JCBI
Journal of Computing & Biomedical Informatics
ISSN: 2710 - 1606
Volume 10 Issue 02
2026

Research Article
https://doi.org/10.56979/1002/2026/1341

A Hybrid HHO-BA Feature Selection Framework for High-Accuracy Malicious URL Detection Using LightGBM

Motasem M. Elshourbagy[1*], Ahmed G. Mabrouk[2], Salahudin Ayubi[3], Abdullah T. Elgammal[4], and Mohamed Ghetas[5]

[1]Physics and Engineering Mathematics Department, Mattaria Faculty of Engineering, Capital University, Cairo, Egypt.
[2]Department of Computer Science, High Institute of Computer Science and Information Systems, October City, Cairo, Egypt.
[3]Faculty of Information Technology and Computer Science, University of Central Punjab, Lahore, Pakistan.
[4]Electrical Eng. Dept., Benha Faculty of Engineering, Benha University, Benha, Egypt.
[5]Faculty of Computer Science and Engineering, Galala University, Suez, Egypt.
[*]Corresponding Author: Motasem M. Elshourbagy. Email: MOATASEM_ALSHURBAGY@m-eng.helwan.edu.eg

Received: December 04, 2025 **Accepted:** February 20, 2026

Abstract: Malicious URLs are frequently used as delivery channels for malware and continue to represent a challenge in cybersecurity. The following paper propose URL detection framework based on using a combination of the Harris Hawks Optimizer (HHO) and Bat Algorithm (BA) using a union feature selection strategy. The goal is to build an informative and diverse subset of features by using the features that were chosen using two complementary metaheuristic search methods. LightGBM and Naive Bayes classifiers are used to evaluate the selected features on ISCX-URL2016 dataset. As it has been experimentally found, LightGBM has a higher accuracy of 99.52 % and Naive Bayes has an accuracy of 81.12 % indicating a distinct difference in the capability of modelling structured URL features. The proposed framework is competitive or better accurate when compared to some of the past studies. The results demonstrate that the HHOUBA union approach is successful in minimizing feature redundancy and discriminative information is maintained, which results in higher learning performance and a steady classification performance. The suggested solution is a solid solution to malicious URL detection.

Keywords: Malicious URL Detection; Feature Selection; Harris Hawks Optimizer; Bat Algorithm

1. Introduction

The Internet has become one of the basic elements in our society allowing international connectivity and provision of a variety of the necessary services, including communication, business, cloud computing and data transfers [1-5]. The internet services constitute the fundamental backbone of critical infrastructures in the health sector, finance sector, education sector and also in the government sector and hence its reliability and its availability are very critical in the growth of the economy as well as social stability. With the continued reliance on these services, the security of these services has become a significant concern [6-10].

Nevertheless, the proliferation of the Internet services has also increased the area of attack in case of cyber threats [11-15]. The internet-based systems are often hit by several cyber-attacks such as spam attack, phishing campaigns, and malware attacks [16-19]. Such attacks can be meant to cause disruption of service availability, theft of sensitive information and destruction of system integrity usually with a massive loss of money and disruption of business. Malware is one of the most dangerous and constant attack vectors of these threats [20- 22].

Malware can rapidly spread across networks, malicious email, breached websites and removable media, which allows it to infect masses on a large scale in a short period of time. According to recent

statistics, volume and sophistication of malware attacks has been on a steady increase since their inception and the damages are worth billions of dollars each and every year around the globe. The conventional malware control systems, including signature-based antivirus software and rule-based intrusion detection systems, are not able to keep up with these emerging threats. These methods are not successful with obfuscated malware, polymorphic variants and zero-day attacks, and more dynamic defense mechanisms are required [23-26].

Machine learning has become a promising substitute to malware detection by making it possible to analyze data and identify patterns automatically. Complex and previously unknown attack behaviors can be detected using learning-based models, which are better robust than a static approach [27-30]. Nevertheless, machine learning models are highly dependent on quality of input features that can be used to improve accuracy, efficiency and generalization that make feature selection a critical step in the process of enhancing machine learning models [31-33].

In this context, bio-inspired optimization algorithms offer complementary search capabilities for identifying informative feature subsets. Examples of such algorithms include the Harris Hawks Optimizer, Bat Algorithm, Narwhal Optimizer, Grey Wolf Optimizer, white shark optimizer [34-38]. To overcome these issues, this paper suggests a hybrid model combining the HHO and BA on the basis of a union-based feature selection approach (HHOUBA) with customized Naive Bayes and LightGBM classifiers to detect malware via URLs. Table 1 gives a comparison of the HHO and BA algorithm with regard to feature selection [37-40]. The analysis of the suggested framework is conducted on the ISCX-URL2016 dataset.

Table 1. Comparison of the HHO and BA algorithms

Aspect	HHO	BA
Biological Motivation	Inspired by the cooperative hunting strategy of Harris hawks, particularly surprise pounce and encircling behaviors.	Based on the echolocation and sonar-based navigation behavior of bats during prey search.
Search Strategy	Models dynamic transitions between exploration and exploitation using prey escape energy and besiege strategies.	Explores the search space through frequency-controlled velocity updates and stochastic local walks.
Exploration–Exploitation Balance	Achieves adaptive balance by switching strategies according to prey energy decay.	Regulates balance by decreasing loudness and increasing pulse emission rate over iterations.
Effectiveness in Feature Selection	Effective in identifying highly discriminative feature subsets with strong global search capability.	Capable of refining feature subsets locally, capturing weak but informative features.
Computational Demand	Moderate computational cost due to multiple adaptive hunting strategies.	Relatively low computational overhead with simple update equations.
Applicability to Attack Detection	Well suited for complex, nonlinear detection tasks in high-dimensional feature spaces.	Effective for attack detection problems requiring fine-grained feature refinement.
Scalability to Large Feature Sets	Demonstrates robust performance on large feature spaces due to dynamic search control.	Scales adequately, though performance depends on proper parameter tuning in high dimensions.

