International Journal of Engineering & Technology, 7 (4) (2018) 4448-4454

International Journal of Engineering & Technology

Website: www.sciencepubco.com/index.php/IJET
doi: 17788
Research paper

Ensemble-based framework for intrusion detection system

Pullagura Indira priyadarsini [1] *, K Nikhila [1], P Manvitha [1]

[1] Dept. of Information Technology, Vardhaman College of Engineering, Shamshabad, Hyderabad, Telangana State, India
**Corresponding author E-mail: indupullagura@gmail.com*

Abstract

In this digital age, data is growing as faster as unimaginable. One common problem in data mining is high dimensionality which impacts the quality of training datasets and thereafter classification models. This leads to a high risk of identifying intrusions for Intrusion Detection System (IDS).The probable solution for reducing dimensionality is feature selection. Another considerable cumbersome task for constructing potent classification models from multiclass datasets is the class imbalance. This may lead to a higher error rate and less accuracy. Therefore to resolve these problems, we investigated ensemble feature selection and ensemble learning techniques for IDS. The ensemble models will decrease the hassle of selecting the wrong hypothesis and give a better approximation of the true function. In this paper Prudent Intrusion detection system (PIDS) framework, focusing on ensemble learning is given. It is a two-phase approach. Firstly, the merging of two filtering approaches is done with Ensemble Feature Selection (EFS) algorithm. The proposed EFS algorithm is implemented based on fuzzy aggregation function *Height* with two filtering methods: Canberra distance and city block distance. Later on, classification with Ensemble Classification (EC) algorithm is done with the unification of Support Vector Machines (SVM), Bayesian Network (BN) and K nearest neighbor (KNN). The proposed ensemble method has attained a substantial improvement in accuracy compared to single classifiers. The experiments were performed on EFS+SVM, EFS+BN, EFS+KNN and proposed framework EFS+EC. SVM recorded an accuracy rate of 81% where K-NN recorded 82.8%, Bayes network recorded 84% and our proposed EFS+EC recorded 92%. It is evidenced from the end results that this PIDS framework excels IDS and prevail the pitfalls of SVM, Bayes network and K-NN classifiers.

Keywords: *Accuracy; Bayesian Network; Canberra Distance; City Block Distance; Data Mining; Feature Selection; Fuzzy Logic; FAR; K Nearest Neighbor; Support Vector Machines; Prudent Intrusion Detection System; Precision.*

1. Introduction

Currently, accustoming data mining techniques has fully-grown in Intrusion Detection System edifice. Internet applications need protection against threats from more advanced cyber-attack techniques and computer malware. Eliciting effective and adaptive security mechanisms such as encryption techniques, Firewalls, Intrusion Detection systems (IDSs) are exploited to control them. Among them, IDS has predominated a major role for most of the security infrastructures [1].Therefore, a noteworthy wideness of investigations has been persisting to build intelligent IDSs, which help in achieving superior network security. In the latest researches, Support Vector Machines (SVMs), Neural Networks, fuzzy logic systems are accomplished to be outstanding in many significant prospects of intrusion detection [2] [3].

Machine learning is a wise discipline inclined with the perspective of learning automatically for recognizing complex patterns and making intellectual choices based on data. In reality, there are some problems related to the size of the dataset. Providing learning systems with the whole features creates serious complications to several machine learning algorithms in view of scalability and learning performance. Thus, feature selection is endured to be one of the current demands in machine learning for high-dimensional data sets [4].High dimensional data is a veritable upstream to many prevailing feature selection approaches in terms of efficacy. Feature selection procedures are inevitable in downsizing the dimensionality of the feature space and eliminating the curse of dimensionality risk. They can enhance the potency of a technique to disparities in the training set. Meanwhile, if the number of features is enormous, filter model is adopted, as it is the state-of-the-art model, scalable and rapid [5]. The filtering methods work in relevance independently of the learning algorithm [6] [7]. Thus handling appropriate feature selection methods renders the models to make them feasible to construe, reduce the training times and augment the generalization [8]. There are several feature selection methods in existence which produce better feature subsets among the whole feature space [9]. In view of this, we intelligibly project ensemble approach as a mode to integrate independent feature subsets in order to hopefully get a resilient feature subset. In the recent study Seijo-Pardo et.al [10], ensemble feature selection was given by designing two different ensembles on the feature selection process. In the previous works, Indira et.al has proposed a robust feature selection method. It worked by deploying Robust Feature selection (RFS) algorithm which was an ensemble of three filtering methods namely Euclidean distance, chi-square distance, and correlation coefficient. It can be implemented for reducing the computational complexity and improving the classification accuracy in IDSs [11]. Distance metrics can be used for exploring the best feature subset in Intrusion Detection system domain. In a recent study made by Weller Fahy et.al, various distance measures were reviewed to detect anomalous activities [12].

