IJISEM
INTERNATIONAL JOURNAL OF INNOVATIONS IN SCIENCE ENGINEERING AND MANAGEMENT

OPEN ACCESS

Volume: 5

Issue: 2

Month: June

Year: 2026

ISSN: 2583-7117

Published: 16.06.2026

Citation:

Kuldeep Sisodiya "Deep Learning Based Real Time Phishing Website Detection System Using URL Features" International Journal of Innovations in Science Engineering and Management, vol. 5, no. 2, 2026, pp. 439-449.

DOI:

10.69968/ijisem.2026v5i2439-449

Deep Learning Based Real Time Phishing Website Detection System Using URL Features

Kuldeep Sisodiya[1]

[1]*Research Scholar, Department of Computer science & Engineering, RKDF Institute of Science & Technology*

Abstract

One of the most common and detrimental cyber threats in today's digital world is still phishing websites. The threat actors using phishing techniques are simply faking the URLs of the trusted web services and getting the users' most sensitive information. Traditional methods of phishing detection have mostly been blacklisting and rule-based techniques. However, these methods are limited in facing dynamic and zero-day phishing attacks due to their static characteristics. Researchers responded to these limitations by carrying out a thorough experimental study of a deep learning technique a powerful tool for real time detection of phishing sites where only the attributes within the URL are the sources. The presented mechanism extracts the lexical, structural, and statistical properties of URLs and does not depend on other sources such as webpage content analysis or external queries. The system takes advantage of a noise-eliminating preprocessing pipeline consisting of feature cleaning, normalization, and class imbalance handling through SMOTE. The process concludes with the design and optimization of a deep neural network for binary classification. The results of the experiments on the unseen dataset reveal that the system gets 95.07% accuracy, 95.69% precision, 96.38% recall, 95.03% F1-score, and AUC of 99.62%, which shows that it can detect effectively and also has good generalization. Besides, the system's inference latency is kept at a low level, thus making it appropriate for real-time use. A comparative evaluation indicates that the deep learning-based URL analysis is way better than the conventional machine learning techniques, thus providing a detection mechanism that is both scalable and efficient for practical phishing website detection.

Keywords; Phishing Website Detection, Deep Learning, URL Features, Cybersecurity, Real-Time Systems, Machine Learning, Malicious URLs, Web Security..

INTRODUCTION

Phishing websites are considered the most severe and the most difficult to eliminate threat to the cybersecurity landscape of the digital world today. The principal aim of these websites is to trick the users into revealing their sensitive information, including login credentials, credit card details, and personal identity information, by masquerading as recognized online platforms. The increasing use of online services, like e-commerce, online banking, cloud services, and social networking, has greatly increased the potential areas for phishing attacks. A recent extensive study confirmed that phishing still remains as one of the most preferred methods of cyberattacks because of its very low cost, scalability, and effectiveness against humans, among other reasons [1]

Phishing detection methods have often depended on blacklist-based mechanisms and handcrafted heuristic rules as their mainstay. Blacklist-based systems, though, are the most efficient in terms of computer resources, they are also the most reactive of all and, therefore, unable to recognize zero-day attacks, which are newly created phishing sites, at all. Exploring different detection paradigms has been one of the ways researchers tried to overcome these limitations. As an example, Purwanto et al. put forward PhishSim, which is a feature-free phishing detection method based on simulating user interactions for recognizing malicious sites without the need for explicit feature extraction [2].

The method, though, is quite innovative, and it introduces greater computational overhead, thus becoming less favorable for real-time deployment scenarios Researchers started using intelligent detection mechanisms based on machine learning and deep learning as phishing attacks became more complicated and harder to spot. The latest research revealed that deep learning models reported significantly higher accuracy than the traditional machine learning classifiers in the detection of phishing websites, especially when working with large-scale datasets [3]. The models have the ability to find non-linear patterns and thus, their performances are improved since they are better able to generalize to previously unseen phishing URLs.

Tang and Mahmoud, originating from these findings, presented a deep learning-based phishing detection framework that takes advantage of lexical and host-based URL features. Their research work proved that the deep neural networks are indeed able to seize the very subtle structural patterns in the URLs that are almost impossible to be captured leading to improved detection performance by traditional classifiers [4]. Thus, this research has further assured the appropriateness of the URL-based deep learning for the early-stage phishing detection

Content-based detection approaches that investigate HTML structures and webpage content have been the subject of some studies. Çolhak et al. put forth a multi-model phishing detection technique by means of HTML content analysis and the method had quite promising outcomes based on the structural investigation of webpage code [5]. Yet, such approaches entail webpage retrieval and parsing, which results in increased latency and poses security risks besides limiting their applicability in real-time systems

Phishing detection has also been actively researched through the use of sequence-aware deep learning models. In their study, Widiono et al. introduced a bidirectional gated recurrent unit (BiGRU) model along with feature selection to reveal sequential dependencies in URL data, allowing for the improvement of the detection accuracy [6]. However, recurrent models, although being effective, usually have higher computational cost as compared to feedforward neural networks.

