IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 15, Issue 9, September 2026

DOI: 10.17148/IJARCCE.2026.15910

Lightweight and Explainable Real-Time Detection of Social Media-Borne Phishing URLS: A Critical Review for Indian Context

Pushpam Kumari

Department of MCA, Babasaheb Bhimrao Ambedkar Bihar University, Muzaffarpur, Bihar, India

Abstract: Cybercrime in India has undergone a fundamental transformation in the last five years. Phishing, which was once limited to fraudulent emails, has now shifted aggressively to social media platforms such as WhatsApp, Telegram, Instagram and Facebook. Cybercriminals exploit trust by circulating deceptive links that promise free mobile recharge, government jobs, scholarship schemes, KYC updates and lottery prizes. Academic researchers have responded with a large number of machine learning based detection systems that report very high accuracy, often exceeding 98 percent. However, these systems are developed in isolation from real-world deployment constraints. Critical practical metrics such as inference latency, model file size, throughput and explainability for non-technical users are completely ignored. Most models depend on external services such as WHOIS lookup, DNS query and PageRank fetching, which require network calls and introduce significant delay. This paper presents a comprehensive critical theoretical review of six prominent recent studies published between 2025 and 2026.

Keywords: Phishing Detection, Social Media Security, Lightweight ML, Explainable AI, Real-Time Detection

1. INTRODUCTION

1.1 Background:

India is largest digital market in terms of social media users. Affordable smartphones and cheap data plans have brought crores of first-time internet users online. For many users WhatsApp is primary communication channel. This digital inclusion has also created huge attack surface.

1.2 Social Media Phishing:

Phishing is form of fraud where attacker creates fake website that looks exactly like legitimate website and tricks users into entering sensitive information such as bank details, OTP and passwords. In social media context attackers register deceptive domains using free domain providers, create fake pages that look like original service providers, and circulate messages creating urgency. Within hours lakhs of users receive it and thousands become victims.

1.3 Difference Between Email and Social Media Phishing:

Traditional detection was focused on email. Email URLs are typically long and analyzed with headers and sender reputation. Social media phishing URLs are fundamentally different. Attackers intentionally use URL shorteners to hide real destination, use typosquatting, and use HTTPS to show lock icon to gain trust. Therefore models trained on email datasets fail on social media URLs.

1.4 Need for Real-Time Lightweight Detection:

Most critical requirement is real-time detection. When user clicks link, security system has less than 100 milliseconds to analyze URL and warn user before malicious scripts execute. If detection system needs to fetch WHOIS, DNS, PageRank from third-party server it will add 300 to 500 milliseconds delay. Model size is equally important. Chrome extension cannot bundle 50MB deep learning model. Users in rural areas use low-RAM phones with 2GB RAM.

2. LITERATURE REVIEW

2.1 Deep Learning Approach:

This approach treats URL as character sequence and uses Bi-Directional LSTM. Bi-LSTM learns patterns automatically without manual feature engineering. Theoretically powerful but practically it has millions of parameters, model size more than 50MB, needs GPU, inference time more than 150ms on CPU, black-box with no explainability.

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 15, Issue 9, September 2026

DOI: 10.17148/IJARCCE.2026.15910

2.2 Ensemble Learning Approach:
Ensemble combines many classifiers to improve accuracy. One work combined 13 models including Random Forest, XGBoost, Decision Tree and SVM with Redis cache for faster lookup. While accuracy improves, complexity increases dramatically. Maintaining 13 models and Redis server makes system expensive and difficult to scale. Suitable for enterprise server not for lightweight deployment.

2.3 Pure Lexical Feature Approach:
This approach emphasizes simplicity and speed. It extracts only lexical features directly from URL string without any network call. Features include URL length, domain length, dot count, hyphen count, at symbol presence, IP address presence, subdomain count, digit ratio, entropy, HTTPS presence. Extraction takes less than 5 milliseconds. Uses simple classifier like LDA. Model size less than 100KB. Truly real-time and lightweight. Limitation is moderate accuracy.

2.4 Survey Based Works:
Survey papers analyze many papers. One survey reviewed 15 phishing detection papers from 2024 to 2026. Conclusion is that high computational cost is main barrier preventing academic models from real-world deployment. Lack of standardized evaluation including latency and size is highlighted.

2.5 Heavy Feature Based Approach:
Some papers use more than 50 features including URL lexical, host-based like WHOIS creation date, DNS record, SSL issuer and reputation like Alexa rank. Use Random Forest and XGBoost and report up to 99.99 percent accuracy. While accuracy impressive methodology flawed for real-time claims. WHOIS lookup takes 300 to 400ms, DNS takes 50 to 100ms, total extraction exceeds 500ms violating real-time definition.

2.6 Explainable AI Approach:
Techniques like SHAP assign importance value to each feature. Crucial for user trust. However current explainable systems use 87 features including PageRank and domain reputation requiring external API calls making them slow.

