IJARCCE

ISSN (Online) 2278-1021
ISSN (Print) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Vol. 9, Issue 6, June 2020

Performance Evaluation of Error Rate in Immune Inspired Concepts with Neural Network for Intrusion Detection in Cybersecurity

Ukam James Joseph[1] and Adeniji Oluwashola David[2]

Department of Computer Science, University of Ibadan, Ibadan, Nigeria[1,2]

Abstract: Intrusions that normally occur in computing systems are often meshed towards accessing, changing or damaging sensitive data or information. It is against this background, that various research has been carried out with the aim of solving detection and preventing such intrusive attacks.The similarity between the problem of computer security that is faced by Immune System (IS) can be shown by translating the language of immunology into computer security terms, Also the IS detects abuses of an implicitly specified policy, and responds to those abuses by counter-attacking the source of the abuse. However, Artificial Immune System (AIS) define the way the Human Immune System (HIS) responds to threats or attacks in the body. AIS and HIS are combined together by researchers to solve intrusion problems in Cybersecurity. The Negative Selection Algorithm (NSA) is an algorithm that divide the problem space into self and non-self which was used to build the model. In this study a model based on AIS concepts that will find a significant application in cybersecurity was developed and evaluated. The developed model called NNET NSA (Neural Network Negative Selection Algorithm) used the NSLKDDCup1999 dataset to test the model. The results from the developed model shows that the model NNET NSA achieved Receiver Operating Characteristics (ROC) showing 90% Area under the Curve (AUC) proportion of accuracy in detection of cyber-crime. The Error rate evaluation of NNET NSA classification of cyber-crime detection was the less by 0.05%, naïve Bayes by 0.16% and SVM by 0.22%. respectively on the R console.

Keywords: Artificial Immune System, Cybersecurity, Intrusion detection, Error Rate**.**

I. INTRODUCTION

Cyber attacks are usually focused at accessing, changing, or destroying sensitive information; extorting money from users; or interrupting normal business processes. Although, potential cyber threats can vary from logic bomb, to virus, social engineering, out of service attacks, DOS, DDOS, seizing web pages and critical infrastructure systems, Trojan horse, malware, illegal control over computing systems, capturing and changing confidential data.The problem that the Human Immune System addresses is similar to the problem faced by computer security systems: the immune system protects the body from pathogens, and analogously, a computer security system should protect computers from intrusions of any form. This analogy can be made more concrete by understanding the base concepts faced by computer security systems. Integrity is when data is protected from corruption, whether malicious or accidental. It is essential to preserve the integrity of critical systems, like information used by emergency services. Confidentiality being a way of allowing access to restricted or confidential data only to authorized users, for example, it is imperative for military institutions to limit knowledge of classified information.

II. RELATED WORKS

The danger theory proposes that the cells of the innate immune system can actively suppress an immune response in the absence of danger and in the presence of molecular signals produced when cells die normally [1] . This is because, it was suggested, that the activating “danger signals” do not come from external sources but produced by the cells of the body when a cell dies unexpectedly [2] [3]. Thus, the central idea in the Danger Theory is that the immune system does not respond to non-self but to danger and just like the self, non-self-theories, it fundamentally supports the need for discrimination. However, it differs in the answer to what should be responded to. Instead of responding to foreignness, the immune system reacts to danger [4] [5]. In an interview by Lauren Constable [6], Polly Matzinger confirmed that the danger model makes the prediction that the immune system will respond to molecules that enter the body and do damage, causing the damaged tissues to release immune-stimulating alarm signals. A side consequence of this she stated, will be that, if we give the body a molecule at the time that something else has caused damage, the immune system will associate that new molecule with the (un associated) damage, and respond to it. Many novel algorithms have been developed using the idea of the Danger Theory as a backbone. The Dendritic Cell Algorithm is one of such. DCA was developed as part

IJARCCE
ISSN (Online) 2278-1021
ISSN (Print) 2319-5940
International Journal of Advanced Research in Computer and Communication Engineering
Vol. 9, Issue 6, June 2020

of the Danger project.The project based on a model of the function of dermal dendritic cells of the human body and their ability to discriminate between healthy and infected tissue as shown in figure 1.

