Computing, Information Systems, Development Informatics & Allied Research Journal
Vol. 7 No. 4, December, 2016 - www.cisdijournal.net

Influence of Feature Selection on Multi-Layer Perceptron Classifier for Intrusion Detection System

Mabayoje, M.A., Balogun, A.O., Ameen, A.O. & Adeyemo, V.E.
Department of Computer Science,
University of Ilorin
Ilorin.
Kwara State, Nigeria
mmabayoje@gmail.com, balogun.ao1@unilorin.edu.ng, shaksoft@yahoo.com,

ABSTRACT

The usage of the most popular neural network – Multilayer perceptron, as gained ground for the purpose of detecting intrusion. A lot of researchers had used it judiciously but there exist problem of slow training time and data over-fitting. This paper reviews the various data mining techniques for applied in the area intrusion detection, categories of attacks, and techniques for feature selection. This paper proposes an architecture where information gain is used for feature selection and multilayer perceptron (MLP) for classification on KDD'99 dataset. Evaluation of the performance of the MLP classifier on the KDD'99 dataset and also on the reduced dataset was conducted.

Keywords: Data mining and IDS, Intrusion Detection System, MLP, Classification Techniques for IDS.

CISDI Journal Reference Format
Mabayoje, M.A., Balogun, A.O., Ameen, A.O. & Adeyemo, V.E. (2016): Influence of Feature Selection On Multi-Layer Perceptron Classifier for Intrusion Detection System. Computing, Information Systems, Development Informatics & Allied Research Journal. Vol 7 No 4. Pp 87-94
Available online at www.cisdijournal.net

1. INTRODUCTION

The rapid growing rate of networks of computer around the world poses security as fundamental issue of computer technology. Hence, security of data as well as data integrity, data confidentiality and data availability is being provide by technology expert [1]. Intrusion detection and prevention is a developing field as it tends to be of a consideration nowadays due to the prevalent activities of hacker. Moreover, the usage of an intrusion detection system for securing a network is not of utmost importance as the rate at which the said IDS can effectively and efficiently performing it duties. Several Intrusion Detection Systems have been employed to classify attacks into several categories, but the researcher is keen on how an intrusion detection system can improve its performance when performing its duties.

Multi-layer Perceptron algorithm is a supervised learning technique. A classification based IDS is capable of classifying a multi-class dataset and can also classify all the network traffic into either normal or malicious [2]. Influence of feature selection on multi-layer perceptron classifier for detecting intrusion in a system is a medium to reveal how an improvement can be made on an existing machine learning (classification) algorithm (i.e. multi-layer perceptron) to ensure accuracy and correctness with lesser resources. In this paper, MLP classifier will be trained and tested on KDD dataset, the research will try to detect attacks on the four attack categories: DoS (Denial of service), Probe (information gathering) R2L (remote to local) and U2R (user to root), [3], not forgetting the Normal category too.

These four attacks have distinct unique execution dynamics and signatures, which motivates the researcher to discover if in fact certain, but not all, features will majorly participate in ensuring correct classification of the type of attack. These features that contribute majorly are selected out of the available feature and all others are discarded, then a comparison of the results of the performance of the MLP is carried out after feeding it with the reduced dataset and the full dataset separately.

This reminder of this paper is organized as follows: A concise review of related work was made in Section 2. Section 3 will detail about our simulation study (proposed system architecture, evaluation setup, implementation of MLP and application of early stopping validation technique, and performance comparison). The results of the MLP classifier when feed with both the full and reduced dataset as input will be analyzed and compared in Section 4. And finally, Section 5 will conclude our study and discuss the future works.

2. RELATED WORKS

Huy and Deokjai (2008), conducted a research on ten classification algorithms used for intrusion detection and evaluated their performances using the KDD99 dataset. Based on the attacks category, they chose the best algorithms and proposed a two classifier algorithm selection models [3]. Yogendra and Upendra (2012), studied and analyzed few data mining classification algorithms (NB, OneR, BayesNet, and J48) in order to detect intrusions and thereafter compare their relative performances. They pointed out that J48 decision tree out-performed other three algorithms [4].

