Article

JOURNAL OF UNIVERSITY OF BABYLON
For Pure and Applied Sciences (JUBPAS)

Vol.32; No.2.| 2024

Network Traffic Classification Against Adversarial Attacks Based Deep Learning

Haneen Mohammed Hussein[1*], Zaid Haitham Qasim[2] and Ammar Ali Mustafa[3]

[1]Division of Construction and Projects, Mustansiriyah University, haneen.mh@uomustansiriyah.edu.iq, Baghdad, Iraq.

[2]Division of Construction and Projects, Mustansiriyah University, zaid_alkhazaali@uomustansiriyah.edu.iq, Baghdad, Iraq.

[3]Division of Construction and Projects, Mustansiriyah University, ammar.ali@uomustansiriyah.edu.iq, Baghdad, Iraq.

***Corresponding author email:** haneen.mh@uomustansiriyah.edu.iq; mobile: +9647711021298

تصنيف حركة مرور الشبكة ضد الهجمات العدائية بناءً على التعلم العميق

حنين محمد حسين[1*]، زيد هيثم قاسم [2]، عمار علي مصطفى [3]

قسم الاعمار والمشاريع ، الجامعة المستنصرية ، haneen.mh@uomustansiriyah.edu.iq ,بغداد,العراق

قسم الاعمار والمشاريع ، الجامعة المستنصرية ، zaid_alkhazaali@uomustansiriyah.edu.iq , بغداد,العراق

قسم الاعمار والمشاريع ، الجامعة المستنصرية ، ammar.ali@uomustansiriyah.edu.iq ,بغداد,العراق

Accepted: 29/4/2024 **Published:** 30/6/2024

ABSTRACT

Background:

Cybersecurity is a prominent concern in today's interconnected world, encompassing both local and remote wireless and wired access across diverse communication technology platforms. It is important to recognize the threat posed by hackers who are currently compromising organizational functionalities, bypassing security measures, and stealing hypersensitive information. Common hacking techniques such as Portscan, Distributed Denial of Service (DDoS) attacks, and the use of Botnets, are frequently utilized by hackers.

Materials and Methods:

The authors explore the advantages of using Machine Learning (ML) to classify attacks such as Port Scanning, DDoS, Botnet, and Botnet-Attempt in a mixture of both benign and attack traffic. They use various Artificial Neural Networks (ANN) structures, each having distinct properties, to train and test on a benchmark dataset (CICIDS2017). The aim is to identify the most effective ANN model and the optimal number of input features required to classify data that contains events of Portscan, DDoS, Botnet, and Botnet-Attempt attacks.

Results:

Various features are used as inputs for an ML model with single and multiple hidden layers, each with different neurons, to evaluate their impact on classification accuracy using a Python language. The best accuracy obtained is 99.71%, achieved by using all features of the dataset and 4 hidden layers, while the accuracy obtained using only 7 features is 97.6%.

Conclusion:

ANN models can perform well in classifying network traffic against adversarial attacks by using an optimal combination of features as input and hidden layers.

Keywords: Network Traffic; Cybersecurity; Machine Learning; Deep Learning; Intrusion Detection.

ISSN: 2312-8135 | Print ISSN: 1992-0652

info@journalofbabylon.com | jub@itnet.uobabylon.edu.iq | www.journalofbabylon.com

مجلة جامعة بابل للعلوم الصرفة والتطبيقية مجلة جامعة بابل للعلوم الصرفة والتطبيقية

ISSN: 2312-8135 | Print ISSN: 1992-0652
info@journalofbabylon.com | jub@itnet.uobabylon.edu.iq | www.journalofbabylon.com

INTRODUCTION

Internet technology has rapidly spread and entered a phase of accelerated expansion. Personal life has become easier due to the use of the internet, but this has also introduced hazards such as cyber-attacks [1-3]. Many networks have specific security vulnerabilities. Malicious actors can exploit these vulnerabilities using malware to launch attacks. Additionally, these attacks can manipulate network information and cause significant harm. Such attacks affect not only productivity but also have far-reaching impacts on future business, finance, and the overall perception of the brand [4, 5].

Artificial Intelligence, ML, and Deep Learning (DL) are longstanding concepts that have been implemented or considered for implementation numerous times over the past few decades. The aim is to enable machines to perform tasks that humans can do without explicit instructions. ML employs various algorithms to tackle data-related challenges, and it's important to acknowledge that no single algorithm is universally the best for all problems. The selection of an algorithm depends on factors such as the specific problem, the number of variables involved, and the most suitable model for the task [6, 7].

Data plays a crucial role in ML, originating from a wide array of sources including social networks, logs, blogs, and various sensors connected in a network such as temperature, current, and humidity sensors. The network is vulnerable to several different types of attacks, the most common ones include denial-of-service attacks, botnet attacks, and forged information injection. ML systems predicted these kinds of attacks with an extremely high degree of accuracy [8].

Supervised learning is used according to the data it deals with. There are various ML techniques available, each suited to the type of data that needs processing in ML. Supervised learning includes classification algorithms that use a dataset with predefined classes for each data point, enabling the computer to learn how to classify future data [6, 9].

ANN significantly alleviate the burden on humanity and society in solving complex problems with high efficiency. These networks mimic brain functions, utilizing acquired training samples for a wide range of applications, including classification, regression, prediction, smart grid management, natural language processing, image processing, medical diagnosis, and more [10, 11]. ANNs are a powerful tool in ML and have significantly contributed to the advancement of AI technologies. They continue to evolve and find new applications across diverse fields [12, 13].

