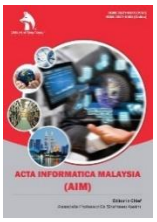
ZIBELINE INTERNATIONAL™ PUBLISHING

ISSN: 2521-0874 (Print)
ISSN: 2521-0505 (Online)
CODEN: AIMCCO

Acta Informatica Malaysia (AIM)

DOI: http://doi.org/10.26480/aim.02.2022.55.61



RESEARCH ARTICLE

ARTIFICIAL NEURAL NETWORK ANALYSIS OF SOME SELECTED KDD CUP 99 DATASET FOR INTRUSION DETECTION

Samuel Olorunfemi Adams[a]*, Ednah Azikwe[b], Mohammed Anono Zubair[a]

[a]Department of Statistics, University of Abuja, Abuja, Nigeria.
[b]Department of Computer Science, Captain Elechi Amadi Polytechnic Port-Harcourt, River State, Nigeria
**Corresponding Author Email: samuel.adams@uniabuja.edu.ng*

ARTICLE DETAILS

Article History:

Received 07 July 2022
Accepted 09 August 2022
Available online 15 August 2022

ABSTRACT

Due to the growing number of intrusions in local networks and the internet, it has become so universal that institution increasingly implements many structures that investigate information technology security violations. This study aimed to process, classify and predict the intrusion detection accuracy of some selected network attacks using the artificial neural network (ANN) technique. Five important attacks, namely; Buffer overflow, Denial of Service (DoS), User to Root Attack (U2R), Remote to Local Attack (R2L) and PROBE were chosen from the KDD CUPP'99 information and intrusion identification accuracy was investigated with artificial neural network (ANN) modeling technique. Findings from the classification show that out of the procedures utilized to establish the ANN model, 27262 of the 45528 buffer overflow are classified appropriately, 7903 of the 45528 DoS attacks are arranged appropriately, 1371 of the 45528 U2R are classified appropriately, 431 of the 45528 R2L are arranged appropriately and, 8304 of the 45528 PROBE are classified appropriately. Comprehensively, about 99.1% of the training proceedings are arranged properly, equivalent to 0.9% erroneous classification while the testing specimen assisted to confirm the model with 99.1% of the attacks were appropriately arranged by the ANN equation. This support that, comprehensively, the ANN equation is precise about the classification and prediction of the five attacks investigated in this study.

KEYWORDS

Network Attacks; Intrusion detection system; Multilayer perception; Neural Network; Neuron; predictors.

1. INTRODUCTION

With the wide adoption of artificial intelligence (AI) systems, a construction project in both of engineering field, as well as management field, is facing the challenge of rising digital transformation (Pan and Zhang, 2021). By taking into account that artificial intelligence (AI) systems can be a significant solution for many issues in construction projects. Moreover, the topic of artificial intelligence (AI) system has currently become the research domain to focus on, it enquires to be more clear and comprehensively established (Allal-Chérif et al., 2021). The population of human beings has been continued to increase and economic development in massive ways propelled energy and material consumption to a greater degree than threatens the very existence of our Earth in near future (Wang and Srinivasan, 2017).

In data security, intrusion disclosure refers to the demonstration of determining activities which endeavor to adjust the classification, accessibility and uprightness of an asset. At the point when Intrusion detection takes a precautionary dimension without direct human mediation, then, it can be referred to as an intrusion detection framework. Intrusion detection can be performed physically or naturally. Physical intrusion discovery could occur by analyzing log documents or other proof for signs of intrusion, in addition to network traffic. A framework that performs computerized intrusion detection is referred to as an Intrusion Detection System (IDS). An IDS can be host-based, assuming it monitors framework logs or calls, or network-based on the off chance that it screens the progression of network packets (McHugh J., 2000). An Artificial Neural Network (ANN) involves the compilation of strongly interconnected processing components and transforming a bunch of inputs into a bunch of wanted outputs. The thought behind the use of soft computing methods and especially ANNs in utilizing IDSs is to involve a clever specialist for the system that is fit for uncovering the latent examples in strange and typical connection review records and to generalize the examples to new (and somewhat unique) connection reports similar to a class. In the current study, a disconnected intrusion detection framework is executed by utilizing Multi-Layer Perceptron (MLP) artificial neural network (see Figure 1). While in numerous past studies the executed framework is a neural network with the ability to identify normal or intrusion connections. This element empowers the system to recommend legitimate actions against potential attacks. The promising outcome of this study indicates the possible relevance of ANNs for creating reasonable IDSs.

