IJIRSET
International Journal of Innovative Research in Science | Engineering | Technology

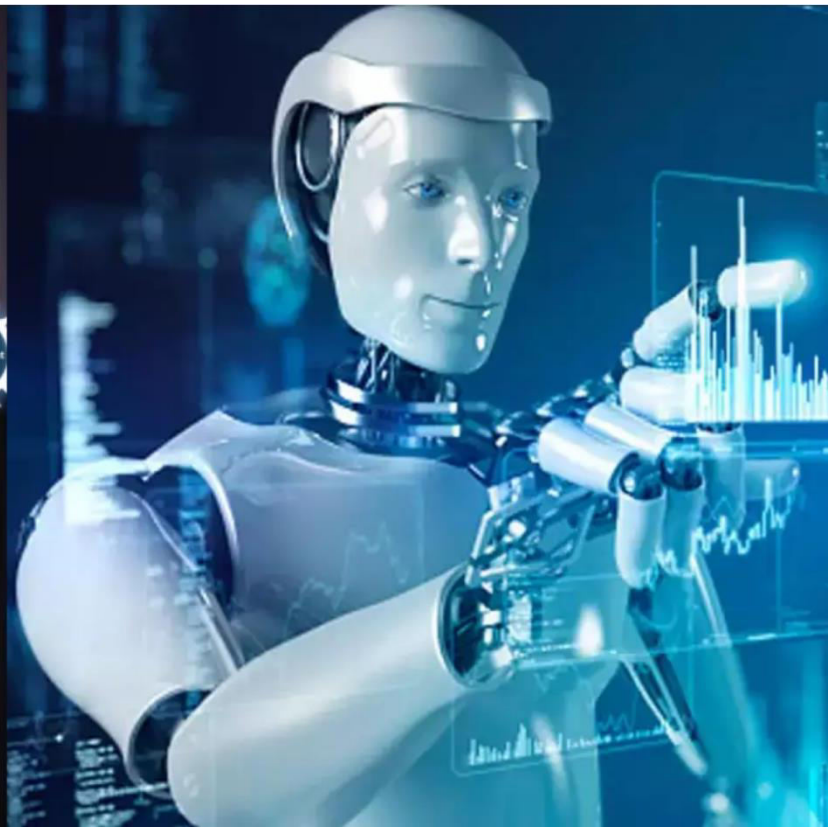
ISSN(O): 2319-8753
ISSN(P): 2347-6710



International Journal of Innovative Research in Science Engineering and Technology (IJIRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)





Impact Factor: 9.935

Volume 15, Issue 10, October 2026

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 9.935

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 15, Issue 10, October 2026

|DOI: 10.15680/IJIRSET.2026.1510090|

Phishing Webpage Detection using Machine Learning: Threat Analysis and Technique Evaluation

Prof. Vinutha[1], Harish P S[2], Sharath R[3], Swati[4], Akash R[5]

Assistant Professor, Department of Computer Science and Engineering, Vidya Vikas Institute of Engineering and Technology, Mysuru, Karnataka, India[1]

Student, Department of Computer Science and Engineering, Vidya Vikas Institute of Engineering and Technology, Mysuru, Karnataka, India[2-5]

ABSTRACT: Phishing pages are cheap to create and usually disappear within hours, so defences that wait for a site to be reported and blocklisted tend to act after the damage is done. This paper describes PhishGuard, a web-based system that judges a URL on the spot and tells the user why it reached its verdict. For each submitted address the system computes thirty binary indicators drawn from the URL string, its host, and, where the page can be retrieved, its HTML and redirect chain. A Multi-Layer Perceptron (MLP) and three classical learners score this vector, and the averaged probability is returned through a FastAPI service to a React client that offers three interface styles for different kinds of users. On a balanced set of 8,000 synthetic feature vectors the MLP reached 95.9% accuracy with 97.0% recall, while logistic regression (97.3%) and a plain count-of-flags rule (97.1%) did at least as well, showing that this benchmark is close to linearly separable. Character-level models evaluated across data sources reached 93.4–94.6% accuracy. Model inference takes well under a millisecond, and the median end-to-end scan, which is dominated by page retrieval, took 1.76 s. We report these results together with the limits of the evaluation, and argue that the value of the system lies in its transparent evidence and complete pipeline rather than in headline accuracy.

