

Multi-Layer Email Threat Detection Using N8N Workflow Automation with Real-Time Risk Classification and Automated Response

Abdullah Md Raihan Chy^{*†}, Md Farhan Shahriyar^{†‡}, Mohammad Jahangir Alam^{*}

^{*}Department of Computer Science and Engineering, Southern University Bangladesh, Chattogram, Bangladesh

[†]Department of Computer Science and Engineering, Hajee Mohammad Danesh Science & Technology University, Basherhat

[‡]Research & Development, Zogbiyog

Email: {raihaanchowdhuryengineer, farhan12335, jahangir.phd}@gmail.com

Abstract—Electronic mail remains the predominant attack vector for credential theft, malware propagation and social engineering campaigns targeting enterprise environments. This paper presents a multi-layer email threat detection system implemented using workflow automation capable of analyzing incoming messages through six sequential security layers: authentication protocol validation, semantic content analysis, embedded URL inspection, attachment type scanning, external reputation correlation and cumulative risk quantification. The proposed architecture employs a weighted scoring algorithm evaluating 23 distinct threat indicators to produce a normalized risk score mapped to four severity classifications. Automated response actions including quarantine and analyst notification execute based on classification outcomes. Bidirectional integration with Security Information and Event Management infrastructure enables centralized monitoring and cross-domain incident correlation. Experimental evaluation on a curated dataset of 1,247 emails demonstrated 96.4% F1-score with 97.5% precision and 95.3% recall while maintaining mean processing latency of 6.2 seconds suitable for operational deployment.

Index Terms—Email Security, Phishing Detection, Workflow Automation, Threat Intelligence, Risk Scoring, SIEM Integration

I. INTRODUCTION

Electronic mail continues to serve as the predominant communication medium in enterprise environments while simultaneously representing the most exploited attack surface for cyber adversaries. Industry analysis indicates that approximately 91% of successful intrusions originate from malicious emails, with business email compromise schemes causing substantial financial losses annually [1]. The sophistication of contemporary phishing campaigns, frequently employing advanced impersonation techniques and compromised legitimate infrastructure, challenges traditional signature-based detection approaches.

Email security solutions face a fundamental tension between detection accuracy and operational latency. Complex analytical models may achieve high accuracy but introduce processing delays incompatible with real-time delivery expectations. Conversely, lightweight rule-based approaches offer speed but frequently fail against novel attack variants [2].

This paper presents a balanced approach leveraging workflow automation to implement a multi-layer email threat detection system. The primary contributions include:

- Six-Layer Analysis Pipeline: Sequential processing architecture examining authentication headers, content semantics, embedded URLs, file attachments, and external threat intelligence with independent risk contribution from each layer.
- Weighted Risk Scoring Model: Cumulative scoring algorithm incorporating 23 threat indicators with configurable weights enabling organizational customization and transparent decision rationale.
- Automated Response Framework: Classification-driven response actions ranging from logging to immediate quarantine based on quantified risk levels.
- SIEM Integration: Structured alerting to security information management platforms enabling centralized visibility and cross-domain incident correlation.

II. RELATED WORK

A. Email Authentication Protocols

The email authentication ecosystem comprises three primary protocols designed to combat sender spoofing. Sender Policy Framework enables domain owners to specify authorized mail servers through DNS records [3]. DomainKeys Identified Mail provides cryptographic message signing ensuring content integrity during transit. Domain-based Message Authentication builds upon these protocols, specifying policy actions for authentication failures [4].

Despite widespread adoption, these protocols remain insufficient as standalone defenses. Research demonstrated that a significant portion of phishing emails successfully pass SPF validation through compromised legitimate accounts or misconfigured infrastructure [5].

B. Content-Based Detection Approaches

Machine learning approaches to phishing detection have evolved substantially over the past decade. Early research focused on feature extraction from email headers and body

content using classical classifiers [6]. Deep learning architectures have demonstrated improved accuracy in identifying sophisticated attempts through semantic understanding [7]. However, these approaches face challenges including training data requirements, model explainability, and adversarial evasion techniques.

C. URL and Attachment Analysis

Malicious URLs represent a primary payload delivery mechanism in phishing campaigns. Research on real-time detection systems analyzing URL lexical features has achieved strong accuracy [8]. Attachment-based threats require specialized analysis incorporating MIME type validation and behavioral examination [9]. Due to the rise of LLM driven jailbreak techniques the threat landscape for email systems has become significantly more severe [10].