[14] provided a comprehensive comparison of various ML algorithms on the CICIDS2017 dataset. The authors train and test different ML algorithms on the dataset to identify the best performing algorithms for classifying vectors of Portscan and DDoS attacks. The classification results show that all variants of discriminant analysis and Support Vector Machine (SVM) provide good testing accuracy, with more than 90% accuracy. The paper mentions that tree-based models, K-Nearest Neighbors KNN, and most ensemble classifier-based algorithms exhibit inefficient performance in the range of 49-69. The subspace discriminant variant of

ensemble classifier also showed competitive testing accuracy of 85.5%. 21 features were selected as the most significant features for classification based on the configured value of the correlation coefficient.

In the Ref. [15], ML technique is utilized, specifically the SVM algorithm, to detect cyber-attacks in networks. Other algorithms such as Random Forest, Convolutional Neural Network (CNN), and ANN are also mentioned in the paper. The classification accuracy of the models is evaluated to assess their performance and CICIDS2017 dataset is used. The paper shows that These algorithms can achieve accuracies of 93.29 for SVM, 63.52 for CNN, 99.93 for Random Forest, and 99.11 for ANN.

An approach to network traffic classification has been presented which achieves an impressive accuracy rate of 98%. This approach combines CNNs and recurrent neural networks (RNNs) and is capable of managing extensive and intricate datasets, making it suitable for real-world network traffic classification tasks. The method begins by using CNNs to extract features from the raw network traffic data, after which RNNs are employed to classify the extracted features [16].

The authors emphasize the need for reliable datasets in intrusion detection and prevention systems [17]. They demonstrate that datasets DARPA98, KDD99, ISC2012, and ADFA13, are outdated, unreliable, and lack necessary features. They aim to create a dependable dataset containing benign and seven common attack network flows that meet real-world criteria. The dataset comprises more than 80 network flow features, categorized based on the day captured, with both normal and attack traffic. The paper evaluates the performance of ML algorithms on the dataset to identify the most effective features for detecting specific attack categories. The name of dataset is CIC-IDS2017 [17].

This paper investigated the potential features in the CICIDS2017 dataset that can be useful for detecting port scanning DDoS, Botnet and other attacks by using ANN with different numbers of hidden layers and neurons within it. The rest of the paper is organized as follows. In materials and methods section, methodology, algorithms, models, and equations are discussed, then the results of the models' accuracy and selected features are discussed in results and discussion section. Conclusions of the authors are presented in the end of paper.

ISSN: 2312-8135 | Print ISSN: 1992-0652
info@journalofbabylon.com | jub@itnet.uobabylon.edu.iq | www.journalofbabylon.com

ISSN: 2312-8135 | Print ISSN: 1992-0652
info@journalofbabylon.com | jub@itnet.uobabylon.edu.iq | www.journalofbabylon.com

MATERIALS AND METHODS

The author's methodology is clarified in the following steps:

1) CICIDS2017 dataset on Friday data is used. The data contains 81 features or attributes which consist the instance or record that will be used as input to the ANN model, the total records are 547557 each includes one of five cases which are port scan, DDoS, Botnet and Botnet – Attempted attacks and Benign traffic. These classes are labeled from 0 to 4 and they are as follows
   - 288544 records are labeled Normal (Benign).
   - 736 records are labeled Botnet.
   - 4067 records are labeled Botnet – Attempted
   - 95144 records are labeled DDoS.
   - 159066 records belong to Portscan attacks

   Figure (1) shows the histogram of Friday data of dataset.

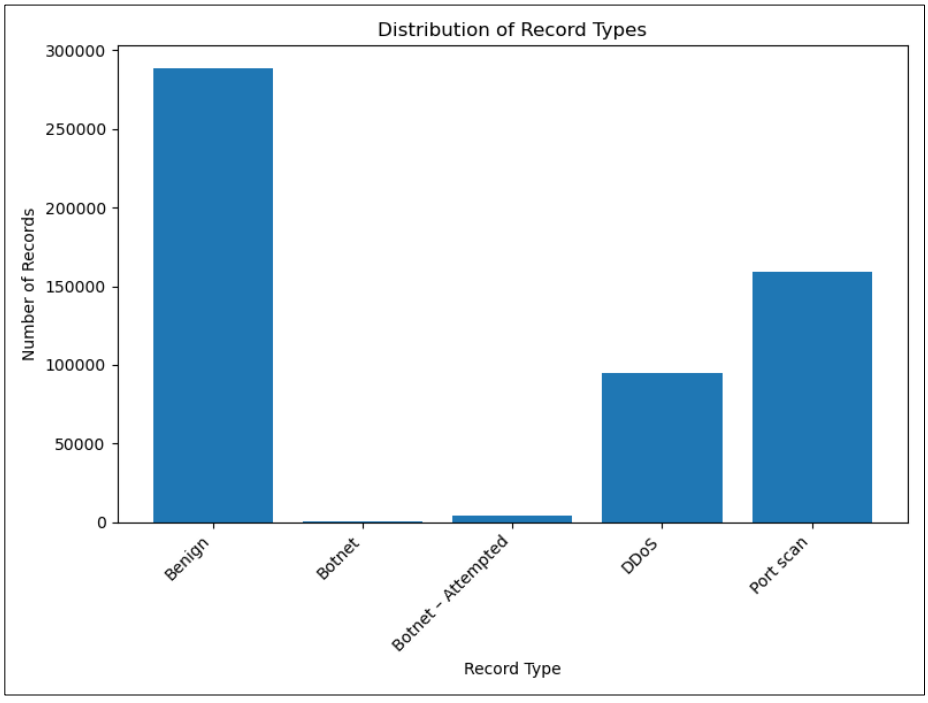

Figure 1. Dataset's classes histogram

2) The shuffling process is applied to input data to enhance the robustness, generalization, and efficiency of the training process, resulting in better model performance on unseen data.
3) Input and output data is converted to tensor datatype to increase efficiency, parallelization, and simplicity in implementing DL models. Tensors provide a powerful and flexible foundation for expressing and manipulating data flows in neural networks.