Moradi and Zulkernine proposed a neural network approach to intrusion detection, they used MLP for detecting intrusion based on an off-line analysis approach. They focused on solving a problem of multiclass in which the type of attack is also detected by the neural network aside from classification of records in one of the two general classes – normal and attack [5].bPurva and Priti in 2013 developed an application software for detecting intrusion through the usage of MLP algorithm based on Back Propagation. They proposed a system that not only detect attacks but also classify them in 6 groups with the accuracy of approximately 83% with the two hidden layers for the neural networks. And also see how live detection of ICMP attacks using Snort IDS [6].

Adsul, Danke, Jagdale, Chaudlhari, and Jadhav (2014) in their work, made a new approach of intrusion detection system based on artificial neural networks. They utilized MLP for detecting intrusion, and the designed system detect the attacks and classify them in six groups with the two hidden layers of neurons in the neural network [7]. This work is similar to that of Devikrishna and Ramakrishna (2014) [2]. Aida, Ahmed and Tamer (2010) in their work stated that an IDS's responsibility is to detect suspicious or unaccepted system and network activity and to alert a systems administrator to this activity. They evaluated the performance of nine NNs based classifiers, based on a selected group of features. They reveal in their result that; the Multilayer perceptron (MLPs) based classifier had about 99.63% true positive thereby having the best results with [8].

Heba, Sherif and Mohamed (2012) designed a multi-layer intrusion detection model, aimed at improving the detection and classification rate accuracy and also achieving high efficiency. They made used of the Naïve Bayes, Multilayer Perceptron neural network, and C5 decision tree; gain ratio was used to best selected features for getting high intrusion detection performance. The results they presented indicated that the proposed model achieved higher classification rate accuracy, and less false alarm rate than Naïve Bayes and MLP. They also pointed out that Gain Ratio enhanced the accuracy of U2R and R2L for the three machine learning techniques significantly. The classification rate of MLP was high when using the whole 41 features in Dos and Probe layers [9].

3. PROPOSED SYSTEM ARCHITECTURE

The figure 3.1 shows the proposed architecture for detecting and classifying attacks.

- **The dataset** use was the 10% of KDD99 which is the mostly widely used data set containing 42 features (with label). This dataset is being feed into the MLP classifier for training and testing.
- **The training and testing layer** made used of cross validation technique (10 folds) which divided the dataset into 10 segments in which 9 segments are used for training and the last one for testing
- **The classifier layer** involved the usage of MLP algorithm for detecting and classifying intrusion.
- **Feature selection layer** provided the removal of redundant and not important attributes in the dataset. Feature selection is used in order to decrease the dimensionality of a dataset and increase its accuracy and performance of the MLP classifier
- **Result analysis** layer provide the performance evaluation process for the MLP classifier when being feed with all features as input and also when being feed with the reduced dataset.

