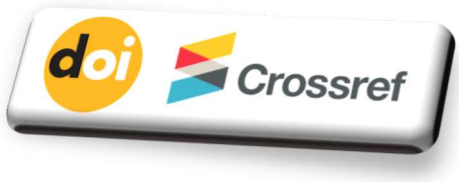
Journal, Advances in Mathematical & Computational Sciences
Vol. 9 No. 2, June 2021



Journal of Advances in Mathematical & Computational Sciences
An International Pan-African Multidisciplinary Journal of the SMART Research Group
International Centre for IT & Development (ICITD) USA

Available online at https://www.isteams.net/ mathematics-computationaljournal.info
DOI: dx.doi.org/10.22624/AIMS/MATHS/V9N2P5
CrossREF Member Listing - https://www.crossref.org/06members/50go-live.html

Application of Reactive Artificial Intelligence Model to Predict Malicious Activities

Yamcharoen P[1], Folorunsho O.S[2], Bayewu A[3] & Ojo T.P[4]
[1] Washington University of Science and Technology, Vienna, VA, 22182, USA
[2] Washington University of Science and Technology, Vienna, VA, 22182, USA
[4]Northumbria University, New Castle, NE1 8ST, UK
[4]University of Indianapolis, IN, 46227, USA
E-mail: [1]ami.yamcharoen@wust.edu; [2]omowunmisesekinatf@yahoo.com
Hadeola_oyeyipo@yahoo.com; [4]titilikesyou@gmail.com

ABSTRACT

The critical infrastructure in the United States have been attacked countless time, and the federal, state, and private entities are concerned with securing their network infrastructure and sensitive data to prevent revenue loss and intellectual properties. Billion dollars are lost to cyberattacks globally, and the United States tops the list of countries that recorded more cyber-attacks. The healthcare industries have lost revenue due to litigation from compromised patient data and revenue paid as a ransom to cyber attackers. Integrating a reactive artificial neural network model into their cybersecurity architecture will monitor the activities of users' activities internally or externally accessing the organization's resources. The Newton method algorithm was applied to predict the malicious activities and the performance of the model support derivative matrix from the input variables that were used to train the model. The validation result of the algorithm has proven that an artificial neural network model can be deployed and integrated into the cybersecurity solution to predict the likelihood of an attack.

Keywords: Healthcare, Cybersecurity, Artificial Intelligence, Threat Actors, Risk Assessment.

Yamcharoen P., Folorunsho O.S., Bayewu A. & Ojo T.P. (2021): Application of Reactive Artificial Intelligence Model to Predict Malicious Activities. Journal of Advances in Mathematical & Computational Science. Vol. 9, No. 2. Pp 61-68. DOI: dx.doi.org/10.22624/AIMS/MATHS/V9N2P5
Available online at www.isteams.net/mathematics-computationaljournal.

1. INTRODUCTION

The United States of America's critical infrastructures have been the primary target of threat actors across the globe simply because the compromised data are in high demand on the dark web. Numerous websites sell credit card information, social security numbers, date of birth, and other valuable information to third parties for personal gain.

Threat actors attack the critical infrastructure with artificial intelligence-generated ransomware that is very difficult to track if the organization's business operation runs on a weak cybersecurity architecture. However, most organizations in the United States have beefed up their cybersecurity monitoring solution and integrated machine learning algorithms to scan their cybersecurity environment seamlessly. Some have implemented a cybersecurity risk assessment as one of their strategic cybersecurity plans, yet the threat actors are hacking into these critical infrastructures to steal sensitive transactional and customer data, which causes the United States to lose billions of dollars in revenue to threat actors annually. One of the largest healthcare institutions was used as a case study in this research. The organization's leadership aims to align the cybersecurity goals and business operation goals to ensure their cybersecurity tenant is secured and protected from threat actors.

The average revenue lost annually due to cyberattacks is more than a million dollars, and revenue is lost due to litigation from patients whose data were compromised. To reduce revenue loss to threat actors in the form of ransom and litigation costs arising from states and federal law sue, the leadership is interested in integrating an effective and robust artificial intelligence model that will analyze internal and external activities on the organization tenant to predict the likelihood of an attack. The log data from the organization systems, applications, portals, and databases, both on-premises and cloud, will be analyzed to determine the risk impact. The log will be analyzed using Log duration, access frequency, password attempt, IP Address Status, Resource Assess, Risk Value, Malicious Status.