The ISCX-URL2016 dataset that is created by the Canadian Institute of Cybersecurity (CIC) is a publicly accessible benchmark, which is used in research on URL-based security. It includes over 114,000 labeled URLs of five categories: Benign, Phishing, Malware, Spam, and Defacement. In this paper, a binary sample of 6,712 malware and 7,781 benign URLs is used. The benign samples are based on the domains with lists of Alexa, and the malware URLs are gathered with the help of the DNS-BH blacklist and verified with the help of the expert analysis and VirusTotal reports. The dataset includes 79 lexical features based

on URL components, which allow the effective supervised learning without using content-based inspection. Table 2 presents the ISCX-URL2016 feature set [41,42].

Table 2. Features of the ISCX-URL2016 Malware Dataset

#	Feature Name	Description
F1	Querylength	Length of the URL query string
F2	domain_token_count	Number of tokens in the domain
F3	path_token_count	Number of tokens in the path
F4	avgdomaintokenlen	Average length of domain tokens
F5	longdomaintokenlen	Length of the longest domain token
F6	avgpathtokenlen	Average length of path tokens
F7	tld	Length of the top-level domain
F8	charcompvowels	Number of vowels in the URL
F9	charcompace	Character continuity measure
F10	ldl_url	Longest digit sequence in URL
F11	ldl_domain	Longest digit sequence in domain
F12	ldl_path	Longest digit sequence in path
F13	ldl_filename	Longest digit sequence in filename
F14	ldl_getArg	Longest digit sequence in arguments
F15	dld_url	Digit distribution length in URL
F16	dld_domain	Digit distribution length in domain
F17	dld_path	Digit distribution length in path
F18	dld_filename	Digit distribution length in filename
F19	dld_getArg	Digit distribution length in arguments
F20	urlLen	Total URL length
F21	domainlength	Length of the domain
F22	pathLength	Length of the path
F23	subDirLen	Length of sub-directory
F24	fileNameLen	Length of filename
F25	this.fileExtLen	Length of file extension
F26	ArgLen	Length of arguments
F27	pathurlRatio	Path-to-URL length ratio
F28	ArgUrlRatio	Argument-to-URL length ratio
F29	argDomanRatio	Argument-to-domain ratio

F30	domainUrlRatio	Domain-to-URL ratio
F31	pathDomainRatio	Path-to-domain ratio
F32	argPathRatio	Argument-to-path ratio
F33	executable	Executable file indicator
F34	isPortEighty	Port 80 usage indicator
F35	NumberofDotsinURL	Number of dots in URL
F36	ISIpAddressInDomainName	IP address used in domain
F37	CharacterContinuityRate	Character repetition continuity
F38	LongestVariableValue	Longest query variable value
F39	URL_DigitCount	Digit count in URL
F40	host_DigitCount	Digit count in host
F41	Directory_DigitCount	Digit count in directory
F42	File_name_DigitCount	Digit count in filename
F43	Extension_DigitCount	Digit count in extension
F44	Query_DigitCount	Digit count in query
F45	URL_Letter_Count	Letter count in URL
F46	host_letter_count	Letter count in host
F47	Directory_LetterCount	Letter count in directory
F48	Filename_LetterCount	Letter count in filename
F49	Extension_LetterCount	Letter count in extension
F50	Query_LetterCount	Letter count in query
F51	LongestPathTokenLength	Longest path token length
F52	Domain_LongestWordLength	Longest domain word
F53	Path_LongestWordLength	Longest path word
F54	sub-Directory_LongestWordLength	Longest sub-directory word
F55	Arguments_LongestWordLength	Longest argument word
F56	URL_sensitiveWord	Sensitive keyword indicator
F57	URLQueries_variable	Number of query variables
F58	spcharUrl	Special character count
F59	delimeter_Domain	Domain delimiter count
F60	delimeter_path	Path delimiter count
F61	delimeter_Count	Total delimiter count
F62	NumberRate_URL	Digit ratio in URL
F63	NumberRate_Domain	Digit ratio in domain
F64	NumberRate_DirectoryName	Digit ratio in directory
F65	NumberRate_FileName	Digit ratio in filename
F66	NumberRate_Extension	Digit ratio in extension
F67	NumberRate_AfterPath	Digit ratio after path
F68	SymbolCount_URL	Symbol count in URL
F69	SymbolCount_Domain	Symbol count in domain
F70	SymbolCount_Directoryname	Symbol count in directory

F71	SymbolCount_FileName	Symbol count in filename
F72	SymbolCount_Extension	Symbol count in extension
F73	SymbolCount_Afterpath	Symbol count after path
F74	Entropy_URL	Entropy of URL
F75	Entropy_Domain	Entropy of domain
F76	Entropy_DirectoryName	Entropy of directory
F77	Entropy_Filename	Entropy of filename
F78	Entropy_Extension	Entropy of extension
F79	Entropy_Afterpath	Entropy after path

2. Related Works

Q. Abu Al-Haija and M. Al-Fayoumi [43] create a smart URL classification system with the help of ensembles of machine learning assessed on the ISCX-URL2016 dataset. They employ a two-step detection system: Firstly, binary classification is used to identify benign and malware URLs, and secondly, multi-class is used to detect URLs as benign, spam, phishing, malware, and defacement. The ensemble of bagging trees (En_Bag) has been the best performing variant of the ensemble, with the highest score of 99.3% accuracy in binary classification and 97.92% accuracy in multi-class assessment. This paper establishes the ability of ensemble learning to offer high classification accuracy on a large dataset of URLs.

L. Chen and L. Meng [44] emphasis on metadata-enhanced transformer learning to detect malicious URLs with the help of fine-tuning a RoBERTa-Large model on a balanced set of benign, phishing, defacement, and malware URLs. They use contextualized subword embedding and structural metadata features, including URL length and entropy, to enhance detection in their hybrid model. The proposed approach has an overall classification accuracy of 98 percent when evaluated on a benchmark test set containing different types of URLs. The article shows the importance of integrating transformer models with lightweight URL metadata to enhance interpretability and performance of large-scale tasks on malicious URL detection.