ISSN: 2312-8135 | Print ISSN: 1992-0652
info@journalofbabylon.com | jub@itnet.uobabylon.edu.iq | www.journalofbabylon.com

4) The normalization of input data is a crucial preprocessing step. It contributes significantly to the stability, efficiency, and generalization of the model during training and deployment. The following formula describe the normalization method that is used which is MinMaxScalar [18].

$$\text{Scaled value} = \frac{\text{original value} - min}{(\max - \min)} \qquad (1)$$

Where:

Scaled value: Transformed value after applying the scaling.

Original value: Value from the original data before scaling.

Minimum (min): The minimum value in the training data for a specific feature.

Maximum (max): The maximum value in the training data for a specific feature.

and the new range of scaled values is between 0 and 1.

5) Split process on input and output data to make training phase and testing phase for machine learning technique.
6) Preprocessing categorical variables using one hot encoding which is essential for neural networks to learn and make predictions on non-numeric data. The output label will be represented as binary vector. Table (1) shows a one hot encoding example for 5 output data categories.

Table 1. One hot encoding vector

Category	label	One hot encoding vector
Benign	0	[1,0,0,0,0]
Botnet	1	[0,1,0,0,0]
Botnet – Attempted	2	[0,0,1,0,0]
DDoS	3	[0,0,0,1,0]
Portscan	4	[0,0,0,0,1]

7) Build the model with specific number of layers, neurons and training/testing data percentage and calculate the time of training and classification accuracy of the model.

These steps are shown as flowchart in Figure (2).

ISSN: 2312-8135 | Print ISSN: 1992-0652
info@journalofbabylon.com | jub@itnet.uobabylon.edu.iq | www.journalofbabylon.com

RESULTS AND DISCUSSION

Different amounts of input features, hidden layers, and neurons within layers are employed in ANN models. First, all features were chosen to be entered into the model. In second phase 7

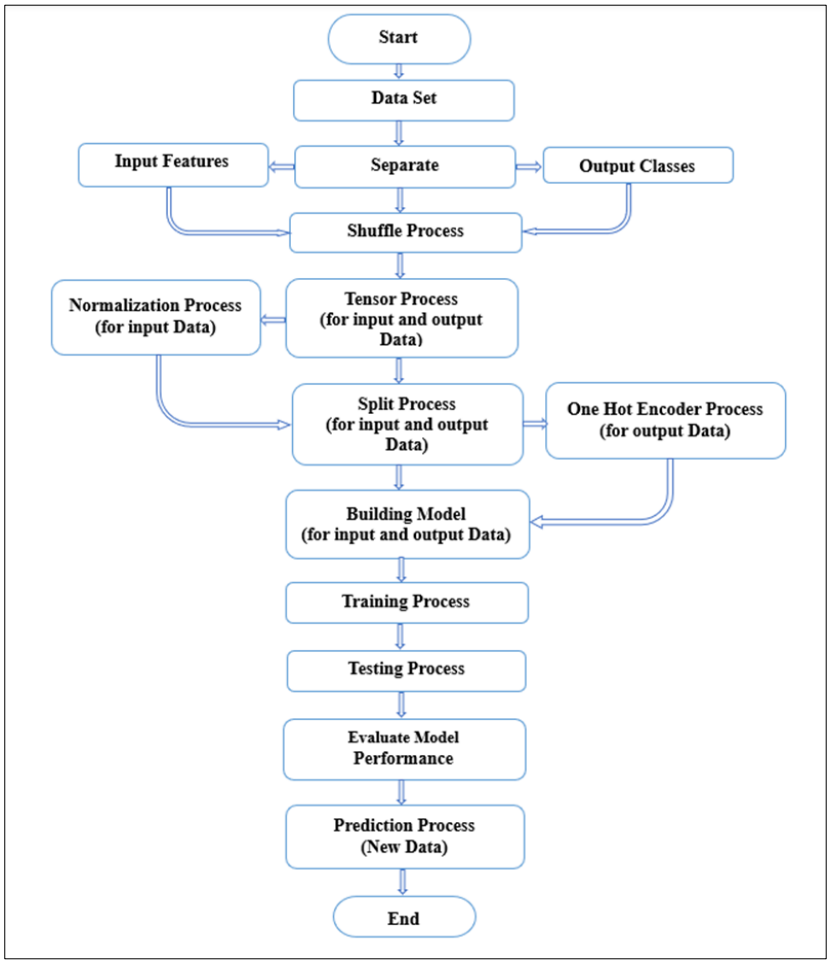

Figure 2. Flowchart of Proposed Model

features were selected and the results are listed in this section, finally only 5 features were selected as input. Models are built using Python language.

In this experiment, 81 features are taken as input features to ANN from the CICIDS2017 dataset, specifically Friday data and tried to see the effect on accuracy and time when neurons and layers are changed. 80% taken for the training phase and 20% taken for the testing phase with batch size 32.

One hidden layer was used with varying numbers of neurons, and the results were recorded in Table (2). The first column represents the number of neurons in the first layer which is changed from 1 to 64 and the second column represents the accuracy of the training model. The accuracy increased proportionally with the number of neurons, as shown in Figure (3). The time taken was

approximately 10 seconds, with a minor variation in the millisecond range. The best accuracy of 99.2% was achieved with 64 neurons.