2. LITERATURE REVIEW

Cybersecurity experts have designed numerous artificial intelligence (AI)-based solutions for a broad range of cyber security applications, partly due to the growing understanding of organizations in the importance of AI in mitigating cyber threats (Zhang et al., 2022). AI-based approaches to nonlinear model problems have been shown to perform well in nonlinear classification (Ozsen et al., 2009), which can also be used to facilitate cyber threat classification. Interests in AI-based solutions are also partly driven by advances in computing capabilities. It requires a reasonable amount of time to train large-scale log data using a machine learning-based classification system on cloud infrastructure to predict the likelihood of an attack in a decentralized wide-area network architecture (Donepudi, 2015). Computing power for AI-based approaches reportedly doubles every three months, surpassing Moore's law. Such capabilities can be utilized to improve AI-based cyber security solutions' performance.

Understandably, machine intelligence cannot replace human intelligence, and the next generation of AI will probably combine human and machine intelligence (Akinsola et al., 2021). Artificial Intelligence and Blockchain intersection is growing rapidly in healthcare and digitalization of healthcare operation has increase the risk level of their business operation (Muheidat & Tawalbeh, 2021). The constant change in the cybersecurity landscape and pattern of attack by threat actors requires effective AI tools to facilitate and enhance human work and not replace the work of physicians and other healthcare staff. AI is ready to support healthcare personnel with various tasks, from administrative workflow to clinical documentation and patient outreach, as well as specialized support such as image analysis, medical device automation, and patient monitoring.

There are different opinions on the most beneficial applications of AI for healthcare purposes. Forbes stated that the most important areas of artificial intelligence would be administrative workflows, image analysis, robotic surgery, virtual assistants, and clinical decision support. However, with the digitalization of healthcare operations, artificial intelligence has been applied broadly to support automating cybersecurity monitoring processes to detect malicious activities in real time (Mar 2018).

Also, a report by Accenture mentioned the same areas and included connected machines, dosage error reduction, and cybersecurity (Kallis et al., 2018). 2019 McKinsey report states important areas are connected and cognitive devices, targeted and personalized medicine, robotics-assisted surgery, and electroceuticals (Singhal & Carlton, 2019). In centralized systems and applications like healthcare, data access and processing in real-time among various information systems is a bottleneck. Blockchain's decentralized database architecture, secure storage, authentication, and data sharing would solve this problem. Besides, Artificial Intelligence can live at the top of Blockchain and generate insights from the generated shared data used to make predictions (Podder et al., 2021). Blockchain is a high-level cybersecurity technology that forms chains that connect the existing blocks stored in nodes and the new block chronologically by mutual agreements between nodes.

Technology convergence accelerates various industries' growth, such as banking, insurance, cybersecurity, forecasting, medical services, cryptocurrency (Soni, 2020). Recently, new information and communication technologies have changed the way of operations in all fields of life, such as intelligent transportation systems, agriculture, education, and healthcare systems (Qureshi et al., 2014). Artificial Intelligence (AI) has important applications for medicine and healthcare. AI can be useful in clinical decision-making (Signorini et al., 1999). Researchers have studied and designed intelligent agents, analyze the environmental data to learn and perceive an environment's risk factors and takes actions that maximize its chances of success (Komninos, 2009). Artificial Intelligence has the computational capability to provide results or examples of behavior that are characteristics of human intelligence (Charniak et al., 2014). AI has many advantages, such as flexibility, adaptability, pattern recognition, and fast computing (Kumar et al., 2010). The study of AI started approximately 25 years ago, and since that time, many brilliant computer scientists have performed AI research and have developed this field.

More recently, AI has been used by medical doctors to attempt to resolve problems faced by the medical community. Based on the knowledge that has already been gained, it appears auspicious that we can find solutions to some of the pressing problems faced today in medicine. Machine learning is a very important discipline in AI (Bratko et al., 2009). Artificial neural networks (ANN) are nonlinear numerical-based data processing systems that function by simulating the human brain's neural networks (Haykin, 1994). In the ANN method, several parameters governing relationships and dependencies are learned and trained for guessing a targeted phenomenon by processing observational data. The cybersecurity goals were mapped with the attributes determining the likelihood of an attack.

