International Information and Engineering Technology Association
Advancing the World of Information and Engineering

International Journal of Safety and Security Engineering
Vol. 15, No. 9, September, 2025, pp. 1809-1817
Journal homepage: http://iieta.org/journals/ijsse

Network Anomaly Activity Detection Model Based on Feature Correlation Analysis

Yohanes Priyo Atmojo[1], I Made Darma Susila[1], Eva Hariyanti[2], Dandy Pramana Hostiadi[3*], Gede Angga Pradipta[3], Putu Desiana Wulaning Ayu[4]

[1] Department of Informatics and Computer, Institut Teknologi dan Bisnis STIKOM Bali, Bali 80234, Indonesia
[2] Department of Information Systems, Universitas Airlangga, Surabaya 60115, Indonesia
[3] Department of Magister Information System, Institut Teknologi dan Bisnis STIKOM Bali, Bali 80234, Indonesia
[4] Department of Information Technology, Politeknik Negeri Bali, Bali 80361, Indonesia

Corresponding Author Email: dandy@stikom-bali.ac.id

https://doi.org/10.18280/ijsse.150905

Received: 1 August 2025
Revised: 3 September 2025
Accepted: 15 September 2025
Available online: 30 September 2025

Keywords:
anomaly detection, intrusion detection system, correlation, network security, machine learning, intersection analysis, feature selection

ABSTRACT

Anomalous activity in computer networks can disrupt communication services between computers and potentially lead to attacks. Several previous studies have introduced machine learning-based anomaly detection models and have optimized them using feature selection methods. However, the feature selection process requires a correlation analysis to assess the strength of feature correlations, thereby improving the performance of the detection model. This paper proposes a new approach to detecting anomalous activity that potentially indicates malicious activity in computer networks. It aims to analyze improvements in the classification model's detection performance using correlation intersection analysis with the Pearson and Kendall correlation methods. The contribution lies in the approach of selecting correlated features using both correlation approaches, yielding the best results with eight features. In the experiment, the model uses the UNSW NB-15 public dataset and is limited to three classification methods. The Decision Tree classification method achieved optimal performance, with a detection accuracy of 96.63%, an F1-score of 94.51%, a recall of 97.96%, and a precision of 91.29%. Network administrators can utilize the proposed model to expedite the analysis of anomalous activity and integrate it with intrusion detection systems.

1. INTRODUCTION

In the cyber era, handling and anticipating attacks is a particular concern [1]. Anomalous activity detection techniques use several approaches to identify activity patterns that indicate malicious threats, including malware use [2]. Malicious activities can include threats, system destruction, and theft of critical data or information, resulting in financial losses [3]. Furthermore, malware use can lead to crucial data theft, account identity theft, forgery, and Denial of Service (DoS) attacks [4, 5]. Malware attacks are frequently found in computer network infrastructure communications, mobile devices or smartphones, and Internet of Things-based smart telecommunications [6, 7].

The detection of malicious activity in computer networks has been studied in several previous studies, including anomaly-based [8], signature [9, 10], mining [11-14], and case-based [15, 16]. The previous studies introduced the model detection through the analysis of network traffic data [17-19] and system log data [20]. However, the detection analysis techniques used require complex methodologies, long analysis times, and, at times, suboptimal performance. Attack detection methods developed in previous studies often introduce detection analysis with a mining approach, involving classification methods [21], clustering [22, 23], graph-based [17], and correlation analysis approaches [24, 25].

The detection model, which utilizes classification methods, can be optimized using feature selection techniques. Appropriate feature selection techniques can enhance detection performance, including accuracy, precision, recall, and computational time [17, 26-31]. Some techniques used include PCA and chi-square [32], wrapper [33], probability analysis [34], ANOVA [35], and manual analysis based on the degree of importance [36]. These selection methods can improve classification performance, but are still carried out independently and affect the characteristics of the data used. Furthermore, previous feature selection studies have not measured feature correlations, which are needed to identify correlated features, demonstrate causal relationships, and assess feature influence. Thus, correlation analysis during the implementation of an anomaly activity detection model can achieve optimal detection performance and identify important features in anomalous activity data.

This paper proposes a new approach for anomaly activity detection by analyzing the header traffic that may indicate attacks or malicious activity. The proposed model adopts correlation theory. The objective of this study is to investigate improvements in anomaly detection performance by measuring feature correlation intersections. The main contribution of this paper lies in the use of the intersection technique, obtained from two correlation approaches. It aims to improve model classification performance by identifying important features using a feature selection technique. Determining the best features for classification begins by selecting the most highly ranked features from two correlation methods, Pearson and Kendall. Then, the model performs an intersection analysis of feature appearance across the two correlation methods. Thus, each resulting feature is a pair of causally correlated features that are important for analyzing anomalous activity in the dataset. Therefore, the results of the best feature analysis can improve detection performance, making it accurate and optimal. Network administrators can utilize the findings of this study to facilitate the analysis of anomalous activity that may compromise communication services in computer networks.

This paper is organized as follows: Section II explains related studies focusing on anomaly activity detection using feature selection. Section III addresses the proposed method. Section IV discusses the experimental results and critical evaluation. Finally, Section V provides the paper's conclusions and future work.

2. RELATED WORKS

The study of anomalous activity detection has challenges and has been carried out by previous researchers. Anomalous activity detection models that have the potential to be malicious attack activities were developed based on anomaly-based analysis [1, 37], signature [9, 10], mining [11-14], and case-based [15, 16, 21]. There are several well-known datasets used for testing anomaly activity model detection, including KDD CUP [17, 38], DARPA [39], Kyoto dataset, and UNSW dataset [40]. Mining-based techniques, such as machine learning classification-based detection models, have been implemented and successfully detected anomalous activities and even attack activities [14, 28, 41, 42].