In the proleptical research, ensemble classification has shown theoretically and realistically better accuracy than any single classifier. The idea of combining responses produced by multiple classifiers into a single response is known as the ensemble ap-

proach [13]. This methodology, named ensemble learning, is built on the hypothesis that merging the output of multiple experts is advantageous than using the output of any single expert. Ensemble learning algorithms have high variance, less over fitting and improvise prediction accuracy. Even though the selection of a suitable grouping technique is a crucial task, several ensemble techniques for combining classifiers have been developed, which is still a heuristic effort [14]. There are mainly two types of ensembles. An ensemble produced from classifiers trained from the same classification algorithm which is known as homogeneous. Example of the homogeneous ensemble will be bagging and boosting. The other is the one generated from classifiers trained from different classification algorithms is called a heterogeneous ensemble. Example of the heterogeneous ensemble is stacking. The outcome of an ensemble classifier relies on the variety of its results of component classifiers and on the selection of the method for combining these outputs into a single one [15].
In recent decades, anomaly-based intrusion detection and many other classification problems have benefited from the idea of combining multiple classifiers. In the work done by Hamed Haddad Pajouh et.al, inferred a two layer model. It is constructed to identify intrusive activities in IoT backbone networks, mainly for identifying low-frequency attacks [16]. A single IDS developed with weak classification algorithm has identified lower detection rate [17].
Right away in this paper, we have built a Prudent Intrusion Detection System (PIDS) framework using Ensemble Feature Selection (EFS) algorithm and Ensemble classification (EC) algorithm. Analyzing advantages of different feature selection methods, the ensemble of two methods is done to a get potent feature set which is certainly useful for better classification. The method used for combining the outputs is based on fuzzy logic. Its main perspective is to select the most optimistic features in KDD cup 99 dataset. An eminent intrusion evaluation dataset, KDD Cup 99 [18] is a classic example of large-scale datasets. Ensemble Classification (EC) is for the process of classifying attack and normal data, through Support Vector Machine (SVM), Bayes Network (BN) and K nearest neighbor classifiers (KNN). Therefore by this ensemble classification method, we have attained better accuracy and lower False Alarm Rate (FAR). The paper is organized as follows. In section 2, related works were given. Followed by, the Methodology for construction of Prudent Intrusion Detection system (PIDS) in section 3. Then in section 4, experiments made and outcomes obtained were discussed precisely. The Last section gives the conclusions and discussions.

2. Related work

In general, several approaches can be used for improving intrusion detection performance, and one of these is classification along with feature selection. Amiri et.al [19] proposed a forward feature selection algorithm using the mutual information method to measure the relation among features. In view of selecting the optimal feature set, with ensemble feature selection techniques, we are truly motivated by the works done by Olsson et.al [20]. They have given ensemble of multiple feature ranking techniques which combines three commonly used filter based feature ranking techniques like document frequency thresholding, infor mation gain, and the chi-square method for text classification problems [20]. Wang et.al has combined ensemble of six commonly used filter based rankers and achieved remarkable results [21]. Former tests presented that integrating feature selection methods could feasibly improve classification accuracy [22]. Two statistical methods viz. Linear Discriminant Analysis (LDA) and Logistic Regression (LR) are applied to develop new intrusion detection models [23].
Ensemble learning has been successfully applied to classification problems. Applying fuzzy methods for the development of IDS is more reliable compared to the classical approaches [24].In recent decades, anomaly-based intrusion detection and many other classification problems have benefited from the idea of combining multiple classifiers. Bukhtoyarov et al. [25] has given ensemble based on Genetic Programming known as (GPEN) which is applied to KDD cup 99 dataset and their goal is to classify the input intrusions as PROBE or non-PROBE attacks, with nine of the 41 features. Borji [26] has proposed an ensemble classification methodology for intrusion detection by using four base classifiers SVM, ANN, k-NN and decision trees to be fused with three combination strategies namely majority voting, belief measure and Bayesian averaging. In the works done in [27], [28] Bayesian network and Random Tree have implemented probity than other classification processes in some aspects. Conversely, we trained and tested with Bayesian network, the results are the mistrial. With the superiority of ensemble learning methods in the global scope and for raising the classifiers assessment we made the proposed investigations.

3. Methodology

The stimulating aspect of applying feature selection techniques is choosing the applicable method for Intrusion Detection System domain. The goal of Prudent Intrusion Detection System (PIDS) is to obtain a system with better accuracy and low false alarm rates. PIDS yields a complete structure for selecting the best features from the KDD cup 99 data set. It thoroughly monitors and classifies the data as an attack and normal. PIDS construction is based on ensemble learning approaches. They are Ensemble Feature Selection (EFS) and Ensemble Classification (EC). It is illustrated by the flowchart in figure 1.

3.1. Retrieval of kdd1 data set

The KDD cup 99 dataset is employed for testing the proposed algorithm. It is the benchmark dataset widely used for IDS evaluation. This is the dataset taken from only 10 percent of the original data set. It contains 494021 instances. The process of retrieving the target KDD1 data set is illustrated in the corresponding figure 2 below.