KEYWORDS: phishing detection, URL classification, multi-layer perceptron, explainable security, feature engineering, FastAPI, React

I. INTRODUCTION

Banking, shopping, and workplace tools have moved almost entirely to the browser, and with them the credentials that attackers want. Phishing exploits this by presenting a convincing copy of a trusted page and persuading the victim to type a password, card number, or one-time code into it. Industry reporting, such as the quarterly trend reports of the Anti-Phishing Working Group [11], has for years placed phishing among the most frequently observed attack types, and individual phishing sites are typically short-lived.

Most deployed protection is list-based: a URL is blocked once it has been seen, reported, and added to a database. Such lists are precise for known threats but blind to a domain registered shortly before a campaign. Hand-written heuristics react faster but are easy to sidestep with look-alike characters, punycode hostnames, or randomly generated subdomains. Learning-based classifiers offer a middle path, since they can generalise from the surface properties of a URL to sites never seen before [1], [2].

Accuracy alone, however, does not make a tool usable. A warning that offers no reason is easily dismissed, and a classifier that cannot learn from its mistakes slowly falls behind attacker behaviour [7], [9]. PhishGuard was built to address these practical gaps. Its contributions are: (i) a thirty-indicator extraction pipeline in which every indicator is a human-readable yes/no signal; (ii) an evidence panel that separates signals actually observed from those assumed because the page could not be loaded, so that users can see how much of a verdict rests on real evidence; (iii) a REST service with a feedback endpoint and a downloadable report; and (iv) a candid evaluation that compares the neural model against simple linear and rule baselines and states where the benchmark falls short of real traffic.

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 9.935

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 15, Issue 10, October 2026

|DOI: 10.15680/IJIRSET.2026.1510090|

II. RELATED WORK

Sahingoz et al. [1] showed that lexical properties and word-level cues extracted from the URL are enough to train competitive classifiers, although rules tied to particular words transfer poorly to obfuscated addresses. Mohammad et al. [2] earlier examined which website features carry the most signal, establishing the thirty-attribute style of feature set that many later studies, including ours, build on. Moving away from hand-crafted features, Aljofey et al. [3] trained a character-level convolutional network directly on the URL string; in later work [4] they added hyperlink and HTML-text features from the fetched page and classified them with XGBoost, gaining accuracy at the cost of a page download before every decision.

Ensembles are a recurring theme. Alsariera et al. [6] wrapped an Extra-Trees base learner in several meta-learners and reported accuracy of at least 97%, and Kalabarige et al. [5] combined boosting-based feature selection with a stacked ensemble. Surveys by Basit et al. [7] and Odeh et al. [8] note that few published detectors reach deployment and that explanation and feedback mechanisms are usually missing. Ejaz et al. [9] addressed drift through continual learning, and Uddin et al. [10] paired boosting models with SHAP attributions to make predictions interpretable. PhishGuard draws on these strands but focuses on the complete user-facing loop of verdict, evidence, report, and feedback.

III. SYSTEM ARCHITECTURE

PhishGuard follows a three-tier design (Fig. 1). The presentation tier is a React application, built with CRACO and Radix UI components, that offers three themes over one shared component library: Studio for analysts, Arcade for casual users, and Terminal for keyboard-driven use. The application tier is a Python FastAPI service running under Uvicorn. Its scan endpoint validates the input, retrieves the page through an outbound-fetch guard with a 15-second read timeout, calls the feature extractor, and returns the verdict, the probability, and the full indicator vector. A second endpoint records user corrections in a feedback ledger for later retraining, and an HTML report can be generated for any scan. The intelligence tier consists of the feature extractor and four scikit-learn [12] models stored as joblib files and loaded once at start-up. No GPU is required; the service runs on an ordinary two-core, 4 GB host, and models can be replaced without changing the API contract.

