Baghdad Science Journal

Volume 23 | Issue 5 Article 27

5-29-2026

# Classification of Phishing Data Using Hybrid MI-AN Feature Selection Method

Damodar Patel
*Department of CSIT, Guru Ghasidas Vishwavidyalaya Bilaspur, Bilaspur, India*,
damodarpatel7497@gmail.com

Amit Kumar Saxena
*Department of CSIT, Guru Ghasidas Vishwavidyalaya Bilaspur, Bilaspur, India*,
amitsaxena65@rediffmail.com

Wutiphol Sintunavarat
*Department of Mathematics and Statistics, Faculty of Science and Technology, Thammasat University, Pathum Thani, Thailand*, wutiphol@mathstat.sci.tu.ac.th

Abhishek Dubey
*College of Computing and Information Sciences, Information Technology Department, University of Technology and Applied Sciences, Salalah, Sultanate of Oman*, abhishekdubey003@gmail.com



## How to Cite this Article

Patel, Damodar; Saxena, Amit Kumar; Sintunavarat, Wutiphol; and Dubey, Abhishek (2026) "Classification of Phishing Data Using Hybrid MI-AN Feature Selection Method," *Baghdad Science Journal*: Vol. 23: Iss. 5, Article 27.
DOI: https://doi.org/10.21123/2411-7986.5311

BAGHDAD SCIENCE JOURNAL 2026;23(5):1920–1940



RESEARCH ARTICLE

# Classification of Phishing Data Using Hybrid MI-AN Feature Selection Method

**Damodar Patel** [1], **Amit Kumar Saxena** [1], **Wutiphol Sintunavarat** [2,*], **Abhishek Dubey** [3]

[1] Department of CSIT, Guru Ghasidas Vishwavidyalaya Bilaspur, Bilaspur, India
[2] Department of Mathematics and Statistics, Faculty of Science and Technology, Thammasat University, Pathum Thani, Thailand
[3] College of Computing and Information Sciences, Information Technology Department, University of Technology and Applied Sciences, Salalah, Sultanate of Oman

## ABSTRACT

Web phishing attacks have been continually evolving over the past few years, which has led customers to lose their trust in online services and e-commerce. To identify phishing data, a variety of methods and systems based on a blacklist of phishing websites are used. However, the rapid development of technology has given rise to increasingly complex techniques for creating user-attracting websites. Therefore, current blacklist-based techniques are unable to identify the recently launched phishing data, such as zero-day phishing websites. Machine learning techniques have been used in several recent studies to detect phishing data and function as an early warning system for such attacks. However, majority of these methods selected the most significant features of websites based on frequency analysis or human experience. In order to improve the classification of phishing data, this research proposes the hybrid MI-AN method for intelligent phishing data classification using three phishing datasets. We have used six machine learning models with the proposed MI-AN feature selection method. The experimental results indicated that the proposed MI-AN method achieved significant improvements in classification accuracy. Compared to other models, the multi-layer perceptron achieves higher accuracy (97.07%) on dataset 1, the random forest model achieves higher accuracy (96.60% and 97.78%) on both datasets 2 and 3. We have also compared proposed method to a previously published research work, and our method achieves better performance. Overall, the proposed MI-AN method consistently enhanced model performance across all phishing datasets, demonstrating robustness, adaptability, and effectiveness in eliminating irrelevant features to achieve better generalization.

**Keywords:** ANOVA feature selection, Hybrid feature selection, Machine learning, Mutual information, Phishing datasets

## Introduction

In the last few years, the number of web users who use e-banking, online shopping, and online services has been exponentially increasing due to their comfort, flexibility, and ease of use. The huge increase in the usage of online services and e-business has motivated several phishers and cybercriminals to create and publish false and phishing websites [1,2] in an attempt to get users' financial and personal information. As a result, online phishing attempts are becoming a major issue for both commercial websites and web users. [3]

Phishers can access user information using several kinds of innovative methods, including forums, Internet Relay Chat, keyloggers, trojans, IMs (instant messaging), and screen captures. Another kind of phishing attack is DNS-based phishing, often known as pharming, in which hackers change the host's files or the domain name system. As a consequence, a fake

Received 25 August 2025; revised 26 October 2025; accepted 16 November 2025.
Available online 29 May 2026

* Corresponding author.
E-mail addresses: damodarpatel7497@gmail.com (D. Patel), amitsaxena65@rediffmail.com (A. K. Saxena), wutiphol@mathstat.sci.tu.ac.th (W. Sintunavarat), abhishekdubey003@gmail.com (A. Dubey).

https://doi.org/10.21123/2411-7986.5311

address is returned in response to URL queries, and any further correspondence is directed to a phishing website. Another severe kind of targeted phishing attack is spear-phishing via email, in which phishers send fake emails to certain people within an organization pretending to be corporate authorities in order to get important information about the company. [4,5]

This study focuses on phishing website-based attacks, which are among the most harmful types of phishing attempts. The phisher creates a fake or phishing website that duplicates a genuine website in order to trick the victims into providing their financial and personal information. Then, using email, social media, or ads on other websites, the phisher attempts to get a large number of victims to this phishing page. When victims visit a phishing website, they are tricked into believing it is a trustworthy website and provide their financial and personal information.

Machine learning techniques[6] have been used in several recent studies to detect phishing websites and function as an early warning system for such attacks. However, the majority of these methods selected the most significant features using feature selection methods. Feature selection[7–10] is the process of selecting the important features and removing redundant features. Feature selection may be divided into four categories: filter, wrapper, embedded, and hybrid. Filter methods are computationally efficient because they rank features using statistical metrics such as correlation, Laplacian score (see more details in[11]), mutual information, and ANOVA without requiring learning models. Wrapper methods,[12,13] which are computationally costly and prone to overfitting, evaluate feature subsets using a predictive model while taking interactions into account like optimization techniques.[14] Feature selection is integrated into model training using embedded methods,[15] such as Lasso or tree-based algorithms, which balance accuracy and efficiency. Combining filter and wrapper/embedded techniques, hybrid methods capitalize on their advantages to provide robust and efficient selection.

Machine learning (ML) models are powerful tools for identifying and preventing phishing attempts. In this study, three phishing datasets with different features and instances are used, and six supervised ML models are employed: Decision Tree (DT),[16] K-Nearest Neighbor (K-NN),[17] Logistic Regression (LR),[18] Multi-Layer Perceptron (MLP),[19] Naïve Bayes (NB),[20] and Random Forest (RF).[21] These models are evaluated using performance metrics such as classification accuracy (CA), F1-score, precision, recall, and specificity. The primary objective of this study is to enhance the performance of phishing website classification by analyzing its features and selecting the most relevant feature set. To achieve this, we propose a hybrid supervised filter-based feature selection method. The effectiveness of the proposed method is assessed by comparing it with recent state-of-the-art approaches.

The structure of the paper is as follows: Section 2 provides a few feature selection techniques. Section 3 presents the proposed model. The experiments are shown in Section 4. The results and the discussion are shown in Section 5. The last section summarizes the conclusion and future work.

## Literature review

An overview of relevant research on phishing datasets is provided in this section. The study proposed by Babagoli et al.[22] a phishing website detection technique, which includes feature selection and a nonlinear regression algorithm based on meta-heuristics. This study selected the optimum feature subset using two feature selection techniques: decision trees and wrappers. The two meta-heuristic algorithms, harmony search (HS), which is based on the nonlinear regression approach, and support vector machine (SVM), have been effectively developed for predicting and identifying fake websites after the feature selection method. The proposed HS algorithm creates new harmony by adjusting the pitch dynamically. This proposed model harmony search achieves 92.80% CA.

Chiew et al.[23] introduced the Hybrid Ensemble Feature Selection (HEFS) architecture for ML-based phishing detection methods, implemented using an RF classifier. HEFS is more computationally efficient and can determine the optimal number of features for a given dataset. For the UCI phishing dataset, their approach achieved an accuracy of 96.17%.

In order to improve the classification of phishing websites, in 2012, Taha[24] proposed an intelligent ensemble learning method based on weighted soft voting. According to the experimental findings, the proposed intelligent method for phishing website classification performed better than the soft voting method and base classifiers, achieving a maximum CA of 95% and an area under the curve (AUC) of 98.8%.

Later, Sharma et al.[25] introduced a hybrid algorithm for identifying between legitimate and phishing websites. In order to map a continuous search space to a discrete search space, the S-shaped and V-shaped transfer functions are used. 97.04% CA was obtained. The proposed algorithm has been contrasted with six filter techniques and seven metaheuristic feature selection algorithms. At the same time, Siddiq et al.[26]

employed a supervised learning method for phishing website detection utilizing most of the deep learning techniques. This proposed method obtained 93.6% CA with the CNN (Conv2D) model and 94.8% CA with the standard neural network model.

In a subsequent study, Fadheel et al.[27] explored three approaches, evaluated using LR, SVM, and K-NN classification algorithms: complete features, KMO test as a feature selection method, and PCA as a dimensionality method. The CA of the PCA method is 92% for LR and 94% for SVM.

More recently, Karim et al.[28] investigated hybrid machine learning models that were trained using three SVM models and four boosting algorithms after a custom pre-processing step to exclude marginally correlated data. Following hyperparameter settling in, XGBoost outperformed the other models in the test, achieving a CA of 97.05%.

Most recently, Medelbekov et al.[29] examined phishing detection using machine learning models such as Random Forest, SVM, Naive Bayes, and Decision Trees. Random Forest shows more effectiveness, obtaining a CA of 96.7%.

Most recently, Singh et al.[30] proposed several ML classifiers, including DT, RF, LR, KNN, MLP, LGBM, AdaBoost, and XGB, when combined with a PSO-based feature selection technique for the identification of phishing emails. The author determined that the XGB classifier had the best accuracy of 94.69%. Most recently, Davoudi and Yari[31] introduced a machine learning ensemble model using a GA for the classification of phishing websites. The proposed model integrates three machine learning algorithms: Decision Trees, Random Forest, and Support Vector Machines. In conclusion, the findings indicate that the proposed approach outperforms the existing study with an accuracy of 96.04%. In a subsequent study, Ali and Jain[32] employed various machine learning approaches for classifying phishing URLs, with R achieving the highest accuracy of 95.2%.

At the same time, Alazaidah et al.[33] addressed two primary goals are examined. The first is to determine which of the 24 classifiers representing the 6 learning strategies is the best at detecting phishing. The second goal is to determine the most effective feature selection technique for phishing datasets from websites. The findings demonstrated the superiority of Random Forest, Filtered Classifier, and J-48 classifiers in identifying phishing websites using two phishing-related datasets (dataset1 having 30 features and dataset2 having 9 features) with distinct samples and taking into consideration eight evaluation metrics. Additionally, out of the four feature selection techniques taken into consideration, the InfoGainAttributeEval approach performed the best.

Finally, Taha et al.[34] applied a variety of methods, including logistic regression, DT decision trees, adaptive boosting, RF random forest, and XG Boost, to develop machine learning models for the real-time classification of phishing websites. According to experimental data, Random Forest is the most accurate model, with a CA of 96.89%, followed by Decision Trees, which have a CA of 94.57%. In this study, an 80/20 train-test data split is applied for effectively evaluating model performance.

## Proposed MI-AN model

### *Proposed feature selection method*

Fig. 1 illustrates the workflow of the proposed MI-AN feature selection method. In this approach, we introduce a hybrid method, termed MI-AN, which integrates the Mutual Information (MI) technique[35] with the ANOVA-based feature selection method.[36] The MI-AN method takes the complete set of original features as input and generates an optimized subset comprising the most relevant features with reduced dimensionality. To provide a clearer understanding of the procedure, the detailed steps of the proposed MI-AN feature selection method are presented in Algorithm 1.

### *Mutual information feature selection*

Mutual Information evaluates the amount of information about another variable that is included in one variable. It measures the degree of dependency between two variables. Two variables have zero mutual information if they are independent of each other, which means that one variable doesn't reveal anything about the other. The degree to which each feature provides information about the target variable may be measured using mutual information in feature selection. Features are often more relevant when they have a higher mutual information with the target.

In feature selection, Y would be the target variable, and $X$ would be a feature. The mutual information for X and Y, two discrete random variables, is given by:

$$MI(X;Y) = \sum_{x\in X}\sum_{y\in Y} P(x,y)\, log\frac{P(x,y)}{P(x)\,P(y)}, \qquad (1)$$

where $P(x,y)$ is the joint probability distribution of $X$ and $Y$, $P(x)$ and $P(y)$ are the marginal probability distributions of $X$ and $Y$, respectively, and the sum is taken over all values $x \in X$ and $y \in Y$.