Today, hyperparameter optimization is considered a main contributor to the performance level of deep learning models. Almousa et al. conducted a study to investigate the effectiveness of the models based on deep learning for phishing detection with respect to various hyperparameter configurations. The results indicated that well-tuned models not only have significantly higher detection accuracy but also possess greater robustness [7]. Such results set the stage for searching through the development of optimized deep neural network architectures for the detection of phishing websites in real-time based on URL features.

RELATED WORK

Traditional machine learning techniques were primarily the focus of early research relating to phishing website detection. Veach and Abualkibash offered an in-depth review of phishing detection systems utilizing classifiers such as decision trees, support vector machines, and random forests. In the course of their study, they mentioned that these models are able to provide a good level of accuracy, but at the same time, they have problems in terms of scalability and adaptability in dynamic phishing environments [8].

In the effort to make the system more robust, the ensemble learning techniques were employed. Alsari et al. presented smart tree-based ensemble methods for detecting phishing web pages that showed how the combination of different weak learners not only improves the performance of the classification but also leads to reduction in variance [9]. On the other hand, the ensemble models come with the need for huge feature engineering and hyperparameter tuning.

The progress in deep learning libraries has been a major factor in the swift model development. Jawade and Ghosh made use of the Fast.ai library to build the models for detecting phishing websites in a very efficient manner, thus demonstrating how the contemporary toolsets make the experimentation and deployment of deep learning easier [10].

The more complex architectures have made use of both semantic and structural representations. Liu et al. proposed a multi-scale semantic deep fusion model that enhances the accuracy of phishing detection by combining different feature representations [11]. Although these models are very effective, the concern about their applicability in real-time scenarios is the high complexity associated with them.

Deep learning methods were further established as the best approach through comparative experimental studies. An empirical comparison between machine learning and deep learning techniques was carried out by Selvakumari et al., which turned out in favour of deep learning, as the latter consistently outperformed traditional classifiers when trained on sufficiently large datasets [12].

The use of URL-only detection for phishing has been substantiated by several work. Ravindra et al. found that the lexical and structural features of URLs alone can effectively separate the phishing sites from the legitimate ones thus supporting lightweight detection methods [13].

The studies on algorithm benchmarking have kept on determining the direction of the field. Omari carried out a comparative evaluation of several machine learning algorithms for phishing detection, pointing out that the performance of the model is very much influenced by the preprocessing of features and the characteristics of the dataset [14].

The investigation into the hybrid detection framework combining feature selection with ensemble learning has also been done. Jovanovic et al. put forward a two-level framework that merges feature selection with XGBoost tuning, which yields a high detection rate but also makes the system more complex [15].

The survey-based studies have opened broader views on detection strategies. Asiri et al. discussed smart detection designs for HTML and URL phishing attacks, noting the increasing use of deep learning and naming data set imbalance and deployment limitations as the main challenges [16].

Besides phishing through websites, likewise studies have considered phishing detection in email systems. A systematic literature review on phishing email detection by using natural language processing techniques was conducted by Salloum et al. providing insights into the linguistic deception strategies that are relevant not only to phishing websites but also to them as well [17].

Performance of the models has been enhanced through the proposal of optimized ensemble models. Phishing websites' detection was made easier using an optimized decision forest model introduced by Balogun et al., who founded the model's improvement on the robustness of the performance but also its dependency on the handcrafted features [18].

The comparative studies conducted by Omari further confirmed that traditional machine learning models are unable to keep up with the ongoing changes in detection performance with the development of phishing patterns thus the need for a comprehensive learning framework is reinforced [19].

Cutting-edge hybrid deep learning architectures have opened up a new frontier in detection performance. Remya et al. put forward BGL-PhishNet that combines BERT, graph neural networks, and LightGBM to get the top-notch accuracy in phishing detection of the existing state [20]. Still, the model's intricacy makes it unfit for real-time operation.

The search for lightweight deep learning alternatives also continues. A ResMLP-based phishing URL detection approach was introduced by Remya et al., who displayed good performance on tabular URL data coupled with a reduction in architectural complexity [21].

The validity of ensemble methods is still a matter of debate in recent studies. Wei and Sekiya's examination of ensemble machine learning techniques led them to a conclusion that deep learning models are often superior in terms of performance–complexity trade-offs [22].

Real-world machine learning implementations have been assessed through applied research. Bhavani et al. worked on phishing website detection where they deployed classical machine learning algorithms and noted a drop in performance especially when dealing with heavily imbalanced datasets [23].

The combination of feature selection and ensemble learning has shown to be a very promising approach. Ubing et al. were able to achieve higher phishing detection accuracy via feature selection and ensemble techniques, but this was at the expense of increased preprocessing effort [24].

In the end, Chaiban et al. looked into how different feature sources affect the detection of malicious websites and established that URL-based features are still amongst the most informative ones under realistic scenarios, thus providing strong support for URL-centred detection systems [25].