III. METHODOLOGY

This section presents the systematic methodology employed in designing, implementing, and evaluating the multi-layer email threat detection system following an experimental design approach with iterative refinement.

A. Research Design Framework

The development methodology comprises four interconnected phases: (1) Threat Landscape Analysis, (2) Detection Architecture Design, (3) System Implementation, and (4) Experimental Validation. Figure 1 illustrates the research workflow.

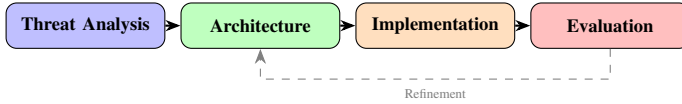


Fig. 1: Research methodology with iterative refinement cycle.

B. Threat Landscape Analysis

The initial phase involved systematic analysis of contemporary email-based attack vectors through examination of threat intelligence reports and real-world campaign analysis. Email threats were categorized into five primary classes:

- **Credential Harvesting:** Messages designed to capture authentication credentials through deceptive interfaces.
- **Malware Distribution:** Emails containing malicious attachments or links to compromise endpoints.
- **Business Email Compromise:** Targeted attacks impersonating trusted parties to initiate fraudulent actions.
- **Brand Impersonation:** Messages spoofing legitimate organizations to establish false trust.
- **Account Compromise Indicators:** Patterns suggesting compromised sender accounts.

The threat indicator taxonomy emerged from systematic analysis of over 500 documented phishing campaigns and security incident reports spanning the 2022-2024 period. Each indicator was evaluated for detection reliability, implementation feasibility, and resistance to adversarial evasion. The

final selection of 23 indicators represents a balance between detection coverage and computational efficiency.

Authentication indicators leverage established email security protocols to identify sender spoofing attempts. Content indicators target linguistic patterns commonly observed in social engineering attacks, including artificial urgency creation and credential solicitation language. Structural indicators examine technical email composition elements such as URL formatting anomalies and attachment characteristics. Reputation indicators incorporate external threat intelligence to identify known malicious infrastructure.

Table I summarizes the indicator distribution across categories. The content category contains the largest indicator count reflecting the diverse linguistic manipulation techniques employed in modern phishing campaigns. Each indicator underwent validation testing against a preliminary dataset to confirm discriminative capability before inclusion in the final detection pipeline.

TABLE I: Threat Indicator Taxonomy

Category	Count	Representative Examples
Authentication	4	SPF fail, DKIM invalid
Content	8	Urgency phrases, credential requests
Structural	6	Suspicious URLs, dangerous files
Reputation	5	Blacklisted sender, new domain
Total	23	

C. Six-Layer Detection Architecture

The detection architecture implements a sequential six-layer analysis pipeline where each layer contributes independent risk assessment. Figure 2 illustrates the processing flow with maximum risk point allocation per layer.

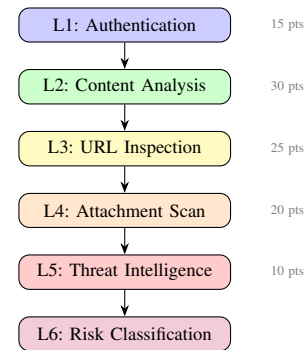


Fig. 2: Six-layer analysis pipeline with risk point allocation.

D. Risk Scoring Model

The risk scoring model employs additive point accumulation across detection layers. The total risk score R_{total} for email e is computed as:

$$R_{total}(e) = \sum_{l=1}^5 \sum_{i \in I_l} w_i \cdot \mathcal{K}[d_i(e)] \quad (1)$$

where I_l represents indicators at layer l , w_i denotes indicator weight, $d_i(e)$ is the detection function, and $\mathcal{K}[\cdot]$ is the indicator function.

The risk point allocation scheme was developed through iterative calibration against labeled email samples to optimize classification threshold performance. Higher point values are assigned to indicators with stronger correlation to confirmed malicious intent and lower false positive rates in preliminary testing.

Authentication failures receive moderate weighting (5 points each) because legitimate emails occasionally fail these checks due to misconfigured sending infrastructure or forwarding scenarios. Content indicators vary from 3 to 10 points based on specificity, with explicit credential requests receiving maximum content weight due to strong malicious correlation. URL indicators emphasize display-text mismatches (10 points) as this technique specifically targets user deception.