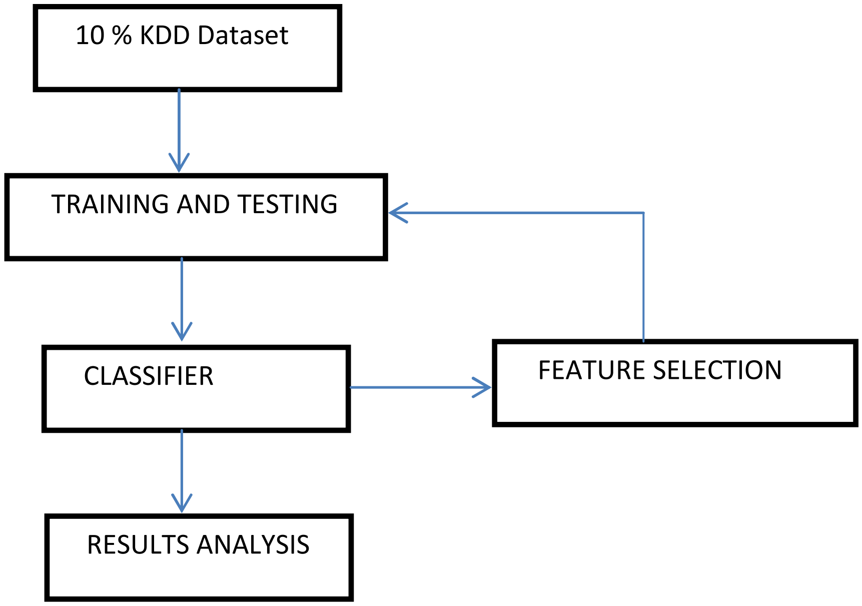

Figure 3.1: System Architecture

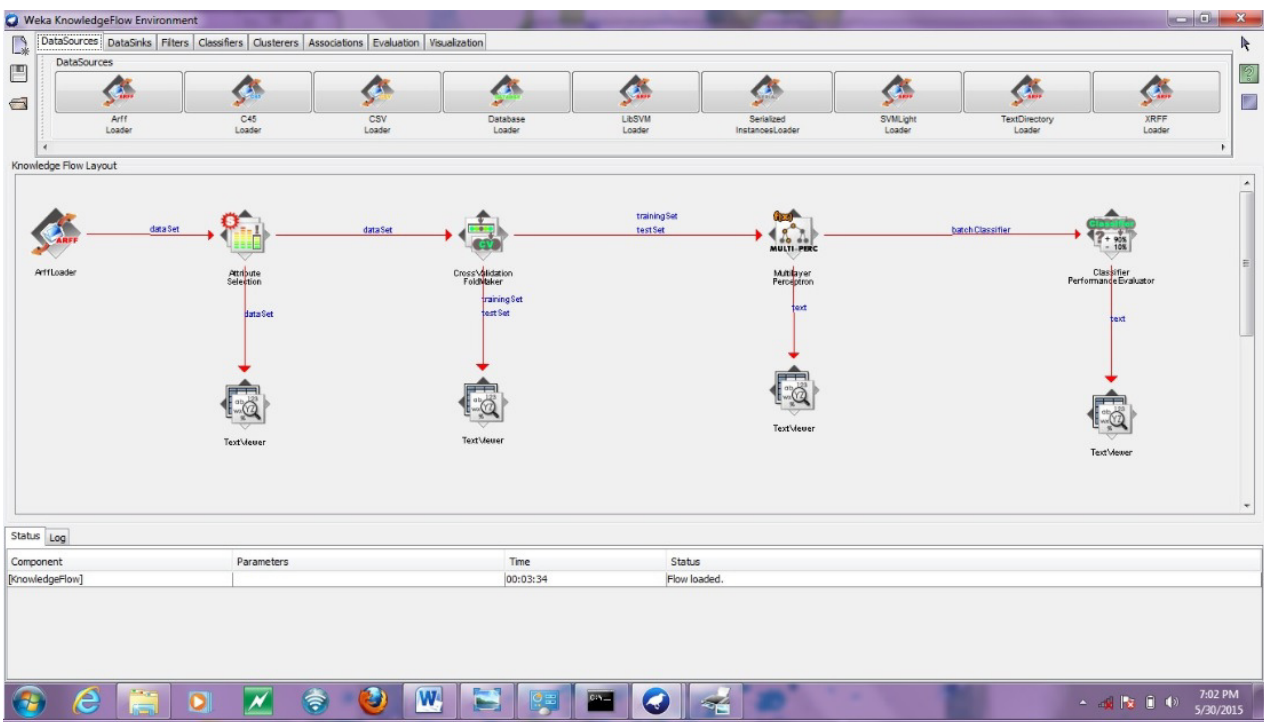

Figure 3.2 Proposed System Architecture (WEKA).

3.1 Evaluation Setup

The experiments were carried out on a HP probook 6470b laptop with the following configurations Intel(R) Core(TM)i5-3230M, CPU 2.60GHz, 6GB RAM (5.55 GB usable), 64-bit operating system whose platform is Microsoft Windows7 Professional (Service Pack 1). The latest Weka – an oen source machine learning package was used for setting up the experimental and evaluation environment (Weka 3.6.11). Weka is a software that holds machine learning algorithms for data mining tasks containing tools for visualization, data preprocessing, regression, classification, association rules, and clustering.

3.2 Implementation of MLP and Application of Early Stopping Validation

According to Moradi and Zulkernine, a common problem that can occur while training neural network is over-fitting. An over fitted ANN training set as error i.e. number of incorrectly classified patterns, that is driven to a very small value, however, when new data is presented, the error become large. In these cases, the ANN has memorized the training examples but has not learnt to generalize the solution to new situations.