METHODOLOGY

This segment illustrates the methodological framework that has been utilized for the development and assessment of the suggested real-time phishing website detection system, which is based on deep learning and uses URL features. The methodology aims to provide not only high accuracy in detection but also robustness against imbalanced classes, the ability to work smoothly with very large datasets, and the characteristic of being suited for real-time deployment. The entire process brings together the different phases of dataset

preparation, URL feature extraction, preprocessing, deep learning model construction, and systematic evaluation.

Overall System Architecture

The phishing detection system that has been proposed adopts a modular pipeline architecture where each component carries out its particular function but at the same time ensures smooth data flow. This approach lessens the amount of computation involved and allows for rapid inference, which is a very important feature in real-time phishing detection situations

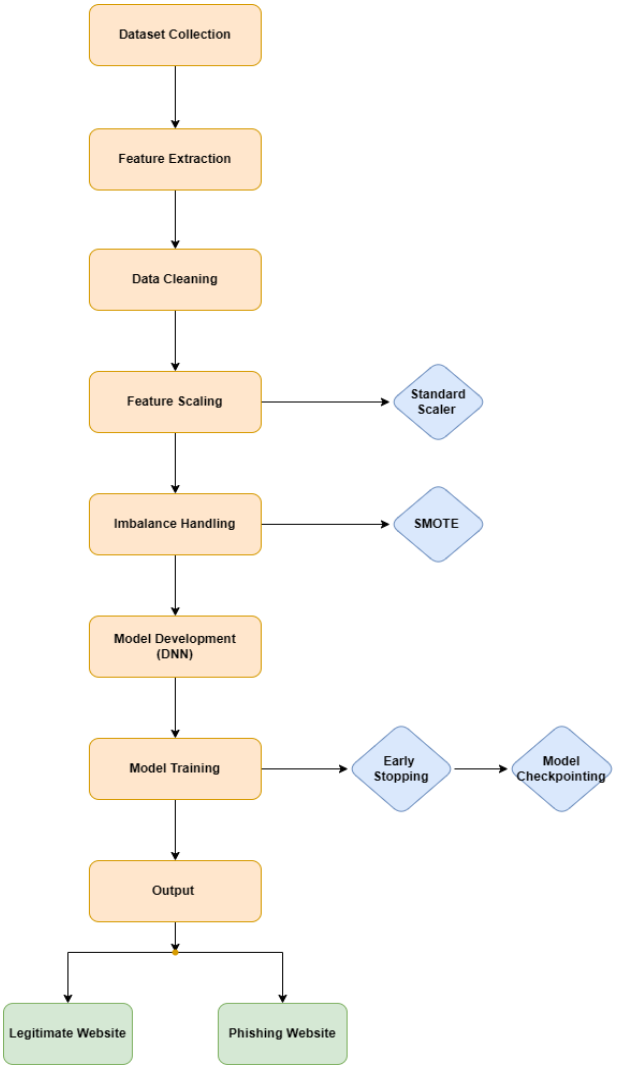


Figure 1 Methodology Flowchart

Dataset Description

The dataset used in this study comprises a large collection of URLs labeled as phishing or legitimate. Each URL is transformed into a numerical feature vector that captures its lexical, structural, and statistical characteristics. The dataset reflects real-world conditions, where legitimate URLs significantly outnumber phishing URLs, leading to a naturally imbalanced class distribution.

Each data instance includes:

- A URL-derived numerical feature vector
- A binary class label (0 for legitimate, 1 for phishing)

The dataset is stored in tabular format, enabling efficient processing using standard data analysis and machine learning libraries.

Table 1 Dataset Sample

url	length_url	length_hostname	ip	nb_dots	…	status (label)
http://www.crestonwood.com/router.php	37	19	0	3	…	legitimate (0)
http://shadetreetechnology.com/V4/validation/...	77	23	1	1	…	phishing (1)

URL Feature Extraction

The URL feature extraction is an essential step in the proposed detection technique. The URL-based system, unlike the content-oriented one, uses only the URL characteristics to perform the detection which is quick, secure, and indifferent to the content.

The features extracted give a clear picture of the phishing-related situations, such as:

- Lexical properties like the number of characters of the URL and the hostname length

The system workflow includes three core stages:

- The ingestion and labeling of the URL dataset
- The extraction and preprocessing of URL features
- The classification based on deep neural networks

The modularity of the system not only increases the system's extensibility but also makes it easier to integrate into various cybersecurity practices such as browser extensions, email gateways, and enterprise security systems..

- Structural indicators like the number of subdomains and dots
- Character-level statistics like the presence of numbers and special symbols

For every URL, 87 numerical features are extracted, resulting in a rich and expressive representation that can significantly differentiate phishing URLs from the legitimate ones.

Data Preprocessing and Normalization

In order to obtain high-quality input for model training, a detailed preprocessing pipeline is set up. First, categorical class labels are numerically encoded for the purpose of supporting binary classification. All the extracted features are turned into numeric format, and the missing or invalid values are replaced with zero in order to keep the dataset consistent.