The network mapping will be simple to remove noisy data that might create complexity between input and output variables to predict unknown outputs from known inputs. The ANN will estimate linear and nonlinear functions and future events (Jain et al., 1999). The active nodes and perceptrons will identify the hidden layer that might affect the model's performance. A series of connections link inputs to perceptrons and perceptrons to outputs. Data (or signals) pass through the connections between perceptrons (or neurons), and the connection weight (w) controls the strength of the transmitted signal (Yu et al., 2018). The input layer is the responsible layer for providing data. The output layer (or the last layer) contains the values predicted by the network and therefore introduces the model output. The middle or hidden layers are made up of processing neurons meant for data processing, as shown in Figure 2. The ANN model's optimum structure, i.e., the number of neurons and hidden layers in each hidden layer, is commonly defined by trial and error (Nadiri et al., 2013). During the data preparation stage, the noisy data were removed, and the sample dataset is shown in Table 1.

Table 1: Sample Data

User ID	User Type	Log Duration (mins)	Access Freqency	IP Address	Resource Assessed	Password Attempt	Risk Value	Malicious Status
1	External	200	10	Anonymous	Database	15	High	Yes
2	Internal	20	2	Known	Portal	1	Low	No
3	External	48	2	Known	System Application	2	Medium	Yes
4	External	120	22	Anonymous	System Application	10	High	No

3. METHODOLOGY

The Cross Industry Standard Process for Data Mining (CRISP-DM) was adopted as the research methodology as shown in Figure 1. The CRISP-DM model is a cyclic and iterative model that consists of six steps (Business Understanding Phase, Data Understanding Phase, Data Preparation Phase, Modeling Phase, Evaluation Phase, and Deployment). The organization's cybersecurity goals and objectives identified at the business understanding phase are ensuring zero downtime should an attack occurs, mitigating risks that might impact business operations, and detecting malicious activities in real-time. T

he organization has a centralized location at the security operation center where they track and store the activities log of the organization. The activities log includes log files from systems, applications, portals, and other third-parties technologies integrated into the organization's cybersecurity architecture to track usage within the organization and external logs accessing the organization's tenant.

During the data understanding phase, the duration for each logged and unlogged activity is captured, as the frequency of access from the location, the IP address of the location, the type of resource assessed by the user, the number of times the password was attempted, the risk value of the cybersecurity architecture and the user acts malicious or not. The target variable is a nominal attribute (Malicious Status) which is directly dependent on the user type, IP Address, Risk Value and Log duration.

The ANN will take inputs which are predicting variables (user type, IP Address, Risk Value and Log duration) from the organization incident data repository and numbered it from 1 to N. The input i is called x_i and the associated weight is called W_j. The total input to a unit is the weighted sum over all inputs, $\sum_{I=1}^{N} w_i x_i = w_1x_1+w_2x_2+. . .+ W_NX_N$

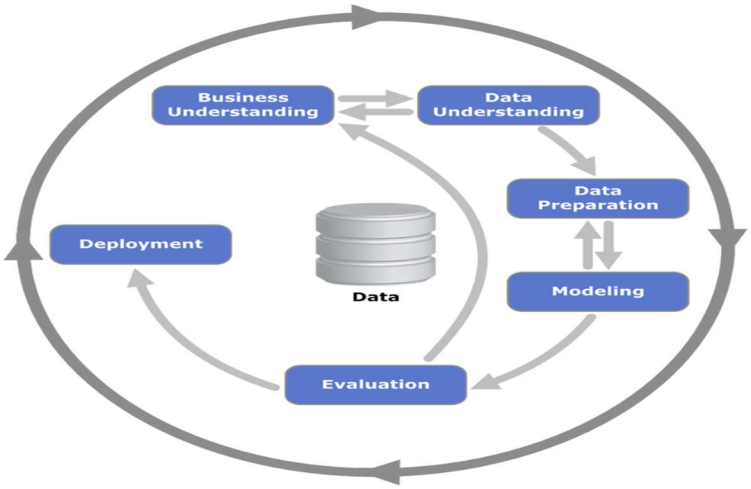


Figure 1: Cross Industry Standard Process for Data Mining

If the weighted sum is below the threshold *Y*, the output unit would be 1 and 0 otherwise, thus the output of the output will be expressed as g ($\sum_{I=1}^{N} w_i x_i = Y$). The model will be trained based on the workflow shown in Figure 2.

