Sustainable Machine Intelligence Journal

Journal Homepage: sciencesforce.com/smij

Sustain. Mach. Intell. J. Vol. 11 (2025) 11–19

Paper Type: Original Article

An Efficient Phishing Detection Framework Based on Hybrid Machine Learning Models

Mohamed Elkholy [1], Mohamed Sabry [2] and Hussam Elbehiery [3,*]

[1] Faculty of Computer Science and Engineering, Alamein International University, Egypt; melkholy@aiu.edu.eg.
[2] Department of Computer Science, October 6th University, Egypt; mo7amed.sabry.hussien@gmail.com.
[3] Vanridge University, USA; drhussam@vru-edu.net.

Received: 09 Oct 2024 **Revised:** 01 Mar 2025 **Accepted:** 31 Mar 2025 **Published:** 03 Apr 2025

Abstract

The paper discusses an improved phishing detection system that uses hybrid machine learning for analyzing URL-based features. Further improvement of prior work using better feature selection as well as ensemble methods is discussed. The proposed model enhances classification accuracy, precision, and recall by utilizing an innovative hybrid ensemble approach that integrates Logistic Regression, Support Vector Machine, and Decision Tree, alongside newly incorporated evaluation techniques. The results manifest high improvement in performance metrics as compared to the previous methods. Various comparisons and benchmarks with other methods have further proved the robustness of our proposed system for detecting malicious URLs.

Keywords: Machine Learning; Uniform Resource Locator (URL); Protocol; Cybersecurity; Social Networks.

1 | Introduction

In this modern era, the Internet remains the backbone of communication, business activity, and information sources. Billions of computers worldwide are connected through various telecommunication technologies, including phone lines, fiber optics, wireless networks, and satellite systems. The exchange of data is done. Through Internet Protocol/Transmission Control Protocol (IP-TCP), which enables inter-computer communication. Communication is to take place. The fact that there is no centralization in internet management and contribution by organizations, research agencies, and universities is incredibly significant for everyone. Areas such as entertainment, education, e-commerce, medicine, and so forth [1]. While the Internet has good sides, it is also the same site to develop many cybercrimes like phishing. This was aimed at unsuspecting subjects, stealing their valuable information.

Phishing is a new source of high-security problems that use uniform resource locators (URL) techniques to lead a user to follow a false website. Detection of phishing URLs would be leading to keep the threat at bay without any data breach or monetary loss to the user [2]. As visible in Figure 1, a URL is a primary resource identifier of the web. It comprises different. Internal parts include the specified protocol HTTP or HTTPS, hostname, top-level domain, primary domain, and path. An understanding of the individual and collective roles of these components in identifying and accessing web resources is critical. The protocol must

Corresponding Author: drhussam@vru-edu.net
https://doi.org/10.61356/SMIJ.2025.11525

authenticate clients with servers before they can communicate with each other safely, and thus the format of the URL. It must help differentiate between authentic and malicious websites. Analyzing URLs brings in a sophisticated machine-learning methodology for the detection of phishing attacks by understanding. Abnormal and suspicious patterns emerge in the structure of URLs.

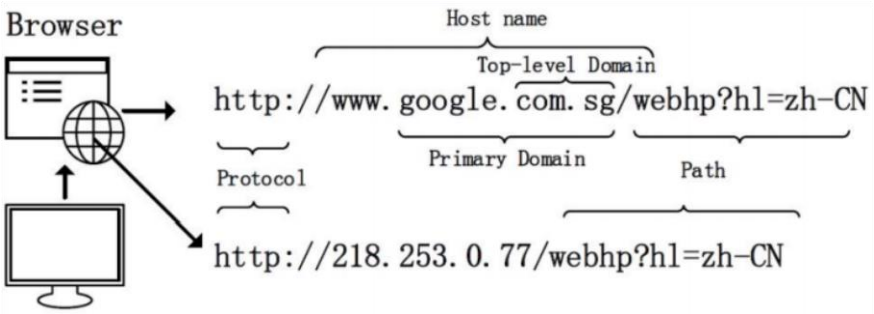

Figure 1. URL presentation based on HTTP.

