www.ijprems.com
editor@ijprems.com

INTERNATIONAL JOURNAL OF PROGRESSIVE RESEARCH IN ENGINEERING MANAGEMENT AND SCIENCE (IJPREMS)
(Int Peer Reviewed Journal)
Vol. 05, Issue 01, January 2025, pp : 1691-1699

e-ISSN : 2583-1062
Impact Factor : 7.001

PROACTIVE NEWTORK DEFENSE

Srisudha Garugu[1], P. Vamsi Vihari[2], M. Pradeep Reddy[3], R. Chethan[4]

[1,2,3,4]Institute/Organization: ACE Engineering College

ABSTRACT

The speedy development of the World Wide Web and the wild stream of arrange activity have come about in a ceaseless increment of organize security dangers. Cyber aggressors look for to abuse vulnerabilities in arrange design to take important data or disturb computer assets. Organize Interruption Discovery Framework (NIDS) is utilized to viably identify different assaults, hence giving opportune assurance to arrange assets from these assaults. To execute NIDS, a stream of administered and unsupervised machine learning approaches is connected to identify inconsistencies in organize activity and to address arrange security issues. Such NIDSs are prepared utilizing different datasets that incorporate assault follows. Be that as it may, due to the headway in modern-day assaults, these frameworks are incapable to distinguish the rising dangers. Hence, NIDS needs to be prepared and created with a present day comprehensive dataset which contains modern common and assault exercises. This paper presents a system in which distinctive machine learning classification plans are utilized to identify different sorts of organize assault categories. Five machine learning calculations: Irregular Timberland, Choice Tree, Calculated Relapse, K-Nearest Neighbors and Fake Neural Systems, are utilized for assault location. This think about employments a dataset distributed by the College of Modern South Grains, a generally unused dataset that contains a expansive sum of organize activity information with nine categories of arrange assaults. The comes about appear that the classification models accomplished the most noteworthy precision of 89.29% by applying the Irregular Timberland calculation.

Keywords – Machine learning, neural networks, Network intrusion detection, NIDS, Artificial Intelligence .

1. INTRODUCTION

In order to tackle threats brought on by an expanding number of sophisticated hackers in a digital environment, the next generation of intrusion detection systems (IDSs) demand network intrusion detection methods that are intelligent as well as automated, and . Specifically, there has been a high demand for automated IDS strategies that are agent-based and need as little involvement of humans as possible while still having the ability to improve and evolve (for example, by taking the right actions at the right time in the environment given) and to be more resilient and robust to unforeseen and potential threats or difficulties that were not seen by it before.

Using Machine Learning (ML) algorithms has gained popularity as a method for identifying and categorizing various attacks and cyber threats.

The following is a summary of our work's significant contributions: The modern methods are presented for detecting network intrusions that integrate deep feed-forward neural networks which is based on Machine learning. Our suggested model has the ability to continuously learn how to interact in a network environment so that it can identify various kinds of network invasions. Our model's selflearning capabilities enable it to continuously improve its detecting abilities. the tasks related to classification involved in many network intrusion classes.

2. LITERATURE SURVEY

Moustafa & Kill (2015) created a demonstrate that centered on the classification of assault families accessible in the UNSW-NB15 dataset.

The think about utilized the Affiliation Run the show Mining procedure for highlight selection.

For classification, Expectation–Maximization (EM) calculation and NB have been utilized.

However, the exactness of both calculations for recognizing uncommon assaults was not essentially higher as the Naïve Bayes had an exactness of 78.06% and the precision of EM was 58.88%.

Moustafa & Kill (2016) advance expanded their work in 2016 and utilized relationship coefficient and pick up proportion for highlight choice in their work. From that point, five classification calculations of NB, DT, ANN, LR and EM were utilized on the UNSW-NB15. Comes about appeared that 85% exactness was accomplished utilizing DT with 15.75 Wrong Alert Rate (Distant). This investigate utilized a subset of UNSW-NB15; in any case, discovery precision was not satisfactory.

For identifying botnets and their tracks, Koroniotis et al. (2017) displayed a system utilizing machine learning strategies on a subset of the UNSW-NB15 dataset utilizing organize stream identifiers. Four classification calculations were utilized i.e., Affiliation Run the show Mining (ARM), ANN, NB and DT. The comes about appeared that the DT gotten the most noteworthy precision of 93.23% with a Untrue Positive Rate (FPR) of 6.77%.

In 2019, Meftah, Rachidi & Assem (2019) connected a two-stage anomaly-based NIDS approach to identify arrange assaults. The proposed strategy utilized LR, Slope Boost Machine (GBM) and Bolster Vector Machine (SVM) with the Recursive Highlight Disposal (RFE) and RF highlight choice procedures on a total UNSW-NB15 dataset. The comes about appeared that the exactness of multi-classifiers utilizing DT was roughly 86.04%, respectively.

Kumar et al. (2020) proposed an coordinates calcification-based NIDS utilizing DT models with a combination of clusters made utilizing the k-mean calculation and IG's highlight determination method. The inquire about utilized as it were 22 highlights and four sorts of organize assaults of UNSW-NB15 dataset, and the RTNITP18 dataset, which served as a test dataset to test the execution of the proposed show.

The result appeared an exactness of 84.83% utilizing the proposed demonstrate and 90.74% utilizing the C5 show of DT.