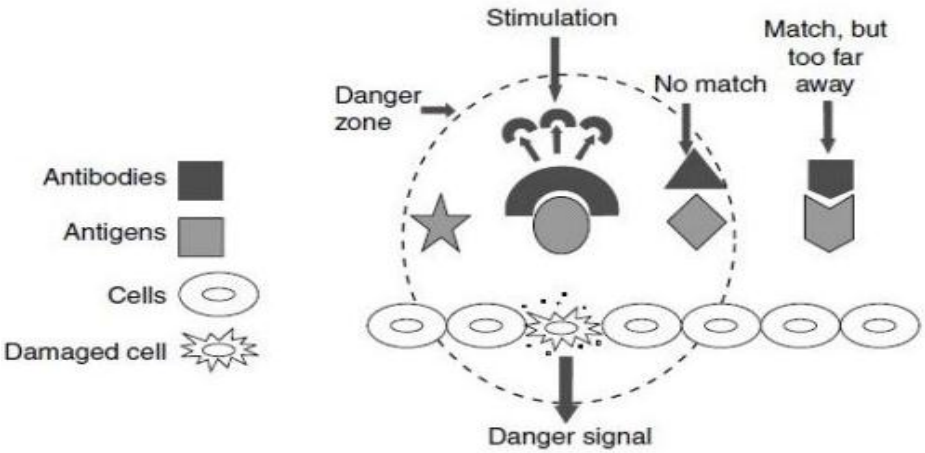

Fig. 1: Danger Theory Illustration Cayzer, 2002.

DCA has found application in both port scan and insider attack detection, botnet zombie machine detection, standard machine learning intrusion datasets, robotic security, schedule overrun detection in embedded systems, sensor networks, and other real-time, dynamic problems [7]. A recent improvement over DCA called dDCA (Deterministic Dendritic Cell Algorithm) was put forward by [8]. From an expounded study, Artificial Immune Systems seem to be most suitable for computer security problems. Although [9] [10] listed various major and minor application areas of AISs viz Clustering/classification, Bio-informatics Anomaly detection, Image processing, Computer security, Control, Numeric function optimization, Robotics, Combinatorial optimization, Virus detection, Learning, Web mining, fraudulent transactions and hardware faults to mention the few.

III. INTRUSION DETECTION

Attacks in MANET which aim at slowing or stopping the flow of information between communicating nodes. Attacks on WSN which prevents the sensors from detecting and transmitting information through the network. There is a need for scalable and energy-efficient routing, data gathering and aggregation protocols in these WSN environments in [11]. In detecting intrusions in host computers or network links, strategies must be employed. A novel algorithm was designed that employed techniques from AIS with ANN and applied to intrusion detection.[12]. Host Based Intrusion Detection Systems (HBIDS) is a strategies that reside on and monitor an individual host machine. analyze system activity, looks for events or sets of events that match a predefined pattern of events that describes a known attack. The basic idea is to use the knowledge of known attack patterns and apply such knowledge to identify attacks in various sources of data being monitored. However, SDN implements network protocols that take years of testing, standardization and interoperability in [13]. The tradeoff between the two protocols can provide a significant impact on the networks. Furthermore, one potential choice of selecting any of the protocols can increase or decrease the degree of application in used as adapted in [14]. Therefore, this strategy attempts to detect only known attacks based on predefined attack characteristics. Signature based approach is strategy that uses the semantic characteristics of an attack. An experiment was performed using deep learning approach. . Honeynet hardware was setup to collect zero-day attack. Bidirectional recurrent neural network algorithm was used for the analysis of the data set at different level of granularity in [15] Zero Day attack Prediction. The sematic are analyzed and details used to form attack signatures. These attack signatures are formed in such a way that they can be searched using information in audit data logs produced by computing systems. Anomaly Detection Strategy identifies abnormal or unusual patterns on a host or network after defining the normal user behaviour. However, the resource management of Multihoming in nested mobile network raises new issues in the host mobility of ipv6 network.in [16]. Because attacks are different from legitimate user' activities, these differences can be easily spotted by these systems. Further information on an intelligentsias-scammer filter mechanism using bayesian techniques [17] was reviewed. The internet service driven network is a new approach to the provision of network computing that concentrates on the services you want to providee as adopted in [18]. The significant roles of encryption algorithms are numerous and essential in information security.[19] **in** Comparative Study of Symmetric Cryptography Mechanism .Every cyber-attack has some form of financial implication on the organization, (Paul et al., 2011) has calculated risk as the product of threat occurrences (expected events/year) and their resultant losses in dollars/event. They thus estimated annual risk to be