This study focuses on phishing detection by leveraging machine learning algorithms to analyze URLs. The goal is to enhance the accuracy and efficiency of identifying phishing websites by using. A dataset [3] comprising over 11,000 phishing URL attributes. Various machine learning models, including Decision Trees, Random Forests, Naïve Bayes, Gradient Boosting Machines, and Support Vector Classifiers have been employed to classify URLs as phishing or legitimate. Additionally, a hybrid model combining Logistic Regression, Support Vector Classifiers, and Decision Trees (LSD) Soft and hard voting mechanisms are proposed for improved performance.

The major contributions of this study include:

- Proposing a phishing URL detection system to safeguard users from cyberattacks.
- Developing a dataset with phishing URL attributes for effective classification.
- Implementing advanced machine learning models and a hybrid approach for accurate threat detection.
- Using cross-fold validation, grid search, and feature selection techniques to enhance prediction results.
- Evaluating the proposed methodology using metrics such as accuracy, precision, recall, specificity, and F1-score.

2 | Related Work

Phishing continues to be an excellent challenge in networking and Internet security. A lot of efforts have been put forth by various researchers to counter the elimination of phishing. Attacks using mechanisms on machine learning and deep learning or towards approaches with lists, such as blacklists and whitelists. Thus, phishing detection systems could be classified into two categories: list-based and machine learning-based systems. This section reviews earlier works. Of the two types.

2.1 | List-Based Phishing Identification Systems

Phishing detection based on lists uses blacklists and whitelists to check if a website is a cybercrime. Website or a legitimate one. Whitelist-based systems guarantee users' safe browsing by maintaining. Lists of trusted websites. For example, [4] presented a system that holds IP addresses of websites. Having login interfaces and alerting users while accessing an unrecognized site; similarly, [5] came up. With a system that automatically updates whitelists based on link attributes sourced from source code and domain-IP associations to improve accuracy. The system achieved an 86.02. On the other hand, the blacklist-based systems collect and store those already identified as phishing. URLs from sources including user reports, spam detection systems, and third-party authorities. Examples would be Phish Net [6] and Google Safe Browsing API [7], which rely on blacklists to detect phishing in near real-time with an extremely high success rate. But then, these types of

systems usually require regular updates to be able to work well against attacks like zero-day, in addition to other disadvantages [8]. The approximate matching algorithms used in phishing URL detection. Will be taxing for a system because of the rapid growth in the entries found in the blacklist.

2.2 | Machine-Learning-Based Identification Systems

Analyzing URLs using machine learning has become the most favored way of differentiating malicious ones. Websites from harmless ones. In contrast to listening approaches, machine learning can be revealed. New phishing attempts that have yet to impress into blacklists. As an illustration, [9] built a system. Named CANTINA, which uses the term frequency-inverse document frequency (TF-IDF) to extract. Keywords for classifying websites. However, its performance is tied to the fact that it will be very. Sensitive to the English language. CANTINA improved by adding on [10] to include HTML. Attributes and attained a 92 percent accuracy level but with an extremely high false positive rate. Phish-WHO [11] proposed a three-layer scheme for phishing detection that considers keyword and domain relationships, while other studies use other URL feature set-based methods. For example, [12] utilized SVM (support vector machines) to classify sites based on directory structure, special. Characters as well as URL length. In comparative results, adaptive regularization algorithms tended to. To outperform others with minimal resource consumption.

3 | Methodology

Our methodology involves extracting key URL features to differentiate between phishing and legitimate sites. We use feature importance analysis with Random Forest to identify the most impactful predictors, such as HTTPS presence and anchor tags. A hybrid ensemble model combining Decision Tree, Random Forest, SVC, and XGBoost improves classification accuracy through majority voting. Finally, the system enables real-time phishing detection by processing URLs and classifying them instantly.

3.1 | Dataset

The dataset comprises 11,000 records with 33 URL features, sourced from the Kaggle [3] repository. Key attributes include URL length, presence of special characters, domain registration length, and HTTPS usage. The dataset is divided into training (70%) and testing (30%) subsets. The given dataset is moderately balanced, leaning more toward legitimate cases, though. That means. A little deviation is there between the classes, which again is good for machine learning. Models, as it reduces bias due to class imbalance, as shown in Figure 2. However, it might still require addressing because the little imbalance depends on the model used.

a. Legit: This class represents legitimate cases, accounting for 55.7% of the data. It suggests that these cases. Are they non-phishing or authentic?

b. Phishing: This class represents phishing cases, making up 44.3% of the datasets. These are instances identified as fraudulent or malicious.