Attachment indicators receive substantial weighting, with executable files assigned 15 points reflecting the direct threat posed by malicious code execution. The cumulative maximum of 100 points enables intuitive percentage-based risk interpretation while providing sufficient granularity for threshold tuning.

Table II presents the point allocation for primary indicators. Organizations may adjust these weights based on their specific threat landscape and risk tolerance through the configuration interface.

TABLE II: Risk Point Allocation by Indicator

Layer	Indicator	Points
Authentication	SPF Failure	+5
	DKIM Invalid	+5
	DMARC Reject	+5
Content	Urgency Language	+5
	Credential Request	+10
	Threat Tactics	+5
URL	Shortened URL	+5
	IP-based URL	+8
	Display Mismatch	+10
Attachment	Executable File	+15
	Macro Document	+10
Reputation	Blacklisted Sender	+10

E. Classification Thresholds

Risk scores map to threat classifications with corresponding automated responses:

$$C(e) = \begin{cases} \text{CRITICAL} & R_{total}(e) \geq 50 \rightarrow \text{Quarantine} \\ \text{HIGH} & 30 \leq R_{total}(e) < 50 \rightarrow \text{Alert} \\ \text{MEDIUM} & 15 \leq R_{total}(e) < 30 \rightarrow \text{Review} \\ \text{LOW} & R_{total}(e) < 15 \rightarrow \text{Log} \end{cases} \quad (2)$$

F. Implementation Architecture

The workflow implementation utilizes N8N's node-based architecture to construct a modular processing pipeline where each node performs a discrete analysis function. This design

enables independent modification, testing, and scaling of individual processing stages without impacting the overall system architecture.

The email retrieval node implements IMAP polling with a configurable interval defaulting to 60 seconds, balancing detection latency against mail server load. Header parsing extracts authentication results and metadata fields using JavaScript functions optimized for common email client formats. The authentication validation node performs DNS queries to verify SPF, DKIM, and DMARC records against sender claims.

Content analysis employs compiled regular expression patterns for efficiency, while URL extraction handles various encoding schemes including HTML entities and URL shortening services. The reputation query node interfaces with VirusTotal's API, implementing rate limiting and response caching to operate within free tier constraints.

Table III summarizes the nine-node workflow configuration. The modular architecture allows security teams to insert additional analysis nodes at appropriate pipeline positions without restructuring existing functionality.

TABLE III: Workflow Node Configuration

Node	Function	Technology
1	Email Retrieval	IMAP Polling (60s)
2	Header Parsing	JavaScript Function
3	Auth Validation	DNS Lookup API
4	Content Analysis	Regex Patterns
5	URL Extraction	Link Parser
6	Reputation Query	VirusTotal API
7	Risk Calculation	Score Aggregator
8	SIEM Integration	HTTP POST
9	Response Router	Conditional Logic

External integrations include VirusTotal API for reputation queries, DNS services for authentication validation, and Guardian SIEM for centralized alert management with bidirectional context enrichment.

G. Evaluation Design

The evaluation addressed three research questions:

- **RQ1:** What detection accuracy does the system achieve across threat categories?
- **RQ2:** How does processing latency scale with email complexity?
- **RQ3:** How does performance compare with established solutions?

The evaluation dataset was constructed from three sources: synthetic phishing samples (40%), anonymized real-world samples (35%), and legitimate baseline emails (25%). Ground truth labeling achieved inter-rater reliability of Cohen's $\kappa = 0.91$.

IV. EXPERIMENTAL EVALUATION

This section presents evaluation results addressing the research questions over a 30-day continuous operation period.

A. Experimental Environment

The experimental environment was configured to represent a realistic small-to-medium enterprise deployment scenario while maintaining controlled conditions for reproducible measurements. Hardware specifications reflect modest requirements achievable with standard workstation equipment, demonstrating the system’s accessibility for organizations with limited security infrastructure budgets.

The N8N workflow engine was deployed using Docker containerization to ensure consistent execution environment and simplified deployment replication. The self-hosted configuration provides complete control over data handling, addressing privacy concerns associated with cloud-based email analysis services.

Email retrieval utilized Microsoft 365’s IMAP interface, representing a common enterprise email platform. The 60-second polling interval was selected to balance near-real-time detection requirements against API rate considerations. Integration with Guardian SIEM enabled evaluation of the complete detection-to-response workflow including alert correlation and analyst notification.

