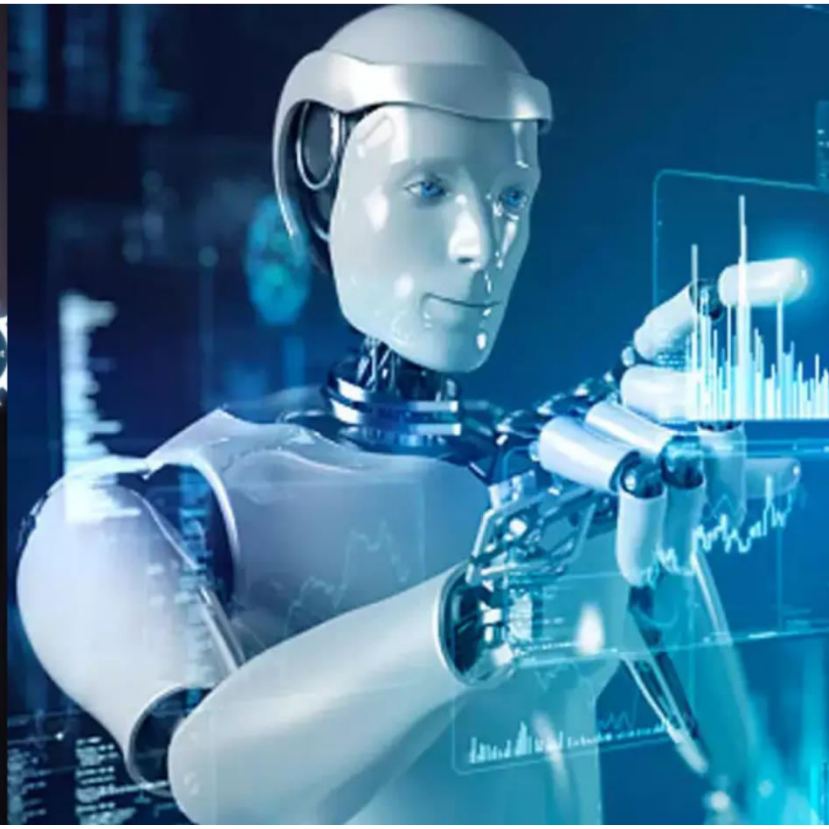
IJIRSET
International Journal of Innovative Research in Science | Engineering | Technology

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: 8.699

Volume 15, Issue 2, February 2026

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

Impact Factor 8.699

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

Volume 15, Issue 2, February 2026

|DOI: 10.15680/IJIRSET.2026.1502064|

A Comprehensive Review of URL-Based Phishing Detection Techniques: From Classical Machine Learning to Transformer-Based Models

Er. Akshat Vedvias[1], Dr. Avni Sharma[2]

P.G. Student, Department of Computer Science and Engineering, Himachal Pradesh Technical University, Hamirpur, Himachal Pradesh, India[1]

Faculty, Department of Computer Science and Engineering, Himachal Pradesh Technical University, Hamirpur, Himachal Pradesh, India[2]

ABSTRACT: Phishing is a common security threat on the internet, which uses fraudulent URLs to undermine user privacy and security. The conventional blacklisting and rule-based methods cannot resist the fast changes in phishing techniques, such as the obfuscated, zero-day, and AI-generated URLs. The present review paper gives an overview of the latest phishing URL detection systems, including classical machine learning algorithms, ensemble classifiers, explainable features selection strategies, deep learning systems, and transformer-based systems. We are catego- rizing and comparing more than a dozen representative studies according to features representation, learning paradigms, detection accuracy, computational efficiency, and implementation. The latest developments that include hybrid deep learning models, uncertainty-aware neural networks, knowledge- distilled transformers, and large-scale hybrid datasets are presented. A consistent comparison is made to investigate all important trade-offs between detection performance and real-world applicabil- ity. As well, pressing issues such as bias in datasets, adversarial robustness, uncertainty quantification, multi-class classification and real-time deployment limitations are also determined. Lastly, future research directions are also promised to facilitate the creation of an efficient, high-quality, and scalable phishing URL detection systems. The survey will be used to act as a practical guide to both researchers and practitioners in cybersecurity.