Table 1: Comparative Analysis of Phishing Detection Papers

Paper (Year)	Features Type	Model	Net Call?	Latency?	Size?	Mobile?
Baskota 2025	Auto Char Seq	Bi-LSTM	No	Not Reported	Not Reported	No
Parimi 2025	20+ Lex+Cache	13 Ensemble	Redis Yes	No	No	No
Devi 2025	13 Lex Only	LDA	No	No	No	Yes
Saxena 2026	15 Papers Review	Survey	-	No	No	-
Rehman 2025	54 WHOIS DNS	RF XGB	Yes	No	No	No
Yi 2026	87 PageRank	XGB SHAP	Yes	No	No	No
Our Proposed	**15-18 Lex Only**	**Tiny LR/DT**	**No**	**Yes <50ms**	**Yes <500KBs**	**Yes**

4. RESEARCH GAPS

Gap 1: No latency reporting. Real-time claim without proof is invalid. All papers claim real-time but no paper reports milliseconds.
Gap 2: No model size reporting. Cannot judge deployability on low-RAM phones or as Chrome extension.
Gap 3: Network Dependency Paradox. Claim real-time but use WHOIS DNS PageRank which are slow and need internet.
Gap 4: Indian Context Ignored. Existing datasets do not contain Indian fraud patterns like free recharge and government scheme fraud which are common in Bihar UP Jharkhand.
Gap 5: Explainability for Common User Missing. No simple explanation for non-technical users in Hindi English mix.
Gap 6: Short URL Handling Missing. Social media heavily uses bit.ly tinyurl to hide destination. No paper handles this.

International Journal of Advanced Research in Computer and Communication Engineering
Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 15, Issue 9, September 2026
DOI: 10.17148/IJARCCE.2026.15910

Gap 7: Throughput Not Measured. Number of URLs per second system can handle is important for deployment.
Gap 8: Low RAM Device Testing Missing. No testing on 2GB RAM phone or as extension.

5. PROPOSED WORK – THEORETICAL FRAMEWORK

Layer 1 - Data Collection:
Two sources. Source A Standard datasets 450K and 235K for benchmarking. Source B Custom Indian Social Media Phishing Dataset containing 1200 plus real fraud links manually collected from WhatsApp groups and Telegram channels active in Bihar UP Jharkhand.

Layer 2 - Feature Engineering:
Extract 15 to 18 features computable in less than 5ms via string operations. Structural: URL length, domain length, path length, dot count, hyphen count, slash count, subdomain count. Security: at symbol presence, IP address presence, HTTPS presence, digit ratio, entropy, short URL detection. Social Engineering: suspicious keyword count with keywords like free recharge job lottery KYC prize.

Layer 3 - Model Selection:
Strict size limit less than 500KB. Logistic Regression 5KB, Decision Tree depth 10 15KB, Random Forest 20 trees 200KB, Tiny LightGBM 25 estimators 300KB. Final model less than 500KB runnable on Chrome extension offline.

Layer 5 - Explainability:
Use SHAP to get top 3 contributing features and translate to simple user understandable explanations with rule templates. Example: Presence of at symbol indicates suspicious pattern commonly used in phishing.

Layer 4 - Holistic Evaluation:
1. Accuracy, Precision, Recall, F1-Score
2. Latency in milliseconds
3. Model Size in KB
4. Throughput - URLs per second
5. Explainability User Study

Layer 6 Deployment:
1. Flask REST API that returns result in less than 50ms
2. Chrome Extension for real-time warning
3. Android WebView Test on 2GB RAM phone

6. CONCLUSION - DETAILED

This critical review clearly demonstrates that although recent research on phishing URL detection has achieved remarkable accuracy of 98 to 99 percent this accuracy has been achieved at cost of practical deployability. Majority of existing models are heavy complex network-dependent and lack explainability. They depend on slow external services like WHOIS DNS and PageRank which add 300 to 500ms delay and violate real-time requirement of less than 100ms. Furthermore none of surveyed works report model size inference latency or throughput which are essential metrics for deployment on low-end devices such as 2GB RAM smartphones widely used in India. Another major limitation is ignorance of Indian social media context. Fraud patterns like free recharge government job scams KYC update fraud and lottery scams are not represented in existing datasets. Our proposed lightweight framework addresses all these gaps by using only 15 to 18 lexical features that can be extracted in less than 5ms without any network call using tiny models under 500KB and providing SHAP-based simple explanations for non-technical users. Future work will focus on building custom Indian dataset with 1200+ real fraud URLs implementing Flask API Chrome extension and testing on low-RAM devices.

REFERENCES

[1]. Baskota et al. Deep Learning Based Phishing URL Detection Using Bi-LSTM. 2025.
[2]. Parimi et al. Ensemble Learning with Redis Cache for Real-Time Phishing Detection. 2025.
[3]. Devi et al. Lexical Feature Based Lightweight Phishing Detection Using LDA. IJARCCE Vol 14 2025.
[4]. Saxena et al. A Survey on Phishing Detection Techniques. IEEE Access Survey 2026.
[5]. Rehman et al. PhiUSIIL: A Large Phishing URL Dataset with 54 Features. Data in Brief Elsevier 2025.
[6]. Yi et al. Explainable Phishing Detection Using XGBoost and SHAP. Expert Systems with Applications 2026.

