ISSN: P-2455-0515
E- 2394-8450

Educreator Research Journal

OPEN ACCESS

VOLUME–XII, SPECIAL ISSUES–II JAN – FEB 2025

Original Research Article

MULTIMODAL STACKING FRAMEWORK FOR EFFECTIVE PHISHING DETECTION INTEGRATING URL AND TEXT FEATURES

** C. V. Swetha, **Dr. P. Ramila Rajaleximi & *** Dr. Sibi Shaji*

** Research Scholar, School of Computational Sciences & IT, Garden City University, Bangalore and Assistant Professor, Presidency College (Autonomous), Bangalore, India*
*** Sanpada College of Commerce and Technology,Sector 2, Sanpada,Navi Mumbai 400705.*
****School of Computational Sciences & IT, Garden City University, Bangalore 560049, India*

Abstract:

This study proposes a stacking-based multimodal framework for phishing detection that integrates text content and URL data. The primary objective is to enhance phishing detection systems by leveraging the advantages of unstructured text content and structured URL features. The framework employs a fine-tuned BERT model to classify text content and a random forest (RF) classifier to process URL features. To improve detection accuracy, the predictions from both models are combined using a meta-classifier. Results demonstrate that the proposed model outperforms individual conventional classifiers such as multilayer perceptron, logistic regression, and RF, achieving an impressive true positive rate of 0.999 and a minimal false positive rate of 0.001. The effectiveness of the model in identifying phishing attempts while reducing errors is highlighted by consistently high precision, recall, and F-measure values, all exceeding 0.994. These findings support adopting the proposed approach in real-time phishing detection systems for platforms like websites, SMS, and emails. By combining URL and text-based features, this research offers a novel approach to combating evolving phishing threats.

Keywords: *Phishing Detection, Multimodal Framework, Stacking-Based Generalization, URL Features, Text Classification, Cybersecurity Systems.*

Introduction:

With the increasing use of digital communication channels such as social media, email, and SMS, cybercriminals exploit these platforms to deceive naive consumers into disclosing private information (Goenka et al., 2024). The complexity and advancement of phishing attacks, which involve the illicit use of links and messages to trick victims into revealing login credentials, financial information, or personal data, are continuously increasing (Ayeni et al., 2024). This evolution poses significant challenges for organisations and cybersecurity experts in detecting phishing attempts effectively. Researchers have also discovered that phishing emails exploit themes of authority, urgency, and fear, with almost all emails using urgency cues and fear appeals, highlighting the need for increased cybersecurity awareness to combat such threats (Mishra & Soni, 2022).

ISSN: P-2455-0515
E- 2394-8450

Educreator Research Journal

OPEN ACCESS

VOLUME–XII, SPECIAL ISSUES–II JAN – FEB 2025

Original Research Article

Phishing detection methods often rely on machine learning and deep learning algorithms to process vast amounts of data and identify patterns indicative of phishing activity (Maddireddy & Maddireddy, 2022). Conventional systems have primarily focused on one-dimensional features, such as text content or URL data, independently. However, as phishing attacks grow more sophisticated, integrating data from various sources is essential to improving detection accuracy (Alkhalil et al., 2021). While URL analysis can be effective, it may overlook contextual cues in text, and text analysis may fail to detect specific URL attributes. Many classification models continue to rely on single data modalities, such as text content analysis or URL features, limiting their effectiveness (Xu et al., 2024).

Several research gaps hinder the development of efficient and broadly applicable phishing detection models (Kritika, 2024). Most existing research focuses on distinct feature types (e.g., text or URL features) without exploring the potential for combining these disparate feature sets (Ozcan et al., 2023). Traditional methods, such as rule-based analysis and blacklisting, struggle to detect emerging phishing threats and adapt to the dynamic nature of attacks (Qabajeh et al., 2018). The challenge has necessitated the adoption of machine learning classifiers, which can better identify and adapt to the evolving nature of phishing scams (Cui et al., 2018).

