How to Cite:
Tamilarasi, M., Karthick, R., Sanjeev, R., & Sasidharan, S. (2022). Intrusion detection in healthcare domain using machine learning. *International Journal of Health Sciences*, *6*(S4), 5861–5872. https://doi.org/10.53730/ijhs.v6nS4.9460

Intrusion detection in healthcare domain using machine learning

Dr. M Tamilarasi
Associate Professor, Department of Information Technology, K.S. Rangasamy College of Technology, Tiruchengode, Namakkal, Tamil Nadu, India

R Karthick
Department of Information Technology, K.S. Rangasamy College of Technology, Tiruchengode, Namakkal, Tamil Nadu, India

R Sanjeev
Department of Information Technology, K.S. Rangasamy College of Technology, Tiruchengode, Namakkal, Tamil Nadu, India

S Sasidharan
Department of Information Technology, K.S. Rangasamy College of Technology, Tiruchengode, Namakkal, Tamil Nadu, India

***Abstract*---**Intrusion detection system is considered as one of the main sources for protection of important information and communication technologies especially in healthcare networks. Intrusion detection system must be updated frequently because of intrusions that improves periodically. Due to particular limitations such as resource-constrained devices, limited memory and battery capacity of nodes, and unique protocol stacks, traditional intrusion detection algorithms must be updated and improved for application to the Internet of Things. To address this issue, a lightweight attack detection approach based on supervised machine learning–based FIDS (Frothy Disturbance Intrusion Detection System) was developed to detect an adversary attempting to inject unneeded data into the network. FIDS can play an important role in intrusion detection. The proposed model was trained with KDD (Knowledge Discovery in Database) dataset using SVM (Support Vector Machine) algorithm. AODV (Ad-hoc- on-demand Distance Vector) routing protocol was used for routing to make it more energy efficient since the energy usable was limited. The process of packet transmission was done with the help of clustering algorithm to separate the sensors into groups with each group containing a cluster head. Simulation results showed that the proposed FIDS -based classifier, aided by a combination of

Manuscript submitted: 27 March 2022, Manuscript revised: 18 May 2022, Accepted for publication: 9 June 2022

two or three in complex features can perform satisfactorily in terms of classification accuracy and detection time.

***Keywords*---**intrusion detection, healthcare domain, machine learning.

Introduction

Network traffic and cyber-attacks have been more complicated and diverse as a result of the rapid development of 5G, IoT, and Cloud Computing. IDS (Intrusion Detection System) should be able to accurately identify malicious network attacks, provide real-time monitoring and dynamic protection measures, and design strategies as the second line of defense after the firewall. As a result, the IDS should be updated to reflect new threats and incursions. According to Iman Sharafaldin et al., the exponential growth in the size of computer networks and created applications has resulted in a large increase in the potential damage that can be produced by launching assaults. In the meantime, IDS and Intrusion Prevention Systems (IPSs) are two of the most significant protection mechanisms against increasingly sophisticated network threats. Anomaly-based techniques in intrusion detection systems struggle with appropriate deployment, analysis, and evaluation due to a lack of adequate dataset.

The concept suggested by Amirhossein Gharib et al. is that the growing number of security risks on the Internet and computer networks need extremely trustworthy security solutions. Meanwhile, IDS and IPS play a critical role in the design and development of a secure network architecture capable of detecting and preventing a wide range of threats. They looked at existing datasets for testing and evaluating IDSs and proposed a new framework for evaluating datasets that have the following characteristics: Attack Diversity, Anonymity, Available Protocols, Complete Capture, Complete Interaction, Complete Network Configuration, Complete Traffic, Feature Set, Heterogeneity, Labeled Dataset, and Metadata. However, there hasn't been enough research done on how to evaluate and appraise the datasets themselves.

