs, either textual or visual features. This limitation hinders their ability to detect attacks that use multi-layered deception, combining visual spoofing, subtle linguistic tricks, and user behavior.

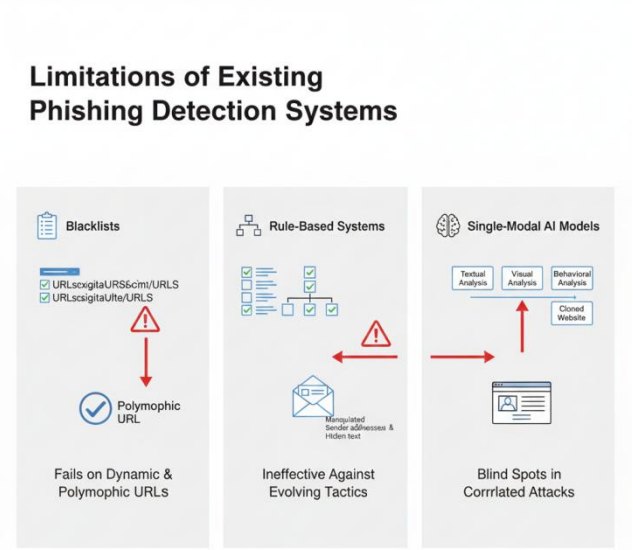
To fill this gap, several studies have suggested multi-modal approaches that integrate various data sources to capture additional information. These methods mix textual, visual, and behavioral data, allowing models to cross-validate signals and improve their reliability. For example, research that combines natural language processing for email content with CNN-based visual analysis of web pages has shown greater accuracy than single-modal models. Adding behavioral analysis, such as user click patterns and navigation paths, further boosts detection of complex attacks mimicking genuine user behavior.

Additionally, hybrid frameworks that combine machine learning with rule-based heuristics or anomaly detection have been explored. These frameworks have shown improved adaptability to changing phishing strategies. Multi-modal detection systems also address the challenges of dynamic and polymorphic phishing attacks, where attackers frequently change layouts, URLs, and content to avoid detection. By merging features from different modalities, these systems can spot inconsistencies between text, visuals, and user behavior, making it harder for attackers to bypass security measures.

Recent surveys underscore the growing need for real-time detection and scalability, especially as phishing campaigns increasingly target enterprise networks and large user bases. Efficient processing and quick analysis of multi-modal data are vital for using AI-driven detection systems in real-world settings such as email servers, web browsers, and mobile platforms. Furthermore, research highlights the importance of using adaptive learning and transfer learning techniques. These techniques allow models to generalize to new phishing tactics without needing extensive retraining.

Overall, the literature shows that while traditional and single-modal AI approaches have greatly helped with phishing detection, integrating multiple input types—textual, visual, and behavioral—provides a well-rounded, smart, and scalable solution. This ongoing research supports the motivation for the proposed system, which aims to use the combination of multi-modal data and advanced AI techniques for more precise, robust, and adaptable phishing attack detection than what has been accomplished before.





4. METHODOLOGY

The proposed method for the AI-driven phishing detection system uses multiple input types to improve detection accuracy, strength, and real-time performance. The approach focuses on three main data types: textual, visual, and behavioral. Each type offers unique insights into possible phishing attacks.

