Indian Journal of Science and Technology, Vol 9(42), DOI: 10.17485/ijst/2016/v9i42/101555, November 2016
ISSN (Print) : 0974-6846
ISSN (Online) : 0974-5645

Nature Inspired Feature Selection Approach for Effective Intrusion Detection

Rajinder Kaur[1], Monika Sachdeva[1] and Gulshan Kumar[2]

[1]Department of Computer Science Engineering, Shaheed Bhagat Singh State, Technical Campus, Ferozepur, India; khalsarajkaur@gmail.com, monasach1975@gmail.com
[2]Department of Computer Applications, Shaheed Bhagat Singh State, Technical Campus, Ferozepur, India; gulshanahuja@gmail.com

Abstract

Objectives: To reduce the dimensionality of network traffic dataset by selecting the relevant and irredundant features for accurate and quick intrusion detection. To achieve the target, we proposed a new feature selection approach based on the nature. **Methods/Statistical Analysis**: The proposed Modified CuttleFish Algorithm (MCFA) approach plays a crucial role in intrusion detection by selecting appropriate subset of most relevant features from huge amount of dataset. Griewank fitness function is used to calculate the fitness of the modified cuttlefish algorithm. Naive bayes classifier is employed at the generated subset of features from benchmark KDD 99 dataset in WEKA data mining tool Compare the results of proposed approach with the existing approaches of WEKA and Improved Cuttlefish Algorithm (ICFA), with different performance metrics. **Findings:** As per the outcomes are obtained by the WEKA Experimenter with the 9 feature selection approaches on KDD 99 10% training dataset, it has been observed that the Consistency Subset feature selection approach with Greedy stepwise search method gives higher accurate results than other approaches and from the literature survey has been found that the ICFA performs well, but still there is problem of low True positive (TP) rate and False negative (FN) rate. These problems are addressed by the proposed feature selection approach which outer perform best from others in accuracy rate (91.79%), True positive rate (0.947), false positive rate (0.025) and ROC area (0.9982) with the minimum amount of time at 19 relevant subset of features instead 41 features. As the consequence, the proposed approach is novel from existing approaches to increase the intrusion detection rate and discard the redundant and irrelevant subset of features. **Application/Improvements:** MCFA has improved intrusion detection rate by increasing the TP rate and decreasing the FP rate, so MCFA can be used for the real time applications of intrusion detection system.

Keywords: Accuracy, Dimensionality, Feature Selection Approach, FP Rate, Intrusion Detection, Modified Cuttlefish Algorithm, TP Rate

1. Introduction

According to the popularity and fast growth of internet, the possibility of network attacks has been increased significantly in recent years. Therefore, to provide more secure information channels much attention has been needed. Distinguish the attack action from the normal network action, is not a simple process. This problem is overcome with the proposed concept of Intrusion Detection in 1980[1]. Intrusion Detection System (IDS) provides a defense mechanism in computer networks against the various attacks and criminal activities. Intrusions are a set of activities and actions that are made an attempt to compromise the security objectives like integrity, comprehensibility and availability of the information resources[2]. The main role of IDS is to recognize the unusual access, attacks and activities to make sure the more protection in the network channels. When any potential attack is detected in network traffic, an alarm is triggered and taking an appropriate action against the attacks by IDS. During the building of IDS many challenges are needed to consider such as collection of data, data processing,

*Author for correspondence

dimensionality of data, classification accuracy. IDS deal with huge amount of data which contains large number of features (attributes). Some of irrelevant, redundant, unnecessary and noisy features in dataset, plays no more important role in IDS system. High dimensionality of data in IDS may lead to decline the classification accuracy of the system. So, to remove these problems, the feature selection approach is very essential in IDS system.

1.1 Feature Selection

Feature Selection (FS) is a process of dimensionality reduction, which opt the best possible subset of features from the large amount of data that represents the whole dataset. It discard the unnecessary, irrelevant and redundant features, it also increase the classification accuracy and improve the efficiency in terms of storage costs, measurement costs and computational costs. A feature selection selects the most favorable subset of features and keeps the original features as such. From given dataset of m number of features the FS selects the minimal n (n<m) number of features and increase the accuracy rate. In feature selection process has four working steps listed in Figure 1. The full amount of dataset is supplied to subset generation step; it generates a subset of optimal features from the whole dataset. Generated subset of features is evaluated with certain predefined criteria and compared with previous best subset of features. Process is going on up to it doesn't meet the predefined criteria. In validation step, checking the selected subset of feature is a relevant subset or not. In this paper, we have purposed a feature selection approach based on the modified cuttlefish algorithm.