The IP-based Ubiquitous Sensor Network (IP-USN), suggested by Syed Obaid Amin et al, is an attempt to establish the "Internet of Things." Advantage of well-established IP network tools and technology is utilized by using IP for low-power networks. Securing IP-USN, along with many other unresolved difficulties, is a major priority for researchers in order to meet future market satisfaction and desires. It's difficult to imagine an IP-USN realm without robust security measures, both reactive and proactive. In this work, a design for IP-RIDES (Robust Intrusion Detection System) was offered. RIDES is a hybrid intrusion detection system that uses both signature and anomaly detection to detect intrusions. This work only explored the implementation of a distributed pattern matching algorithm with the help of signature-code, a dynamically formed attack signature identifier, for the signature-based intrusion detection component. Other issues, such as rule formulation, are not addressed. The results of the evaluation suggest that storing the Snort signature set takes up about 95% less space than standard storage requirements.

Eung Jun Cho et al. Botnet has recently been used to launch spam, keylogging, and DDoS attacks. A botnet is a group of bots controlled by an attacker. To detect Botnet on a wired network, a variety of detection approaches have been proposed. Botnet attacks, on the other hand, are undetectable in IP-based sensor networks. The Internet of Things (IoT) is a new paradigm that combines the internet with things from various domains of human civilization (such as smart homes, health care, smart grids, manufacturing processes, product supply networks, and environmental monitoring) (IoT). However, because of their widespread use and extensive range, IoT networks are vulnerable to cyber-attacks.

The system administrator puts IDS agents in network entities with large computing and storage capacities, but the memory, processing power, and battery energy-capacity limits of IOT network nodes that host IDS agents are difficult to overcome. End systems are directly connected to specific nodes (e.g., Wireless access points, switches, and routers) in traditional networks, which are responsible for delivering packets to their destination. IOT networks, on the other hand, have numerous hops. Regular nodes can forward packets and act as end systems at the same time. The topology of the network changes on a regular basis (e.g. Vanets, mobile sinks, dynamic selection of cluster heads). The topology's uniqueness presents new hurdles for ids The protocols used in IoT networks differ from those used in traditional networks. To overcome the above drawbacks, an hybrid model is used with the cluster based architecture and AODV routing. Which will be explained below.

The advantage of the proposed solution is,

- The model requires less energy to detect attacks.
- Each ids agent relies on a policy that minimizes packet transmission.
- Cluster head is used to reduce energy consumption, amount of data in the entire network and to increase network lifetime.
- Intrusion detection model determines whether an intrusion has occurred, and classifies the type of attacks.

Previous work

In the studies of network intrusion detection primarily based by machine learning, pupils specially primarily traffic detection was distinguished from abnormal network by spatiality reduction, cluster and classifying, which all serve to realize the identifying the malicious attacks. In Pervez paper a current technique used for feature choices and classifying merging of the multi-class NSL-KDD dataset exploitation Support Vector Machine (SVM) and mentioned the classification accuracy of classifiers underneath utterly totally different dimension options. city studied some new technologies to reinforce CANN intrusion detection methods' classification performance and evaluated their performance on the NSL-KDD dataset. He classified the data using the K Farthest Neighbour (KFN) and also the K Nearest Neighbour (KNN), and he been used the Second Nearest Neighbour (SNN) as soon as the closest and farthest buddies had the same beauty label. The end result suggests the CANN detection fee and decreases the failure fee. The alert fee is progressed or gives the identical performance. Bhattacharya projected a device studying version supported through hybrid main aspect

analysis (PCA)–Firefly. The dataset used the open dataset gathered through Kaggle. Firstly, the version plays one key mystery write for transforming the IDS dataset, then makes use of the hybrid PCA-Firefly rule to scale back the dimension, and additionally the Boost set of rules classifies the decreased dataset. In latest years, the effective cap potential of computerized function extraction, deep studying has created exceptional achievements within the fields of laptop vision and autonomous driving (AD). Several students follow deep mastering to intrusion detection for visitor's type, which has come to be a warm spot of cutting-edge research. The method of deep mastering is to mine the ability traits of high-dimensional records thru a training version and redecorate community visitor's anomaly detection right into a type drawback. Through an oversized style of pattern facts training, adaptatively mastering among conventional community visitors and ordinary community visitors efficaciously enhances the period of time intrusion processing. Torres et al. initialized reborn community site visitor's traits into a sequence of characters, then used repeated Neural Network (RNN) to find out the temporal characteristics, which were more accustomed of the observing malicious network traffic. Wang et al. proposed a malicious software package traffic classification rule supported by convolutional neural networks (CNN). By mapping the network traffic detection characteristics to pixels, the network detection image is generated, and also the image to employed as the input of the CNN to appreciate traffic classification.