Table 2. Results with one hidden layer in ANN

Neuron Layer 1	Accuracy %
1	96.5
2	97
4	97.3
8	98.3
16	98.6
32	98.7
64	99.2

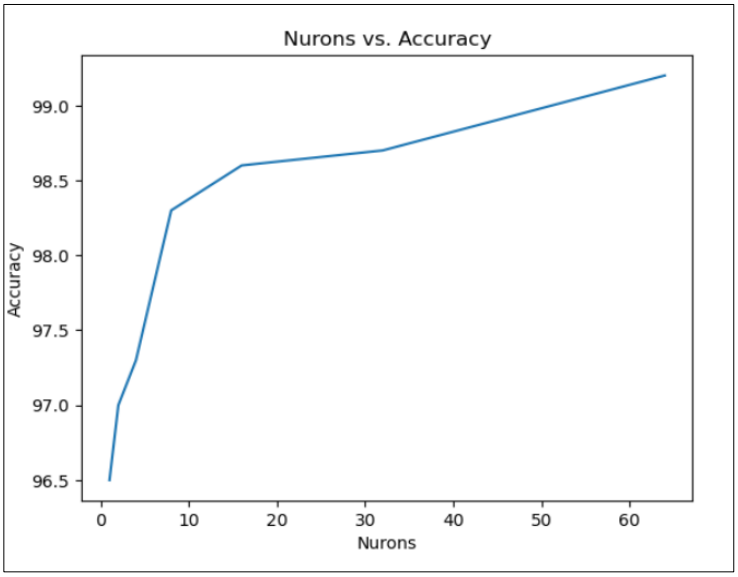

Figure 3. Accuracy obtained using One hidden layer

Two and three layers are tested too. Our results have shown that, for two layers, the optimal configuration is 16 neurons in layer 1 and 32 neurons in layer 2, while for three layers, the optimal configuration is 16 neurons in layer 1, 32 neurons in layer 2, and 64 neurons in layer 3. These configurations produced accuracy levels of around 99%.

ISSN: 2312-8135 | Print ISSN: 1992-0652
info@journalofbabylon.com | jub@itnet.uobabylon.edu.iq | www.journalofbabylon.com

ISSN: 2312-8135 | Print ISSN: 1992-0652
info@journalofbabylon.com | jub@itnet.uobabylon.edu.iq | www.journalofbabylon.com

Table 3. Results with two hidden layers in ANN

Neuron Layer 1	Neuron Layer 2	Accuracy %
1	2	93.67
2	4	97
4	8	98.04
8	16	98.59
16	32	99.07
32	64	99.37
64	128	99.56

Furthermore, authors observed a gradual increase in accuracy as the number of neurons in each layer was increased. The tests were conducted multiple times, and the results were consistent. The time required for both tests was approximately 11 seconds with a tiny variation in millisecond range. These results shown in Tables (3), (4) and Figures (4), (5).

Table 4. Results used three hidden layers in ANN

Neuron Layer 1	Neuron Layer 2	Neuron Layer 3	Accuracy %
1	2	4	94.8
2	4	8	96.87
4	8	16	98.4
8	16	32	98.85
16	32	64	99.29
32	64	128	99.53
64	128	256	99.68

ISSN: 2312-8135 | Print ISSN: 1992-0652
info@journalofbabylon.com | jub@itnet.uobabylon.edu.iq | www.journalofbabylon.com

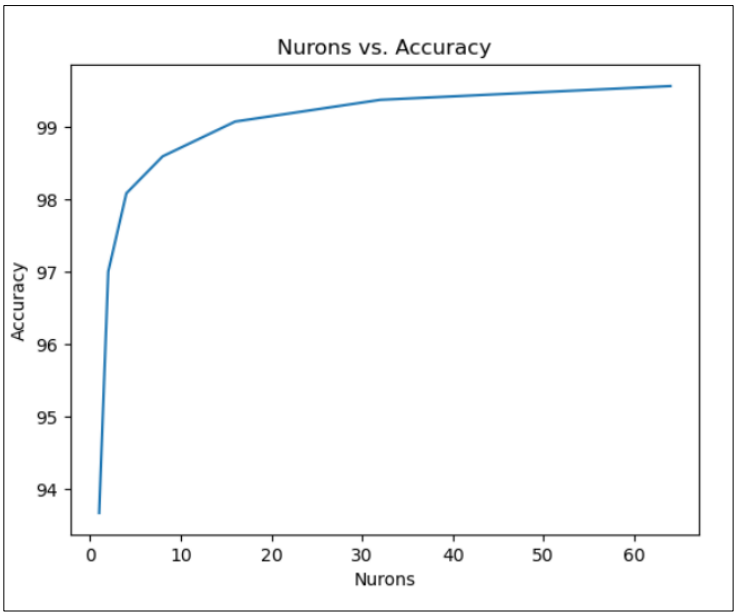

Figure 4. Accuracy obtained using two hidden layers

Table (5) shows the results of using four layers with different numbers of neurons. The model took around 12 seconds to train, but for last record in which the combination of neurons is (64, 128, 256, 512) it took 19 seconds. The accuracy increased gradually, reaching 99% with 8 neurons in layer 1, 16 in layer 2, 32 in layer 3, and 64 in layer 4. The accuracy shown in Figure (6).