The solution to the over-fitting problem of ANN is to find the suitable number of training epochs by trial and error, and another more reasonable method for improving generalization is called early stopping. This technique divides the available data is divided into three subsets. The first subset is used for training and updating the ANN parameters called "training set", the second subset is the validation set whose error is monitored during the training process (the validation error will normally decrease during the initial phase of training similar to the training set error) and the last subset is the "test set".

However, when the ANN begins to over-fit the data, the error on the validation set will typically begin to rise and when the validation error increases for a specified number of iterations, the training is stopped, and the weights that produced the minimum error on the validation set are retrieved [5]. In this paper, the training-validation strategy was adopted. MLP was made up of three layer feed-forward network, each layer serve as input, hidden and output, having the following parameters set for the model are momentum =0.2; ; validationThreshold =20; randomSeed = 0 ; learning rate = 0.3; epoch = 50

3.3 Performance Comparison

The performance of MLP on each dataset i.e. the full (containing all the features and the reduced dataset (containing 12 features plus label), will be evaluated and measured via the following parameters: incorrectly classified instances (%),correctly classified instances (%),root mean squared error, relative absolute error, kappa statistics, root relative squared error and measured via the following parameters: TP (True Positive) rate, FP (False Positive) rate, Precision, Recall, F-Measure and TT (Training Time of the algorithm on each dataset), and AA (Average Accuracy = Total correctly classified instances/Total instances).

4. ANALYSIS OF RESULTS

The Tables I,II,III,IV displays the performance of MLP based on the two distinct dataset mentioned earlier, and the table V is derived from all the previous tables.

Table I: Performance evaluation of MLP on the full dataset

PARAMETERS	DOS	NORMAL	PROBING	R2L	U2R
CORRECTLY CLASSIFIED INSTANCES (%)	99.9930	100	99.0748	97.7798	71.1538
INCORRECTLY CLASSIFIED INSTANCES (%)	0.0062	0	0.9252	2.2202	28.8462
KAPPA STATISTICS	0.9998	1	0.9866	0.8683	0.4564
MEAN ABSOLUTE ERROR	0.0001	0.0001	0.0035	0.0061	0.0411
ROOT MEAN SQUARED ERROR	0.0021	0.0004	0.0269	0.0375	0.1331
RELATIVE ABSOLUTE ERROR	0.3297	629.333	5.8164	36.1037	66.0513
ROOT RELATIVE SQUARED ERROR	1.5411	633.909	15.5228	42.7331	79.9266

Table II: Performance measurement of MLP on the full dataset

PARAMETERS	DOS	NORMAL	PROBING	R2L	U2R
TP RATE	1	1	0.991	0.978	0.712
FP RATE	0	0	0.002	0.086	0.259
PRECISION	1	1	0.991	0.962	0.655
RECALL	1	1	0.991	0.978	0.712
F-MEASURE	1	1	0.991	0.969	0.667
ROC AREA	1	0	0.999	0.99	0.845
Training Time	4170.33secs	926.9secs	49.42secs	18.14secs	40.59secs

Table III: Performance evaluation of MLP on the reduced dataset

PARAMETERS	DOS	NORMAL	PROBING	R2L	U2R
CORRECTLY CLASSIFIED INSTANCES (%)	99.9938	100	97.0782	97.6909	73.0769
INCORRECTLY CLASSIFIED INSTANCES (%)	0.0062	0	2.9218	2.3091	26.9231
KAPPA STATISTICS	0.9998	1	0.9577	0.8617	0.5117
MEAN ABSOLUTE ERROR	0.0001	0.0003	0.0054	0.0065	0.0423
ROOT MEAN SQUARED ERROR	0.0021	0.007	0.0391	0.0396	0.1291
RELATIVE ABSOLUTE ERROR	0.2974	1294.665	9.0433	38.4633	67.9958
ROOT RELATIVE SQUARED ERROR	1.5502	1305.3079	22.5764	45.1331	77.5406

Table IV: Performance measurement of MLP on the reduced dataset

PARAMETERS	DOS	NORMAL	PROBING	R2L	U2R
TP RATE	1	1	0.971	0.977	0.731
FP RATE	0	0	0.008	0.111	0.236
PRECISION	1	1	0.97	0.961	0.679
RECALL	1	1	0.971	0.977	0.731
F-MEASURE	1	1	0.971	0.969	0.7
ROC AREA	1	0	0.999	0.981	0.89
Training Time	2057.11secs	131.5secs	26.55secs	10.66secs	10.98secs

Table V: Summary of the results derived from the tables above.