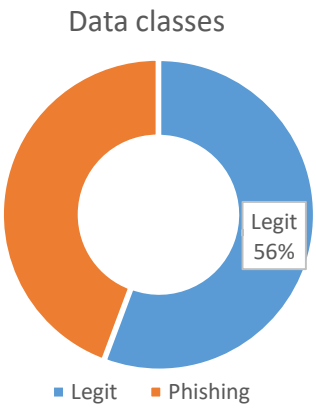

Figure 2. Data classes.

3.2 | Automated Feature Extraction

A personalized scraping mechanism would automatically extract different primary features from user-provided URLs. This mechanism analyzes various properties of the URLs, including important ones. Key attributes like the presence of HTTPS, domain-related parameters, length of URL, and subdomain, etc. These features would then be crucial for differentiating genuine from phishing URLs. From the basis of machine learning models.

3.2.1 | Feature Importance Analysis

To determine the contribution of each feature to the classification task, A random forest model was used to analyze the importance of features. As shown in Figure 3, features are ranked based on their importance, identifying the key predictors of phishing detection.

From the analysis:

a. HTTPS: The presence of HTTPS is the most critical feature, significantly distinguishing between phishing and legitimate URLs.

b. Anchor URL: The usage of anchor tags in URLs also plays a key role, as phishing sites often manipulate anchor attributes.

c. Website Traffic and Subdomains: High traffic and fewer subdomains are typical characteristics of legitimate websites, making these features highly indicative.

d. Other key features include Links in Script Tags, Server Form Handler, and Prefix-Suffix Usage. Which further contributes to accurate predictions.

Lower-ranked features, such as Disable Right Click, I-Frame Redirection, and Status Bar Customization still hold value but have less predictive power in comparison.

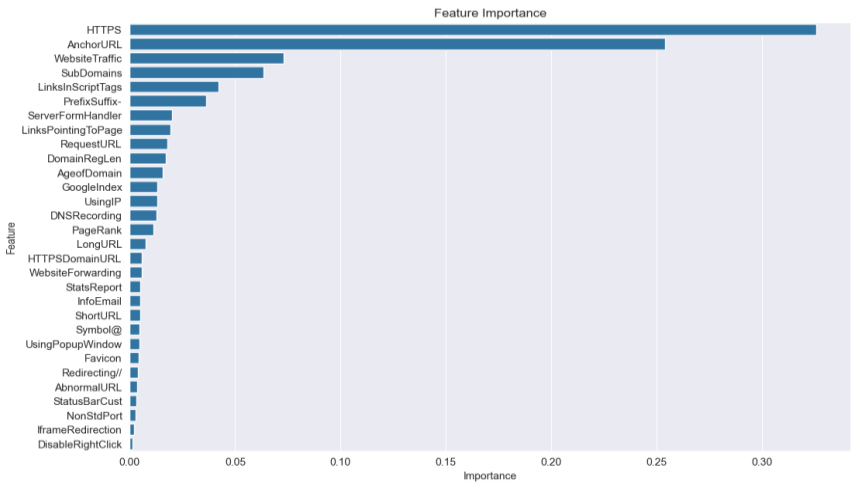

Figure 3. Feature importance ranking.

With the help of all features, including URL structure, content, and metadata, we concluded. The successful building of a phishing detection system relies primarily on these features. The most important predictors include HTTPS and Anchor URLs, which indicate that URL features are the most important. Essential in phishing detection [13].

3.3 | Hybrid Model Ensemble

- Several machine learning classifiers, including Decision Tree (DT), Random Forest (RF), Support Vector Classifier (SVC), and XGBoost are individually trained using the extracted data. URL features.

- A majority-voting [12] scheme integrates the predictions of classifiers. The final classification, the decision of the hybrid that weighs the different classifiers' contributions, is based on performance. Metrics calculated from the confusion matrix. Thus, hybridization results in a more robust system and an accurate outcome. Perkel, Evgenia, and Burgess in the second edition of K-Systems McGraw-Hill (2004).
- The combined predictions by the classifiers are used in a majority voting scheme. Further, the classification decision of the hybrid model is weighted against the performance of the classifiers, which are evaluated based on confusion matrix metrics. This hybridization further ensures good robustness and accuracy of the overall system.

