International Journal of
INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING
ISSN:2147-6799 www.ijisae.org Original Research Paper

Distance and Clustered Feature Selection for the Pattern Recognition of Intrusion Detection in Communication Networks

N. Chitra[1*], Dr. Safinaz S. [2]

Submitted: 25/09/2023 **Revised**: 16/11/2023 **Accepted**: 28/11/2023

Abstract: Information related to various types of sensor networks, satellite networks, and communication networks stores a large pool of data in the cloud increasing its usage. As more people access the data, traffic increases and intrusion into the system and detection are unavoidable and can be avoided using intrusion detection system (IDS) to avoid attacks. The effectiveness of IDS is improved by giving the input which has no noise or attacks so that the performance can be improvised. The feature selection concept which reduces the noise and gives the better objects can be deployed in the IDS which make the system to work smoothly in the network. The proposed algorithm Distance- clustered feature selection is based on the distance between the features and proves the effectiveness of Distance- clustered feature selection (DCFS) using C4.5 classifiers, taking the KDDCUP 99 dataset as input. Modified mutual information feature selection algorithm (MMIFSA), dynamic mutual information feature selection algorithm(DMIFSA), and redundant penalty feature mutual information algorithm (RPFMI) models are simulated using the KDDCUP 99 dataset, and their outcomes in comparison with Distance- clustered feature selection are observed. The distance- clustered feature selection outcomes, when compared with the other mutual information-based feature selection algorithms, proved to be more effective. Various performance metrics to build the distance- clustered feature selection model are evaluated and has shown better improvement in TPR and accuracy of 99.948%, indicating the proposed algorithm is effective and all the parameters are improved compared to other algorithms.

Keywords: *Distance-clustered feature selection, classifier, feature selection, mutual Information, true positive rate, accuracy, intrusion detection System, C4,5 classifier*

1. Introduction

Nowadays, society is more prone to the use of technology, and to cater to the needs, communication networks like satellites sense weather changes and store their information in the cloud. Similarly, all the communication networks, like radar, telecom, and sensor networks, depending on their necessity, store the data in storage. As a result, the flow of traffic increases as more users retrieve the data. During this process, unknown persons try to access the networks and have a hold over them, causing damage to the network [1]. So to overcome these attacks, the intrusion detection system (IDS) is designed and implemented in the system. IDS can be placed at the gateway of a network or on [2] the router so that it can alert the network regarding attacks of various types. On the grounds of the labels available the [3] learning methods are classified as 1) Supervised learning 2) Semi-Supervised learning 3) Unsupervised learning. For selecting features of discrimination and related to various class and dataset having labeled data is considered as supervised learning. Several Supervised learning methods that are proposed by authors in literature repository are studied. Some data can possess labeled and unlabeled data and the distinguished features can be identified by using the Semi-Supervised learning. Unsupervised learning can be applied wherein it does not possess any labeled data and it is complex without any reference class to reduce the data. The clustering process is used to remove the unwanted features based on the distance parameter. In clustering the decision making of the final set of objects is difficult as the objects may belong to different clusters compared to supervised learning as it has labeled data to set the condition on selection of object.

Clustering types: In this section it is discussed about the types of clustering. There are different approaches prevailing in the clustering [4] due to fact that no specific method is fixed and followed. Clustering is mainly classified as hierarchical and partition type as shown in Figure1. Further hierarchical is classified as agglomeration and divisive type. The formation of cluster starts with single object from bottom and later on moves to and merging with other features until all the objects are over is called as agglomeration clustering. Divisive method the groups are formed by splitting into smaller from top to bottom till all the objects are over. While agglomeration is bottom- top strategy and divisive is top- bottom strategy. The group formation in above mentioned can happen at a single object and slowly merging others or all the objects are taken at a time or average objects are considered termed as single,

[1] *Research Scholar, Department of Electronics and Communication Engineering, SOE, Presidency University, Bangalore, India*
E-mail: chitravadde24@gmail.com
[2] *Associate Professor, Department of Electronics and Communication Engineering, SOE,Presidency University, Bangalore, India E-mail:safinazs@presidencyuniversity.in*

complete and average. In partition clustering the data is split based on the evaluation of objects using Euclidean distance that evaluates minimum distance between objects that are available in the group. Partition clustering can be evaluated by distance using the probability, KL distance and other distance-based evaluations.