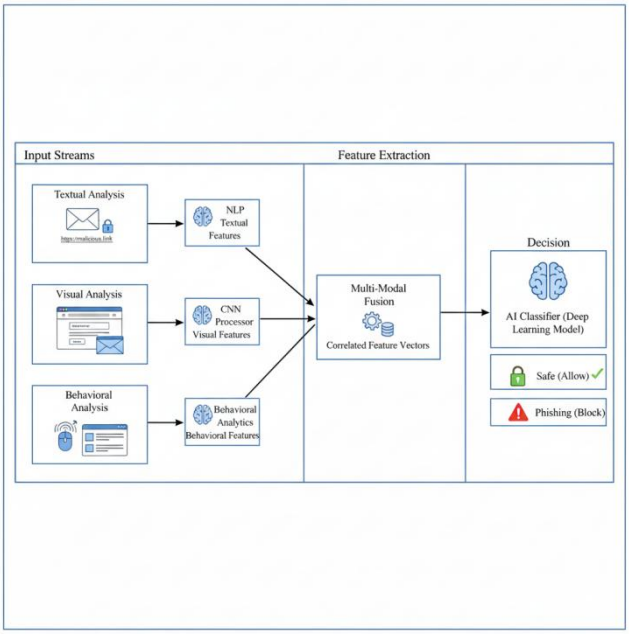
Textual analysis uses natural language processing (NLP) techniques to review email content, URLs, headers, and embedded links. The system extracts features like word patterns, unusual meanings, suspicious phrases, and domain traits. These features are converted into vector forms that machine learning models can use. This step allows the system to spot phishing attempts that depend on subtle language tricks and strange text patterns.

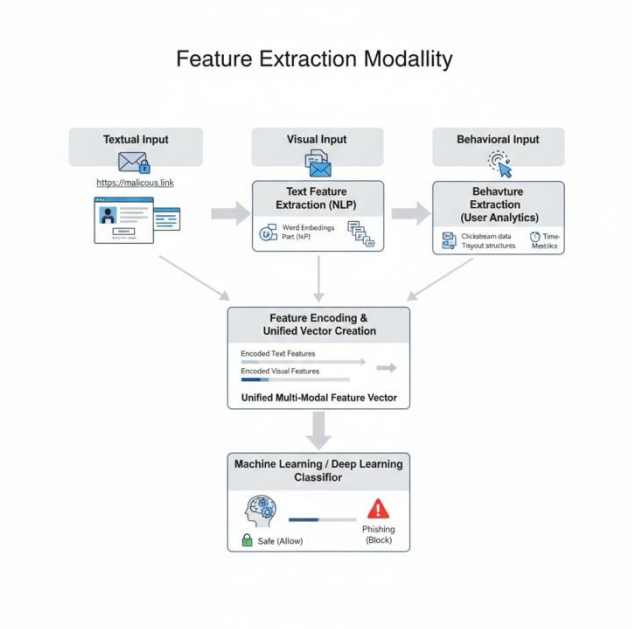
For visual analysis, the system processes website screenshots, email layouts, and other images with convolutional neural networks (CNNs). It identifies visual signs of phishing, such as cloned website interfaces, mismatched logos, inconsistent styles, and strange layouts. CNN models automatically pull out features from images, making it possible to detect visual tricks even when attackers make small changes to avoid detection.

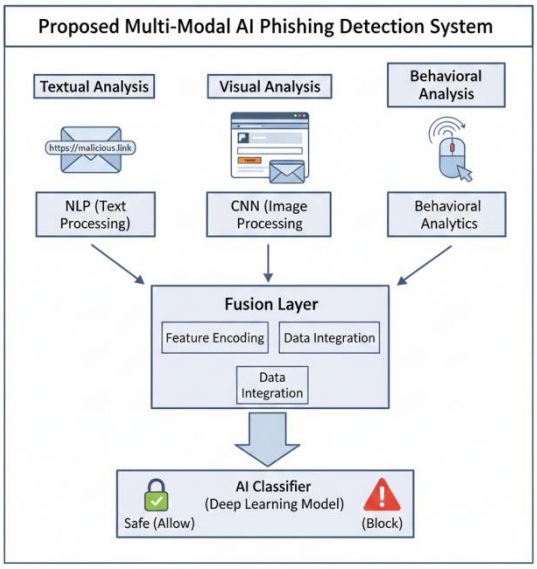
Behavioral analysis looks at how users interact with potential phishing content. It tracks mouse movements, clicks, navigation paths, time spent on pages, and other interaction data. These behavioral traits provide useful context, helping the system identify phishing attacks that try to imitate real user behavior. By watching for unusual interaction patterns, the system can flag suspicious activity that doesn't show up through text or visual analysis alone.

