TEST Engineering & Management

May-June 2020
ISSN: 0193-4120 Page No. 3889-3899

Hybridization of Ruzchika MAP Estimated Naïve Bayes and Chicken Swarm optimization based dichotomous Regression Classifier for Intrusion Detection

[1]A.Shanthi sona, [2]N. Sasirekha

[1]Assistant Professor, PG and Research Department of Computer Science, TiruppurKumaran College for women, Tirupur, India, shanthisona@yahoo.in

[2]Associate Professor, Department of Computer Science, Vidyasagar College of Arts and Science, Udumalpet, India, nsasirekhaudt@gmail.com

Article Info
Volume 83
Page Number: 3889-3899
Publication Issue:
May-June 2020

Abstract

Intrusion detection is the method of observing and analyzing the events occurred in a network in order to solve the security problems. With the extensive growth of the network, the entire computer suffers from security vulnerabilities. Therefore the Intrusion Detection System (IDS) plays a major role in identifying the anomalies or attacks in the network. In order to improve the Intrusion Detection accuracy, Hybrid Ruzicka Naive Bayes Chicken Swarm Optimized Feature Selection and Dichotomous Regression Classifier (HRNBCSOFS-DRC) model is introduced. The main objective of HRNBCSO-DRC model is to identify the intrusion through the optimal feature selection and classification. The hybrid technique starts with the initialization of Chickens (i.e., features) populations in the search space. Then the fitness of each chicken is calculated based on the Ruzicka similarity measure to identify the current best. Based on the fitness value, the chickens are ranked.. Then the roulette wheel selection technique is applied to each group for choosing the attributes with higher fitness by applying the MAP estimated naïve Bayes probabilistic rule. The chicken's with higher fitness is considered as rooster and the minimum fitness are considered as chicks and the remaining chickens are considered as hens. After finding the best feature, the rooster position gets updated along with the position of other chicken's (i.e. hens and chicks) .This process gets iterated until a termination condition is met. Followed by, the hybrid model uses Dichotomous Regression function to analyze the selected feature value (i.e. training data) with the testing data. Then the data is classified as normal or abnormal based on the correlation coefficient value.. Experimental evaluation is performed with NSL-KDD dataset using different metrics such as Intrusion detection accuracy, precision, recall, F-measure and Intrusion detection time.

Article History
Article Received:*19 November 2019*
Revised: *27 January 2020*
Accepted: *24 February 2020*
Publication: *12 May 2020*

Keywords: *Intrusion detection, Feature selection, Ruzicka similarity, Chicken Swarm Optimization, MAP estimated naïve bayes probabilistic rule, Dichotomous Regression function, and classification.*

1. Introduction

Intrusion Detection Systems (IDS) is essential in network security infrastructure to monitor and recognize the unwanted and malicious (i.e. unauthorized system access). Due to the complex and time-varying network, conventional techniques are difficult to extract features of intrusion behaviour from the high-dimensional data set and processing of such kinds of data leads to high false detection rates. Security has become a major concern with the extensive usage of network- based services. Intrusion creates a serious risk to network security. Intrusion is a kind of attack or interruption that affects the security mechanism. Therefore, an efficient intrusion detection system (IDS) provides the solution to prevent the network intruders for effective communication. The intrusion detection system is a security management system that helps to monitor the network for detecting the malicious actions. Based on the motivation, a machine learning technique is integrated with IDS to identify the normal or anomalous behaviours with lesser time. Therefore, the desirable features are analyzed and selected by the intrusion detection system to minimize the dimensionality and false detection.

Paper Outline

The rest of this paper is organized into five different sections. Section 2 elaborates the issues and challenges of intrusion detection in the related works. Section 3 describes the HRNBCSOFS-DRC model for intrusion detection based on feature selection and classification. In section 4, experimental evaluation is carried out with dataset and the performance results of various metrics are discussed in section 5. Finally, section 6 provides the conclusion of the work.

2. Related Works

A Deep Belief Network-based Intrusion Detection System (DBN-IDS) was developed in [1] using particle swarm optimization. But, the intrusion detection accuracy was not improved. An Adaptive network intrusion detection method based on a selective ensemble of kernel extreme learning machines with random features(ANID-SEoKELM) was developed in [2]. Though the method achieves higher detection accuracy, the intrusion detection time was not minimized.

An anomaly-based IDS for hierarchical data was developed in [3] to minimize the false positive rate (FPR) of intrusion detection. But the designed method failed to increase the quality of discovering the intrusion detection. An artificial neural network was developed in [4] to detect the intrusion with higher accuracy. The designed method failed to select the optimal features for minimizing time complexity in intrusion detection.

A Text-Convolution Neural Network and Random Forest-based intrusion detection (TR-IDS) technique was developed in [5] with statistical features. But the performance of intrusion detection accuracy and other parameters remained unaddressed. An artificial neural network-based intelligent intrusion detection system was developed in [6] with less number of features. But the system failed to use the fast converging learning algorithms for achieving the accurate detection rate.

A PSO optimization-based fast learning network was introduced in [7] for intrusion detection. But the technique provides less accuracy for a certain number of classes due to the limited amount of training data. The hybrid feature selection and two-stage meta classifier were developed in [8] for intrusion detection. The performance of intrusion detection time was not minimized using a hybrid technique.

Efficient intrusion detection using a hybrid data optimization technique was developed in [9] with minimum time cost. The designed technique failed to achieve higher detection accuracy. A hybrid classification method named Artificial Bee Colony (ABC) and Artificial Fish Swarm (AFS) algorithms were developed in [10] for increasing the intrusion detection rate. Though the method minimizes the computational complexity, the performance of precision and recall remained unsolved.