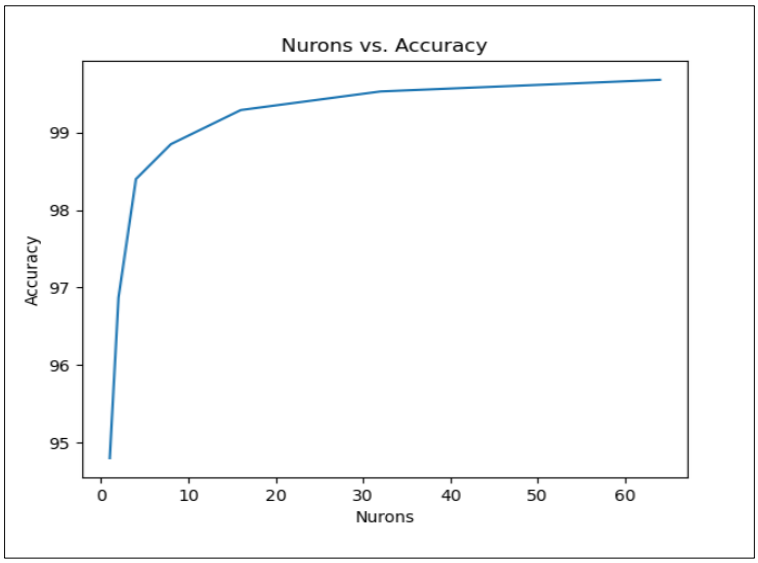

Figure 5. Accuracy obtained using three hidden layers

ISSN: 2312-8135 | Print ISSN: 1992-0652
info@journalofbabylon.com | jub@itnet.uobabylon.edu.iq | www.journalofbabylon.com

Table 5. Results with four hidden layers in ANN

Neuron Layer 1	Neuron Layer 2	Neuron Layer 3	Neuron Layer 4	Accuracy %
1	2	4	8	96.55
2	4	8	16	97.96
4	8	16	32	98.53
8	16	32	64	99.33
16	32	64	128	99.44
32	64	128	256	99.64
64	128	256	512	99.71

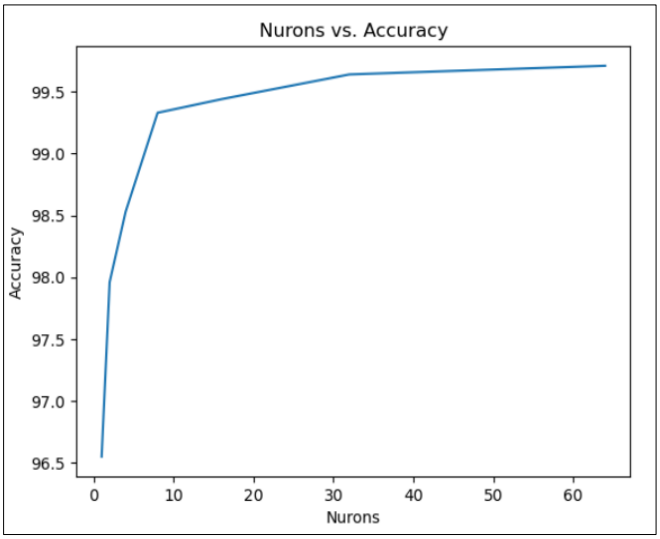

Figure 6. Accuracy obtained using four hidden layers

In this section, seven features and five features are used with 1 layer and 4 layers to investigate the difference between them. The selected seven features are Source port, Destination port, Backward Packet Length Max, Backward Packet Length Mean, Protocol, Packet Length Mean, and Packet Length Std. The accuracy increases proportionally with respect to the number of neurons. The time was approximately 10 seconds with a tiny variation in millisecond domain. 64 neurons produce a better accuracy equal to 96.3%. Table (6) and figure (7) shows the accuracy using one hidden layer while in table (7) and figure (8) the hidden layers are four.

ISSN: 2312-8135 | Print ISSN: 1992-0652
info@journalofbabylon.com | jub@itnet.uobabylon.edu.iq | www.journalofbabylon.com

Table 6. Results with one hidden layer and 7 features

Neuron Layer 1	Accuracy %
1	83.51
2	89.79
4	92.01
8	94.43
16	94.79
32	95
64	96.3

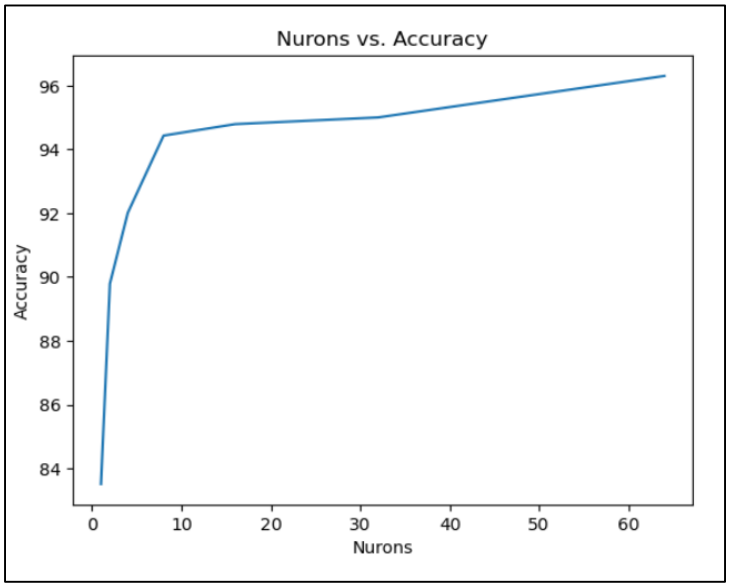

Figure 7. Accuracy obtained using one hidden layer and 7 features

The experiments involved using four layers in the hidden layer, each with different numbers of neurons. The results are presented in Table (7). The corresponding accuracy values for different numbers of neurons are presented in Figures (8) the time taken to train the models with 64, 128, 256, and 512 neurons combination is 19 seconds. Based on the results, using 8 neurons in the first layer, 16 neurons in the second layer, 32 neurons in the third layer, and 64 neurons in the fourth layer produced higher accuracy, around 97%.