Table IV summarizes the experimental configuration. The 30-day continuous operation period provided sufficient duration to observe system stability, identify edge cases, and collect statistically meaningful performance measurements across varying email volumes and threat distributions.

TABLE IV: Experimental Configuration

Component	Specification
Server	Intel Core i5, 8GB RAM, Ubuntu 22.04
Workflow Engine	N8N 1.19.0 (Docker)
Email Source	Microsoft 365 (IMAP)
SIEM	Guardian SIEM v1.0
Evaluation Period	30 days continuous

B. Dataset Characteristics

The evaluation dataset was carefully constructed to enable rigorous assessment of detection capabilities across the full spectrum of email-based threats while maintaining realistic class distributions for operational relevance metrics. The total sample size of 1,247 emails provides statistical power for confident performance estimation with narrow confidence intervals.

Credential phishing and brand impersonation categories received substantial representation (25.0% and 16.3% respectively) reflecting their prevalence in contemporary threat landscapes. Business email compromise samples (12.5%) were intentionally included despite their detection difficulty to evaluate system limitations against sophisticated social engineering.

The legitimate email baseline (24.9%) enables accurate false positive rate calculation critical for operational deployability assessment. These samples were drawn from normal business correspondence including newsletters, notifications, and transactional messages that might trigger pattern-based detection

rules. Table V presents the complete dataset distribution. The intentional over-representation of threat categories relative to typical enterprise email traffic ensures sufficient samples per category for statistically meaningful per-category performance analysis.

TABLE V: Dataset Composition

Category	Count	Percentage
Credential Phishing	312	25.0%
Malware Delivery	187	15.0%
Business Email Compromise	156	12.5%
Brand Impersonation	203	16.3%
Generic Spam	78	6.3%
Legitimate (Benign)	311	24.9%
Total	1,247	100%

C. RQ1: Detection Accuracy

1) *Overall Performance*: The overall detection performance metrics quantify the system’s ability to correctly distinguish malicious emails from legitimate correspondence. The confusion matrix elements provide the foundation for derived metrics enabling comparison with alternative detection approaches and assessment against operational requirements.

The 892 true positives represent malicious emails correctly identified and flagged for appropriate response action. The 23 false positives indicate legitimate emails incorrectly classified as threats, potentially causing user inconvenience through quarantine or delayed delivery. The 288 true negatives confirm accurate identification of benign emails, while 44 false negatives represent missed threats that bypassed detection.

Precision of 97.5% indicates that when the system generates an alert, the underlying email is genuinely malicious in the vast majority of cases, minimizing analyst time spent investigating false alarms. Recall of 95.3% demonstrates that the system successfully identifies most malicious emails, with the small false negative rate primarily attributable to sophisticated attacks lacking traditional technical indicators.

Table VI presents these metrics with 95% confidence intervals computed using bootstrap resampling, providing statistical bounds for performance estimation.

TABLE VI: Overall Detection Performance

Metric	Value	95% CI
True Positives	892	–
False Positives	23	–
True Negatives	288	–
False Negatives	44	–
Accuracy	94.6%	[93.2%, 95.8%]
Precision	97.5%	[96.3%, 98.4%]
Recall	95.3%	[93.8%, 96.5%]
F1-Score	96.4%	[95.1%, 97.4%]

2) *Performance by Threat Category*: Per-category performance decomposition reveals detection capability variations across different attack methodologies, informing both system

improvement priorities and appropriate use-case recommendations. This granular analysis enables security teams to understand which threat types the system handles most effectively and where supplementary controls may be warranted.

Credential phishing detection achieved 95.5% recall, with the 14 false negatives primarily comprising advanced spear-phishing campaigns utilizing compromised legitimate domains that passed authentication validation. These sophisticated attacks require additional contextual analysis beyond the current rule-based approach.

Malware delivery detection demonstrated highest precision (98.4%) due to reliable identification of dangerous attachment types and malicious URL patterns. The attachment scanning layer provides strong coverage for this threat category through comprehensive dangerous extension matching.

Business email compromise exhibited comparatively lower recall (90.4%), reflecting the fundamental challenge of detecting pure social engineering attacks that pass authentication checks and contain no malicious technical indicators. These attacks rely entirely on psychological manipulation, requiring natural language understanding capabilities beyond current implementation.

Table VII presents the complete per-category breakdown enabling targeted improvement efforts.