KEYWORDS: Phishing, URL, Random Forest, Support Vector Machine, Deep Learning.

I. INTRODUCTION

The use of phishing attacks has remained a major menace to users, as well as organizations, globally because it uses false URLs to steal sensitive information to profit financially and carry out identity theft. The more classic signature and blacklist-based methods are unable to detect new phishing URLs and those that are obfuscated and thus research is being conducted on intelligent methodologies of detection [1], [2]. The first classical machine learning models implemented on phishing URL detection were Logistic Regression (LR), Support Vector Ma- chines (SVM), and Random Forests (RF) which used handcrafted lexical and host-based features and content features [4], [8]. These methods were proven to have sufficient accuracy on benchmarks, but were not generalizable as they depended on manually engineered features.

Ensemble learning and feature selection strategies were investigated to make them more robust, and a combination of numerous classifiers and explainable artificial intelligence (XAI) tools were used to find discriminative features [5], [7], [12]. Nevertheless, feature-based mod- els cannot go without constant updates to keep up with the changing phishing tactics. The deep-learn- ing revolution led to the automatic representation learning directly on raw URL sequences using models like Convolutional Neural Networks (CNNs), Recurrent Neural Networks (RNNs) and hybrid models, which have enhanced the quality of detecting features without the need of significant feature engineering [9], [10]. More recently, models based on transformers, including BERT and ELECTRA, have reached state- of-the-art performance by modeling contextual dependencies between URL characters, and methods like knowledge distillation and uncertainty quantification have also made real- time models deployable and allowed making reliable decisions [11], [12]. In spite of these developments, several problems still exist, such as bias in datasets, adversarial resistance,

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

Impact Factor 8.699

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

Volume 15, Issue 2, February 2026

|DOI: 10.15680/IJIRSET.2026.1502064|

and restrictions of deployment in real-time [3], [13]. The review synthesizes the existing literature and assesses the trade-offs of methods, as well as determining the relevant directions in the future work.

II. LITERATURE SURVEY

The development of phishing URL detection has gone beyond the conventional feature-based approaches of machine learning to the deep learning and transformer-based approaches. This part of the paper summarizes the literature by collating them into classical methods of machine learning, adaptive and ensemble-based systems, deep learning models, and transformer-based methods.

Classical Machine Learning–Based Phishing URL Detection

Early phishing URL detection methods were based on lexical and structural features that were manually constructed and the use of classical machine learning learners. Zouina and Outtaj [1] suggested a phishing detection system that is based on Support Vector Machines (SVM) and the similarity- ity index based on string distance metrics. Their study had indicated that lexical URL features of simplicity could be used with used to obtain a decent level of detection accuracy and still be computationally efficient. Nevertheless, it did not apply well in a large-scale real world system because of the need to have such knowledge regarding target legitimate URLs.

Other studies also investigated classical supervised learning algorithms i.e. the Logistic Re- gression (LR), Decision Trees (DT), Random Forests (RF) and Support Vector Machines (SVM). Javeed et al. [8] compared several machine learning models with handcrafted URL features and found that Random Forests classifiers performed better with bench- mark datasets. On the same note, Bani Hani et al. [4] established the efficacy of machine learning classifiers to malicious URL recognition using lexical or host-based characteristics. Despite their effectiveness, such mechanisms rely on manual feature engineering and are not very flexible to changing phishing tactics.

Adaptive and Reinforcement Learning Approaches

The adaptive learning tech- niques were developed to solve the inability of the traditional machine learning models to be updated so that they become dynamic. Smadi et al. [2] suggested a phishing email detection system based on reinforcement learning on a dynamically changing neural network. The model did not require any interruption and kept refreshing its structure and parameters according to the feedback and allowed detecting zero- day phishing attacks better. Though this method added flexibility to phishing detection, it used handcrafted email characteristics and had a heavy computational cost which limited its scalability.

Ensemble Learning and Explainable Feature Selection

Ensemble learning methods learn using several classifiers in order to enhance the ability to withstand detection and misclassification errors. The article by Khedekar et al. [5] offered a voting-based ensemble framework of malicious URL detection that combined the Random Forest, XGBoost, and Logistic Regression classifiers. Their system proved to be more robust and real time deployment with the use of web based interface. Nevertheless, the methodology was still feature-based.