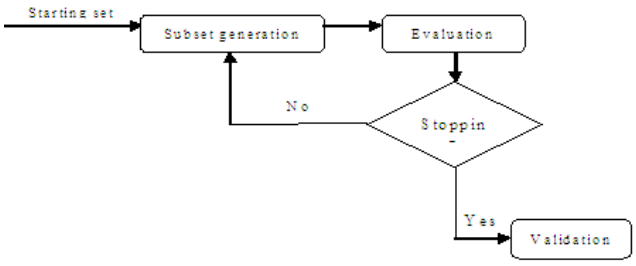

Figure 1. General process of feature selection process

On building the efficient IDS by reducing the input features, the proposed Feature Vitality Based Reduction Method[3] (FVBRM) gives best results at 24 features while comparing it with Correlation-based Feature Selection, Information Gain and Gain Ratio methods. NSL-KDD dataset is applied with classification algorithm Naive bayes at WEKA 3.6 Machine learning tool. FVBRM method improves the results for intrusion particularly with U2R attacks. A hybrid feature selection algorithm in combination of filter selection process and wrapper method[4] indicate the best performance with Least Square Support Vector Machine (LSSVM) classifier, which is used to guide the selection process and retain optimized set of 25 features. The proposed technique gives less accuracy results. Mutual Information (MI) is validated empirically by comparing the results with Gain Ration method, Relief method, CFS method, FCBF method and Feature selection with dynamic mutual information method[5]. Empirical results show that the proposed technique gives best result at 10 features with 56.56% accuracy and ROC 0.796. Moreover, the Modified Mutual Information-based Feature Selection (MMIFS) method[6] is applied at KDD Cup 99 dataset with LSSVM classifier in order to gain the higher accuracy with subset of 14 features. Detection rate increases with increases the number of features in proposed method.

On ranking the various feature selection algorithms like InfoGain, Gain Ratio, OneR, RELIEF, with J48 classifier proposed a new approach[7]. Experimental results at KDDcup99 dataset for intrusion detection showing that the proposed method gives the best accuracy performance (66.807%) at 12 features, yet the improved accuracy rate is very less with selected features. A new feature-selection approach based on the cuttlefish optimization algorithm[8] is proposed for the search strategy at optimal subset of features and then Decision Tree (DT) classifier is used for classification purpose. Classification is a data mining task that maps the data into predefined groups and classes. It is also known as supervised learning[9]. Empirical results indicate that the proposed approach gives best result in Detection Rate (DR) and Accuracy Rate (AR) at subset of 20 features. The dimensionality of reduced dataset is large and it shows that the DR and AR are decreases with number of selected features are less than 20 features. A novel feature selection technique based upon Cuttlefish algorithm to get more classification accuracy in appropriate training time for intrusion detection has been proposed[10]. In ICFA the Rosenbrock function is used to evaluate the fitness of the generated subset at 20 features in KDDCup 99 dataset. The best Feature subset at 20 features gives best classification accuracy and ROC curve. Selected feature subset gives small variation in accuracy rate.

The paper is prepared as follow: firstly in selection 1 gives an overview at the introduction part including Feature selection and related work. Section 2 describes the Research methodology with modified cuttlefish algorithm and experiment setup. Experimental results and analysis have been discussed in section 3. Finally, the paper is concluded with the future work in section 4.

2. Research Methodology

After reviewing the literature, it has been evident that feature selection has been an essential step to the increase the accuracy rate in the intrusion detection. The Cuttlefish algorithm[11] is a best approach to select the relevant subset of feature. In this paper, we have modified the cuttlefish algorithm to gain the best result in detection rate.

