UNIVERSITI TEKNOLOGI MARA

MATHEMATICS LETTERS
https://sites.google.com/tmsk.uitm.edu.my/mathematicsletters

Innovative Machine Learning Model for Malicious URL Detection

[1,2,*]Boon-Chian Tea & [3]Nur Irdina Hassan
[1,*]Center of Mathematical Sciences, Faculty of Computer and Mathematical Sciences, Universiti Teknologi MARA (UiTM), 40450 Shah Alam, Selangor, Malaysia.
[2]Asymmetric Division, Malaysia Cryptology Technology and Management Centre, c/o Universiti Putra Malaysia (UPM), 43400 Serdang, Selangor, Malaysia.
[3]BP Business Service Centre Asia Sdn Bhd, Level 9, Tower 5, Avenue 7, The Horizon, Bangsar South City, No. 8, Jalan Kerinchi, 59200 Kuala Lumpur, Malaysia
[1,2,*]teaboonchian@uitm.edu.my
[3]irdinalin@gmail.com
*Corresponding author

Received: 20/3/2026 Revised: 31/3/2026 Accepted: 30/4/2026 Published: 30/4/2026

ABSTRACT

Malicious URLs impose significant risk when access, including scams, attacks, and fraud either intentionally or unintentionally. This paper utilizes machine learning techniques to safeguard these threats. The research reviews current research surrounding machine learing and employs an ensemble approach to improve detection accuracy. Top models such as Decision Tree, Random Forest, AdaBoost, and XGBClassifier will be combined, in addition Voting Classifier with hard voting will also be used. The machine learning model's performance and its effectiveness are evaluated in the aspects of accuracy, precision, recall, and F1-score.

Keywords: Malicious URL Detection, Feature Extraction, Machine Learning, Ensemble Model, Voting Classifier.

INTRODUCTION

The advancement of information technology and internet have caused the number of malicious Uniform Resource Locators (URLs) to increase. These web addresses created impose significant risk including scams and fraud when accessed, either intentionally or unintentionally. According to a report by Cybersecurity Malaysia in 2023 (Jan to June), about 85.23% of cyber incidents in Malaysia involved malware, while 14.77% were related to phishing, including phishing URLs, IP addresses, and domains. Traditional methods such as updating lists of known malicious URLs manually, are insufficient to keep up with the constant evolution of new cyber threats. In order to address these challenges, researchers consider machine learning, which uses data-driven approach to identify patterns and traits that suggest malicious intent, enabling a more effective detection of new and evolving malicious URLs.

Ensemble learning has been widely applied to detect phishing and malicious URLs due to its strengths in combining several classifiers in improving accuracy and robustness. For instance, Al-Haija and Al-Fayoumi (2023) demonstrated such ability by achieving 99.3% accuracy using an ensemble of bagging trees, while Kalabarige et al. (2022) introduced a multilayered stacked ensemble (MLSELM) that achieved accuracy rates between 96.79% and 98.90% across multiple datasets. In addition, Ghaleb et al. (2022) demonstrated a two-stage ensemble model using Random Forest (RF) for preliminary classification and Multilayer Perceptron (MLP) for final decision-making, showing a 7.8% accuracy improvement when using cyber threat intelligence features. Ghaleb's model however posed noise in high dimension cases.

Ghaleb et al. (2022) developed a phishing detection model using genetic-based ensemble classifiers. Base classifiers such as AdaBoost, Bagging, GradientBoost, Random Forest, XGBoost, and LightGBM were trained to identify malicious URL features. These classifiers are optimized using a Genetic Algorithm for better accuracy and were combined with stacking for final predictions. On the other hand, Sameen et al. (2020) introduced a phishing detection system called PhishHaven, that leverages lexical analysis with URL HTML encoding. This model integrated URL Hit approach to address tiny URL detection and utilized an unbiased voting mechanism to reduce potential false classification. Mondal et al. (2021) proposed an ensemble learning-based approach that integrates multiple classifiers with a novel thresholding feature to further refine predictions by selecting the highest-probability label for final classification.

Rashid et al. (2020) proposed a method that integrates automated web page collection, preprocessing, feature extraction, and dimensionality reduction via PCA with SVM, achieved 95.66% accuracy, proving to be the most effective in distinguishing phishing from legitimate sites. Korkmaz et al. (2020) designed a phishing detection using Random Forest model and showed that their work has the highest accuracy performance of 94.59%. Recent studies have suggested some promising results in detecting malicious and phishing URLs using deep learning model. For instance, Wang et al. (2021) developed a CNN-based model with vector embedding, dynamic convolution, and block extraction modules, achieving 98% accuracy by combining automated and manual feature extraction. Afzal et al. (2021) proposed URLdeepDetect, a hybrid framework using LSTM and k -means clustering for semantic and lexical analysis, reaching 98.3% and 99.7% accuracy, respectively.