Some of the past research that used utilized artificial neural networks for KDD CUP'99 data set efficiency, detection value, and wrong classification were determined to discover the viability of the ANN equation that was observed to produce great outcomes for obtrusion determination. Current research showed that intrusion determination with deep literacy is an artificial neural network (ANN) calculation which is a development of normal machine literacy that is the instruction derivation and literacy are disconnected function. In contrast to the broadly utilized current intrusion determination technique which creates a standard as well as an equation that vindictive intrusion designs, it discovered connections

Quick Response Code	Access this article online	
	Website: www.actainformaticamalaysia.com	**DOI:** 10.26480/aim.02.2022.55.61

Cite The Article: Samuel Olorunfemi Adams, Ednah Azikwe, Mohammed Anono Zubair (2022). Artificial Neural Network Analysis of Some Selected KDD CUP 99 Dataset for Intrusion Detection. *Acta Informatica Malaysia, 6(2): 55-61.*

straightforwardly from secured information to detect strange dangers. It has concentrated on a arrangement recognized intrusion utilizing combined machine literacy through consolidating K-means using security vulnerability management (Tahir et al., 2015). The suggested a network security library and KDD with the disadvantages of the KDD Cup-99 information (Tavallaee et al., 2009). The research likewise illuminated the likelihood dispersion of network congestion reports with their conduct in the test and training group information. The study uncovered some particular intrusion class are excluded from the initial dataset though contained in the second group. Thus, utilization of KDD Cup'99 produce an assignment of attack determination more useful. Shin, et al., 2016 utilized a K-means design that is always involved with non-stratified combination to search for the comparison in the information and this manner recognized a specification ready to determine a DoS invasion and an infiltrated invasion simultaneously. The proposed a ready relationship technique to diminish the number of alerts and give a brief undeniable level perspective on network attacks (Kavousi and Akbaris, 2012). The intrusion exercises designed advanced in this way can help with connecting alerts, reproducing attack situations, and forecasting conceivable future attacks in a real-time framework. Chung and Kim developed and tried the invasion determination equation by utilizing a single or mixture of several machine literacy designs, for example, the decision tree, support vector machine (SVM), and Bayesian classification. Has a explored an artificial intelligence (AI) invasion determination framework by utilizing a deep neural network applying the KDD Cup 99 data in light of consistently developing network attacks (Kim et al., 2017). The DNN design was implemented on the information enhanced using initialization to make a literacy model, with the whole KDD Cup 99 information was utilized as a confirmation. Gao et al., (2014) concentrated on intrusion detection by utilizing deep belief networks (DBNs), applying the KDD Cup 99 information indicated that precision is improved by more than 6 percent of the current ANN and SVM equation. That looked at the support vector machine and forward additive neural network (Jo et al., 2016). The forward additive neural network is a calculation which compensates for the shortcomings of the current back-proliferation calculation. In this review, an interruption identification study was analyzed utilizing FANN and the outcome was contrasted and the current SVM model. It was seen that FANN exhibited higher exactness and a preferable identification rate over the SVM.

This study aimed to process, classify and predict the intrusion detection accuracy of some selected network attacks using the artificial neural network (ANN) technique. The remaining section of the study is arranged as follows: materials and methods are given in section two. The results are given in part three. Section 4 concludes the study with the conclusion and recommendation.

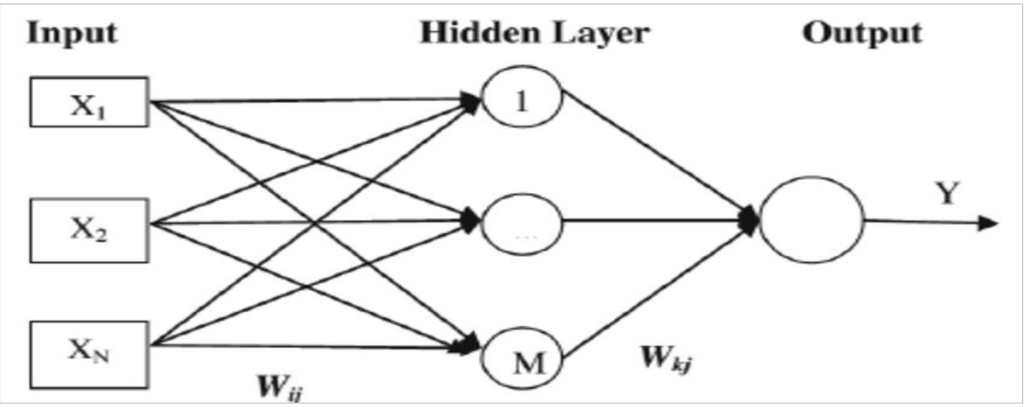


Figure 1: Architecture of Multilayer Feed-Forward Neural Network Used for Predictions

2. Materials and Methods

This subdivision provides the data as well as the method used in the study. The artificial neural network (ANN) model was applied to intrusion detection using both mathematical analysis and a graphical model-building approach. The methodology that will be adapted in the process of the analysis, design, and development of the system is discussed.

2.1 KDD CUP'99 Data set

The data used in this study is from the Knowledge Discovery and Data Mining Tools Competition (KDD Cup'99). It is a famous set of data, publically accessible for study purposes. It is openly obtained from the University of California-Irvine Machine Learning depository on URL http://archives.ics.uuci.edu/ml/datasets/KDD+Cup+1999+Data. It has been involved in earlier review design for network intrusion detection systems (NIDS) along with intrusion prevention systems (IPS) in network security. In this review, an objective examination of the KDD Cup'99 set of data has improved comprehension of primary and hierarchical perspectives on the elements/traits of network traffic information. The comprehensive KDDCup'99 dataset has been ordered into three essential parts (Lippmann, et al., 2000).

i. 10 percent of KDD CUPP-99 numbered training information: This excerpt of data is regarded as train information which involves 97278 regular reports from the entire 494021 reports.

ii. KDD Cup-99 test information: This proportion of the information is regarded as a data for test that is additionally altered with small repetitive traffic information parcels which is referred to as amended KDD having 60593 ordinary reports amidst a total of 311029 reports.

iii. The entire KDD: This contains the entire set of KDD Cup'99 data having 972780 typical reports from 4898431 reports.