Kocher and Kumar [29] conducted detection using several classification methods, including Stochastic Gradient Descent (SGD), Random Forest (RF), Naïve Bayes (NB), k-Nearest Neighbors (k-NN), and Logistic Regression (LR) classifiers. Their research provides an optimization classification model for feature selection in the feature selection stage using the chi-square method. The model was tested on the UNSW-NB15 dataset, achieving a detection accuracy of 99.64% with the RF method. However, in this study, the analysis of computational time and the features used could not show a correlation.

In the study [17], an attack detection model was introduced using graph-based analysis. The study used the KDD Cup dataset, with the detected attack type being DoS. Feature analysis employed a graph neural network-based feature correlation approach. The detection results showed a detection accuracy of 98.85%. However, this study focused only on one type of attack: DoS. In fact, real-life attacks can exhibit variants that can be detected through feature correlation, enabling identification of their causality.

Umar et al. [30] proposed a detection model with a machine learning classification model, involving six classification methods such as NB, Artificial Neural Network (ANN), RF, Support Vector Machine (SVM), k-NN, and Deep Neural Network (DNN). This study used several datasets, including CSE-CIC-IDS2018, UNSW-NB15, and NSL-KDD. The feature selection optimization technique used a wrapper and performed data normalization. The detection results showed good performance: RF on the NSL-KDD dataset achieved 99.86%, and ANN achieved the best accuracy on CSE-CIC-IDS2018 with 95.43%. However, this model still produces high-dimensional features and achieves little feature reduction, resulting in high computational costs.

Previous research has applied various feature selection techniques to improve detection performance. However, these studies relied solely on a single feature selection method and failed to demonstrate feature correlations, thereby missing important interrelated features. Changing the feature selection approach in a detection model can alter the ranking of important features and make them uncorrelated. In contrast, our model introduces a novel approach by selecting features based on their correlation, thus demonstrating causal relationships. When correlated features are identified as important, detection performance improves, particularly by reducing false positives. Furthermore, the intersection of these two correlation-based feature selection approaches underscores the importance of selected features in analyzing anomalous activity in computer networks.

3. PROPOSED METHOD

This paper focuses on improving the performance of a classification model for anomaly activity detection by optimizing feature selection via correlation intersection analysis. The proposed method is shown in Figure 1, which is divided into three main stages: pre-processing, anomaly detection, and model performance evaluation.

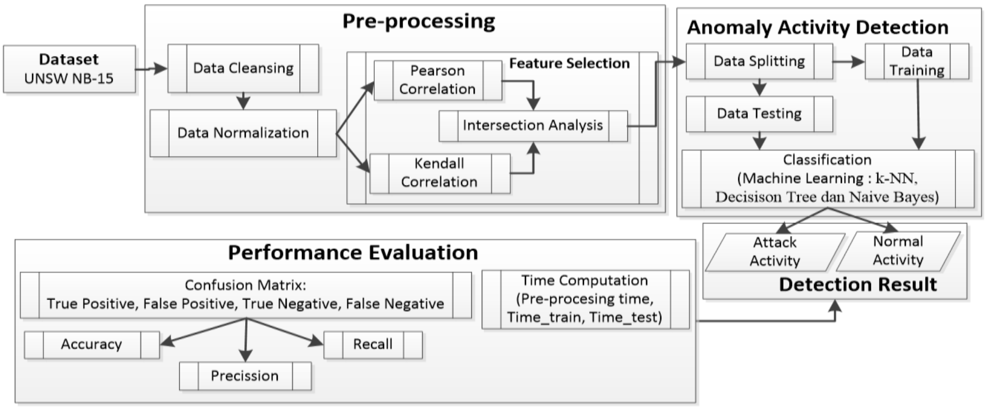


Figure 1. Proposed method

3.1 Dataset description

This research uses the UNSW-NB-15 public dataset, where the dataset developed by the Australian Center for Cyber Security at UNSW Canberra, offering a realistic blend of legitimate modern network traffic and nine distinct types of synthesized cyberattacks, namely, Exploits, Fuzzers, Reconnaissance, Backdoors, Analysis, DoS, Shellcode, Generic, and Worms. This dataset produces 2,540,044 labeled raw data records totaling 100 GB in .pcap format, recorded with tcpdump, and was generated in the Cyber Range Lab using the IXIA PerfectStorm device. The raw data is extracted using Argus and Bro-IDS (Zeek) tools, resulting in 49 descriptive features, including flow-based, basic, content, timing, and connection-context features, along with binary (normal vs. malicious) and categorical class labels. The dataset is organized into four CSV files, with separate testing and train splits of 82,332 and 175,341 records, and includes auxiliary ground truth and event log files to aid validation and analysis [43].

3.2 Preprocessing phase

In the pre-processing stage, the model has three sub-processes: data cleansing, data normalization, and feature selection.

3.2.1 Data cleansing

Data cleansing is the data cleaning stage, where null data is removed from the dataset records and data is checked for redundancy or data that does not meet the Intrusion Detection Message Exchange Format (IDMEF) writing standards [44].

3.2.2 Data normalization

Data Normalization is the stage that normalizes the values and data types of each feature. First, categorical values are transformed into numeric values. Second, the values of each numeric data point are standardized to a range of 0 to 1 using min-max scaling, as shown in Eq. (1):

$$z_i' = \frac{z_i - \min(Z)}{\max(Z) - \min(Z)} \tag{1}$$

where, z_i is the original (pre-normalization) value of the i-th feature data, $min\ (Z)$ and $max\ (Z)$ denote the minimum and maximum values across all features Z features. Then, z_i' denotes the normalized value of each feature, scaling the original data to the range 0 to 1.