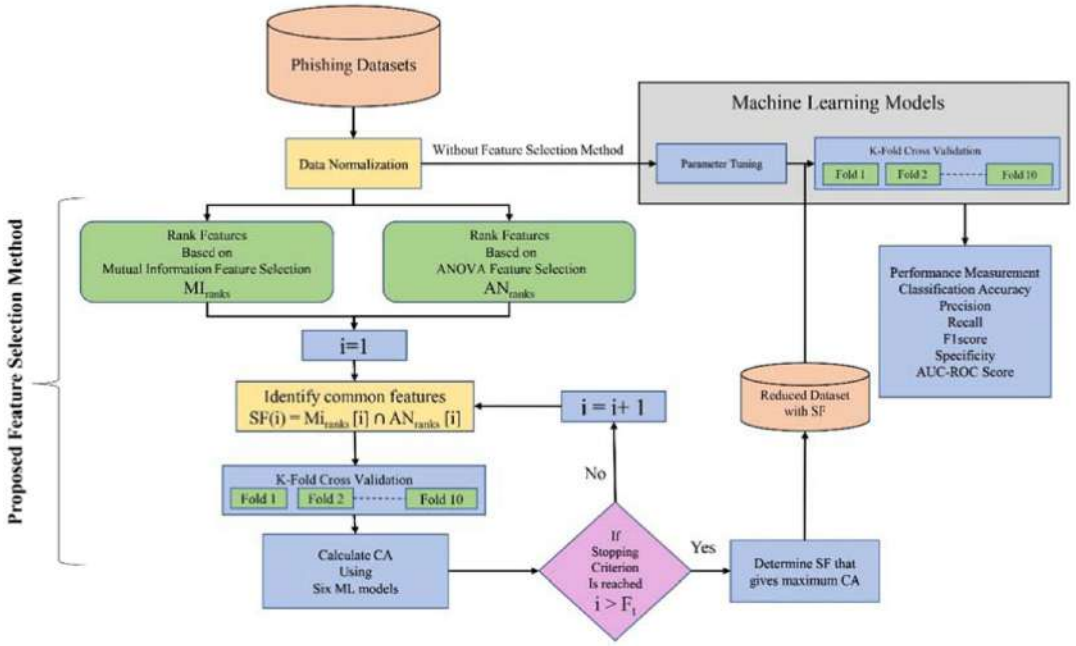


**Fig. 1.** Workflow of the proposed model.

**Algorithm 1:** Proposed MI-AN Feature Selection method for any one dataset

**Input:** D: Dataset, Ft: Total number of features, n: Number of samples(also called instances or patterns)
**Output:** SF: Reduced Feature Set
**Step 1:** SF={}, i=0
**Step 2**: Compute the rank of each of the Ft features of D using MI and AN separately, and place the ranked features in two lists, namely MIranks and ANranks, respectively, in ascending order. Thus, the highest - ranked feature will be placed first in either of the lists, followed by the remaining low-ranked features.
**Step 3:** i=i+1
**Step 4:** Determine the common features from both the lists MIranks, and ANranks from among the first i features. Add these features to SF{i}.
**Step 5:** Calculate CA for six ML models using the features of SF{i} obtained in Step 4.
**Step 6:** If i = Ft then go to Step 7.
Else go to Step 3.
**Step 7:** Determine the SF{i} that produces the maximum CA. Assign features of SF{i} to SF.
**Step 8:** Calculate all performance evaluation measurements, namely precision, recall, F1 score, specificity, and ROC curve, with the features of SF.

### *ANOVA for feature selection*

ANOVA (Analysis of Variance) is a statistical technique used to determine whether the means of different groups differ significantly from one another. It identifies if there is a statistically significant difference between the means across groups. In feature selection, ANOVA is frequently applied to evaluate the relationship between a continuous target variable and a categorical feature. For classification problems with a categorical target, ANOVA can be employed to select the most relevant features. Features with a higher F-statistic are considered more important, as the F-statistic compares the variation between groups to the variation within groups. The primary objective of ANOVA is to maximize the variation between groups while minimizing the variance within groups. Thus, a higher F-statistic indicates that the feature has stronger discriminating power, making it an excellent candidate for selection.

The F-statistic in ANOVA is calculated as:

$$F\text{-}Statistic = \frac{\frac{1}{k-1}\sum_{i=1}^{k} n_i\left(Y_i' - Y'\right)^2}{\frac{1}{N-k}\sum_{i=1}^{k}\sum_{j=1}^{n_i}\left(Y_{ij} - Y_i'\right)^2}, \quad (2)$$

where $k$ is the number of groups (categories of the feature), $n_i$ is the number of samples in the group $i$,

**Table 1.** Summary of hyperparameters and optimal values across models.

| Model | Parameter Name | Values Tested | Optimal Values for Dataset 1 | Optimal Values for Dataset 2 | Optimal Values for Dataset 3 |
|---|---|---|---|---|---|
| DT | max_depth | None, 10, 20, 30, 40, 50 | 50 | 40 | 30 |
| DT | min_samples_split | 2, 5, 10 | 2 | 2 | 2 |
| DT | min_samples_leaf | 1, 2, 4 / 1 | 1 | 1 | 2 |
| DT | criterion | gini, entropy | entropy | entropy | entropy |
| K-NN | n_neighbors | 3, 5, 7, 9, 11, 13, 15, 17, 19 | 7 | 7 | 5 |
| K-NN | weights | uniform, distance | distance | uniform | distance |
| K-NN | algorithm | auto, ball_tree, kd_tree, brute | auto | auto | brute |
| K-NN | leaf_size | 10, 20, 30, 40, 50, 60, 70, 80, 90 | 8 | 80 | 10 |
| K-NN | p | 1, 2 | 1 | 1 | 1 |
| RF | n_estimators | 5, 7,9, 10, 20, 30, 40, 50, 100, 150, 200, 250 | 7 | 70 | 150 |
| RF | min_samples_split | 2, 5, 7, 9, 10 | 9 | 5 | 2 |
| RF | min_samples_leaf | 1, 2, 3, 4 | 1 | 2 | 1 |
| RF | max_features | sqrt, log2, None | sqrt | sqrt | sqrt |
| RF | 'bootstrap' | True, False | True | True | False |
| RF | max_depth | None, 10, 20, 30, 40, 50 | 20 | 10 | 20 |
| MLP | max_iter | 100, 500, 1000, 2000, 5000 | 1000 | 1000 | 1000 |
| MLP | hidden_layer_sizes | (50,), (100,), (50, 50), (100, 50), (100, 100) | 50, | 50,50 | 100, |
| MLP | alpha | 0.0001, 0.001, 0.01, 0.1 | 0.001 | 0.1 | 0.001 |
| MLP | activation | tanh, relu | relu | relu | relu |
| MLP | learning_rate | constant, adaptive | constant | constant | constant |
| LR | C | 0.1, 1.0 | 1.0 | 1.0 | 1.0 |
| LR | penalty | l1, l2, elasticnet, none | none | elasticnet | L2 |
| LR | solver | saga, liblinear | saga | saga | liblinear |
| NB | Gaussian Naive Bayes | | | | |

**Table 2.** Summary of key properties of the phishing websites dataset 1 used in the experiments.

| Properties | Description |
|---|---|
| Number of features | Total: 30 features. There are 12 features that are based on the address bar, 6 are based on abnormalities, 5 are based on HTML and JavaScript, and 7 are based on domain categories. |
| Number of data or records | 11055 (4898Phishing websites + 6157Legitimatewebsites data) |
| Percentage of phishing websites data | 44 % |
| Percentage of legitimate websites data | 56 % |
| Classes | 2 (Legitimate and phishing website) |

$Y_i'$ is the mean of the group $i$, $Y'$ is the overall mean of the target variable, $Y_{ij}$ is the value of the target for the $j^{th}$-sample in the $i^{th}$-group, $N$ is the total number of samples.

## Experiment

### Experiment setting

A Windows 10 machine with an Intel (R) Core i3 processor (2.30 GHz), 8 GB RAM, and a 500 GB SSD was used to evaluate the proposed model, implemented in Python 3.12.1. The evaluation was performed using the labels provided in the data source and six classification models, namely DT, KNN, LR, MLP, NB, and RF, with hyperparameter tuning. Table 1 presents the summary of hyperparameters and their optimal values across the models for each phishing dataset. The dataset was normalized using a standard scaler, and null values were removed prior to training. To address the imbalanced classification problem, the SMOTE algorithm was applied. The experiment employed a 10-fold cross-validation, and the resulting performance metrics, including CA, F1 score, precision, recall, specificity, and ROC curve across the 10-fold cross-validation, are shown in Tables 5 to 8.[37]

### Dataset description

In this study, three phishing datasets were used for the experiments from the UCI Machine Learning Repository,[38] Mendeley Data,[39] and the Kaggle website data. Tables 2 to 4 summarize the key properties of the phishing datasets used in the experiments and evaluation.

### Visualizing feature importance using MI and ANOVA

Figs. 2 to 7 show the feature importance of the MI and ANOVA methods across three phishing datasets. To interpret the contribution of individual features,

**Table 3.** Summary of key properties of the phishing URLs dataset 2 used in the experiments.

| Properties | Description |
|---|---|
| Number of features | Total: 87 features. There are 56 extracted from the structure and syntax of URLs, 24 extracted from the content of their corresponding pages, and 7 are extracted by querying external services. |
| Number of data or records | 11430 (5715 Phishing URLs + 5715 Legitimate URLs) |
| Percentage of phishing URLs data | 50% |
| Percentage of legitimate URLs data | 50% |
| Classes | 2 (Legitimate URLs and phishing URLs) |

**Table 4.** Summary of key properties of the phishing webpages dataset 3 used in the experiments.

| Properties | Description |
|---|---|
| Number of features | Total: 48 features |
| Number of data or records | 10000 (5000 Phishing webpages + 5000 Legitimate webpages) |
| Percentage of phishing webpages data | 50% |
| Percentage of legitimate webpages data | 50% |
| Classes | 2 (Legitimate webpages and phishing webpages) |