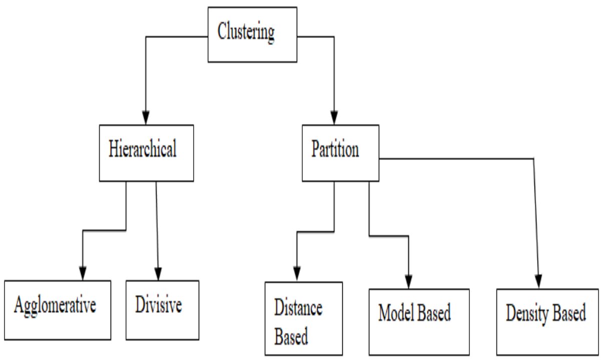

Fig.1 Classification of clustering

When there are many features in a machine, grouping related characteristics together is a technique used to minimize the number of features. The grouping of features is more advantageous and minimizes the dimensional features when attempting to reduce the enormous dataset. Estimation variance is decreased, and feature selection stability is increased. To study performance and choose the best qualities, the data is grouped. The clustering approach is used to divide the data into several groups with related data. The number of characteristics is chosen based on the choice of grouping or clustering. There are two ways to cluster data: hard clustering and soft clustering.

The group's characteristics are unique to that group and do not apply to others. The breakup of partition clustering is as follows

1) K-Means Clustering

2) PAM (partition around Medians)

K-Means Calculates when the target class is unavailable, clustering is utilized for unsupervised learning and identifies the groups based on the cluster group indicated by K times. After iterations are finished, or if the centroid has not changed even after iterations, the feature closest to the centroid is assigned.

2. Literature Survey

In DMIFS [5] information content between variables and class labels is used for finding the best-performing variables, and the algorithm is a straight-forward feature selection method. MMIFS [1] proposed by Jing Ping Song et al. uses the KDDCUP99 dataset to design the IDS model. It is an improvisation of DMIFS and introduces the second set of features. It showed an increase in accuracy over DMIFS. The RPFMI proposed by Fei Zhao et al. [6] is based on Batti's algorithm [7]. The computation is done based on the relation between the selected features and the given features [8]. By doing so, the good features contributing to feature selection are selected, increasing accuracy. A normalized MI feature selection algorithm is proposed by [9,10, 11] Mohammad A. Ambusaidi et al. describe an enhancement of the MMIFS(Modified Mutual information feature selection), MIFS(Mutual information feature selection), and Battis [9] algorithms. The algorithms differ only in selecting a second set of features. In this one, the ratio of mutual information of features (f_i) and first selection features (f_s) to the entropy of features is adapted as the stopping function. In NMIFS (Normalized mutual information feature selection) [10], there is no need to set the redundancy parameter beta as required in MMIFSA, RPFMI, and Batti algorithms. In all the above algorithms, information content between the features, called mutual information content, is used to find the relation between the features. The science that quantifies the extent of the[5,6,7,8,9] relationship between the variables is the information theory. Mutual information relates the variables and gives the uncertainty of their occurrence. Assume X and Y are random variables related to the probability of discrete or continuous events. The random variables contain the samples$X = (x_1, x_2, x_3 \cdots\cdots x_N) Y = (y_1, y_2, y_3 \cdots\cdots y_N)$ where $x_i$and $y_i$are the elements of X and Y, respectively. Shannon Entropy H(X) and H(Y) (Equ.1, Equ.2) is the average information [9] content of the random variables for discrete data.

$$\mathrm{H(X)} = -\sum_{\mathrm{i=i}}^{\mathrm{N}} \mathrm{P(x_i) log\ P(x_i)} \qquad (1)$$

$$\mathrm{H(Y)} = -\sum_{\mathrm{i=1}}^{\mathrm{N}} \mathrm{P(y_i) logP(y_i)} \qquad (2)$$

Where P (x_i) =possible number of instants of x_i/Total number instants (N).