To detect modern phishing attacks, machine learning models such as k-nearest neighbour (KNN), random forest (RF), support vector machine (SVM), naïve Bayes (NB), and decision trees (DT) have been employed (Adewole et al., 2019). While these models provide useful insights, they are often insufficient for addressing the multifaceted nature of phishing attacks. Text-based classification approaches for SMS phishing detection have been studied, but they often include phishing links, leading to high false-positive rates and lower detection accuracy (Sharaff et al., 2022). Stacking techniques, which combine predictions from multiple classifiers, could improve detection accuracy but remain underexplored (Al-Sarem et al., 2021). Additionally, many models rely on restricted datasets and fail to adapt to new phishing techniques, negatively impacting long-term performance (Jain & Gupta, 2022). This study aims to address the growing threat by identifying a robust model to combat phishing effectively.

The objective of this research is to develop an advanced multimodal framework for phishing detection using stacking that incorporates URL features and text content. By integrating the power of RF for URL classification and Bidirectional Encoder Representations from Transformers (BERT), a transformer-based model, for textual content analysis, this approach seeks to overcome the limitations of conventional models. The proposed framework improves detection accuracy and robustness by aggregating predictions from both classifiers through a meta-classifier, utilising stacking-based generalisation. Additionally, this research aims to compare the efficacy of the proposed model to that of traditional machine learning models. A stacking-based multimodal framework that combines structured URL features and unstructured text content could enhance phishing detection accuracy and generalisation for real-world applications.

Proposed Methodology: Figure 1 illustrates the proposed stacking-based multimodal framework for phishing detection. This model integrates URL data and text content to distinguish phishing attempts from legitimate entries and applies stacking or stacked generalisation.