CLASSIFIER	ATTACK TYPES				
	DOS	NORMAL	PROBING	R2L	U2R
MLP (full dataset) Accuracy	99.9930	100	99.0748	97.7798	71.1538
Training Time	4170.33secs	926.9secs	49.42secs	18.14secs	40.59secs
MLP (reduced dataset) Accuracy	99.9938	100	97.0782	97.6909	73.0769
Training Time	2057.11secs	131.5secs	26.55secs	10.66secs	10.98secs

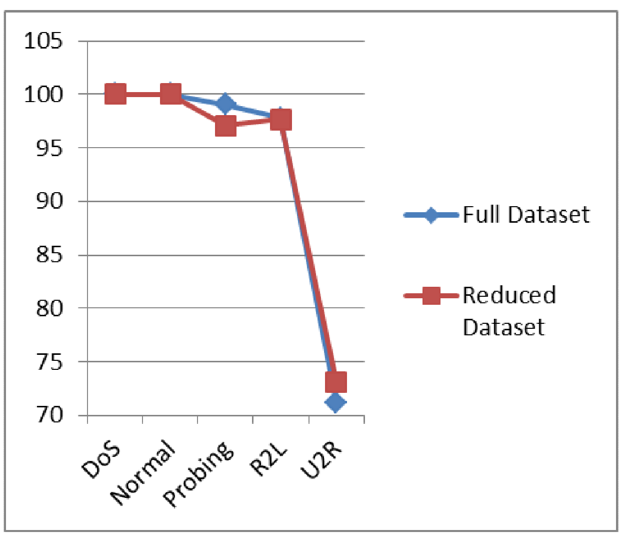

Figure 3.2: Comparison of MLP accuracy on full and Reduced dataset.

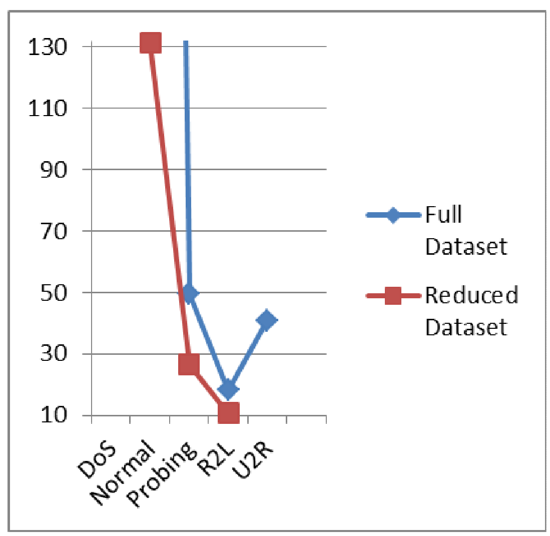

Figure 3.3: Comparison of MLP training time On full and reduced datas

5. CONCLUSIONS AND FUTURE WORKS

For this study, the dimensionality of the dataset was reduced using the information gain technique for reduction of the attributes in a dataset. This study approached the influence of this feature selection technique on the classification of attack by the Multilayer Perceptron algorithm. Our simulations showed that, the dataset that has its attributes filtered has the lowest training time and in most cases had an improved accuracy compared to the full KDD'99 dataset when being classified by MLP. It is considered that for future research, another type of feature selection technique can be used for filtering the dataset.