This set of date includes five million reports with named training information, marked test information, and unmarked entire information. Each of the traffic information has forty-one highlights, and these highlights are sorted into three classes, namely; fundamental, content, and time-based traffic features (Stolfo, 2000; Kayacik et al., 2003; Tavallaee et al., 2009). The general element of the data in KDD Cup'99 has been ordered into discrete and continuous categories. The network traffic information that is distinguished as a 'bad' association and powerless against the security of the system is referred to as an attack, and this kind of obstruction information is not regarded as a regular class (Karthikeyan and Indra, 2010; Kayacik et al., 2005; Lee and Stolfo, 2000). KDD Cup'99 invasion consist of four classifications, namely; U2L, DoS, probe, and R2L. The entire obstruction reports of invasion classifications are grouped in at least one out of the four accompanying classifications (Tavallaee, et al., 2009; Mukkamala et al 2002).

This study researched five significant network attacks; Denial of Service, Buffer overflow, User to Root Attack (U2R), PROBE and Remote Local Attack (R2L). According to Kalavadekar and Sane, (2017), Lakshmi and Babu, (2015), Mukkamala, Janoski, Sung (2002), Sharma, Jain, and Sharma, (2016) a wide range of attacks and their corresponding frequencies with rates in the training cases of KDD Cup'99 consist of sum of training with 24 types of intrusions which can be expanded to fourteen kinds of information test. An important observation is that, the training group is not derived from a similar probability distribution as the test category. In the same vein, it is observed that the test group usually add explicit invasion classes which are not in the information of the training category. A few assaults are just present in the test group yet not present in the training group as well as the other way around. Moreover, the HTTP burrow attack is depicted as a U2R class of attack in (Horng *et al*. 2011), yet (Lippmann *et al*.2000) added it for the R2L class. This study also added an HTTP tunnel to the R2L cases of invasion. The KDD-99 information depends on the association of traffic information kept in DARPA'98. The study from portrayed that, the fundamental hindrance that characterized DARPA-98 information synthetic (McHugh, 2000).

2.2 Neural Network Models

Fundamentally an ANN is a numerical model of interaction developed exactly as opposed to utilizing mass and energy adjusts around the

process. Similar to neurons in the brain, a neural network comprises a network of to some degree associated handling components or nodes, organized in layers. They further incorporate interconnections between the nodes of progressive layers. A schematic setup of the essential design of a simple neuron or a node inside a neural network model is delineated in Figure 2, with inputs, an enactment function, and a solitary output.

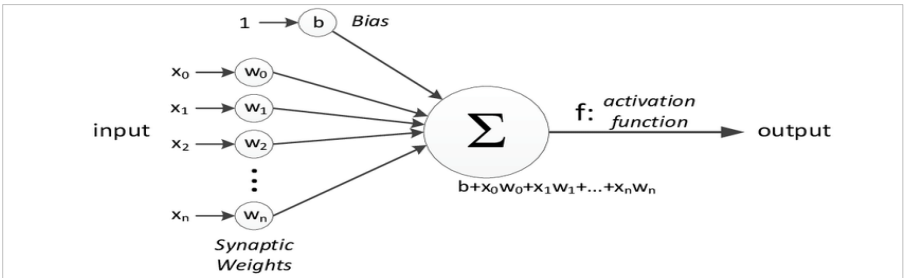


Figure 2: A Mathematical Model of a Neuron

The associations between nodes are determined by qualities called weights. The weight is the ''strength'' of association between neurons, and Yi is the result. Every neuron in the hidden layer gets weighted inputs in addition to predisposition from every neuron in the past layer, as given by;

$$Z_i = \left(\sum_{k=1}^{N_j - 1} X_k^{j-1} W_{k,i} - b_k \right) \quad (1)$$

An artificial neural network (ANN) can be depicted as a bunch of interconnected units developing in time and working equally; the units address axons and dendrites and each connection (j, i) from unit j to unit i has a weight μ_{ij} that balances impact of unit j on unit i. In this manner, an ANN is a weight-directed graph in which to each node i have related a predisposition or limit s_i and a transfer function f_i, so that unit i will create an output y_i of the form:

$$y_i = f_i \left(\sum \mu_{ij} x_j - s_i \right) \quad (2)$$

Where x_j is the jth contribution of this unit and μ_{ij}, x_j is the amount of all its weighted data sources. On the off chance that this aggregate is more noteworthy than the edge s_I, unit i is enacted for delivering the output y_i; in any case unit i is in a dormant state. The parameters μ_{ij} and s_I can be adjusted so that the neural network delivers some ideal way of behaving. The neural network can be prepared to accomplish some specific occupation by adjusting the weight and bias parameters. The exchange functions widely utilized are nonlinear, smooth, expanding, and limited like sigmoid functions (alleged from their "*S*" shape). In any case, at times the exchange work is an identity function. When $fi(x) = 1 \; if \; x > 0 \; and \; fi(x) = 0$, in any case, unit i is known as an edge entryway. As limit functions are irregular, they are much of the time replaced by sigmoidal transfer functions that are consistent and differentiable, such as $f(x) = arctan\,(x)$ and $f(x) = tanh\,(x)$, or by other exchange functions like $f(x) = 1 / (1 + e - x)$. One disadvantage of this neuron model was seen when it was utilized to portray what electrochemical triggering phenomena take place at the active cell membranes of biological neurons. It was noticed that the description of signal transformations in complicated neural networks needs an analysis computationally too heavy. Based on this observation, suggested the following simple nonlinear dynamic model a neuron observation Kohonen (1998).