2.1 Modified Cuttlefish Algorithm

A MCFA is a programming approach that mimics the mechanism of Cuttlefish as an evaluation of search strategy to find the optimal subset of features. Cuttlefish has an ability to change its colour according to the environment circumstances. Cuttlefish produce the patterns and colours from the reflected light that is passing trough different layers of cells including chromatophores, iridophores and leucophores. MCFA has main two processes: Refection process and Visibility process. Refection process is working as simulate the mechanism of light reflection. Whereas Visibility process working as simulate the mechanism of matching patterns. Refection and Visibility processes are used as a search approach to evaluate the global optimal solution. New solution is generated from the combination of both Refection and Visibility process. The whole process of colour changing in cuttlefish is based upon the cells of chromatophore, iridophore and leucophore layers. Colour changing mechanism of cuttlefish at different layers, used as a search strategy of feature selection in IDS on basis of fitness value shown in Figure 2.

Initialize the MCFA algorithm with population of P, which has N number of initial solution that are randomly generated, P[N]= $\{p_1, p_2, p_3, p_4, p_5, \ldots\ldots\ldots\ldots p_n\}$. In our experiment, we have used the benchmark KDD dataset which contains 41 features. So, the value of N is 41. New generated solution is dividing into two categories, selected subset of features and unselected subset of features. The size of both subsets is less than the P[N]. The size of selected subset is fixed in the starting of the algorithm. We have used the Griewank fitness function to calculate the fitness of the random selected subset from the whole population. Then the best fittest solution is divided into two subsets i.e. average best subset and best subset, where the size of best subset is one less than the average best subset. Next colour changing six cases mechanism of cuttlefish at different layers of cells is used as simulation of subset generated in feature selection from dataset as shown in Figure 3.

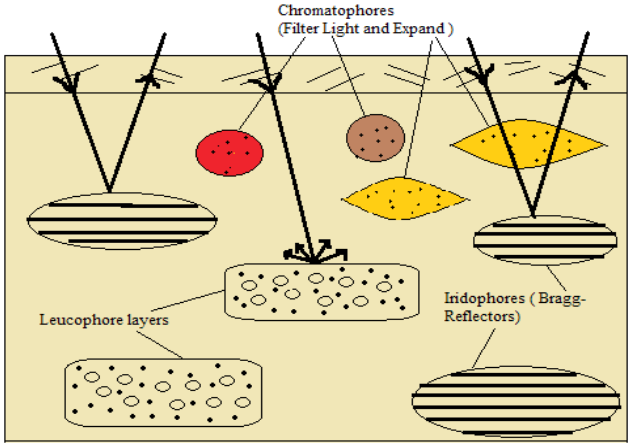

Figure 2. Three layer skin structure of cuttlefish

$$\text{New subset} = \text{Reflection} + \text{Visibility} \quad (1)$$

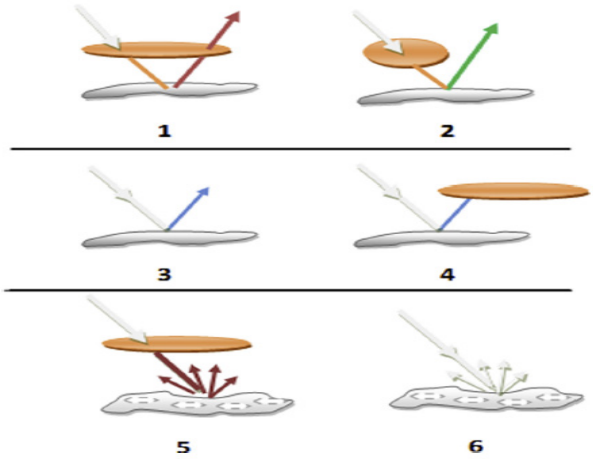

Figure 3. Six cases mechanism at different layers of Cuttlefish

Simulation of case 1&2 – In MCFA two processes reflection and visibility is used to find out the new subset. In case 1&2 light reflected mechanism is occurred due to interaction between the chromatophore and iridophore

cells. Chromatophore cell will contract or slow down its muscles to extend or minimize its saccule and iridophore cell will reflect the light coming from chromatophore cells. The stretch and contract process in chromatophore and reflected light in iridophore and visibility of matching patterns of cuttlefish are used as to find out the new solution. From this mechanism we have used these equations in our work.