An improved PCA integrated with Gaussian Naive Bayes technique was developed in

[11] for feature selection and classification to quickly detecting the intrusion behaviour. Though the method minimizes the detection time, the false positive rate was not minimized. A nonparametric Bayesian model was developed in [12] for unknown intrusions (or anomaly detection). The model failed to handle larger scale IDS-based datasets to offer

A Bayesian-based Markov chain Monte Carlo (MCMC) was developed in [13] for anomaly intrusion detection. But the designed method failed to handle larger scale datasets to offer a detection system. Sparse Logistic Regression model was introduced in [14] to select the more relevant features for intrusion detection. The various classification techniques were not used for accurately detecting the intrusion.

An efficient SVM ensemble with feature augmentation method was developed in [15] for intrusion detection. But the performance of various accuracy parameters such as true positive, false negative rate remained unaddressed. A hybrid classification and feature selection were introduced in [16] for intrusion detection. The method failed to achieve a better accuracy rate using the hybrid method.

A wrapper approach based feature selection and classification were introduced in [17] for network intrusion detection. The designed approach failed to increase the accuracy of a classification and minimize the misclassified instances. A Feed forward deep neural networks (FFDNNs) combined with a feature selection developed in [18]. Though the method improves the accuracy, the performance of classification time remained unaddressed.

A support vector machine and extreme learning machine was developed in [19] to increase the efficiency of identifying the attack. The method failed to construct more effective classifiers for classifying attacks with higher performance. A Group-wise Principle Component Analysis (GPCA) technique was introduced in [20] for intrusion detection. However, the overall performance of intrusion detection was not increased.

The most important issues identified from the above-said literature are overcome by introducing a novel technique. The brief explanation of HRNBCSOFS-DRC model is presented in the next section.

3. Methodology

The HRNBCSOFS-DRC model is developed for malicious behaviour (i.e. intrusion). The hybrid model accurately performs intrusion detection by selecting the optimal features. The relevant feature selection process in HRNBCSOFS-DRC model minimizes the amount of time taken by an algorithm to detect the malicious behaviour in the network. Compared to existing chicken swarm optimization, the proposed HRNBCSOFS-DRC model uses Ruzicka similarity for fitness calculation and roulette wheel selection technique for finding the optimal features. The HRNBCSOFS-DRC model also uses Map estimated naïve Bayes probabilistic rule than the conventional naïve Bayes for optimal feature selection. In addition, the hybrid model also used dichotomous regression function to classify normal or abnormal data. The architecture diagram of the proposed HRNBCSOFS-DRC model is shown in figure1.

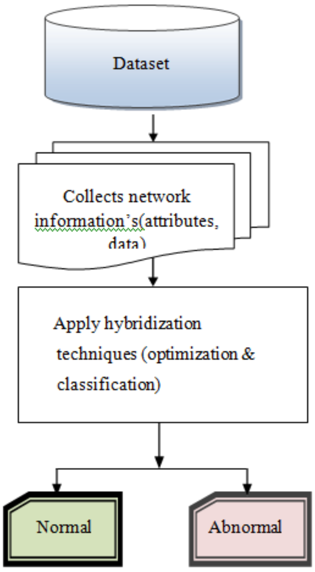

Dataset
Collects network information's(attributes, data)
Apply hybridization techniques (optimization & classification)
Normal
Abnormal

Figure 1: HRNBCSOFS-DRC Model

Fig.1 shows the architecture diagrams of the proposed model to identify the intrusion in the network by applying the hybridization technique. Initially, the intrusion detection dataset $(D_{ids})$is considered. The network information's such as attributes (i.e. features) A1, A2, A3, … An and their data D_1, D_2, D_3, … . $D_n$are collected from the dataset. After collecting the data, the hybridization technique initially performs the feature selection
with classification to improve the intrusion detection accuracy in a network.

The proposed HRNBCSOFS-DRC model uses the novel bio-inspired algorithm called Chicken Swarm Optimization (CSO). The Chicken Swarm Optimization is the meta heuristic which provides better solutions with minimal computational effort.CSO algorithm is types of random search algorithm and the whole chicken

= {A1, A2, A3, … An} ∈ D_{ids} (1) Where, A_i is the set of the attributes
{A1, A2, A3, …An} in the intrusion detection
datasetD_{ids}. After the initialization, the fitness of each attribute in calculated. On the contrary to the existing CSO, the proposed algorithm uses the Ruzicka similarity for calculating the fitness of the each individual i.e. attributes to select the more
as optimal for intrusion detection. Therefore, the similarity is mathematically calculated as follows,

$$f_c = \beta = \frac{A_i \cap O_v}{\sum A_i + \sum O_v - A_i \cap O_v} \quad (2)$$

Where, represents a fitness of the attributes, β represents the Ruzicka similarity coefficient, A_i represents attributes in the dataset,$O_v$denotes an objective function, $A_i \cap O_v$is a mutual dependence between the attributes and a objective function,$\sum A_i$ is the sum of$A_i$score,$\sum O_v$is the sum of O_v score. The Ruzicka similarity coefficient provides the similarity value between 0 and 1. Based on the similarity value, the fitness of each attributes is calculated.
After computing the fitness, the attributes are ranked based on fitness. The whole chicken swarm population is partitioned into a number of subgroups. Each subgroup comprises the rooster, several hens, and chicks. For each group, the best individual is chosen based on fitness using the roulette wheel selection technique. On the contrary existing CSO, the proposed algorithm uses the MAP estimated naïve Bayes probabilistic rule to accurately find the higher priority features which results minimize the intrusion detection time.