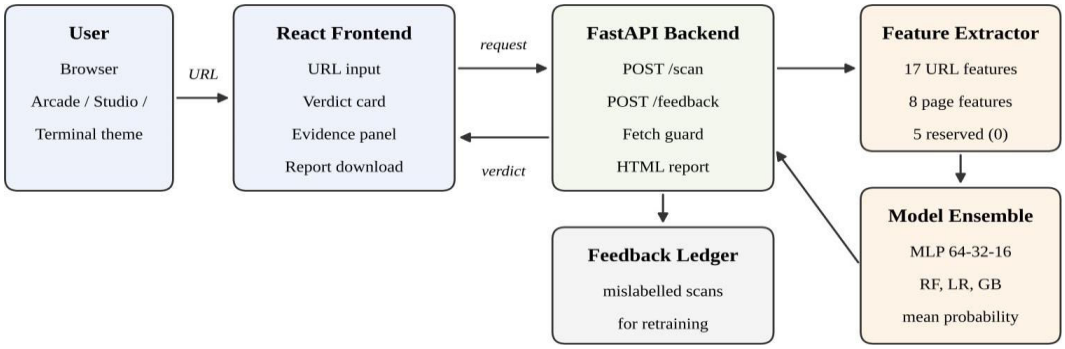


Fig.1. PhishGuard architecture and request flow.

IV. METHODOLOGY

4.1 Datasets

Two datasets were prepared. For the thirty-indicator models we generated 8,000 synthetic vectors, 4,000 per class, by sampling each indicator from a class-specific probability profile. This was a deliberate safety choice: building a comparable corpus from real pages would have meant downloading thousands of live phishing sites. For the character-level models we assembled 2,055 curated URLs, of which 1,121 are legitimate addresses drawn from the Tranco [13] and Alexa top-site lists and 934 are phishing addresses written to follow patterns documented by PhishTank, OpenPhish, and the APWG (none of them live), plus 4,000 template-generated URLs. After removing duplicates 4,264 unique URLs remained. Both datasets were divided 80/20 with stratification.

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 9.935

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 15, Issue 10, October 2026

|DOI: 10.15680/IJIRSET.2026.1510090|

4.2 Feature Extraction

Every URL is mapped to thirty binary indicators, where 1 means suspicious. Seventeen come from the address itself: lexical cues such as overall and hostname length, path depth, digit ratio, character entropy, and special characters, and structural cues such as a raw IP address as host, missing HTTPS, an explicit port, an '@' symbol, and excessive subdomains. Eight come from the retrieved page and its redirect chain: hidden iframes, mouse-over scripts, disabled right-click, too many redirects, login forms, an external favicon, an empty title, and forms that submit to another domain. If the page cannot be fetched, these eight default to suspicious, which is a fail-safe choice whose side effects are discussed in Section 5.3. The remaining five (DNS record, domain age, expiry window, web traffic, WHOIS age) are reserved for future WHOIS and traffic-rank feeds and are currently fixed at zero.

4.3 Classifiers

The primary model is an MLP with hidden layers of 64, 32, and 16 ReLU units, trained with the Adam optimiser, an L2 penalty of 0.0001, and at most 300 iterations. Alongside it we train a random forest (200 trees, depth 8), a logistic regression, and a gradient-boosting model (200 stages, depth 3) on the same split. The phishing probability shown to the user is the mean of the four model outputs. These probabilities are not calibrated, so the interface presents them as a confidence indication rather than an exact likelihood. As a reference point we also report a rule that flags a URL when six or more indicators are set, with the threshold chosen on training data only.

4.4 Evaluation Protocol

Models are compared on a stratified 20% hold-out set (random seed 42) and with stratified five-fold cross-validation, treating phishing as the positive class. For the character-level models, a cross-source test trains on template-generated URLs and tests on curated ones, so that memorised templates cannot inflate the score. A live check scans twenty well-known legitimate sites through the running API. Fifteen automated tests cover the backend endpoints. All numbers were reproduced from the project code on a Windows 11 laptop with Python 3.13.