$$\text{Reflection}_j = R^* G_1[i].\text{Points}[j] \quad (2)$$

$$\text{Visibility}_j = V^*(\text{BestPoints}[j] - G_1[i].\text{Points}[j]) \quad (3)$$

In (2) and (3) G_1 signify a group of chromatophore cells that are used in case 1 & 2. i is i^{th} cell in group G_1. Points[j] represent j^{th} point at i^{th} cell and BestPoints[j] represents the best solution points. R signifies the reflection degree and V signifies the final view of the patterns. The value of R and V are evaluated as:

$$R = \text{random}()^*(r_1 - r_2) + r_2 \quad (4)$$

$$V = \text{random}()^*(v_1 - v_2) + v_2 \quad (5)$$

Where, random ( ) function is a function used to generate a random numbers between (0, 1) and r_1, r_2 are two constant values that are applied to find the stretch interval of the chromatophore cells. v_1 and v_2 are two constant values that are applied find out the interval of the visibility's degree to the final view point of the pattern. For the simulation of these cases ,we sorting the P in descending order on the basis of fitness values and choosing the first k subsets from P, where k is a random number between (1, N/2), N is the size of P. The new subset is generated from each subset in k using the subsets of Reflection set and Visibility set. If the new generated subset is improved than the average best subset then restore the average best subset with new subset.

Simulation of case 3&4: The iridophore cells are light reflecting cells, the iridophore cells will reflect the inward light coming from the outside (environment), and the reflected colour is a particular colour. Iridophore cells are used to hide the organs. So, it is supposed that the hide organs are represented by the best solution. In this case the value of R is fixed to 1 and the value of V is calculated as case 1 and 2.

$$\text{Reflection}_j = R^* \text{Best.Points}[j] \quad (6)$$

From these two cases, we simulate that a single feature-exchanging operator is used to produce a new solutions from the best Subset. A random feature is selected from selected Features to be exchanged with another random feature selected from it. If new produced solution is better than best Subset, then replace best Subset with the new subset.

Simulation of case 5: In this case, the light is reflected from leucophore cells are almost same that is coming from the chromatophore cells. From the similarity between the reflected light and incoming light, we assumed that best solution is incoming light and reflected colour may be any colour near the best solution set as average best solution. So, the difference between the best solution points and the average solution points is worked as a small area produced in the region of best solution consider as a new search area.

$$\text{Reflection}_j = R^* \text{Best.Points}[j] \quad (7)$$

$$\text{Visibility} = V^*(\text{Best.Points}[j] - AV_{best}) \quad (8)$$

Where, The AV_{best} is the average value of the *Best* points. The value of *R* is set to 1, while the value of *V* will be calculated. So, as a simulation in MCFA, generate p new subsets from the average best subset by eliminating one feature each time. Evaluate each subset using Griewank fitness function. If any subset is improved the results than the best subset then restore best subset with it.

Simulation of case 6: In this case, the leucophore cells mechanism as a mirror and will reflect the inward light from the environment. Here, the cuttlefish is capable to merge itself into the environment. As a simulation here, we have assumed that any inward light coming from the environment will be reflected the same as it and can be represented by any random solution. So, In this case the remaining solutions of P will be generated randomly. If the new solution is better than AV best subset then put it into AV best solution.

End the algorithm process when it returns the best subset of features.

2.2 Griewank Function

Griewank Function is used to calculate the fitness of the selected subset in Modified Cuttlefish algorithm. The Griewank function has been broadly used to check the convergence of the optimization algorithm. Its number of minima rises exponentially as its number of dimensions enhances. Its range fluctuate between the values $[-600,600]^n$, where n is defined as dimensions of the function. The function is defines as:

$$f(x) = \sum_{i=1}^{d} \frac{x_i^2}{4000} - \prod_{i=1}^{d} \cos\left(\frac{x_i}{\sqrt{i}}\right) + 1$$