Let us consider a circular wheel. The circular wheel is divided into 'n' number of segments, where 'n' is the number of attributes in the population. In this selection, all the attributes in the population are positioned on the roulette wheel with their fitness value. The selection of the best individual is shown in fig. 2.

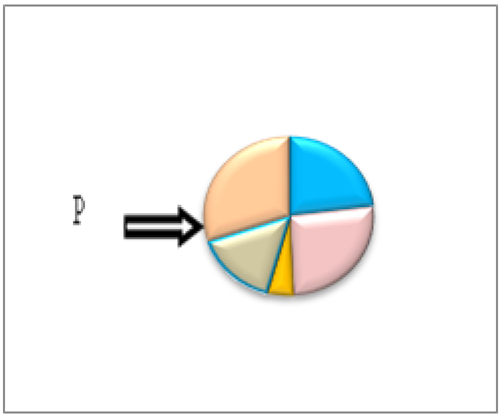


Figure 2: Roulette wheel based feature selection

Fig.2 shows the best feature selection using a roulette wheel with the help of the wheel pointer. From the figure, the different color segment indicates the fitness value of the different attributes. Then the roulette wheel is rotated. The individual of the wheel which comes facing the wheel pointer(P) is selected as an attribute with higher fitness. The wheel pointer uses MAP estimated naïve bayes probabilistic rule for selecting the best individuals with maximum probability from the population. The maximum probability of the attribute selection is expressed as follows,

$$W = \arg\max\, P\,(A_i|O_v) \quad (3)$$

$$P\,(A_i|O_v) = \frac{f_i}{\sum_{j=1}^{n} f_j} \quad (4)$$

In (3), (4),Wdenotes a selection output, $\arg\max$ denotes a argument of a maximum function using MAP rule, $P\,(A_i|O_v)$ is the selection probability, f_i denotes a fitness of individual 'i' in the population 'j', 'n' is the number of attributes in the population. Based on the MAP estimated naïve bayes decision rule, attributes with higher fitness has a maximum probability being selected for the intrusion detection. The higher fitness chickens are selected as roosters and the minimum fitness values is designated as chicks, the rest of the chickens treated as hens. All the chickens (i.e. hens and chicks)in the subgroups followed their rooster to search their food. The movements of all the chickens are updated based on the roosters with higher fitness values. Since, the rooster with better fitness has higher priority for accessing the food than the other ones with lesser fitness values. Therefore, the position of the rooster is updated as follows,

$$p_r(t+1) = p_r(t) * (1 + r\, n\,(0, d^2)) \quad (5)$$

Where, $p_r(t+1)$ denotes an updated position of the rooster, $p_{i,j}\,(t)$ denotes a current position of the rooster, rdenotes a uniform random number from 0 to 1, $n\,(0, d^2)$ represents a Gaussian distribution with mean '0' and standard deviation (d). Based on the updated position of the rooster, the hen's positions are updated as follows.

$$p_h(t+1) = p_h(t) * \omega_1 * r * \left(p_{v_1}(t) + p_h(t)\right) + \omega_2 * r\ * \left(p_{v_2}(t) + p_h(t)\right) \quad (6)$$

In (6),$p_h(t+1)$ represents the updated position of the hens, $p_h(t)$ is the current position of hen, ω_1, ω_2 are the fitness evaluation parameters, v_1 is the index of the rooster [1,2, 3, …N], v_2 is the index of the hen [1,2, 3,…N], $v_1 \neq v_2$, r is the uniform random number from 0 to 1.Based on the updated position of the hen, the chicks' positions in the groups are updated as follows.

$$p_c(t+1) = p_c(t) + f_n * (p_m\,(t) - p_c(t)) \quad (7)$$

In (7),$p_c(t+1)$ is the updated position of the chicks, $p_c(t)$ is the current position of chicks, $p_m(t)$ is the position of the chicks mother$m$ is [1,2,3, ….N], f_n is the parameter $(f_n \in (0,2))$ it means that the chick follows its mother i.e. hen for hunting the food. Evaluate the new solution and obtain thefeature with better fitness than the previous one. This process is repeated until the maximum iteration is reached. Finally, the optimal relevant features are obtained. With the selected features, the hybrid technique finds the intrusion behaviours by analyzed with the testing feature values using Dichotomous regression function. Regression is the process of determining the relationship between the training data and testing data. Dichotomous means that the regression function provides the two outcomes namely normal, abnormal.

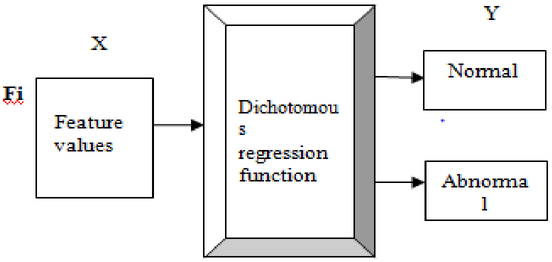


Figure 3: Dichotomous regression analysis based intrusion detection

As shown in figure 3, Dichotomous regression analysis based intrusion detection is described. The regression function takes the input selected feature value (i.e. training data) and related with the testing feature value (i.e. testing data)hence it provides the two outcomes results as normal or abnormal (i.e. intrusion). The correlation between the values of the features are measured as follows,

$$\rho = \frac{\sum_{i=1}^{n}(D_i - D_m)\,(Dt_i - Dt_m)}{\sqrt{\sum_{i=1}^{n}(D_i - D_m)^2 \sum_{i=1}^{n}\,(Dt_i - Dt_m)^2}} \quad (8)$$

$$D_m = \frac{1}{n}\sum_{i=1}^{n} D_i \quad (9)$$

$$Dt_i = \frac{1}{n}\sum_{i=1}^{n} Dt_i \quad (10)$$

Where, ρ denotes a correlation coefficient, D_i is the training data, Dt_i is the testing data, n denotes a number of data, D_m is the sample mean of the training data, Dt_m is the sample mean of the testing data. The correlation coefficient provides the outcomes either '+1' or '-1' where the '+1' indicates that the normal whereas '-1' indicates the abnormal (i.e. intrusions).