Other hybrid and multi-model approaches are also worth its mention. For examples, Somesha et al. (2020) integrated DNN, LSTM, and CNN with feature extraction from URL obfuscation, hyperlinks, and third-party information, achieving accuracy above 99.4% across models, with LSTM peaking at 99.57%. Khukalenko et al. (2023) focused on lexical and network features, comparing multiple classifiers, and found XGBoost and Random Forest to be the most effective. In short, these studies underscore the effectiveness of CNNs, LSTMs, and ensemble methods in detecting phishing URLs, with constant and high accuracies of about 98%. These indicate both the potential and computational challenges of deep learning in cybersecurity.

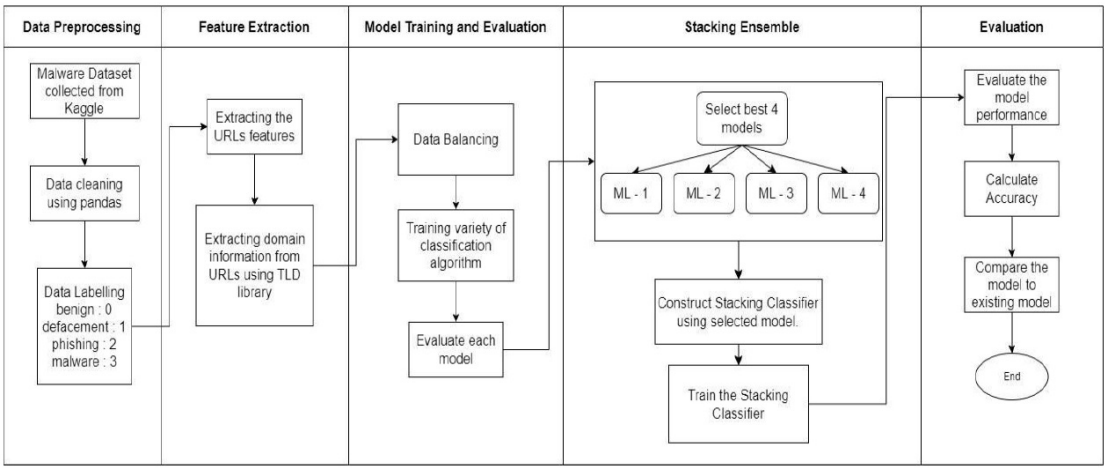
In this paper, we construct a high-performance machine learning model and perform malicious URL classification using ensemble models for better detection and further categorizing them into different types. We then evaluate the model's performance when implementing a balanced and imbalanced dataset. The remainder of this paper is organized as follows: Section 2 delves into the research methodology. Section 3 outlines the results and discussion. Finally, we summarize the key findings, limitations, and potential future research directions in Section 4.

METHODOLOGY

The following Figure 1 depicts our proposed model, which is geared towards optimizing the classification process by integrating the most effective classifiers through a voting mechanism.

Figure 1

Flowchart of Proposed Model.



In the model training and evaluation phase (third phase in Figure 1), the accuracy is computed via

$$\text{Accuracy} = \frac{\text{True Positive} + \text{True Negative}}{\text{True Positive} + \text{True Negative} + \text{False Positive} + \text{False Negative}} \tag{1}$$

On the other hand, the evaluation phase (fifth phase in Figure 1), the precision, recall, and F1-score are calculated based on the following formulation:

$$\text{Precision} = \frac{\text{True Positive}}{\text{True Positive} + \text{False Positive}} \tag{2}$$

$$\text{Recall} = \frac{\text{True Positive}}{\text{True Positive} + \text{False Negative}} \tag{3}$$

$$\text{F1-Score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \tag{4}$$

RESULTS AND DISCUSSION

In this section, we present our findings obtained from machine learning for malicious URL detection via our proposed model. The simulation is carried out using Python 3.11.17 (*myenv*) and libraries such as TensorFlow, Keras, NumPy, and Matplotlib. Visual Studio Code serves as the primary development environment due to its versatility and ease of use.

The following Table 1 to Table 4 layout the results for imbalanced dataset, balanced dataset with balanced random forest, balanced dataset under sampling and balanced dataset over sampling, respectively. It should be noted that each class represents categories where Class 0 is for benign, Class 1 for defacement, Class 2 for phishing and Class 3 for malware respectively.