ISSN: 2312-8135 | Print ISSN: 1992-0652
info@journalofbabylon.com | jub@itnet.uobabylon.edu.iq | www.journalofbabylon.com

Table 7. Results with four hidden layers and 7 features

Neuron Layer 1	Neuron Layer 2	Neuron Layer 3	Neuron Layer 4	Accuracy %
1	2	4	8	76.85
2	4	8	16	93.65
4	8	16	32	96.57
8	16	32	64	97.06
16	32	64	128	97.4
32	64	128	256	97.52
64	128	256	512	97.6

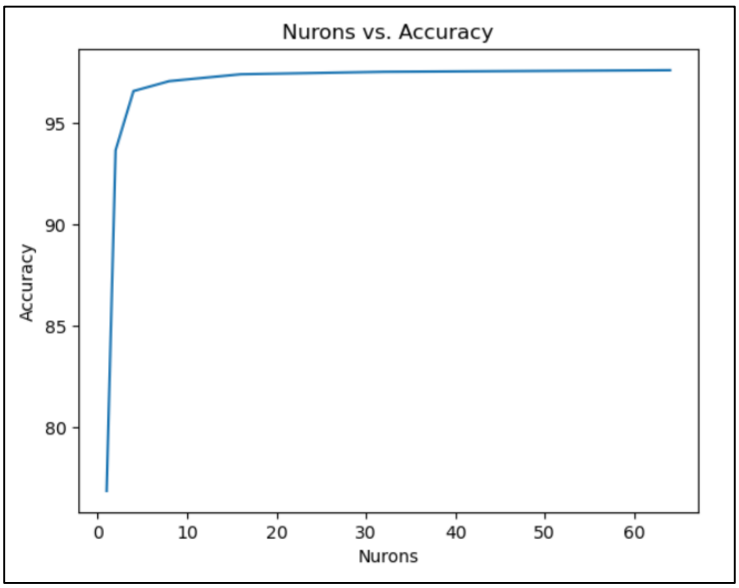

Figure 8. Accuracy obtained using four hidden layers and 7 features

Table 8. Results with one hidden layer and 5 features

Neuron Layer 1	Accuracy %
1	57.03
2	88.5
4	90.94
8	93.57
16	94.9
32	95.18
64	95.55

five input features are used: Destination port, Backward Packet Length Max, Backward Packet Length Mean, Packet Length Mean, and Packet Length Std. Two model structures were tested: one with a single layer and another with four layers. The One-layer model took around 10 seconds to train with a tiny variation in the millisecond range. At 32 neurons, the model achieved an accuracy of approximately 95%, as depicted in Table (8) and Figure (9). The four- layer model took a maximum of 19 seconds to train, with an accuracy of about 96%, as shown in Figure (10) and Table (9).

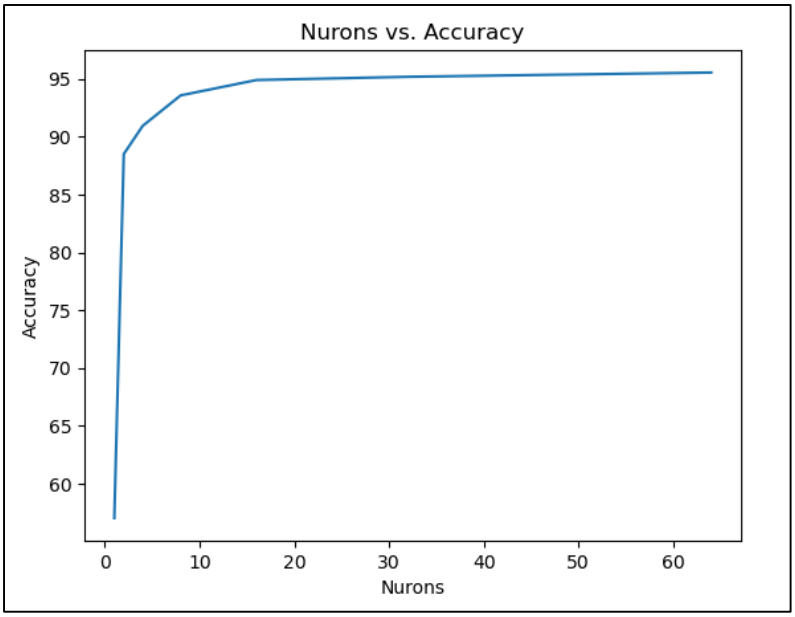

Figure 9. Accuracy obtained using one hidden layer and 5 features

Table 9. Results with four hidden layers and 5 features

Neuron Layer 1	Neuron Layer 2	Neuron Layer 3	Neuron Layer 4	Time/Sec	Accuracy %
1	2	4	8	12	52.71
2	4	8	16	12	95
4	8	16	32	12	95.72
8	16	32	64	12	96.25
16	32	64	128	12	96.62
32	64	128	256	13	96.75
64	128	256	512	18	96.8

ISSN: 2312-8135 | Print ISSN: 1992-0652
info@journalofbabylon.com | jub@itnet.uobabylon.edu.iq | www.journalofbabylon.com

ISSN: 2312-8135 | Print ISSN: 1992-0652
info@journalofbabylon.com | jub@itnet.uobabylon.edu.iq | www.journalofbabylon.com