In Figure 3, y_p and j_p are nonnegative scalar factors, the input activation I_i is some capacity of the y_p and of a few interior boundaries. The function $w(j_p)$ is the spillage term, a nonlinear function of result movement. To ensure great security in criticism networks g should be convex (for example its second derivative concerning j_p must be positive). The spillage term $w(j_p)$ considers every single different loss and dead-time impact in the neuron, as a dynamic function of movement.

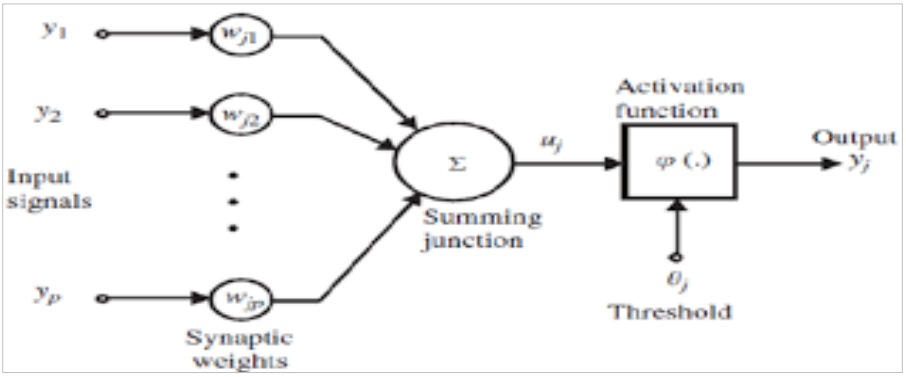


Figure 3: A Nonlinear Dynamic Model of a Neuron

3. RESULT

The process summary cases presented in Table 1 show that 45695 cases was allotted to the training group and 19711 added to testing group. The 129 cases rejected amidst the investigation are the forthcoming network intrusion. Table 2 shows the network information table regarding the artificial neural network which is valuable as long as guaranteeing the determinations are right. Note specifically that; the quantity of nodes in the entry layer is the amount of covariates in addition to the complete amount of element degrees; a different nodes is made from every classification of intrusion label as well as the classifications are generally not thought of "repetitive" units as is run of the mill in several modeling algorithms. Moreover, a different output node is made for every classification of attacks, a total of five nodes in the outer layer. The mechanical architecture determination selected nine nodes in the latent layer. Any remaining system data refers to the defect of the method, the model synopsis in Table 3 showcases details regarding the aftereffects of training and assigning the last structure to the test. The result shows that; Cross entropy error of 42.822 is shown in the light of the fact that the output layer utilizes the softmax actuation function. The assessment algorithm halted because the greatest number of epochs was achieved. Preferably, training ought to stop because the error has merged. This brings up issues about whether something went wrong during training and is something to remember while further investigating the output.

The classification report displayed in Table 4 indicated the pragmatic consequences of utilizing the structure. In every group, the forecasted variable is; buffer overflow, DoS, U2R, R2L, and PROBE if the cases forecast pseudo-probability is higher than 0.5 (see Figure 4). For every example, units within the slanting of the cross-classification of groups are the right prediction. Cells off the askew of the cross-classification of cases are wrong forecasts. Of the cases utilized to make the model, 27262 of the 45528 buffer overflow are arranged accurately, 7903 of the 45528 DoS attacks are classified accurately, 1371 of the 45528 U2R are characterized accurately, 431 of the 45528 R2L are classified accurately and, 8304 of the 45528 PROBE are arranged accurately. Generally speaking, 99.1% of the

training cases are ordered accurately, relating to the 0.9% wrong classification displayed in the model synopsis table. This demonstrates that the model has accurately identified the cases. The classification has given the cases utilized to make the model will generally be as well "hopeful" as their order rate is inflated. The testing group assists with approving the model; here 99.1% of these cases were accurately ordered by the model. This affirmed the fact that the model is right about the five attacks.

Table 1: Case Processing Summary

		N	Percent
Sample	Training	45695	69.9%
	Testing	19711	30.1%
Valid		65406	100.0%
Excluded		129	
Total		65535	

Table 2: Network Information

Input Layer	Factors	1	Label
	Covariates	1	http
		2	Finger
		3	Smtp
		4	telnet
		5	erc_i
		6	domain_u
		7	ftp
		8	7_data
	Number of Units[a]		13
	Rescaling Method for Covariates		Standardized
Hidden Layer(s)	Number of Hidden Layers		1
	Number of Units in Hidden Layer 1[a]		9
	Activation Function		Hyperbolic tangent
Output Layer	Dependent Variables	1	Attacks2
	Number of Units		5
	Activation Function		Softmax
	Error Function		Cross-entropy

a. Excluding the bias unit

Table 3: Model Summary

Training	Cross-Entropy Error	42.822
	Percent Incorrect Predictions	0.0%
	Stopping Rule Used	Training Error Ratio Criterion (.001) Achieved
	Training Time	0:00:03.28
Testing	Percent Incorrect Predictions	0.0%
	Dependent Variable: Attacks	