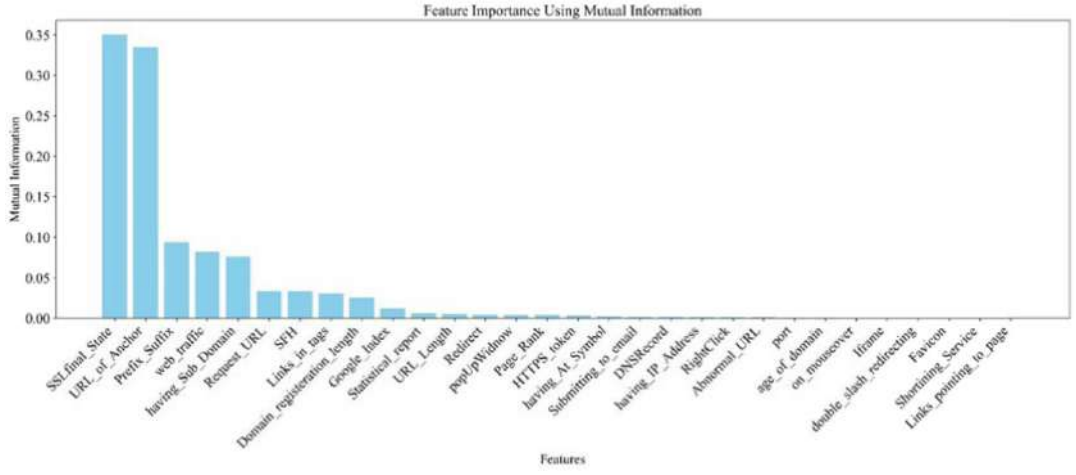


**Fig. 2.** Feature importance using MI in phishing dataset 1.

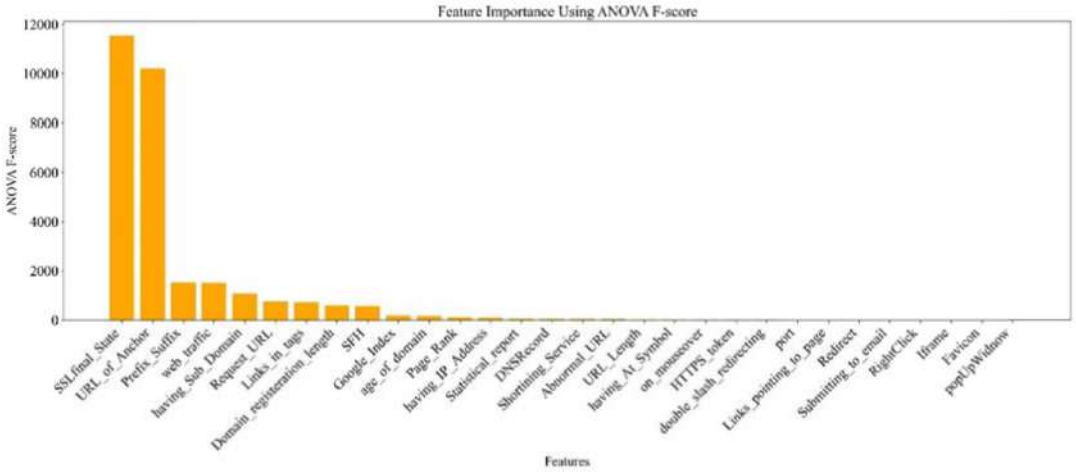


**Fig. 3.** Feature importance using ANOVA in phishing dataset 1.

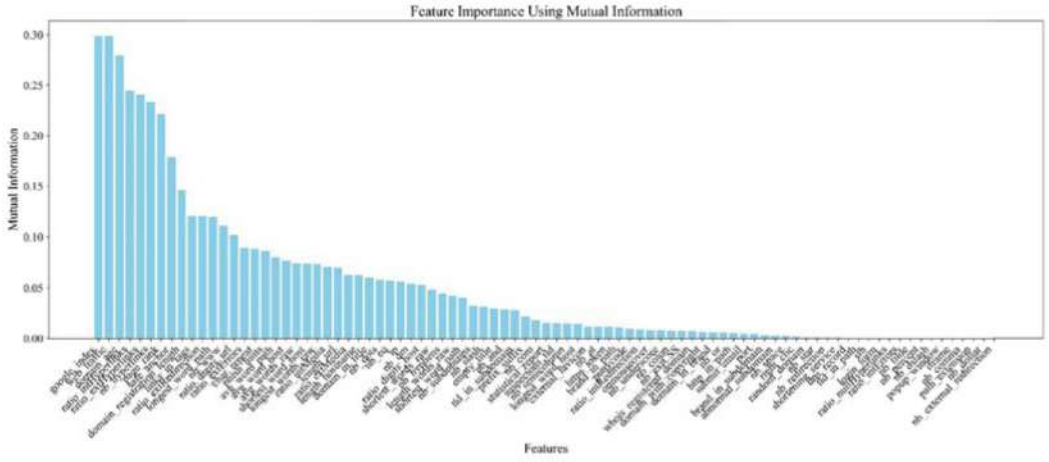


**Fig. 4.** Feature importance using MI in phishing dataset 2.

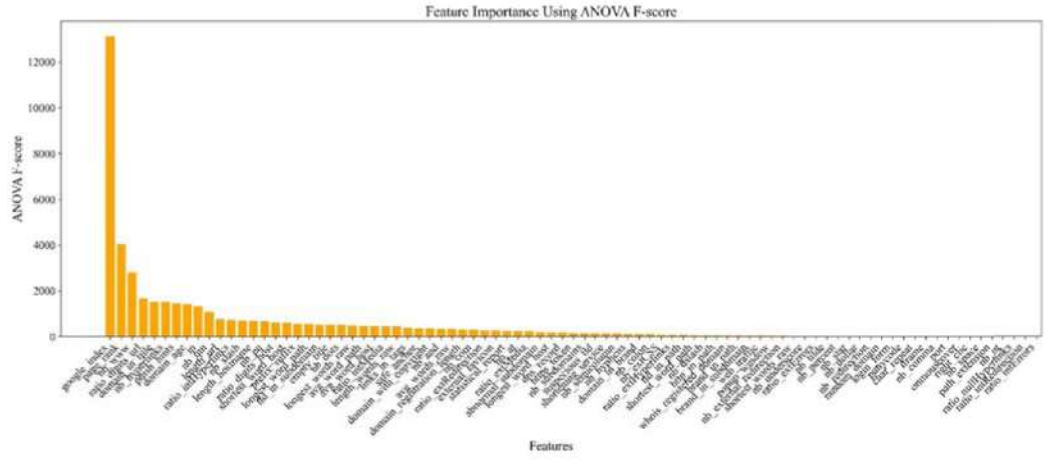


**Fig. 5.** Feature importance using ANOVA in phishing dataset 2.

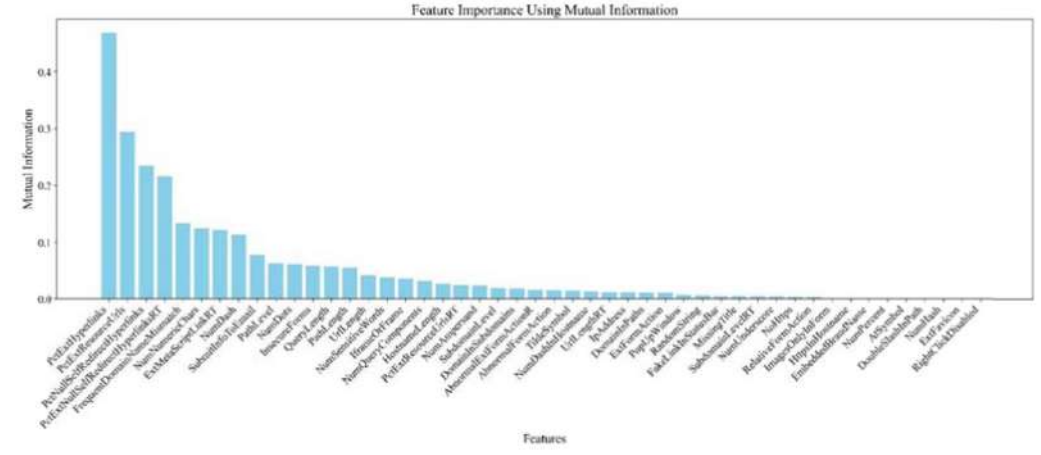


**Fig. 6.** Feature importance using MI in phishing dataset 3.

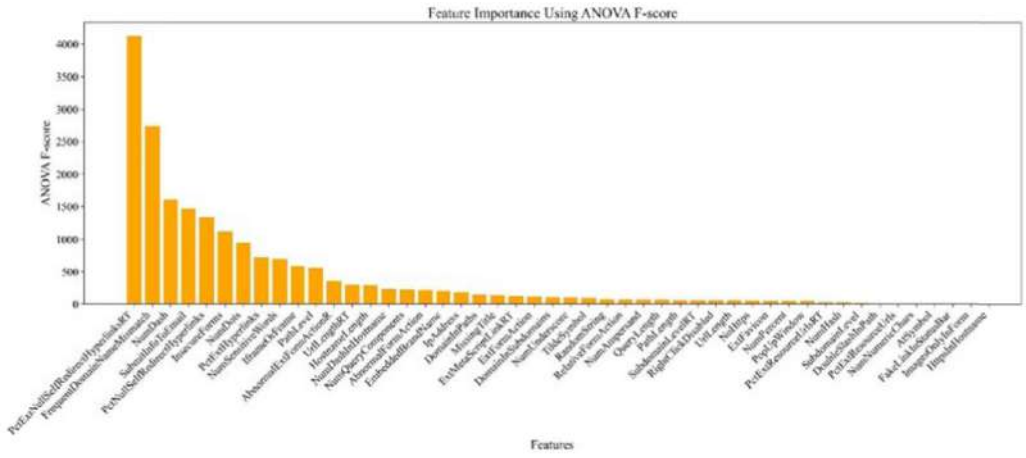


**Fig. 7.** Feature importance using ANOVA in phishing dataset 3.

we visualize feature importance using Mutual Information (MI) and ANOVA F-scores. These plots rank features by predictive power, highlighting the most influential in each dataset.

### *Machine learning models*

Hybrid feature-selection schemes have shown consistent gains in diverse domains, from Arabic sentiment analysis to network intrusion detection.[40] Motivated by these results, we adopt a simple, scalable filter–filter hybrid based on Mutual Information and ANOVA (MI–AN), and evaluate it with six widely used learners described below.

1) **Decision Tree:** A supervised machine learning approach for classification and regression purposes is the Decision Tree (DT).[16] A tree-like structure is used to represent decisions and their potential consequences, including nodes standing for decisions or attribute tests, branches for results, and leaves for final predictions or outcomes. The tree divides according to the best attribute after beginning at a root node.
2) **K-Nearest Neighbor:** The K-Nearest Neighbor (K-NN) classifies test data according to the known labels of the k neighbors using a distance metric such as Euclidean distance.[17] The class of the test patterns is determined by a majority of the class labels that the training patterns predict.
3) **Logistic Regression:** The supervised machine learning approach known as Logistic Regression (LR)[18] is mostly used for binary classification tasks, while it may also be utilized for multi-class issues. The logistic (sigmoid) function is used to map input features to a probability score, which is subsequently used to estimate the probability that an event will occur. Since the output falls between 0 and 1, one can use it to predict "yes/no" or "true/false" outcomes.
4) **Multilayer Perceptron:** A particular kind of artificial neural network (ANN) called a Multilayer Perceptron (MLP)[19] has been developed for supervised learning tasks like regression and classification. It consists of an input layer that takes in feature data, one or more hidden layers that use weighted connections and activation functions (such as ReLU, Sigmoid, and Tanh) to compute and extract patterns, and an output layer that generates the final prediction. Weighted edges are used to link neurons in one layer to neurons in the next layer; each connection has a bias. Through repeated weight adjustments during training, MLP minimizes prediction errors by combining gradient descent with the backpropagation technique.
5) **Naive Bayes:** The supervised machine learning algorithm known as Naive Bayes (NB) is based on Bayes' Theorem and is mainly used for classification tasks. Considering the class label, it makes the often-unrealistic assumption that all features are independent of one another, which significantly simplifies computation.[20]
6) **Random Forest:** The popular machine learning algorithm Random Forest (RF) is part of the supervised learning approach. It is used for machine learning issues that include both regression and classification. The basis for it comes from the concept of ensemble learning. A single conclusion is obtained by combining the results of several decision trees. Because of its simplicity and adaptability, it is becoming more and more popular. The random forest uses forecasts from each tree to anticipate the final outcome, and the majority votes in favor of those predictions.[21]

### *Performance evaluation*

In this study, we evaluated performance in many experiments. We calculated the confusion matrix to determine the true positive (TP), false positive (FP), true negative (TN), and false negative (FN). The definitions and formulae to evaluate the model's performance, like CA, precision, recall, f1 score, and specificity, are provided below.

1) **Classification Accuracy:** Classification Accuracy (CA) is defined as the ratio of correctly predicted events to all dataset observations. Eq. (3) was used to compute CA. The total TP and TN divided by the total FP, FN, TN, and TP numbers provides the CA as follows:

$$CA = \frac{TP + TN}{TP + TN + FP + FN} \quad (3)$$

2) **Precision:** Precision is defined as the percentage of correctly predicted positive observations to all expected positives. Eq. (4) was used to calculate precision. The precision is obtained by dividing the total number of TP by the total number of TP and FP as follows:

$$Precision = \frac{TP}{TP + FP} \quad (4)$$

3) **Recall:** Recall is defined as the percentage of correctly predicted positive observations to all of the dataset's actual positive observations. Eq. (5) was used to calculate recall. The recall is obtained by dividing the total number of TP by the total number of TP and FN as follows:

$$Recall = \frac{TP}{TP + FN} \quad (5)$$

4) **F1 Score:** The F1 score is the harmonic mean of recall and precision. It offers a harmony between recall and precision. Eq. (6) was used to calculate the F1 score:

$$F1\ score = \frac{2 * Recall * Precision}{Recall + Precision} \quad (6)$$

5) **Specificity:** Specificity measures the model's ability to accurately detect negative samples. Specificity is calculated by dividing TN by the sum of FP and TN, as defined in Eq. (7) below:

$$Specificity = \frac{TN}{TN + FP} \quad (7)$$

6) **AUC-ROC:** The AUC-ROC curve is a significant evaluation metric for binary classification problems. How well a model differentiates between the two classes is shown by the area under the curve. A model is considered excellent if its AUC is around 1, which signifies that it has a good measure of separability.

## Results and discussion

Table 5 presents the experimental results of the proposed model for phishing dataset 1. The first column of Table 5 presents the selected number of features, and the second to seventh columns present the CA of the DT, K-NN, LR, MLP, NB, and RF models, respectively. The last row of Table 5 presents the CA with all features or the CA before feature selection (base CA).