ISSN: P-2455-0515
E- 2394-8450

Educreator Research Journal

OPEN ACCESS

VOLUME–XII, SPECIAL ISSUES–II

JAN – FEB 2025

Original Research Article

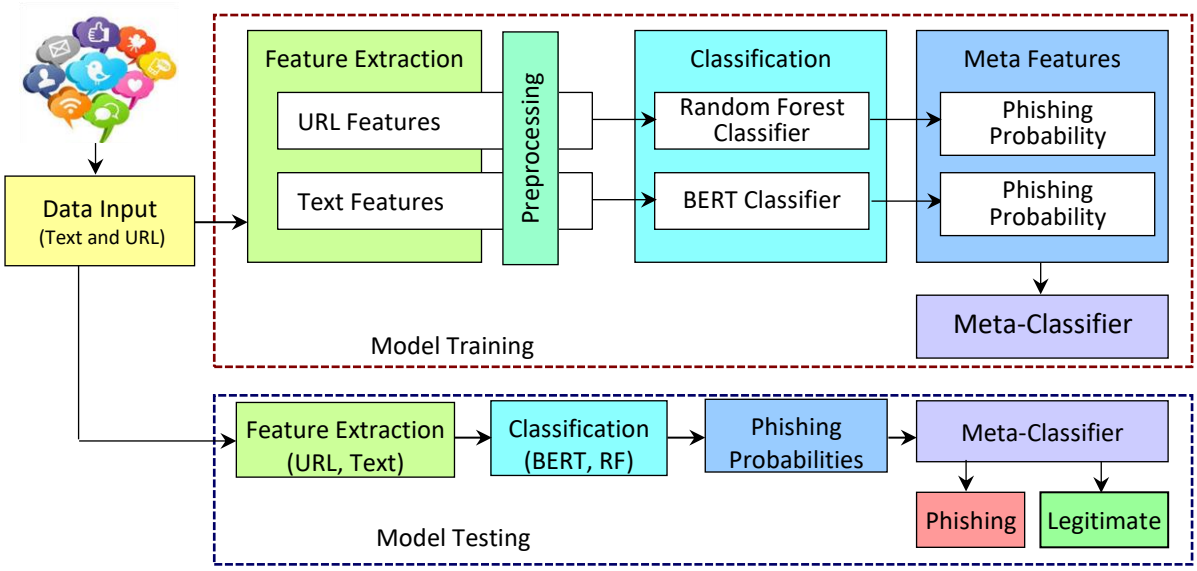

Figure 1. Stacking based Multimodal Phishing Detection Framework

Initially, features are extracted from URLs and processed using an RF classifier to predict phishing probabilities. Text data is then extracted, preprocessed, and analysed with a fine-tuned BERT model for phishing detection. Predicted probabilities from RF and BERT are combined into a stacked feature set, and a meta-classifier optimises predictions by training on the stacked features and true labels, enhancing detection accuracy.

Feature Extraction and Preprocessing:

The study utilises data from emails, social media posts, and messages to distinguish between phishing and benign activities, categorising URLs and text as phishing or non-phishing. Various lexical, structural, and domain-related features are extracted from the URLs. Lexical features include URL length, the number of dots, dashes, underscores, and character randomness. Domain-related features involve checking for a legitimate domain name as a subdomain, the use of an IP address instead of a domain name, HTTPS in the hostname, and inconsistencies. Structural features cover the number of subdomain levels, URL path depth, query parameter count, hostname length, and the percentage of hyperlinks redirecting to # or null targets. Extracted features undergo preprocessing and min-max normalisation to ensure consistency and equal contribution, enhancing model performance and stability. Text content accompanying URLs is preprocessed through stopword removal, tokenisation, and stemming. These steps reduce redundancy, simplify word structures, and improve computational efficiency, enabling the model to detect suspicious language patterns and inconsistencies critical for identifying phishing activities.

Stacking Generalization:

Stacking is an ensemble learning technique that combines multiple classification models using a meta-classifier to generate final predictions. Individual models are trained on the full training set, and the meta-classifier is trained on their outputs. In the proposed framework, RF and BERT are used as base classifiers for structured and

ISSN: P-2455-0515
E- 2394-8450

Educreator Research Journal

OPEN ACCESS

VOLUME–XII, SPECIAL ISSUES–II

JAN – FEB 2025

Original Research Article

textual data, respectively, with logistic regression (LR) as the meta-classifier to integrate predictions and model relationships among the base classifiers.

Base Classifiers:

Random Forest (RF), an ensemble learning method, is used for phishing detection with URL features (Yang et al., 2021). It combines predictions from multiple decision trees, reducing overfitting and improving generalisation. RF is robust in handling high-dimensional, noisy data, outliers, and complex features, making it suitable for large-scale phishing detection tasks. Hyperparameter tuning, such as tree number and depth, optimises its performance, with feature importance rankings supporting interpretability. For text analysis, BERT, a transformer-based model for natural language understanding, uses bidirectional attention to capture both semantic and syntactic nuances. Pre-trained on large corpora like Wikipedia, BERT is fine-tuned with labelled text data to effectively detect phishing attempts and subtle linguistic patterns.

Meta Classifier:

A stacking-based framework for phishing detection uses LR as a meta-classifier, combining predictions from RF and BERT for final classification decisions. LR is a statistical model ideal for binary classification tasks, utilizing a sigmoid function to predict outcomes based on input features, producing probabilities between 0 and 1. LR in stacking is a user-friendly tool that effectively integrates predictions from various models, providing clear insights into feature significance through its coefficients. Additionally, LR is a reliable choice for stacking models due to its lower overfitting risk and improved generalization, as it uses prediction probabilities from base classifiers for accurate final classification decisions.

Result Analysis:

The multimodal phishing detection model, using Random Forest and BERT classifiers as base learners and Logistic Regression as a meta-classifier, is implemented in Python 3 within the Anaconda environment. Development is done in Jupyter Notebook on a 64-bit OS with 256 GB of RAM and an Intel Pentium CPU. Tools used include Pandas and NumPy for data manipulation, scikit-learn for machine learning, transformers and Torch for BERT fine-tuning, and TensorFlow for deep learning.

