International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 11 , November - 2026

Multimodal Detection and Analysis of Impersonation-Based Cyber Scams

Mayur Dongare[1], Om Charthal[2], Vaishnavi Wankhade[3], Shakshi Sharma[4], Swarali Jadhav[5], Prof. Dr. P. S. Deshmukh

Department of Computer Science & Engineering,
Prof. Ram Meghe Institute of Technology & Research, Amravati-Badnera-444701 (MS), Amravati, Maharashtra, India.

Abstract - Impersonation-based cyber scams exploit the trust associated with legitimate individuals, organizations, brands, and public institutions. Such attacks are distributed through multiple digital channels and may combine deceptive messages, suspicious URLs, copied visual identities, QR codes, documents, and video-based content. Detection systems that examine only one evidence type can overlook cross-modal indicators or fail to explain why an input appears risky. This paper presents a proposed multimodal framework for detecting and analyzing impersonation-based cyber scams through modality-specific processing and evidence-aware risk assessment. Text and email content will be analyzed using contextual language models such as BERT; URL characteristics will be evaluated using structured features and an XGBoost-based classifier; image and screenshot evidence will be examined through OCR and visual similarity analysis; and QR codes and documents will be decoded or parsed to extract actionable text and links. Optional video processing will use bounded frame sampling and, where feasible, speech transcription. The resulting modality-level outputs will be combined through a calibrated evidence-fusion stage that considers supporting indicators, contradictions, unavailable modalities, and uncertainty. The system is intended to produce a risk band, supporting evidence, and safe recommendations rather than a definitive legal or factual judgment. Since the project is proposed and implementation/evaluation are ongoing, this paper defines the research problem, framework, methodology, and evaluation protocol; it does not report fabricated experimental results.

Keyword - Impersonation-Based Scams, Multimodal Cybersecurity, Phishing Detection, BERT, XGBoost, OCR, Visual Similarity, QR-Code Analysis, Evidence Fusion, Explainable AI.

I. INTRODUCTION

Digital communication has made it easier for attackers to imitate trusted people, organizations, financial services, brands, and public agencies. Impersonation-based scams exploit this trust to persuade recipients to disclose credentials, transfer money, open malicious links, install software, or provide sensitive information. The impersonation may appear in a short message, email, webpage, screenshot, QR code, document, or video, and a single incident may contain several of these formats.

Phishing research has developed from content- and URL-based classifiers toward approaches that incorporate visual similarity, webpage behavior, contextual knowledge, and multimodal models [1], [2], [4], [7]. Email and message classification remain important because linguistic patterns can reveal urgency, threats, requests for payment, and credential harvesting [5], [6], [9]. URL analysis contributes complementary signals from domain structure, path tokens, and other lexical or metadata features [7], [10], [11]. Visual approaches can identify webpages that imitate legitimate brands or interfaces, while multimodal methods seek to combine text, images, and contextual information [2], [4], [12].

Despite this progress, scam evidence is often fragmented. A message may contain no suspicious URL while an attached QR code redirects to a deceptive destination. A screenshot may imitate a trusted organization but include a different sender or payment request. Conversely, a suspicious-looking feature may be benign when examined in context. This motivates a framework that analyzes available modalities separately, preserves the evidence produced by each component, and combines the outputs without treating missing evidence as proof of safety.

This paper proposes a multimodal detection and analysis framework for impersonation-based cyber

(This work is licensed under a Creative Commons Attribution 4.0 International License.)

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 11 , November - 2026

scams. Its contribution is the structured integration of text, URL, visual, QR/document, and optional video evidence into an explainable risk assessment. The framework emphasizes evidence traceability, uncertainty handling, and safe output presentation. It is a research proposal and implementation plan; empirical claims will be added only after reproducible experiments are completed.

A. Research Problem

How can heterogeneous evidence associated with impersonation-based cyber scams be processed and combined to produce a useful, explainable risk assessment while accounting for conflicting signals, missing modalities, and limitations of individual detectors?