To improve the transparency of phishing detection systems, explainable artificial intelligence (XAI) has been embraced more of late. An explainable feature selection framework was presented by Shafin [7] that integrated SHAP, aggregated LIME explanations to rank phishing features globally and locally significantly. The experiment revealed that explainability-based feature selection has the potential to enhance the accuracy of classification and inference time. Despite this, these approaches remain based on hand-crafted characteristics and are not end-to-end representation learners.

Deep Learning–Based Phishing URL Detection

The use of deep learning was a significant change in phishing URL detection as it allowed automatic learning of features based on raw URLs. Barik et al. [9] suggested a CNN-based phishing URL detection model that has been optimized through variational autoencoders and improved grid search methods. They had a highly accurate model that was based on an optimized model but entailed a complex preprocess- ing pipeline that included feature extraction and dimensionality reduction.

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

Impact Factor 8.699

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

Volume 15, Issue 2, February 2026

|DOI: 10.15680/IJIRSET.2026.1502064|

The hybrid deep learning designs achieved even better detection results by integrating more than two neural components. The hybrid deep learning model suggested by Aljofey et al. [10] combines multi-scale CNNs, BiLSTM networks and gated MLP architectures. Their design reached state of the art accuracy with less computational overhead than the transformer based models. Such deep learning models are usually not capable of estimating uncertainty or explicitly resistant to adversarial phishing attacks despite their effectiveness.

Transformer-Based and Uncertainty-Aware Models

Recently, transformer-based models have become the cutting-edge phishing URL detection systems thanks to their capacity to achieve a long-range structure and contextual regularities in web addresses. Ghalechyan et al. [11] introduced a phishing URL detection frame-work with BERT and complemented and enhanced the approach with probabilistic neural networks to determine prediction uncertainty. Their model found confirmation in the real-world data of production and the significance of the uncertainty-aware decision-making in cybersecurity systems was observed.

To solve the high cost of computation of transformers, Jishnu and Arthi [12] proposed a real-time phishing URL detecting framework with knowledge-distilled ELECTRA. Their method of com- pressing the transformer model obtained almost state-of-the-art accuracy, and allowed deployment via browsers and constant improvement with user feedback. This paper has shown that phishing systems that rely on transformers can be successfully modified to work in real time.

Dataset-Centric and Real-Time Detection Studies

Recent researches have highlighted the relevance of diversity and scalability of the datasets in phishing detection. Sameen et al. [3] introduced a phishing URL detector called PhishHaven which was an ensemble-based system of phishing URL detection where both human-written phishing URLs and AI-generated ones could be recognized with the help of lexical and real-time deployment features. Their publication emphasized the increase in the risk of AI-driven phishing attacks.

Agoylo Jr. and Cerna [13] worked on the size of hybrid datasets built by combining various sources of phishing data like PhishTank and Kaggle in order to enhance generalization of the model. Their experiment showed that lightweight machine learning models that were trained using different kinds of data were able to deliver robust performance and at the same time were high inference velocity to be applied in real-time phishing protection system

III. ARCHITECTURE OF URL-BASED PHISHING DETECTION

The URL-based phishing detection systems are configured to detect the malicious web links basing on the nature of URLs prior to accessing the related web contents by the users. Compared to content or visual phishing detection techniques, URL techniques have benefits of speed, scalability and real-time applicability. The following section is a generalized ar- chitecture of URL-based phishing detection systems, which is synthesized on the basis of existing literature, and how various machine learning and deep learning methods fit into this framework.

Architecture Overview General

A standard URL based phishing detection system has five primary elements, namely, (i) data collection, (ii) preprocessing, (iii) feature repre- repentration, (iv) classification, and (v) decision and deployment layer. Previous works have given figure-wise descriptions of similar architectures [3], [9], [12], [13]. The general pipeline starts with acquisition of URLs and concludes with binary or multi-class decision which is an indication of legitimacy or maliciousness of the URL.