In Table 5, it clearly shows in the DT model we achieve a base CA of 88.86%, whereas we achieve a CA of 96.74% with 23 SF; in the K-NN model we achieve 96.46% of base CA, whereas we achieve 96.85% of CA with 25 SF; in the LR model we achieve 91.99% of base CA, whereas we achieve 92.18% of CA with 25 SF; in the MLP model we achieve 96.81% of base CA, whereas we achieve 97.07% of CA with 25 SF; in the NB model we achieve 63.92% of base CA, whereas we achieve 90.08% of CA with only 2 SF; and in the RF model we achieve 93.88% of base CA, whereas we achieve 96.87% of CA with 25 SF. Compared to all models, the MLP model performed better in terms of CA, and compared to the base

**Table 5.** Classification accuracy achieved by the proposed method with different ML models for phishing dataset 1 (in %).

| SF | DT | K-NN | LR | MLP | NB | RF |
|---|---|---|---|---|---|---|
| 1 | 88.40 | 88.40 | 88.40 | 88.40 | 88.40 | 88.40 |
| 2 | 90.86 | 90.17 | 90.47 | 90.86 | 90.08 | 90.86 |
| 3 | 91.27 | 91.27 | 90.90 | 91.45 | 61.90 | 91.27 |
| 4 | 91.79 | 90.46 | 90.88 | 91.46 | 61.90 | 91.79 |
| 5 | 93.33 | 91.56 | 90.47 | 93.11 | 61.90 | 93.32 |
| 7 | 93.68 | 92.83 | 91.12 | 93.36 | 61.90 | 93.69 |
| 9 | 94.10 | 93.45 | 91.87 | 93.93 | 63.12 | 94.14 |
| 10 | 94.36 | 93.63 | 91.91 | 94.45 | 62.95 | 94.52 |
| 11 | 94.75 | 93.85 | 92.01 | 94.62 | 63.02 | 94.94 |
| 12 | 95.02 | 94.43 | 91.97 | 94.90 | 63.24 | 95.17 |
| 14 | 95.20 | 95.01 | 91.65 | 95.29 | 63.48 | 95.67 |
| 15 | 95.31 | 95.15 | 91.72 | 95.61 | 63.57 | 95.70 |
| 16 | 95.55 | 95.40 | 91.84 | 95.80 | 63.67 | 96.00 |
| 17 | 95.87 | 95.72 | 91.90 | 96.00 | 63.72 | 96.14 |
| 19 | 96.23 | 96.14 | 91.92 | 96.32 | 63.76 | 96.42 |
| 21 | 96.64 | 96.72 | 92.09 | 96.82 | 63.71 | 96.69 |
| 22 | 96.71 | 96.70 | 92.17 | 97.05 | 63.68 | 96.84 |
| 23 | 96.74 | 96.77 | 92.15 | 96.94 | 63.66 | 96.84 |
| 25 | 96.54 | 96.85 | 92.18 | 97.07 | 63.72 | 96.87 |
| 26 | 88.26 | 96.35 | 92.05 | 96.75 | 63.70 | 93.91 |
| 27 | 88.16 | 96.35 | 92.04 | 96.69 | 63.81 | 94.21 |
| 29 | 88.72 | 96.43 | 92.00 | 96.85 | 63.94 | 94.19 |
| 30 | 88.86 | 96.46 | 91.99 | 96.81 | 63.92 | 93.88 |

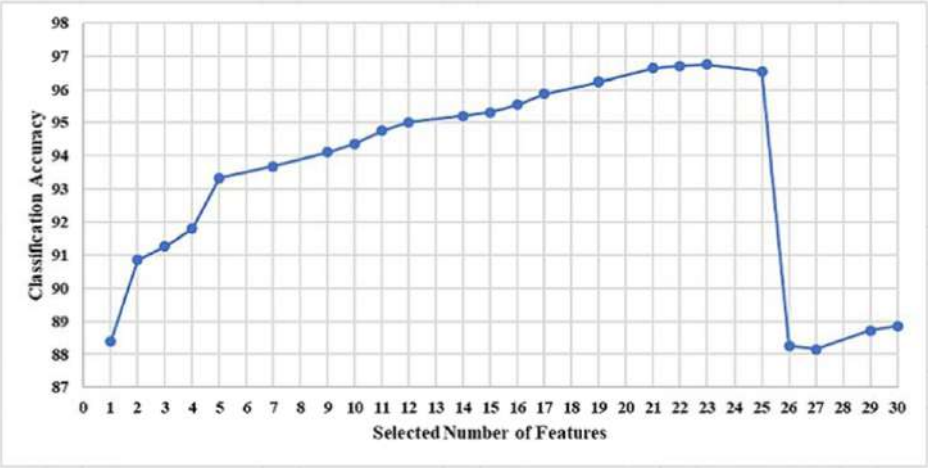


**Fig. 8.** Classification accuracy of proposed model with DT model for phishing websites dataset 1.

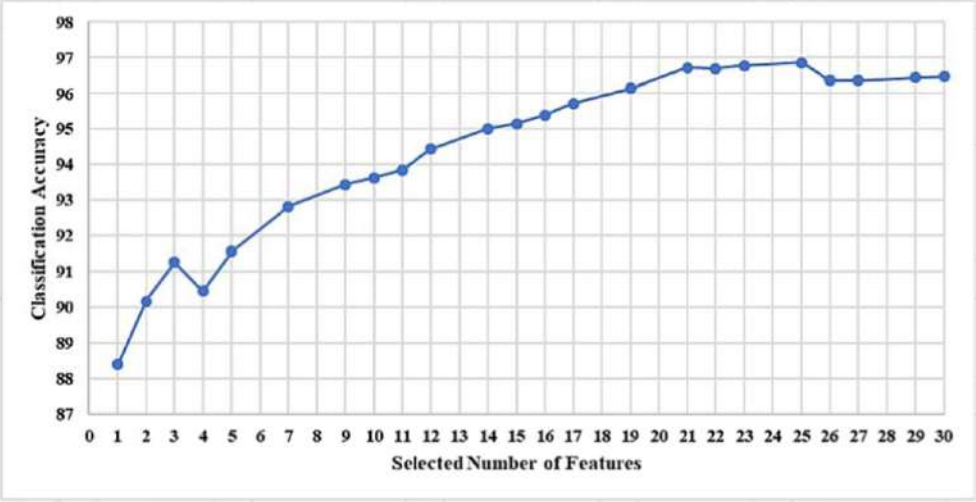


**Fig. 9.** Classification accuracy of proposed model with K_NN model for phishing websites dataset 1.

CA and selected features of CA, we achieved better performance in selected features.

Figs. 8 to 13 present the graphical representation of the results in Table 5 across different ML models. In Fig. 8, the proposed feature selection technique with 23 selected features (SF) achieves the highest classification accuracy (CA) in the DT model. Similarly, Figs. 9 to 11 and 13 show that the technique with 25 SF yields the best CA for the K-NN, LR, MLP, and RF models. In contrast, Fig. 12 demonstrates that the NB model achieves its best performance with only 2 SF, reaching about 90% accuracy with just 1–2 SF, while performance drops to around 60% as the number of SF increases from 3 to 30. This highlights the strong discriminative power of the proposed method in the NB model. Overall, when comparing Figs. 8 to 13, the MLP model shows comparatively better CA with the proposed feature selection technique.

Tables 6 and 7 present the experimental results of the proposed model on phishing datasets 2 and 3. The first column of Tables 6 and 7 represents the range of selected features (e.g., 1–10, 11–20) corresponding to their best classification accuracies. Although experiments were conducted using all features, only the most relevant ranges are presented to avoid redundancy and unnecessary length. The second to seventh columns present the CA of the DT, K-NN, LR, MLP, NB, and RF models, respectively. The last row of Tables 4 and 5 presents the CA with all features or the CA before feature selection (base CA). Figs. 8 and 9 present the graphical representation of the results in Tables 6 and 7 across the best ML models.

In Table 6, and Fig. 14, it clearly shows in the DT model we achieve a base CA of 93.66%, whereas we achieve a CA of 96.98% with 43 SF; in the K-NN model we achieve 95.68% of base CA, whereas

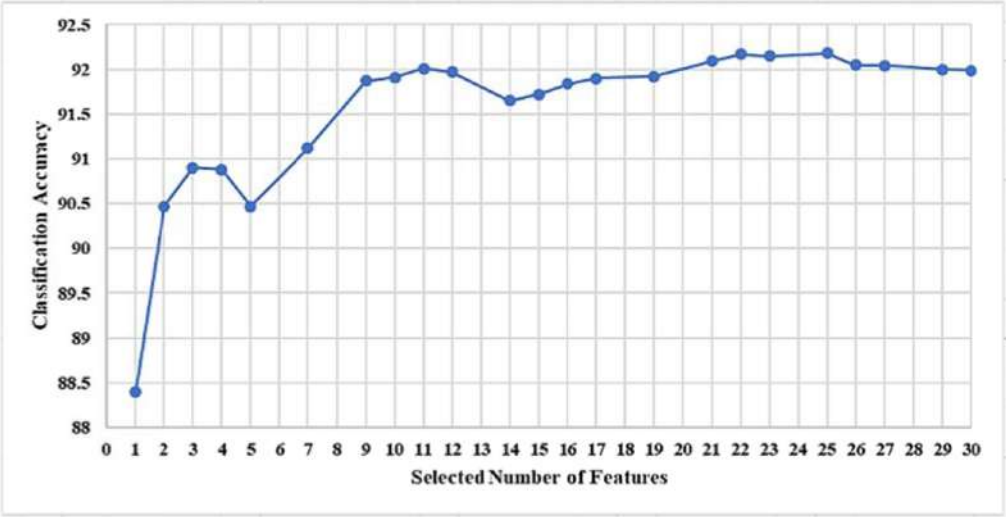


**Fig. 10.** Classification accuracy of proposed model with LR model for phishing websites dataset 1.

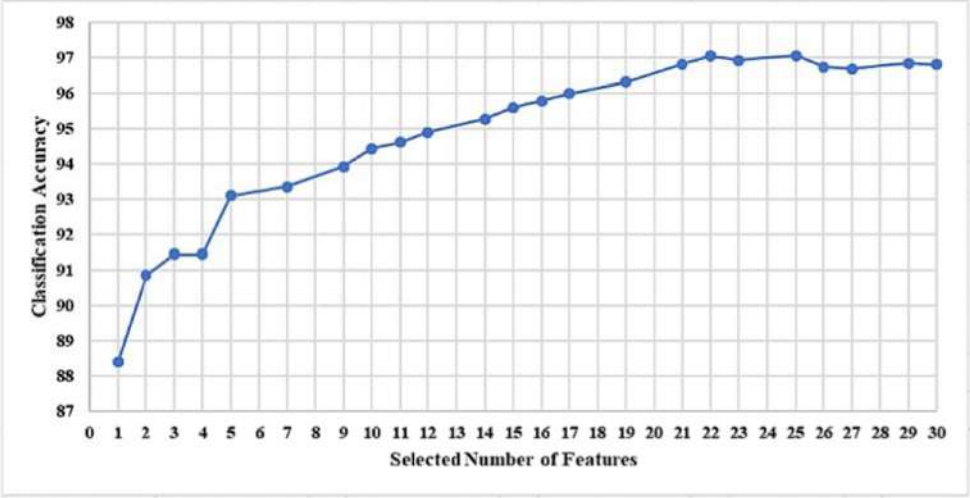


**Fig. 11.** Classification accuracy of proposed model with MLP model for phishing websites dataset 1.

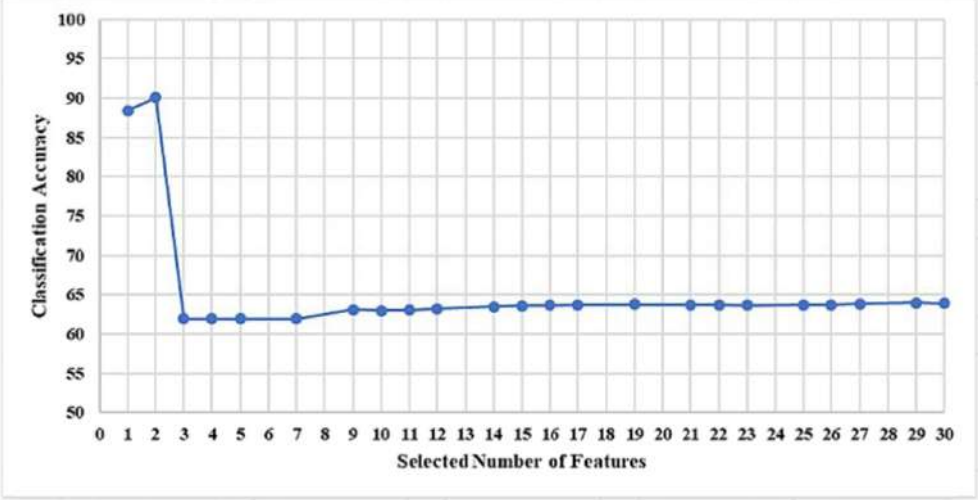


**Fig. 12.** Classification accuracy of proposed model with NB model for phishing websites dataset 1.

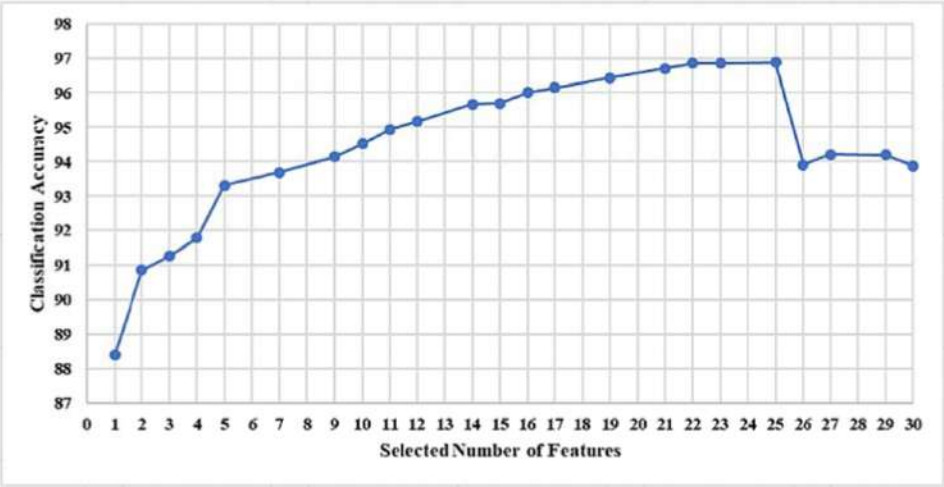


**Fig. 13.** Classification accuracy of the proposed model compared with the RF model on phishing website dataset 1.

**Table 6.** Classification accuracy achieved by the proposed method with different ML models for phishing dataset 2 (in %).