$R = \sum_i E_i * L_i$ (C) ……..Equation (1)

Where

R= risk in dollars per year

IJARCCE
ISSN (Online) 2278-1021
ISSN (Print) 2319-5940
International Journal of Advanced Research in Computer and Communication Engineering
Vol. 9, Issue 6, June 2020

i = index representing the different threats facing the firm
E_i = the expected number of security events of type i per year
Li(C) = expected dollar loss caused by security event i given the current set of countermeasures C. They then referred to the term $L_i(C)$ as the single-event loss expectancy and noted three major types of single-event loss expectancies; brand damage which represents damage to a company's image, regulatory fines and production losses due to disruption of IT resources which support production.

IV. METHODOLOGY

The negative selection module generates set of detectors based on the self-strings. The detector set is represented using prefix directed acyclic graphs. The Developed Model was coded in the R programming language with the: Input module, Network Decoder module (optional module as seen from the methodology), Negative selection module and Classification module.R-contiguous matching technique is a direct adaptation of the technique followed by the T-cells. According to this scheme, the detector string matches with any given string, if and only if the detector matches with the given string in at least r continuous positions.

Development of NNET detection generation: The matching characters has to occur at the same indices at the detector and training strings. An r-chunk matching scheme was used when comparing detectors with the test traffic as a part of the methodology. Since the r-chunk detectors are represented as strings along with indices, prefix tree is a data structure that is more suited to represent them. Its introduction has proven to achieve linear running times for classification. In this research, the r-chunk detectors are represented as prefix trees and prefix directed acyclic graphs, b oth of these concepts are explained in the following section.

Algorithm 1 : Construction of Prefix Tree
A tree can be called a prefix tree T if it satisfies the following conditions:
1. It has exactly one root node (a node with no inbound edges) and can have one or more leaf nodes (nodes with no outbound edges)
2. All the edges are labelled with characters of the alphabet Σ
3. Each node cannot have more than one edge that is labelled with the same element aϵΣ
4. For a string s, one could say sϵT if there exists a path in T from the root to the leaf node with labels the same as the characters of s.
5. Language L(t) is a set consisting of strings with one or more characters appended to the end of elements of the set T. In other words, each of the strings in L(T) has a prefix string s'ϵT. A prefix string s' of a string s is made of the first n-r characters of s, where r<n, n being the length of the string.
For the below example, s='bb'ϵT because there is a path from the root node to the leaf with these characters as labels. Similarly, s='ab'ϵT and s='aa'$\notin$T
Input: S **C S** (self-set), M C Σ^l (monitor set)
Output: Set of Prefix trees
Begin
For all I ϵ { 1,…, l – r + 1 } **do**
Generate an empty prefix tree T_i
End for
For all T_i **do**
For all s ϵ S **do**
Insert s[I,…I + r-1] into T_i
End for
For all non-leaf nodes n in T_i **do**
Create a new leaf n' labelled with every aϵ Σ, provided that no other outbound edge from n is labelled with a.
End for
For all non-leaf nodes n in T_i **do**
Delete n if there is no path from n to any of the newly generated nodes
End for
End for

Fig 2:Algorithm 1 : construction of prefix tree

Algorithm 2 : C onstruction of Prefix DAG
A prefix directed acyclic graph (prefix DAG), denoted as D is similar to a prefix tree except that it is an acyclic graph as opposed to a tree. The following conditions are to be satisfied for it to be called a prefix DAG:

IJARCCE
ISSN (Online) 2278-1021
ISSN (Print) 2319-5940
International Journal of Advanced Research in Computer and Communication Engineering
Vol. 9, Issue 6, June 2020

1. The prefix DAG can have more than one root and leaf nodes
2. All the edges are labelled with characters of the alphabet Σ
3. Each node cannot have more than one edge that is labelled with the same element a ∈ Σ
4. A string s is said to belong to D, if and only if there is a path from a root to a leaf node labeled the same edges as the string. Note that as opposed to the prefix tree, the prefix DAG can have more than one root and leaf nodes.
5. Language L(d,n) is a set consisting of strings with one more characters appended to the end of elements of the set D for a given root node n. In other words, each of the strings in L(T) has a prefix string s'∈T. A prefix string s' of a string s of length n is made of the first n-r characters of s, where r<n.