Zhang [45] has introduced a malware detection technique using a Multilayer Perceptron with Principal Component Analysis to reduce features. It is applied to 48 structural and metadata features that are derived using PDF files and evaluated on a real-world PDF dataset of Sophos consisting of 105,000 samples. PCA is used to eliminate redundancy and keep the relevant components, which is used to enhance training efficiency without losing the detection performance. The experiment compares the model on the full set of features and the reduced set of features such as PCA10, PCA28, and PCA32 in a shallow MLP. Findings indicate that PCA32 has a TPR of 93.17% and an FPR of 0.08% whereas PCA10 has a TPR of 98% which means that it is suitable in the detection of PDF malware.

Mohaisen et al. [46] introduce AMAL, a framework of large-scale malware analysis that does not focus on the static analysis of malware but on the behavior-based one. The system is created by integrating the AutoMal and the MaLabel. The behavior of malware is monitored in the sandboxed environments and logs activities involving file system, registry, memory and network. Out of the executions, 65 normalized behavioral features are obtained and measured by machine learning models. The Recursive Feature Elimination is used to rank the features. The 10-fold cross-validation is used in performance evaluation. The experimental outcomes indicate that a linear SVM has accuracy of 97.93%, and a polynomial kernel SVM achieves the highest accuracy of 99.22%.

3. Proposed Malware Detection Framework

3.1. Data Preprocessing

The ISCX-URL2016 malware dataset used in this work contains 14,493 samples represented by 79 features. Before training, preprocessing was applied to handle noisy and missing values. The adopted strategy included data cleaning, imputation, row removal, and feature elimination. Nine features were affected by NaN or infinite values. Median imputation was applied to avgpathtokenlen, Entropy_Filename, and Entropy_Extension due to its robustness to outliers. Records containing isolated invalid values in argPathRatio, NumberRate_DirectoryName, NumberRate_FileName, and

NumberRate_AfterPath were removed because they represented less than 0.01% of the data. Two features, NumberRate_Extension and Entropy_DirectoryName, showed high missing rates and were therefore excluded to avoid introducing noise and bias. After that, Min-max normalization is applied to scale all features to [0, 1], ensuring balanced feature influence, stable training, and faster convergence. Table 3 presents examples of feature values before and after normalization [47-49].

This section describes the proposed malware detection framework, including data preprocessing, feature selection using the HHO and BA algorithms, and classification with the selected machine learning models. Figure 1 illustrates the overall structure of the framework.

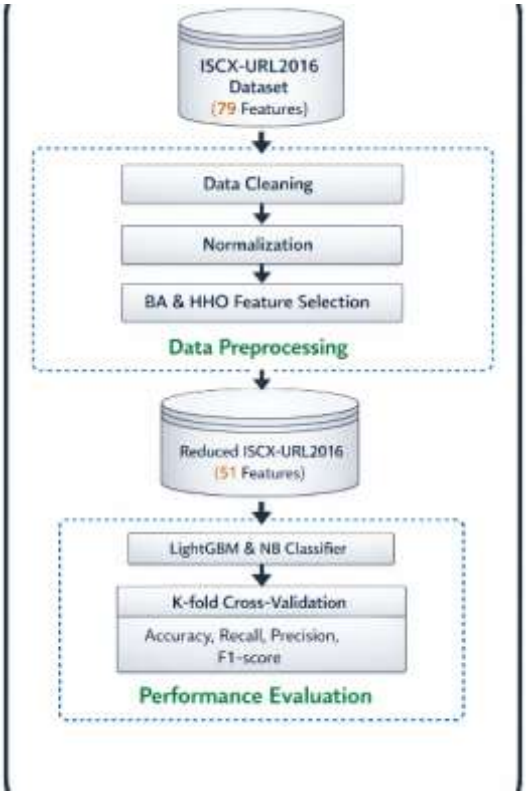

Figure. 1. The proposed Malware detection framework

Table 3. Sample of the dataset before and after normalization

Before normalization	After normalization
2, 2.5, 3, 4.5555553, 6, 9	0, 0.036585365, 0.029411765, 0.055555552, 0.063829787, 0.106060606
2, 3.5, 4, 6, 11, 15	0, 0.109756095, 0.058823529, 0.096774194, 0.136363636, 0.159574468
2, 4.5, 7, 10, 15	0, 0.182926825, 0.147058824, 0.161290323, 0.159574468
2, 4.5, 6, 9, 12, 24	0, 0.182926825, 0.117647059, 0.14516129, 0.151515152, 0.255319149
2, 4, 5, 11, 16	0, 0.14634146, 0.088235294, 0.177419355, 0.170212766
2, 2.25, 5, 9, 10	0, 0.01953125, 0.080645161, 0.205882353, 0.121212121
3, 3.5833333, 8, 10, 12	0.166666667, 0.040364583, 0.129032258, 0.235294118, 0.151515152
2, 4.1, 9, 10, 11, 15	0, 0.0484375, 0.205882353, 0.121212121, 0.177419355, 0.159574468
2, 4.0833335, 7, 8, 12, 15	0, 0.048177086, 0.112903226, 0.176470588, 0.151515152, 0.159574468

3.2. Proposed feature selection methods

Feature selection is the process of identifying and selecting a subset of the most relevant features from a dataset to improve model performance, reduce complexity, and eliminate redundant or irrelevant

information [50-52]. This paper presents a hybrid feature selection method that is founded on union concepts. The union method takes two sets of features derived with the help of two metaheuristic algorithms and merges them into one set. The purpose of this strategy is to take advantage of the complementary advantages of the two algorithms by keeping more informative features. It assists in decreasing bias that the use of one individual search process may result in and aids in better balancing between the minimization of features and the maintenance of information. It is therefore important to choose the algorithms which have different and complementary search behaviors. The reasons why HHO and BA were selected as good candidates of the union-based feature selection strategy are explained in Table 4 [37-40].

The proposed feature selection approach integrates the outputs of two metaheuristic algorithms. HHO and BA are applied independently to the ISCX-URL2016 dataset to select relevant feature subsets. The final feature set is obtained by combining the outputs of both algorithms using the union operation, referred to as HHO∪BA. Table 5 shows the feature subsets selected by HHO, BA, and HHO∪BA. This strategy preserves complementary features selected by each algorithm while reducing redundancy [37-40].