V. RESULTS AND DISCUSSION

5.1 Thirty-Indicator Classifiers

Table 1 and Fig. 2 summarise the hold-out results. Logistic regression and gradient boosting score highest, at 97.3% and 97.2% accuracy. The MLP achieves the best recall among the non-linear models (97.0%) but the lowest precision (95.0%), producing 41 false alarms against 24 missed phishing vectors. Notably, the six-flag rule reaches 97.1% accuracy and outperforms both the MLP and the random forest. This is the expected outcome for this benchmark: because each indicator is sampled independently given the class, the best possible decision is essentially a weighted count of flags, which linear models and counting capture directly. Every model exceeds 0.99 AUC.

Table.1. Thirty-indicator classifiers on the hold-out set (n = 1,600; values in %)

Model	Accuracy	Precision	Recall	F1	AUC	5-fold CV F1
MLP	95.94	94.98	97.00	95.98	0.994	96.75 ± 0.53
Random forest	96.00	96.46	95.50	95.98	0.994	96.17 ± 0.47
Model	**Accuracy**	**Precision**	**Recall**	**F1**	**AUC**	**5-fold CV F1**
Logistic regression	97.31	97.49	97.12	97.31	0.997	97.81 ± 0.39
Gradient boosting	97.19	97.60	96.75	97.18	0.997	97.57 ± 0.55
Rule (≥ 6 flags)	97.06	97.24	96.88	97.06	0.996	–

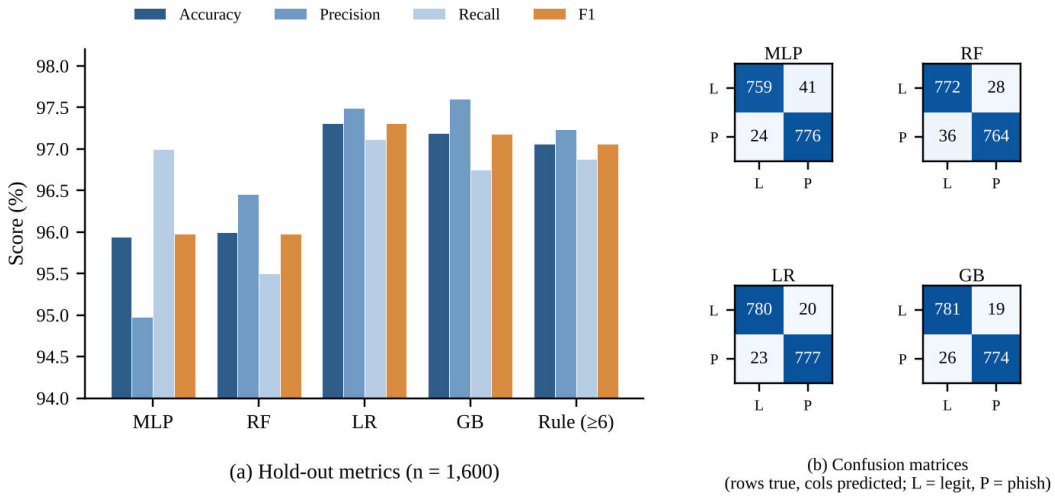
International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 9.935

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 15, Issue 10, October 2026

|DOI: 10.15680/IJIRSET.2026.1510090|



(a) Hold-out metrics (n = 1,600)

(b) Confusion matrices
(rows true, cols predicted; L = legit, P = phish)

Fig. 2. (a) Hold-out metrics of the four classifiers and the rule baseline; (b) confusion matrices of the four classifiers.

5.2 Character-Level Models

Table 2 reports models that read the raw URL instead of the engineered indicators. On the 853-URL test split the n-gram models are perfect and the one-dimensional CNN reaches 98.7%. These figures say more about the corpus than about the models: 98.1% of the curated phishing URLs use plain HTTP, so the single test "no HTTPS" already scores 99.1%. The cross-source test is the more meaningful figure, with 94.6% accuracy for logistic regression and 93.4% for the n-gram MLP.