Since no dataset exists for phishing detection that combines message content with URL features, this study integrates two distinct datasets. The Kaggle Phishing Dataset for Machine Learning was used to assess URL feature-based phishing detection, utilizing 50 features and 10,000 samples, including 5,000 phishing and 5,000 non-phishing instances (https://www.kaggle.com/datasets/shashwatwork/phishing-dataset-for-machine-learning). The SMS phishing dataset from Mendeley, containing 5,971 text messages, was used for analysis, including 1,127 spam and smishing messages and 4,844 non-spam messages (Mishra & Soni, 2022) (https://data.mendeley.com/datasets/f45bkkt8pr/1). The messages are repeated to align with the first dataset for consistency.

The performance of the proposed model is evaluated by comparing it with various standard classifiers, including Naïve Bayes (NB), Logistic Regression (LR), Multilayer Perceptron (MLP), k-nearest Neighbour (KNN), Rule-Based Learner (RBL), One Rule Classifier (ORC), Partial Decision Tree (PART), Random Forest (RF), and

ISSN: P-2455-0515
E- 2394-8450

Educreator Research Journal

OPEN ACCESS

VOLUME–XII, SPECIAL ISSUES–II

JAN – FEB 2025

Original Research Article

Random Tree (RT). The evaluation uses different performance metrics, including Correctly Classified Instances (CCI), Incorrectly Classified Instances (ICI), Precision, Recall, F-Measure, Area Under the Curve (AUC), Precision-Recall Curve (PRC), True Positive Rate (TPR), False Positive Rate (FPR), Mean Absolute Error (MAE), and Root Mean Squared Error (RMSE).

Table 1 presents the average performance of the proposed model, which incorporates both URL features and text features, alongside other classifiers that utilise only URL features. The results indicate that NB and KNN demonstrate the lowest performance among the classifiers. In contrast, rule-based and tree-based classifiers such as RBL, ORC, PART, RT, and MLP show significantly improved performance, achieving accuracy levels exceeding 98%. Classifiers like RF and LR further enhance performance, delivering approximately 99% classification accuracy. Notably, the proposed model outperforms the other models under comparison. Furthermore, the model effectively integrates URL and text features, achieving superior accuracy and minimal errors, demonstrating its robustness for phishing detection compared to traditional classifiers.

Table 1. Performance Comparison with Other Classifiers

Classifier	CCI (in %)	ICI (in %)	Precision	Recall	F-Measure	AUC	PRC	MAE	RMSE
NB	95.12	4.88	0.952	0.951	0.951	0.982	0.975	0.0509	0.2001
KNN	97.71	2.29	0.977	0.977	0.977	0.977	0.966	0.0230	0.1513
RBL	98.96	1.04	0.990	0.990	0.990	0.990	0.985	0.0104	0.102
ORC	98.25	1.75	0.981	0.984	0.982	0.985	0.987	0.0016	0.0141
PART	98.65	1.35	0.985	0.983	0.984	0.985	0.987	0.0009	0.0190
RT	98.84	1.26	0.988	0.988	0.988	0.989	0.989	0.0017	0.0405
MLP	98.81	1.19	0.988	0.988	0.989	0.988	0.989	0.0026	0.0392
RF	99.05	0.95	0.990	0.991	0.991	0.994	0.991	0.0125	0.0358
LR	98.97	1.03	0.987	0.986	0.987	0.986	0.985	0.0003	0.0173
Proposed	99.97	0.03	0.997	0.996	0.996	0.998	0.999	0.0003	0.0173

Table 2 presents the performance analysis, highlighting the class-wise evaluation. It demonstrates superior phishing detection accuracy and reliability of the proposed model. Outperforming all other classifiers, the proposed approach achieves an impressive TPR of 0.999 and an FPR of just 0.001 for both phishing (P) and legitimate (L) classes. With consistent precision, recall, and F-measure values above 0.994, it exhibits remarkable resilience in accurately detecting phishing and legitimate cases with minimal errors. Classifiers like RF and MLP achieve high accuracy but are marginally inferior to the proposed model, except for NB and KNN, which show improved phishing detection performance. The proposed model, with AUC and PRC nearing 1.0, effectively handles unbalanced data and reduces misclassifications, setting a new benchmark in phishing detection.