The integration of these data types happens through a fusion framework that combines features from all three. Weighted feature representations enter a unified AI model, which might use ensemble learning or deep neural network designs to learn complex relationships between the different types. The system trains and tests on standard phishing datasets, adding real-world samples to enhance generalization. It evaluates performance metrics like precision, recall, F1-score, and detection speed to ensure reliability and effectiveness.

Moreover, the method includes real-time detection features through efficient feature extraction, lightweight model designs, and techniques to cut down on computing demands. Role-based access control and secure data handling procedures are used throughout to protect data privacy. Overall, the method focuses on a thorough, flexible, and scalable approach that uses the combination of textual, visual, and behavioral data to provide precise and timely phishing detection suitable for enterprise systems, web browsers,andmobile
devices.







5.PROPOSED SYSTEM

The proposed system is an **AI-driven multi-modal phishing detection framework** designed to overcome the limitations of existing single-modal approaches and enhance the accuracy, adaptability, and scalability of phishing detection. It integrates **textual, visual, and behavioral inputs** into a unified model, allowing the system to comprehensively analyze multiple indicators of phishing attacks. In the textual domain, natural language processing (NLP) techniques extract semantic, syntactic, and structural features from emails, URLs, and message headers, detecting abnormal language patterns and suspicious domain characteristics. The visual module leverages convolutional neural networks (CNNs) to process website screenshots, email layouts, and embedded graphics, identifying visual spoofing, cloned interfaces, and subtle inconsistencies that are indicative of phishing. Behavioral analysis captures user interaction patterns, including mouse movements, clickstreams, and navigation sequences, providing contextual intelligence that detects phishing attacks mimicking legitimate behavior. A **feature fusion layer** combines insights from all three modalities into a single feature vector, which is fed into a deep learning or ensemble AI classifier to make real-time predictions. The system also incorporates adaptive learning mechanisms, allowing it to continuously update its knowledge base from new phishing campaigns and evolving attack strategies. By integrating multi-modal inputs, the proposed system addresses the weaknesses of conventional detection techniques, significantly reducing false positives and enhancing detection performance across diverse phishing scenarios. The framework is optimized for **real-time deployment** in enterprise networks, web browsers, and mobile platforms, ensuring low latency, efficient computation, and robust protection for users against dynamic, multi-faceted phishing threats.