3.4 | Real-Time Detection

- When a URL is provided, the system first processes it through the feature extraction module to gather the necessary features.
- The extracted features are then input into the ensemble model, which classifies the URL in real time, providing a rapid and reliable detection outcome.

3.5 | System Specifications

Experiments were conducted on a system with the following specifications:

- Processor: AMD Ryzen 7
- Memory: 16GB RAM
- GPU: NVIDIA RTX 3050 (4GB)

4 | Proposed Model Architecture

The classification methodology follows a structured approach, as illustrated in Figure 4. The process consists of several key stages:

- Data Collection and Pre-Processing: Raw URL data is collected and undergoes rigorous cleaning, including quality checks, outlier removal, and imputation. Essential features such as URL structure and length are extracted to ensure meaningful representation.
- Dataset Splitting: The dataset is divided into training (75%) and testing (25%) subsets to maintain balance and enable unbiased model evaluation.
- Training Phase: Machine learning algorithms are trained on carefully selected features. Robust techniques such as feature selection are applied to emphasize key predictors while reducing noise.
- Cross-Validation and Refinement: Models are iteratively fine-tuned using cross-validation to optimize their parameters and enhance predictive accuracy.
- Ensemble Learning with Majority Voting: To improve classification performance, predictions from multiple models, including Random Forest and Gradient Boosting, are combined using a majority voting approach. This strategy leverages the strengths of individual models, resulting in greater accuracy and robustness.
- Evaluation and Model Selection: Performance metrics such as accuracy, precision, recall, and F1-score are used to compare models, ensuring that the most effective one is selected.
- Final Model Deployment: The optimized ensemble model, incorporating majority voting, is deployed as the final phishing detection system, providing reliable and accurate classification results.

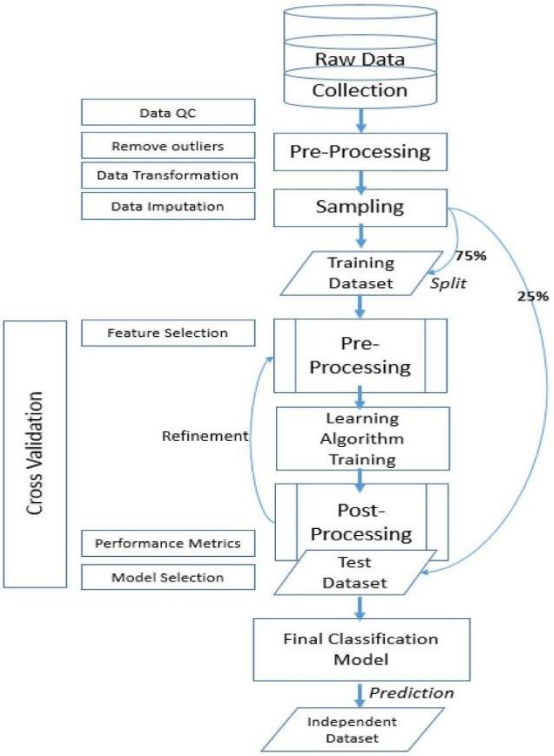

Figure 4. Classification methodological structure.

Figure 4 illustrates the overall classification methodology, outlining each stage of the process from data collection to final deployment. The structured approach ensures effective feature extraction, model training, and evaluation, leading to a robust phishing detection system. By leveraging ensemble learning and cross-validation, the model achieves higher accuracy and reliability in distinguishing phishing URLs from legitimate ones.

5 | Evaluation Metrics

Machine learning performance must be evaluated using several key evaluation metrics. These Metrics assess the number of true and false predictions made by the model for both legitimate. And phishing classes. Evaluation parameters such as accuracy, precision, recall, specificity, and the F1-score were employed in this study.