P (y_i) =possible number of instants of y_i/Total number of instants (N)

$$\mathrm{H(X|Y)} = \sum_{\mathrm{x \in X}} \sum_{\mathrm{y \in Y}} \mathrm{P(x, y) logP(x|y)} \qquad (3)$$

Where P(x, y) in Equ.3 is the joint distribution function of x and y. H (X|Y) is the joint entropy (equ.3) of X when Y variable is given and P(x |y) is conditional probability of x with y is given.

$$H(X, Y) = H(X) + \mathrm{H(Y|X)} \qquad (4)$$

$$\mathrm{H(X, Y)} = \mathrm{H(Y)} + \mathrm{H(X|Y)} \qquad (5\ )$$

H(X,Y) in Equ.4 and Equ.5 is the joint entropy of X and Y variables and is used to find the information content between variables X and Y

$$I(X,Y) = H(X) - H(Y|X) \qquad (6)$$

$$\mathrm{I(X,Y)} = H(Y) - H(X|Y) \qquad (7)$$

$$\mathrm{I(X;Y)} = -\sum_{y\epsilon Y}\sum_{x\epsilon X} \mathrm{P(x,y)} \log \frac{\mathrm{P(x,y)}}{\mathrm{P(x)P(y)}} \qquad (8)$$

I(X,Y) Equ.6, Equ.7, Equ.8 is the information content between the random variables X and Y. Mutual information [1,2,5,6,7, 9,13] of X and Y variables is defined as the extent of information contained in both X and Y. The more the information content between them MI will be maximum and both the variables are said to be closely related to each other.

Correlation

Let x(n) and y(n) are the sequences of the data given then cross-correlation between them is given by [17,18] as C(x(n), y(n)) $=\mathrm{C}\big(\mathrm{x\ (n); y(n)}\big) = \sum_{n=-\infty}^{\infty} x(m)y(m+n)$. (9)

The correlation between the sequences x(n) and y(n) is represented as C(x(n),y(n)) as shown in Euq.9If the correlation between two sequences are maximum then they are more correlated and if correlation is less than they are not closely related. Further the concept of correlation is used and modified by [16] to find the best features as shown in Equ.10

$$\mathrm{M_s} = \frac{\mathrm{KC_{fc}}}{\sqrt{\mathrm{K + K(K-1)C_{ff}}}} \qquad (10)$$

$\mathrm{C_{fc}}$ is the correlation between the feature and class

$\mathrm{C_{ff}}$ is the correlation between the subset features and K is number of features.

In PAM the process is similar to K-Means but the difference is the centroid in the cluster is the input feature but in K-Means the average of feature data does not belong to input feature data. In K-Means clustering the centroid is selected by the expectation maximization and objective function in Equ.11 ruling is given as

$$\mathrm{J} = \sum_{i=1}^{m}\sum_{k=1}^{k} W_{ik} \ ||\, x_i - \mu_k||^{\,2} \qquad (11)$$

W_{ik} =1 and $\mathrm{x_i}$ is the data and μ_k is centroid which is average, $||\, x_i - \mu_k||$ will be evaluated till it gives any change.

3. **Proposed Algorithm:**

The relation between feature and class is evaluated using correlation between them. If the value of the correlation is maximum then those features are closely related to class and the minimum value is considered as weakly related to class. Highest valued one is selected and also called as selected feature and they are put into set S. Then the remaining features are called as instant features $\mathrm{f_{i.}}$ Sometimes weakly related features may also be useful for the sake of finding exact relationship between features the distance between the remaining features, selected feature and class is evaluated. This set contains redundant features which do not convey any information and to eliminate them the distance between variables of the data is given as input to the K-Means clustering algorithm.

Algorithm1: Distance Clustered feature selection (DCFS)

Consider the input dataset as D=T (F, C)

Output: Final set of features as selected features

Let $S \leftarrow \varphi, f_i \leftarrow \text{instant features}\ ,\ f_s \leftarrow selected\ features$

F ←set of features, C ← $Class\ label f_i = F - f_s$

1. Initialize set S=φ (null set)

2. Evaluate the correlation between the feature and class. Find the maximum values feature as first set of features.

3. The number of features will be reduced and the present features are $f_i = F - f_s$

4. For the evaluation of the next set of features find the distance between the first set of features, class and the instant features.