3.2.3 Feature selection

In this paper, both correlation analyses, Kendall's and Pearson's, are used to select the best features and generate correlated feature pairs. The two techniques are used for their superior filter-based properties, simplicity, speed, and effectiveness in assessing the relevance of each feature to the target class, thereby yielding features that demonstrate causality [24, 25]. In this study, feature selection began by examining the best candidate features from the Pearson and Kendall methods. Each method ranked the best features by analyzing the sum of two features up to a total of n, where n is the total amount of features in the dataset. Then, the candidate results from both correlation techniques yielded the best feature set, and intersection analysis was employed to examine the occurrence of similar features across the two correlation methods. Any feature that intersects with both methods was considered the best feature and processed in the classification phase. Algorithm 1 describes the process of selecting the best feature from the intersection analysis between Kendall and Pearson.

Algorithm 1. Feature Selection – Intersection between Pearson and Kendall Correlation

```
INPUT:
    Matrix $X \in \mathbb{R}^{n \times m}$
    Feature numbers → $m$; Index of features → $j$
    Number of data → $n$; Index of data → $i$
    Label Vector $y \in \{0,1\}^n$
    Number of selected features → $k$
OUTPUT:
    Selected features→ $F_{selected}$

DEFINITION:
    Calculate the mean of list $a$ → $mean(a)$
    Value of $j$-th feature in $i$-th data → $x_{ij}$
    Mean value of $j$-th feature → $\bar{x_j}$
    Mean value of label → $\bar{y}$
    Sort list a in descending → $sort_{desc}(a)$

Step 1: Initialize empty list for Pearson and Kendall scores
$PearsonScores \leftarrow []$
$KendallScores \leftarrow []$

Step 2: Obtain $PearsonScores$
for each feature $j$ in $\{1, \dots, m\}$ do:
    Compute Pearson correlation score $r_j$
    $r_j \leftarrow \frac{\sum_{i=1}^{n}(x_{ij} - \bar{x_j})\,(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_{ij} - \bar{x_j})^2} \times \sqrt{\sum_{i=1}^{n}(y_i - \bar{y})^2}}$

    $PearsonScores_j \leftarrow |r_j|$
end for

Step 3: Sort and get top $k$ of $PearsonScores$ (descending)
$F_{pearson} \leftarrow$ indices of top- $k$ features from $sort_{desc}(PearsonScores)$

Step 4: Obtain $KendallScores$
Initialize: $C \leftarrow 0, D \leftarrow 0$
    for each pair $(i, l); such\ that\ i < l$ do:
        if $(x_{ij} - x_{lj})(y_i - y_l) > 0$ then:
            $C \leftarrow C + 1$
        else:
            $D \leftarrow D + 1$
        end if
    Compute Kendall correlation score $\tau_j$
    $\tau_j \leftarrow \frac{(C - D)}{\frac{1}{2}n(n-1)}$
    $KendallScores_j \leftarrow |\tau_j|$
    end for

Step 5: Sort and get top $k$ of $KendallScores$ (descending)
$F_{kendall} \leftarrow$ indices of top- $k$ features from $sort_{desc}(KendallScores)$

Step 6: Compute intersection
$F_{Selected} \leftarrow F_{Pearson} \cap F_{Kendall}$

return $F_{Selected}$
```

3.3 Anomaly activity detection

The anomaly activity detection stage is the process of detecting anomalous activity, indicated as malicious activity, using a machine learning-based classification model. The detection stage involves dividing the data into two parts: 70% for training and 30% for testing. The training model adopts three machine learning algorithms: Naive Bayes, k-NN, and Decision Tree. The testing data is then classified using the classification model trained with the training data. In this paper, we consider these three classification algorithms because they have demonstrated high detection performance on large datasets for detecting activity anomalies and attacks in computer networks [45, 46]. Furthermore, these methods apply to high-dimensional classification processes that involve both numeric and categorical data [13]. The classification method uses Naive Bayes, as shown in Eq. (2):

$$P(y|x) = \frac{P(y)P(x|y)}{P(x)} \tag{2}$$

where, x denotes the vector comprising the features selected by $F_{selected}$. $P(y)$ represents the prior probability of the anomaly class, and $P(x_i|y)$ is the probability of a particular feature x_i given that the instance belongs to the anomaly class. The classification method using the Decision Tree uses Eq. (3):

$$E(S) = \sum_{i=1}^{c} -p_i . log_2 p_i \tag{3}$$

And the calculation of Information Gain on the attribute A uses Eq. (4):

$$Gain(S, A) = E(S) - \sum_{v \in values(A)} \frac{|S_v|}{S} E(S_v) \tag{4}$$

where, S denotes the subset of data labeled as anomalies that might represent attacks. p_i is the proportion of data belonging to the attack class j. Then A stands for the attributes selected through feature selection, and v refers to a possible value of an attribute in A. S_v is the subset of data with the value v. The classification method with k-NN uses Eq. (5):

$$\delta(a, b) = \sqrt{\sum_{k=1}^{d} (X_k - Y_k)^2} \tag{5}$$

where, $\delta(a, b)$ represents the classification proximity measure between the anomaly data from the training set and the anomaly data from the test set. d denotes the total number of data points classified. The parameter k refers to the number of nearest neighbors used in voting for each anomaly class. X denotes training data, and Y denotes the testing data.