| SF(Rang) | DT | K-NN | LR | MLP | NB | RF |
|---|---|---|---|---|---|---|
| 1–10 | 92.91 | 94.74 | 92.46 | 94.35 | 91.22 | 95.38 |
| 11–20 | 93.32 | 95.67 | 93.04 | 95.21 | 86.48 | 95.82 |
| 21–30 | 93.72 | 95.69 | 93.55 | 95.47 | 81.99 | 96.07 |
| 31–40 | 93.84 | 95.84 | 94 | 95.69 | 81.14 | 96.26 |
| 41–50 | 93.98 | 95.84 | 94.13 | 95.52 | 75.55 | 96.6 |
| 51–60 | 93.82 | 95.8 | 94.28 | 95.57 | 73.56 | 96.49 |
| 61–70 | 93.88 | 95.91 | 94.38 | 95.71 | 71.92 | 96.54 |
| 71–80 | 93.96 | 95.7 | 94.43 | 95.65 | 71.44 | 96.6 |
| 81–87 | 93.93 | 95.71 | 94.47 | 95.49 | 70.19 | 96.54 |
| 87 | 93.66 | 95.68 | 94.47 | 95.65 | 68.70 | 96.55 |

**Table 7.** Classification accuracy achieved by the proposed method with different ML models for phishing dataset 3 (in %).

| SF(Rang) | DT | K-NN | LR | MLP | NB | RF |
|---|---|---|---|---|---|---|
| 1–10 | 95.55 | 94.54 | 89.32 | 95.62 | 82.56 | 96.35 |
| 11–20 | 95.74 | 95.03 | 92.32 | 96.12 | 83.26 | 97.41 |
| 21–30 | 95.72 | 94.6 | 93.06 | 96.28 | 84.21 | 97.78 |
| 31–40 | 95.9 | 94.69 | 93.67 | 96.65 | 84.12 | 97.75 |
| 41–48 | 95.96 | 95.06 | 93.74 | 96.9 | 84.49 | 97.78 |
| 48 | 95.96 | 94.99 | 93.71 | 96.8 | 83.51 | 97.75 |

we achieve 95.91% of CA with 63 SF; in the LR model we achieve 94.47% of base CA, whereas we achieve 94.47% of CA with 85 SF; in the MLP model we achieve 95.65% of base CA, whereas we achieve 95.71% of CA with 63 SF; in the NB model we achieve 68.70% of base CA, whereas we achieve 91.22% of CA with only 8 SF; and in the RF model we achieve 96.55% of base CA, whereas we achieve 96.60% of CA with 48 SF. Compared to all models, the RF model performed better in terms of CA, and compared to the base CA and selected features of CA, we achieved better performance in selected features.

Similarly In Table 7, and Fig. 15, it clearly shows in the DT model we achieve a base CA of 95.96%, whereas we achieve a CA of 95.96% with 48 SF; in the K-NN model we achieve 94.99% of base CA, whereas we achieve 95.06% of CA with 44 SF; in the LR model we achieve 93.71% of base CA, whereas we achieve 93.74% of CA with 44 SF; in the MLP model we achieve 96.8% of base CA, whereas we achieve 96.90% of CA with 42 SF; in the NB model we achieve 83.51% of base CA, whereas we achieve 84.49% of CA with 42 SF; and in the RF model we achieve 97.75% of base CA, whereas we achieve 97.78% of CA with 27 SF. Compared to all models, the RF model performed better in terms of CA, and compared to the base CA and selected features of CA, we achieved better performance in selected features.

Table 8 presents a performance evaluation of the proposed model against other ML models. The first column of Table 8 presents the datasets. The second column lists the model names. The third to Seventh columns present the CA, precision, recall, f1score, and specificity (in %) of the proposed method, and the last column presents the number of selected features in the proposed method.

For phishing dataset 1, Table 1 and Fig. 16 show that in the DT model, we achieved a CA of 96.74%, a precision of 96.52%, a recall of 96.98%, a specificity of 96.51%, and an F1 score of 96.52% with 23 SF. In the K-NN model, we achieved CA, precision, recall, f1 score, and specificity of 96.85%, 96.62%, 97.09%, 96.86%, and 96.61%, respectively, with 25 SF. In the LR model, we achieved CA, precision, recall, f1score, and specificity of 92.18%, 91.34%, 93.19%, 92.26%, and 91.16%, respectively, with 25 SF. In the MLP model, we achieved CA, precision, recall, f1 score, and specificity of 97.07%, 96.76%, 97.40%,

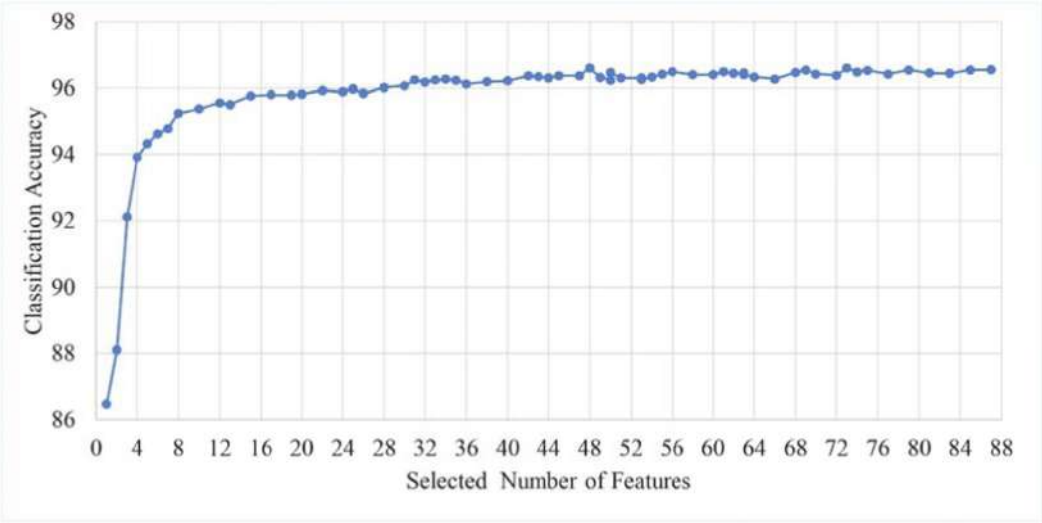


**Fig. 14.** Classification accuracy of proposed model with RF model for phishing dataset 2.

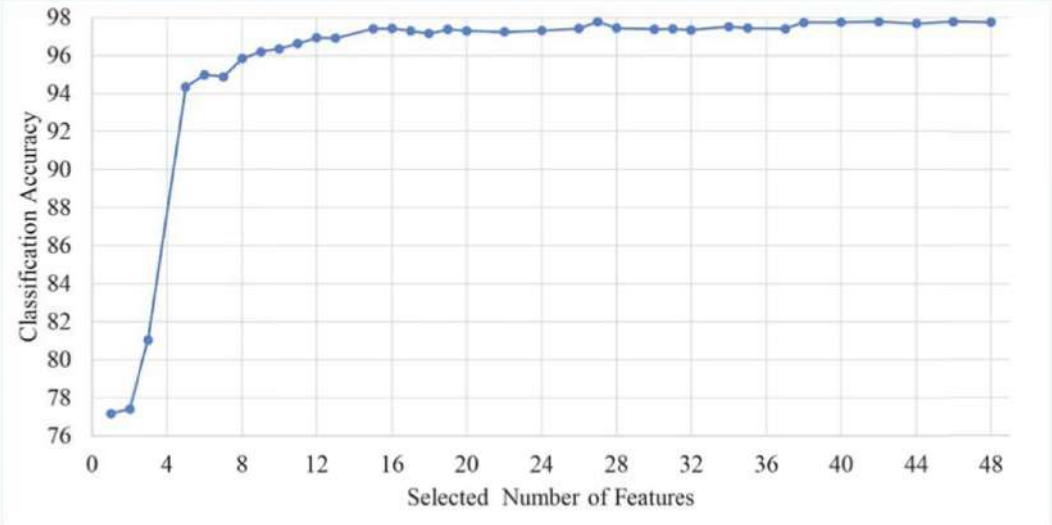


**Fig. 15.** Classification accuracy of the proposed model with the RF model for phishing dataset 3.

97.08%, and 96.74%, respectively, with 25 SF. In the NB model, we achieved CA, precision, recall, f1 score, and specificity of 90.08%, 89.48%, 90.86%, 90.16%, and 89.31%, respectively, with only 2 SF. In the RF model, we achieved CA, precision, recall, f1score, and specificity of 96.87%, 96.29%, 97.48%, 96.88%, and 96.25%, respectively, with 25 SF. Comparison of all models: the MLP model performed better with each performance evaluation (except recall). Similarly, for phishing dataset 2, the RF model we achieved CA, precision, recall, f1score, and specificity of 96.6%, 96.6%, 96.59%, 96.6% 96.61%, respectively, with 48 SF. Comparison of all models for phishing dataset 2: the RF model performed better with each performance evaluation. Lastly, for the phishing dataset 3, the RF model performed better compared to all other models. We achieved CA, precision, recall, f1score, and specificity of 97.78%, 98.20%, 97.34%, 97.77%, 98.22%, respectively, with 27 SF.