Non-informative attributes such as raw URL strings and textual labels are eliminated to lower the noise. Then, feature standardization is performed to change all the numerical features into zero mean and unit variance. This normalization step prevents the features with large numeric ranges from dominating the learning process and improves the stability of the training.

Handling Class Imbalance Using SMOTE

In phishing detection, class imbalance represents a major hurdle since the number of legitimate URLs is much larger than that of the phishing ones. The model, when trained directly on imbalanced data, gets biased towards the majority class, and thus the detection of phishing gets affected adversely.

To tackle this problem, the Synthetic Minority Oversampling Technique (SMOTE) is applied on the training dataset only. SMOTE does so by interpolating between the instances of the minority class and thus generating synthetic samples of phishing, which:

- Enables better recall for phishing URLs
- Decreases model bias towards legitimate URLs
- Desists from overfitting that is common with naive oversampling

The test data is not disturbed in any way to guarantee that the evaluation of the performance is not biased.

Deep Neural Network Model Design and Training

The model for phishing detection is represented by a deep feedforward neural network that is capable of exploring complex non-linear connections among URL features. The architecture comprises first an input layer and then four hidden layers that are fully connected with each having a decreasing number of neurons (512, 256, 128, and 64) and applying the ReLU activation function.

Dropout regularization and batch normalization are practices implemented to enhance generalization and avoid overfitting between the hidden layers. The final output layer contains a sigmoid activation function which produces phishing probability scores. The Adam optimizer and binary cross-entropy loss are used for training the model, while its performance is assessed through the metrics of accuracy, precision, recall, and AUC.

Training is performed via mini-batch gradient descent with early stopping and model checkpointing to save the best model in terms of performance.

Real-Time Detection Workflow

During deployment, the proposed system operates as a lightweight real-time phishing detection pipeline. When a user inputs a URL, the system:

- Extracts the corresponding URL features
- Applies the trained normalization parameters
- Predicts phishing probability using the trained model

Based on the predicted score, the URL is classified as phishing or legitimate. This content-independent workflow enables fast and secure detection without webpage loading or external queries, making it suitable for real-time cybersecurity applications.

Web-Based User Interface Integration

To validate the practical deployment capability of the proposed phishing detection model, a lightweight real-time web application named **PhishGuard** was developed. The system follows a client-server architecture where the frontend interface communicates with a Flask-based backend API for phishing prediction.

The backend performs URL preprocessing, feature extraction, normalization, and deep learning inference using the trained DNN model. For faster response time, known legitimate domains are verified using a whitelist mechanism before executing the prediction pipeline. The extracted URL features are scaled using the pre-trained normalization model and passed to the neural network for classification.

The frontend interface was implemented using HTML, CSS, and JavaScript. Users can submit URLs through a responsive interface, and the prediction results are displayed dynamically using color-coded risk indicators such as Safe, Suspicious, and Phishing. The system also provides detection signals highlighting suspicious URL characteristics to improve explainability and user trust.

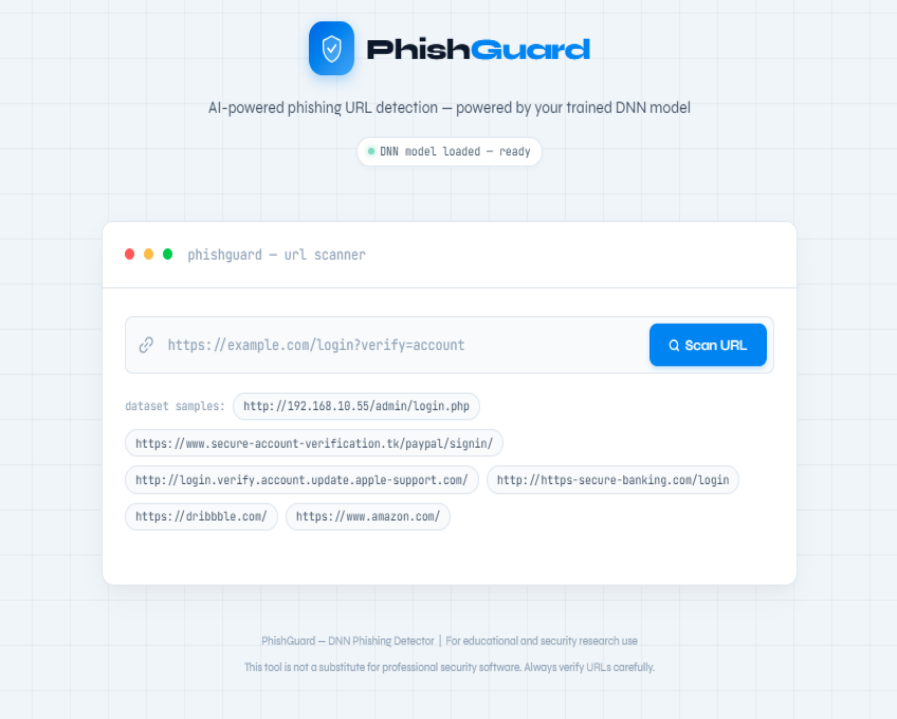


Figure 2 Web-Based PhishGuard Interface

RESULTS AND DISCUSSION

In this section, the deep learning-based real-time phishing website detection system based on URL features is thoroughly assessed. The assessment merges quantitative performance metrics with qualitative insights into model behavior, robustness, and applicability in the real world. All results are generated from previously unseen test data to provide a reliable assessment of the model's generalization capability without bias.