B. Research Objectives

- Develop a unified input workflow for text, email, URLs, images/screenshots, QR codes, documents, and optional video.
- Apply modality-appropriate extraction and analysis methods, including contextual text classification, structured URL classification, OCR, visual similarity, QR decoding, and document parsing.
- Design an evidence-fusion mechanism that retains the provenance and confidence of individual outputs and handles missing or contradictory evidence.
- Generate an explainable risk band, supporting indicators, uncertainty notes, and safe user recommendations.
- Evaluate component performance, fusion behavior, false-positive and false-negative patterns, computational cost, and robustness through a reproducible protocol.

II. RELATED WORK AND BACKGROUND

A. Text and Email-Based Detection

Email phishing detection has been studied using linguistic analysis and deep learning, with research examining message content, semantic patterns, and classification pipelines [5], [9], [17], [19]. Smishing detection extends this problem to short mobile messages, where limited context, abbreviated language, and urgent calls to action complicate classification [6]. Transformer-based language models provide contextual representations that may capture relationships beyond keyword counts; however, model output can still be sensitive to domain shift, class imbalance, and unfamiliar scam phrasing. The proposed framework therefore treats text analysis as one evidence source rather than the complete decision.

B. URL and Webpage Analysis

URL detection commonly relies on lexical, host, path, and structural properties. Machine-learning methods and explainability techniques have been used to classify suspicious URLs and interpret influential features [7], [10], [11]. Website-level approaches also examine global and local visual similarity, including brand or webpage appearance [12]. These approaches motivate separate URL and visual branches in the proposed framework. A URL classifier's prediction will not be interpreted as a guarantee that a destination is malicious or safe; reputation signals, if included, will be identified as external evidence and evaluated for freshness and reliability.

C. Visual and Multimodal Detection

Visual phishing research demonstrates the value of comparing webpage appearance and branding cues, including hybrid deep-learning methods [12]. Recent multimodal phishing systems combine webpage text, visual content, and contextual knowledge, illustrating the potential benefit of analyzing multiple evidence types [2]–[4]. Yet multimodal systems introduce practical concerns: modality alignment, data availability, computational cost, and the risk of allowing one unreliable component to dominate the final decision. The proposed design retains per-modality outputs and explicit uncertainty so that the fusion stage can be audited.

D. Research Gap and Proposed Position

The reviewed literature supports text, URL, and visual detection as useful but complementary directions [2], [7], [9]. The project focuses on the practical evidence chain across multiple user-submitted formats: extraction, modality-specific inference, evidence provenance, fusion, and user-facing explanation. In contrast to a single classifier output, the intended result records what was observed, which component produced it, whether evidence was unavailable, and how the combined assessment was formed. This paper does not claim that the proposed combination is novel in every individual component; its research contribution will be evaluated as an integrated and uncertainty-aware workflow.

(This work is licensed under a Creative Commons Attribution 4.0 International License.)

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 11 , November - 2026

III. PROPOSED MODEL

A. Objective and Scope

The objective is to design and implement a multimodal system that assists users in examining possible impersonation-based scams. The system will accept evidence in several formats and route it to relevant extraction and analysis components. The scope includes text/email analysis, URL feature analysis, image and screenshot OCR/visual analysis, QR decoding, document text/link extraction, and optional video frame or transcript analysis. The system will provide an advisory risk assessment with traceable supporting evidence.

The framework will not automatically open decoded QR destinations, execute embedded files, submit credentials, or declare a person or organization fraudulent. Destination retrieval, if implemented, will use controlled and security-conscious processing. Any external reputation or verification service will be treated as an additional source with its own provenance and limitations.

B. Framework Components

The overall architecture of the proposed framework is shown in Fig. 1. Submitted evidence passes through input validation and preprocessing, is routed to modality-specific analysis modules, and is then combined in a cross-modal evidence integration layer that feeds the results dashboard and the optional storage layer. The individual components are described below.