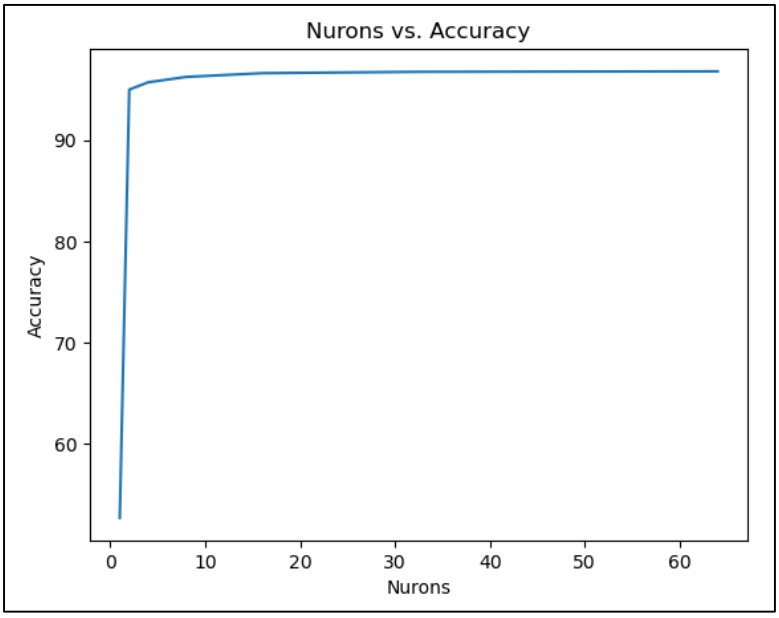

Figure 10. Accuracy obtained using four hidden layers and 5 features

CONCLUSION

Through conducting experiments, it was demonstrated that ANN performs well in identifying adversarial attacks such as Portscan, DDoS, Botnet, and Botnet-Attempted attacks. The following combinations of neurons and layers have been found to provide satisfactory results: (64), (16, 32), (16, 32, 64), and (8, 16, 32, 64), all of which deliver approximately 99% accuracy. 7 features show a good tradeoff between the size of inputs to the model and the accuracy results.

ACKNOWLEDGMENTS

Acknowledgment and appreciation to everyone who assisted to complete this research paper.

Conflict of interests.

There are non-conflicts of interest.

References

[1] A. M. Gamundani and L. M. Nekare, “A Review of New Trends in Cyber Attacks: A Zoom into Distributed Database Systems”, in *2018 IST-Africa Week Conference (IST-Africa)*, 2018, p. Page 1 of 9-Page 9 of 9.

[2] D. Arnaldy and A. R. Perdana, “Implementation and Analysis of Penetration Techniques Using the Man-In-The-Middle Attack”, in *2019 2nd International Conference of Computer and Informatics Engineering (IC2IE)*, IEEE, 2019, pp. 188–192. doi: 10.1109/IC2IE47452.2019.8940872.

[3] N. Zhu, X. Chen, and Y. Zhang, "Construction of Overflow Attacks Based on Attack Element and Attack Template", in *2011 Seventh International Conference on Computational Intelligence and Security*, IEEE, 2011, pp. 540–544. doi: 10.1109/CIS.2011.125.

[4] A. M. Kandan, G. JaspherWillsie Kathrine, and A. R. Melvin, "Network Attacks and Prevention techniques - A Study", in *2019 IEEE International Conference on Electrical, Computer and Communication Technologies (ICECCT)*, IEEE, Feb. 2019, pp. 1–6. doi: 10.1109/ICECCT.2019.8869077.

[5] J. Fan, D. Mu, and Y. Liu, "Research on Network Traffic Prediction Model Based on Neural Network", in *2019 2nd International Conference on Information Systems and Computer Aided Education, ICISCAE 2019*, Institute of Electrical and Electronics Engineers Inc., Sep. 2019, pp. 554–557. doi: 10.1109/ICISCAE48440.2019.221694.

[6] B., Mahesh, "Machine Learning Algorithms - A Review", International Journal of Science and Research (IJSR), Volume 9 Issue 1, pp. 381-386, January 2020, https://www.ijsr.net/getabstract.php?paperid=ART20203995.

[7] R. Bhardwaj, A. R. Nambiar, and D. Dutta, "A Study of Machine Learning in Healthcare", in *Proceedings - International Computer Software and Applications Conference*, IEEE Computer Society, Sep. 2017, pp. 236–241. doi: 10.1109/COMPSAC.2017.164.

[8] S. Tufail, H. Riggs, M. Tariq, and A. I. Sarwat, "Advancements and Challenges in Machine Learning: A Comprehensive Review of Models, Libraries, Applications, and Algorithms", *Electronics (Switzerland)*, vol. 12, no. 8. MDPI, Apr. 01, 2023. doi: 10.3390/electronics12081789.

[9] P. Louridas and C. Ebert, "Machine Learning," *IEEE Software*, vol. 33, no. 5. Institute of Electrical and Electronics Engineers (IEEE), pp. 110–115, Sep. 2016. doi: 10.1109/ms.2016.114.

[10] A. Goel, A. K. Goel, and A. Kumar, "The role of artificial neural network and machine learning in utilizing spatial information", *Spatial Information Research*, vol. 31, no. 3. Springer Science and Business Media B.V., pp. 275–285, Jun. 01, 2023. doi: 10.1007/s41324-022-00494-x.

[11] M. Madhiarasan and M. Louzazni, "Analysis of Artificial Neural Network: Architecture, Types, and Forecasting Applications," *Journal of Electrical and Computer Engineering*, vol. 2022. Hindawi Limited, pp. 1–23, Apr. 18, 2022. doi: 10.1155/2022/5416722.