ISSN: P-2455-0515
E- 2394-8450

Educreator Research Journal

OPEN ACCESS

VOLUME–XII, SPECIAL ISSUES–II

JAN – FEB 2025

Original Research Article

Table 2. Class-wise Performance Comparison with Other Classifiers

Classifiers	Class	TPR	FPR	Precision	Recall	F-Measure	AUC	PRC
NB	P	0.931	0.029	0.97	0.931	0.95	0.982	0.979
	L	0.971	0.069	0.934	0.971	0.952	0.983	0.972
KNN	P	0.975	0.02	0.979	0.975	0.977	0.977	0.967
	L	0.98	0.025	0.975	0.977	0.977	0.977	0.967
RBL	P	0.989	0.003	0.989	0.988	0.988	0.987	0.989
	L	0.979	0.003	0.987	0.986	0.986	0.985	0.982
ORC	P	0.988	0.012	0.985	0.989	0.987	0.989	0.989
	L	0.977	0.023	0.976	0.980	0.978	0.982	0.985
PART	P	0.989	0.011	0.989	0.987	0.986	0.989	0.989
	L	0.984	0.016	0.982	0.980	0.981	0.982	0.985
RT	P	0.989	0.002	0.990	0.988	0.989	0.988	0.989
	L	0.988	0.002	0.987	0.988	0.989	0.989	0.989
MLP	P	0.989	0.012	0.989	0.989	0.989	0.989	0.989
	L	0.987	0.013	0.987	0.987	0.987	0.989	0.989
RF	P	0.991	0.009	0.990	0.991	0.989	0.998	0.992
	L	0.995	0.010	0.991	0.990	0.991	0.989	0.991
LR	P	0.990	0.011	0.989	0.990	0.990	0.990	0.984
	L	0.989	0.010	0.990	0.989	0.990	0.990	0.985
Proposed	P	0.999	0.001	0.998	0.996	0.997	0.998	0.999
	L	0.999	0.001	0.997	0.995	0.994	0.997	0.998

Although the results are encouraging, limitations highlight the need for further research. The model's generalisability requires evaluation with other advanced text classification models beyond BERT. Real-time datasets were not used, limiting real-world applicability. Broader comparisons with state-of-the-art approaches are needed. Future datasets should integrate raw message content and URL features.

Conclusion:

The study developed a stacking-based multimodal framework for phishing detection, effectively combining text content and URL data to distinguish between legitimate entries and phishing attempts. The framework enhances phishing detection accuracy by combining RF classifier for URL features and refined BERT model for text content, with the meta-classifier aggregating predictions for improved performance. The model outperforms conventional classifiers with exceptional accuracy and resilience, achieving a TPR of 0.999 and low FPR of 0.001, and consistently surpassing precision, recall, and F-measure values. The proposed framework effectively detects phishing in real-time across various platforms, but further improvements could involve comparing it with other text classifiers and real-world scenarios. The robustness of the model in addressing phishing threats could be enhanced by developing a new dataset combining raw message content and URL features. Comparing the model with advanced learning algorithms will offer valuable insights into its strengths and limitations, enabling its further development and refinement in real-world applications.