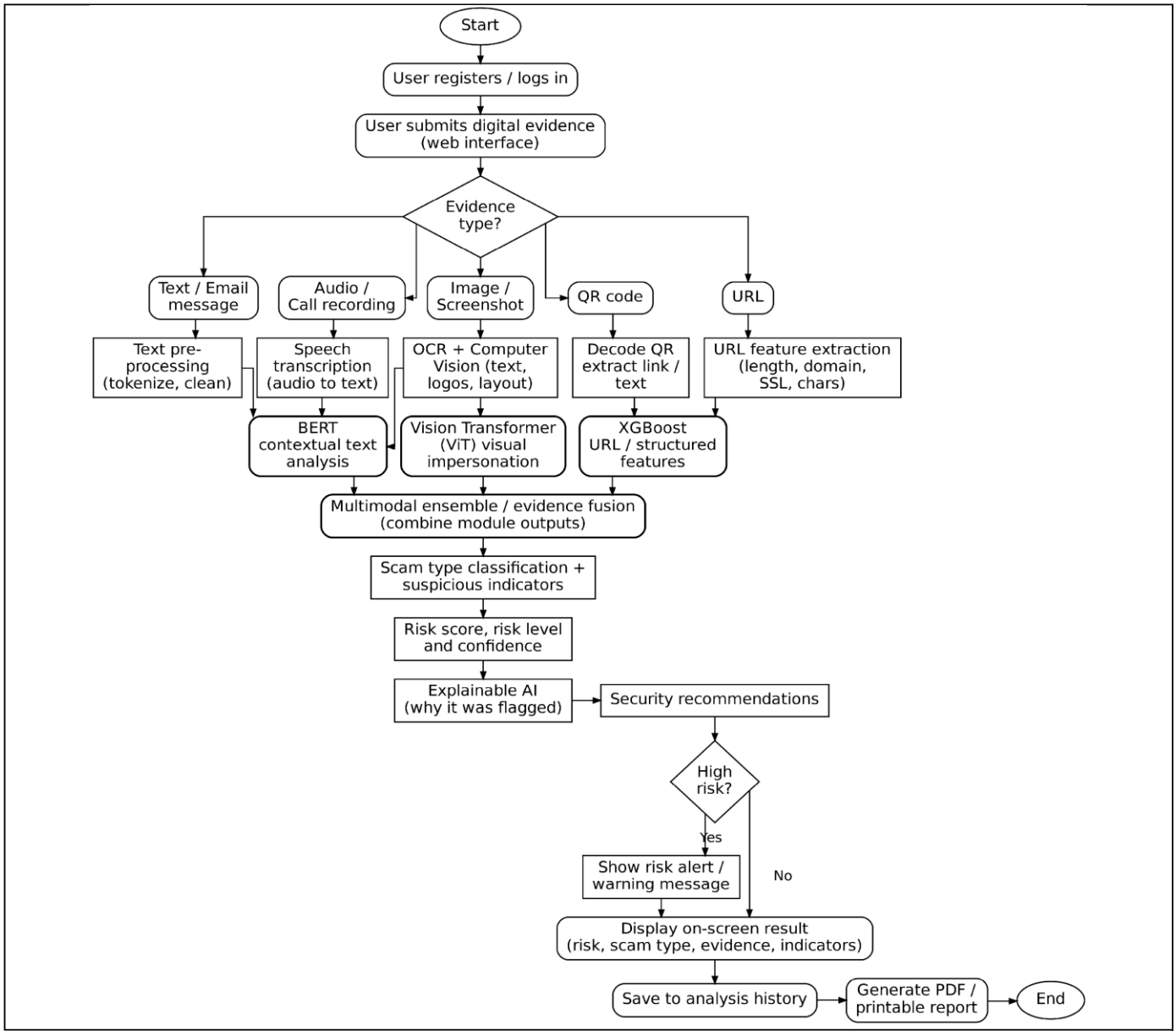


Fig. 1. System architecture of the proposed multimodal impersonation-scam detection framework.

(This work is licensed under a Creative Commons Attribution 4.0 International License.)

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 11 , November - 2026

Text and Email Analyzer: Normalize the submitted message and extract linguistic/contextual features. A BERT-based classifier is proposed for message-level classification. The output will include a predicted class or risk score, confidence, and relevant textual indicators.

URL Analyzer: Parse the URL into host, subdomain, path, query, and lexical/structural features. An XGBoost classifier is proposed for structured feature analysis. The module will preserve the original URL and extracted feature values for explanation.

Image and Screenshot Analyzer: Use OCR to extract visible text and a visual-analysis component to examine branding, layout, or similarity cues. A Vision Transformer (ViT) or suitable visual embedding model is a candidate for visual representation; its feasibility and performance will be assessed before final selection.