Table 4: Classification

Sample	Observed	Predicted					
		Buf. overflow	DoS	U2R	R2L	Probe	% Correct
Training	Buf. overflow	27262	0	273	12	19	98.9%
	DoS	7	7903	0	0	1	99.9%
	U2R	90	0	1371	0	0	93.8%
	R2L	21	0	0	431	0	95.4%
	Probe	1	0	0	0	8304	100.0%
	Overall %	**59.9%**	**17.3%**	**3.6%**	**1.0%**	**18.2%**	**99.1%**
Testing	Buf. overflow	11733	0	117	5	14	98.9%
	DoS	8	3339	0	0	0	99.8%
	U2R	31	0	609	0	0	95.2%
	R2L	6	0	0	200	0	97.1%
	Probe	2	0	0	0	3647	99.9%
	Overall %	**59.8%**	**16.9%**	**3.7%**	**1.0%**	**18.6%**	**99.1%**

Dependent Variable: Attack

Figure 5 displayed the use of the Multilayer perception (MLP) invasion determinate for neural structure. This issue refers to an element of the estimations that limit the error in foreseeing intrusion. This design is the pre-compensation architecture that is associations in the structure stream forward within the entry layer to the exist layer with no criticism circles. In Figure 5 the information layer contains the explanatory variable while the latent layer consist of undetectable nodes. The worth of each latent unit is several capacity within the predictors; the specific type of some function relies on a limited extent of the network class and some degree next to client configurable specification. The exit layer involves the reactions and historical backdrop from intrusion detection with a grouped variable having five (5) classes, it is reported as five response variables with each output nodes in most model within the hidden nodes. The Multilayer perception structure permits an extra latent layer; all things considered, every unit of the second hidden layer is a function of the units in the primary hidden layer, and every response is a model of the nodes in the extra latent layer. The ROC curve in Figure 6 provides a detailed presentation of the specificity as well as the sensitivity for every potential breakoff within a simple graph, which is a lot neater and very impressive above the tables of progression. The chart displayed in the study indicates five (5) attacks and curves. Since there are five classes, the non-linear plots are proportional around the 45-degree (not shown) within the higher left corner of the outline to the lesse right. The lift plot (Figure 7) derived from the cumulative gain diagram; qualities on the y pivot relate to the proportion of the accumulated profits for every nonlinear plot to the gauge. Consequently, for instance the lift at 10% for the R2L is 10%/10% = 1.0. It gives one more perspective on data in the cumulative gains diagram.