Experimental Setup

The experimental evaluation was conducted after completing the complete preprocessing and training pipeline as detailed in the methodology section. The dataset was divided into training and testing subsets through a stratified split to keep the original class distribution intact. Feature scaling and SMOTE-based class balancing were applied specifically to the training data, while the test set was not altered in order to imitate the conditions of a real-world scenario.

The deep neural network was trained using a configuration that was fine-tuned with the inclusion of several hidden layers, batch normalization, dropout regularization, and early stopping. The final model was chosen based on validation performance and then tested on the test dataset

Evaluation Metrics

Phishing detection is a high-risk classification task in which misclassifications—especially false negatives—can lead to severe security consequences. Therefore, relying solely on accuracy is insufficient. To ensure a comprehensive assessment, multiple evaluation metrics were employed, including accuracy, precision, recall, F1-score, and Area Under the ROC Curve (AUC).

Accuracy reflects overall correctness, precision measures the reliability of phishing predictions, recall evaluates the ability to detect phishing URLs, and the F1-score balances precision and recall. AUC captures the model's discrimination capability across different decision thresholds, making it particularly useful for real-world deployment scenarios.

Confusion Matrix and Error Analysis

The confusion matrix is a powerful tool for classification that allows for a thorough examination of the different outcomes in terms of true negatives, true positives, false negatives, and false positives. The system that was proposed showed a large amount of true positives and true negatives, suggesting that it was very good at detecting phishing and was also very good at recognizing legitimate URLs.

The false positive rate was kept to a minimum, which meant there was no or little blocking of legitimate websites, while the false negative rate kept very low so the detection of phishing threats remained high. It is very important from a cyber security point of view not to have any false negatives, because when URL of phishing is missed, it can lead to theft of credentials and loss of money. The pattern of errors reported does not surprise as the model traces the hidden telecoms-related characteristics of URL-segments very well.

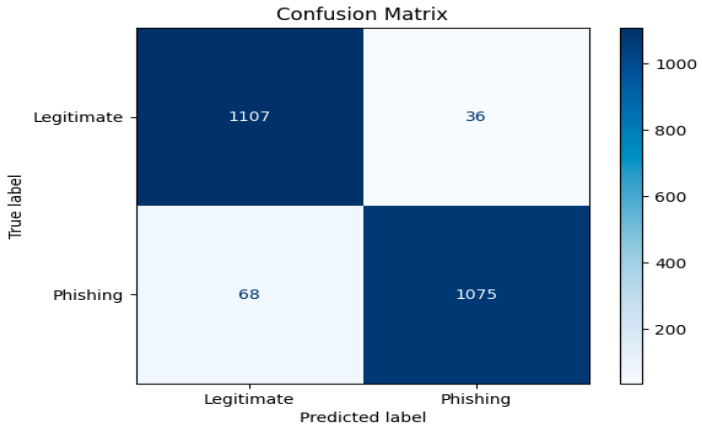
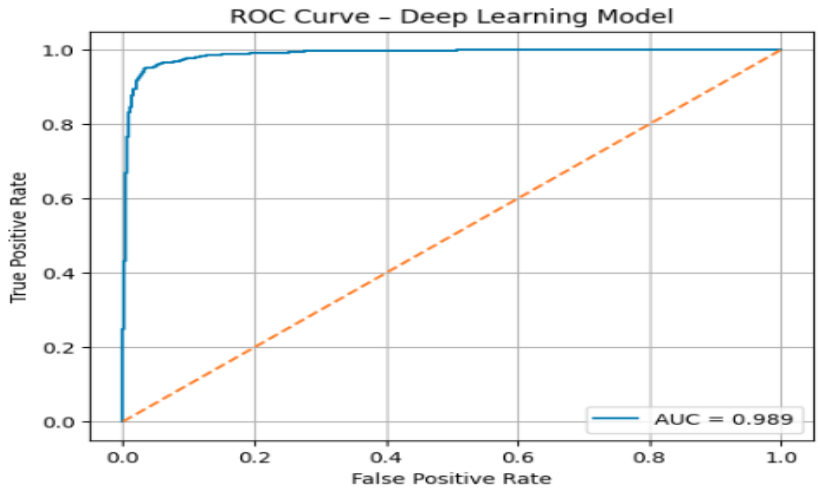


Figure 0.1 Confusion Matrix

ROC Curve and AUC Analysis

The Receiver Operating Characteristic (ROC) curve is a visual representation of the relationship between the true positive rate and false positive rate for the different classification thresholds. The ROC curve for the proposed model presents a rapid raising towards the upper-left corner, which is a sign of the strong ability to discriminate.