Table.2. Character-level models (test n = 853, curated subset n = 400; values in %)

Model	Accuracy	Precision	Recall	F1	AUC	Curated acc.
n-gram MLP	100.00	100.00	100.00	100.00	1.000	100.00
n-gram RF	100.00	100.00	100.00	100.00	1.000	100.00
n-gram LR	100.00	100.00	100.00	100.00	1.000	100.00
Conv1D CNN	98.71	98.22	99.60	98.91	0.999	97.75

5.3 Live-Site Check and Latency

Twenty popular legitimate sites, among them google.com, wikipedia.org, amazon.in, sbi.co.in, github.com/login, and paypal.com/signin, were scanned through the running service. Nineteen were correctly marked legitimate with confidence between 0.957 and 0.993. The twentieth, irctc.co.in, did not answer within the timeout; its eight page indicators therefore defaulted to suspicious and the ensemble labelled it phishing with 0.985 confidence. The same mechanism had caused four of fourteen legitimate sites to be misclassified in an earlier run, when the fetch guard wrongly rejected IPv6 NAT64 addresses. The fail-safe default thus trades false negatives for false positives on unreachable or bot-blocking sites, which is why the evidence panel marks such signals as assumed. No live phishing pages were scanned, for safety reasons.

Inference time for a single URL is negligible: 0.05 ms for the MLP and logistic regression, 0.14 ms for gradient boosting, and 30 ms for the random forest. Over the nineteen successful live scans the median

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 9.935

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

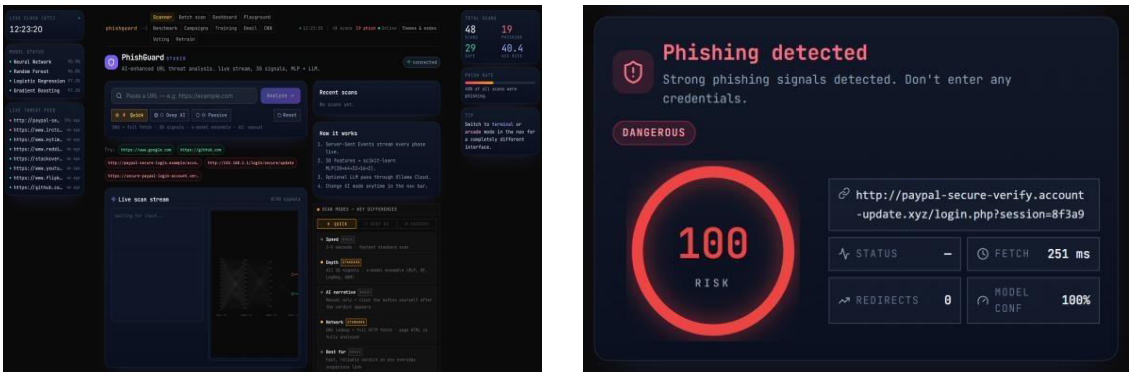
Volume 15, Issue 10, October 2026

|DOI: 10.15680/IJIRSET.2026.1510090|

end-to-end time was 1.76 s (range 0.84–29.0 s), almost all of it spent on DNS resolution and page retrieval. All fifteen backend tests pass.

VI. APPLICATION SNAPSHOTS

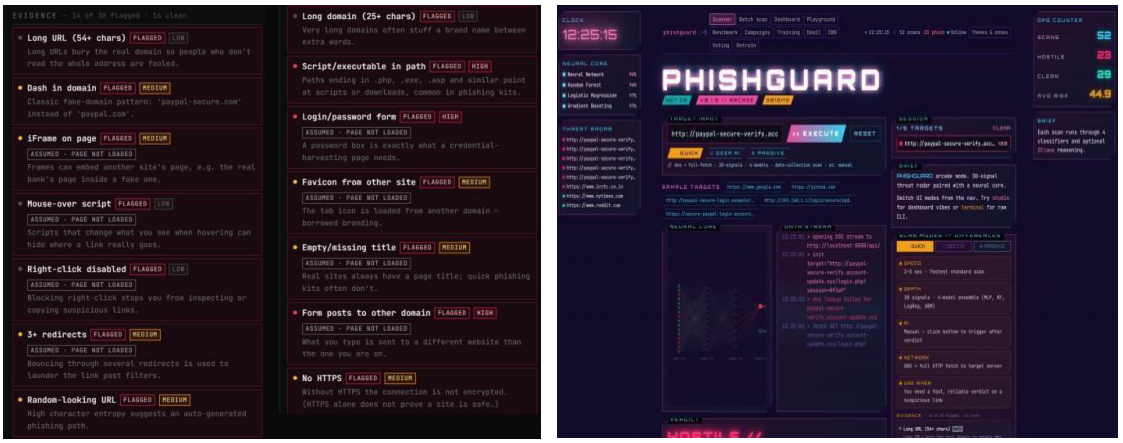
Figs. 3 and 4 show the running application. The example input is a constructed phishing-style URL that does not resolve, so in the evidence panel (Fig. 4a) the page-level checks are labelled assumed, page not loaded rather than shown as observed.