QR-Code Analyzer: Decode QR payloads without automatically visiting the destination. Classify decoded content as URL, text, or another payload type, then route URLs or text to the corresponding analyzer. Decoder errors and unsupported payloads will be reported.

Document Analyzer: Extract text, hyperlinks, and metadata from supported document formats. For scanned pages, OCR will be used where appropriate. Embedded URLs will be routed to URL analysis, while extracted text will be routed to the text analyzer.

Optional Video Analyzer: Sample a bounded number of frames and optionally extract speech using a speech-to-text component. Frame and transcript analysis will be treated as optional due to processing cost, temporal coverage limits, and implementation priority.

C. Evidence Fusion and Risk Assessment

Each active analyzer will return a structured result containing the component identifier, input span or artifact reference, prediction or extracted indicator, confidence where available, processing status, and explanation. The fusion stage will combine available results using a validation-calibrated strategy. Candidate approaches include weighted score aggregation or a trained fusion model, subject to adequate labeled multimodal data. Weights and decision thresholds will not be fixed from intuition alone; they will be selected using training/validation data and evaluated on held-out samples.

Fig. 2 illustrates how indicators are correlated across modalities. Textual, URL, visual, and identity signals are extracted from the submitted evidence and compared. Where multiple signals support the same risk hypothesis they are combined directly; otherwise contradictions, missing evidence, and uncertainty are assessed first, and the combined evidence yields a calibrated risk assessment with explainable findings and recommended safer actions.

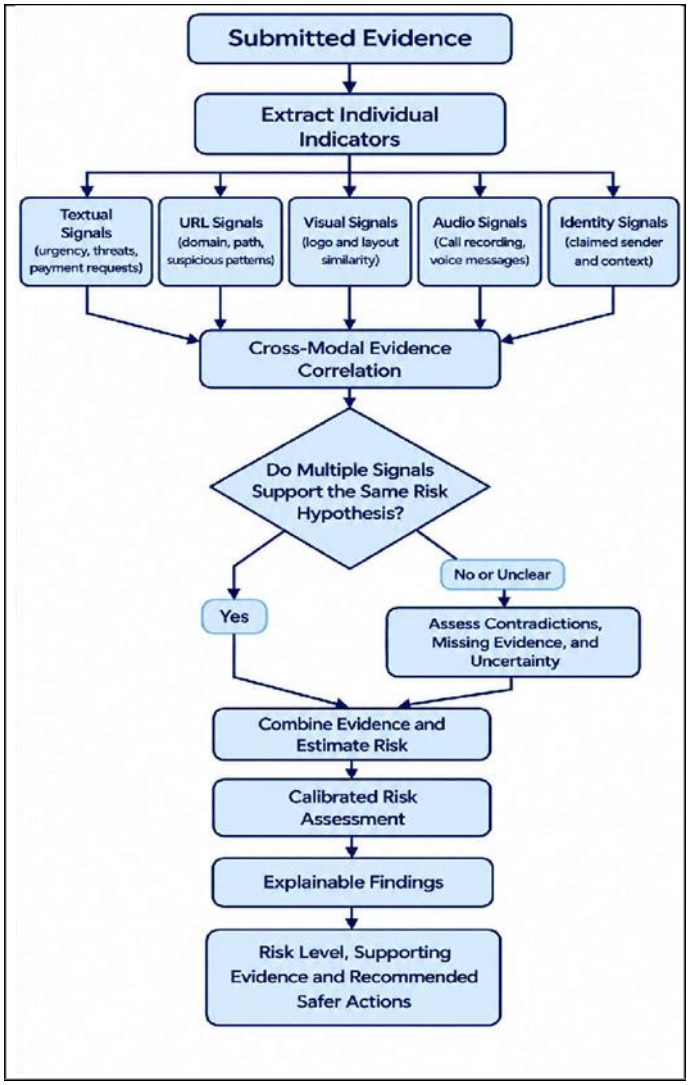


Fig. 2. Cross-modal evidence correlation and risk assessment.

Missing modalities will be represented as unavailable rather than as benign evidence. Contradictory signals will be retained and surfaced in the explanation. The final output will include a risk band, supporting indicators, contradictory or missing evidence, and safe actions. The risk score will be advisory and will not be presented as a calibrated probability unless calibration is empirically demonstrated.