Data Collection and URL Acquisition

The initial step is to get the URLs on different locations, such as publicly available repositories like PhishTank, Kaggleasets, OpenPhish, and benign sources like Alexa or Majestic Million [3], [11], [13]. The URLs can be retrieved in real-time either using browser extensions, email gateways or network traffic monitoring systems [12]. Recent researchers note the importance of hybrid datasets that have been built out of a variety of information sources to minimize dataset bias and enhance generalization [13].

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

Impact Factor 8.699

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

Volume 15, Issue 2, February 2026

|DOI: 10.15680/IJIRSET.2026.1502064|

Table 1: Data Sources

Types	Data Source
Phishing URLs	PhishTank, OpenPhish, PhishStorm, DeepPhish
Legitimate URLs	Alexa Top Sites, Majestic Million, Google Search, Kaggle

Evaluation Metrics

The accuracy of URL-based phishing detection systems is typically measured in terms of standard classification measures based on the confusion matrix where **TP, TN, FP**, and **FN** are the true positives, true negatives, false positives, and false negatives, respectively.

Accuracy

Accuracy is a measure of the overall correctness of the classifier, which is also commonly reported both in classical, deep learning, and transformer-based literature.

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}$$

Precision

Precision is the number of URLs that are considered to be phishing out of the total URLs that are phishing.

$$\text{Precision} = \frac{TP}{TP + FP}$$

Recall (True Positive Rate)

Recall is used to measure how effective a model is in identifying phishing URLs accurately and it is an important measure in security applications.

$$\text{Recall} = \frac{TP}{TP + FN}$$

F1-Score

F1-score is a balanced measure, which is used as a combination of precision and recall.

$$\text{F1-score} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

False Positive Rate (FPR)

FPR is the ratio of valid URLs which are mistakenly identified as phishing.

$$\text{FPR} = \frac{FP}{FP + TN}$$

False Negative Rate (FNR)

FNR shows the ratio of the phishing URLs that the detection system failed to identify.

$$\text{FNR} = \frac{FN}{FN + TP}$$

Area Under the Curve (AUC)

AUC is a metric that provides an evaluation of the capability of the classifier to differentiate between phishing and legitimate URLs at varying thresholds.

$$\text{AUC} = \int_0^1 \text{TPR}\, d(\text{FPR})$$

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

Impact Factor 8.699

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

Volume 15, Issue 2, February 2026

|DOI: 10.15680/IJIRSET.2026.1502064|

Feature Representation and Encoding

The representation of the features is a critical issue in URL-based phish- ing detection and it differs according to the underlying learning model. In classical machine method of learning, the handcrafted features are identified by URLs, and known lexical features (URL length, dots, the presence of special characters), host based features (domain age, DNS information), content related indicators are used where available [4], [8]. The most discriminative features and better interpretability have been achieved by explainable feature selection techniques [7].

Deep learning models minimize the use of manual feature engineering, through direct learning of a representation of raw URL strings. Convolutional Neural Networks (CNNs) and recurrent models encode URLs as character (or token) sequences, which encode spatial and temporal patterns [9], [10]. Hybrid deep learning models also optimize representation learning through integration of two or more neural parts [10].

Models that utilize transformers utilize contextual embeddings that are created using tokenized URLs, which allows the formation of long-range dependencies and semantic relationships [11], [12]. Compressed representations are used in knowledge-distilled trans- formers to achieve a trade-off between accuracy and computational efficiency [12].

Classification and Detection Models

The classification layer is used to identify the level of phishing and legitimate URL. For their simplicity and efficiency, classical classifiers, including Logistic Regression, Support Vec-tor Machines, Random Forests, and XGBoost, have been at large scale usage [4], [8], [13]. Ensemble classifiers also enhance robustness through prediction by combining models [5], [3].

Classifiers based on deep learning are end to end classifiers that use learned representations to do end-to-end classification. The CNN-based as well as hybrid deep learning models have shown high detection accuracy over traditional methods [9], [10]. The latest trend in classifiers is transformer based classifiers which are capable of a high degree of accuracy in encoding intricate URL patterns [11], [12]. Other modern systems also add uncertainty estima- tion mechanisms to give the level of confidence in the predictions to enhance credibility in security-important systems [11].