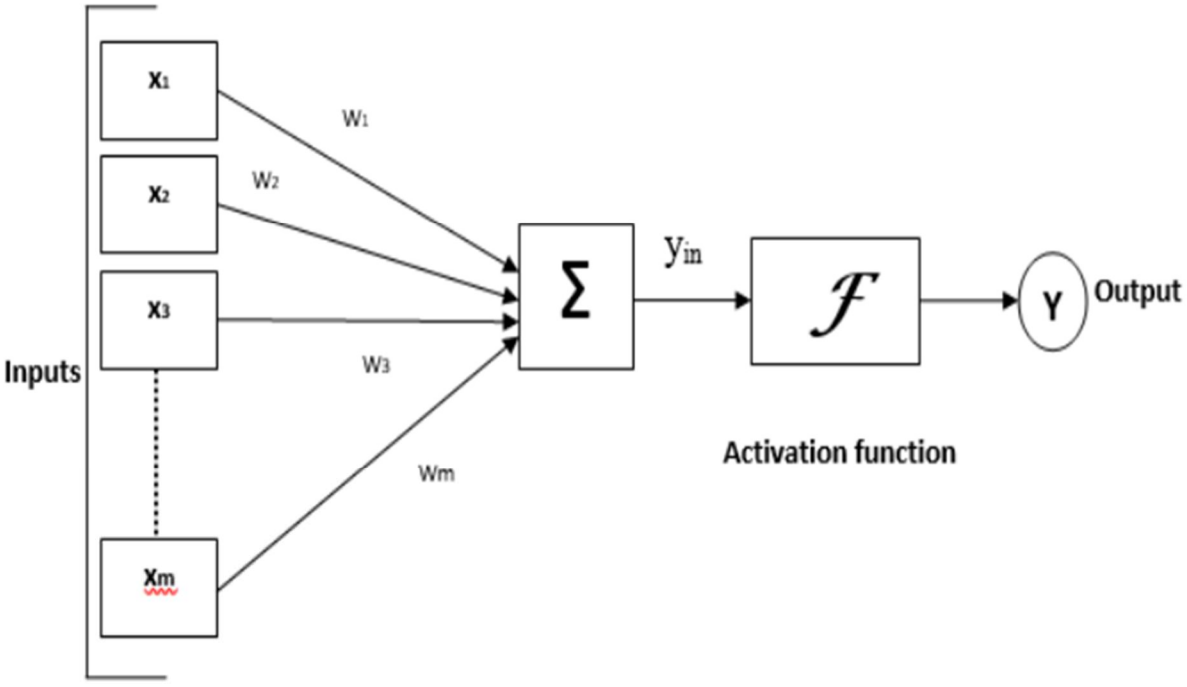


Fig 2: Workflow Basis for Training Model .

4. RESULTS AND DISCUSSION

Over two million log records were pulled from the organization's central repositories. After cleaning and data manipulation, a back propagation learning algorithm was applied for feed-forward networks with continuous output, as shown in Figure 2. The model's training was done by setting the weights in the network using the user type, log durations, and the number of passwords attempts to generate a pattern between the IP Address and the predicting variables.

The square difference between the output and desired output was measured to determine the correct class of malicious activities. The weights of the random values generated were used to minimize the total error and improve the precision and accuracy of the model. The backpropagation was used to optimize the model performance so that the results obtained from the training model depend on the initial value of the weights of the inputs. However, overfitting is our major challenge during the training model, and the result of the trained model is shown in Table 2.

Table 2: Model Result

	Newton Method	Quasi-Newton Method
Training	22.78	21.49
Testing	21.54	19.87
Validation	19.06	17.86
All Data	23.67	21.61

The ANN model was trained with two different algorithms: Newton's and quasi-newton methods. The Newton method performs better in predicting if a user's activity is malicious or not. Newton's methods' testing and validation values are better than the quasi-newton methods simply because the Newton method uses the gradient.