Table 4. Complementary characteristics of HHO and BA supporting the union-based feature selection strategy

Aspect	HHO Characteristics	BA Characteristics
Exploration strategy	Emphasizes global exploration through random jumps, Levy flight, and energy-driven transitions.	Provides moderate exploration using frequency control and stochastic velocity updates.
Update formulation	Updates positions based on prey escape energy and adaptive besiege strategies.	Updates positions using frequency-modulated velocity and stochastic local search.
Search dynamics	Strong ability to escape local regions through adaptive switching between exploration and exploitation phases.	Gradual refinement around promising regions through controlled random walks.
Feature selection tendency	Prioritizes highly discriminative features with strong individual impact.	Preserves weak but cumulatively informative features through local refinement.
Interaction with fitness landscape	Reacts dynamically to changes in the fitness function using energy modulation.	Smoothly adapts to local fitness variations via frequency and pulse adjustments.
Suitability for union strategy	Contributes diverse and high-impact features from global search behavior.	Contributes complementary features identified through fine-grained local search.

Table 5. Feature subsets selected by HHO, BA, and HHO∪BA

Method	Selected Features
HHO	F2, F3, F5, F6, F15, F23, F24, F29, F30, F41, F45, F46, F50, F52, F53, F56, F57, F60, F61, F63, F69, F70, F71, F73, F74, F75, F78
BA	F1, F2, F4, F5, F7, F8, F9, F10, F16, F19, F20, F25, F29, F32, F33, F37, F38, F40, F43, F44, F45, F46, F49, F50, F52, F55, F56, F58, F59, F61, F68, F69, F70, F71, F72, F75, F78, F79
HHO∪BA	F1, F2, F3, F4, F5, F6, F7, F8, F9, F10, F15, F16, F19, F20, F23, F24, F25, F29, F30, F32, F33, F37, F38, F40, F41, F43, F44, F45, F46, F49, F50, F52, F53, F55, F56, F57, F58, F59, F60, F61, F63, F68, F69, F70, F71, F72, F73, F74, F75, F78, F79

3.3. LightGBM and NB for Classification

LightGBM is a gradient boosting classifier that builds decision trees sequentially using gradient optimization, allowing efficient learning on large and high-dimensional datasets. Its leaf-wise growth strategy balances accuracy and computational cost. In contrast, Naive Bayes is a probabilistic classifier based on Bayes' theorem that assumes feature independence, leading to fast training and inference with

minimal overhead. This makes NB suitable for large datasets and real-time tasks. Table 6 compares the LightGBM and NB classifiers [53-56].

Table 6. Comparison of the LightGBM and NB classifiers.

Aspect	LightGBM	NB
Learning Approach	Gradient boosting ensemble learning based on decision trees.	Probabilistic learning based on Bayes' theorem.
Working Mechanism	Builds trees sequentially using gradient-based optimization with leaf-wise growth.	Estimates class probabilities assuming conditional independence among features.
Strengths	High accuracy, fast training, and efficient handling of large-scale and high-dimensional data.	Very fast inference, low computational cost, and effective for real-time applications.
Performance on Attack Detection	Strong capability in modeling complex and nonlinear attack patterns.	Effective for simple patterns, but limited in capturing feature dependencies.
Computational Efficiency	Optimized for speed and memory through histogram-based splitting.	Extremely lightweight with minimal computational overhead.
Default Hyperparameters	num_leaves=31, learning_rate=0.1, n_estimators=100, max_depth=-1	alpha=1.0 (Laplace smoothing, depending on NB variant)

4. Performance evaluation

The performance of the proposed model is evaluated using four commonly adopted metrics: accuracy, recall, precision, and F1-score. These metrics are selected to provide a comprehensive assessment of classification effectiveness. All evaluation measures are derived from the confusion matrix, which consists of four outcomes: true positive (TP), true negative (TN), false positive (FP), and false negative (FN). Table 7 presents the definitions and equations of the evaluation metrics [57-61].

Table 7. Evaluation metrics

Metric	Definition	Purpose	Equation
Accuracy	Ratio of correctly classified samples to total samples	Measure overall prediction correctness	(TP + TN) / (TP + TN + FP + FN)
Recall	Ratio of correctly detected positives	Evaluate detection capability	TP / (TP + FN)
Precision	Ratio of correct positive predictions	Reduce false positive errors	TP / (TP + FP)
F1-score	Harmonic mean of precision and recall	Balance recall and precision	2 x (Precision x Recall) / (Precision + Recall)

Accuracy, precision, recall, and F1-score were used to report the performance of LightGBM and Naive Bayes in figures 2-5. LightGBM scores 99.52%, 99.48%, 99.48%, and 99.48% on accuracy, precision, recall and F1-score respectively, which means that it can achieve very high and consistent classification accuracy. On the other hand, Naive Bayes achieves an accuracy of 81.12%, precision of 67.36%, recall of 89.24% and F1-score of 76.77%. Naive Bayes has less accuracy and precision, which is an indication that it cannot model dependent features, whereas LightGBM is useful in modeling more intricate feature interactions. The findings also validate that the proposed feature selection strategy gives very informative features, especially to the advantage of LightGBM and resulting in a strong malware detection.

The performance gains are further supported by the proposed union-based feature selection method HHO∪BA. By combining the complementary search behaviors of Harris Hawks Optimizer and Bat Algorithm, the method selects informative and diverse features while reducing redundancy. This balanced feature subset enhances model learning and aligns well with LightGBM, which benefits from rich feature

interactions. The results confirm that HHOUBA effectively supports robust and accurate malware detection.

Figure 6 is a comparison of the accuracy of the proposed approach with some of the existing studies. LightGBM classifier offers an accuracy of 99.52, which is higher than Ref [43] at 99.30%, Ref [44] and Ref [45]at 98% and Ref [46] at 99.22. The NB classifier on the other hand has an accuracy of 81.12% which is not very effective in this task. The union-based HHOUBA by LightGBM further contributes to the better performance of LightGBM since the selection method of features (HHOUBA) identifies different and informative features without redundancy. This sub-set of balanced features increases the capability of LightGBM to capture complicated interactions of features and also increases the general detection reliability.