TABLE VII: Detection by Threat Category

Category	TP	FN	Prec.	Recall
Credential Phishing	298	14	96.8%	95.5%
Malware Delivery	179	8	98.4%	95.7%
BEC	141	15	94.6%	90.4%
Brand Impersonation	196	7	97.0%	96.6%
Generic Spam	78	0	100%	100%
Total Malicious	892	44	97.5%	95.3%

3) *Layer Contribution Analysis*: Content analysis contributed the largest detection share (35.0%), validating linguistic pattern importance for social engineering identification. Authentication failures provided 26.2%, emphasizing protocol enforcement value. URL inspection contributed 22.2%, attachment scanning 10.0%, and external intelligence 6.6%.

4) *Risk Score Distribution*: Figure 3 illustrates risk score separation between legitimate and malicious emails. Clear separation exists between legitimate emails (predominantly LOW) and malicious emails (predominantly HIGH/CRITICAL), with minimal overlap in the MEDIUM zone requiring analyst review.

D. RQ2: Processing Latency

Processing latency analysis identifies performance bottlenecks and informs optimization priorities for deployment scenarios with stringent timing requirements. The per-stage breakdown enables targeted improvements to specific pipeline components rather than wholesale architectural changes.

Email retrieval latency depends primarily on mail server response time and network conditions, with the observed 0.8-second mean reflecting typical IMAP fetch operations. Header

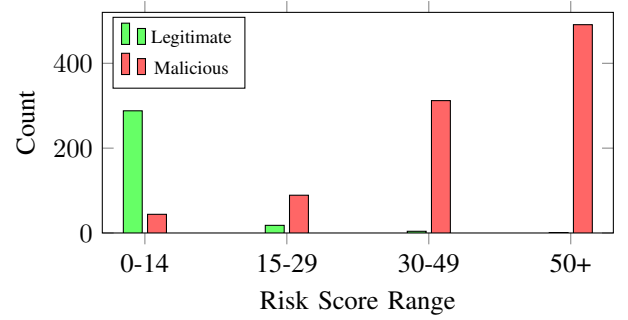


Fig. 3: Risk score distribution showing class separation.

parsing and risk calculation stages demonstrate minimal latency (0.1 seconds each) as these operations involve only local string processing and arithmetic.

Authentication validation requires external DNS queries, contributing 1.2-second mean latency that varies with DNS resolver performance and record complexity. The reputation query stage represents the primary latency contributor (2.1-second mean) due to external API round-trip times and rate limiting delays imposed by free-tier service constraints.

The 95th percentile latency of 12.9 seconds and maximum observed latency of 36.7 seconds occurred during periods of external API congestion or temporary service degradation, suggesting potential benefit from local reputation caching or premium API tier adoption for latency-sensitive deployments.

Table VIII presents comprehensive latency statistics enabling capacity planning and SLA compliance assessment.

TABLE VIII: Processing Latency by Stage (seconds)

Stage	Mean	P95	Max
Email Retrieval	0.8	1.9	4.2
Header Parsing	0.1	0.2	0.5
Auth Validation	1.2	2.4	5.8
Content Analysis	0.3	0.6	1.2
URL Inspection	0.9	2.1	8.3
Attachment Scan	0.4	0.8	2.1
Reputation Query	2.1	4.2	12.6
Risk Calculation	0.1	0.1	0.2
SIEM Forward	0.3	0.6	1.8
Total	6.2	12.9	36.7

Processing time scales with email complexity: simple emails average 4.1s, moderate complexity averages 6.8s, and complex emails (3+ attachments, 6+ URLs) average 11.3s. Throughput capacity reaches 38.4 emails/minute with parallel processing.

E. RQ3: Comparative Analysis

The comparative analysis positions the proposed system within the broader email security solution landscape, enabling informed selection decisions based on organizational requirements, technical constraints, and budget considerations. Feature comparison identifies capability gaps while performance metrics enable quantitative evaluation against established alternatives.