3.4 Performance evaluation

The classification model's performance was evaluated using a confusion matrix, which was constructed by counting the true positives (TP), false positives (FP), false negatives (FN), and true negatives (TN). True positives refer to anomalous malware events correctly identified as attacks, whereas false positives correspond to benign activities wrongly classified as attacks. False negatives indicate actual malware events that the model missed, and true negatives represent normal activities correctly recognized as benign. Then, the accuracy is calculated using Eq. (6), precision using Eq. (7), recall using Eq. (8), and F1-score using Eq. (9):

$$Accuracy = \frac{TP + TN}{TP + FP + FN + TN} \tag{6}$$

$$Precision = \frac{TP}{TP + FP} \tag{7}$$

$$Recall = \frac{TP}{TP + FN} \tag{8}$$

$$Accuracy = 2 \times \frac{Precision \times Recall}{Precision + Recall} \tag{9}$$

4. RESULT AND DISCUSSION

This study proposes an anomaly detection model for potential malware-related attacks and applies it to the public UNSW NB-15 dataset of computer networks. The proposed model focuses on intersection analysis during the feature selection stage, using two correlation measures to determine the best features: Pearson and Kendall.

The UNSW NB-15 dataset contains 44 features, two of which are used as class labels and two as attack categories. In this study, the model used 42 basic features based on their relevance to network activity and their potential to indicate anomalies, excluding features such as 'id' and 'is_sm_ips_ports'. Correlation analysis, using Pearson and Kendall correlation coefficients, was employed to determine the optimal number of feature pairs as a correlation pair. To identify feature pairs, this study limited the maximum number to 20 pairs. The total number of feature pair combinations formed was 1,929,894,318,331. Then, from the resulting combination pairs, the model selected the best pair by testing each pair with a classification model.

The Pearson correlation analysis revealed that the best feature combination was obtained using 12 features across three classification models: Decision Tree, Naive Bayes, and k-NN. The experiment showed that the optimal classification method using Pearson correlation analysis is a Decision Tree, achieving 97.7786% recall, 85.5366% accuracy, and 80.2611% precision. The evaluation results are shown in Table 1.

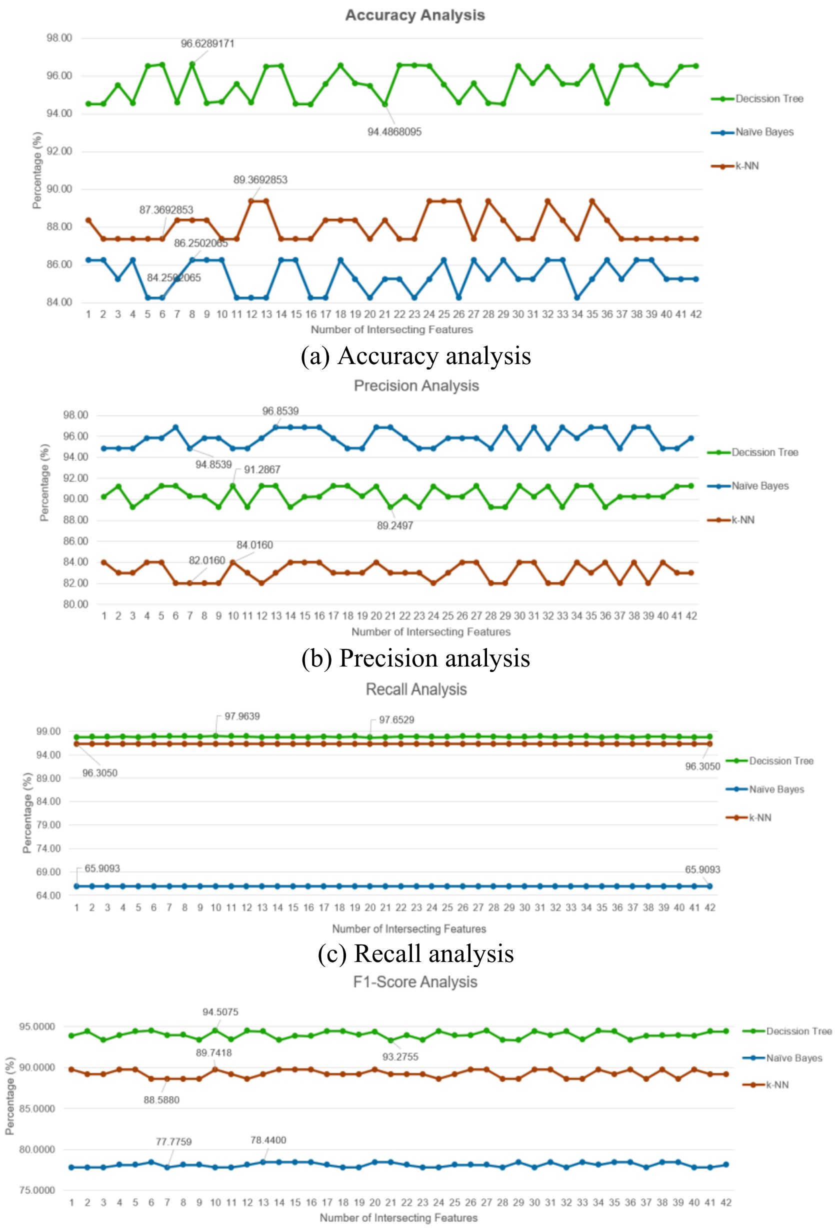
Meanwhile, Kendall's correlation produced 16 best features that were correlated with each other, as shown in Table 2. The results of the detection evaluation using Kendall's tau showed that the best classification method was the Decision Tree, with an accuracy of 85.5002%, precision of 80.2570%, and recall of 97.6992%. Both methods, Kendall and Pearson correlation, produced different numbers of correlated features, but they did overlap. The overlapping results yielded the best combination of 42 features, with each achieving classification performance values exceeding 50% for accuracy, precision, and recall. The results of the 42 features overlapping combinations between Pearson and Kendall correlation are shown in Figure 2. From the feature overlapping results from both correlation methods, the eight best features were identified: 'dmean', 'swin', 'sttl', 'dload', 'ct_dst_sport_ltm', 'dwin', 'state', and 'ct_state_ttl'. The classification results from the three models showed that the decision tree model achieved the highest performance, with an accuracy of 96.6289%, a recall of 97.9639%, a precision of 91.2867%, and an F1-score of 94.5075%. The NB classification method performed best with eight intersecting features, achieving 86.2502% accuracy, 78.44% F1-score, 97.9639% recall, and 96.8539% precision. Meanwhile, the k-NN classification method performed best with 12 intersecting features, obtaining an accuracy of 89.3693%, a recall of 96.3050%, an F1-score of 89.7418%, and a precision of 84.016%.