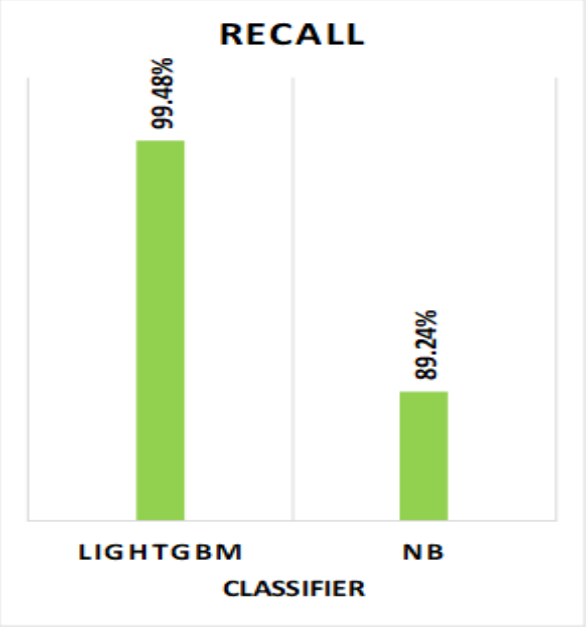

Figure 2. Malicious URL Detection Recall

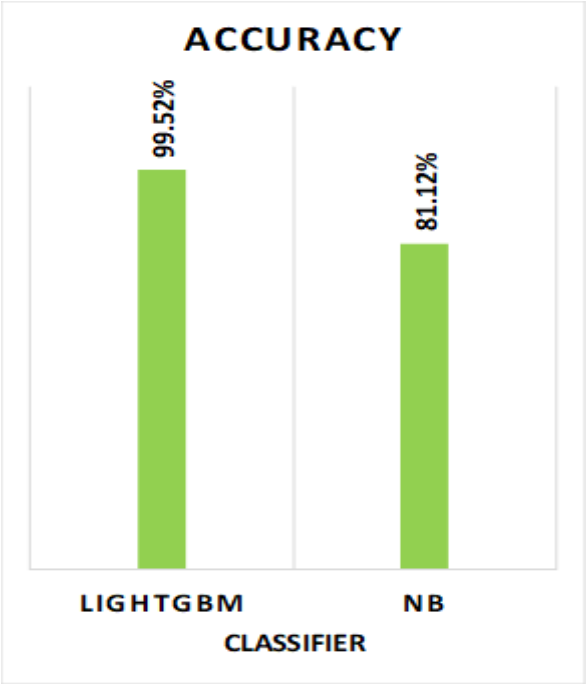

Figure 3. Malicious URL Detection Accuracy

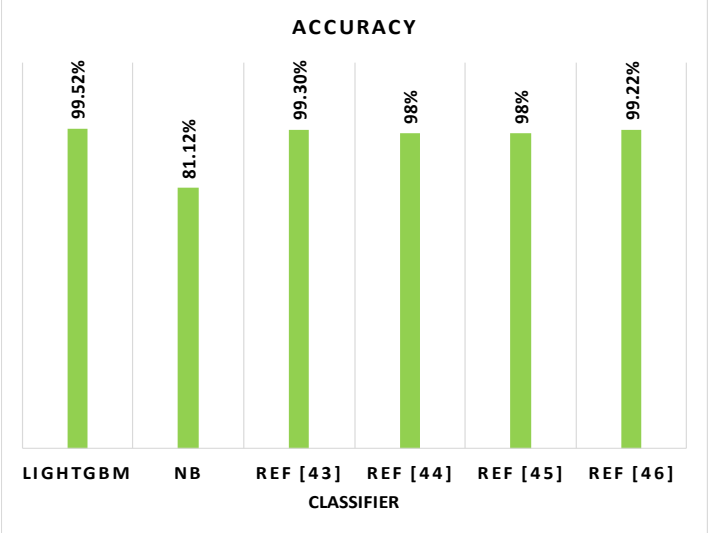

Figure 4. Accuracy of the proposed method compared to existing works

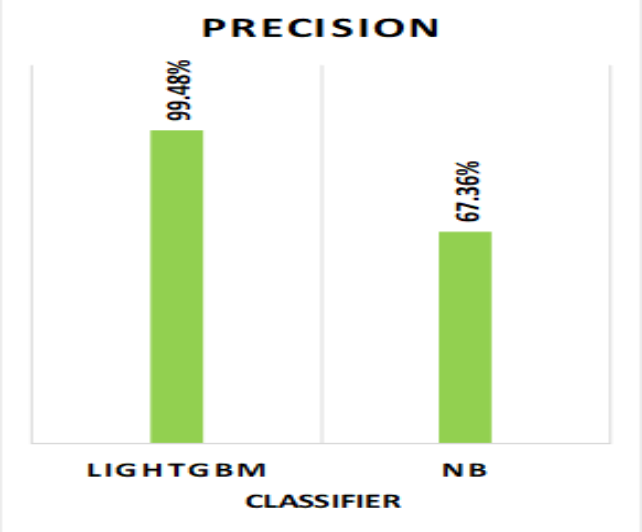

Figure 5. Malicious URL Detection Precision

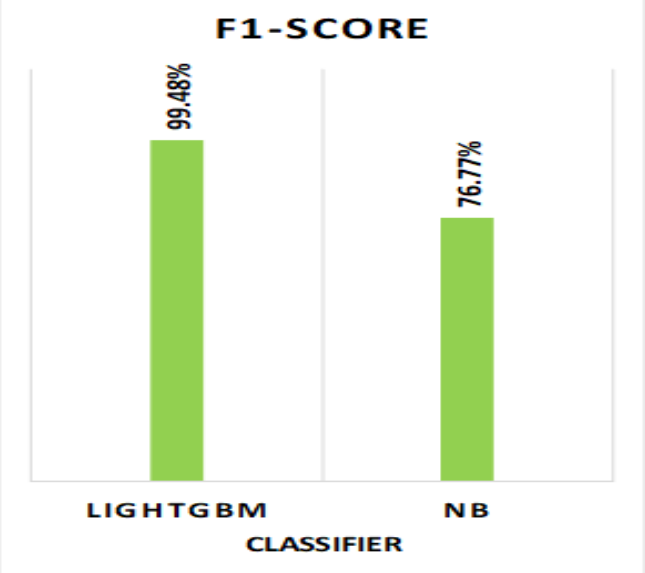

Figure 6. Malicious URL Detection F1-Score

5. Conclusion

In this paper, a malware URL detection framework was introduced that is based on union-based feature selection and machine learning classification. The HHO∪BA approach proposed combines feature subsets produced by HHO with feature subsets produced by BA to create a single set of features that embrace two distinct types of information and reduce redundancy. It was experimentally tested on the ISCX-URL2016 dataset that LightGBM was able to reach a 99.52% accuracy compared to NB and other related studies. The findings substantiate the fact that tree-based ensemble models have the advantage of using diverse and informative feature subsets since the models are able to use feature interactions and nonlinear patterns to their advantage. Naive Bayes, on the other hand, was limited by the assumption of conditional independence, which restricted its use on structured URL data. All in all, the research indicates that the union-based metaheuristic feature selection leads to a high level of reliability in detection and efficiency in learning.

References

1. T. M. Ghazal, "Privacy-Preserving Transformer-Based Meta-Learning Algorithm for Zero-Day Intrusion Detection in IoT-Enabled Smart Grids," Journal of Internet Services and Information Security (JISIS), vol. 15, no. 4, pp. 62–79, Nov. 2025. doi: 10.22667/JISIS.2025.11.30.062
2. M. Al Batahari, N. S. A. Shukor, A. A. Aziz, and T. M. Ghazal, "Dynamic Load Balancing in Cloud Computing Using Machine Learning," in Proc. 2025 3rd International Conference on Business Analytics for Technology and Security (ICBATS), 2025. doi: 10.1109/ICBATS66542.2025.11258155
3. H. Al-Ahmed et al., "Exploring the Effectiveness of Personalized Marketing Strategies in E-commerce Platforms," in Artificial Intelligence and Digital Transformation in Business, Springer, 2025, pp. 567–579. doi: 10.1007/978-3-031-95280-7_53
4. Y. A. Maz, M. Anbar, S. Manickam, M. M. Abualhaj, S. A. Almalki, and B. A. Alabsi, "Transfer Learning-Based Approach with an Ensemble Classifier for Detecting Keylogging Attack on the Internet of Things," Computers, Materials & Continua, vol. 82, no. 1, pp. 1–10, Jan. 2025. doi: 10.32604/cmc.2025.068257
5. A. I. Karajeh, T. A. AL-Momani, M. A. Almomani, O. Almomani, and M. H. Tarawneh, "Forecasting future earnings via e-business information: Financial implications for investment decisions in the era of big data," Investment Management and Financial Innovations, vol. 22, no. 4, pp. 237-246, Nov. 2025. doi: 10.21511/imfi.22(4).2025.19
6. I. A. Alshafei et al., "Digital Transformation in Design Education: Exploring the Challenges and Opportunities in Jordanian Higher Education," Computers, vol. 14, no. 12, Art. no. 535, Dec. 2025. doi: 10.3390/computers14120535