2.3 Experiment Criteria

In this research, implementation of the proposed approach is done in C++ programming language at Microsoft visual studio platform. We have compared the proposed feature selection approach with 10 existing feature selection approaches like Cfs Subset Evaluator, ChiSquared Attribute Evaluator, Consistency Subset Evaluator, Filtered Attribute Evaluator, Filtered Subset Evaluator, Gain Ratio, InfoGain, OneR and Symmetrical Uncert Method with using different performance metrics[12]. Analysis and comparison of these approaches has done in our previous work[13]. The Benchmark KDD-99 intrusion data set is used with different feature selection approaches. Knowledge Discovery in Databases (KDD) is the procedure of exploring the helpful knowledge from a large database by using data mining algorithm through specifications of threshold[14,15]. KDD-99 dataset contains 41 features and are ladled with normal and attack class[16]. Reduced dataset suggested by different approaches is classified using Naive Bayes classifier in WEKA 3.6.9 Environment tool[17]. Different performance metrics like TP rate, FP rate, Precision, ROC area, Kappa Statistic, Classification accuracy, Training time are computed for various feature selection approaches. Choose the best approach from the comparison of existing approaches and compare the results with proposed approach. Result analysis is done with the help of Microsoft Excel. The several experiments are conducted on Intel Core i3 CPU 380 2.40 GHz Processor with 6 GB RAM with 64 bit operating system at Window 10 Professional. Figure 4 shows the operation of Modified Cuttlefish algorithm according to which MCFA is implemented into our experiment.

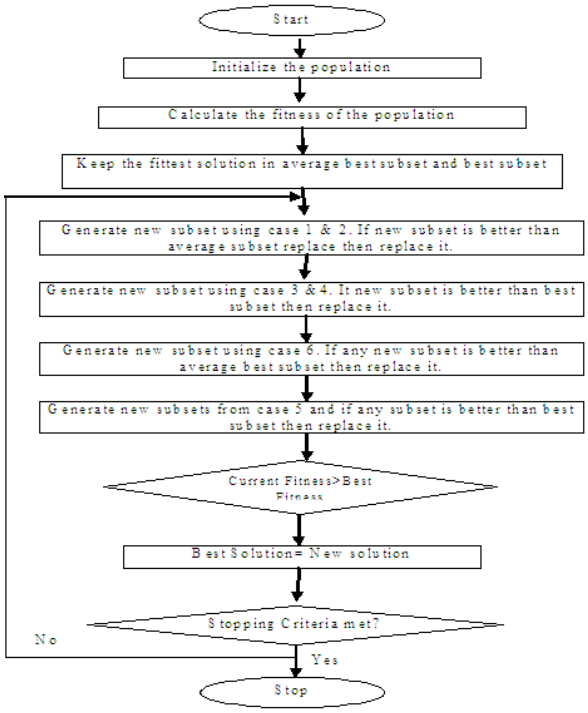

Figure 4. General flow chart of MCFA

Table 1. Confusion matrix

Class	Predicted Normal	Predicted Attack
Actual Normal	TN	FP
Actual Attack	FN	TP

3. Experiment Results and Analysis

3.1 Performance Metrics

The proposed approach MCFA and existing approaches of FS are evaluated based on the performance metrics of True positive rate, False positive rate, Precision, F measure, ROC area, Kappa statics, Accuracy and Training Time[18,19]. Confusion matrix of Table 1 summarizes the number of instances to calculate the normal or abnormal by the classification model. Performance metrics are dependents on the results of confusion matrix. Brief description of different performance metrics are as followed.

True Negative: TN evaluates the number of normal features those are detected as normal feature.

- False Negative: FN evaluates the number of attack features those are detected as normal feature.
- True Positive: TN evaluates the number of attack features those are detected as attack feature.
- False Positive: TN evaluates the number of normal features those are detected as attack feature.
- Classification Accuracy: To measure the performance of the classifier the classification accuracy (CA) is most required[20]. It concludes the fraction of correctly classified Instances over the full amount of instances.

$$CP = \frac{TP + TN}{TP + TN + FP + FN} * 100$$

- True positive Rate (TP Rate)- TPR is defined as the ratio of number of classified attack connections and full amount of normal connections.

$$\text{TP Rate} = \frac{TP}{TP + FN}$$

- False Positive Rate (FP Rate)- FPR is defined as the ratio of the number of misclassified normal connections and full amount of normal connections.