$\mathrm{Dis\ (f_s\ , C, f_i)} = \sqrt{\mathrm{f_s * C}} + \sqrt{\mathrm{C * \ f_i}} + \sqrt{\mathrm{f_i * f_s}}$

The final value is calculated as $\mathrm{D_{f_s f}} = \frac{1}{\mathrm{Dis(f_s\ ,C\ ,\ \ f_i)}}$

5. The minimum value of $\mathrm{D_{f_s f}}$ is calculated and applied for K-Means clustering for k=3 to find the second set of features

6. Evaluate the final set of features $\mathrm{f_{final}}$= $\min \left[\frac{\mathrm{Dis(f_s\ ,f_i,C)}}{\sum \mathrm{corr(f_s,C)}}\right]$

Where $\mathrm{corr(f_sC)}$ is the correlation of selected feature and class and $\mathrm{F_{final}}$ are final set of features

7. Final set of features

The experimental setup is organized as:

1) Selection of Set of features which are high information content

2) Selection of Feature set that is next level of information content

3) Model building

4) Comparison of the Performance parameters

The experiment is conducted for feature selection of best features that reduces the classification time. It is done in two steps. Firstly, the high information content between class type and each feature is evaluated and highest valued feature is put into the empty set S. Secondly proposed algorithm is applied to find second set of features. The features which are left after first set of features also can be related to the features, so the left over are applied the distance formula with f_s(selected features) and left over features are called as f_i and the minimum value of it is considered and taken as second set of features. This is forward feature selection supervised selection method. The model build using DCFS and performance of parameters is done. Similarly, second set of features for other algorithms are done and models are built and compared. After classification the instances the status of correct Prediction of attacks and mistaken prediction of normal data as attacks and various possibilities are given in the matrix form called as confusion matrix. Table.1 gives the distribution of the favored and unfavored predictions of the objects.

Table 1. Confusion Matrix

Class	Predicted outcomes as Negative Class	Predicted outcomes as Positive Class
Actual outcomes as negative class	True Negative (TN)	False Positive (FP)
Actual outcomes as Positive class	False Negative (FN)	True Positive (TP)

4.2 Performance metrics [1][2][15][18]:

True positive (TP):

Those instances correctly predicted as not attacks are called true positives.

False Positive (FP):

Predicting the instances incorrectly as attacks even though they are not attacks is said to be False positive

True Negative (TN): As the number of attacks increases, those instances are predicted as attacks and are said to be True Negative.

False Negative (FN): The number of attacks that occur and the number of incorrectly predicted as no attack samples are known as False Negative.

Detection rate (DR) or TPR (true positive rate): Ratio of instances that are correctly predicted as no attack to the sum of all samples of correct predictions and wrong predictions

DR or TPR = TP/(TP+FN)

False-positive rate (FPR): The ratio of samples that are predicted as attacks to the quantity of all attacked samples. This metric reflects the ability to identify incorrectly predicted attacks. The FPR reflects the detection rate. The FPR value is usually low for the best intrusion detection models.

FPR=FP/ (TN+FP)

Precision (P): The proportion of samples that are truly predicted as attacks to all samples that are attacked. It relates the ratio of the correct classification to the incorrect classification.

P=TP/(TP+FP)

Accuracy: The ratio of samples [14] that are correctly predicted to all predicted values. This means it gives the overall evaluation and returns the proficiency of the detection model in detecting the attacks and differentiates between an attack and no attack.

Accuracy = (TP+TN)/(TN+TP+FN+FP)

Sensitivity or recall (R): the proportion of instances that are correctly identified to the sum of correctly identified and misidentified instances.

R = TP/(TP+FN)

F-measure: The average of the harmonics of precision and sensitivity is F-measure.

F= 2*P*R/(P+R)

4. Dataset:

The KDDCUP 99 dataset [10] is built based on 7 weeks of network traffic, which consists of 5 million connections. It has 22 attacks, 41 features, and a total of 494021 records. The records fall into the following categories:1.Denial-of-service (DoS) attack: The attacker consumes the user's memory or overburdens computing resources, or illegally denies the user access to the user's resource.2.User-to-Root attack (U2R): The attacker starts to gain access to the normal user account by taking control over the resource by stealing the password and can illegally access the user.3Remote-to-local attack (R2L): The attacker can send the data packets on the network by gaining the access of the user of that.4Probing attacks: This attack pertains to gathering information about the topography, architecture, and security of the communication network for future.