S. Bhattacharya and P. K. R. Maddikunta Proposed an intrusion detection set of rules primarily based totally on long-time period memory (LSTM), which detects DoS assaults related to probe assaults with one-of-a-kind facts inside the KDD dataset. Kwon et al. offer applicable deep gaining knowledge of model analysis, specialising in data simplification, dimension reduction, classification, and different technologies, and propose a very convolutional network (FCN) model. By scrutiny with the quality machine learning technology, it' verified that the FCN model is helpful for network traffic analysis. D. Kwon projected an anomaly-based IDS supported by a two-stage of the meta-classifier that uses a hybrid feature selection technique to urge correct feature representations. They conducted tests on the projected methodology on the NSL-KDD and UNSW-NB15 intrusion datasets and improved detection rates.[1]

Class equalization strategies within the field of machine learning, the matter of sophistication imbalance network traffic has constantly been a challenge. Therefore, intrusion detection faces large challenges in the network traffic detection system with extraordinarily unbalanced categories. Therefore, several students have begun to review the ways to improve the intrusion recognition accuracy of unbalanced network traffic knowledge. Abdulhammed et al. wear down the unbalanced dataset with KDD exploitation information up sampling and down sampling ways and by Deep Neural Networks, Random Forest, Voting, Variational Autoencoder, and Stacking Machine Learning classifiers to the datasets. In their projected method, the accuracy will reach 99.99%. Recently, Chuang and Wn trained the depth automatic encoder to establish AN data generation model to induce low-cost knowledge needed to create a balanced knowledge set.[2]

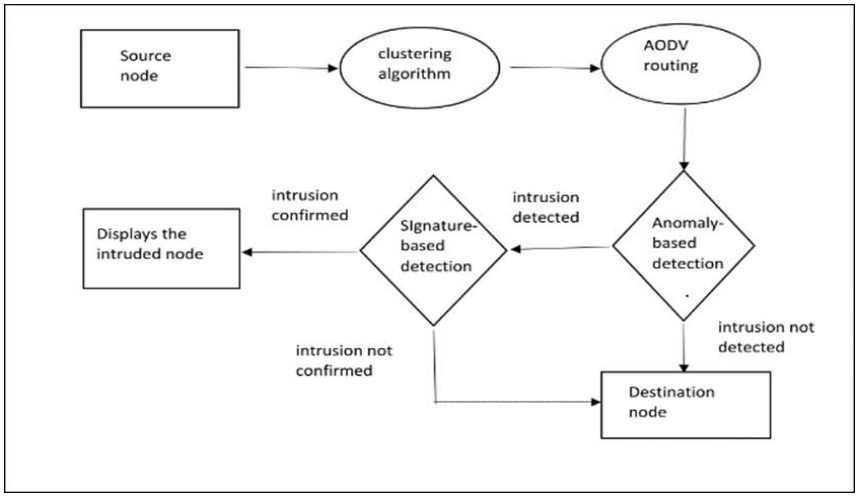


Figure 1 Architecture of the Proposed System

Piyasak planned a way to spice up the accuracy of minority categorification. This method combines the synthetic Minority Over-sampling Technique (SMOTE) associated Complementary Neural Network (CMTNN) to imbalanced information categorification. Experiments on the UCI dataset show that the proposed combination technique can improve class imbalance problems. Yan proposed an improved native adaptative composite minority sampling algorithmic programmed (LA-SMOTE) to subsume the network traffic imbalance downside thus supported by the deep learning GRU neural network to get the network traffic anomaly.[4]

Proposed system

In the proposed solution, a supervised machine learning model was trained with KDD datasets using SVM algorithm to detect intrusion. This model had given a target by observing and learning from the KDD dataset. The given target will be compared with signature rules (A set of known threats) and determined whether it is an intrusion or not. The process of sending packet from source node to destination node is done with the help of clustering algorithm where AODV is used for finding the best route with less energy cost. While transferring data from one node to another if an intrusion is found the model will stop the transfer and alert the user about where the intrusion is detected.