$$\rho = \begin{cases} +1, & normal \\ -1, & abnormal \end{cases} \quad (11)$$

Whereρ denotes a classification outcomes of the regression function. In this way, the hybrid technique classifies the normal or abnormal with higher accuracy. Based on the classification results, the hybrid technique effectively finds the intrusion in the network. The algorithmic process of the hybrid technique is described as follows,

Input: Data set D_{ids}, attributes $A_i = \{\mathrm{A}_1, \mathrm{A}_2, \mathrm{A}_3, \ldots \mathrm{A}_n\}$, data $\mathrm{D}_1, \mathrm{D}_2, \mathrm{D}_3, \ldots \mathrm{D}_n$
Output: Normal and abnormal data

```
Begin
Initialize chickens population $\{A_1, A_2, A_3, \ldots A_n\}$
  for each $A_i$
  Evaluate fitness $f_c$ based on similarity
    While (t < maximum iteration)
    Rank attributes based fitness values
    Divide the whole population into groups
    for each group
      Apply Map estimated naïve Bayes probabilistic rule i.e.$\arg\max\ P(A_i|O_v)$
      Select current best attribute as a rooster with higher fitness
      for each $A_i$
        if ($A_i$ = rooster)
          update its position $p_r(t+1)$
        else if ($A_i$ = hen)
          update its position $p_h(t+1)$
        else if ($A_i$ = chick)
          update its position $p_c(t+1)$
        end if
        Evaluate new solution
      end for
    end for
    t= t+1
    end while
  end for
Select optimal features and remove the other features
for the values of each selected feature
  Measure correlation '$\rho$'
    if ($\rho = +1$) then
      Data is classified as 'normal'
    Else
      Data is classified as 'abnormal'
    end if
end for end
```

Algorithm1 Hybrid Ruzicka Naive Bayes Chicken Swarm Optimized Feature Selection and Dichotomous Regression Classifier

Algorithm 1 describes the step by step process of hybrid technique for intrusion detection in the network. The populations of 'n' number of features are initialized randomly in search space. Then the fitness is calculated for each attribute based on the Ruzicka similarity. Followed by, every chickens are sorted based on the fitness value. Then the subgroups are created from the total populations based on their fitness and select the current best individual though the roulette wheel and the probabilistic rule. The rule is used to select the features with maximum probability for intrusion detection. Followed by, the positions of the chickens are updated along with the newly selected features. This process gets repeated until the maximum iteration is reached. In this way, the optimal features are selected from the dataset and removed the other features. In addition, the hybrid model also measures the correlation between the training and testing sets. The correlated results are used for classifying the data as normal or abnormal. This helps to improve intrusion detection accuracy and minimize the time.

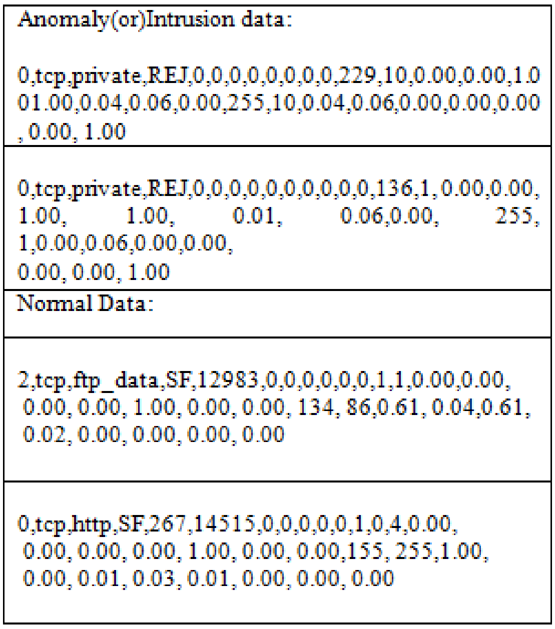
4. Experimental evaluation

Experimental evaluations of proposed HRNBCSOFS-DRC model and existing methods namely DBN-IDS [1] and ANID-SEoKELM[2] are implemented using Java language using NETBEANS 8.2 IDE tool. For the experimental consideration, NSL-KDD dataset is used and taken from https://www.kaggle.com/hassan06/nslkdd/version/1 This dataset is an enhancement of KDD'99 dataset from which duplicate instances were removed and improved the classification results. The dataset comprises 42 attributes. In NSL-KDD dataset, 42 attributes (i.e., features) are taken as input for feature selection process. Among the 42 attributes, the 30 attributes are selected as relevant (i.e. optimal) based on measuring maximum probability rule. In feature selection process, the relevant features are selected and irrelevant features are removed. Then this result of selected 30 attributes is taken as input for classification process shown in Table1: Relevant Features

Table1: Relevant Features

protocol_type:symbolic
service: symbolic
flag: symbolic
src_bytes: continuous
dst_bytes: continuous
land: symbolic
wrong_fragment: continuous
urgent: continuous
hot:continuous
num_failed_logins:continuous
logged_in:symbolic
root_shell: continuous
su_attempted:continuous
num_root:continuous
num_file_creations continuous
num_access_files: continuous
num_outbound_cmds: continuous
is_host_login:symbolic
is_guest_login:symbolic
count:continuous
srv_count:continuous
serror_rate:continuous
same_srv_rate:continuous
diff_srv_rate:continuous
srv_diff_host_rate:continuous
dst_host_count:continuous
dst_host_srv_count: continuous
dst_host_same_srv_rate: continuous
dst_host_diff_srv_rate: continuous

Here, correlation between the training and testing sets is measured. If the result of correlation is '+1', then the data is classified as normal. If the result of correlation is '-1', then the data is classified as abnormal as shown in table 2.