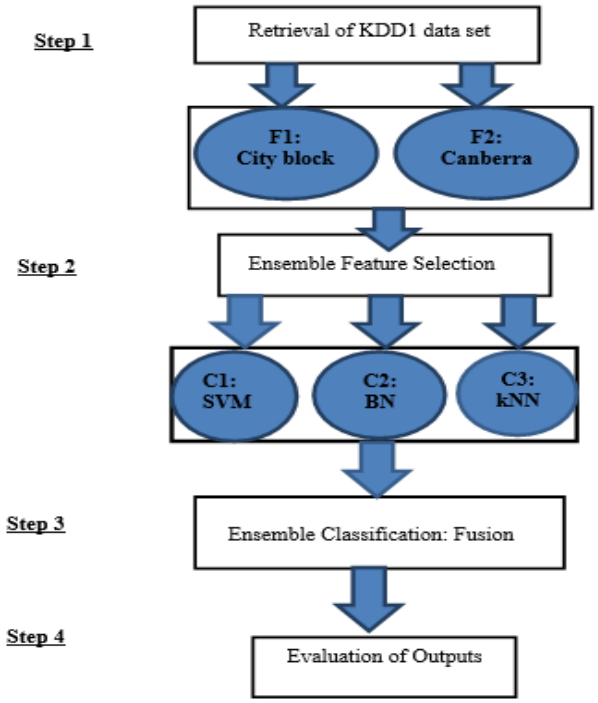

Fig. 1: PIDS Framework.

Firstly collection of KDD cup 99 dataset is done then it is converted to a .csv file. Certain pre-processing techniques like normalization are applied to it. Preserving the values of features in the dataset is done by applying transformation techniques like the discretization of continuous variables. Symbolic values of three fea-

tures have been given numeric ranging from 1 to N. Interquartile range (IQR) has been operated to eliminate noise and outliers in the data set. From this dataset, a portion of it is taken containing 14207 instances with a relative size of records as in KDD cup99 data set. It contains 3000 Normal instances, 10000 DoS instances, 574 probe instances, 401 R2L, and 52 U2R instances. Then Feature rescaling is done for every feature individually. Each class is converted to a numeric value. They will be assigned as "0" for U2R, R2L, Probe, DoS and "1" for Normal. Features are taken as $\{F_1, F_2 \ldots F_{41}\}$. Finally, it is termed as KDD1 dataset.

3.2. Ensemble feature selection

Combining feature selection methods is done to get more stable and robust outputs. Ensembles can be achieved by the aggregation operations. It is done to achieve the merits of two filtering methods such as Canberra distance and city block distance. Fuzzy logic is applied for attaining best features from the dataset. The main idea behind using fuzzy logic is backtracking. In general, some of the features may be dropped out in the conventional method where some threshold is used. Hence, here we give weight to all values.
Aggregation of both filters is done by making use of Height operation of the fuzzy set. From the data set taken Canberra distance is calculated for all the features. It is shown by the EFS algorithm in figure 3 below. Now for the same data set, city block distance is calculated for all the features. For an input pattern {I, J} where I= number of instances and J=number of features i.e., $\{F_1, F2 \ldots F_{41}\}$.

For any particular feature $F_i \in J$, the Canberra distance is computed as

$$D_{ca} = \sum_{i=1}^{I} \frac{mod\ (xi-yi)}{xi+yi} \quad (1)$$

Likewise, city block distance is also computed for the same data set. It is given as

$$D_{cb} = \sum_{i=1}^{I} mod\,(xi - yi) \quad (2)$$

Where x_i is an individual feature in J and y_i is the class label. Here (1) and (2) were calculated for all the 41 features in the data set taken.
So, we get 41 D_{ca} values and 41 D_{cb} values. Then the transformation of those values into fuzzy values is done. This is known as fuzzification. They are termed as fuzzy sets namely f_can_dist, f_cit_dist. The process of transforming is done using trapezoidal membership function. A special case of trapezoidal is L-Function. Suppose x is the element to be transformed then f_x will be (i.e. fuzzy conversion for x) $\frac{x-a}{b-a}$.Here 'a' and 'b' are minimum and maximum values in the whole set. Then in the process of ensemble feature selection, Height of the fuzzy sets 'A' and 'B' is considered as shown in steps 8 and 9 of figure 3 below. It returns the membership value of 1 attained by any point. After that, Union of the resultant values of ς1 and ς2 is taken.