Table 1. Classification performance using Pearson correlation

Number of Features	Feature Name	Classification Method	Pearson Correlation		
			Accuracy (%)	Precision (%)	Recall (%)
12	['dmean', 'swin', 'sttl', 'ct_dst_sport_ltm', 'dload', 'ct_state_ttl', 'dwin', 'state', 'dtcpb', 'stcpb', 'rate', 'ct_src_dport_ltm']	Decision Tree Naïve Bayes k-NN	85.5366 75.2502 78.3693	80.2611 85.8539 73.0160	97.7786 65.9093 96.3050

Table 2. Classification performance using Kendall correlation

Number of Features	Feature Name	Classification Method	Kendall Correlation		
			Accuracy (%)	Precision (%)	Recall (%)
16	['dttl', 'dur', 'sttl', 'dload', 'spkts', 'ct_dst_sport_ltm', 'dpkts', 'ct_state_ttl', 'synack', 'dmean', 'is_sm_ips_ports', 'swin', 'dbytes', 'dwin', 'state', 'sload']	Decision Tree Naïve Bayes k-NN	85.5002 75.2502 78.3693	80.2570 85.8539 73.0160	97.6992 65.9093 96.3050



(a) Accuracy analysis

(b) Precision analysis

(c) Recall analysis

(d) F1-score analysis

Figure 2. Analysis performance classification model for 42 intersection features combination

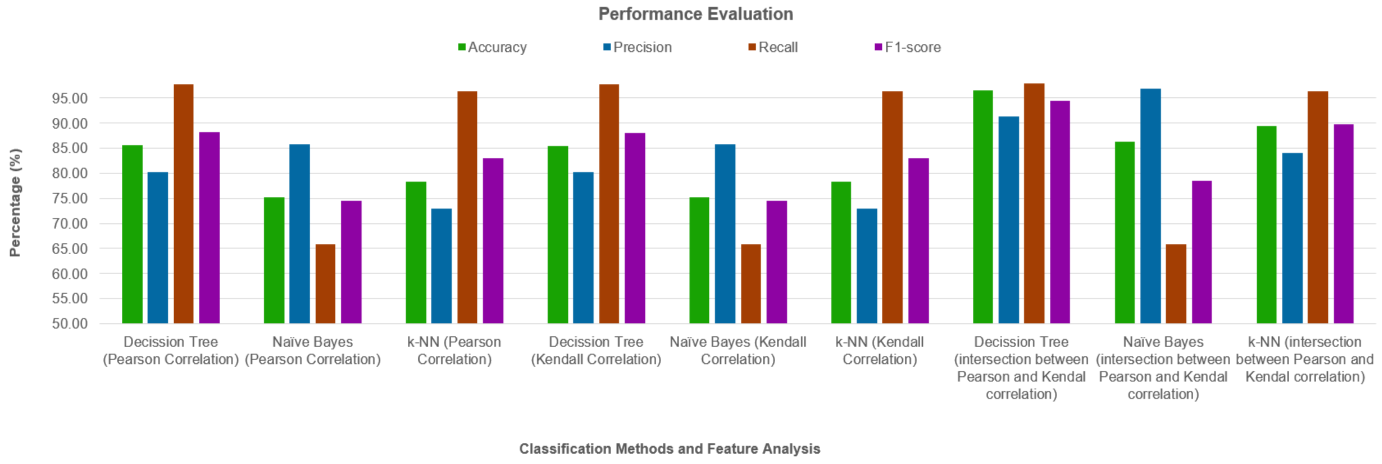


Figure 3. Comparison of classification performance between Pearson, Kendall, and intersection correlation

Table 3. Comparison of performance detection between previous studies and the proposed model

Author	Method	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)
Kocher and Kumar [47]	Linier Regression	93.23	92	99	95
	Naïve Bayes	48.03	100	23	38
	Random Forest	95.43	96	97	97
	Stochastic Gradient Descent	93.29	91	100	95
	k-NN	93.71	94	96	95
	Decision Tree	94.2	93	98	96
Elmrabit et al. [45]	Linier Regression	75.3	85.8	73.5	79.2
	Gaussian Naïve Bayes	71.6	69.3	99.7	81.8
	k-NN	82.9	85.1	88.7	86.9
	Decision Tree	88.5	91.4	90.6	91
	Random Forest	87.7	84.4	99.1	91.2
Proposed model	Decision Tree	96.63	91.29	97.96	94.51
	Naïve Bayes	86.25	96.85	65.91	78.44
	k-NN	89.37	84.02	96.31	89.74

Experimental results showed that combining the two feature correlation techniques improved classification model performance, particularly for the decision tree model. The results of each correlation method, compared with the intersection, are shown in Figure 3. Compared to the decision tree algorithm, the other two methods, k-NN and NB, performed lower due to the different number of feature intersections generated from the best features of the Kendall and Pearson correlations. Decision Tree had the best number of intersections of 8, while k-NN had the best number of intersections of 12, and NB had 7. In addition, the features generated from the intersection in the decision tree method were those with a smaller feature bias value than those obtained from the k-NN and NB methods using both Kendall and Pearson correlation techniques. The bias value referred to is the proximity of the values of each feature.