D. Expected Output

- Overall risk band and score, if calibration supports score interpretation.
- Detected scam indicators and the source modality for each indicator.

(This work is licensed under a Creative Commons Attribution 4.0 International License.)

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 11 , November - 2026

- Extracted text, decoded QR payload, or URL features relevant to the assessment.
- Uncertainty notes, unsupported input types, and conflicting signals.
- Safe next steps such as avoiding credential entry, verifying through an independently obtained official channel, and not transferring money based solely on the message.

IV. METHODOLOGY

A. Data Collection and Annotation

The project will require a dataset containing scam and benign examples across supported modalities. Candidate sources include publicly available phishing URL datasets, phishing email/message corpora, screenshot or webpage-image datasets, and safely collected synthetic or manually curated examples. Dataset selection will depend on license, availability, label quality, and compatibility with the project's threat scope. The paper does not assert a final dataset size or class distribution because these have not yet been established.

Records will be annotated using a consistent label scheme, such as scam/suspicious versus benign, with more specific categories where reliable labels exist (for example, brand impersonation, credential theft, payment request, or delivery-related scam). Annotation guidelines will distinguish observed evidence from inferred intent. Duplicate detection, source overlap, and temporal leakage will be checked before splitting data.

B. Preprocessing and Feature Extraction

- **Text:** preserve the original input; normalize encoding and whitespace; identify language where relevant; tokenize according to the chosen BERT tokenizer; and retain message-level context such as sender claims and request type when available.
- **URL:** parse scheme, host, subdomain, path, query, length, suspicious token patterns, and selected structural features. Sensitive query values will be removed or masked before storage.
- **Image:** run OCR to obtain text regions and confidence; optionally extract visual embeddings or similarity features. OCR output will retain page/region provenance where feasible.
- **QR/document:** decode or parse in a sandboxed workflow; record payload type and extracted content; route text and URLs to their respective analyzers.
- **Audio:** sample frames under a defined cap and optionally transcribe speech; record timestamps and sampling policy so limited temporal coverage is explicit.

C. Model Training and Inference

The text branch will evaluate a BERT-family model against a suitable baseline. The URL branch will train XGBoost on structured features and compare it with simple baseline classifiers. Visual analysis will compare a pretrained or fine-tuned vision representation against a simpler baseline if labeled images are available. The exact model variants, training schedule, and hyperparameters will be selected after data inspection and resource profiling.

Training will use a reproducible split strategy. Where samples originate from related campaigns, domains, brands, or near-duplicate messages, group-aware splitting will be considered to reduce leakage. A validation set will support hyperparameter selection and threshold calibration; the test set will be held out until final evaluation. If dataset size is insufficient for fine-tuning, pretrained inference or a reduced experimental scope will be documented.

D. Fusion Procedure

- Collect outputs from all available modality-specific analyzers and mark failed or unsupported modules explicitly.
- Normalize component scores only where their meaning and scale are documented; retain raw component outputs.
- Combine evidence using a validation-selected fusion strategy and account for modality availability.
- Apply thresholds selected on validation data; include an uncertainty/abstention state for low-confidence or conflicting cases.
- Generate an explanation that links each risk indicator to its originating input and analyzer.

E. Mathematical Formulation

Let the input evidence be $X = \{x_1, x_2, ..., x_m\}$, where each x_i is an available modality. A modality-specific analyzer produces an output $z_i = f_i(x_i)$, which may include a score, label, and extracted indicators. Let $a_i \in \{0,1\}$ denote whether modality i is available and successfully processed. A candidate weighted fusion score is:

$$S(X) = (\Sigma\ a_i\ w_i\ s_i) / (\Sigma\ a_i\ w_i)$$
$$\text{when } \Sigma\ a_i\ w_i > 0$$

where s_i is a compatible modality score and w_i is a weight selected using validation data. This expression is a candidate

(This work is licensed under a Creative Commons Attribution 4.0 International License.)

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 11 , November - 2026

formulation, not a finalized or calibrated model. If no valid modality score is available, the system should return an insufficient-evidence state rather than a low-risk score. Contradiction indicators and confidence values may be incorporated into a learned fusion model after sufficient labeled data are obtained.