Proposed Algorithm

A cluster of nodes is created to manage high number of tasks with each cluster containing a cluster head to manage the task inside the cluster. AODV routing used for efficient packet transmission. Hybrid model is used to classify whether an intrusion is happened or not.

Step 1 cluster formation

A cluster of nodes was created to distribute the work load among the network to make use of the limited energy source efficiently with each group containing a

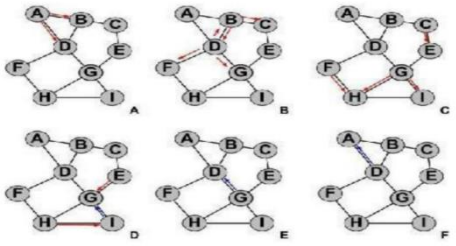
cluster head to manage the tasks efficiently without overloading a single node with too much tasks.

- BS obtains information about the location of all nodes in the network in the first round. It splits the network into 10 zones based on the density and geographical architecture of the network. The goal of this strategy is to ensure that CHs are selected consistently throughout the network's structure
- Because it has been assumed that all nodes have the same maximum energy (Emax) from the start, nodes in each zone have a probability of becoming a CH of p (1/number of nodes in the zone). As a result, a cluster head (CH) is chosen at random from each zone
- Once the CHs are formed, they broadcast their identities to the rest of the network, requesting that they accept their joining request and form true clusters. CH7 and CH12, for example, advertise their identities to all nodes in the network
- The signal strength of the request signal is analyzed by the nodes that receive the joining request. The CH request's signal intensity is determined by the distance between the CH and the node, as well as the physical barrier between the CH and the node. Each node delivers an acknowledgement to the most desired CH based on the strength of the signal. Each CH awaits a joining request from the nodes in its vicinity.
- The CH creates and distributes the data sending schedule to the cluster's members.
- Each node provides data to the CH, which the CH compresses and sends to the BS.

Step 2 AODV routing to find best route

AODV algorithm was used to find the best route among the clusters to manage the energy usage efficiently. So that, the packet transfer will be held with the shortest and most efficient route without any wastage of energy and it also helps against route looping.

Figure 2 Route Request



- I require a packet from A.
- A packet called a 'ROUTE REQUEST' will be created by A and delivered to B and D. (a). B and D add A as a reverse route to their routing table and forward the ROUTE REQUEST packet to their neighbors (b).

- B and D were unconcerned about the packet and instead exchanged each other's (as duplicates). While no route is known, the forwarding process continues (c). I create the ROUTE REPLY packet and transmits it to the node from which it received the ROUTE REQUEST from G (d).
- Duplicate packets are still discarded, while the ROUTE REPLY packet uses already established reverse routes to get to A in the lowest time possible (e and f).
- After a time, the reverse routes built by other nodes that were not used for the ROUTE REPLY are erased. Once G and D get the ROUTE REPLY packet, they will add the route to I.

Step 3 Intrusion detection

Hybrid model was used to classify the data have been intruded or not with the help of anomaly-based and signature-based intrusion detection. The detection of attack is done by the following algorithm.