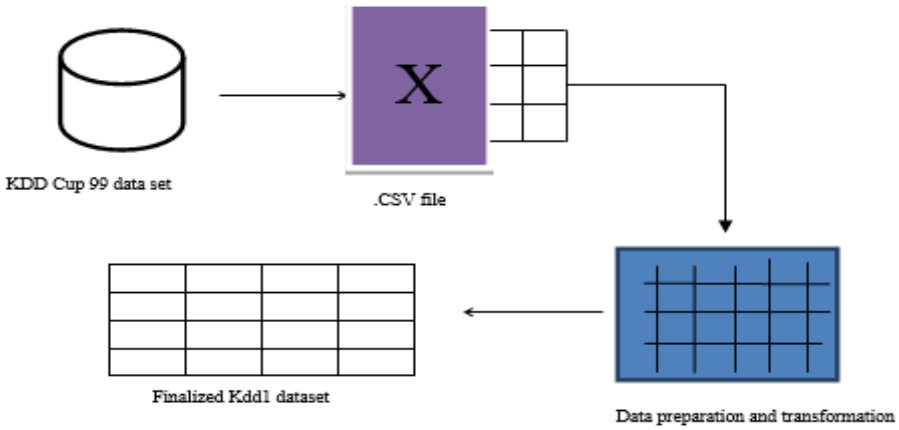

Fig. 2: Collection of KDD1 Dataset for Testing.

Ensemble Feature Selection Algorithm:

Input: KDD1 data set, i, can_dist, cit_dist, f_can_dist, f_cit_dist, A, B, ς1, ς2, ς, α
Output: α

Start:
1. *Take KDD1 dataset.*
2. *Apply necessary pre-processing techniques like normalization to KDD1 dataset*
3. *For i=1 to N do*
4. *can-dist = calculate (feature, class);*
5. *cit_dist = calculate (feature, class);*
   End For
6. *A=Convert can_dist to f_can_dist;*
7. *B=Convert cit_dist to f_cit_dist;*
8. *ς1 =Height (A;*
9. *ς2 =Height (B);*
10. *ς = ς1 U ς2;*
11. *α = {x |x is a corresponding feature of ς/ x ∈ KDD1 dataset};*

End

Fig. 3: EFS Algorithm.

3.3. Ensemble classification

Following the feature selection, Ensemble classification is done. As discussed earlier it combines three classifiers into one. It performs well even though the data set is large. A group of weak classifiers can overrule a strong classifier. So we have taken Bayesian network, SVM, and KNN as base classifiers and built an ensemble.

a) Bayesian classification

The Bayesian network (BN) has been widely used in classification in many areas like pattern recognition, medical diagnosis and information retrieval etc. BN is constructed with directed acyclic graphs (DAGs) and indicate probabilistic relationships between variables in a symbolic pattern. It contains nodes which represent variables and the edges specify dependencies between them. Another main trait of Bayesian networks is that their proficiency to

learn [29]. They are capable of supporting both the graph and the probabilistic tables or functions. A significant risk with Bayesian networks, associated with any others in their formation process. This is commonly accurate when the data on which learning can be done is scanty or missing. Another limitation is that it is their impotence for applying causal loops. This is the reason for complications in encoding certain real-life situations.

b) Support Vector Machine (SVM)

SVM is basically supervised machine learning method anticipated for binary classification. It requires labeled information for efficient learning. These are used in this process since they have an eminent classifying ability with good generalization power. Support Vector Machines (SVMs) are machines that make classification process based on support vectors. These are introduced by the Vapnik [30], [31]. These are built based on Statistical Learning Theory (SLT). They are explicit on training samples and have excellent generalization ability on testing samples. SVMs can create linear and non-linear decision margins using an optimization problem.

c) K-Nearest Neighbor (KNN)

K-Nearest Neighbor is the simple method that can be applied for the data with variable dimensions [32]. The basic idea of KNN is finding n objects from the training data that are nearer to the data testing. It works based on the nearest distance. So it is very essential to pick the number of k-nearest neighbors which has the larger effect on predicted results in KNN. Small values of k can produce a great variety on the prediction results, whereas a large value of k can lead to a large bias of models.

A dataset KDD1 with n no. of tuples and α no. of features is given as an input to the EC algorithm. The class label will be 0 or 1.EC algorithm is stated in figure 4 below. From the EC algorithm, three models are built and fused. There are three local decisions namely y_1,y_2, y_3. Each local decision y_i is labeled as xx1 or xx2. Here xx1 stands for attack and xx2 stands for non-attack. Then y_i can be 0 or 1. If y_i =0 it is attack (means it can be either DoS or Probe or U2R or R2L). Whereas y_i =1 indicate normal.

Then the final decision in the proposed ensemble classification (EC) is obtained by combining the models by exploiting the weighted average voting method. After building ensemble classification (EC), we predict the class labels based on the predicted probabilities p for classifier.

$$y=argmax \sum_{j=1}^{m} w_j p_{ij} \quad (1)$$

Where w_j is the weight that can be assigned to the j^{th} classifier and 'm' is the number of classifiers and i={0,1}.

At the testing part, instances of the KDD1 data set are fed to the suggested Ensemble Classification (EC) process by leaving their class-label to which they exist. This ensemble classifier anticipates the network traffic data as normal or as an attack.