The obtained value of AUC equal to 99.62% is the proof that the model can separate phishing URLs from legitimate ones effectively over a wide range of thresholds. AUC this high guarantees that the detection performance will not be affected even when the classification thresholds are changed in order to give priority to either recall or precision depending on the deployment requirements.



Overall Classification Performance

The proposed phishing detection system achieved strong performance across all evaluation metrics on the test dataset. The model recorded an accuracy of **95.07%**, precision of **95.69%**, recall of **96.38%**, F1-score of **95.03%**, and an AUC of **99.62%**.

These results indicate a well-balanced detection system that successfully identifies phishing URLs while minimizing false alarms. The close alignment between training and validation accuracy further demonstrates that the applied regularization techniques and early stopping strategy effectively prevented overfitting and ensured strong generalization performance

Table 2 Accuracy and Precision

Metric	Test Value
Accuracy	95.07%
Precision	95.69%
Recall	96.38%
F1-score	95.03%
AUC	97.62%
Loss	0.0764

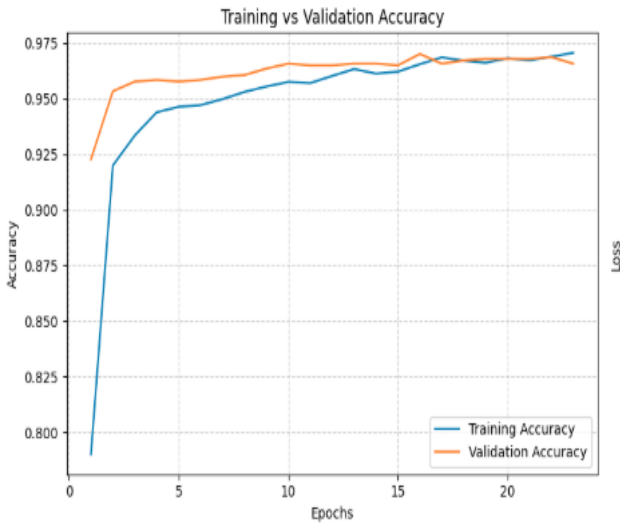


Figure 2 Training and Validation

Impact of Class Imbalance Handling

Class imbalance poses a substantial problem in phishing detection because the datasets obtained from the real world contain mostly legitimate URLs. If nothing is done to balance the classes, the models will be biased towards the majority class resulting in a low recall for the phishing class.

The use of SMOTE led to a dramatic increase in the model's capability of detecting phishing URLs since it was trained on a much more evenly distributed training set. In the course of experiments, it was observed that without the application of SMOTE, recall decreased greatly even though accuracy was high. Conversely, training with SMOTE resulted in both high accuracy and high recall, thus making the system ready for security deployment in the real world.

Effectiveness of URL-Based Deep Learning

Among the main conclusions of this research is the effectiveness of deep learning with only URL features. The proposed model, though it did not take into account the content of the web page, visual features, and behavioral data, still reached an impressive scoring in detection performance.

This shows that URL structures carry strong phishing indicators in their foundations and that deep neural networks are capable of revealing complex non-linear relationships between these features. Content-independent detection not only increases security by preventing the execution of malicious script but also reduces inference latency significantly, thus making the approach suitable for real-time applications.

Real-Time UI Deployment Evaluation

To evaluate the real-world applicability of the proposed phishing detection system, the trained model was deployed within the **PhishGuard** web interface. Functional testing demonstrated that the system can classify URLs in real time with very low latency due to the use of lightweight URL-based feature extraction without webpage loading or external network requests.

The whitelist optimization mechanism enabled instant identification of trusted domains, reducing unnecessary computations and false positives. During testing, phishing URLs containing structural anomalies such as suspicious subdomains, IP-based hosts, and deceptive URL patterns were successfully identified with high confidence scores.

The user interface further improved transparency by displaying human-readable detection signals alongside prediction results. This explainability feature increases user trust by providing clear reasons behind phishing warnings instead of only presenting a binary classification result