The algorithm is based on the distance evaluation between the features and KDDCUP 99 dataset is given as input which contains 41 features, 23attacks as class. The class is an attack happened on the network to disrupt the usual functioning of network. Actually 23 types of attack happened on the network and it is as whole summarized as single class having normal and attacks as two class .The models using MMIFS, REMIFSA, DMIFS and RPFMI are built using C4.5 Decision tree classifier and the performances are compared. Fei Zhao et.al [4] RPFMI has built using DOS classifier whereas in this paper it is evaluated using C4.5classifier.Performance metrics are evaluated for Anomaly and Normal class separately and results are compared. The Clustering algorithm is used in the feature reduction process to assess the ability of dataset. Features having the same parameters are grouped together and it becomes easy for identifying the best object. First the correlation between the class and features are performed. Features 3,5,23 in Fig.2 shown to possess the maximum value of correlation. Those features are selected features and play major role in selecting other features for contribution of the better performance of the detection model**.**

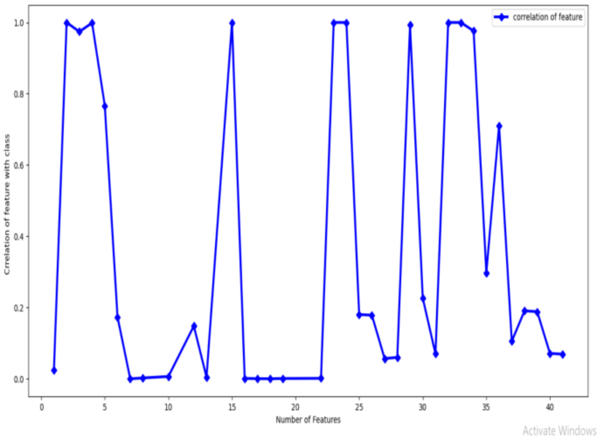

Fig.2. Correlation of features

The features after removing the selected features are divided into first half of features as 1,2,3,4,5,6,7,8,10,12,15,16,17,18,19,21,23 is given as input to K-Means clustering algorithm to get the best features as shown in Figure 3. K-Means algorithm helps in simplifying the problem by giving the features having same features as a group. The data is split into three clusters centered on the centroid, as seen red in Fig. 3. Three centroid with values of 0.59, 0.78, and 0.81 form a cluster with similar inherent properties. The clusters having more distanced features, moderate distanced features, and lowest distanced features form separate clusters. Less distant features are selected, and 1, 2, 3, 4, 5, 6, 14, 15, and 23 are selected from the first half of features.

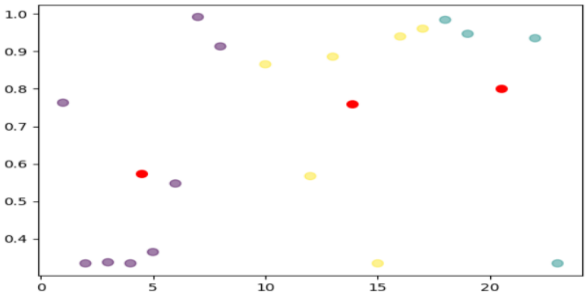

Fig.3. K-Means clustering of first half of features

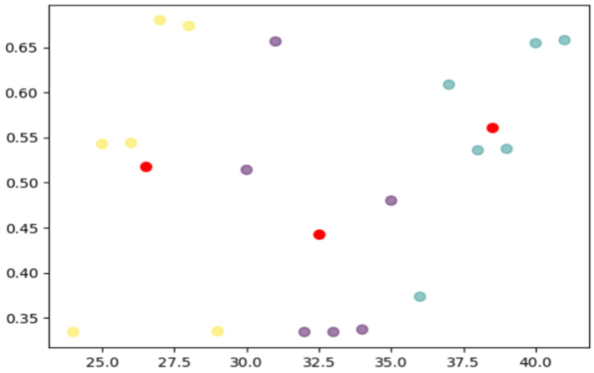

Fig.4. K-Means clustering of second half of features

Second half of features as 24,25,26,27,28,29,30,31,32,33,34,35,36,37,38,39,40,41 is given as input to the K-Means cluster algorithm for the sake of simplification is in seen in Fig. 4. Three clusters with centroids in red color are formed around 0.52,0.45,0.56 and different clusters of similar characteristics are seen and second set of features as 24, 29,31,32,33,34,36,38,39 having the lowest values surrounding the centroid are selected.