Ensemble Classification Algorithm:

Input: *KDD1 dataset with 'α' features, class label*
y_1, y_2, y_3 ,xx1=attack, xx2= non-attack, Final decision

Output: *accuracy, FAR, DR, specificity, precision*

Start:

1. *Take KDD1(n x α)data set;*
2. *Classify KDD1dataset:*
3. *Pass the data records to Bayes classification Algorithm;*
4. *Iterate step 3 on remaining classification algorithms SVM, KNN;*
5. *For every input data instance, there are three local decisions y_1,y_2, y_3 from three classifiers.*
6. *The local decision y_i is labeled as xx1 or xx2*
7. *If (y_i ="DOS"|"PROBE"|"R2L"|"U2R")*
8. *y_i= xx1;*
9. *Else*
10. *y_i =xx2;*
11. ***Fusion***: Final decision=argmax$\sum_{j=1}^{m} w_j p_{ij}$ *i*
12. *Evaluation Measures: accuracy, FAR, DR, specificity, precision*

End

Fig. 4: EC Algorithm.

4. Experiments conducted & result obtained

For performing experiments, we have used KDD1 dataset. The procurement of KDD1 dataset is mentioned in the earlier section 3.1. In the entire experiments conducted 10 fold cross validation is made for testing our PIDS model. The 10 fold cross validation is also known as rotation estimation. It is a recommended method over holdout method and leave-one-out methods for estimating a classifier. The dataset has been split at random into ten parts of the same size. Every part is kept out in turn and the training is conducted on remaining 9 parts, then the testing is made on holdout set. The training is made totally 10 times on different training sets and lastly, the average of ten error rates is considered for attaining complete error estimate. Totally, we have made four experiments. 1. With EFS outputs given to SVM, 2.With EFS outputs given to Bayes network classification, 3.With EFS outputs given to K-NN and 4. The Proposed PIDS (EFS+EC).All the four experiments were made using WEKA machine learning tool [33].

From the EFS algorithm, the fuzzy outcomes for filter1 i.e. Canberra distance and filter2 i.e. city block distance as mentioned in lines 6 & 7 of the EFS algorithm are given with the graphical interpretation below in figures 5 and 6.

Then the resultant values ς1and ς2 will be obtained by considering height of fuzzy set A and fuzzy set B respectively. So, we get ς1=duration, service, src_bytes, inum_file_creations, dst_host_srv_count and dst_host_srv_rerror_rate. ς2=dos,num_failed_logins, isu_attempted, inum_outbound_cmds and is_host_login. Therefore, ς will be obtained as the union of ς1, ς2. The selected features are 11. They are duration, service, src_bytes, dos, num_failed_logins, isu_attempted, inum_file_creations, inum_outbound_cmds, is_host_login, dst_host_srv_count and dst_host_srv_rerror_rate.

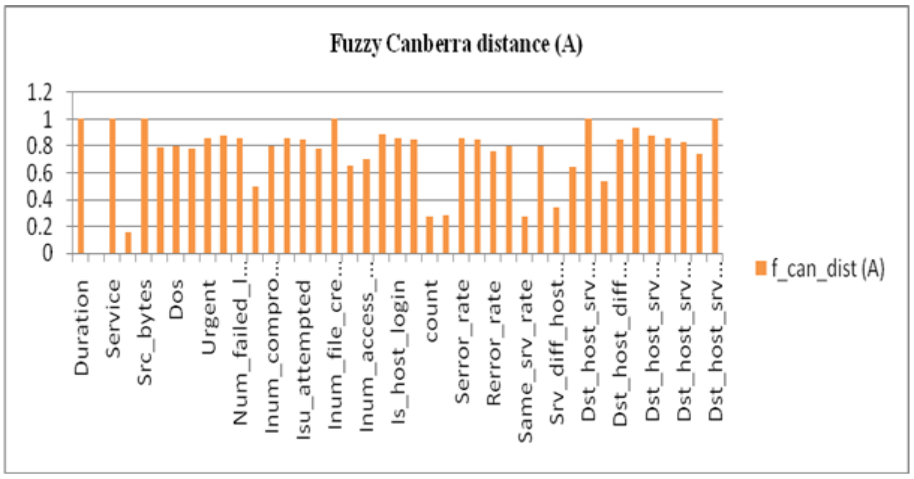

Fig. 5: Fuzzy Values Obtained on Canberra Distance for 41 Features.

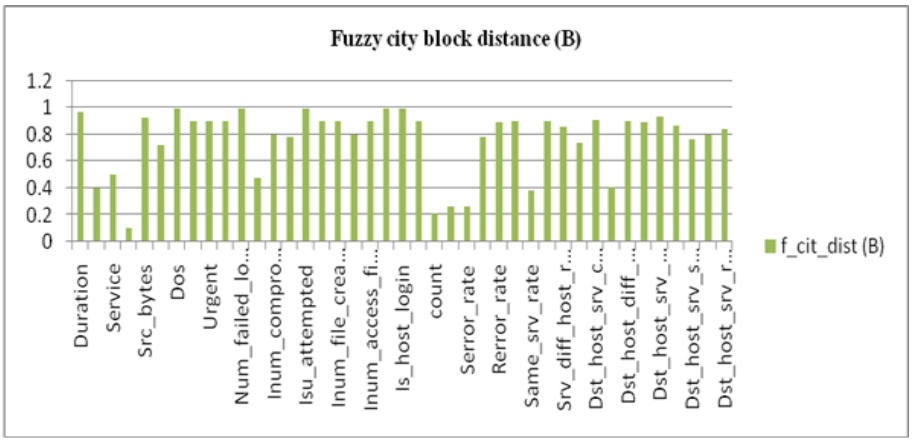

Fig. 6: Fuzzy Values on City Block Distance for All the Features.