$$\text{FP Rate} = \frac{FP}{FP + TN}$$

- Precision: This metric is defined with respect to the intrusion class. It should be high for more accuracy in IDS system.

$$\text{Precision} = \frac{TP}{TP + FP}$$

- Receiving Operating Characteristic (ROC Area)- ROC is applied to draw a curve between TP Rate and FP Rate and the area contained under the curve is known as AUC that gives the value of the ROC.
- F-Measure: The F-measure is defined as a weighted harmonic mean of recall and precision. It is high when both the recall and precision are high.
- Kappa Statistic: This is a statistic which calculates the inter-rater contract for qualitative or categorical items. The value of the kappa statistic lies between 0 to 1 ranges. 0 means totally disagree and 1 means full agreement.
- Training time: It is total time used by Classifier to build the model on a specified dataset. It is frequently calculated in seconds.

3.2 Results

In this research work, an effective feature selection approach is proposed to increase the accuracy rate with appropriate subset of features. Proposed feature selection approach is based on Modified Cuttlefish Algorithm and shows the best results at 19 features rather than 41 features. These selected 19 features are applied on benchmark KDD-99 dataset in WEKA data mining tool. Effectiveness and feasibility of proposed approach is validated by several experiments on dataset. Various performance metrics like TP rate, FP rate, Precision, Kappa statistic, classification accuracy, training time are computed with Naive bayes

Table 2. Comparative analysis of existing feature selection approaches

Feature Selection Approaches	No. of Selected Features	TP Rate	FP Rate	Precision	F Measure	ROC Area	Kappa Statistic	Accuracy	Training Time (Sec.)
Full Feature	41	0.890	0.109	0.803	0.802	0.867	0.7803	88.25%	467.73
CFS Subset + BFS	6	0.82	0.204	0.852	0.813	0.937	0.6303	82.01%	56.84
CFS Subset + Genetic search	15	0.898	0.112	0.905	0.898	0.955	0.794	89.85%	176.29
Chi Squared	12	0.901	0.106	0.903	0.901	0.967	0.8002	90.12%	118.44
Consistency Subset + Greedy stepwise	11	0.904	0.105	0.908	0.903	0.972	0.8054	90.39%	133.56
Consistency Subset + Linear Forward Selection	10	0.891	0.121	0.898	0.889	0.968	0.7777	89.05%	121.23
Filtered Attribute	12	0.887	0.122	0.892	0.886	0.962	0.7714	88.72%	113.85
Filtered Subset + Greedy stepwise	6	0.82	0.204	0.852	0.813	0.937	0.6303	82.01%	85.81
Gain Ratio	12	0.897	0.114	0.903	0.896	0.965	0.7903	89.67%	117.1
Info Gain	12	0.887	0.122	0.892	0.886	0.962	0.7714	88.72%	125.14
One R	12	0.901	0.106	0.903	0.901	0.967	0.8002	90.12%	125.69
Symmetrical Uncert	12	0.876	0.138	0.887	0.874	0.963	0.747	87.57%	126.89

Table 3. Comparative analysis of the proposed approach with best existing approach and ICFA

Feature Selection Approaches	No. of Selected Features	TP Rate	FP Rate	Precision	F Measure	ROC Area	Kappa Statistic	Accuracy	Training Time (Sec.)
Consistency Subset + Greedy stepwise	11	0.904	0.105	0.908	0.903	0.972	0.8054	90.39%	133.56
ICFA	20	0.909	0.027	0.939	0.901	0.998	0.8669	90.74%	54.00
Proposed MCFA	19	0.947	0.025	0.941	0.92	0.9982	0.8763	91.79%	50.57

classifier. From existing approaches CFS, Chi-squared, Consistency subset evaluator, Filtered attribute evaluator, Filtered attribute evaluator, Gain ratio, Info Gain, OneR, Symmetrical liner uncertain feature selection approaches the consistency subset evaluator approach is best approach. Table 2 describes the comparative analysis of different existing feature selection approaches with Naïve Bayes Classifier.