The Hessian matrix of the second derivate of the Y function in the process flow formula was minimized to improve the model's performance. However, in quasi-newton methods, the Hessian matrix is auto computed, which does not support the re-calibration of the model to improve performance. One of the challenges encountered during the modeling phase of the project is overfitting.

5. FUTURE WORKS AND CONCLUDING REMARKS

Future work should focus on reducing errors generated when the model is trained. It is essential to reduce the model's weight to improve performance and optimization to ensure that malicious user activities are detected in real-time to reduce the model processing time. The user activity validation is critical, and they should be enough memory to process the incident data so that threat actors will not compromise the organization's cybersecurity architecture. Conclusion: The Newton Method algorithm works with a back propagation algorithm for feeding the networking with continuous output using the weight generated from the output to train the model. The validation value of the model has proven that user activities can be detected if the model is well-calibrated, and the target variable is directly dependent on the predicting variables. The comparative analysis of the model has proven that a reactive artificial neural network can be developed and deployed for healthcare institutions if the model can be optimized and regularized to reduce the computing error that might hinder the model's performance. When trained with Newton Method, the design supports an iterative matrix, which should help the model handle exponential growth in the incident data If overfitting is addressed.

REFERENCE:

1. Akinsola, J. E. T., Akinseinde, S., Kalesanwo, O., Adeagbo, M., Oladapo, K., Awoseyi, A. & Heimgartner, R. (2021). *Application of artificial intelligence in user interfaces design for cyber security threat modeling* (pp. 1-28). IntechOpen.
2. Bratko, I., Michalski, R. S., and Kubat, M. (1999). Machine learning and data mining: methods and applications, 1999
3. Donepudi, P. K. (2015). Crossing point of Artificial Intelligence in cybersecurity. *American Journal of Trade and Policy*, *2*(3), 121-128.
4. Haykin, S. (1994). Neural networks: A comprehensive foundation. Prentice Hall PTR.
5. Kalis, B., Collier, M., & Fu, R. (2018). 10 promising AI applications in health care. *Harvard business review*.
6. Komninos, N. (2009). Intelligent cities: towards interactive and global innovation environments. *International Journal of Innovation and Regional Development*. 1:337–355, 2009.
7. Kumar, G., Kumar, K., & Sachdeva, M. (2010). The use of artificial intelligence-based techniques for intrusion detection: a review. *Artif. Intell. Rev.* 34:369–387, 2010.
8. Jain, S. K., Das, A., & Srivastava, D. K. (1999). Application of ANN for reservoir inflow prediction and operation. *Journal of water resources planning and management*, *125*(5), 263-271.
9. Marr, B. (2018). How is AI used in healthcare-5 powerful real-world examples that show the latest advances. *Forbes, July 27*.
10. Muheidat, F., & Tawalbeh, L. A. (2021). Artificial intelligence and blockchain for cybersecurity applications. In *Artificial Intelligence and Blockchain for Future Cybersecurity Applications* (pp. 3-29). Cham: Springer International Publishing.
11. Ozsen S, Gunes S, Kara S, Latifoglu F (2009) Use of kernel functions in artificial immune systems for the nonlinear classification problems. IEEE Trans Inf Technol Biomed 13(4):621–628
12. Podder, P., Bharati, S., Mondal, M., Paul, P. K., & Kose, U. (2021). Artificial neural network for cybersecurity: A comprehensive review. *arXiv preprint arXiv:2107.01185*.
13. Qureshi, K. N., Abdullah, A. H., & Anwar, R. W. (2014) Wireless sensor-based hybrid architecture for vehicular ad hoc networks, *TELKOMNIKA (Telecommunication Computing Electronics and Control),* vol. 12, pp. 942–949, 2014.

14. Singhal, S., & Carlton, S. (2019). The era of exponential improvement in healthcare. *McKinsey & Company*.
15. Yu, H., Wen, X., Feng, Q., Deo, R. C., Si, J., & Wu, M. (2018). Comparative study of hybrid-wavelet artificial intelligence models for monthly groundwater depth forecasting in extreme arid regions, Northwest China. *Water resources management*, *32*, 301-323.
16. Zhang, Z., Ning, H., Shi, F., Farha, F., Xu, Y., Xu, J., & Choo, K. K. R. (2021). Artificial intelligence in cyber security: research advances, challenges, and opportunities. *Artificial Intelligence Review*, 1-25.