V. IMPLEMENTATION PLAN

A. Planned Software Components

- **Backend and orchestration:** Python API service for input validation, module routing, inference orchestration, and result serialization. The final framework choice will be confirmed during implementation.
- **Text analysis:** Transformers-based BERT inference/training pipeline with tokenizer and preprocessing.
- **URL analysis:** Python URL parsing and feature extraction with XGBoost classifier.
- **Image analysis:** OCR library plus candidate ViT/visual embedding model; compare implementation feasibility and latency.
- **QR and document processing:** QR decoder and format-specific text/link parsers with OCR fallback for scans.
- **Frontend:** a user interface for submitting evidence, viewing module outputs, risk assessment, explanation, and safe recommendations.
- **Storage and security:** minimize retained personal data, restrict access, avoid logging secrets, and store only necessary analysis history.

B. Processing Workflow

The user submits one or more evidence items. The application validates file type and size, identifies available modalities, and performs extraction. Extracted text and URLs are routed to their corresponding analyzers. Each module returns a structured result with provenance and status. The fusion stage evaluates the available evidence and produces a risk assessment. Finally, the interface presents the result with supporting indicators, limitations, and safe actions. Processing failures will be displayed as failures or unavailable evidence rather than silently treated as negative detections.

C. Security and Privacy Considerations

Because inputs may contain personal information, credentials, financial details, or malicious files, the implementation will apply input validation, file-size limits, safe parsing, access control, and data minimization. Uploaded content will not be executed. QR destinations and URLs will not be opened automatically. Any URL fetching feature will require controlled retrieval, protection against server-side request forgery, redirect limits, and isolation from internal network resources. Logs and stored results will avoid retaining unnecessary secrets.

VI. EXPERIMENTAL SETUP AND EVALUATION PLAN

A. Evaluation Metrics

The system will be evaluated using accuracy, precision, recall, F1-score, confusion matrices, and ROC-AUC or PR-AUC where appropriate. Because scam detection can involve class imbalance and asymmetric costs, precision-recall behavior and false-positive/false-negative rates will be reported alongside accuracy. Calibration metrics and reliability plots will be used if scores are presented as probabilities. Runtime, memory use, and modality coverage will also be measured.

B. Component-Level Evaluation

- **Text branch:** compare BERT with a baseline such as TF-IDF plus Logistic Regression, using identical data splits.
- **URL branch:** compare XGBoost with baseline classifiers and report performance by URL feature group where feasible.
- **Visual branch:** evaluate OCR text extraction separately from visual similarity classification; report OCR failures and image-quality effects.
- **QR/document branch:** measure decoding or extraction success rate and downstream classification performance on extracted content.
- **Video branch:** if implemented, evaluate sampled-frame coverage, transcript quality where labels exist, and incremental performance relative to non-video inputs.

C. Multimodal and Ablation Evaluation

The multimodal evaluation will compare individual modality results with fusion results on examples for which multiple evidence types are available. Ablations will remove one modality at a time, disable OCR, disable visual similarity, or compare fusion strategies. These experiments will test whether each component contributes measurable value and whether fusion improves detection without unacceptable false-positive increases. Missing-modality and contradictory-evidence subsets will be analyzed separately.

(This work is licensed under a Creative Commons Attribution 4.0 International License.)

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 11 , November - 2026

D. Generalization and Robustness

Where dataset structure permits, evaluation will use domain-, campaign-, brand-, or time-separated splits to test generalization beyond near-duplicate samples. Robustness checks may include obfuscated URLs, spelling variation, image resizing/compression, OCR noise, and benign messages containing urgency or financial terminology. Results will report limitations and failure cases, not only aggregate metrics.

E. Results Status

At the time of writing, the project's final dataset, completed training runs, measured performance values, and ablation outcomes have not been supplied. Therefore, this manuscript intentionally does not provide numerical results or claim that a particular model achieved a specific accuracy. The Results section should be completed after experiments are run, with the test protocol fixed and raw outputs preserved.

VII. RESULT

Table I reports the classification performance of each analysis subsystem and of the overall multimodal risk fusion engine, evaluated using five-fold cross-validation on an empirical test set of N = 2,500 instances. The reported measures are accuracy, precision, recall, specificity, F1-score, AUC-ROC, and average latency.