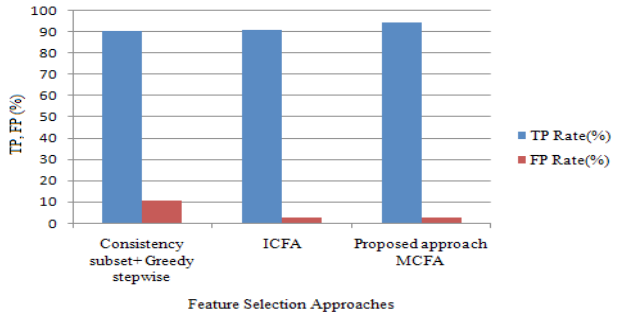

Figure 5. TP rate and FP rate of proposed approach MCFA with existing approaches

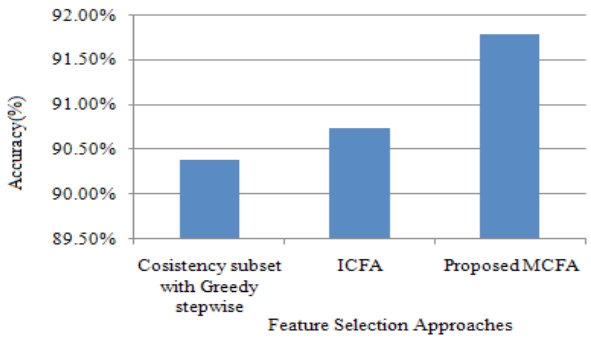

Figure 6. Classification accuracy of proposed approach MCFA with existing approaches

Table 3 depicts the information of the comparative analysis of Consistency subset evaluator, ICFA and MCFA feature selection Consistency subset evaluator approach with greddy stepwise search engine gives 90.39% accuracy and 0.8054 ROC with subset of 11 features. The results of MCFA also compared with results of ICFA (Kaur et al. 2015). ICFA gives best result at subset of 20 features with 90.74% classification accuracy and 0.998 ROC areas. Whereas our Proposed approach MCFA gives 91.79% classification accuracy at optimal subset of features while at 41 features the classification accuracy is 88.25%. Empirical results indicate that the proposed approach outperform best from all the other existing approaches in accuracy rate, ROC area and it also takes less training time. So, MCFA approach gives best result in detection rate by selecting the relevant subset of features form large number of features. Future Figure 5 and 6 show the comparison analysis of True positive rate, false positive rate and accuracy rate of the proposed approach Modified Cuttlefish Algorithm Feature Selection Approach (MCFA) with existing feature selection approaches.**Table 3**. Comparative analysis of the proposed approach with best existing approach and ICFA

4. Conclusion

In this paper, we have implemented the Modified Cuttlefish algorithm to search the optimal subset of features for intrusion detection. To measure the performance of proposed approach, we used the standard benchmark KDD99 dataset and obtained the best result of detection rate. We examined the performance of different feature selection approaches that are existing in the WEKA environment tool like CFS, Chi-squared, Consistency subset evaluator, Filtered attribute evaluator, Filtered attribute evaluator, Gain ratio, Info Gain, OneR, Symmetrical liner uncertain with search methods. From comparative analysis of existing feature selection approaches Consistency subset evaluator approach with greedy stepwise search method is best approach from all existing approaches. To

gain more accuracy and detection rate results, we proposed a feature selection approach based on the Modified Cuttlefish Algorithm (MCFA). MCFA is implemented for selecting the relevant subset of features at benchmark KDD99 dataset. The selected subset of relevant features by proposed approach is employed in WEKA with Naive bayes classifier. The effectiveness of proposed approach is validated by comparing the results with existing feature selection approaches. Empirical results indicate that the proposed approach MCFA gives the relevant subset of 19 features outperforms the other representative feature selection approaches on benchmark KDD 99 dataset for intrusion detection. The fitness of proposed approach is evaluated by Griewank function. Proposed approach gives best result in True positive rate, false positive rate, ROC curve and classification accuracy as compared the other existing approaches. MCFA also reduces the training and testing time with reduced dataset at subset of 19 features instead of 41 features. Proposed approach plays a great role in selection of appropriate subset of features for effective intrusion detection. The Future works include to faster the feature selection process to making the Modified Cuttlefish Algorithm parallel. Moreover, the classification at generated subset of features may be done with bayes net, Part, J48 classifiers.