7. N. K. Al-Okbi et al., "Unsupervised Learning: Discovering Patterns without Labels: Health Care, E-Commerce, and Cybersecurity," in Mastering the Minds of Machines, L. Abualigah, Ed. CRC Press, 2025, pp. 34–41. doi: 10.1201/9781032840634-3
8. N. M. Alaskar, M. Hussain, S. J. Almheiri, Atta-ur-Rahman, A. Khan, and K. M. Adnan, "Big Data-Driven Federated Learning Model for Scalable and Privacy-Preserving Cyber Threat Detection in IoT-Enabled Healthcare Systems," Computers, Materials & Continua, early access, Dec. 2025. doi: 10.32604/cmc.2025.074041
9. Z. Awais et al., "ISCC: Intelligent Semantic Caching and Control for NDN-Enabled Industrial IoT Networks," IEEE Access, vol. 13, pp. 169881-169898, 2025. doi: 10.1109/ACCESS.2025.3614984
10. M. M. Abualhaj, S. N. Al-Khatib, M. Al-Zyoud, I. Qaddara, M. O. Hiari, and S. M. A. Aldossary, "Enhanced Network Communication Security Through Hybrid Dragonfly-Bat Feature Selection for Intrusion Detection," Journal of Communications, vol. 20, no. 5, pp. 607-618, 2025. doi: 10.12720/jcm.20.5.607-618
11. N. K. Al-Okbi et al., "Defending Digital Integrity: Advances in Media Forgery Analysis Research and Cybersecurity Development," SN Computer Science, vol. 6, no. 8, pp. 1–18, 2025. doi: 10.1007/s42979-025-04462-8
12. G. A. Hindi et al., "Hacking Back: Using Genetic Algorithms to Outsmart Hackers," Journal of Computer Science, vol. 21, no. 9, pp. 2049–2064, Oct. 2025. doi: 10.3844/jcssp.2025.2049.2064
13. M. Abualhaj et al., "Comparative Analysis of LSTM-Based Variant Models for Detecting Attacks in IoT Networks," Journal of Computing & Biomedical Informatics, vol. 10, no. 1, 2025.
14. Aljaidi, M., Samara, G., Singla, M. K., Alsarhan, A., Hassan, M., Safaraliev, M., ... & Tavlintsev, A. (2024). A particle swarm optimizer-based optimization approach for locating electric vehicles charging stations in smart cities. International Journal of Hydrogen Energy, 87, 1047-1055.
15. Al-Na'amneh, Q., Aljaidi, M., Nasayreh, A., Gharaibeh, H., Al Mamlook, R. E., Jaradat, A. S., ... & Samara, G. (2024). Enhancing IoT device security: CNN-SVM hybrid approach for real-time detection of DoS and DDoS attacks. Journal of Intelligent Systems, 33(1), 20230150.
16. M. M. Abualhaj et al., "Spam Detection Boosted by Firefly-Based Feature Selection and Optimized Classifiers," International Journal of Advances in Soft Computing and Its Applications (IJASCA), vol. 17, no. 3, pp. 1–19, 2025. doi: 10.15849/IJASCA.251130.01
17. M. M. Abualhaj, S. N. Al-Khatib, A. A. Abu-Shareha, A. Hyassat, and M. Sh. Daoud, "Smart Firewall for Phishing Detection Powered by Bio-Inspired Algorithms," Journal of Advances in Information Technology, vol. 16, no. 11, pp. 1529-1539, 2025. doi: 10.12720/jait.16.11.1529-1539
18. Abualhaj, M.M.; Hiari, M.O.; Al-Khatib, S.N.; Abu-Shareha, A.A.; Daoud, M.S.; Faheem, M.R.; Al-Allawee, A.; Anbar, M. A Robust IDS System for Intelligent Phishing Website Detection. Int. J. Electr. Electron. Eng. Telecommun. 2026, 15, 1-12.
19. Almomani, O.; et al. A Robust Model for Android Malware Detection via ML and DL classifiers. Mesopotamian J. Big Data 2025, 261–277. doi.org