Anomaly(or)Intrusion data: 0,tcp,private,REJ,0,0,0,0,0,0,0,0,229,10,0.00,0.00,1.0 01.00,0.04,0.06,0.00,255,10,0.04,0.06,0.00,0.00,0.00 , 0.00, 1.00
0,tcp,private,REJ,0,0,0,0,0,0,0,0,0,0,136,1, 0.00,0.00, 1.00, 1.00, 0.01, 0.06,0.00, 255, 1,0.00,0.06,0.00,0.00, 0.00, 0.00, 1.00
Normal Data:
2,tcp,ftp_data,SF,12983,0,0,0,0,0,0,1,1,0.00,0.00, 0.00, 0.00, 1.00, 0.00, 0.00, 134, 86,0.61, 0.04,0.61, 0.02, 0.00, 0.00, 0.00, 0.00
0,tcp,http,SF,267,14515,0,0,0,0,0,1,0,4,0.00, 0.00, 0.00, 0.00, 1.00, 0.00, 0.00,155, 255,1.00, 0.00, 0.01, 0.03, 0.01, 0.00, 0.00, 0.00

Table 2: Classification of data

Performance analysis technique HRNBCSOFS-DRC model are compared with existing results with certain parameters listed below,

- Intrusion detection accuracy
- Precision
- Rcall
- F-measure
- Intrusion detection time

5. Results and Discussions

In this section, results of the proposed HRNBCSOFS-DRC model and existing methods namely DBN-IDS [1] and ANID-SEoKELM [2]are discussed in this section. The description of various metrics such as intrusion detection accuracy, precision, recall, F-measure and intrusion detection time are presented with a number of data to show the performance analysis of the proposed HRNBCSOFS-DRC model than the existing methods. The metrics are evaluated as given below.

Performance analysis of Intrusion detection accuracy

Intrusion detection accuracy is defined as ratios of a number of instances (i.e. data) are classified as normal or abnormal to the total number of data taken for the experimental evaluation. The formula for calculating the Intrusion detection accuracy is given below,

$$IDA = \left(\frac{Data\ as\ normal\ or\ abnormal}{Total\ number\ of\ data}\right) * 100 \quad (12)$$

WhereIDA denotes an intrusion detection accuracy which is measured in terms of percentage(%).

Table 3: Intrusion detection accuracy

Instances	Intrusion detection accuracy (%)		
	DBN-IDS	ANID-SEoKELM	HRNBCSOFS-DRC
1000	85	83	89
2000	87	80	93
3000	88	83	95
4000	86	81	96
5000	89	84	94
6000	88	82	95
7000	90	86	94
8000	89	85	96
9000	90	87	94
10000	89	85	95

Table 3 shows the intrusion detection accuracy of three methods The ten different results of intrusion detection accuracy are reported with respect to instances in the range of 1000 to 10000 collected from the dataset. The result in the graphical representation. show that the proposed HRNBCSOFS-DRC model achieves higher accuracy compared to existing techniques.

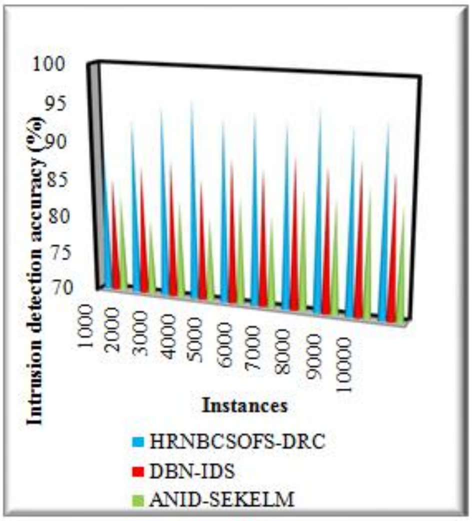


Figure 4: Intrusion detection accuracy

Fig.4. given above illustrates the graphical representation of intrusion detection accuracy with respect to a number of data. As shown in the figure, 'x' axis refers to the instances and y axis refers to the accuracy of intrusion detection.. The graphical result evidently proves that the HRNBCSOFS- DRC model achieves higher intrusion detection accuracy than the existing techniques. This is because of the fact that Dichotomous Regression function is applied for Intrusion Detection. The HRNBCSOFS-DRC model initially performs the feature selection through the optimization technique. With the selected feature value, the classification is done with the help of regression analysis. The regression function measures the correlation between the testing feature value and training feature value. If both the feature values get highly correlated, then the data is classified either normal or abnormal. Hence, the accuracy rate of HRNBCSOFS-DRC model is said to be improved. The average of ten results shows that the of intrusion detection accuracy is said to be improved using HRNBCSOFS-DRC model by 7% compared to DBN-IDS [1] and 13% compared to ANID- SEKELM [2].

Performance analysis of Precision

Precision is defined as a number of data correctly classified to the total number of data taken for the experimental evaluation. The formula for precision is calculated as follows

$$Precision = \left(\frac{TP}{TP+FP}\right) * 100 \quad (13)$$

In (13),denotes a true positive, FP represents the false positive. True positive is the numbers of data are correctly classified and the false positive rate is the numbers of data are incorrectly classified as an attack. Precision is measured in terms of percentage(%).