(a) Home page

(b) Scan result with risk score and model confidence

Fig.3. Studio theme: home page and scan result.



(a) Per-feature evidence panel

(b) Arcade theme after a scan

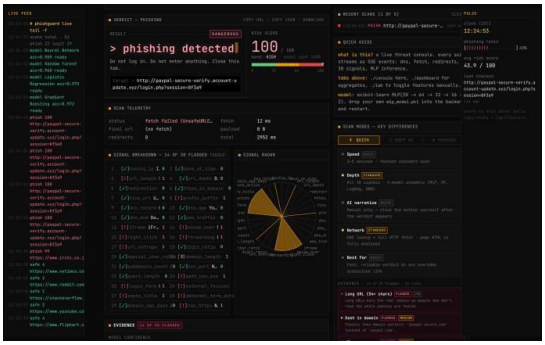
International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

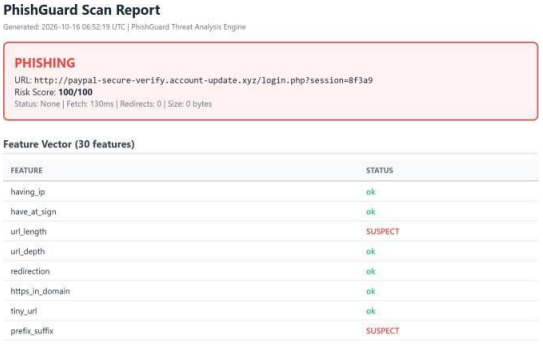
Impact Factor 9.935

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 15, Issue 10, October 2026

|DOI: 10.15680/IJIRSET.2026.1510090|





(c) Terminal theme with signal breakdown and radar (d) Downloadable HTML scan report

Fig.4. Evidence panel, Arcade and Terminal themes, and the downloadable scan report.

VII. LIMITATIONS

Three points qualify the results. First, the thirty-indicator models were trained and tested on synthetic vectors, so their scores measure agreement with an assumed feature profile rather than performance on live traffic. Second, the five reserved indicators are always zero at inference time even though they carry signal in the training data; permutation importance ranks domain age, web traffic, expiry window, and DNS record as the MLP's four most influential inputs, which means the deployed models partly depend on information they never receive. Third, the curated URL corpus can be separated by protocol alone, whereas most present-day phishing sites use HTTPS. A time-split evaluation on a live feed is required before the accuracy figures can be generalised.

VIII. CONCLUSION AND FUTURE WORK

PhishGuard shows that a complete, explainable phishing checker can run on commodity hardware without external threat-intelligence services: an address is scored in well under a millisecond once its indicators are known, and the user sees exactly which signals drove the verdict and which were only assumed. Our own baselines also show that, on this benchmark, the MLP offers no accuracy advantage over logistic regression or a simple flag count. The contribution of the work therefore lies in the end-to-end pipeline, the observed-versus-assumed evidence display, the feedback path, and an honest account of its evaluation. Future work will (i) evaluate on time-split PhishTank and OpenPhish data and implement the reserved WHOIS, DNS, and traffic-rank indicators; (ii) explore transformer-based URL encoders for homoglyph and punycode attacks; (iii) package the detector as a Chrome and Firefox extension and as an ONNX build for mobile devices; and (iv) add support for Indian regional languages and federated retraining from user feedback.

IX. ACKNOWLEDGMENT

The authors thank the management and the Department of Computer Science and Engineering, Vidya Vikas Institute of Engineering and Technology, Mysuru, for the facilities and support provided for this work.