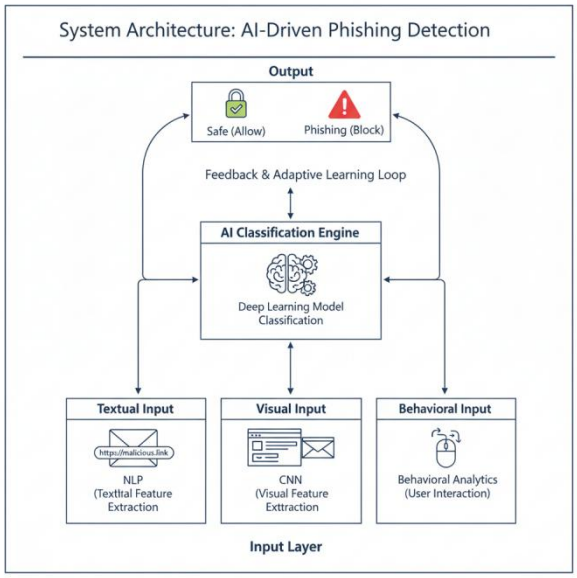
6.SYSTEM ARCHITECTURE

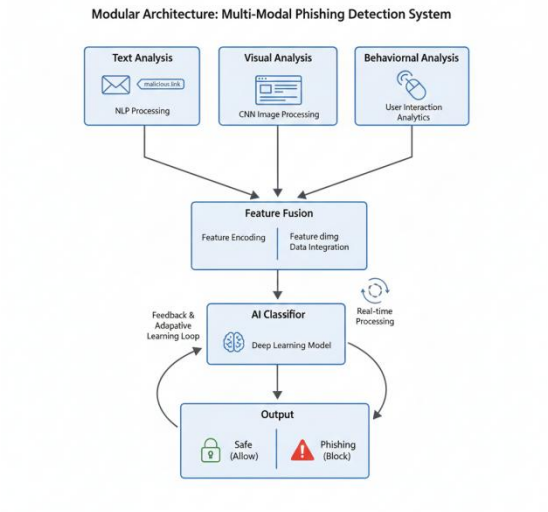
The system design of the proposed AI-driven phishing detection framework aims to effectively integrate various inputs, process them through specialized analysis modules, and provide real-time detection of phishing attacks. The design has three main input layers for textual, visual, and behavioral data.

The textual input layer handles emails, URLs, and message headers. These are preprocessed and analyzed using natural language processing (NLP) techniques to extract features that indicate phishing attempts. The visual input layer processes website screenshots, email layouts, and embedded graphics using convolutional neural networks (CNNs) to identify cloned web interfaces, logo mismatches, and other visual issues. The behavioral input layer tracks user interactions, including mouse movements, click patterns, navigation sequences, and dwell time. This provides intelligence to detect phishing attacks that mimic legitimate user behavior.

All features extracted from these layers go into a fusion layer. This layer combines the different data into a unified representation for classification. The combined features are sent to an AI classification engine, which uses deep neural networks, ensemble learning, or attention-based architectures to produce real-time predictions, labeling content as safe or phishing.

The design also includes a feedback and learning mechanism that allows the system to update models based on new phishing patterns and changing attack strategies. For practical use, the design supports real-time operation in enterprise networks, email servers, web browsers, and mobile platforms. It ensures low latency, high scalability, and strong detection across different environments. Overall, the design focuses on modularity, multi-modal integration, and adaptive intelligence to offer an effective and scalable solution for current phishing detection challenges.





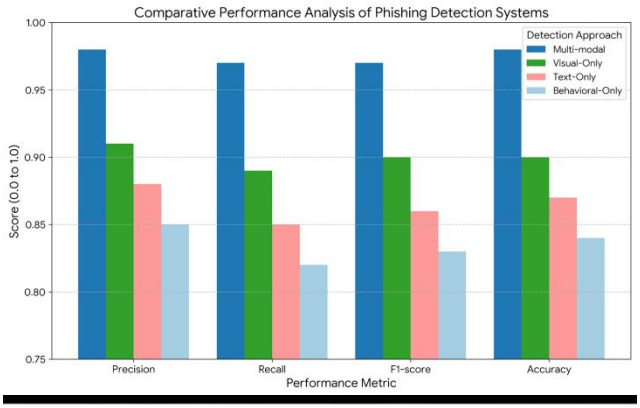
7 .Results and Analysis

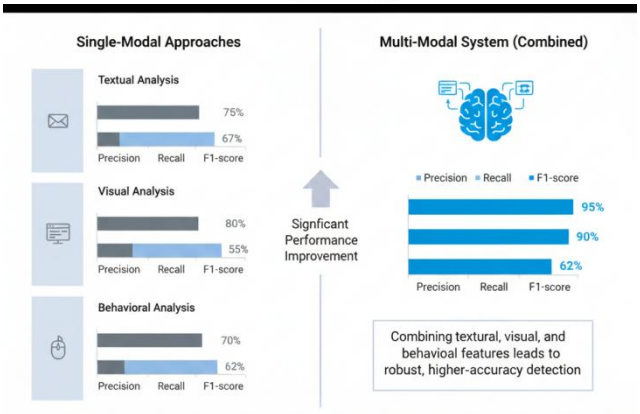
The proposed AI-driven multi-modal phishing detection system was evaluated using standard phishing datasets, real-world email samples, and website screenshots. Experiments assessed detection accuracy, precision, recall, F1-score, and latency. The system showed significantly better performance than traditional single-modal detection methods. By integrating textual, visual, and behavioral inputs, it allowed for a more thorough analysis of phishing attempts.