Table 4: Precision versus Instances

Instances	Precision (%)		
	DBN-IDS	ANID SEKELM	HRNBCSOFS-DRC
1000	89	87	90
2000	92	89	95
3000	92	90	96
4000	94	90	98
5000	94	91	96
6000	93	91	97
7000	94	93	96
8000	94	92	98
9000	95	94	97
10000	94	91	97

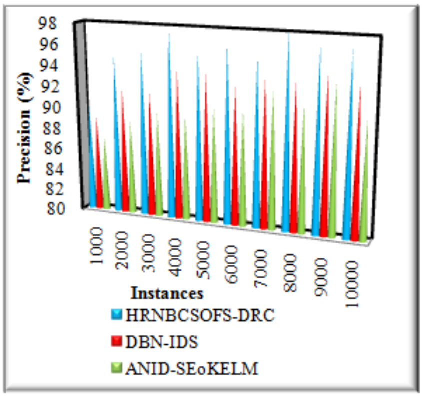


Figure 5: Performance results of precision

Table 4 and Fig.5 shows the performance results of precision versus number of data. For each iteration, the various precision results are obtained with respect to a number of data. The results of various methods are shown in the above graph and it represents three different colors. The graphical results are inferred that the precision of HRNBCSOFS-DRC model is improved than the other methods. This is because of the proposed hybridization technique accurately classifies the data and minimizes the incorrect data classification through the regression analysis. The correlation between the testing and training values minimized the false positive rate and improves the accurate data classification. The ten various precision results of HRNBCSOFS-DRC model is compared with the existing results. Then the average of comparison results clearly shows that the precision is considerably increased by 3% and 6% than the state-of-the-art methods.

Performance analysis of Recall

The recall is defined as a ratio of the rue positives to the sum of the true positive results and false-negative results from the data taken for the experimental evaluation. The mathematical formula for the recall is calculated as follows,

$$Recall = \left(\frac{TP}{TP+FN}\right) * 100 \quad (14)$$

negative is the numbers of data are incorrectly classified as normal. Recall also measured in terms of percentage (%).

Table 5: Recall versus Instances

Instances	Recall (%)		
	DBN-IDS	ANID-SEoKELM	HRNBCSOFS-DRC
1000	92	91	96
2000	92	87	97
3000	94	90	98
4000	93	89	98
5000	93	90	97
6000	94	88	98
7000	95	91	96
8000	93	91	98
9000	94	91	97
10000	93	92	98

The above table values are described that the experimental results of recall using three different techniques namely HRNBCSOFS-DRC model, DBN-IDS [1] and ANID-SEoKELM [2]. The numbers of data are taken as input for calculating the recall.
The reported result proves that the performance of recall is said to be improved using the proposed HRNBCSOFS-DRC model than the other methods.

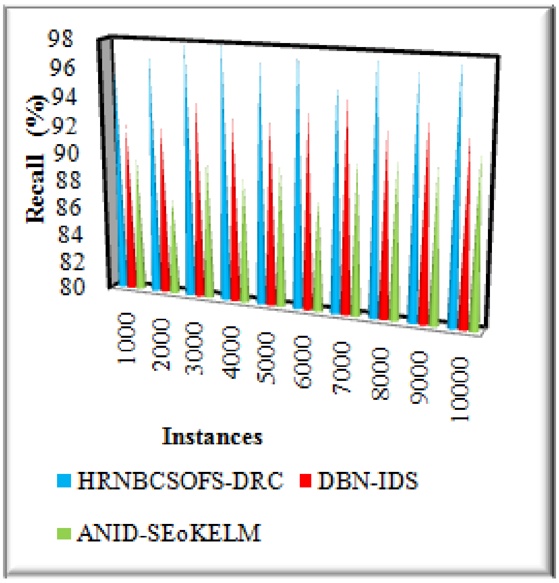


Figure 6: Performance results of Recall

The experimental results for recall estimation of three methods are plotted in Fig. 6. The recalls calculated with the true positives and false-negative results of classification. For all three methods, the HRNBCSOFS-DRC model has an ability to provide higher recall values. This is because of analyzing the feature values with the help of correlation and the regression provides the two possible results either normal or anomaly. This is evidently proved using mathematical calculation. Let us consider the 1000 data, the true positive result of HRNBCSOFS-DRC model is 750 and the false negative is 30. Therefore, the resultant value of recall is 96%. Whereas, the recall value of the two existing methods are 92% and 91% respectively. Similarly, the various results are obtained with respect to a number of data. The comparative analysis of the various results confirms that the recall is found to be improved by 4% and 8% using HRNBCSOFS-DRC model than the existing DBN-IDS [1] and ANID-SEoKELM [2].

Performance analysis of F-measure

F-measure is a measure of an evaluation accuracy of the classification algorithm and it is the average of the precision and recall. The F- measure is mathematically calculated as follows,

$$F - measure = 2 * \left(\frac{Precision * Recall}{Precision+Recall}\right) * 100 \quad (15)$$

Table 6: F-measure versus Instances

Instances	F-measure (%)		
	DBN-IDS	ANID-SEoKELM	HRNBCSOFS - DRC
1000	90	89	93
2000	92	88	96
3000	93	90	97
4000	93	89	98
5000	93	90	96
6000	93	89	97
7000	94	92	96
8000	93	91	98
9000	94	92	97
10000	93	91	97