Input: set of Prefix Trees
Output: Prefix DAG, D
Begin
For i=1 to l-r+1 **do**
For all nodes 'n' in the tree T_i **do**
Choose a ϵ Σ with no other outgoing node with the same label for n
Let 's' be the string formed from root to n. if the same string exists in $T_i + 1$ for a node n', form an edge (called failure link) from node n of T_i to the node n' of $T_i + 1$.
Note that this path will have labels that correspond to s'[2… | s' |]
End for
End for

Fig 2: Algorithm 2 : construction of prefix DAG

Algorithm 3 : NNET Detection Generation

The NNET detection generation was developed as follows [21] [22]:

Detect_set_gen (S, x, r_s)
S: self-sample
X: number of detectors
r_s: self-radius
1. Let D be from an empty space
2. repeat
3. t be random samples from the space $[0,1]^n$
4. Repeat for all s_i C S …….{i=1,2,3…}
5. d← r-chunk
6. for every l-r+1
7. do
8. while d ≤ r_s, repeat step 2
9. D←D U t
10. until D=m
11. return D

Fig 3: Algorithm 3 : NNET detection generation

The equation below was used to define the cost when evaluating the performance normalization of 0 and 1.

$$PCF(+) = \frac{p(+)c(-|+)}{P(+)c(-|+) + p(-)c(+|-)} \quad \text{……(Equation1)}$$

Where:
c(-|+) = cost of misclassifying a positive profile as negative
c(+|-) = cost of misclassifying a negative profile as positive
p(+) = probability of a positive profile and
p(-) = 1 – p(+).

4.0: Result And Discussion

The result that was gathered during the experiment was adapted in RStudio development platform. The NSLKDDCup dataset for artificial immune system for intrusion detection was loaded. Then execute the command by typing the file name "NLC KDD+_20Percent.arff"" on the R console. The "NLC KDD+_20Percent.arff"" initiates the execution of the negative selection algorithm by loading the pre-processed dataset as shown below in fig 4.0.

IJARCCE

ISSN (Online) 2278-1021
ISSN (Print) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Vol. 9, Issue 6, June 2020

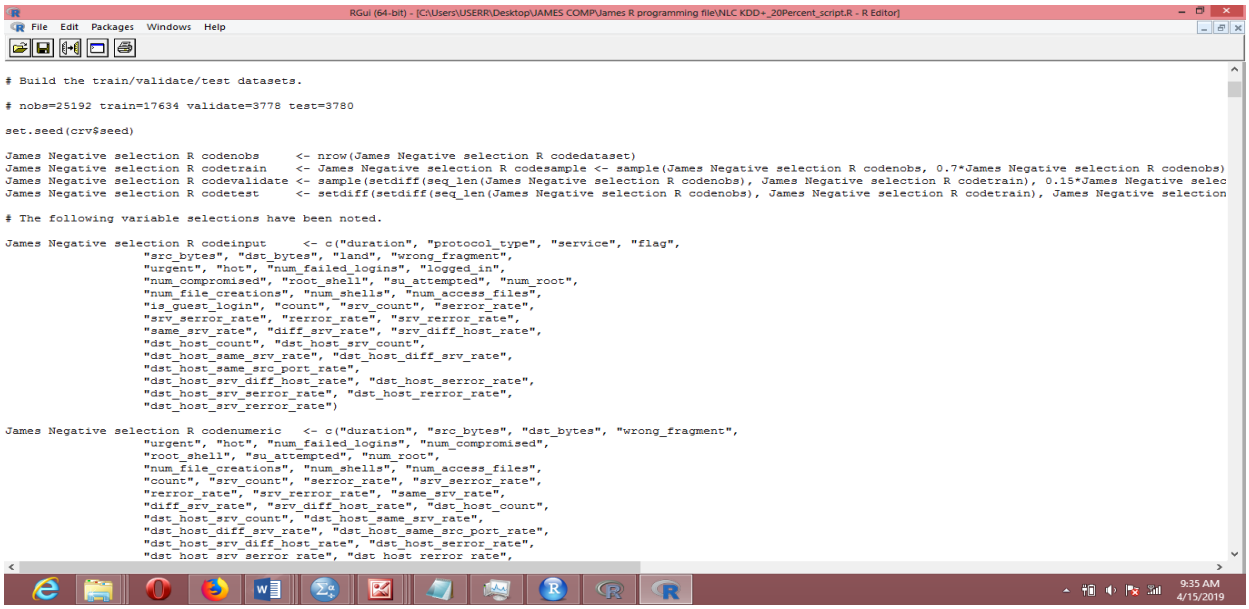

Figure 4.0: NNET NSA generating attributes and features

There are altogether seven major nodes displayed here and corresponding twenty seven minor nodes with respect to the attributes of the dataset used. Each of this node has an equivalent weight assigned to them. This result helps us further understand where intrusions will likely take place as shown in fig 4.1.