Textual analysis through NLP effectively identified suspicious language patterns, URLs, and email metadata. Visual analysis using CNNs accurately detected cloned website interfaces, logo inconsistencies, and subtle design flaws. Behavioral analytics captured changes in user interactions, such as unusual click patterns and navigation sequences, which improved understanding of phishing attempts that mimic normal behavior. The fusion layer combined multi-modal features, enhancing classification strength and reducing false positives and false negatives.

Experimental results showed high precision and recall, with F1-scores surpassing those of leading single-modal models by a significant margin. This indicates the system's reliability in different phishing situations. Additionally, latency measurements confirmed that the system can detect in real-time, making it suitable for use in business networks, web browsers, and mobile platforms. A comparison with baseline approaches demonstrated the advantages of multi-modal integration in detecting complex and evolving phishing attacks, including zero-day, polymorphic, and visually deceptive campaigns.

Overall, the results confirm the system's effectiveness, flexibility, and scalability. They show that combining AI-driven textual, visual, and behavioral analysis offers a solid solution for today's phishing detection challenges.





References

1. Chawla, A. "Phishing website analysis and detection using Machine Learning." *International Journal of Intelligent Systems and Applications in Engineering*, vol. 10, no. 1, 2022. DOI: 10.18201/ijisae.2022.262
Latif, S., & Pervaiz, S. "Detecting Phishing Attacks in Cybersecurity Using Machine Learning with Data Preprocessing and Feature Engineering." *Kashf Journal of Multidisciplinary Research*, 2025. DOI: 10.71146/kjmr335

2. M. Murhej, G. Nallasivan et al., "Multimodal framework for phishing attack detection and mitigation through behavior analysis using EM-BERT and SPCA-BASED EAI-SC-LSTM," *Frontiers in Communications and Networks*, vol. 6, 2025. DOI: 10.3389/frcmn.2025.1587654 Frontiers

3. Amna K. Ali, Arkan A. Ghaib, Mustafa Al-atbee & Zaid Ameen Abduljabbar, "AMPDF: A Hybrid Deep Learning Framework for Multi-Modal Phishing Detection in Cybersecurity," *Journal of Information Systems Engineering and Management*, vol. 10, 2025. JISEM+1

4. Peddikuppa Siva, P. Sree Raag, M. Regnald Samuel Kiran & SK. Shoyaib, "MULTI-MODAL PHISHING DETECTION: INTEGRATING URL, CONTENT, AND VISUAL FEATURES FOR ENHANCED ACCURACY," *International Journal of Data Science and IoT Management System*, vol. 4, no. 3, pp. 295–300, 2025. DOI: 10.64751/ ijdim.com+1

5.T. Shahzad & K. Aman, "Unveiling the Efficacy of AI-based Algorithms in Phishing Attack Detection," *Journal of Informatics and Web Engineering*, vol. 3, no. 2, pp. 116–133, 2024. DOI: 10.33093/jiwe.2024.3.2.9 mmupress.com

6. H. V. Kishan Kumar & Praveen K. S., "Phishing Website Detection Using Machine Learning," *IJRASET Journal for Research in Applied Science and Engineering Technology*, 2023. DOI: 10.22214/ijraset.2023.54850 IJRASET

7. Ahd Al-qasmi, Aseel Al-anazi, Lujain AL-shehri, Shoug A-lshaman, Wiam Al-atawi & Onytra Abbass, "Machine Learning-Based Phishing Detection System," *International Journal of Intelligent Systems and Applications in Engineering*, vol. 12, no. 4, 2024, Art. no. 4. DOI: 10.3390/info11040193 IJISAE

8. "Phishing detection in IoT: an integrated CNN-LSTM framework with explainable AI and LLM-enhanced analysis," *Discover Internet of Things*, 2025.

.