Decision Layer and Deployment

The last phase in the architecture is decision making and implementation. Depending on the classifier output, URLs are categorized as legitimate or phishing with corresponding measures being taken i. e. blocking ac- cess, warning or recording suspicious behavior. The strategies of deploying in real time encompass browser extensions, email filtering systems and backend security APIs [3], [12], [13].

The considerations related to deployment, i.e., inference latency, computational cost, and scalability are essential to a practical adoption. Knowledge-distilled trans- formers Lightweight machine learning models and knowledge-distilled trans- formers are usually favored in real-time applications because they have fewer resource demands [12], [13]. Also, other systems have user feedback systems to facilitate ongoing learning and adaptation to new phishing methods [12].

IV. SUMMARY AND OPINION

This review examined the recent developments in URL-based phishing detection including the classical machine learning methods up to deep learning and transformer-based models. Initial feature-driven models with classifiers like Logistic Regression, Support Vector Machines, and Random Forests performed reasonably well but lacked the ability to perform adaptive feature engineering to changing phishing attacks as well as manual feature engi- neering [1], [4], [8]. The methods of ensemble learning and explainable feature selection enhanced robustness and interpretability, but were mostly in nature stagnant [5], [7]. The deep learning models have markedly improved the detection accu- racy by training on raw URLs, without the need to rely on handcrafted features [9], [10]. Transformer-based methods also developed the field by capturing con- textual URL patterns and providing uncertainty-aware decision making, and knowledge dis- tillation helped to deploy the application in real-time [11], [12]. The models, however, bring increased complexity to computation. Generally, there is no specific method that will consider all the issues in phishing URL detection. Lightweight machine learning models can still be taken to the real-time and resource-bounded environment, but the deep learning and transformer-based models can be taken to further levels in providing improved accuracy and

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

Impact Factor 8.699

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

Volume 15, Issue 2, February 2026

|DOI: 10.15680/IJIRSET.2026.1502064|

flexibility. The new systems must combine efficient transformer structure, explainability and continuous learning to present robust and deployable phishing detection systems.

V. CONCLUSION

This review provided the detailed evaluation of URL-based phishing detection algorithms, including classical machine learning, ensemble, deep learning architectures, and transformer- based algorithms. Although conventional feature-based solutions are efficient and have simplicity in deploying them, they are not effective against changing phishing techniques. Transformers are capable of outperforming other detection methods such as deep learning because they learn intricate features of images.

raw URLs and have increased computational overhead. Recent developments including knowledge distillation, uncertainty-sensitive models and large-scale hybrid datasets indicate some fruitful avenues to scaling to practice. In general, the accurate, efficient, understandable, and adaptable combination of accuracy, explainability, and efficiency is necessary to achieve effective phishing URL detection. Future investigations ought to revolve around lightweight, explainable, and continually learning models in order to improve real- world phishing defense systems.

Table 2: Comparison of Evaluation Results in Existing Phishing Detection

Ref.	Year	Approach Type	Features Used	Model(s)	Dataset(s)	Classification	Reported Accuracy / Performance
[1]	2017	Classical ML	Lexical URL + similarity index	SVM	PhishTank, Alexa	Binary	95.8%
[2]	2018	Adaptive ML (RL)	Handcrafted email features	Dynamic NN + RL	PhishingCorpus, SpamAssassin	Binary	98.6%
[3]	2020	Ensemble ML (Real-time)	Lexical URL features	Voting-based Ensemble	PhishTank, Alexa, DeepPhish	Binary	~98%
[4]	2024	Classical ML	Lexical + host features	RF, SVM, DT	PhishTank, Kaggle	Binary	~99%
[5]	2025	Ensemble ML	Lexical + TF-IDF	Voting Classifier	Public URL datasets	Multi-class	~94%
[6]	2025	Classical ML	URL, HTML, domain features	RF, LR	UCI Phishing Dataset	Binary	97.7%
[7]	2025	XAI-based ML	URL + content features (selected via SHAP & LIME)	RF, XGBoost	Web Phishing Dataset	Binary	97.4%
[8]	2025	Classical ML	Handcrafted URL features	RF, SVM, AdaBoost	UCI Dataset	Binary	97.7%
[9]	2025	Optimized DL	TF-IDF + VAE	CNN + EGSO	Kaggle, PhishTank	Binary	99.4%
[10]	2026	Hybrid Deep	Raw URLs	Multi-scale CNN +	Multiple URL datasets	Binary	99.8%