Experimental results show improved classification model performance across three models: Decision Tree, Naive Bayes, and k-NN, using the Pearson-Kendall correlation intersection analysis approach. Compared to the Pearson correlation technique, the intersection results improved accuracy by 11.0308%, precision by 11.0085%, recall by 0.0618%, and F1-score by 5.6338%. Meanwhile, the Kendall correlation method improved accuracy by 11.0429%, precision by 11.0099%, recall by 0.0882%, and F1-score by 5.6454%. Therefore, this study concludes that feature selection through feature correlation analysis and intersection analysis can improve classification model performance. In this study, the previous studies are compared with the proposed model's result using the same dataset, UNSW NB-15, as shown in Table 3.

The model proposed in this study performs better than those reported in the literature [45, 47]. In terms of accuracy, the proposed model achieves the highest classification performance with the decision tree method, at 96.63%, compared to the decision tree method [47], which obtained only 94.2%. However, the recall, F1-score, and precision values have lower values, namely 96.85% for precision, 97.96% for recall, and 94.51% for F1-score, compared to research [47], which was able to produce a precision value of 100%, a recall of 100% and an F1-score of 97%. On the other hand, the proposed model has a higher value in terms of accuracy of 96.63% and an F1-score of 94.51% with the Decision Tree method compared to research [45], which obtained the highest accuracy value of 88.5% with the Decision Tree method and an F1-score value of 91.2% with the RF method. However, the proposed model has smaller precision and recall values, namely 91.29% and 97.96% compared to research [45], which has a recall and precision of 99.7% and 91.4%, respectively.

5. CONCLUSIONS

This study proposes an intersectional analysis of two correlation methods, namely Pearson and Kendall, to identify the critical features that are correlated with each other. The intersection analysis used in the selection technique aims to improve the performance of classification methods.

Classification models are used in this paper, namely Decision Tree, NB, and k-NN, in detecting network anomalies indicative of malware attacks. The test results show that the intersection analysis of the two correlation methods improves the performance of the classification model, with an average increase in accuracy of 11.0308%, recall of 0.0618%, precision of 11.0085%, and F1-score of 5.6338% for the Pearson Correlation method. Using the Kendall Correlation method, a model was obtained with an accuracy of 11.0429%, recall of 0.0882%, an F1-score of 5.6454%, and precision of 11.0099%. The best intersections are obtained with the eight best features: 'dmean', 'swin', 'sttl', 'dload', 'ct_dst_sport_ltm', 'dwin', 'state', 'ct_state_ttl'. The best performance was achieved by the Decision Tree method, with a recall of 97.96%, accuracy of 96.63%, precision of 91.29%, and F1-score of 94.51%. The experiment results show the highest performance in evaluation accuracy and F1-score compared to the previous studies [45, 47]. The results can help network administrators facilitate and accelerate the analysis of anomalous activity.

The proposed model will be developed as future work by measuring weights based on the frequency of occurrence of resulting feature pair combinations, such as using the correlation-based Principal Component Analysis method. The future model will analyze scalability and computational complexity in subsequent research. Furthermore, time-based activity analysis will be performed using a time-windowing approach, enabling more accurate analysis and allowing real-time implementation.

ACKNOWLEDGMENT

This work was supported by Institut Teknologi dan Bisnis STIKOM Bali from the initial stages of the research through to publication.

REFERENCES

[1] Ma, W., Hou, Y., Jin, M., Jian, P. (2024). Anomaly based multi-stage attack detection method. PLOS ONE, 19(3): e0300821. https://doi.org/10.1371/journal.pone.0300821

[2] Subhi, M., Rashid, O.F., Abdulsahib, S.A., Hussein, M.K., Mohammed, S.M. (2024). A novel anomaly intrusion detection method based on RNA encoding and ResNet50 model. Mesopotamian Journal of CyberSecurity, 4(2): 120-128. https://doi.org/10.58496/mjcs/2024/011

[3] Khaleel, Y.L., Habeeb, M.A., Albahri, A.S., Al-Quraishi, T., Albahri, O.S., Alamoodi, A.H. (2024). Network and cybersecurity applications of defense in adversarial attacks: A state-of-the-art using machine learning and deep learning methods. Journal of Intelligent Systems, 33(1). https://doi.org/10.1515/jisys-2024-0153

[4] Song, J., Choi, S., Kim, J., Park, K., Park, C., Kim, J., Kim, I. (2024). A study of the relationship of malware detection mechanisms using artificial intelligence. ICT Express, 10(3): 632-649. https://doi.org/10.1016/j.icte.2024.03.005

[5] Niveditha, S., Prianka, R., Sathya, K., Shreyanth, S., Subramani, N., Deivasigamani, B., Karthikeyan, S. (2024). Predicting malware classification and family using machine learning: A cuckoo environment approach with automated feature selection. Procedia Computer Science, 235: 2434-2451. https://doi.org/10.1016/j.procs.2024.04.230

[6] Belkacem, S. (2024). Simultaneous botnet attack detection using long short term memory-based autoencoder and XGBoost classifier. International Journal of Safety and Security Engineering, 14(1): 155-163. https://doi.org/10.18280/ijsse.140115

[7] Rasheed, M.M., Faieq, A.K., Hashim, A.A. (2020). Android botnet detection using machine learning. Ingenierie des Systemes d'Information, 25(1): 127-130. https://doi.org/10.18280/isi.250117

[8] Venu, N., Kumar, A.A. (2023). An IoT based smart home automation system and safe home 2.0. Innovations, No. 74.