Overall, across the three phishing datasets, model performance varied based on data features. On Dataset 1 in Fig. 16, the MLP attained the highest accuracy across all methods; however, on Datasets 2 and 3 in Figs. 17 and 18, the RF outperformed due to its effective handling of feature interactions and reduced overfitting. Compared to the base accuracy using all features, the proposed MI-AN feature selection method significantly improved performance across all datasets, indicating that the proposed MI-AN effectively eliminates irrelevant features and improves model generalization. The proposed MI-AN approach

**Table 8.** Performance evaluation of proposed model with different ML models.

| Datasets | ML models | CA | Precision | Recall | F1score | Specificity | SF |
|---|---|---|---|---|---|---|---|
| Phishing Dataset 1 | DT | 96.74% | 96.52% | 96.98% | 96.75% | 96.51% | 23 |
| | K-NN | 96.85% | 96.62% | 97.09% | 96.86% | 96.61% | 25 |
| | LR | 92.18% | 91.34% | 93.19% | 92.26% | 91.16% | 25 |
| | MLP | 97.07% | 96.76% | 97.40% | 97.08% | 96.74% | 25 |
| | NB | 90.08% | 89.48% | 90.86% | 90.16% | 89.31% | 2 |
| | RF | 96.87% | 96.29% | 97.48% | 96.88% | 96.25% | 25 |
| Phishing Dataset 2 | DT | 93.98% | 93.78% | 94.21% | 93.99% | 93.75% | 43 |
| | K-NN | 95.91% | 96.21% | 95.59% | 95.9% | 96.24% | 63 |
| | LR | 94.47% | 94.83% | 94.07% | 94.45% | 94.87% | 85 |
| | MLP | 95.71% | 95.63% | 95.8% | 95.72% | 95.63% | 63 |
| | NB | 91.22% | 91.19% | 91.27% | 91.23% | 91.18% | 8 |
| | RF | 96.6% | 96.6% | 96.59% | 96.6% | 96.61% | 48 |
| Phishing Dataset 3 | DT | 95.96% | 96.59% | 95.28% | 95.93% | 96.64% | 48 |
| | K-NN | 95.06% | 94.07% | 96.18% | 95.11% | 93.94% | 44 |
| | LR | 93.74% | 93.02% | 94.58% | 93.79% | 92.90% | 44 |
| | MLP | 96.90% | 97.43% | 96.34% | 96.88% | 97.46% | 42 |
| | NB | 84.49% | 93.74% | 73.92% | 82.66% | 95.06% | 42 |
| | RF | 97.78% | 98.20% | 97.34% | 97.77% | 98.22% | 27 |

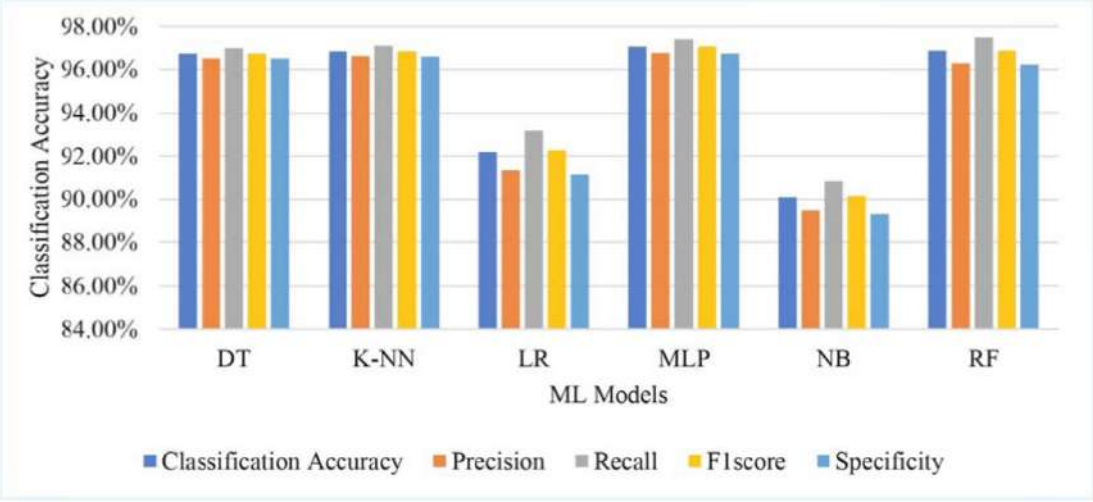


**Fig. 16.** Performance evaluation of proposed model with different ML models for dataset 1.

demonstrates robustness and adaptability, achieving outstanding results across various phishing datasets.

Figs. 9 to 14 illustrate the receiver operating characteristic (ROC) curves of the ML models enhanced by the proposed method for phishing dataset 1, providing an evaluation of its effectiveness. The ROC curves depict the trade-off between false positives (costs) and true positives (benefits), thereby establishing the quality of the proposed method.

The ROC curve of the proposed MI-AN method, which is commonly utilized in evaluating a classification model's effectiveness, is shown in Figs. 19 to 24. An example of a model's performance is a figure that shows the true positive rate on the Y-axis and the false positive rate on the X-axis. It also includes the ROC curve of DT is 0.96, K-NN is 0.97, LR is 0.92, MLP is 0.97, NB is 0.90, and RF is 0.97. High true positive rates and low false positive rates are characteristics of an ideal model; the ROC curve demonstrates how the two relate to one another. Additionally, Figs. 19 to 24 show the ROC curve for each ML model with the MI-AN feature selection method.

Similarly, Figs. 25 and 26 show the ROC curves for the best-performing ML models on Phishing Datasets 2 and 3. Although ROC curves were plotted for all models, only the best-performing ones are presented to avoid redundancy and unnecessary repetition. For Dataset 2, the proposed MI-AN method enhanced the model's performance, with the Random Forest (RF) achieving an AUC–ROC value of 0.97, indicating excellent classification capability with a high true positive rate and a very low false positive rate. Likewise, for Dataset 3, the MI-AN model achieved an AUC–ROC value of 0.95 for the RF classifier,