- Accuracy measures the overall performance of the model in terms of the number of correct predictions. Predictions (true positives and true negatives) relative to the total number of predictions. It is Mathematically defined as shown in Eq. (1):

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN} \tag{1}$$

- Precision, as noted in [13], evaluates the frequency with which the model is correct. Predicting the positive class. Precision indicates the proportion of true positive predictions among all positive predictions, highlighting the extent to which the model correctly classifies phishing URLs. Precision is calculated using Eq. (2):

$$\text{Precision} = \frac{\mathbf{TP}}{\mathbf{TP+FP}} \tag{2}$$

- Recall, another important metric referenced in [14], measures the proportion of actual positives. Instances that the model successfully identifies. It evaluates how well the model detects phishing. And legitimate URLs. The recall metric is defined in Eq. (3):

$$\text{Recall} = \frac{\text{TP}}{\text{TP+FN}} \tag{3}$$

- The F1-score is the harmonic mean of precision and recall. It is particularly useful when there. There is a need to balance precision and recall, ensuring both metrics are optimized. The F1-score is. Calculated using Eq. (4):

$$F1 = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (4)$$

These evaluation metrics provide a comprehensive framework for assessing the performance of machine learning models in phishing detection tasks, enabling the identification of the most dependable. And an effective classifier for the dataset.

5.1 | Performance Comparison

Tables 1 and 2 present the performance metrics of various machine learning models. Metrics such as accuracy, precision, recall, and F1-score were evaluated on the test dataset.

Table 1. Performance metrics of models - current study.

Model	Accuracy (%)	Precision (%)	F1-Score (%)
Logistic Regression	92.44	92.00	91.93
DecisionTreeClassifier	98.38	96.00	95.96
RandomForestClassifier	98.38	97.00	97.00
SupportVectorClassifier	94.65	95.00	94.95
XGBoost Classifier	97.91	97.00	97.00
Hybrid Ensemble Model	94.79	95.00	94.95

Table 2. Performance metrics of models - previous study.

Model	Accuracy (%)	Precision (%)	F1-Score (%)
Decision Tree	95.41	95.80	95.91
Random Forest	96.77	96.73	97.12
Support Vector Machine	71.80	96.34	65.67
Gradient Boosting Machine	70.34	99.65	64.10
Hybrid (LR+SVC+DT) Hard	94.09	93.31	94.79
Proposed Approach	98.12	97.31	95.89

5.2 | Key Insights and Observations

The comparative analysis of machine learning models demonstrates significant differences in performance. Across various classifiers. Key insights are outlined below:

a. Logistic Regression: Achieved consistent performance with a test accuracy of 91.92%, demonstrating its reliability for baseline phishing detection.

b. Decision Tree Classifier: Provided strong predictive performance with a test accuracy of 95.99%, but slightly lower generalization compared to ensemble models.

c. Random Forest Classifier: Delivered robust results with a high test accuracy of 97.08%, highlighting its effectiveness in handling complex URL features.

d. Support Vector Classifier (SVC): Balanced precision and recall effectively, achieving a test accuracy of 94.51%.

e. XGBoost Classifier: Outperformed most individual classifiers with a test accuracy of 96.89%, leveraging gradient boosting for superior learning from features.

f. Hybrid Ensemble Model: Combined multiple classifiers for robust predictions, achieving competitive performance with a test accuracy of 94.63% and an F1-score of 94.95%.

5.3 | Performance Highlights

- The Random Forest Classifier emerged as the best-performing model with the highest test accuracy (97.08%) and an F1-score of 97%, demonstrating its suitability for phishing detection.
- The XGBoost Classifier also showed excellent performance, achieving near-perfect precision and recall metrics, making it a strong contender for deployment.
- The Hybrid Ensemble Model provided competitive results, effectively leveraging the majority voting for robust phishing URL classification.

5.4 | Key Improvements

The proposed ensemble model showed several advancements:

- Improved recall, significantly reducing false negatives compared to individual classifiers.
- Enhanced robustness through majority voting, achieving balanced precision and recall.
- High computational efficiency due to optimized hyperparameter tuning and feature selection.

These results underscore the importance of ensemble methods for phishing detection systems, combining. The strengths of multiple classifiers for optimal performance.

6 | Conclusion

This study demonstrates that hybrid machine-learning approaches significantly enhance the performance of phishing detection systems. By leveraging URL-based features and applying advanced algorithms in combination with ensemble methodologies, Notable improvements in accuracy, precision, recall, and F1-score are achieved by the proposed model.

A key contribution of this work is the automated feature extraction and important analysis, which ensures the model focuses on the most relevant attributes for phishing classification. Additionally, the introduction of a hybrid ensemble model, combining Logistic Regression, Support Vector Machines, and Decision Trees through a weighted majority voting mechanism, leads to a more balanced and effective detection system.