TABLE I. MODEL PERFORMANCE AND EVALUATION CRITERIA

Architecture / Subsystem	Accuracy	Precision	Recall	Specificity	F1-Score	AUC-ROC	Avg Latency
BERT Transformer *(Text & SMS)*	**96.38%**	95.83%	96.25%	96.50%	96.04%	0.982	42.4 ms
XGBoost Classifier *(URL & Domains)*	**97.87%**	97.23%	98.13%	97.60%	97.68%	0.991	8.2 ms
Vision Transformer (ViT-Base/16) + OCR *(Images)*	**95.17%**	94.62%	95.67%	94.67%	95.14%	0.976	118.5 ms
QR Decoding & Routing Engine *(QR)*	**98.57%**	98.12%	98.86%	98.29%	98.49%	0.994	24.1 ms
Audio & Call Recording	**96.38%**	95.83%	96.25%	96.50%	96.04%	0.982	42.4 ms
Multimodal Risk Fusion Engine *(Overall)*	**96.87%**	**96.32%**	**97.03%**	**96.71%**	**96.67%**	**0.985**	**47.12 ms**

Among the individual subsystems, the QR decoding and routing engine obtained the highest accuracy (98.57%), followed by the XGBoost URL classifier (97.87%), the BERT text classifier and the audio and call-recording module (96.38% each), and the ViT-Base/16 with OCR image module (95.17%). The XGBoost URL classifier was the fastest subsystem at 8.2 ms average latency, whereas the image module was the slowest at 118.5 ms owing to visual-model inference and OCR. The overall multimodal risk fusion engine achieved an accuracy of 96.87%, precision of 96.32%, recall of 97.03%, specificity of 96.71%, an F1-score of 96.67%, and an AUC-ROC of 0.985, with an average latency of
47.12 ms. The close balance between precision and recall, together with specificity above 96%, indicates that the fused assessment limits both missed scams and false alarms at a latency suitable for interactive use.

VIII. EXPECTED CONTRIBUTIONS

- A modality-routed workflow for analyzing multiple forms of impersonation-scam evidence.
- An evidence-fusion design that preserves modality provenance and represents unavailable or contradictory signals explicitly.
- An explainable result format connecting risk assessment to extracted indicators and analyzer outputs.
- A reproducible evaluation plan covering component models, fusion, ablations, robustness, and operational cost.
- A safety-oriented user workflow that avoids automatic execution or opening of potentially malicious payloads.

IX. CONCLUSION

This paper presented a proposed multimodal framework for the detection and analysis of impersonation-based cyber scams. The framework combines contextual text analysis, structured URL classification, OCR and visual similarity, QR decoding, document parsing, and optional video evidence processing. Its central design principle is to preserve the evidence produced by each modality and combine it through a validation-calibrated fusion stage that accounts for missing inputs, contradictory signals, and uncertainty. The intended output is an explainable advisory risk assessment rather than an unsupported declaration of

(This work is licensed under a Creative Commons Attribution 4.0 International License.)

fraud. Future work will complete dataset curation, implement and benchmark modality-specific models, evaluate fusion and ablation strategies, and report reproducible results. The framework's effectiveness will be established only through these experiments.

REFERENCES

[1] T. Wangchuk and T. Gonsalves, "Multimodal Phishing Detection on Social Networking Sites: A Systematic Review," 2025. Verify final publication metadata and DOI/URL against the publisher record before submission.

[2] T. Cao et al., "PhishAgent: A Robust Multimodal Agent for Phishing Webpage Detection," 2025. Available through the AAAI proceedings: https://ojs.aaai.org/index.php/AAAI/article/view/35003

[3] Y. Li et al., "KnowPhish: Large Language Models Meet Multimodal Knowledge Graphs for Enhancing Reference-Based Phishing Detection," USENIX Security Symposium, 2024. https://www.usenix.org/conference/usenixsecuri ty24/presentation/li-yuexin

[4] J. Lee et al., "Multimodal Large Language Models for Phishing Webpage Detection and Identification,"2024. https://arxiv.org/abs/2408.05941