20. Alsarhan, A., & Agarwal, A. (2009, December). Spectrum sharing in multi-service cognitive network using reinforcement learning. In 2009 First UK-India International Workshop on Cognitive Wireless Systems (UKIWCWS) (pp. 1-5). IEEE.
21. Alabdallat, R.; Abualhaj, M.; Abu-Shareha, A. Android Malware Detection Using a Modified Dwarf Mongoose Algorithm. Int. J. Intell. Eng. Syst. 2025, 18. doi.org
22. Alabdallat, R.; Abualhaj, M.; Abu-Shareha, A. Enhanced Multiclass Android Malware Detection Using a Modified Dwarf Mongoose Algorithm. Int. J. Anal. Appl. 2025, 23, 1–23. doi.org
23. Abualhaj, M.M.; et al. Intelligent Malware Detection through Bio-Inspired Optimization and Gradient Boosting. J. Adv. Inf. Technol. 2026, 17, 1-13.
24. Abualhaj, M.M.; et al. A Bio-Inspired Hybrid Optimization Framework for Efficient and Accurate Real-Time Malware Detection. Sci. Rep. 2026, 16.
25. Abualhaj, M.M.; Al-Khatib, S.; Al-Zyoud, M.; Hiari, M.O.; Al-Allawee, A.; Alsharaiah, M.A. A Customized Machine Learning Model for Improving Malware Detection. Int. J. Comput. Netw. Inf. Secur. 2026, 18, 1–17. doi.org
26. Hasan, R.; Biswas, B.; Samiun, M.; et al. Enhancing malware detection with feature selection and scaling techniques using machine learning models. Sci. Rep. 2025, 15, 9122. https://doi.org/10.1038/s41598-025-93447-x
27. Almobaideen, W.; Abu Alghanam, O.; Abdullah, M.; et al. Comprehensive review on machine learning and deep learning techniques for malware detection in android and IoT devices. Int. J. Inf. Secur. 2025, 24, 110. https://doi.org/10.1007/s10207-025-01027-x
28. Khan, M.A.; Abbas, S.; Rehman Sakhawat, A.; Ali, A. Pioneering Adaptive Machine Learning Solutions for Smart Grid Integrity. ResearchGate 2025. doi.org
29. Chopra, M.; Reddy, R.; Chopra, S. Leveraging LSTM Neural Networks and ARIMA Models for Enhanced Real-Time Sales Forecasting in Dynamic Retail Environments. JoAIMLR 2024, 1. https://joaimlr.com/index.php/v1/article/view/35
30. Ayodeji, D.C.; et al. Advanced risk assessment using machine learning and sentiment analysis on log data. Int. Multidiscip. Res. J. 2025, 1.
31. M. M. Abualhaj, Q. Y. Shambour, A. A. Abu-Shareha, S. N. Al-Khatib, and A. Amer, "Enhancing malware detection through self-union feature selection using gray wolf optimizer," Indonesian Journal of Electrical Engineering and Computer Science, vol. 37, no. 1, p. 197, Jan. 2025, doi: https://doi.org/10.11591/ijeecs.v37.i1.pp197-205.
32. M. M. Abualhaj, M. O. Hiari, A. Alsaaidah, and M. M. Al-Zyoud, "Comparative analysis of whale and Harris Hawks optimization for feature selection in intrusion detection," Indonesian Journal of Electrical Engineering and Computer Science, vol. 37, no. 1, p. 179, Jan. 2025, doi: https://doi.org/10.11591/ijeecs.v37.i1.pp179-185.
33. Panda, B.; Bisoyi, S.S.; Panigrahy, S.; Mohanty, P. Machine learning techniques for imbalanced multiclass malware classification through adaptive feature selection. PeerJ Comput. Sci. 2025, 11, e2752. doi.org
34. Masadeh, R.; Almomani, O.; Zaqebah, A.; Masadeh, S.; Alshqurat, K.; Sharieh, A.; Alsharman, N. Narwhal Optimizer: A Nature-Inspired Optimization Algorithm for Solving Complex Optimization Problems. Comput. Mater. Contin. 2025, 85, 2. doi: 10.32604/cmc.2025.63798
35. Almomani, O.; Al-Zyoud, M.; Abu-Shareha, A.A.; Almomani, A.; Salloum, S.A.; Alomari, K.M. Adaptive Enhanced Grey Wolf Optimizer for Efficient Cluster Head Selection and Network Lifetime Maximization in Wireless Sensor Networks. Comput. Mater. Contin. 2026, 87, 33. doi: 10.32604/cmc.2025.66621
36. Fakhouri, H.N.; Al-Shamayleh, A.S.; Istiwi, A.; Makhadmeh, S.N.; Hamad, F.; Fakhouri, S.N.; Alyasseri, Z.A.A. Hybrid white shark optimizer with differential evolution for training multi-layer perceptron neural network. *J. Intell. Syst.* 2025, *34*, 20240362. doi: 10.1515/jisys-2024-0362
37. Abualhaj, M.M.; Al-Khatib, S.N.; Hiari, M.O.; Shambour, Q.Y.; Al-Allawee, A.; Almomani, O.; Daoud, M.S.; Anbar, M. An Efficient Feature Selection Technique to Enhance Spam Email Detection. TEM J. 2026, 15, 91-101. doi.org
38. Mohammad, R.M. Android malware detection using a novel binary Firefly Bat feature selection algorithm. Inf. Secur. J. Glob. Perspect. 2026, 1–27. doi.org
39. Laamari, M.A.; Kamel, N. A New Multi-Objective Binary Bat Algorithm for Feature Selection in Intrusion Detection Systems. Concurr. Comput. Pract. Exp. 2025, 37, e70000. doi.org
40. Azeem, S.; Javed, S.; Naseer, I.; Ali, O.; Ghazal, T.M. A New Hybrid PSO-HHO Wrapper Based Optimization for Feature Selection. IEEE Access 2025, 13, 87090-87099. doi.org
41. Almomani, O.; Alsaaidah, A.; Abu-Shareha, A.A.; Alzaqebah, A.; Almaiah, M.A.; Shambour, Q. Enhance URL Defacement Attack Detection Using Particle Swarm Optimization and Machine Learning. J. Comput. Cogn. Eng. 2025. doi.org