The drastic improvements in the true positive rate of the model as the number of features reach to 10 features as shown in Table 2. With features 2,3,4,5 the proposed model detection accuracy has reached to 99.895%. When the feature 15 is included in the dataset and it is contributing more for the system later on its accuracy reached to 99.9385%.

Table 2: Simulation of DCFS IDS model

Features	Anomaly		Normal		Accuracy (%)
	TPR (%)	FPR (%)	TPR (%)	FPR (%)	
4	98.84	1.15	99.11	0.88	99.895

5	97.45	0.02	99.77	0.2	99.9385
6	97.45	0.02	99.77	0.2	99.9385
7	97.44	0.03	99.87	0.1	99.9460
8	99.96	0.03	99.86	0.1	99.9466
9	99.96	0.03	99.96	0.1	99.946
10	99.95	0.04	99.90	0.1	99.948

As the system is given feature 23 the Intrusion detection system-maintained accuracy as consistent that indicates feature 23is redundant one. Feature 24 and 29 could improve the performance so they are considered more contributing features and finally features 32 and 33 also contributed to the building the model though they seem to be less correlated to the class label. To find the final set of features the f_{final} is evaluated from the Distance between the selected, instant features and class and sum of Distance of selected feature and class.

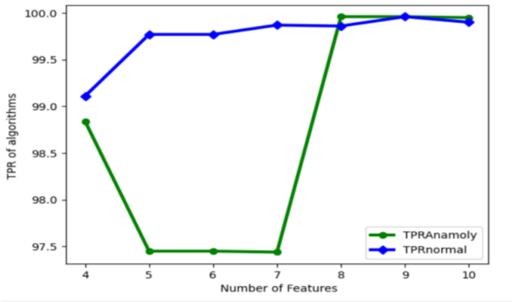

Fig.5. TPR of Anomaly and Normal

As seen in Fig. 5, the TPR of the anomaly class increases as the number of features increases and maintains consistency, starting at 8 features and increasing to 10 features, indicating the capability of identifying the positive incident correctly, and it increases drastically. As shown in Fig.5, the TPR of the anomaly class increases as the number of features increases and maintains consistency, starting at 8 features and increasing to 10 features, indicating the capability of identifying the positive incident correctly, and it increases drastically. In Fig. 6, the FPR of the anomaly is shown, indicating the proposed method has the capability to reduce false predictions and the normal class maintains the steady phase of the FPR.

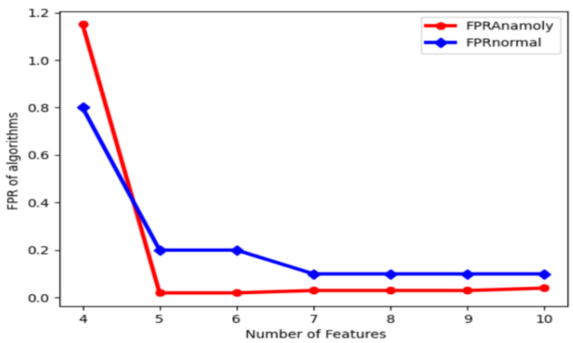

Fig.6. FPR of DCFS for Anomaly and Normal

Table .3 Comparison of algorithm

Algorithms	**Class**	**TPR (%)**	**FPR (%)**	**Accuracy (%)**
C4.5	Anomaly	99.9	0.7	99.90
	Normal	99.3	0.6	
DCFS+C4.5	Anomaly	99.95	0.04	99.948
	Normal	99.90	0.1	
MMIFS+C4.5	Anomaly	99.82	0.17	99.77
	Normal	99.56	0.43	
DMIFS+C4.5	Anomaly	99.96	0.03	99.94
	Normal	99.88	0.11	
RPFMI+C4.5	Anomaly	98.77	1.2	98.946
	Normal	99.86	0.33	

Accuracy of the proposed and other algorithms are compared in Table.3 and observed that TPR and FPR are the better values as compared to the methods indicating the performance metrics are better and the stability of the system is more improved which is designed with distance-based algorithm.