The results highlight the robust performance of Random Forest and XGBoost classifiers, yet the proposed hybrid model offers a more robust and competitive approach, particularly in addressing false negatives and zero-day phishing attacks. This underscores the advantage of hybrid methods in providing scalable and real-time detection against evolving cyber threats.

For future work, integrating deep learning models with real-time streaming data could further enhance adaptability and detection accuracy in dynamic cybersecurity environments. Exploring graph-based approaches or Graph Neural Networks (GNNs) might also offer additional insights into phishing patterns and adversarial attack resistance.

Acknowledgments

The author is grateful to the editorial and reviewers, as well as the correspondent author, who offered assistance in the form of advice, assessment, and checking during the study period.

Author Contributions

All authors contributed equally to this work.

Funding

This research was conducted without external funding support.

Data Availability

The code and dataset used in this study are available in the following GitHub repository: Phishing Site Detection Repository. This repository contains the implementation of the models, data preprocessing scripts, and evaluation metrics used in the study. Researchers and practitioners can access the code to reproduce the results or extend the work further.

Conflicts of Interest

The author declares that there is no conflict of interest in the research.

Ethical Approval

This article does not contain any studies with human participants or animals performed by any of the authors.

References

[1] A. Karim, M. Shahroz, K. Mustofa, et al., Phishing Detection System Through Hybrid Machine Learning Based on URL, IEEE Access, 2023.
[2] R. Ø. Skotnes, "Management commitment and awareness creation—ICT safety and security in electric power supply network companies," Inf. Compute. Secure, vol. 23, no. 3, pp. 302–316, Jul. 2015.
[3] Phishing Website Dataset, Available at: https://www.kaggle.com/datasets/akashkr/phishing-website-dataset.
[4] A. K. Jain and B. Gupta, "PHISH-SAFE: URL features-based phishing detection system using machine learning," in Cyber Security. Switzerland: Springer, 2018, pp. 467–474.
[5] Y. Cao, W. Han, and Y. Le, "Anti-phishing based on automated individual whitelist," in Proc. 4th ACM Workshop Digit. Identity Manage., Oct. 2008, pp. 51–60.
[6] S. Sheng, M. Holbrook, P. Kumaraguru, L. F. Cranor, and J. Downs, "Who falls for phish? A demographic analysis of phishing susceptibility and effectiveness of interventions," in Proc. SIGCHI Conf. Hum. Factors Compute. Syst., Apr. 2010, pp. 373–382.
[7] P. Prakash, M. Kumar, R. R. Kompella, and M. Gupta, "Phish Net: Predictive blacklisting to detect phishing attacks," in Proc. IEEE INFOCOM, Mar. 2010, pp. 1–5.
[8] P. K. Sandhu and S. Singla, "Google safe browsing-web security," in Proc. IJCSET, vol. 5, 2015, pp. 283–287.
[9] S. Sheng, B. Wardman, G. Warner, L. Cranor, J. Hong, and C. Zhang, "An empirical analysis of phishing blacklists," in Proc. 6th Conf. Email Anti-Spam (CEAS), Mountain View, CA, USA. Pittsburgh, PA, USA: Carnegie Mellon Univ., Engineering and Public Policy, Jul. 2009.
[10] Y. Zhang, J. I. Hong, and L. F. Cranor, "Cantina: A content-based approach to detecting phishing web sites," in Proc. 16th Int. Conf. World Wide Web, May 2007, pp. 639–648.
[11] G. Xiang, J. Hong, C. P. Rose, and L. Cranor, "CANTINA+: A feature-rich machine learning framework for detecting phishing web sites," ACM Trans. Inf. Syst. Secure, vol. 14, no. 2, pp. 1–28, Sep. 2011.
[12] C. L. Tan, K. L. Chiew, K. Wong, and S. N. Sze, "Phish WHO: Phishing webpage detection via identity keywords extraction and target domain name finder," Deci's. Support Syst., vol. 88, pp. 18–27, Aug. 2016.
[13] S. C. Jeeva and E. B. Raising, "Intelligent phishing URL detection using association rule mining," Hum. -Centric Compute. Inf. Sci., vol. 6, no. 10, pp. 1–19, 2016.
[14] D. M. W. Powers, "Evaluation: From precision, recall and F-measure to ROC, informedness, markedness & correlation," Journal of Machine Learning Technologies, vol. 2, no. 1, 2011, DOI:10.9735/2229-3981.