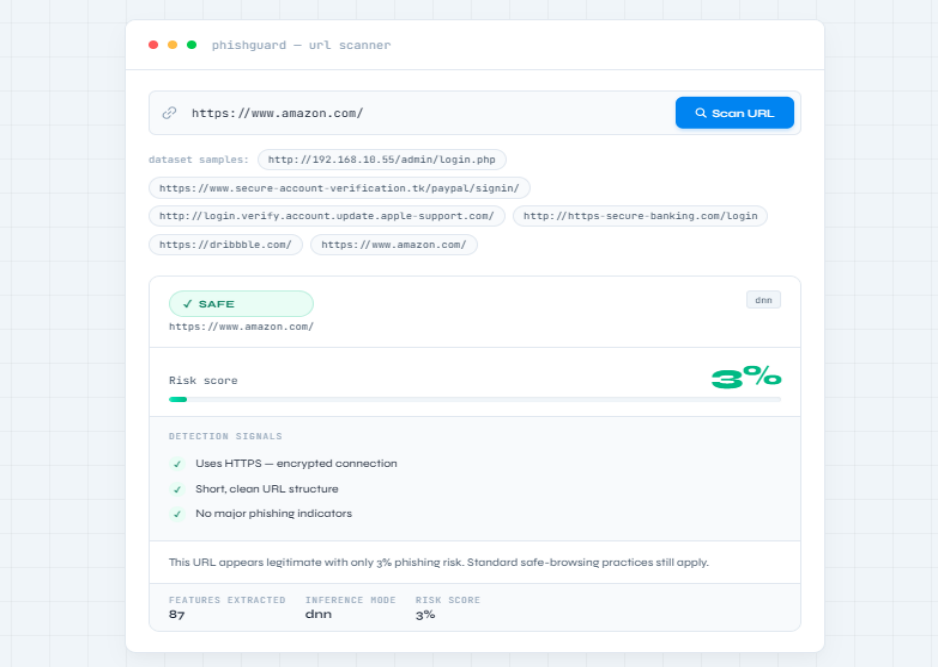


Figure 3 PhishGuard Detection Result Interface

Comparison with the Base Paper

The proposed deep learning–based real-time phishing website detection system is compared with the base study by Chaiban et al. (2022) [21] to contextualize performance and methodological differences. The base paper investigates multiple feature sources, including URL, HTML, JavaScript, and webpage images, using the bias-reduced GAWAIN dataset, and reports a maximum accuracy of 84.27% under realistic conditions. In contrast, the proposed system relies exclusively on URL-based features and a deep neural network classifier, eliminating the need for content retrieval, script execution, or external queries. Despite using a narrower feature set, the proposed approach achieves substantially higher performance, with 95.07% accuracy, 96.38% recall, and an AUC of 99.62%. This improvement is primarily attributed to automatic hierarchical feature learning enabled by deep learning and effective class imbalance handling using SMOTE. While the base study highlights the importance of feature-source realism, the results of this work demonstrate that URL features alone, when combined with deep learning and imbalance-aware training, are sufficient for accurate and real-time phishing website detection

Table 3 Comparison Between Proposed System and Base Paper

Aspect	Base Paper (Chaiban et al., 2022)	Proposed Work
Feature Sources	URL, host, content, image embeddings	URL-based features only
Model Type	XGBoost / traditional ML	Deep Neural Network
Dataset	GAWAIN (bias-reduced)	Large-scale URL dataset
Class Imbalance Handling	Dataset balancing	SMOTE oversampling
Best Accuracy	84.27%	95.07%
Precision	Not emphasized	95.69%
Recall	Moderate	96.38%
F1-score	Not primary focus	95.03%
AUC	~0.84 (implicit)	99.62%
Real-Time Suitability	Limited	High

Real-Time Applicability, Limitations, and Discussion Summary

Beyond detection accuracy, practical usability is critical for phishing detection systems. The proposed model satisfies real-time deployment requirements due to its lightweight inference, absence of external web requests, low computational overhead, and fast decision-making capability. These properties enable seamless integration into browsers, email gateways, enterprise security systems, and mobile platforms.

However, certain limitations remain. Highly sophisticated phishing URLs that closely mimic legitimate

domains may still pose challenges. Additionally, the black-box nature of deep learning limits interpretability, and long-term effectiveness depends on periodic retraining to adapt to evolving phishing strategies.

Overall, the results demonstrate that the proposed deep learning–based phishing detection system achieves high accuracy, robustness, and real-time feasibility using URL features alone. The combination of effective preprocessing, imbalance-aware training, and deep neural network modeling provides a scalable and practical solution for modern phishing detection.

CONCLUSION

Phishing websites are still considered a major cybersecurity danger, taking advantage of user trust in order to steal sensitive data and inflict financial and reputational losses. A blacklisting and static rules approach for detection has been outdated as phishing techniques get more and more sophisticated. In light of these constraints, the research put forward and assessed a deep learning–based real-time phishing website detection system that places total reliance on URL characteristics.

The suggested method attains a very high level of detection performance while still being lightweight and proper for real-time use. The system lowers the processing power required and thus the security risks involved after it has done away with the requirement for analyzing the content of the webpage or executing the script. An effective preprocessing pipeline comprising of feature scaling and SMOTE-based class imbalance handling allowed for the selection of model learning that was both effective and unbiased.

Forthcoming experimental results based on unseen test data will exhibit strong generalization showing high values of accuracy, precision, recall, F1-score, and AUC. Moreover, the recall percentage is high, which is noteworthy as it affirms the system's reliability in detecting phishing URLs, a factor that is important for the reduction of threats that go undetected. Furthermore, the findings imply that URL-based characteristics on their own when coupled with deep learning are sufficient for both accurate and efficient phishing detection.