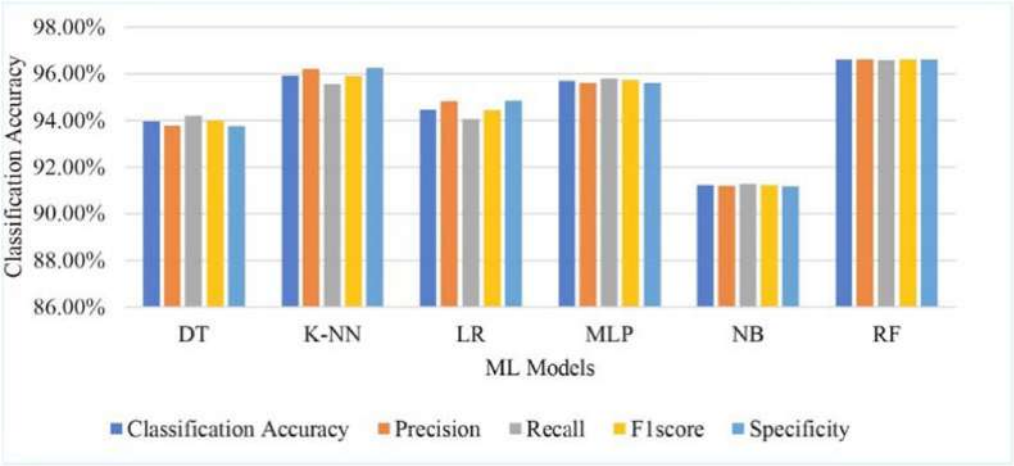


**Fig. 17.** Performance evaluation of proposed model with different ML models for dataset 2.

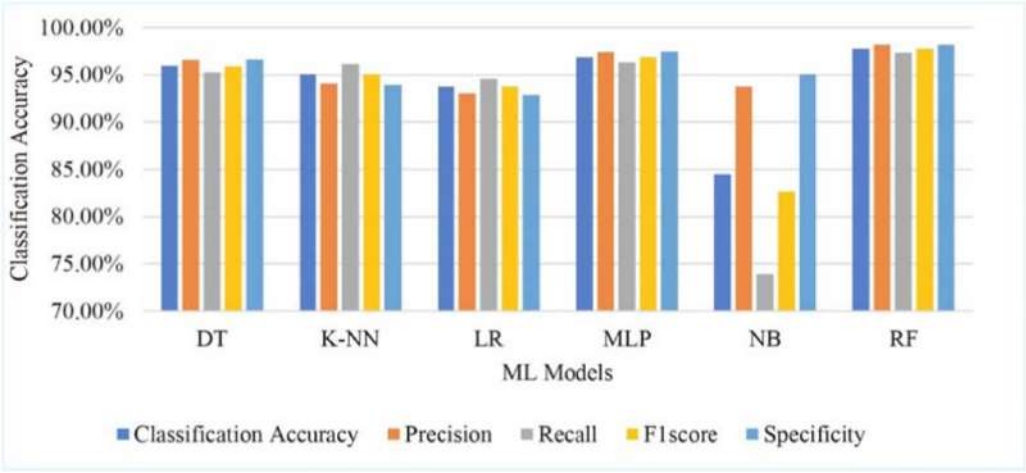


**Fig. 18.** Performance evaluation of proposed model with different ML models for dataset 3.

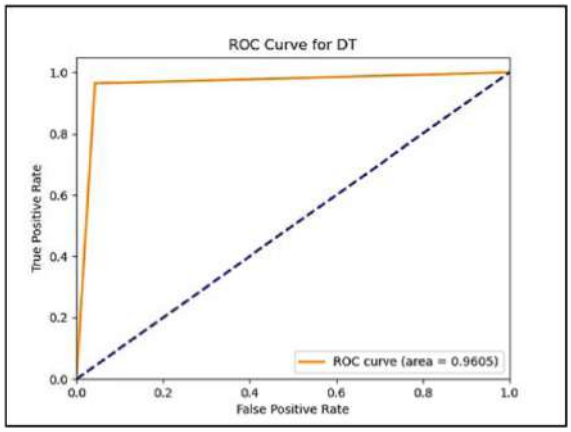


**Fig. 19.** ROC curve of DT model using the proposed MI_AN method for phishing dataset 1.

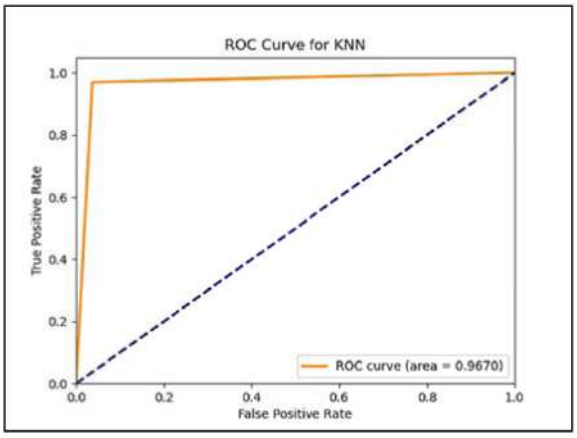


**Fig. 20.** ROC curve of K-NN model using the proposed MI_AN method for phishing dataset 1.

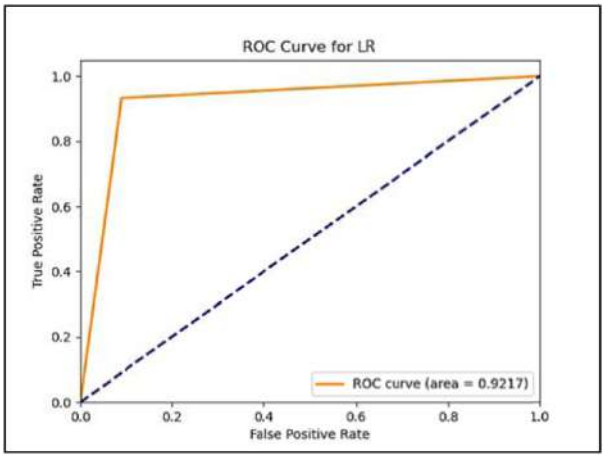


**Fig. 21.** ROC curve of LR model using the proposed MI_AN method for phishing dataset 1.

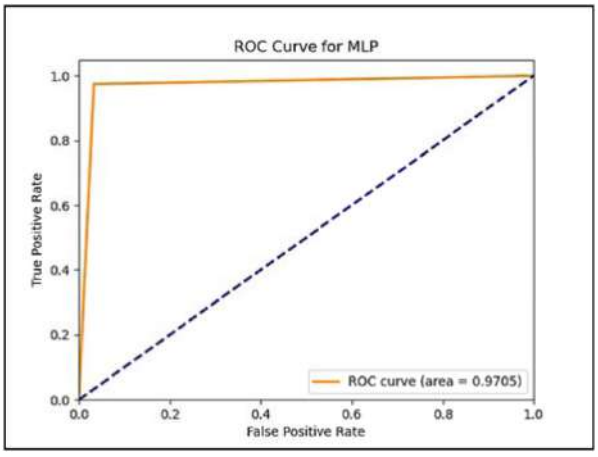


**Fig. 22.** ROC curve of MLP model using the proposed MI_AN method for phishing dataset 1.

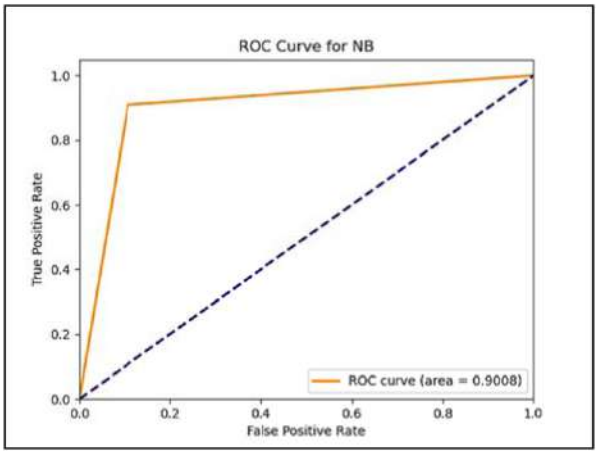


**Fig. 23.** ROC curve of NB model using the proposed MI_AN method for phishing dataset 1.

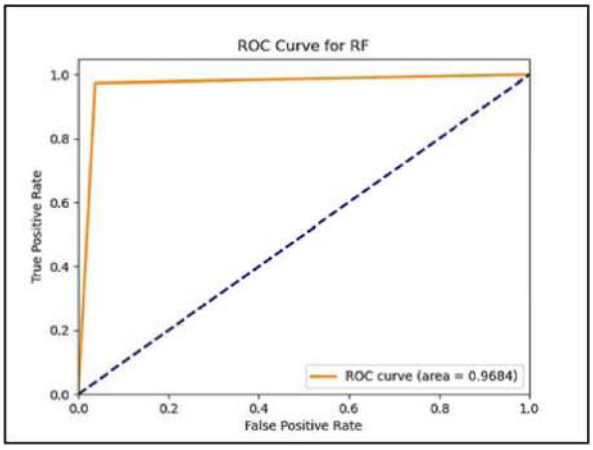


**Fig. 24.** ROC curve of RF model using the proposed MI_AN method for phishing dataset 1.

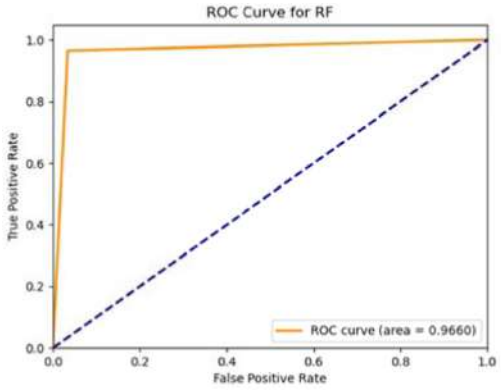


**Fig. 25.** ROC curve of RF model using the proposed MI_AN method for phishing dataset 2.

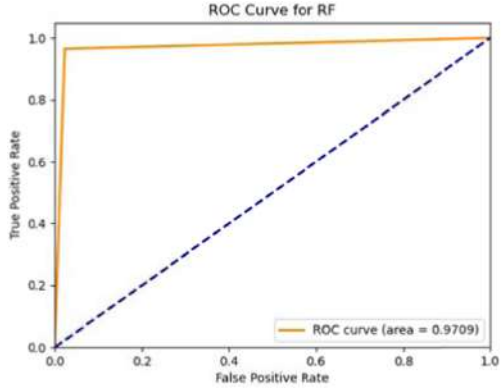


**Fig. 26.** ROC curve of RF model using the proposed MI_AN method for phishing dataset 3.

demonstrating strong discriminative power and reliable detection of phishing instances. These high AUC values confirm the robustness and effectiveness of the proposed approach in improving phishing detection accuracy across multiple datasets.

**Table 9.** Comparison of elapsed time and memory usage for machine learning models using all and selected features across phishing datasets.

| | | All Features | | Selected Feature | |
|---|---|---|---|---|---|
| Datasets | Models | Elapsed Time (s) | Memory Uses (MB) | Elapsed Time (s) | Memory Uses (MB) |
| Phishing Dataset 1 | DT | 0.50 | 0.41 | 0.40 | 3.21 |
| | KNN | 3.40 | 0.18 | 2.95 | 3.66 |
| | LR | 0.41 | 25.96 | 0.38 | 18.75 |
| | MLP | 149.60 | 0.57 | 157.63 | 3.91 |
| | NB | 0.21 | 0.15 | 0.08 | 0.02 |
| | RF | 7.15 | 1.54 | 6.72 | 2.95 |
| Phishing Dataset 2 | DT | 2.70 | 2.75 | 1.93 | 0.4 |
| | KNN | 9.27 | 4.8 | 6.39 | 2.07 |
| | LR | 1.25 | 72.89 | 1.27 | 9.08 |
| | MLP | 171.14 | 5.26 | 164.63 | 1.18 |
| | NB | 0.35 | 1.46 | 0.14 | 2.43 |
| | RF | 20.33 | 2.15 | 17.13 | 1.69 |
| Phishing Dataset 3 | DT | 1.09 | 0.44 | 1.09 | 0.44 |
| | KNN | 3.87 | 1.77 | 3.68 | 3.83 |
| | LR | 0.62 | 34.92 | 0.61 | 3.27 |
| | MLP | 143.09 | 2.14 | 131.21 | 3.62 |
| | NB | 0.21 | 0.14 | 0.26 | 3.5 |
| | RF | 9.05 | 1.58 | 7.80 | 0.29 |

Table 9 presents the computational efficiency of various ML models evaluated on three phishing datasets. The models considered include DT, K-NN, LR, MLP, NB, and RF. For each dataset, we report the total elapsed time (training + prediction) in seconds and the memory usage in megabytes (MB), and compare performance with all features versus selected features obtained using the proposed MI-AN feature selection method.

Elapsed time represents the total duration for a model to complete training and prediction, measured using 10-fold cross-validation. The start time was recorded immediately before model training, and the end time after cross-validated predictions, capturing the computational cost of model execution. Across all datasets, simpler models such as DT and NB consistently require less time, ranging from 0.21 s to 2.70 s with all features, whereas more complex models such as MLP and RF require significantly more time, often exceeding 100 s for MLP and up to 20 s for RF. After feature selection, elapsed time generally decreases, particularly for KNN, DT, and RF. For example, DT on Dataset 2 decreased from 2.70 s to 1.93 s, and KNN from 9.27 s to 6.39 s. MLP exhibits slightly variable behavior, with elapsed time decreasing for some datasets but increasing slightly for Dataset 1, likely due to overhead from smaller feature subsets interacting with the network architecture.

Memory usage indicates the RAM consumed during model training and prediction, calculated as the difference in process memory before and after execution and reported in MB. Memory trends vary across models and datasets. For LR, feature selection markedly reduces memory consumption (e.g., Dataset 2 from 72.89 MB to 9.08 MB), highlighting the efficiency benefits of using fewer features. Other models, such as DT, KNN, and NB, exhibit modest increases in memory usage after feature selection, which can be attributed to the internal handling of resampled and scaled subsets during SMOTE preprocessing. Complex models such as MLPs and RFs generally maintain moderate memory usage, demonstrating that feature selection can optimize computational resources without compromising the models' ability to process high-dimensional data.

Overall, the results across the four datasets indicate that the MI-AN feature selection method successfully balances predictive performance with computational efficiency. By reducing the number of irrelevant or redundant features, models generally execute faster and consume less memory while maintaining accuracy. These findings highlight the practical utility of MI-AN for real-time phishing detection systems and reinforce the importance of evaluating computational efficiency alongside predictive performance.

Table 10 shows the comparison of the proposed method with recent research work. In Table 5, the highest CA from recent research work has been recorded. All the papers have used the same dataset, which consists of 30 features and 11,055 samples.

The comparison and analysis of our proposed model with related research that is currently available in the literature are shown in Table 10. This comparison and analysis show that the proposed model performs better than the other recent research work. Overall, the proposed model provides computational efficiency and robustness when it comes to identifying legitimate and phishing websites.

**Table 10.** Comparison of proposed method with recent research work.

| Author(s) | Year | Model | CA |
|---|---|---|---|
| Babagoli et al | 2019 | Harmony search | 92.80% |
| Chiew et al. | 2019 | RF | 96.17% |
| Taha | 2021 | Intelligent Ensemble Learning Approach | 95% |
| Sharma et al. | 2022 | S-shaped, V-shaped transfer function, K-NN | 97.04% |
| Siddiq et al. | 2022 | Standard neural network | 94.84 |
| Fadheel et al. | 2022 | SVM | 94% |
| Karim et al | 2022 | XGBoost | 97.04% |
| Davoudi and Yari | 2023 | Ensemble model with GA | 96.04% |
| Singh et al. | 2023 | XGB-PSO model | 94.69% |
| Ali and Jain | 2023 | RF | 95.2% |
| Medelbekov et al. | 2023 | RF | 96.7 % |
| Alazaidah et al. | 2024 | RF | 96.58% |
| Taha et al. | 2024 | RF | 96.89% |
| Patel et al. | 2025 | Proposed model for dataset 1 | 97.07% |
| | | Proposed model for dataset 2 | 96.60% |
| | | Proposed model for dataset 3 | 97.78% |

The MI-AN–based phishing detection framework enhances accuracy across multiple datasets but also entails ethical and security considerations. The system must guard against adversarial manipulation of website features and ensure low false-positive rates to avoid misclassifying legitimate sites. Continuous model evaluation, explainability, and adversarial resilience are crucial to maintaining ethical responsibility and user trust in practical deployment.

## Conclusion

Phishing websites attempt to steal identities, obtain credentials, and commit financial fraud. It is important to inform and educate users about these kinds of assaults and implement protections like email filters, browser filters, and multi-factor authentication. To prevent harming themselves, users need to recognize the obvious characteristics of phishing websites and keep to security best practices. The main objective of this study is to build an accurate and computationally efficient model that can classify between phishing and legitimate websites based on data.

The proposed study develops a novel hybrid supervised feature selection method for individual models that more accurately and efficiently determines legitimate websites from phishing websites. In this study, we proposed the MI-AN feature selection method to reduce or optimize the features from the datasets of three phishing datasets and utilize the most significant relevant features to create a computationally efficient model for the highly accurate classification of phishing and legitimate websites. The mutual information and ANOVA feature selection methods are combined in the proposed hybrid MI-AN method. In the MI-AN feature selection method, the multi-layer perceptron performs better than other models on dataset 1. For datasets 2 and 3, random forest models perform better in terms of classification accuracy. It is also compared with previously published papers and performs better in terms of classification accuracy.

In the future, we will use ensemble models and novel feature selection methods to construct a robust model based on ML and deep learning for identifying and classifying harmful URLs and classifying related attacks, such as phishing emails and websites. Furthermore, other phishing websites' datasets should be used to validate and evaluate the proposed MI-AN feature selection method.

## Acknowledgment

The authors are thankful to the referees and the editor for their comments and suggestions, which substantially improve the initial version of the paper.

## Authors' declaration

- Conflicts of Interest: None.
- We hereby confirm that all the Figures and Tables in the manuscript are ours. Furthermore, any Figures and images that are not ours have been included with the necessary permission for republication, which is attached to the manuscript.
- No animal studies are present in the manuscript.
- No human studies are present in the manuscript.
- Ethical Clearance: The project was approved by the local ethical committee at Thammasat University.

## Authors' contributions statement

D P, A K S, W S, and A D contributed equally and substantially to the writing of this paper. All authors reviewed and approved the final manuscript.

## Funding statement

This work was supported by Thammasat University Research Unit in Fixed Points and Optimization.