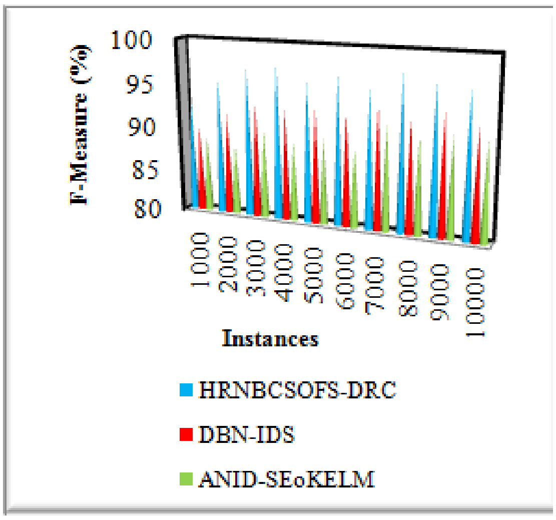


Figure 7: Performance results of F-measure

Table 6 and Fig.7 illustrates the performance results of F-measure using three methods with respect to a varied instances taken in the range of 1000-10000. From the table value, it is proved that the proposed model provides better F-measure results as compared to existing techniques. While considering the input count of the data is 1000, the precision and recall of HRNBCSOFS-DRC model are 90% and 96%. Therefore, the result of F-measure is 93% using the proposed technique and the results of the other two existing methods are 90% and 89% respectively. The performance results show that the HRNBCSOFS-DRC model outperforms well and improved by 4% and 7% as compared to DBN-IDS [1] and ANID-SEoKELM [2].

Performance analysis of Intrusion detection time
Intrusion detection time is defined as an amount of time required to classify the data as normal or abnormal. The Intrusion detection time is mathematically calculated as follows,

$$IDT = n * t\ (classifying\ one\ instances\ )$$

Where, IDT denotes an intrusion detection time, n represents a number of instances, t denotes a time for classifying one instances. Intrusion detection time is measured in terms of milliseconds (ms).

Table 7: Intrusion detection time versus Instances

Instances	Intrusion detection time (ms)		
	DBN-IDS	ANID-SEoKELM	HRNBCSOFS-DRC
1000	26	29	23
2000	32	36	30
3000	36	42	33
4000	40	44	36
5000	45	48	40
6000	48	54	42
7000	50	56	46
8000	56	58	53
9000	59	63	56
10000	62	65	58

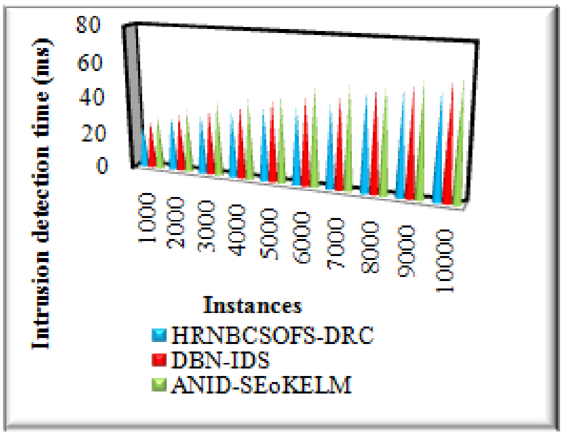


Figure 8: Performance results of Intrusion Detection time

Table 7 and Fig. 8 illustrates the performance results of intrusion detection time using three methods with respect to varied instances taken in the range of 1000-10000. The reported results inferred that the intrusion detection time of proposed HRNBCSOFS-DRC is lesser when compared to other existing methods. The ten various results of intrusion detection time shown in fig. 8While considering 1000 data for calculating the intrusion detection time, the proposed HRNBCSOFS-DRC model takes $23ms$ for classifying the data as normal or abnormal. Whereas $26ms\ and\ 29ms$ are taken by the DBN-IDS [1] and ANID-SEoKELM [2] for classifying the input data. Similarly, the nine various results are obtained with different data. The proposed technique is compared to the existing results. From these results, it is cleared that the intrusion detection time using proposed HRNBCSOFS-DRC model is found to be minimized by 8% compared to [1] and 16% compared to [2].

Screenshots

The Intrusion Detection Regression Classifier with normal and Anomaly data is shown in Fig.9.

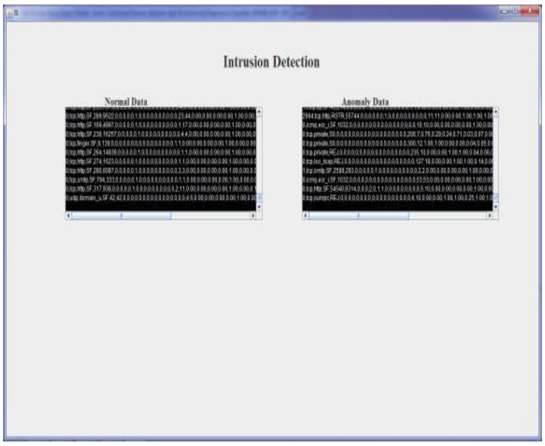


Figure 9: Screenshot of Intrusion Detection: Regression classifier.

6. Conclusion

A novel model called HRNBCSOFS-DRC is introduced for improving the intrusion detection accuracy rate and minimizes the detection time. The features are collected from the datasets and it was given to the input of the hybrid technique. The input features are learned in the search space by applying the optimization technique. Then the feature with higher fitness is selected based on the Ruzicka similarity. This process gets iterated until a termination condition is met. In this way, the optimal features are selected and other features are removed. Then the Dichotomous Regression function analyzes the testing and training data by measuring the correlation. Based on the classification result, normal and attack detection is performed with higher accuracy. Experimental evaluation is carried out using improved version of KDD'99 dataset with the parameters such as intrusion detection accuracy, precision, recall, F- measure and intrusion detection time. The discussed results clearly show that HRNBCSOFS- DRCmodel improves the intrusion detection accuracy with minimum false positive rate and higher true positive rate as well as minimizes the intrusion detection time.