- Two tail times can be directly applied to one tail time in this module. As a result, only the former is taken into account. Separate the two tail times because the scheduling rates are different in these two periods. For online determination of whether it is now tail time, two mechanisms are used.
- The current RRC state is inferred based on power usage using a power-based state inference technique. The cornerstone for distinguishing between the two types of tail time is determining the current RRC state. Furthermore, a high-accuracy power-based state inference mechanism has been demonstrated. The error estimate of this technique is ignored because it can achieve high accuracy (greater than 95 percent).
- The virtual tail time is used to assess if it is currently tail time, and it correlates to the RNC's initial inactivity timers. The inactivity timers are reset after transmitting data in the tail time, causing the physical tail time to be broken. The virtual tail time is the term for the used tail time.
- In the current RRC condition, a timer is required to detect whether it is virtual tail time. The virtual tail timer performs activities that are comparable to those done by the RNC's inactivity timer. The virtual tail times of the DCH and ACK states, which are designated as and, respectively, are represented by two timers.
- Similar to the inactivity timer α, the virtual tail timer γ is activated when the throughput is 0 or under the configured threshold.
- If timer γ is activated when the throughput is 0, Random cluster can start transmitting data after the timer γ is activated and stop when the timer γ expires or is reset.
- If timer γ is activated when the throughput is under the configured threshold but greater than 0, Random cluster cannot transmit data after the timer γ is activated.
- If Random cluster transmits data under this condition, the transmission of real-time data may be ongoing when the timer γ expires, and demotion at this time would trigger additional state promotions. Thus, having no transmission at the second condition would not reset the inactivity timer α, and the state is demoted to the ACK state after the expiry of timer α. When

in the ACK state, the virtual tail timer δ would be activated only when the throughput is 0. Random cluster can start transmitting data after timer δ is activated and stop when the timer δ expires or is reset.

Creating clusters

Cluster algorithm was used to create clusters. Cluster is nothing but making individual node into a group of nodes. Each cluster has a cluster head which gives instruction to the nodes in the cluster. The cluster heads will be chosen randomly, so that every node can get a chance to become a cluster head. The nodes will be distributed evenly to all the clusters such that all the clusters have equal number of nodes. It is done to help the cluster head, so the cluster heads don't have too much receiving and transmitting. The base station is far from the sensors and it is immobile. All nodes are able to send data to the base station. The base station has the information about all the nodes. Clustering algorithm is used to maintain energy consumption as less as possible and it will easily adapt to new examples.

AODV (Ad-hoc On-demand Distance vector) routing

AODV was used for routing. It only builds route between nodes if only asked by the source node, so it is considered as On-demand algorithm. It adds a new attribute, sequence number to each route, so the routing loops will not occur while finding for the best root. To keep the routing table consistent, each node in the ad hoc network updates it with advertisements on a regular basis or whenever important new information comes. Routing table is nothing but a table which contains information about the packets route and destination. It finds the best route where the energy consumption is less. The routing table will be updated periodically because the nodes are dynamical, so they will change the position.

Intrusion Detection

Anomaly-based Detection

The intrusion detection was done by the supervised model that is trained with KDD datasets using SVM algorithm. The developed model was considered as hybrid model because it uses both Signature-based IDS detection and Anomaly Detection. The model builds a normal behavior by learning from the datasets. The model monitored the entire network and compared it with the normal behavior. It was useful to find an unknown threat. If an abnormality is found it will be marked as target and it will be compared with the signature rules.

Signature-based Detection

The signature-based Intrusion Detection System (IDS) detects attacks on the basis of specific pattern such as number of bytes or number of 1's and 0's, it also detects on the basis of known intrusion instruction sequence. If there are such patterns found, it will be an intrusion because it is a pattern only found during intrusion but it is at a disadvantage because this detection cannot find any new type of attacks. If the given target from the anomaly-based detection is confirmed

as a attack then the packet transferring process will stop and give the detail of the node which have been intruded else, the target is not considered as intrusion and the process of packet transmission will proceed further

Experimental Results

The proposed system "Intrusion Detection in Health Care Domain using Machine Learning" was implemented with KDD dataset for identification of nodes. KDD training dataset consists of approximately many records single connection vector each of feature and is labelled as either normal attack or an attack, with exactly one specific attack type. This dataset is used to detect and identify the intruder node of the simulation. Experiments show that the proposed method can accurately determine the sample that need to be expanded in the imbalanced network traffic and improve the attack recognition more effectively and it's been implemented.