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

Impact Factor 8.699

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

Volume 15, Issue 2, February 2026

|DOI: 10.15680/IJIRSET.2026.1502064|

		Learning		BiLSTM + gMLP			
[11]	2024	Transformer-based DL	Raw URLs	BERT + PNN	PhishTank, OpenPhish, Alexa	Binary	97–99%
[12]	2024	Distilled Transformer (Real-time)	Raw URLs	KD-ELECTRA	Large-scale hybrid datasets	Binary	99.7%
[13]	2026	Dataset-centric ML	Character n-grams	LR, RF, XGBoost	PhishTank + Kaggle (Hybrid)	Binary	92.7%
[14]	2025	Ensemble ML	Handcrafted URL features	Ensemble Classifiers	Public URL datasets	Binary	~96–98%

REFERENCES

1. M. Zouina and B. Outtaj, "A novel lightweight URL phishing detection system using SVM and similarity index," Human-centric Computing and Information Sciences, vol. 7, no. 1, 2017.
2. S. Smadi, N. Aslam, and L. Zhang, "Detection of online phishing email using dynamic evolving neural network based on reinforcement learning," Decision Support Systems, vol. 107, pp. 88–102, 2018.
3. M. Sameen, B. Alshammari, M. K. Hasan, and M. A. Hossain, "PhishHaven—An efficient real-time AI phishing URLs detection system," IEEE Access, vol. 8, pp. 83490–83504, 2020.
4. S. Bani Hani, A. Al-Zu'bi, and M. Al-Sarayreh, "Malicious URL detection using ma- chine learning," in Proc. IEEE Int. Conf. on Intelligent Computing and Information Systems (ICICS), 2024.
5. K. Khedekar, S. Patil, and R. Kulkarni, "Malicious URL detection using machine learning," in Proc. IEEE CSITSS, 2025.
6. Y. Pang, X. Li, and J. Wang, "Enhancing phishing website detection with machine learning algorithms," in Proc. IEEE Int. Symp. on Digital Forensics and Security (ISDFS), 2025.
7. S. S. Shafin, "An explainable feature selection framework for web phishing detection with machine learning," Data Science and Management, vol. 8, no. 1, pp. 1–15, 2025.
8. M. Usman Javeed, A. Khan, and S. Ahmad, "Phishing website URL detection using a hybrid machine learning approach," Journal of Computing Biomedical Informatics, vol. 9, no. 1, pp. 25–36, 2025.
9. K. Barik, S. Misra, and R. Mohan, "Web-based phishing URL detection model using deep learning optimization techniques," International Journal of Data Science and Analytics,Springer, 2025.
10. M. Aljofey, A. Al-Qahtani, and A. Hussain, "A hybrid deep learning model for robust phishing URL detection: Integrating multi-scale CNNs, BiLSTM, and gated MLP architec- tures," Computers and Electrical Engineering, Elsevier, 2026.
11. G. Ghalechyan, A. Karapetyan, and V. Ghazaryan, "Phishing URL detection with neural networks: An empirical study," Scientific Reports, Nature Portfolio, vol. 14, 2024.
12. K. S. Jishnu and B. Arthi, "Real-time phishing URL detection framework using knowledge distilled ELECTRA," Automatika, vol. 65, no. 4, pp. 621–637, 2024.
13. J. C. Agoylo Jr. and P. D. Cerna, "From data to defense: Leveraging PhishTank and multi- source hybrid datasets for intelligent website-level phishing protection," International Journal of Engineering Trends and Technology, vol. 74, no. 3, pp. 45–56, 2026.
14. A. Verma and R. Singh, "Malicious URL detection using ensemble machine learning tech- niques," International Journal of Information Security Science, vol. 14, no. 2, pp. 101–112, 20

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