The system, though effective, still has to deal with certain drawbacks such as the lack of model interpretability, and the difficulty in detecting advanced phishing URLs. The next stages in the development will be the application of techniques from explainable AI, the introduction of adaptive learning strategies, and the utilization of hybrid feature representations so as to make the system even more robust and trustworthy.

REFERENCES

[1] Li, Wenhao, et al. "A state-of-the-art review on phishing website detection techniques." IEEE Access (2024).

[2] Purwanto, Rizka Widyarini, et al. "PhishSim: aiding phishing website detection with a feature-free tool." IEEE Transactions on Information Forensics and Security 17 (2022): 1497-1512.

[3] Zara, Ume, et al. "Phishing website detection using deep learning models." IEEE Access 12 (2024): 167072-167087.

[4] Tang, Lizhen, and Qusay H. Mahmoud. "A deep learning-based framework for phishing website detection." IEEe Access 10 (2021): 1509-1521.

[5] Çolhak, Furkan, et al. "Phishing website detection through multi-model analysis of html content." International Conference on Theoretical and Applied Computing. Singapore: Springer Nature Singapore, 2024.

[6] Widiono, Suyud, Achmad Nuruddin Safriandono, and Setyo Budi. "Phishing website detection using bidirectional gated recurrent unit model and feature selection." Journal of Future Artificial Intelligence and Technologies 1.2 (2024): 75-83.

[7] Almousa, May, et al. "Phishing website detection: How effective are deep learning-based models and hyperparameter optimization?." Security and Privacy 5.6 (2022): e256.

[8] Veach, Alexander M., and Munther Abualkibash. "Phishing website detection using several machine learning algorithms: a review paper." International Journal of Informatics, Information System and Computer Engineering (INJIISCOM) 3.2 (2022): 219-230.

[9] Alsariera, YAZAN A., et al. "Intelligent tree-based ensemble approaches for phishing website detection." J. Eng. Sci. Technol 17.1 (2022): 563-582.

[10] Jawade, Jayesh V., and Soma N. Ghosh. "Phishing website detection using Fast. ai Library." 2021

International conference on communication information and computing technology (ICCICT). IEEE, 2021.

[11] Liu, Dong-Jie, Guang-Gang Geng, and Xin-Chang Zhang. "Multi-scale semantic deep fusion models for phishing website detection." Expert Systems with Applications 209 (2022): 118305.

[12] Selvakumari, M., et al. "Phishing website detection using machine learning and deep learning techniques." Journal of Physics: Conference Series. Vol. 1916. No. 1. IOP Publishing, 2021.

[13] Ravindra, Salvi Siddhi, et al. "Phishing website detection based on URL." International Journal of Scientific Research in Computer Science, Engineering and Information Technology (USRCSEIT) 7.3 (2021): 589-594.

[14] Omari, Kamal. "Comparative study of machine learning algorithms for phishing website detection." International Journal of Advanced Computer Science and Applications 14.9 (2023).

[15] Jovanovic, Luka, et al. "Improving phishing website detection using a hybrid two-level framework for feature selection and xgboost tuning." Journal of Web Engineering 22.3 (2023): 543-574.

[16] Asiri, Sultan, et al. "A survey of intelligent detection designs of HTML URL phishing attacks." IEEe Access 11 (2023): 6421-6443.

[17] Salloum, Said, et al. "A systematic literature review on phishing email detection using natural language processing techniques." Ieee Access 10 (2022): 65703-65727.

[18] Balogun, Abdullateef O., et al. "Optimized decision forest for website phishing detection." Proceedings of the Computational Methods in Systems and Software. Cham: Springer International Publishing, 2021. 568-582.

[19] Omari, Kamal. "Comparative study of machine learning algorithms for phishing website detection." International Journal of Advanced Computer Science and Applications 14.9 (2023).

[20] Remya, S., et al. "BGL-PhishNet: Phishing Website Detection Using Hybrid Model-BERT, GNN, and LightGBM." IEEE Access 13 (2025): 47552-47569.

[21] Remya, S., et al. "An effective detection approach for phishing URL using ResMLP." IEEE access 12 (2024): 79367-79382.

[22] Wei, Yi, and Yuji Sekiya. "Sufficiency of ensemble machine learning methods for phishing websites detection." *IEEE Access* 10 (2022): 124103-124113.

[23] Bhavani, P. Amba, et al. "Phishing websites detection using machine learning." *Madhumitha and Likhitha, Pinnam Sree and Sai, Chanda Pranav Sai, Phishing Websites Detection Using Machine Learning (September 2, 2022)* (2022).

[24] Ubing, Alyssa Anne, et al. "Phishing website detection: An improved accuracy through feature selection and ensemble learning." *International Journal of Advanced Computer Science and Applications* 10.1 (2019).

[25] Chaiban, A.; Sovilj, D.; Soliman, H.; Salmon, G.; Lin, X. Investigating the Influence of Feature Sources for Malicious Website Detection. Appl. Sci. 2022, 12, 2806. https://doi.org/10.3390/app12062806