5. References

1. Denning D. An intrusion-detection model, Software Engineering, *IEEE Transactions on SE*, 1987; 13(2): 222-32.
2. Stoneburner G. Underlying technical models for information technology security, NIST Special Publication 800-33, National Institute of Standards and Technology.
3. Mukherjee S, Sharma N. Intrusion detection using naive bayes classier with feature reduction. *Procedia Technology*, 4:119-28. *2nd International Conference on Computer*, Communication, Control and Information Technology (C3IT-2012), 2012.
4. Saad A. An overview of hybrid soft computing techniques for classier design and feature selection. In: *Hybrid Intelligent Systems*, 2008. HIS '08. *Eighth International Conference,* 2008, p. 579-83.
5. Kumar G, Kumar K. An Information theoretic approach for feature selection. *Security and Communication Networks,* 2011; 5(2):178-85.
6. Skaruz J, Nowacki JP, Drabik A, Seredynski F, Bouvry P. Soft computing techniques for intrusion detection of sql-based attacks. In: *Proceedings of the Second International Conference on Intelligent Information and Database Systems: Part I, ACIIDS'10*, p. 33, Berlin, Heidelberg. Springer-Verlag, 2010.
7. Singh R, Kumar H, Singla R. Analysis of feature selection techniques for network trac dataset. In: *Machine Intelligence and Research Advancement (ICMIRA)*, 2013. *International Conference,* 2013, p. 42- 6.
8. Eesa AS, Orman Z, Brifcani AMA. A novel feature-selection approach based on the cuttlefish optimization algorithm for intrusion detection systems. *Expert Systems with Applications*, 2015; 42(5):2670-79.
9. Lakshmi SV, Prabakaran TE. Performance Analysis of Multiple Classifiers on KDD Cup Dataset using WEKA Tool. *Indian Journal of Science and Technology,* 2015 Aug; 8(17):1-10.
10. Kaur R, Kumar G, Kumar K. A Novel Feature Selection Technique based on Improved Cuttlefish Algorithm (ICFA) for Intrusion Detection. *Fourth International Conference on Advances in Information Technology and Mobile Communication*, Bangalore, India, Narosa Publishers.
11. Eesa AS, Orman Z. Cuttlefish Algorithm—A Novel Bio-Inspired Optimization Algorithm. *International Journal of Scientific and Engineering Research,* 2013; 4(9):1978-86.
12. Kaur N, Singh W. Alcoholic Behavior Prediction through Comparative Analysis of J48 and Random Tree Classification Algorithms using WEKA. *Indian Journal of Science and Technology*, 2016 Aug; 9(32):1-7.
13. Kaur R, Sachdeva M, Kumar K. Study and Comparison of Feature Selection Approaches for Intrusion Detection. In: *Proceedings of 4th International Conference on Advancements in Engineering & Technology (ICAET-2016),* India. 2016 Mar .
14. Dash P, Pattnaik S, Rath B. Knowledge Discovery in Databases (KDD) as Tools for Developing Customer Relationship Management as External Uncertain Environment: A Case Study with Reference to State Bank of India. *Indian Journal of Science and Technology,* 2016 Jan; 9(4):1-11.
15. Bafna P, Pillai S, Pramod D. Quantifying Performance Appraisal Parameters: A Forward Feature Selection Approach. *Indian Journal of Science and Technology*, 2016 Jun; 9(21):1-7.
16. NSL-KDD. The NSL - KDD intrusion dataset. http://nsl.cs.unb.ca/NSL-KDD/ .
17. Data Mining Software in Java. http://www.cs.waikato.ac.nz/~ml/weka/.
18. Shaveta E, Bhandari A, Saluja KK. Applying genetic algorithm in intrusion detection system: A comprehensive review, 2014.
19. Tsai CF. Data pre-processing by genetic algorithms for bankruptcy prediction. *IEEE International Conference on Industrial Engineering and Engineering Management,* 2011, p. 1780-83.

20. Sang HV, Nam NH, Nhan ND. A Novel Credit Scoring Prediction Model based on Feature Selection Approach and Parallel Random Forest. *Indian Journal of Science and Technology*, 2016 May; 9(20):1-6.