[9] Kumar, V., Sinha, D., Das, A.K., Pandey, S.C., Goswami, R.T. (2019). An integrated rule based intrusion detection system: Analysis on UNSW-NB15 data set and the real time online dataset. Cluster Computing, 23(2): 1397-1418. https://doi.org/10.1007/s10586-019-03008-x

[10] Dogra, A., Taqdir. (2024). Enhancing DDoS attack detection and network resilience through ensemble-based packet processing and bandwidth optimization. International Research Journal on Advanced Engineering Hub, 2(4): 930-937. https://doi.org/10.47392/irjaeh.2024.0130

[11] Putra, Z.P. (2024). Evaluating the performance of classification algorithms on the UNSW-NB15 dataset for network intrusion detection. Jurnal Ilmiah FIFO, 16(1): 84. https://doi.org/10.22441/fifo.2024.v16i1.009

[12] Rahman, J., Singh, J., Nayak, S., Jena, B., Mohanty, L., Singh, N., Laird, J.R., Singh, R., Garg, D., Khanna, N.N., Fouda, M.M., Saba, L., Suri, J.S. (2024). A generalized and robust nonlinear approach based on machine learning for intrusion detection. Applied Artificial Intelligence, 38(1). https://doi.org/10.1080/08839514.2024.2376983

[13] Sharma, N., Arora, B., Ziyad, S., Singh, P.K., Singh, Y. (2024). A holistic review and performance evaluation of unsupervised learning methods for network anomaly detection. International Journal on Smart Sensing and Intelligent Systems, 17(1): 1-38. https://doi.org/10.2478/ijssis-2024-0016

[14] Rejito, J., Stiawan, D., Alshaflut, A., Budiarto, R. (2024). Machine learning-based anomaly detection for smart home networks under adversarial attack. Computer Science and Information Technologies, 5(2): 122-129. https://doi.org/10.11591/csit.v5i2.pp122-129

[15] Ibrahim Hairab, B., Aslan, H.K., Elsayed, M.S., Jurcut, A.D., Azer, M.A. (2023). Anomaly detection of zero-day attacks based on CNN and regularization techniques. Electronics, 12(3): 573. https://doi.org/10.3390/electronics12030573

[16] Lee, J.H., Ji, I.H., Jeon, S.H., Seo, J.T. (2023). Generating ICS anomaly data reflecting cyber-attack based on systematic sampling and linear regression. Sensors, 23(24): 9855. https://doi.org/10.3390/s23249855

[17] Ko, H., Praca, I., Choi, S.G. (2023). Anomaly detection analysis based on correlation of features in graph neural network. Multimedia Tools and Applications, 83(9): 25487-25501. https://doi.org/10.1007/s11042-023-15635-z

[18] More, S., Idrissi, M., Mahmoud, H., Asyhari, A.T.

(2024). Enhanced intrusion detection systems performance with UNSW-NB15 data analysis. Algorithms, 17(2): 64. https://doi.org/10.3390/a17020064

[19] Alam, S., Alam, Y., Cui, S., Akujuobi, C. (2023). Data-driven network analysis for anomaly traffic detection. Sensors, 23(19): 8174. https://doi.org/10.3390/s23198174

[20] Rahman, R.U. (2024). A novel machine learning-based artificial intelligence approach for log analysis using blockchain technology. Sigma Journal of Engineering and Natural Sciences — Sigma Mühendislik ve Fen Bilimleri Dergisi, 1391-1409. https://doi.org/10.14744/sigma.2024.00004

[21] Ghosh, T., Bagui, S., Bagui, S., Kadzis, M., Bare, J. (2023). Anomaly detection for Modbus over TCP in control systems using entropy and classification-based analysis. Journal of Cybersecurity and Privacy, 3(4): 895-913. https://doi.org/10.3390/jcp3040041

[22] Aziz, M.N., Ahmad, T. (2021). Clustering under-sampling data for improving the performance of intrusion detection system. Journal of Engineering Science and Technology, 16(2): 1342-1355.

[23] Pu, G., Wang, L., Shen, J., Dong, F. (2021). A hybrid unsupervised clustering-based anomaly detection method. Tsinghua Science and Technology, 26(2): 146-153. https://doi.org/10.26599/tst.2019.9010051

[24] Hostiadi, D.P., Atmojo, Y.P., Huizen, R.R., Susila, I.M.D., Pradipta, G.A., Liandana, I.M. (2022). Correlation-based feature selection on botnet activity detection using Kendall correlation. In 2022 International Conference on Computer Engineering, Network, and Intelligent Multimedia (CENIM), Surabaya, Indonesia, pp. 327-332. https://doi.org/10.1109/cenim56801.2022.10037525

[25] Hostiadi, D.P., Ahmad, T. (2022). Hybrid model for bot group activity detection using similarity and correlation approaches based on network traffic flows analysis. Journal of King Saud University - Computer and Information Sciences, 34(7): 4219-4232. https://doi.org/10.1016/j.jksuci.2022.05.004

[26] Amarudin, A., Ferdiana, R., Widyawan, W. (2024). B-DT model: A derivative ensemble method to improve performance of intrusion detection system. Journal of Advances in Information Technology, 15(1): 87-103. https://doi.org/10.12720/jait.15.1.87-103

[27] Jang, J., An, Y., Kim, D., Choi, D. (2023). Feature importance-based backdoor attack in NSL-KDD. Electronics, 12(24): 4953. https://doi.org/10.3390/electronics12244953