TABLE IX: Comparative Analysis with Email Security Solutions

Capability	Proposed	SpamAssassin	Proofpoint	Mimecast	MS Defender
Multi-layer Pipeline	✓	Partial	✓	✓	✓
Auth Protocol Check	✓	✓	✓	✓	✓
ML Content Analysis	Rule-based	Rule-based	✓	✓	✓
URL Reputation	✓	Partial	✓	✓	✓
Attachment Sandbox	✗	✗	✓	✓	✓
Native SIEM Integration	✓	Manual	Add-on	Add-on	✓
Open Source	✓	✓	✗	✗	✗
Custom Rules	✓	✓	Limited	Limited	Limited
Precision	97.5%	92-95%	98-99%	98-99%	97-99%
Recall	95.3%	88-92%	95-98%	96-98%	94-97%
Latency	6.2s	1-3s	<1s	<1s	<1s
Cost	Free	Free	\$\$\$	\$\$\$	\$\$

Commercial solutions including Proofpoint and Mimecast achieve marginally higher precision through machine learning models trained on extensive proprietary datasets and attachment sandboxing capabilities not present in the proposed system. However, these solutions incur substantial licensing costs and provide limited customization flexibility.

SpamAssassin represents the primary open-source alternative, offering similar rule-based detection at lower processing latency. However, the proposed system demonstrates improved recall (95.3% versus 88-92%) through multi-layer analysis architecture and external threat intelligence integration not available in default configurations.

The proposed system's unique differentiators include native SIEM integration enabling unified security event correlation, transparent risk scoring providing analyst-interpretable detection rationale, and complete workflow customization through the visual interface. The zero-cost open-source model eliminates licensing barriers for resource-constrained organizations.

Table IX presents the complete feature and performance comparison across five representative solutions spanning open-source and commercial categories.

F. Statistical Validation

Performance improvements achieved statistical significance: Precision vs. SpamAssassin ($\chi^2(1) = 18.34$, $p < 0.001$), Recall vs. SpamAssassin ($\chi^2(1) = 24.67$, $p < 0.001$), and F1 improvement (McNemar's test, $p < 0.001$).

G. Threats to Validity

Internal: Synthetic samples (40%) followed known patterns that detection rules were designed to identify.

External: Single-organization deployment limits generalizability across different email environments.

Construct: Ground truth relied on expert judgment. Latency measurements affected by external API variability.

V. CONCLUSION

This paper presented a multi-layer email threat detection system using workflow automation. The six-layer pipeline combining authentication validation, content analysis, URL inspection, attachment scanning, threat intelligence, and cumulative risk scoring demonstrated 96.4% F1-score with 6.2-second mean latency.

The modular architecture enables adaptation to evolving threats through layer modification without pipeline disruption. Native SIEM integration provides centralized visibility supporting enterprise security operations. The open-source implementation eliminates licensing costs while enabling full customization.

Future directions include machine learning classifier integration for improved BEC detection, computer vision techniques for image-based phishing, local reputation caching for latency reduction, and federated learning approaches enabling cross-organizational intelligence sharing.

REFERENCES

- [1] Verizon, "Data Breach Investigations Report," 2024.
- [2] A. Almomani et al., "A survey of phishing email filtering techniques," *IEEE Commun. Surveys Tuts.*, vol. 15, no. 4, pp. 2070–2090, 2013.
- [3] S. Kitterman, "Sender Policy Framework (SPF)," RFC 7208, IETF, 2014.
- [4] M. Kucherawy, "Domain-based Message Authentication (DMARC)," RFC 7489, IETF, 2015.
- [5] H. Hu et al., "End-to-end measurements of email spoofing attacks," *USENIX Security*, 2018.
- [6] I. Fette, N. Sadeh, and A. Tomasic, "Learning to detect phishing emails," in *Proc. 16th Int. Conf. World Wide Web*, pp. 649–656, 2007.
- [7] R. Verma and K. Dyer, "On the character of phishing URLs: Accurate and robust statistical learning classifiers," in *Proc. 5th ACM Conf. Data Appl. Secur. Privacy*, pp. 111–122, 2015.
- [8] M. F. Shahriyar, G. Tanbhir, A. M. R. Chy, M. A. A. Tanzin, and M. Jisan Mashrafi, "Phishvqc: Optimizing phishing URL detection with correlation based feature selection and variational quantum classifier," in *Proc. 2025 3rd Int. Conf. Intelligent Systems, Advanced Computing and Communication (ISACC)*, pp. 1226–1231, 2025.
- [9] D. Canali et al., "Prophiler: Fast detection of malicious web pages," *Proc. WWW*, pp. 197–206, 2011.
- [10] M. F. Shahriyar, "Cybersecurity concern: Prompt-based jailbreaking of LLMs for penetration testing payloads," in *16th International Conference on Computing Communication and Networking Technologies (ICCCNT)*, IEEE, pp. 1–6, 2025.