REFERENCES

1. Patel Hemant, Bharat Sarkhedi, and Hiren Vaghamshi (2013) "Intrusion Detection in Data Mining with Classification Algorithm". IJAREEIE, Vol. 2, Issue7, July 2013.
2. Krzystof Grabczewski and Nobert Jankowski, (2012). "Feature selection with Decision Tree Criterion". 2012.
3. Y. Ma, D. Choi, and S. Ata (Eds.): APNOMS 2008, "Application of Data Mining to Network Intrusion Detection: Classifier Selection Model", LNCS 5297, pp. 399–408, 2008.
4. Yogendra Kumar Jain and Upendra (2012), "An efficient Intrusion Detection Based on Decision Tree Classifier Using Feature Reduction". IJSRP, Vol. 2, Issue 1, January 2012.
5. Moradi Mehdi and Zulkernine Mohammed, "A Neural Network Based System for intrusion Detection and Classification of Attacks".
6. Purva Adlakha and Priti Subramanium (2013), "Detecting and Classifying Attacks in Network Intrusion Detection System Using Multi-layer Perceptron Based on Aritifical Neural Network" IJARCSSE, Vol. 3, Issue 6, June 2013.
7. Dr.A.P. Adsul,Pooja Danke,Meghana Jagdale,Kuldeep Chaudhari, and Samarth Jadhav (2014), "Attacks Classification in Network Intrusion Detection System Using ANN", IJAIEM, Vol. 3, Issue 4, April 2014
8. Aida O. Ali, Ahmed Saleh, and Tamer Ramdan (2010), "Multilayer perceptrons networks for an Intelligent Adaptive intrusion detection system", IJCSNS, Vol. 10, No. 2, February 2010.
9. Heba Ezzat Ibrahim, Sherif M. Badr, and Mohamed A. Shaheen, (2012) "Adaptive Layered Approach using Machine Learning Techniques with Gain Ratio for Intrusion Detection Systems". IJCA, Volume 56 – No.7, October 2012.
10. Devikrishna K.S., and Ramakrishna B.B. (2013) "An Artificial Neural Network based Intrusion Detection System and Classification of Attacks" Vol. 3, Issue 4, Jul-Aug 2013.
11. Eitel J.M. Lauria and Giri K. Tayi, (2008) "A Comparative Study Of Data Mining Algorithms For Network Intrusion Detection In The Presence Of Poor Quality Data". ICIQ-03, 2008.
12. E.Kesavulu Reddy, Member IAENG, V.Naveen Reddy, P.Govinda Rajulu, (2011) "A study of Intrusion Detection in Data Mining". WCE 2011, July 6 -8, 2011.
13. Hui Zhao (2013) "Intrusion Detection Ensemble Algorithm based on Bagging and Neighbourhood Rough Set". IJSIA, Vol. 7, No. 6 (2013).
14. Ian H. Witten, Eibe Frank and Mark A. Hall (2012), "Data Mining Practical Machine Learning Tools and Techniques"
15. Mitchell D'silva, Deepali Vora (2013) "Comparative Study of Data Mining Techniques to Enhance Intrusion Detection". IJERA, Vol. 3, Issue 1, January – February 2013.
16. Mohammad Mahmudul Alam Mia, Shovasis Kumar Biswas, Monalisa Chowdhury Urmi, and Abubakar Siddique (2015), "An algorithm for Training Multilayer Perceptron (MLP) for Image Reconstruction using Neural Network without Overfitting", IJSTR, Vol. 4, Issue 2, February 2015.
17. Neha Maharaj and Pooja Khanna (2014), "A Comparative Analysis of Different Classification Techniques for Intrusion Detection System", IJCA, Vol. 95 – No.17, June 2014.
18. Paavo Nieminen (2010), "Classification and Multilayer Perceptron Neural Networks", 2010.
19. Reema Patel, Amit Thakkar, Amit Ganatra (2012) " A Survey and Comparative Analysis of Data Mining Techniques for Network Intrusion Detection Systems".IJSCE, Volume-2, Issue-1, March 2012.
20. Trilok Chand Sharma and Manoj Jain (2013), "WEKA Approach for Comparative Study of Classification Algorithm", IJARCCE, Vol. 2, Issue 4, April 2013.
21. V. Jaiganesh, Dr. P. Sumathi, and A.Vinitha "Classification Algorithms in Intrusion Detection System: A Survey". IJCTA, Vol 4(5), September – October, 2013.
22. V. Jaiganesh, S. Mangayarkarasi, and Dr. P. Sumathi (2013) "Intrusion Detection Systems: A Survey and Analysis of Classification Techniques". IJARCCE, Vol. 2, Issue 4, April 2013.