Table 1

Model Result for Imbalanced Dataset.

Measurement	**Imbalanced Dataset**			
	Class 0	**Class 1**	**Class 2**	**Class 3**
Precision	0.92	0.95	0.85	0.97
Recall	0.98	0.98	0.58	0.88
F1 Score	0.95	0.96	0.69	0.92
Accuracy	91.85%			
Cross-Validation Accuracy	91.69%			

Table 2

Model Result for Balanced Dataset with Balanced Random Forest.

Measurement	**Balanced Dataset with Balanced Random Forest**			
	Class 0	**Class 1**	**Class 2**	**Class 3**
Precision	0.92	0.95	0.85	0.97
Recall	0.98	0.98	0.58	0.88
F1 Score	0.95	0.96	0.69	0.93
Accuracy	91.88%			
Cross-Validation Accuracy	91.69%			

Table 3

Model Result for Balanced Dataset with Random Under Sampling.

Measurement	**Balanced Dataset with Random Under Sampling**			
	Class 0	**Class 1**	**Class 2**	**Class 3**
Precision	0.95	0.92	0.50	0.83
Recall	0.83	0.97	0.74	0.91
F1 Score	0.89	0.95	0.60	0.87
Accuracy	84.42%			
Cross-Validation Accuracy	86.25%			

Table 4

Model Result for Balanced Dataset with Random Over Sampling.

Measurement	Balanced Dataset with Random Over Sampling			
	Class 0	**Class 1**	**Class 2**	**Class 3**
Precision	0.92	0.95	0.85	0.97
Recall	0.98	0.98	0.58	0.88
F1 Score	0.95	0.96	0.69	0.93
Accuracy	87.91%			
Cross-Validation Accuracy	92.88%			

Based on the above results, the proposed model achieved the highest accuracy result of 91.88% by applying Balanced Random Forest. The proposed model also achieved the highest cross-validation mean accuracy of 92.88% by applying Random Over Sampling.

CONCLUSION

In conclusion, we constructed a high-performance machine learning model and used it to detect malicious URLs. We began by studying the existing work on malicious URL detection using machine learning models, which provided valuable insights into the strengths and weaknesses of prior approaches and guided the design of our model. We then constructed our own machine learning model and applied ensemble learning through a voting classifier to enhance detection accuracy. This approach allowed us to perform malicious URL classification effectively and able to categorize URLs into several types, that are benign, defacement, phishing, and malware. Ultimately, we evaluated the performance of the model on both balanced and imbalanced datasets. The results demonstrated that while imbalanced data provided strong baseline accuracy, the use of balancing techniques such as Balanced Random Forest and Random Over Sampling improved overall performance and ensured fairer classification across different classes. These results confirm that ensemble learning, combined with appropriate handling of class imbalance, can produce a reliable and effective model for malicious URL detection.

ACKNOWLEDGEMENT

The authors would like to further express appreciation to the Universiti Teknologi MARA (UiTM), Malaysia Cryptology Technology and Management Centre (PTPKM), and BP Business Service Centre Asia Sdn Bhd respectively for opportunity and support in conducting this project.

Conflict of Interest
The authors have no conflicts of interest to declare.

Non-funded
This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

REFERENCES

Afzal, S., Asim, M., Javed, A. R., Beg, M. O., & Baker, T. (2021). "URLdeepDetect: A Deep Learning Approach for Detecting Malicious URLs Using Semantic Vector Models." Journal of Network and Systems Management, 29(3). https://doi.org/10.1007/s10922-021-09587-8

Alkhudair, F., Alassaf, M., Ullah Khan, R., & Alfarraj, S. (2020). "Detecting Malicious URL." 2020 International Conference on Computing and Information Technology (ICCIT-1441). https://doi.org/10.1109/iccit-144147971.2020.9213792

Dsouza, R. (n.d.). "Malicious URL(s) Classification." MSc Research Project Cloud Computing. https://norma.ncirl.ie/4138/1/rohandsouza.pdf

Ghaleb Al-Mekhlafi, Z., Abdulkarem Mohammed, B., Al-Sarem, M., Saeed, F., Al-Hadhrami, T., T. Alshammari, M., Alreshidi, A., & Sarheed Alshammari, T. (2022). "Phishing Websites Detection by Using Optimized Stacking Ensemble Model." Computer Systems Science and Engineering, 41(1), 109–125. https://doi.org/10.32604/csse.2022.020414

Ghaleb, F. A., Alsaedi, M., Saeed, F., Ahmad, J., & Alasli, M. (2022). "Cyber Threat Intelligence-Based Malicious URL Detection Model Using Ensemble Learning." Sensors, 22(9), 3373. https://doi.org/10.3390/s22093373