The key measure for assessing classifier efficiency is Accuracy: It is given as the ratio of test samples to decorously estimate the label of newly or previously unseen data. TP stands for actual normal instances that were correctly predicted as Normal. TN stands for attacks that were correctly classified as non-normal. FP stands for actual attacks that were incorrectly labeled as Normal. FN stands for Actual normal instances that were incorrectly predicted as attacks. The accuracy rate is specified as:

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

In addition, for evaluating classification results we have used Specificity, Detection Rate (DR), Precision and False Alarm Rate (FAR). Since Sensitivity, Specificity and Precision are suitable alternatives to the accuracy measure particularly when the data set is imbalanced and the main class of interest is in the minority. They are defined as follows:

- Precision: TP/ (TP+FP).It is given as the ratio of items correctly classified as X to all items classified as X.
- Detection Rate: TP/(TP+FN).It is stated as the ratio of items correctly classified as X to all items that belong to class X.
- Specificity: TN/ (TN+FP).It is specified as the ratio of items correctly classified as negatives (not X) to all items that belong to class, not X.
- FAR: FP/ (TN+FP). It is given as FAR = 1-Specificity. And it is the ratio of items incorrectly classified as positives (X) to all items that belong to a class, not X.

Comparison of performance of all the four experiments on the KDD1 dataset using Accuracy, Detection Rate (DR), False Alarm rate (FAR), Precision and Specificity is shown in the below figure 7. The Accuracy, Detection Rate (DR), FAR, Precision and specificity of the proposed PIDS model is illustrated in figure 8.

For the data set chosen, our proposed ensemble framework outperformed K-NN, SVM, and Bayes classifications with good accuracy rate. SVM recorded an accuracy rate of 81% where K-NN recorded 82.8%, Bayes network recorded 84% and our proposed EFS+EC recorded 92%. It is proved from the consequences by our investigational analysis that this PIDS framework performs profusely on Intrusion detection system and seizes the pitfalls of SVM, Bayes network and K-NN classifiers. Proposed work has yielded a high detection rate and lower false alarm rate. Ensemble of SVM, K-NN, and BN has proved that it is successful for IDS.

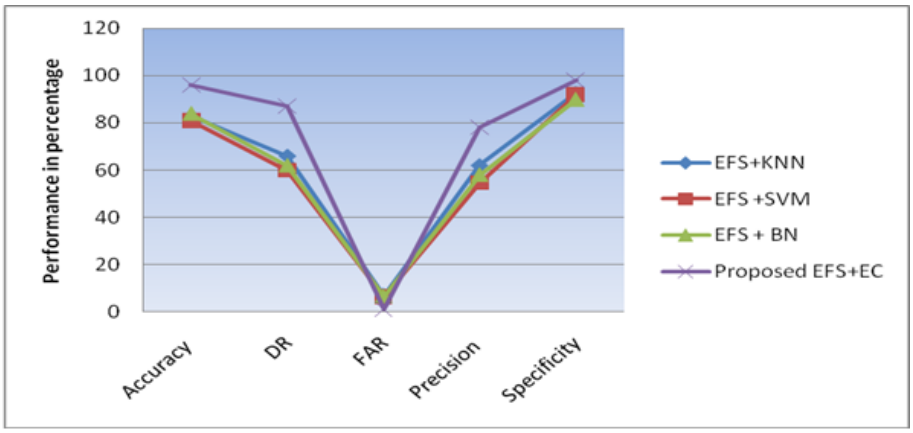

Fig. 7: Performance Evaluation Using Accuracy, DR, FAR, Precision and Specificity of Five Models.