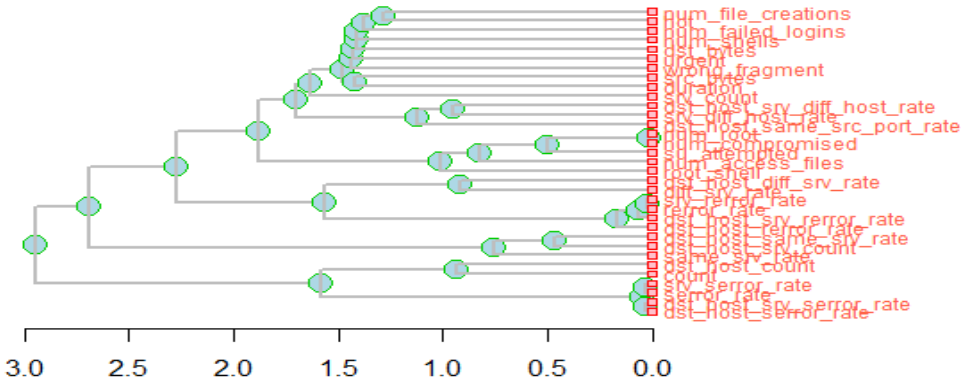

Figure 4.1: When the r-chunk has matched the antigens to create antibodies.

The Cost curves illustrating both false positive and false negative rates is shown in fig 4.2 below.

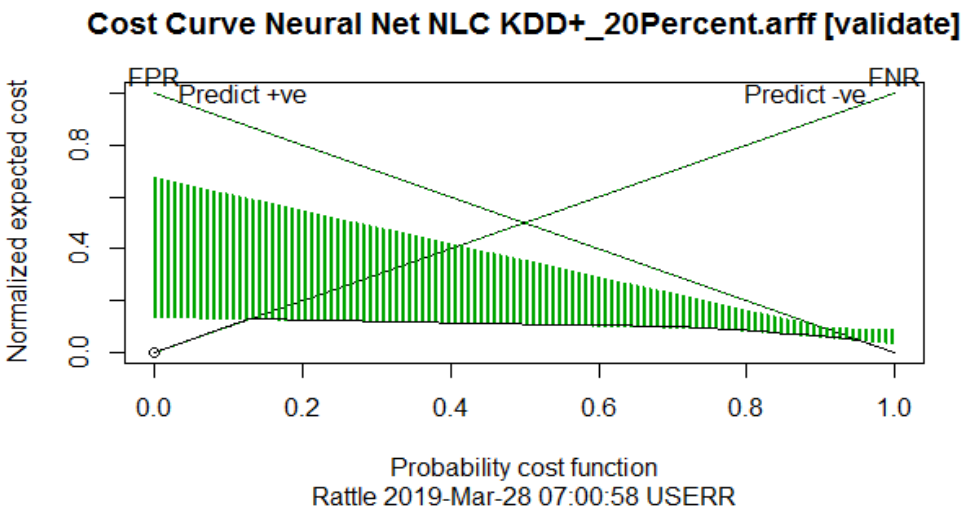

Figure 4.2: Cost curves illustrating both false positive and false negative rates

IJARCCE
ISSN (Online) 2278-1021
ISSN (Print) 2319-5940
International Journal of Advanced Research in Computer and Communication Engineering
Vol. 9, Issue 6, June 2020

Table 4.1 below reveals that NNET NSA correctly classified 85.7% of the cyber-crime as attack, followed by SVM by 77.5%, while Naïve Bayes by 70.4%. On the other hand, NNET NSA Error rate classification of cyber-crime detection was the less by 0.05%, naïve Bayes by 0.16% and SVM by 0,22%. By implication, our model appeared superior in the classification of cybercrime detection status (normal or anomalous).