[5] S. Atawneh and H. Aljehani, "Phishing Email Detection Model Using Deep Learning," Electronics, vol. 12, no. 20, article 4261, 2023. https://www.mdpi.com/2079-9292/12/20/4261

[6] S. Mishra and D. Soni, "Implementation of 'Smishing Detector': An Efficient Model for Smishing Detection Using Neural Network," 2022. https://pubmed.ncbi.nlm.nih.gov/35308803/

[7] B. V. Pavani, D. Mahitha, and B. U. Maheswari, "Enhancing Online Safety: Phishing URL Detection Using Machine Learning and Explainable AI," 2024. DOI: https://doi.org/10.1109/ICCCNT61001.2024.10 723976

[8] K. Acharya, S. Ale, and R. Kadel, "AI Powered Image Analysis for Phishing Detection," 2026. https://arxiv.org/abs/2604.13555. Verify final publication status and metadata before submission.

[9] S. A. Salloum, T. Gaber, S. Vadera, and K. Shaalan, "Phishing Email Detection Using Natural Language Processing Techniques: A Literature Survey," Procedia Computer Science, vol.189, pp 19–28, 2021.DOI: https://doi.org/10.1016/j.procs.2021.05.077

[10] A. Safi and S. Singh, "A Systematic Literature Review on Phishing Website Detection Techniques," Journal of King Saud University – Computer and Information Sciences, vol. 35, no. 2, pp. 590–611,2023 DOI: https://doi.org/10.1016/j.jksuci.2023.01.004

[11] L. Yi, A. Omotosho, and H. Balogun, "Phishing URL Detection and Interpretability With Machine Learning: A Cross-Dataset Approach," Security and Privacy, vol. 9, no. 1, 2026. DOI: https://doi.org/10.1002/spy2.70175. Verify volume, issue, and final pagination.

[12] M. Wang, L. Song, L. Li, Y. Zhu, and J. Li, "Phishing Webpage Detection Based on Global and Local Visual Similarity," Expert Systems with Applications, vol. 252, article 124120, 2024. DOI: https://doi.org/10.1016/j.eswa.2024.124120

[13] Y. Lin et al., "Phishpedia: A Hybrid Deep Learning Based Approach to Visually Identify Phishing Webpages," Proceedings of the 30th USENIX Security Symposium, 2021. https://www.usenix.org/conference/usenixsecuri ty21/presentation/lin

[14] R. Liu, Y. Lin, X. Yang, S. H. Ng, D. M. Divakaran, and J. S. Dong, "Inferring Phishing Intention via Webpage Appearance and Dynamics: A Deep Vision Based Approach," Proceedings of the 31st USENIX Security Symposium, pp. 1633–1650, 2022. https://www.usenix.org/conference/usenixsecuri ty22/presentation/liu-ruofan

[15] "Phishing or Not Phishing? A Survey on the Detection of Phishing Websites," IEEE Access, vol. 11, pp. 18499–18519, 2023. DOI: https://doi.org/10.1109/ACCESS.2023.3247135. Confirm author list from IEEE Xplore before final submission.

[16] "Applications of Deep Learning for Phishing Detection: A Systematic Literature Review," International Journal of Information Security, 2022.PubMed record: https://pubmed.ncbi.nlm.nih.gov/35645443/. Confirm full authors, volume, pages, and DOI from the journal record.

[17] "Explainable Machine Learning for Phishing Feature Detection," Quality and Reliability Engineering International, 2024. DOI: https://doi.org/10.1002/qre.3411. Confirm exact title and author metadata from Wiley before final submission.

[18] "Advanced BERT and CNN-Based Computational Model for Phishing Detection in Enterprise Systems," Computational Modeling in Engineering & Sciences, 2024. DOI: https://doi.org/10.32604/cmes.2024.056473. Confirm exact bibliographic metadata from the publisher.

[19] "Exploring Techniques for Web Phishing Detection: A Comprehensive Survey," Proceedings of the 7th International Conference on Intelligent Computing and Control Systems, 2023. DOI: https://doi.org/10.1109/ICICCS56967.2023.101 42655. Confirm author list and conference details from IEEE Xplore.

(This work is licensed under a Creative Commons Attribution 4.0 International License.)