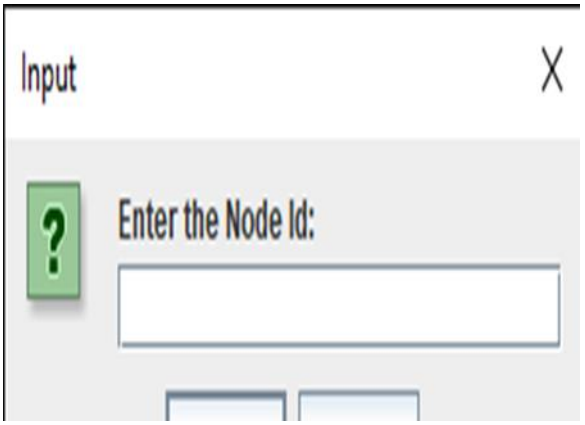


Figure 3 Node Creation

Figure 3 represents the input of an denial of service attack detection based on multivariate correlation analysis on network controller.Here the Node Id can be entered to find the analysis of an attack.

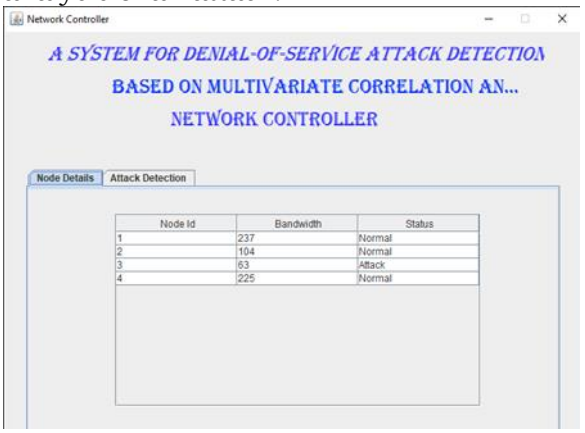


Figure 4 Node details

Figure 4 represents the node details of an system for denial of service attack detection based on multivariate correlation analysis of an Network Controller. The Normal and Attack nodes are mentioned here.This shows the attack status of a bandwidth.

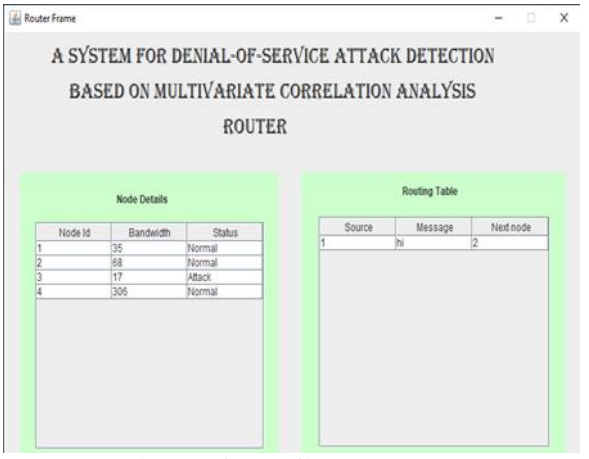


Figure 5 Sending Message

Figure 5 represents the Routing Table details of an system for Denial of Service attack detection based on multivariate correlation analysis router. Attack Node sent to third node so by the time the Node stops at the second Node**.**The data will be blocked before which is transmitted to the attacker node. Routing table shows the attack message with Node ID.

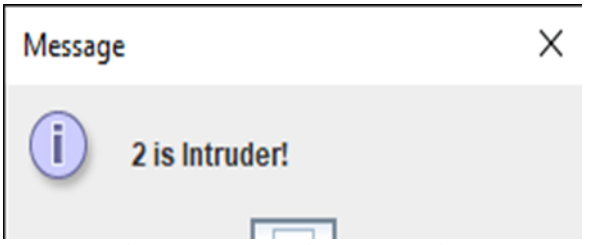


Figure 6 Intrusion Detection

Figure 6 shows the final result of the analysis. It shows the intruder on which node, for our convenience. By using KDD dataset, the Intruder is identified.