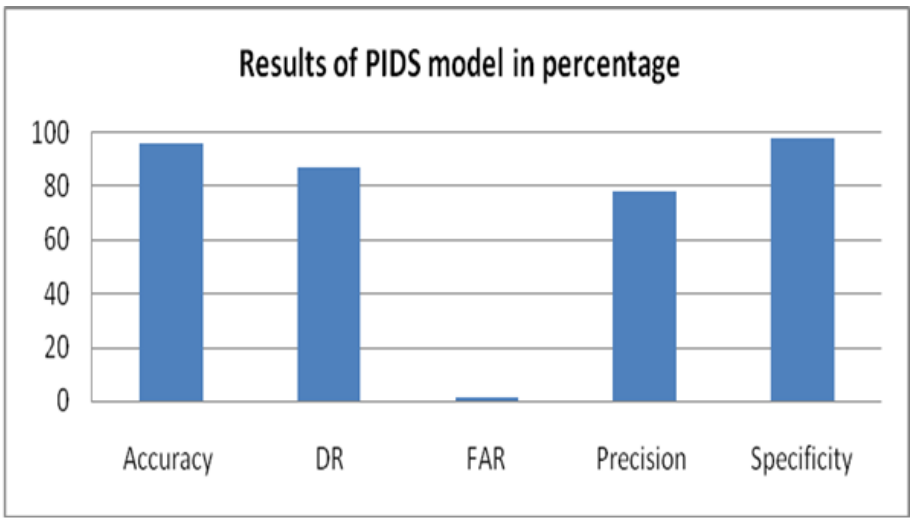

Fig. 8: The Results of Proposed PIDS Framework.

5. Conclusions and discussions

In this article, we have described the feature selection ensemble that is potent than the diverse individual methods. Fuzzy logic is applied for attaining best features from the dataset. The ensemble work with two filtering methods is done with an aggregator using Height to yield a common final output. In this work, we proposed a novel classifier ensemble method for intrusion detection that is diversified by using three different approaches. The comparison of results on EFS on individual classifiers with proposed PIDS has shown with respect to accuracy, Detection Rate, FAR, precision and specificity. The results indicate that proposed ensemble classification overrules SVM, K-NN and Bayes network classifiers. For performing experiments the data set chosen is a KDD Cup 99 data set.

As a future work, it can be further extended to adjust the ensemble size dynamically according to the dimension of the dataset. That is, the decision of a number of base classifiers to be used for constructing ensemble should be done dynamically. Hence adaptively changing the size by analyzing these factors will help to improve performance with relatively less overhead. Future effort will comprise conducting superfluous empirical studies with data from other software projects and application fields and experimentations with other learners.

References

[1] J. McHugh, A. Christie, and J. Allen, "Defending Yourself: The Role of Intrusion Detection Systems", *IEEE Software*, Sept. Oct. 2000, pp. 42-51. https://doi.org/10.1109/52.877859.

[2] Mukkamala S, Janoski G, and Sung AH, "Intrusion Detection Using Neural Networks and Support Vector Machines", *Proceedings of IEEE International Joint Conference on Neural Networks*; 2002, pp. 1702-1707. https://doi.org/10.1109/IJCNN.2002.1007774.

[3] A. Chandrasekhar, K. Raghuveer, "An effective technique for intrusion detection using neuro-fuzzy and radial svm classifier", Computer Networks & Communications (NetCom), Vol. 131, Springer, 2013, pp. 499–507.

[4] A. Kalousis, J. Prados, and M. Hilario, "Stability of feature selectionalgorithms: a study on high-dimensional spaces," Knowl. Inf. Syst., vol. 12, no. 1, 2007, pp. 95–116. https://doi.org/10.1007/s10115-006-0040-8.

[5] KhalidS,Khalil,T, &Nasreen S.,"A survey of feature selection and feature extraction techniques in machine learning", Science and Information Conference (SAI) 372– 378, 2014. https://doi.org/10.1109/SAI.2014.6918213.

[6] Mohammed A.Ambusaidi, Xiangjian He, Priyadarsi Nanda, Zhiyuan Tan,"Building an Intrusion Detection System Using a Filter-Based Feature Selection Algorithm", IEEE Transactions on Computers, Vol. 65, I (10), 2016.

[7] RonKohavi, George H. John "Wrappers for feature subset selection", Artificial Intelligence 97 pp.273-324, 1997. https://doi.org/10.1016/S0004-3702(97)00043-X.

[8] X. Jin, A. Xu, R. Bie1 and P. Guo, "Machine Learning Techniques and Chi-Square Feature Selection", Springer- Verlag Berlin Heidelberg LNBI 3916, (2006), pp. 106 – 115.

[9] T. G. Dietterich, "Ensemble methods in machine learning," in Proceedingsof the First International Workshop on Multiple Classifier Systems. London, UK, UK: Springer-Verlag, pp. 1–15, 2000. https://doi.org/10.1007/3-540-45014-9_1.

[10] B. Seijo-Pardo, I. Porto-Diaz, V. Bolon-Canedo, A. Alonso-Betanzos, "Ensemble Feature Selection: Homogeneous and Heterogeneous Approaches",Knowledge-Based Systems 2016, https://doi.org/10.1016/j.knosys.2016.11.017.

[11] Pullagura Indira priyadarsini, M.SeshaSai, A. Suneetha, M.V.B.T.Santhi "Robust Feature Selection Technique for Intrusion Detection System", International journal of control and automation (IJCA) Vol.11, no.2, 2018 pp.33-44.