[28] Moualla, S., Khorzom, K., Jafar, A. (2021). Improving the performance of machine learning-based network intrusion detection systems on the UNSW-NB15 dataset. Computational Intelligence and Neuroscience, 2021(1): 5557577. https://doi.org/10.1155/2021/5557577

[29] Kocher, G., Kumar, G. (2021). Analysis of machine learning algorithms with feature selection for intrusion detection using UNSW-NB15 dataset. International Journal of Network Security & Its Applications, 13(1): 21-31. https://doi.org/10.5121/ijnsa.2021.13102

[30] Umar, M.A., Chen, Z., Shuaib, K., Liu, Y. (2025). Effects of feature selection and normalization on network intrusion detection. Data Science and Management, 8(1): 23-39. https://doi.org/10.1016/j.dsm.2024.08.001

[31] Kasongo, S.M., Sun, Y. (2020). Performance analysis of intrusion detection systems using a feature selection method on the UNSW-NB15 dataset. Journal of Big Data, 7(1): 105. https://doi.org/10.1186/s40537-020-00379-6

[32] Alshamkhany, M., Alshamkhany, W., Mansour, M., Khan, M., Dhou, S., Aloul, F. (2020). Botnet attack detection using machine learning. In 2020 14th International Conference on Innovations in Information Technology (IIT), Al Ain, United Arab Emirates, pp. 203-208. https://doi.org/10.1109/iit50501.2020.9299061

[33] Hossain, M.I., Eshrak, S., Auvik, M.J., Nasim, S.F., Rab, R., Rahman, A. (2020). Efficient feature selection for detecting botnets based on network traffic and behavior analysis. In 7th International Conference on Networking, Systems and Security, Dhaka, Bangladesh, pp. 56-62. https://doi.org/10.1145/3428363.3428378

[34] Mathur, L., Raheja, M., Ahlawat, P. (2018). Botnet detection via mining of network traffic flow. Procedia Computer Science, 132: 1668-1677. https://doi.org/10.1016/j.procs.2018.05.137

[35] Hostiadi, D.P., Ahmad, T., Putra, M.A.R., Pradipta, G.A., Ayu, P.D.W., Liandana, M. (2024). A new approach of botnet activity detection models using combination of univariate and ANOVA feature selection techniques. International Journal of Intelligent Engineering and Systems, 17(3): 485-502. https://doi.org/10.22266/ijies2024.0630.38

[36] Hostiadi, D.P., Ahmad, T., Wibisono, W. (2020). A new approach to detecting bot attack activity scenario. In International Conference on Soft Computing and Pattern Recognition, pp. 823-835. https://doi.org/10.1007/978-3-030-73689-7_78

[37] Kumar, V.N., Srisuma, V., Mubeen, S., Mahwish, A., Afrin, N., Jagannadham, D.B.V., Narasimharao, J. (2023). Anomaly-based hierarchical intrusion detection for black hole attack detection and prevention in WSN. In Lecture Notes in Networks and Systems, pp. 319-327. https://doi.org/10.1007/978-981-19-8563-8_30

[38] KDD Cup 1999 Data. http://kdd.ics.uci.edu/databases/kddcup99/kddcup99.html.

[39] Khamphakdee, N., Benjamas, N., Saiyod, S. (2015). Improving intrusion detection system based on Snort rules for network probe attacks detection with association rules technique of data mining. Journal of ICT Research and Applications, 8(3): 234-250. https://doi.org/10.5614/itbj.ict.res.appl.2015.8.3.4

[40] Vinayakumar, R., Alazab, M., Soman, K.P., Poornachandran, P., Al-Nemrat, A., Venkatraman, S. (2019). Deep learning approach for intelligent intrusion detection system. IEEE Access, 7: 41525-41550. https://doi.org/10.1109/access.2019.2895334

[41] Yang, Y., Gu, Y.H., Yan, Y. (2023). Machine learning-based intrusion detection for rare-class network attacks. Electronics, 12(18): 3911. https://doi.org/10.3390/electronics12183911

[42] Bagui, S., Walauskis, M., DeRush, R., Praviset, H., Boucugnani, S. (2022). Spark configurations to optimize decision tree classification on UNSW-NB15. Big Data and Cognitive Computing, 6(2): 38. https://doi.org/10.3390/bdcc6020038

[43] Moustafa, N., Turnbull, B., Choo, K.K.R. (2019). An ensemble intrusion detection technique based on

proposed statistical flow features for protecting network traffic of Internet of Things. IEEE Internet of Things Journal, 6(3): 4815-4830. https://doi.org/10.1109/jiot.2018.2871719

[44] Debar, H., Curry, D., Feinstein, B. (2007). The intrusion detection message exchange format (IDMEF). Request for Comments (RFC), 4765. https://doi.org/10.17487/RFC4765

[45] Elmrabit, N., Zhou, F., Li, F., Zhou, H. (2020). Evaluation of machine learning algorithms for anomaly detection. In 2020 International Conference on Cyber Security and Protection of Digital Services (Cyber Security), Dublin, Ireland, pp. 1-8. https://doi.org/10.1109/cybersecurity49315.2020.9138871

[46] Sharma, N., Yadav, N., Sharma, S. (2021). Classification of UNSW-NB15 dataset using exploratory data analysis using ensemble learning. EAI Endorsed Transactions on Industrial Networks and Intelligent Systems, 8(29): 171319. https://doi.org/10.4108/eai.13-10-2021.171319

[47] Kocher, G., Kumar, G. (2020). Performance analysis of machine learning classifiers for intrusion detection using UNSW-NB15 dataset. AIRCC Publishing Corporation, 10(2): 31-40. https://doi.org/10.5121/csit.2020.102004