Conclusion

In today's modern world, cyber-attacks are very commonly theft the data from computer. The proposed system had simulated the network intrusion detection in the cyberspace and demonstrated that the pressure in network intrusion detection was also increasing. The experimented intrusion detection datasets were imbalanced due to natural difference between the number of intrusive and normal samples. DSSTE was applied on NSL-KDD to increase the number of intrusive samples and make the dataset balanced. The effectiveness of KDD CUP 99 datasets was evaluated by running SVM, decision tree, and KDD dataset before

and after using AODV Algorithm. Experiments using various types of attacks and one-vs-one classification were conducted in this study. When compared with existing system, the proposed model requires less energy to detect attacks. Each IDS agents relies on a policy that increases reliable packet transmission.

References

1. S. Bhattacharya, P. K. R. Maddikunta, R. Kaluri, S. Singh, T. R. Gadekallu, M. Alazab, et al., "A novel PCA-firefly based XGBoost classification model for intrusion detection in networks using GPU", *Electronics*, vol. 9, no. 2, pp. 219, Jan. 2020.
2. R. Abdulhammed, M. Faezipour, A. Abuzneid, and A. AbuMallouh, "Deep and machine learning approaches for anomaly-based intrusion detection of imbalanced network traffic," IEEE sensors Lett., vol. 3, no. 1, Jan. 2019, Art. no. 7101404.
3. Sharafaldin, I, Lashkari,A.H and Ghorbani, A.A, "Toward Generating a New Intrusion Detection Dataset and Intrusion Traffic Characterization", 4th International Conference on Information Systems Security and Privacy (ICISSP), Purtogal, (2018).
4. W. Wang, M. Zhu, X. Zeng, X. Ye and Y. Sheng, "Malware traffic classification using convolutional neural network for representation learning", *Proc. Int. Conf. Inf. Netw. (ICOIN)*, pp. 712-717, 2017.
5. Gil, G.D., Lashkari, A.H., Mamun, M. and Ghorbani, A.A., "Characterization of encrypted and VPN traffic using time-related features. In Proceedings of the 2nd International Conference on Information Systems Security and Privacy, pp. 407-414, (2016).
6. Moustafa, N. and Slay, J., "The evaluation of Network Anomaly Detection Systems: Statistical analysis of the UNSW-NB15 data set and the comparison with the KDD99 dataset". Information Security Journal: A Global Perspective, 25(1-3), pp.18-31, (2016).
7. Moustafa, N. and Slay, J., "UNSW-NB15: a comprehensive data set for network intrusion detection systems (UNSW-NB15 network data set). IEEE Military Communications and Information Systems Conference (MilCIS), pp. 1-6, (2015).
8. Pongle, Pavan, and Gurunath Chavan. "A survey: Attacks on RPL and 6LoWPAN in IoT." IEEE International Conference on Pervasive Computing, (2015).
9. Oh, Doohwan, Deokho Kim, and Won Woo R, "A malicious pattern detection engine for embedded security systems in the Internet of Things." Sensors, pp, 24188-24211, (2014).
10. Mangrulkar, N.S., Patil, A.R.B. and Pande, A.S., "Network Attacks and Their Detection Mechanisms: A Review". International Journal of Computer Applications, 90(9), (2014).
11. Suryasa, I. W., Rodríguez-Gámez, M., & Koldoris, T. (2021). The COVID-19 pandemic. *International Journal of Health Sciences*, *5*(2), vi-ix. https://doi.org/10.53730/ijhs.v5n2.2937
12. Suryasa, I. W., Rodríguez-Gámez, M., & Koldoris, T. (2022). Post-pandemic health and its sustainability: Educational situation. *International Journal of Health Sciences*, *6*(1), i-v. https://doi.org/10.53730/ijhs.v6n1.5949

13. Susilo, C. B., Jayanto, I., & Kusumawaty, I. (2021). Understanding digital technology trends in healthcare and preventive strategy. *International Journal of Health & Medical Sciences*, *4*(3), 347-354. https://doi.org/10.31295/ijhms.v4n3.1769