[12] D. J. Weller-Fahy, B. J. Borghetti, and A. A. Sodemann, "A Survey of Distance and Similarity Measures Used Within Network Intru-

sion Anomaly Detection," IEEE Commun. Surv.Tutor, vol. 17, no. 1, 2015, pp. 70–91. https://doi.org/10.1109/COMST.2014.2336610.
[13] M.G. Ouyang, W.N. Wang and Y.T. Zhang, "A fuzzy comprehensive evaluation based distributed intrusion detection", *Proceedings First International Conference on Machine Learning and cybernetics*, China, Beijing, 2002, pp. 281-285. https://doi.org/10.1109/ICMLC.2002.1176757.
[14] M.Govindarajan, "Evaluation of Ensemble Classifiers for Intrusion Detection", World Academy of Science, Engineering and Technolog International Journal of Computer and Information Engineering, Vol: 10, No: 6, 2016.
[15] Y. Chen, M.-L. Wong, H. Li, "Applying Ant Colony Optimization to configuring stacking ensembles for data mining", Expert Syst. Appl. 41 (6), 2014, pp. 2688–2702,. https://doi.org/10.1016/j.eswa.2013.10.063.
[16] Hamed Haddad Pajouh, Reza Javidan, RaoufKhaymi, Ali Dehghantanha and Kim-Kwang Raymond Choo, "A Two-layer Dimension Reduction and Two-tier Classification Model for Anomaly-Based Intrusion Detection in IoT Backbone Networks",IEEE,2016, https://doi.org/10.1109/TETC.2016.2633228.
[17] Christopher et.al."IDS and Correlation, Challenges and solutions",Vol 14,AISC,Springer, 2005.
[18] MahbodTavallaee, Ebrahim Bagheri, Wei Lu, and Ali A. Ghorbani "A Detailed Analysis of the KDD CUP 99 Data Set", Proceedings of the IEEE Symposium on Computational Intelligence in Security and Defense Applications, 2009.
[19] F. Amiri, M. Rezaei Yousefi, C. Lucas, A. Shakery, N. Yazdani, "Mutual information-based feature selection for intrusion detection systems", Journal of Network and Computer Applications 34 (4),1184–1199, 2011. https://doi.org/10.1016/j.jnca.2011.01.002.
[20] J. O. S. Olsson and D.W. Oard," Combining feature selectors for text classification",*CIKM* '06: Proceedings of the *15thACM international conference on Information and knowledge management*, pages 798–799, New York, NY, USA, 2006.
[21] H.Wang, T. M. Khosh goftaar, and K. GAO. "Ensemble feature selection technique for software quality classification", In *Proceedings of the 22nd International Conference on Software Engineering and Knowledge Engineering*, pages 215–220, Redwood City, CA, USA, July 1-3 2010.
[22] Z. Karimi and A. Harounabadi, "Feature Ranking in Intrusion Detection Dataset using Combination of Filtering *Methods", International Journal of Computer Applications* (0975 – 8887), vol. 78, Iss (4), pp. 21–27, 2013.
[23] Basant Subba, S.B., Sushanta Karmakar, "Intrusion Detection Systems using Linear Discriminant Analysis and Logistic Regression", in INDICON. 2015, IEEE.
[24] L. Kuncheva, "Combining Pattern Classifiers: Methods and Algorithms", Wiley-Interscience, 2004. https://doi.org/10.1002/0471660264.
[25] V. Bukhtoyarov, V. Zhukov, "Ensemble-distributed approach in classification problem solution for intrusion detection systems", Intelligent Data Engineering and Automated Learning-IDEAL 2014, Springer, pp. 255–265.
[26] A. Borji. , "Combining Heterogeneous Classifiers for Network Intrusion Detection", in Proceedings of the Annual Asian Computing Science Conference, pp 254-260. Springer, Berlin, Heidelberg, 2007, Dec. https://doi.org/10.1007/978-3-540-76929-3_24.
[27] Sumaiya Thaseen, C.A.K., "An Analysis of Supervised Tree Based Classifiers for Intrusion Detection System" International Conference on Pattern Recognition, Informatics and Mobile Engineering (PRIME), IEEE, 2013, pp. 294-299.
[28] Sumouli Choudhury, A.B., Comparative Analysis of Machine Learning Algorithms along with Classifiers for Network Intrusion Detection", International Conference on Smart Technologies and Management for Computing, Communication, Controls, Energy and Materials (ICSTM), pp. 89-95, 2015.
[29] Gregory F. Cooper and Edward Herskovits,"A Bayesian method for the induction of probabilistic networks from data", Machine Learning, 1992. https://doi.org/10.1007/BF00994110.
[30] Boser, Guyon, and Vapnik, "A training algorithm for optimal margin classifiers", Proceedings of the fifth annual workshop on Computational learning theory.pp.144-152, 1992.
[31] Cortes C., Vapnik V., "Support vector networks, in Proceedings of Machine Learning20: pp.273–297, 1995.
[32] Alkhatib K, Najadat H, Hmeidi I, Shatnawi MKA. Stock Price Prediction Using K-Nearest Neighbor (kNN) Algorithm. International Journal of Business, Humanities and Technology. 3 (3), , 2013,pp.32 – 44.
[33] http://www.cs.waikato.ac.nz/ml/weka/.