REFERENCES

[1]O. K. Sahingoz, E. Buber, O. Demir, and B. Diri, "Machine learning based phishing detection from URLs," Expert Systems with Applications, vol. 117, pp. 345–357, 2019, doi: 10.1016/j.eswa.2018.09.029.
[2]R. M. Mohammad, F. Thabtah, and L. McCluskey, "An assessment of features related to phishing websites using an automated technique," in Proc. 7th Int. Conf. Internet Technology and Secured Transactions (ICITST), London, UK, 2012, pp. 492–497.
[3]A. Aljofey, Q. Jiang, Q. Qu, M. Huang, and J.-P. Niyigena, "An effective phishing detection model based on character level convolutional neural network from URL," Electronics, vol. 9, no. 9, Art. 1514, 2020, doi: 10.3390/electronics9091514.
[4]A. Aljofey, Q. Jiang, A. Rasool, H. Chen, W. Liu, Q. Qu, and Y. Wang, "An effective detection approach for phishing websites using URL and HTML features," Scientific Reports, vol. 12, Art. 8842, 2022, doi: 10.1038/s41598-

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 9.935

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 15, Issue 10, October 2026

|DOI: 10.15680/IJIRSET.2026.1510090|

022-10841-5.
[5]L. R. Kalabarige, R. S. Rao, A. R. Pais, and L. A. Gabralla, "A boosting-based hybrid feature selection and multi-layer stacked ensemble learning model to detect phishing websites," IEEE Access, vol. 11, pp. 71180–71193, 2023, doi: 10.1109/ACCESS.2023.3293649.
[6]Y. A. Alsariera, V. E. Adeyemo, A. O. Balogun, and A. K. Alazzawi, "AI meta-learners and extra-trees algorithm for the detection of phishing websites," IEEE Access, vol. 8, pp. 142532–142542, 2020, doi: 10.1109/ACCESS.2020.3013699.
[7]A. Basit, M. Zafar, X. Liu, A. R. Javed, Z. Jalil, and K. Kifayat, "A comprehensive survey of AI-enabled phishing attacks detection techniques," Telecommunication Systems, vol. 76, no. 1, pp. 139–154, 2021, doi: 10.1007/s11235-020-00733-2.
[8]A. Odeh, I. Keshta, and E. Abdelfattah, "Machine learning techniques for detection of website phishing: A review for promises and challenges," in Proc. IEEE 11th Annual Computing and Communication Workshop and Conference (CCWC), 2021, pp. 813–818, doi: 10.1109/CCWC51732.2021.9375997.
[9]A. Ejaz, A. N. Mian, and S. Manzoor, "Life-long phishing attack detection using continual learning," Scientific Reports, vol. 13, Art. 11488, 2023, doi: 10.1038/s41598-023-37552-9.
[10]K. M. M. Uddin, N. Biswas, S. T. Rikta, M. Nur-A-Alam, and R. Mostafiz, "Explainable machine learning for phishing site detection: A high-efficiency approach using boosting models and SHAP," The Journal of Engineering, vol. 2025, no. 1, Art. e70110, 2025, doi: 10.1049/tje2.70110.

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 9.935

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 15, Issue 10, October 2026

|DOI: 10.15680/IJIRSET.2026.1510090|

[11]Anti-Phishing Working Group (APWG), Phishing Activity Trends Reports, quarterly series. [Online]. Available: https://apwg.org/trendsreports/

[12]F. Pedregosa et al., "Scikit-learn: Machine learning in Python," Journal of Machine Learning Research, vol. 12, pp. 2825–2830, 2011.

[13]V. Le Pochat, T. Van Goethem, S. Tajalizadegan, M. Korczyński, and W. Joosen, "Tranco: A research-oriented top sites ranking hardened against manipulation," in Proc. Network and Distributed System Security Symposium (NDSS), 2019, doi: 10.14722/ndss.2019.23386.



INTERNATIONAL STANDARD SERIAL NUMBER INDIA

INNO SPACE SJIF Scientific Journal Impact Factor

doi cross ref

निस्केयर NISCAIR

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

9940 572 462 6381 907 438 ijirset@gmail.com

www.ijirset.com

Scan to save the contact details