ISSN: P-2455-0515
E- 2394-8450

Educreator Research Journal

OPEN ACCESS

VOLUME–XII, SPECIAL ISSUES–II JAN – FEB 2025

Original Research Article

References:

1. *Adewole, K. S., Akintola, A. G., Salihu, S. A., Faruk, N., & Jimoh, R. G. (2019). Hybrid rule-based model for phishing URL detection. In Emerging Technologies in Computing: Second International Conference, iCETiC 2019, London, UK, August 19–20, 2019, Proceedings 2 (pp. 119-135). Springer International Publishing.*
2. *Alkhalil, Z., Hewage, C., Nawaf, L., & Khan, I. (2021). Phishing attacks: A recent comprehensive study and a new anatomy. Frontiers in Computer Science, 3, 1-23.*
3. *Al-Sarem, M., Saeed, F., Al-Mekhlafi, Z. G., Mohammed, B. A., Al-Hadhrami, T., Alshammari, M. T., Alreshidi, A., & Alshammari, T. S. (2021). An optimized stacking ensemble model for phishing website detection. Electronics, 10(11), 1-18.*
4. *Ayeni, R. K., Adebiyi, A. A., Okesola, J. O., & Igbekele, E. (2024, April). Phishing attacks and detection techniques: A systematic review. In 2024 International Conference on Science, Engineering and Business for Driving Sustainable Development Goals (SEB4SDG) (pp. 1-17). IEEE.*
5. *Cui, B., He, S., Yao, X., & Shi, P. (2018). Malicious URL detection with feature extraction based on machine learning. International Journal of High Performance Computing and Networking, 12(2), 166-178.*
6. *Goenka, R., Chawla, M., & Tiwari, N. (2024). A comprehensive survey of phishing: Mediums, intended targets, attack and defence techniques and a novel taxonomy. International Journal of Information Security, 23(2), 819-848.*
7. *Jain, A. K., & Gupta, B. B. (2022). A survey of phishing attack techniques, defence mechanisms and open research challenges. Enterprise Information Systems, 16(4), 527-565.*
8. *Kritika, E. (2024). A comprehensive literature review on phishing URL detection using deep learning techniques. Journal of Cyber Security Technology, 1-29.*
9. *Maddireddy, B. R., & Maddireddy, B. R. (2022). AI-Based Phishing Detection Techniques: A Comparative Analysis of Model Performance. Unique Endeavor in Business & Social Sciences, 1(2), 63-77.*
10. *Mishra, S., & Soni, D. (2022, December). SMS phishing dataset for machine learning and pattern recognition. In International Conference on Soft Computing and Pattern Recognition (pp. 597-604). Cham: Springer Nature Switzerland.*
11. *Ozcan, A., Catal, C., Donmez, E., & Senturk, B. (2023). A hybrid DNN–LSTM model for detecting phishing URLs. Neural Computing and Applications, 1-17.*
12. *Qabajeh, I., Thabtah, F., & Chiclana, F. (2018). A recent review of conventional vs. automated cybersecurity anti-phishing techniques. Computer Science Review, 29, 44-55.*
13. *Sharaff, A., Allenki, R., & Seth, R. (2022). Deep learning based sentiment analysis for phishing sms detection. In Research anthology on implementing sentiment analysis across multiple disciplines (pp. 864-891). IGI Global.*

ISSN: P-2455-0515
E- 2394-8450

Educreator Research Journal

OPEN ACCESS

VOLUME–XII, SPECIAL ISSUES–II

JAN – FEB 2025

Original Research Article

14. *Xu, X., Li, J., Zhu, Z., Zhao, L., Wang, H., Song, C., Chen, Y., Zhao, Q., Yang, J., & Pei, Y. (2024). A Comprehensive Review on Synergy of Multi-Modal Data and AI Technologies in Medical Diagnosis. Bioengineering, 11(3), 1-51.*
15. *Yang, R., Zheng, K., Wu, B., Wu, C., & Wang, X. (2021). Phishing website detection based on deep convolutional neural network and random forest ensemble learning. Sensors, 21(24), 1-18.*

Cite This Article:

***Swetha C. V., Dr. Rajaleximi P. R.* & Dr. *Shaji S. (*2025).** *Multimodal Stacking Framework for Effective Phishing Detection Integrating URL and Text Features.* **In Educreator Research Journal: Vol. XII (Issue I),** pp. 12–19.