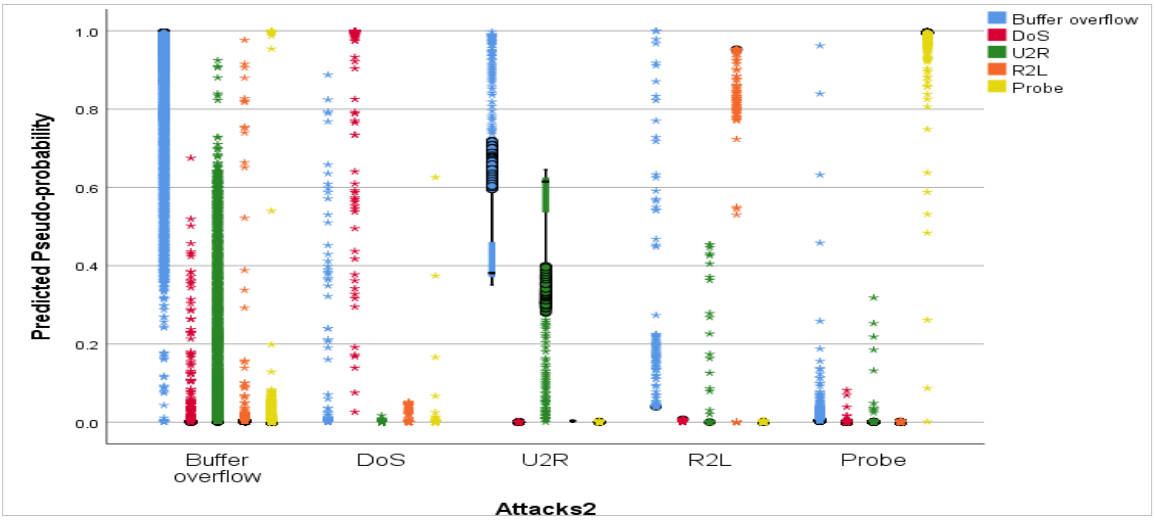


Figure 4: Predicted Pseudo-Probability of the Five Selected KDD CUP'99 dataset

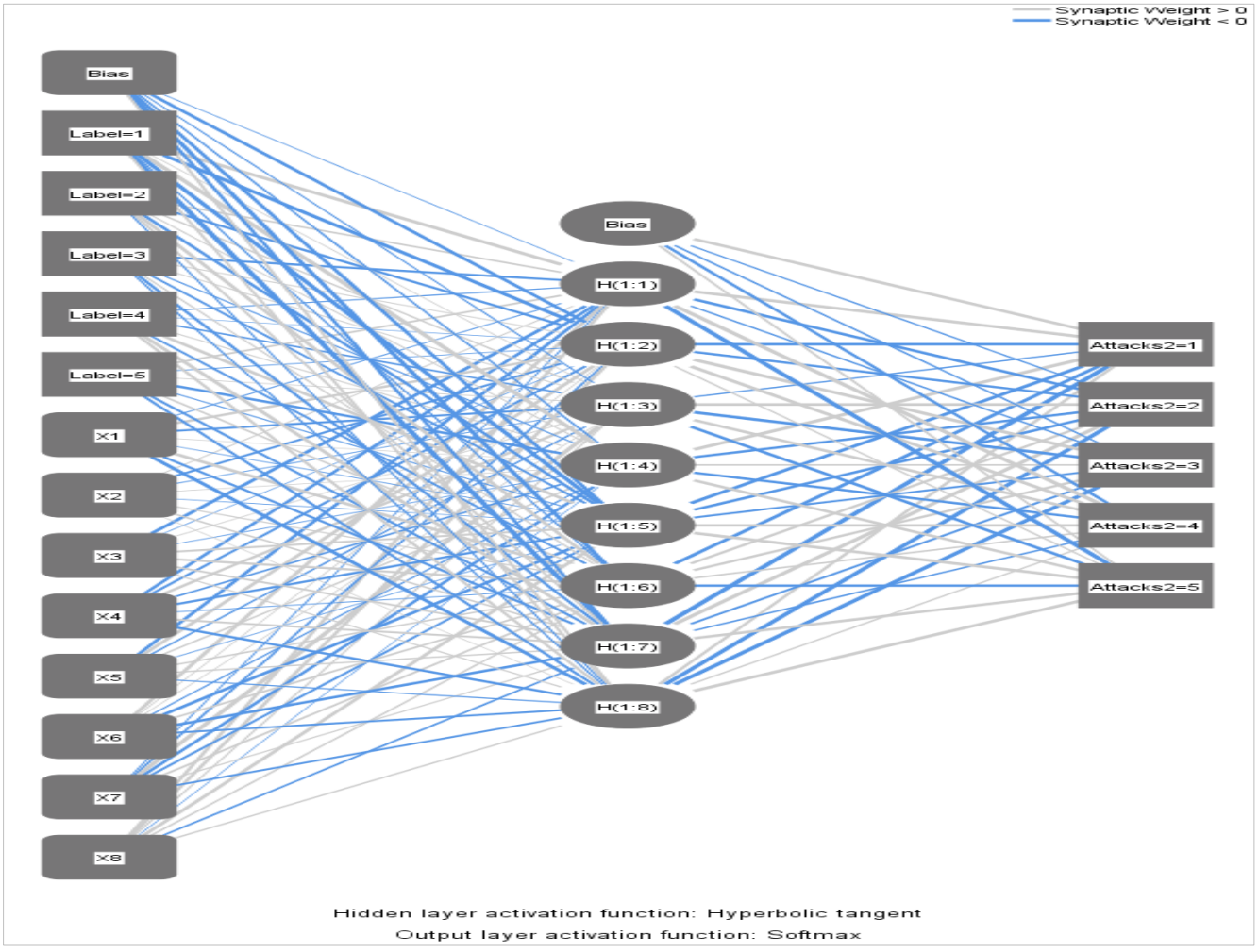


Figure 5: Neural Network Layer of the five Selected Set of KDD CUP99 Data

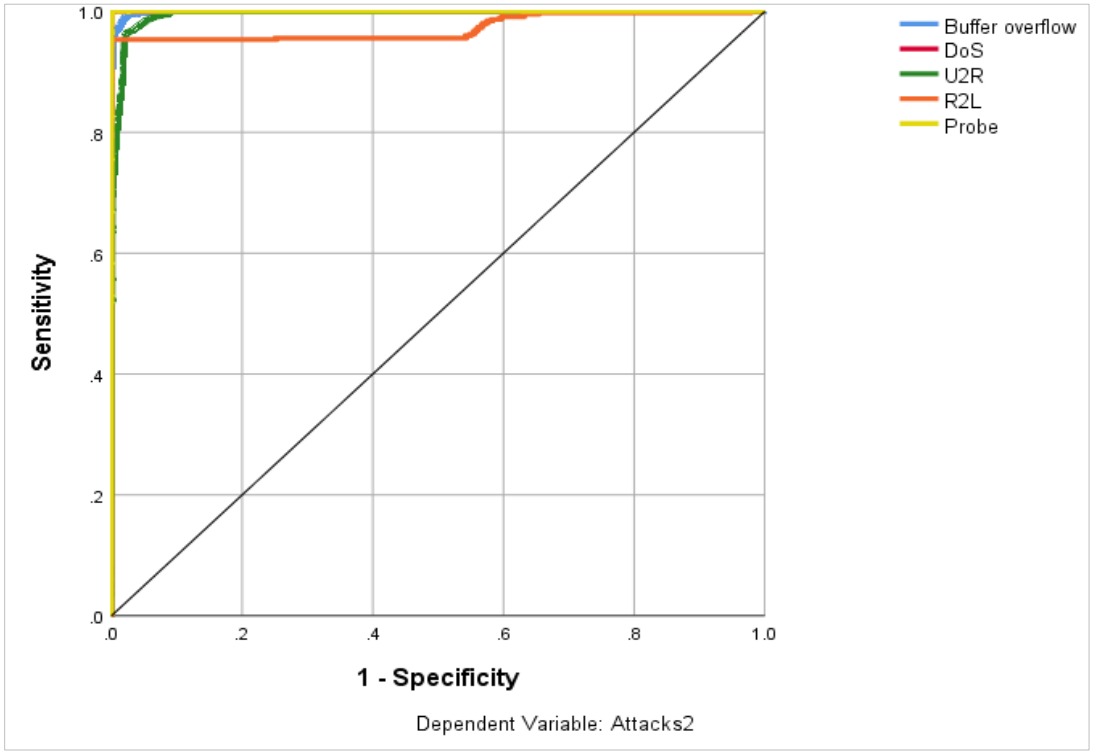


Figure 6: ROC Curve of the Five Selected Set of KDD CUP99 Data

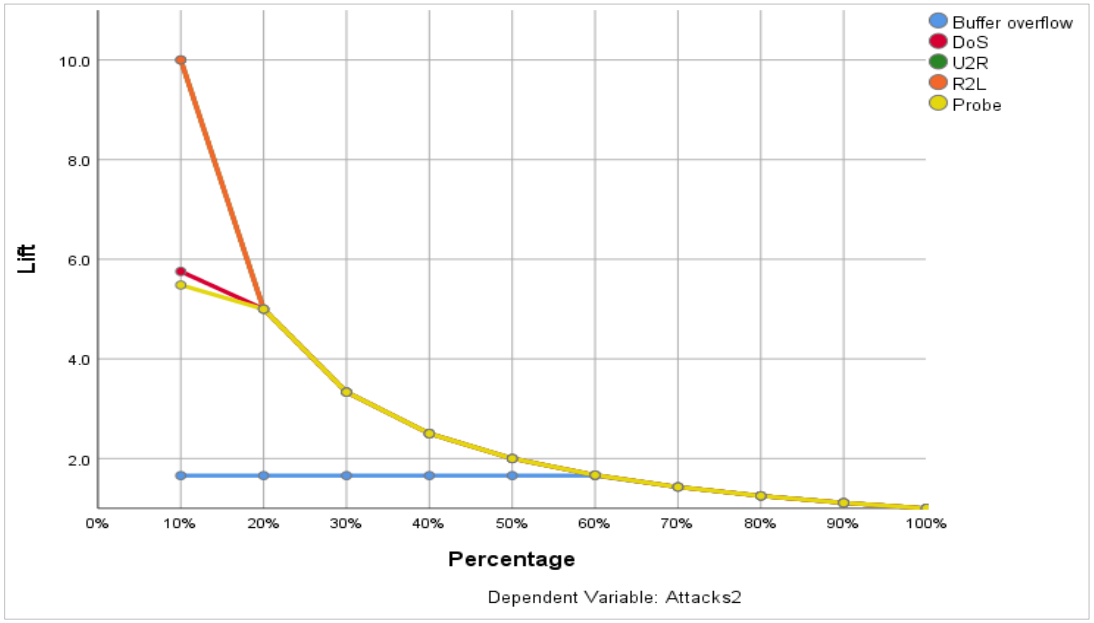


Figure 7: Lift Chart of the five Selected KDD CUP'99 Dataset

4. Conclusion

This paper presents the information processing method, accuracy of attack classification, and prediction of five important attacks, namely: Denial of Service (DoS), Buffer overflow, User to Root Attack (U2R), PROBE and Remote to Local Attack (R2L), extracted from the analyzes of KDD Cup'99 dataset. The case processing demonstrated that 45695 samples were relegated to the first group referred to as training, 19711 assigned to the group called testing while 129 cases were rejected from the imminent network intrusion. The network information indicated that a different output unit is made for every class of attacks, with a total of five arrays insider the outer part. The automatic architecture determination actually picked nine units in the concealed layer while the remaining network material refers to the default for the methodology. The model summary showed data regarding the report of training and administering the last network to the test group. The outcome demonstrated that a cross-entropy error of 42.822 is shown with a rate erroneous percentage of 0.0% for training classes with a training error proportion measure of .001. The little percentage of erroneous predictions demonstrates the effectiveness of the ANN model in intrusion detection.

The classification indicated that; out of about 65535 cases utilized to develop the model, 27262 out of the 45528 buffer overflow are grouped accurately, 7903 of the 45528 DoS attacks are classified accurately, 1371 of the 45528 U2R are classified accurately, 431 of the 45528 R2L are characterized accurately and, 8304 of the 45528 PROBE are classified accurately. In general, 99.1% of the cases in the training group are classified accurately, compared to the 0.9% erroneous classification. This demonstrates that the model has accurately recognized the attacks. The classifications of the attacks utilized to make the model will generally be extremely hopeful because their classification rate is expanded. The testing group assists with approving the equation where about 99.1% of the invasions were accurately classified by the ANN equation. In all, the model is accurate in the five attacks classification. The artificial neural network has presented an appropriate model to detecting intrusion with higher precision for Buffer overflow, User to Root Attack (U2R),Denial of Service (DoS), PROBE and Remote to Local Attack (R2L) invasions.