5. Conclusion:

This work presents a DCFS algorithm based on distance metric filter and embedded method that evaluates without any reference values. This method evaluates using class and without class and it is termed as Semi-supervised learning. The proposed method evaluates the correlation between features and subset features, which eliminates unnecessary features. The dataset used is the KDDCUP 99 dataset, and DMIFSA, MMIFSA, REMIFSA, and RPFMI are modeled and performance metrics are evaluated. TPRs accuracy is observed to be improved over other algorithms. TPR is quite better than others, which indicates the model's capacity to detect the attacks correctly. If the value of incorrectly predicted attacks is very low, it improves accuracy, TPR, and other metrics. This feature selection method achieved the maximum TPR and FPR (less) values for almost all features. In the future, this work can be extended by using the distance method or correlation method for Unsupervised feature selection. This can also be extended to other strategies for feature selection. Other datasets can also be applied, and their model performance can be observed.

Conflicts of Interest

The authors declare no conflicts of interest.

References

[1] Ambusaidi, Mohammed A. "Using Mutual Information for Feature Selection in a Network Intrusion Detection System." *The Third International Conference on Digital Security and Forensics (DigitalSec2016)*. 2016.

[2] Amiri, F., et al. "Yazdani. N,“Mutual information-based feature selection for intrusion detection systems,”." *Journal of network and computer applications* 34.4 (2011): 1184-1199.

[3] Battiti, Roberto. "Using mutual information for selecting features in supervised neural net learning." *IEEE Transactions on neural networks* 5.4 (1994): 537-550.

[4] Estévez, Pablo A., et al. "Normalized mutual information feature selection." *IEEE Transactions on neural networks* 20.2 (2009): 189-201.

[5] Estévez, Pablo A., et al. "Normalized mutual information feature selection." *IEEE Transactions on neural networks* 20.2 (2009): 189-201.

[6] Estévez, Pablo A., et al. "Normalized mutual information feature selection." *IEEE Transactions on neural networks* 20.2 (2009): 189-201.

[7] Cup, Kdd. "Intrusion Detection Dataset Task Description." *University of California Department of Information and Computer Science* (1999).

[8] Peng, Hanchuan, Fuhui Long, and Chris Ding. "Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy." *IEEE Transactions on pattern analysis and machine intelligence* 27.8 (2005): 1226-1238.

[9] Bharill, Neha, and Aruna Tiwari. "A review of clustering techniques and developments." (2017).

[10] Song, Jingping, et al. "Modified mutual information-based feature selection for intrusion detection systems in decision tree learning." *Journal of computers* 9.7 (2014): 1542-1546.

[11] NagaJyothi, Grande, and Sriadibhatla SriDevi. "Distributed arithmetic architectures for fir filters-a comparative review." *2017 International conference on wireless communications, signal processing and networking (WiSPNET)*. IEEE, 2017.

[12] Song, Jingping, Zhiliang Zhu, and Chris Price. "Feature grouping for intrusion detection system based on hierarchical clustering." *Availability, Reliability, and Security in Information Systems: IFIP WG 8.4, 8.9, TC 5 International Cross-Domain Conference, CD-ARES 2014 and 4th International Workshop on Security and Cognitive Informatics for Homeland Defense, SeCIHD 2014, Fribourg, Switzerland, September 8-12, 2014. Proceedings 9*. Springer International Publishing, 2014.

[13] Karthick Raghunath, K. M., et al. "Utilization of IoT-assisted computational strategies in wireless sensor networks for smart infrastructure management." *International Journal of System Assurance Engineering and Management* (2022): 1-7.

[14] Avanija, J., et al. "Designing a Fuzzy Q-Learning Power Energy System Using Reinforcement Learning." *International Journal of Fuzzy System Applications (IJFSA)* 11.3 (2022): 1-12.

[15] Zhao, Fei, et al. "A filter feature selection algorithm based on mutual information for intrusion detection." *Applied Sciences* 8.9 (2018): 1535.

[16] Kartika S. (2016). Analysis of “SystemC” design flow for FPGA implementation. International Journal of New Practices in Management and Engineering, 5(01), 01 - 07. Retrieved from http://ijnpme.org/index.php/IJNPME/article/view/41

[17] Mohammad Hassan, Machine Learning Techniques for Credit Scoring in Financial Institutions , Machine Learning Applications Conference Proceedings, Vol 3 2023.