Indrasiri, P. L., Halgamuge, M. N., & Mohammad, A. (2021). "Robust Ensemble Machine Learning Model for Filtering Phishing URLs: Expandable Random Gradient Stacked Voting Classifier (ERG-SVC)." IEEE Access, 9, 150142–150161. https://doi.org/10.1109/access.2021.3124628

Jalil, S., Usman, M., & Fong, A. (2022). Highly Accurate Phishing URL Detection Based on Machine Learning. Journal of Ambient Intelligence and Humanized Computing. https://doi.org/10.1007/s12652-022-04426-3

Kalabarige, L. R., Rao, R. S., Abraham, A., & Gabralla, L. A. (2022). "Multilayer Stacked Ensemble Learning Model to Detect Phishing Websites." IEEE Access, 10, 79543–79552. https://doi.org/10.1109/access.2022.3194672

Khukalenko, Y., Stopochkina, I., & Ilin, M. (2023). "Machine Learning Models Stacking in the Malicious Links Detecting." Theoretical and Applied Cybersecurity, 5(1). https://doi.org/10.20535/tacs.2664-29132023.1.287752

Korkmaz, M., Sahingoz, O. K., & Diri, B. (2020). "Detection of Phishing Websites by Using Machine Learning-Based URL Analysis." 2020 11th International Conference on Computing, Communication and Networking Technologies (ICCCNT). https://doi.org/10.1109/icccnt49239.2020.9225561

Krishna, D. S., Tutika Lavanya, Kola, R., Samhita Pedaprolu, Basha, S., & Rao, R. S. (2023). "Feature Extraction Based Ensemble Stacking For Combating Cyber Threat in Phishing URLs." https://doi.org/10.1109/icaect57570.2023.10118072

Mondal, D. K., Singh, B. C., Hu, H., Biswas, S., Alom, Z., & Azim, M. A. (2021). "SeizeMaliciousURL: A novel learning approach to detect malicious URLs." Journal of Information Security and Applications, 62, 102967. https://doi.org/10.1016/j.jisa.2021.102967

Qasem Abu Al-Haija, & Mustafa Al-Fayoumi. (2023). "An Intelligent Identification and Classification System for Malicious Uniform Resource Locators (URLs)." https://doi.org/10.1007/s00521-023-08592-z

Sameen, M., Han, K., & Hwang, S. O. (2020). "PhishHaven – An Efficient Real-Time AI Phishing URLs Detection System." IEEE Access, 8, 83425–83443. https://doi.org/10.1109/access.2020.2991403

Sevastianov, L., & Shchetinin, E. (2020). "On Methods for Improving the Accuracy of Multi-Class Classification on Imbalanced Data." https://ceur-ws.org/Vol-2639/paper-06.pdf

Siddhartha, M. (2021). "Malicious URLs Dataset." www.kaggle.com. https://www.kaggle.com/datasets/sid321axn/malicious-urls-dataset

Singh, S., Singh, M. P., & Pandey, R. (2020, October 1). "Phishing Detection from URLs Using Deep Learning Approach." IEEE Xplore. https://doi.org/10.1109/ICCCS49678.2020.9277459

Somesha, M., Pais, A. R., Rao, R. S., & Rathour, V. S. (2020). "Efficient Deep Learning Techniques for The Detection of Phishing Websites." Sādhanā, 45(1). https://doi.org/10.1007/s12046-020-01392-4

Subasi, A., & Kremic, E. (2020). "Comparison of Adaboost with MultiBoosting for Phishing Website Detection." Procedia Computer Science, 168, 272–278. https://doi.org/10.1016/j.procs.2020.02.251

Rashid, J., Mahmood, T., Nisar, M. W., & Nazir, T. (2020). "Phishing Detection Using Machine Learning Technique." IEEE Xplore. https://doi.org/10.1109/SMART-TECH49988.2020.00026

Rawat, S. S., & Kumar Mishra, A. (2023). "The Best ML Classifier(s): An Empirical Study on The Learning of Imbalanced and Resampled Credit Card Data." 2023 Second International Conference on Informatics (ICI), 1–6. https://doi.org/10.1109/ici60088.2023.10421691

Wang, Z., Ren, X., Li, S., Wang, B., Zhang, J., & Yang, T. (2021). "A Malicious URL Detection Model Based on Convolutional Neural Network." Security and Communication Networks, 2021, 1–12. https://doi.org/10.1155/2021/5518528

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), 8281. https://doi.org/10.3390/s21248281