42. Rizk, F.; Rizk, R.; Rizk, D.; Rizk, P.; Chu, C.H.H. KAN-MID: A Kolmogorov-Arnold Networks-Based Framework for Malicious URL and Intrusion Detection in IoT Systems. IEEE Access 2025. doi.org
43. Al-Haija, Q.A.; Al-Fayoumi, M. An intelligent identification and classification system for malicious uniform resource locators (URLs). Appl. Sci. 2023, 13, 4261. doi.org
44. Chen, L.; Meng, L. Metadata driven malicious URL detection using RoBERTa large and multi source network threat intelligence. Sci. Rep. 2025, 15, 34790. doi.org
45. Zhang. Machine learning with feature selection using principal component analysis for malware detection: A case study. arXiv 2019, arXiv:1902.03639. doi.org
46. Mohaisen, A.; Alrawi, O.; Mohaisen, M. AMAL: High-fidelity, behavior-based automated malware analysis and classification. Comput. Secur. 2015, 52, 251–266. doi.org
47. Cabello-Solorzano, K.; et al. The Impact of Data Normalization on the Accuracy of Machine Learning Algorithms: A Comparative Analysis. Lect. Notes Netw. Syst. 2023, 750, 341-351. doi.org
48. Ahmed, H.A. An Investigation on Disparity Responds of Machine Learning Algorithms to Data Normalization Method. ARO-Sci. J. Koya Univ. 2022, 10, 29–37. doi.org
49. Abu-Shareha, A.A.; Amer, A. Improving linear support vector machine classifier using mutual information feature weighting for intrusion detection system. *Proc. SPIE 13731, Seventh International Conference on Image, Video Processing, and Artificial Intelligence* 2025, 137310Z. doi: 10.1117/12.3075315
50. Al-shadeedi, Y.A.; Abualhaj, M.; Abu-Shareha, A.; Yousif, M.; Achuthan, A. Improved Intrusion Detection System Using a Modified African Vultures Algorithm. J. Commun. 2026, 21, 1-11.
51. Al Mosuli, A.; Abualhaj, M.; Abu-Shareha, A.; Yousif, M.; Daoud, M. Enhancing Intrusion Detection System Performance Using a Modified Grey Wolf Optimizer. Ann. Emerg. Technol. Comput. 2026, 10, 21-44. doi.org
52. Sanjalawe, Y.; Fraihat, S.; Al-E'mari, S.; Abualhaj, M.M.; Makhadmeh, S.; Alzubi, E. Smart load balancing in cloud computing: Integrating feature selection with advanced deep learning models. PLOS ONE 2025, 20, e0329765. doi.org
53. Abu-Shareha, A.A.; Abualhaj, M.; Hussein, A.H.; Amer, A.; Achuthan, A.; Halin, A.A. Diabetes Prediction Using Hybrid Supervised and Unsupervised Techniques Based on PIMA Dataset. J. Artif. Intell. Technol. 2025.
54. Abu-Shareha, A.A.; Abualhaj, M.M.; Hussein, A.; Almomani, O.; Amer, A.; Achuthan, A.; Halin, A.A. A comparative study of the diabetes progression prediction techniques. Discov. Artif. Intell. 2025. doi.org
55. Luo, Y.; Chen, R.; Li, C.; Yang, D.; Tang, K.; Su, J. An improved binary simulated annealing algorithm and TPE-FL-LightGBM for fast network intrusion detection. Electronics 2025, 14, 231. doi.org
56. Wei, G.; Yi, L. Predicting the impact of yoga on chronic venous insufficiency: a machine learning approach using Naive Bayes classifier and optimization systems. J. Ambient Intell. Humaniz. Comput. 2025, 16, 109–131. doi.org
57. Majid, M.D.; Anwar, M.; Bilal, S.F.; et al. A hybrid learning framework for automated multiclass electrocardiogram classification with SimCardioNet. Sci. Rep. 2026. doi.org
58. Zubair, M.; Owais, M.; Hassan, T.; et al. An interpretable framework for gastric cancer classification using multi-channel attention mechanisms and transfer learning approach on histopathology images. Sci. Rep. 2025, 15, 13087. doi.org
59. Abu Owida, H.; Arabiat, A.; Al-Ayyad, M.; Altayeb, M. Advancements in machine learning techniques for precise detection and classification of lung cancer. Bull. Electr. Eng. Informatics 2025, 14, 6. doi: 10.11591/eei.v14i6.10527
60. Alsarhan, A., Agarwal, A., Obeidat, I., Bsoul, M., Al-Khasawneh, A., & Kilani, Y. (2013). Optimal spectrum utilisation in cognitive network using combined spectrum sharing approach: overlay, underlay and trading. International Journal of Business Information Systems, 12(4), 423-454.
61. Hussain, M.; Chen, C.; Hussain, M.; et al. Optimised knowledge distillation for efficient social media emotion recognition using DistilBERT and ALBERT. Sci. Rep. 2025, 15, 30104. https://doi.org/10.1038/s41598-025-16001-9.