Table 4.1: Summary of comparison of weighted average and
Error Rate of different classifiers NNET NSA, Naïve Bayes and SVM

Algorithm	**TP Rate**	**FP Rate**	**Precision**	**Recall**	**Error Rate**
Naive Bayes	0.704	0.042	0.512	0.592	0.16
SVM	0.775	0.048	0.420	0.650	0.22
NNET NSA	**0.857**	**0.039**	**0.857**	**1.00**	**0.05**

The performance evaluation of the classification accuracy of **NNET NSA 90%,** Naive Baye 81% and SVM was 65% as shown below in fig 4.3.

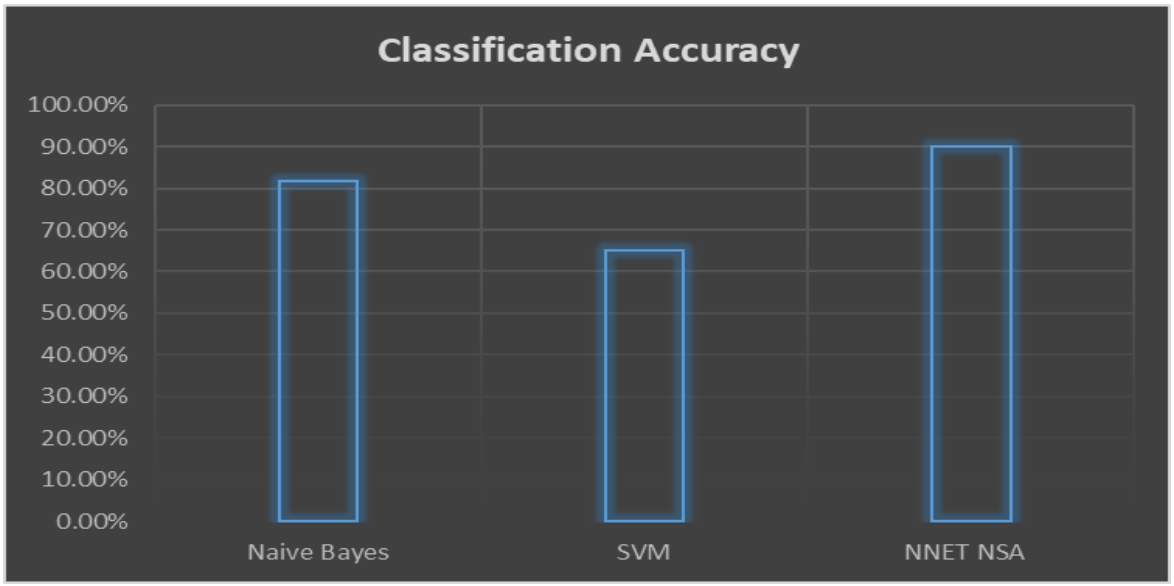

Figure 4.3: Bar-chart showing accuracy difference in different classifiers

V. CONCLUSION

The detection generation phase was computationally cost effective. This shows an overall optimal performance of the model. Furthermore, individual matching threshold of detectors was used for an optimal maximization of the detector coverage area, statistical estimations like the Pearson coefficient were incorporated in the clustering / matching phase to capture the distance between clusters, average error rates were calculated, strongly linked to the way a BIS works. The AG-AB binding process (as defined in our matching function) was scalable and seamless and the similarity patterns were easily detected and classified by the use of the NN capability.

ACKNOWLEDGMENT

The authors wish to thank the Department of Computer Science, University of Ibadan, Nigeria for the support in this research work.

REFERENCES

[1]. Dutt, I., Borah, S., & Maitra, I. "Intrusion Detection System using Artificial Immune System". International Journal of Computer Applications, 44(12), 19–22. (2016).
[2].Julie Greensmith, Amanda Whitbrook, U. A. Artificial Immune Systems. nternational Series in Operations Research & Management Science, 146, 421–448, Springer Dordrecht(2010)..
[3]Prakash, A., & Deshmukh, S. G. "A multi-criteria customer allocation problem in supply chain environment : An artificial immune system with fuzzy logic controller based approach". Expert Systems With Applications, 38(4), 3199–3208..(2011)
[4]. Aickelin, U, Bentley, P., Cayzer, S.,Kim, J., & Mcleod, J. "Danger Theory : The Link between AIS and IDS ", 147–155. (2003).
[5].Cayzer, S. . The Danger Theory and Its Application to Artificial Immune Systems, 141–148.(2002)