## References

1. Ali W, Malebary S. Particle swarm optimization-based feature weighting for improving intelligent phishing website detection. IEEE Access. 2020;8:116766–116780. https://doi.org/10.1109/ACCESS.2020.3003569.
2. Ali W, Ahmed AA. Hybrid intelligent phishing website prediction using deep neural networks with genetic algorithm-based feature selection and weighting. IET Inf. Secur. 2019;13(6):659–669. https://doi.org/10.1049/iet-ifs.2019.0006.
3. Alkanhel R, El-kenawy E S M, Abdelhamid A A, Ibrahim A, Alohali M A, Abotaleb M, Khafaga D S. Network Intrusion Detection Based on Feature Selection and Hybrid Metaheuristic Optimization. Comput. Mater. Contin., 2023;74(2):2677–2693. https://doi.org/10.32604/cmc.2023.033273.
4. Shukla S, Misra M, Varshney G. HTTP header based phishing attack detection using machine learning. Trans Emerg Telecommun Technol. 2024;35(1):e4872. https://doi.org/10.1002/ett.4872.
5. Feng F, Zhou Q, Shen Z, Yang X, Han L, Wang J. The application of a novel neural network in the detection of phishing websites. J Ambient Intell Human Comput. 2024;15:1865–1879. https://doi.org/10.1007/s12652-018-0786-3.
6. Han J, Pei J, Tong H. Data Mining: Concepts and Techniques. 4th edition. USA: Morgan Kaufmann Publishing; 2023.
7. Theng D, Bhoyar KK. Feature selection techniques for machine learning: a survey of more than two decades of research. Knowl Inf Syst. 2024;66(3):1575–1637. https://doi.org/10.1007/s10115-023-02010-5.
8. Patel D, Saxena AK. Feature selection in high dimension datasets using incremental feature clustering. Indian J Sci Technol. 2024;17(32):3318–3326. https://doi.org/10.17485/IJST/v17i32.2077.
9. Abdelhamid A A, El-Kenawy E S M, Ibrahim A, Eid M M, Khafaga D S, Alhussan A A, MIRJALILI S, KHODADADI N, LIM W H, SHAMS M Y. Innovative feature selection method based on hybrid sine cosine and dipper throated optimization algorithms. 2023;11:79750–79776. https://doi.org/10.1109/ACCESS.2023.3298955
10. Atteia G, El-kenawy E S M, Samee N A, Jamjoom M M, Ibrahim A, Abdelhamid A A, Ahmad T A, Khodadadi N, Ghanem R A, Shams MY. Adaptive dynamic dipper throated optimization for feature selection in medical data. Comput. Mater. Contin. 2023;75(1):1883–1900. https://doi.org/10.32604/cmc.2023.031723.
11. Patel D, Saxena AK, Laha S, Ansari GM. A novel scheme for feature selection using filter approach. In: 2022 7th International Conference on Computing, Communication and Security (ICCCS), Seoul, Korea. 2022;1–4. https://doi.org/10.1109/ICCCS55188.2022.10079604.
12. Patel D, Saxena A, Wang J. A machine learning-based wrapper method for feature selection. Int J Data Warehous Min. 2024;20(1):1–33. https://doi.org/10.4018/IJDWM.352041.
13. Myriam H, Abdelhamid A A, El-Kenawy E S M, Ibrahim A, Eid M M, Jamjoom M M, Khafaga D S. Advanced metaheuristic algorithm based on Particle Swarm and Al-biruni Earth Radius optimization methods for oral cancer detection. IEEE Access, 2023;11:23681–23700. https://doi.org/10.1109/ACCESS.2023.3253430.
14. El-Kenawy E S M, Khodadadi N, Mirjalili S, Makarovskikh T, Abotaleb M, Karim F K, Alkahtani H K, Abdelhamid A A, Eid M M, Horiuchi T, Ibrahim A, Khafaga D S. Metaheuristic optimization for improving weed detection in wheat images captured by drones. Mathematics. 2022;10(23):4421. https://doi.org/10.3390/math10234421
15. Hashemi A, Dowlatshahi MB, Nezamabadi-pour H. Ensemble of feature selection algorithms: a multi-criteria decision-making approach. Int J Mach Learn Cybern. 2022;13(1):49–69. https://doi.org/10.1007/s13042-021-01347-z.
16. Charbuty B, Abdulazeez A. Classification based on decision tree algorithm for machine learning. J Appl Sci Technol Trends. 2021;2(1):20–28. https://doi.org/10.38094/jastt20165.
17. Uddin S, Haque I, Lu H, Moni MA, Gide E. Comparative performance analysis of K-nearest neighbour (KNN) algorithm and its different variants for disease prediction. Sci Rep. 2022;12(1):6256. https://doi.org/10.1038/s41598-022-10358-x.
18. Song X, Liu X, Liu F, Wang C. Comparison of machine learning and logistic regression models in predicting acute kidney injury: a systematic review and meta-analysis. Int J Med Inform. 2021;151:104484. https://doi.org/10.1016/j.ijmedinf.2021.104484.
19. Desai M, Shah M. An anatomization on breast cancer detection and diagnosis employing multi-layer perceptron neural network (MLP) and convolutional neural network (CNN). Clin eHealth. 2021;4:1–11. https://doi.org/10.1016/j.ceh.2020.11.002.
20. Maswadi K, Ghani NA, Hamid S, Rasheed MB. Human activity classification using Decision Tree and Naive Bayes classifiers. Multimed Tools Appl. 2021;80(14):21709–21726. https://doi.org/10.1007/s11042-020-10447-x.
21. Palimkar P, Shaw RN, Ghosh A. Machine learning technique to prognosis diabetes disease: Random forest classifier approach. In: Bianchini M, Piuri V, Das S, Shaw, RN, editors. Advanced Computing and Intelligent Technologies. Lecture Notes in Networks and Systems, vol 218. Singapore: Springer;2022. p. 219–244. https://doi.org/10.1007/978-981-16-2164-2_19.
22. Babagoli M, Aghababa MP, Solouk V. Heuristic nonlinear regression strategy for detecting phishing websites. Soft Comput. 2019;23:4315–4327. https://doi.org/10.1007/s00500-018-3084-2.
23. Chiew, K L, Tan C L, Wong K, Yong K S, Tiong W K. A new hybrid ensemble feature selection framework for machine learning-based phishing detection system. Information Sciences. 2019;484:153–166. https://doi.org/10.1016/j.ins.2019.01.064.
24. Taha A. Intelligent ensemble learning approach for phishing website detection based on weighted soft voting. Mathematics. 2021;9(21):2799. https://doi.org/10.3390/math9212799.
25. Sharma SR, Singh B, Kaur M. Improving the classification of phishing websites using a hybrid algorithm. Comput Intell. 2022;8(2):667–689. https://doi.org/10.1111/coin.12494.
26. Siddiq MAA, Arifuzzaman M, Islam MS. Phishing website detection using deep learning. In: Proceedings of the 2nd In-

ternational Conference on Computing Advancements, Dhaka, Bangladesh, March 10 - 12. 2022;83–88. https://doi.org/10.1145/3542954.3542967.
27. Fadheel W, Al-Mawee W, Carr S. On phishing: URL lexical and network traffic features analysis and knowledge extraction using machine learning algorithms (a comparison study). In: 2022 5th International Conference on Data Science and Information Technology (DSIT), Shanghai, China. 2022;1–7. https://doi.org/10.1109/DSIT55514.2022.9943832.
28. Karim MFB, Hasan T, Tazreen N, Hakim SB, Tarannum S. An investigation of ML techniques to detect phishing websites by complexity reduction. In: 2022 IEEE International Conference on Cybernetics and Computational Intelligence (CyberneticsCom), Malang, Indonesia. 2022;144–149. https://doi.org/10.1109/CyberneticsCom55287.2022.9865297.
29. Medelbekov M, Nurtas M, Altaibek A. Machine learning methods for phishing attacks. J Probl Comput Sci Inf Technol. 2023;1(2):13–24. https://doi.org/10.26577/JPCSIT.2023.v1.i2.02.
30. Singh T, Kumar M, Kumar S. Enhancing phishing website detection using particle swarm optimization and feature selection techniques. In 2023 IEEE World Conference on Applied Intelligence and Computing (AIC). 2023:977–982. https://doi.org/10.1109/AIC57670.2023.10263814.
31. Davoudi M R, Yari A R. Improving the feature section method based on genetic algorithm to increase the efficiency of detecting phishing websites. Automatic Control and Computer Sciences. 2023;57(3):213–221. https://doi.org/10.3103/S0146411623030045.
32. Ali M S, Jain A K. Efficient feature selection approach for detection of phishing URL of COVID-19 era. In International Conference on Cyber Security, Privacy and Networking, Springer International Publishing, Cham. 2021:45–56. https://doi.org/10.1007/978-3-031-22018-0_5.
33. Alazaidah R, Al-Shaikh A, Al-Mousa MR, Khafajah H, Samara G, Alzyoud M. Website Phishing Detection Using Machine Learning Techniques. J Stat Appl Probab. 2024;13(1):119–129. https://dx.doi.org/10.18576/jsap/130108.
34. Taha MA, Jabar HD. A Machine Learning Algorithms for Detecting Phishing Websites: A Comparative Study. Iraqi J Comput Sci Math. 2024;5(3):275–286. https://doi.org/10.52866/ijcsm.2024.05.03.015.
35. Cheng J, Sun J, Yao K, Xu M, Cao Y. A variable selection method based on mutual information and variance inflation factor. Spectrochim Acta A Mol Biomol Spectrosc. 2022;268:120652. https://doi.org/10.1016/j.saa.2021.120652.
36. Çakır M, Değirmenci A, Karal O. Exploring the behavioural factors of cervical cancer using ANOVA and machine learning techniques. In: Mirzazadeh A, Erdebilli B, Babaee Tirkolaee E, Weber GW, Kar AK, editors. Science, Engineering Management and Information Technology. SEMIT 2022. Communications in Computer and Information Science, vol 1808. Cham: Springer;2023. p. 249–260. https://doi.org/10.1007/978-3-031-40395-8_18.
37. Nti IK, Nyarko-Boateng O, Aning J. Performance of machine learning algorithms with different K values in K-fold cross-validation. Int J Inf Technol Comput Sci. 2021;13(6):61–71. https://doi.org/10.5815/ijitcs.2021.06.05.
38. Prasad A, Chandra S. PhiUSIIL: A diverse security profile empowered phishing URL detection framework based on similarity index and incremental learning. Comput Secur. 2024;136:103545. https://doi.org/10.1016/j.cose.2023.103545.
39. Hannousse, Abdelhakim; Yahiouche, Salima (2021), “Web page phishing detection”, Mendeley Data, V3, https://doi.org/10.17632/c2gw7fy2j4.3.
40. AL-Jumaili ASA, Huda KT. A hybrid method of linguistic and statistical features for Arabic sentiment analysis. Baghdad Sci J. 2020;17(1(Suppl.)):385–390. https://doi.org/10.21123/bsj.2020.17.1(Suppl.).0385

# تصنيف بيانات التصيّد الاحتيالي باستخدام طريقة اختيار السمات الهجينة MI-AN

**دامودار باتيل[1]، أميت كومار ساكسينا[1]، ووتيفول سينتونافارات[2]، أبهيشيك دوبي[3]**

[1] قسم علوم الحاسوب وتكنولوجيا المعلومات، جامعة غورو غاسيداس فيشفيديالايا، بيلاسبور، الهند.
[2] قسم الرياضيات والإحصاء، كلية العلوم والتكنولوجيا، جامعة ثاماسات، باثوم ثاني، تايلاند.
[3] كلية الحوسبة وعلوم المعلومات، قسم تكنولوجيا المعلومات، جامعة التكنولوجيا والعلوم التطبيقية، صلالة، سلطنة عُمان.

## الخلاصة

شهدت هجمات التصيّد الاحتيالي عبر الويب تطورًا مستمرًا خلال السنوات القليلة الماضية، مما أدى إلى فقدان المستخدمين ثقتهم بالخدمات الإلكترونية والتجارة عبر الإنترنت. وللتعرّف على بيانات التصيّد الاحتيالي، تُستخدم مجموعة متنوعة من الأساليب والأنظمة المعتمدة على القوائم السوداء لمواقع التصيّد. إلا أن التطور السريع للتكنولوجيا أدى إلى ظهور تقنيات أكثر تعقيدًا لإنشاء مواقع إلكترونية جذابة للمستخدمين، ولذلك أصبحت التقنيات الحالية المعتمدة على القوائم السوداء غير قادرة على اكتشاف بيانات التصيّد الحديثة، مثل مواقع التصيّد الاحتيالي من نوع اليوم الصفري (Zero-Day Phishing Websites)..استُخدمت تقنيات تعلّم الآلة في العديد من الدراسات الحديثة للكشف عن بيانات التصيّد الاحتيالي والعمل كنظام إنذار مبكر لمثل هذه الهجمات. ومع ذلك، فإن غالبية هذه الأساليب اعتمدت في اختيار الخصائص الأكثر أهمية للمواقع الإلكترونية على تحليل التكرار أو الخبرة البشرية. ومن أجل تحسين تصنيف بيانات التصيّد الاحتيالي، يقترح هذا البحث الطريقة الهجينة MI-AN لتصنيف بيانات التصيّد الاحتيالي بذكاء باستخدام ثلاث مجموعات بيانات للتصيّد.تم استخدام ستة نماذج من تعلّم الآلة مع طريقة اختيار الخصائص المقترحة MI-AN. وأظهرت النتائج التجريبية أن الطريقة المقترحة حققت تحسينات ملحوظة في دقة التصنيف. وبالمقارنة مع النماذج الأخرى، حقق نموذج الإدراك متعدد الطبقات (Multi-Layer Perceptron) أعلى دقة بلغت (97.07%) في مجموعة البيانات الأولى، بينما حقق نموذج الغابة العشوائية ( Random Forest) أعلى دقة بلغت (96.60% و97.78%) في مجموعتي البيانات الثانية والثالثة على التوالي. كما تمت مقارنة الطريقة المقترحة مع أبحاث منشورة سابقًا، وأظهرت نتائج المقارنة تفوق الطريقة المقترحة من حيث الأداء. وبشكل عام، ساهمت طريقة MI-AN المقترحة بصورة متسقة في تحسين أداء النماذج عبر جميع مجموعات بيانات التصيّد الاحتيالي، مما يبرهن على متانتها وقدرتها على التكيّف وفعاليتها في إزالة الخصائص غير ذات الصلة لتحقيق تعميم أفضل للنماذج.

**الكلمات المفتاحية:** اختيار الخصائص باستخدام تحليل التباين (ANOVA)، اختيار الخصائص الهجين، تعلّم الآلة، المعلومات المتبادلة (Mutual Information)، مجموعات بيانات التصيّد الاحتيالي.