References

Chung, S. K. and Kim, K., 2015. A Heuristic Approach to Enhance the

Performance of Intrusion Detection System using Machine Learning Algorithms. Proceedings of the Korea Institutes of Information Security and Cryptology Conference (CISC-W'15).

Gao, N., Gao, L., Gao, Q., and Wang, H., 2014. An Intrusion Detection Model Based on Deep Belief Networks. Advanced Cloud and Big Data (CBD), 2014 Second International Conference. Pp. 247-252.

Horng S-J, SuM-Y,ChenY-H,KaoT-W, Chen RJ,Lai J-L, PerkasaCD., 2011. A Novel Intrusion Detection System based on Hierarchical Clustering and Support Vector Machines.Expert systems Application, 38(1): Pp. 306–313.

Jo, S., Sung, H. and Ahn, B., 2016. A Comparative Study on the Performance of SVM and an Artificial Neural Network in Intrusion Detection. Journal of the Korea Academia- Industrial cooperation Society, 17(2). 703-711, Pp. 2016.

Kalavadekar, M.P.N. and Sane, S.S., 2017. Effective Intrusion Detection Systems using Genetic Algorithm. International Journal of Engineering Trends, 1(2):Pp. 8316–831.

Karthikeyan, K.R. and Indra, A., 2010. Intrusion Detection Tools and Techniques-A Survey.International Journal of Computer Theory and Engineering, 2(6):Pp. 901–906.

Kavousi F., Akbari B., 2012. Automatic Learning of Attack Behavior Patterns using Bayesian Networks. 6th International symposium on Telecommunications (IST), Tehran, Iran: IEEE; Pp. 999–100.

Kayacik, H.G., Zincir-Heywood, A.N. and Heywood, M.I., 2003. On the Capability of an SOM based Intrusion Detection System. In: Proceedings of the International Joint Conference on Neural Networks, vol 3. Portland, OR, USA, IEEE, Pp. 1808–1813.

Kayacik, H.G., Zincir-Heywood, A.N. and Heywood M.I., 2005. Selecting Features for Intrusion Detection: A Feature Relevance Analysis on KDD 99 Intrusion Detection Datasets. Proceedings of the 3rd Annual Conference on Privacy, Security and Trust, Pp. 85-89

Kim, J., Shin, N., Jo, S.Y. and Kim, S.H., 2017. Method of Intrusion Detection using Deep Neural Network. 2017 IEEE International Conference on Big Data and Smart Computing (BigComp.). https://doi.org/10.1109/BIGCOMP.2017.7881684

Kohonen, T., 1998. The Self-Organizing Map. Neurocomputing, 21, Pp. 1-6. http://dx.doi.org/10.1016/S0925-2312(98)00030-7

Lakshmi, T.V.N. and Babu V.K., 2015. Detection of user to Root Attacks using Machine Learning Techniques. International Journal of Advance Engineering Global Technology, 3(3): Pp.418–424.

Lee W, Stolfo SJ., 2000. A Framework for Constructing Features and Models for Intrusion Detection Systems. ACM Transactions on Information System Security, 3(4): Pp. 227–261

Lippmann R., Haines, J.W., Fried, D.J., Korba, J., and Das, K., 2000. The 1999 DARPA off- Line Intrusion Detection Evaluation. Computer Network, 34(4): Pp. 579–595.

McHugh J., 2000. Testing Intrusion Detection Systems: A Critique of the 1998 and 1999 DARPA Intrusion Detection System Evaluations as Performed by Lincoln laboratory. ACM Transactions on Information System Security. (TISSEC) 3(4): Pp. 262–29.

Mukkamala S, Janoski G, Sung A., 2002. Intrusion Detection using Neural Networks and Support Vector Machines. Proceedings of the 2002 International Joint Conference on Neural Networks, IJCNN'02, Honolulu, HI, USA, IEEE, 2: Pp. 1702–1707

Sharma, C., Jain, S.C., Sharma, A.K., 2016. Explorative Study of SQL Injection Attacks and Mechanisms to Secure Web Application Database-A. International Journal of Advance Computer Science Application, 7(3):Pp.79–87.

Shin, D., Choi, K., Chune, S. and Choi, H., 2016. Malicious Traffic Detection Using K-means, The Journal of Korean Institute of Communications and Information Sciences. 41(2);Pp. 277-284.

Stolfo, S. J., Fan, W., Lee, W., Prodromidis, A. and Chan, P.K., 2000. Cost-based Modeling and Evaluation for Data Mining with Application to Fraud and Intrusion Detection. Proceedings DARPA Information Survivability Conference and Exposition. DISEX'00, 2000: Pp.130-144, 2. https://doi.org/10.1109/DISCEX.2000.821515.

Tavallaee M, Ebrahim E, Lu W, Ghorbani AA., 2009. A Detailed Analysis of the KDD CUP 99 data set. IEEE Symposium on Computational Intelligence in Security and Defense Applications, (CISDA 2009). Ottawa, ON, Canada: IEEE; Pp. 1–6. https://doi.org/10.1109/CISDA.2009.5356528

Tahir, M., Hassan, W., Md Said, A., Zakaria, N., Katuk, N., Kabir, N., Omar, M., Ghazali, O. and Yahya, N., 2015. Hybrid Machine Learning Technique for Intrusion Detection System. 5th International Conference on Computing and Informatics (ICOCI).