[6] Matzinger, P. (2012). The evolution of the danger theory, 8(4), 311–317.
[7]. Greensmith, Julie; Aickelin, U. "The Deterministic Dendritic Dendritic Cell Algorithm". ICARIS, 1–12.(2010).
[8] Balachandran, S., Dasgupta, D., Nino, F., & Garrett, D. " A General Framework for Evolving Multi-Shaped Detectors in Negative Selection. IEEE Explore, 1–15. https://doi.org/10.1109/FOCI.2007.371503(2007)
[9]. Ye, G., Wang, Y., & Sun, Q.).Super Base Station Fault Detection Mechanism Based on Negative Selection Algorithm and Expert Knowledge Base. IOP Conference Series:Materials Science and Engineering, 490(07), 1-6 IOP publishing.(2019)
[10]. Aickelin, Uwe, Dasgupta, D., & Gu, F.. Search Methodologies. Search Methodologies. https://doi.org/10.1007/978-1-4614-6940-7 (2013)
[11]. Ojoawo A.O & Adeniji O.D " Energy Efficient Hierarchical Cluster Head Election Using Exponential Decay Function Prediction." International Journal of Wireless & Mobile Networks (IJWMN) Vol. 10, No. 5, (2018)
[12]. Adeniji O.D. & Ukam, J.J " Immune Inspired Concepts Using Neural Network for Intrusion Detection in Cybersecurity" Proceedings of the 20th iSTEAMS Multidisciplinary Trans-Atlantic GoingGlobal Conference KEAN University, New Jersey, USA Pp 119-126. (2019)
[13] A. A. Olabisi O. D. Adeniji, Abeng Enangha " A Comparative Analysis of Latency, Jitter and Bandwidth of IPv6 Packets Using Flow Labels in Open Flow Switch in Software Defined Network" Afr. J. MIS, Vol.1, Issue 3, pp. 30-36.(2019)
[14] Adeniji Oluwashola David, Adenike Osofisan "Route Optimization in MIPv6 Experimental Test bed for Network Mobility: Tradeoff Analysis and Evaluation." International Journal of Computer Science and Information Security (IJCSIS), Vol. 18, No. 5, pp 19-28.(2020) .
[15]. Adeniji O.d., Olatunji O.O "Zero Day Attack Prediction with Parameter Setting Using Bi Direction Recurrent Neural Network in Cyber Security".International Journal of Computer Science and Information Security (IJCSIS), Vol. 18, No. 3,pp 111-118 (2020).
[16] S.D Adeniji, S Khatun, RSA Raja, MA Borhan' "Design and analysis of resource management support software for multihoming in vehicle of IPv6 Network". Proceedings of the Fifth IASTED International Conference.Vol 607,issue 089.pp 13. (2008).
[17] Adeniji Olushola D , Olubukola Adigun, Omowumi O Adeyemo " An intelligent spam-scammer filter mechanism using bayesian techniques"International Journal of Computer Science and Information Security (IJCSIS), Vol. 10, No. 3 pp 126 (2012)
[18] S.D Adeniji, S Khatun, MA Borhan, RSA Raja, " A design proposer on policy framework in IPV6 network" IEEE International Symposium on Information Technology.Vol 4,pp 1-6,(2008)
[19]. Logunleko K.B., Adeniji. O.D., Logunleko A.M," A Comparative Study of Symmetric Cryptography Mechanism on DES, AES and EB64 for Information Security". International Journal of Scientific Research in Computer Science and Engineering Vol.8, Issue.1, pp.45-51. (2020).
[20]. Paul, L., Deane, J. K., Rakes, T. R., & Baker, W. H. ". Decision support for Cybersecurity risk planning. Decision Support Systems", 51(3), 493–505. https://doi.org/10.1016/j.dss.2011.
[21]. Revathi, S., & Malathi, A."A Detailed Analysis on NSL-KDD Dataset Using Various Machine Learning Techniques for Intrusion Detection. International Journal of Engineering Research and Technology (IJERT), 2(12), 1848–1853. (2013).
[22]. Cui, L., Pi, D., & Chen, C. "BIORV-NSA : Bidirectional inhibition optimization r-variable negative selection algorithm and its application". Applied Soft Computing Journal, 32,544–552. (2015).