References

[1] Peng Wei, YufengLi, Zhen Zhang, Tao Hu, ZiyangLi, Diyang Liu, "An Optimization Method for Intrusion Detection Classification Model Based on Deep Belief Network", EEE Access, Volume 7, 2019, Pages 87593 -8760.

[2] Jinping Liu,Wuxia Zhang,TianyuMa,Zhaohui Tang,Jean-Paul Niyoyita,WeihuaGui, "ANID-SEoKELM: Adaptive network intrusion detection based on a selective ensemble of kernel ELMs with random features", Knowledge-Based Systems, Elsevier, Volume 177, 2019, Pages 104- 116

[3] Ran Yahalom, AlonSteren, Yonatan Nameri, Maxim Roytman, AngelPorgador, Yuval Elovici, "Improving the effectiveness of intrusion detection systems for hierarchical data",Knowledge-Based Systems, Elsevier, Volume 168, 2019, Pages 59-69

[4] Alex Shenfield, David Day, Aladdin Ayesh, "Intelligent intrusion detection systems using artificial neural networks", ICT Express, Elsevier, Volume 4, Issue 2, June 2018, Pages 95-99

[5] ErxueMin, Jun Long, QiangLiu, JianjingCui, and Wei Chen, "TR-IDS: Anomaly-based Intrusion Detection through Text-Convolutional Neural Network and Random Forest", Security and Communication Networks, Hindawi, Volume 2018, July 2018, Pages1-9

[6] Akashdeep, IshfaqManzoor, Neeraj Kumar, "A feature re duce d intrusion detection system using ANN classifier", Expert Systems With Applications, Elsevier, Volume 88,2017, Pages 249–257

[7] Mohammed Hasan Ali, BahaaAbbas DawoodAl Mohammed, Alyani Ismail, MohamadFadliZolkipli, "A New Intrusion Detection System Based on Fast Learning Network and Particle Swarm Optimization", IEEE Access, Volume 6, 2018, Pages 20255 –20261

[8] BayuAdhiTama, Marco Comuzzi, Kyung-Hyune Rhee, "TSE-IDS: A Two-Stage Classifier Ensemble for \Intelligent Anomaly-Based Intrusion Detection System", IEEE Access, Volume 7,2019, Pages 94497-94507.

[9] JiadongRen, JiaweiGuo, Wang Qian, Huang Yuan,XiaobingHao, and Hu Jingjing, "Building an Effective Intrusion Detection System byUsing Hybrid Data Optimization Based on MachineLearning Algorithms", Security and Communication Networks, Hindawi, Volume7 June 2019, 11Pages 1-11

[10] VajihehHajisalem and ShahramBabaie, "A hybrid intrusion detection system based on ABC- AFS algorithm for misuse and anomaly detection", Computer Networks, Elsevier, Volume 136,2018,Pages 37–50

[11] Bing Zhang, ZhiyangLiu,YanguoJia, JiadongRen, and Xiaolin Zhao, "Network Intrusion Detection Method Based onPCA and Bayes Algorithm", Security and Communication Networks, Hindawi, Volume 7, November 2018, Pages1-11

[12] DimitriosPapamartzivanos, Félix Gómez Mármol, GeorgiosKambourakis "Introducing Deep Learning Self-Adaptive Misuse

Network Intrusion Detection Systems”, IEEE Access, Volume 7, 2019,Pages 13546 – 13560.

[13] WajdiAlhakami, Abdullah ALharbi, SamiBourouis,RoobaeaAlroobaea,NizarBouguila, “Network Anomaly Intrusion Detection Using a Nonparametric Bayesian Approach and Feature Selection”, IEEE Access,

[14] Reehan Ali Shah,Yuntao Qian, Dileep Kumar, Munwar Ali and Muhammad BuxAlvi, “Network Intrusion Detection through Discriminative Feature Selection by Using Sparse Logistic Regression”, Future Internet, Volume 9, Issue 4, 2017, Pages 1-15

[15] JieGua,Lihong Wang, Huiwen Wang, Shanshan Wang, “A novel approach to intrusion detection using SVM ensemble with feature augmentation”, Computers & Security, Elsevier, Volume 86, 2019, Pages 53-6

[16] Shadi Aljawarneha, Monther Aldwairia, Muneer Bani Yasseina, “Anomaly-based intrusion detection system through feature selection analysis and building a hybrid efficient model”, Journal of Computational Science, Volume 25, 2018, Pages 152-160

[17] ChaoukiKhammassi, and SaoussenKrichen, “A GA-LR wrapper approach for feature selection in network intrusion detection”, Computers & Security, Elsevier, Volume 70, September 2017, Pages 255-277

[18] Sydney MambweKasongo andYanxia Sun, “A Deep Learning Method With Filter Based Feature Engineering for Wireless Intrusion Detection System”, IEEE Access, Volume 7, 2019, Pages 8597 – 38607

[19] WathiqLaftah Al-Yaseen, Zulaiha Ali Othman, MohdZakree Ahmad Nazri, “Multi-level hybrid support vector machine and extreme learning machine based on modified K-means for intrusion detection system”, Expert Systems with Applications, Elsevier, Volume 67, January 2017, Pages 296-303

[20] José Camacho, Roberto Therón, José M. García-Giménez, Gabriel Maciá-Fernández, Pedro Garcí, “Group-Wise Principal Component Analysis for Exploratory Intrusion Detection”, IEEE Access, Volume 7, 2019, Pages 113081 - 113093