[12] A. Kerim and M. O. Efe, "Recognition of Traffic Signs with Artificial Neural Networks: A Novel Dataset and Algorithm", *International Conference on Artificial Intelligence in Information and Communication (ICAIIC)*. IEEE, Apr. 13, 2021, pp. 171–176. doi: 10.1109/icaiic51459.2021.9415238.

[13] P. Ongsulee, "Artificial intelligence, machine learning and deep learning", in *2017 15th International Conference on ICT and Knowledge Engineering (ICT&KE)*, IEEE, Nov. 2017, pp. 1–6. doi: 10.1109/ICTKE.2017.8259629.

[14] M. Aamir, S. S. H. Rizvi, M. A. Hashmani, M. Zubair, and J. A. Usman, "Machine Learning Classification of Port Scanning and DDoS Attacks: A Comparative Analysis", *Mehran University Research Journal of Engineering and Technology*, vol. 40, no. 1, pp. 215–229, Jan. 2021, doi: 10.22581/muet1982.2101.19.

[15] V. Vamsi Krishna M, et al, "Machine Learning Techniques for Detecting Cyber Attacks in Networks" *International Journal of Research Sciences and Advanced Engineering [IJRSAE]*.

ISSN: 2312-8135 | Print ISSN: 1992-0652
info@journalofbabylon.com | jub@itnet.uobabylon.edu.iq | www.journalofbabylon.com

Thomson Reuters Research ID: D-1153-2018, SJI Listed, Volume 9, Issue 36, PP: 01 - 10, OCT - DEC' 2021.

[16] A. Jenefa *et al.*, "A Robust Deep Learning-based Approach for Network Traffic Classification using CNNs and RNNs", in *ICSPC 2023 - 4th International Conference on Signal Processing and Communication*, Institute of Electrical and Electronics Engineers Inc., 2023, pp. 106–110. doi: 10.1109/ICSPC57692.2023.10125858.

[17] I. Sharafaldin, A. Habibi Lashkari, and A. A. Ghorbani, "Toward Generating a New Intrusion Detection Dataset and Intrusion Traffic Characterization", in *Proceedings of the 4th International Conference on Information Systems Security and Privacy*, SCITEPRESS - Science and Technology Publications, 2018, pp. 108–116. doi: 10.5220/0006639801080116.

[18] S. Soppin, M. Ramachandra, and B. N. Chandrashekar, *Essentials of Deep Learning and AI*. BPB PUBLICATIONS, 2021, ISBN: 978-9391030353.

ISSN: 2312-8135 | Print ISSN: 1992-0652
info@journalofbabylon.com | jub@itnet.uobabylon.edu.iq | www.journalofbabylon.com
مجلة جامعة بابل للعلوم الصرفة والتطبيقية مجلة جامعة بابل للعلوم الصرفة والتطبيقية

Article
JOURNAL OF UNIVERSITY OF BABYLON
For Pure and Applied Sciences (JUBPAS)
Vol.32; No.2.| 2024

الخلاصة

المقدمة:

يعد الأمن السيبراني مصدر قلق بارز في العالم المتصل اليوم، ويشمل ذلك الوصول اللاسلكي والسلكي على حد سواء محليًا وعن بعد عبر منصات تكنولوجيا الاتصالات المتنوعة. من المهم إدراك التهديد الذي يشكله المتسللون الذين يقومون حاليًا بتهديد الوظائف التنظيمية، وتجاوز التدابير الأمنية، وسرقة المعلومات شديدة الحساسية. يتم استخدام تقنيات القرصنة الشائعة من قبل المخترقين بشكل متكرر مثل فحص المنافذ وهجمات منع الخدمة الموزعة (DDoS) واستخدام شبكات الروبوتات الضارة (Botnet).

طرق العمل:

قام المؤلفون باستكشاف فوائد التعلم الآلي لتصنيف فحص المنافذ، وDDoS، وBotnet، وهجمات محاولة تكوين Botnet وسط مزيج من البيانات الطبيعية والبيانات التي تدل على الهجوم. يتم استخدام هياكل مختلفة للشبكات العصبية الاصطناعية لكل منها خصائص مميزة، ويتم تدريبها واختبارها على مجموعة بيانات مرجعية منشورة باسم (CICIDS2017). الهدف هو تحديد نموذج ANN الأكثر فعالية لتصنيف البيانات التي تحتوي على الاحداث لفحص المنافذ ورفض الخدمة الموزعة (DDoS) وBotnet ومحاولات هجمات الروبوتات.

النتائج:

تم استخدام عناصر(الميزات) مختلفة كمدخلات لنموذج تعلم الالة ذو الطبقات المخفية المفردة والمتعددة ولكل منها خلايا عصبية مختلفة لتقييم تاثيرها على دقة التصنيف باستخدام لغة بايثون. أفضل دقة تم الحصول عليها هي 99.71%، تم تحقيقها باستخدام جميع ميزات مجموعة البيانات و4 طبقات مخفية، في حين أن الدقة التي تم الحصول عليها باستخدام 7 ميزات فقط هي 97.6%.

الاستنتاجات:

يمكن لنماذج التعلم الالي تحقيق أداء جيد في تصنيف حركة مرور الشبكة ضد الهجمات العدائية باستخدام العدد الأمثل من الميزات والطبقات المخفية.

الكلمات المفتاحية: حركة مرور الشبكة، الأمن السيبراني، التعلم الآلي، التعلم العميق، كشف التسلل.

ISSN: 2312-8135 | Print ISSN: 1992-0652
info@journalofbabylon.com | jub@itnet.uobabylon.edu.iq | www.journalofbabylon.com