Kasongo & Sun (2020) displayed the NIDS approach utilizing five classification calculations of LR, KNN, ANN, DT and SVM in conjunction with the include determination procedure of the XGBoost calculation. The inquire about utilized the UNSW-NB15 dataset to apply twofold and multiclass classification strategies. In spite of the fact that double classification performed well with an precision of 96.76% utilizing the KNN classifier, multiclass classification didn't perform well as it accomplished the most noteworthy precision of 82.66%.

Kumar, Das & Sinha (2021) proposed Bound together Interruption Discovery Framework (UIDS) to identify typical activity and four sorts of arrange assault categories by utilizing UNSW-NB15 dataset. The proposed UIDS show was planned with the set of rules (R) inferred from different DT models counting k-means clustering and IG's include determination method. In expansion, different calculations such as C5, Neural Organize and SVM were moreover utilized to prepare the show. As a result, the proposed demonstrate moved forward with an precision of 88.92% over other approaches. In any case, other calculations such as C5, Neural Organize and SVM accomplished an exactness of 89.76%, 86.7% and 78.77%, respectively.

It is apparent that more work needs to be done to distinguish the highlights for the families of arrange assaults. There is a require to decide a non specific demonstrate that gives superior precision for all the assaults displayed in the dataset. This inquire about gives a show that decides a common subset of highlights. Hence, by utilizing that highlight subset we would be able to distinguish all assaults, having a place to any category with steady precision. It centers on the execution of a non, specific demonstrate that gives moved forward classification precision. In addition, there is restricted inquire about that has utilized the lesson awkwardness method to adjust occurrences of uncommon assaults display in the dataset.

Reference	Dataset (complete/partial)	Algorithms	Accuracy/ FAR/FPR	Limitations
Moustafa & Slay (2015)	UNSW-NB15 (Partial) KDD99 Dataset	Naïve Bayes and EM Algorithm	Accuracy: Naïve Bayes– 37.4% EM Algorithm: 75.81% FPR: 22.07	This inquire about is deciding as it were five organize assault categories. The issue of course lopsidedness has not been settled. As a result, the calculations are not performing well
Moustafa & Slay (2016)	UNSW-NB15 (Partial) Dataset	Naïve Bayes, decision tree, artificial neural network, logistic regression, and expectation-maximisation	Accuracy: between 78.46% to 85.57% FAR: between 15.76% to 23.78%	In this think about, information pre-processing methods have not been actualized and the issue of lesson awkwardness has not been settled.
Koroniotis et al. (2017)	UNSW-NB15 Dataset	Naive Bayes, decision tree, association rule mining (ARM)	Accuracy: between 63.98% to 93.21%	This work didn't unravel the issue of lesson awkwardness. Subsequently , the calculations did not perform well to identify a few arrangement assaults.

Reference	Dataset (complete/partial)	Algorithms	Accuracy/ FAR/FPR	Limitations
		and artificial neural network	FPR: between 6.76% to 36.03%	
Meftah, Rachidi & Assem (2019)	UNSW-NB15 Dataset	Logistic regression, gradient boost machine, and support vector machine	Accuracy: achieving a multi-classification accuracy of 86.04%	This research didn't address the class imbalance problem.
Kumar et al. (2020)	UNSW-NB15 (Partial) and RTNITP18 Dataset	Decision tree models (C5, CHAID, CART, QUEST)	Accuracy: 84.83% using proposed model and 90.74% using C5 model	This work has anticipated as it were 4 out of 9 categories of the UNSW-NB15 dataset. Moreover, the issue of lesson awkwardness has not been unraveled in this consider.
Kasongo & Sun (2020)	UNSW-NB15 Dataset	Logistic regression, K-nearest neighbors, artificial neural network, decision tree and support vector machine	Accuracy: between 53.43% to 82.66% using multiclass classification scheme	The issue of lesson awkwardness has not been settled in this investigate. Thus, the demonstrate has not accomplished great exactness.
Kumar, Das & Sinha (2021)	UNSW-NB15 (Partial) Dataset	Decision tree models (C5, CHAID, CART, QUEST)	Accuracy: 82.92% using proposed model	Investigate has anticipated as it were 4 sorts of organize assault categories of the UNSW-NB15 dataset. In expansion, the issue of lesson awkwardness has not been settled in this inquire about.

3. PROPOSED SYSTEM

The system utilizes a subset of the NIDSdataset. It comprises of two primary steps. The to begin with step includes information pre-processing, in which standardization and normalization of information are performed. Due to the tall dimensional nature of the dataset, a few highlights that are unimportant or repetitive may lead to decrease the exactness of assault location. To illuminate this issue, highlight choice is utilized, in which as it were the pertinent subset of highlights is chosen to dispense with futile and boisterous highlights from multidimensional datasets. A short time later, we have at that point tended to the lesson awkwardness issue. In the another step, distinctive classifiers are prepared with important highlights to distinguish all categories of assault to get most extreme precision. At last, precision, accuracy, review and F1-score execution measures are utilized to assess the show. The proposed technique that speaks to the by and large system is appeared in Fig.

www.ijprems.com
editor@ijprems.com

INTERNATIONAL JOURNAL OF PROGRESSIVE RESEARCH IN ENGINEERING MANAGEMENT AND SCIENCE (IJPREMS)
(Int Peer Reviewed Journal)
Vol. 05, Issue 01, January 2025, pp : 1691-1699

e-ISSN : 2583-1062
Impact Factor : 7.001

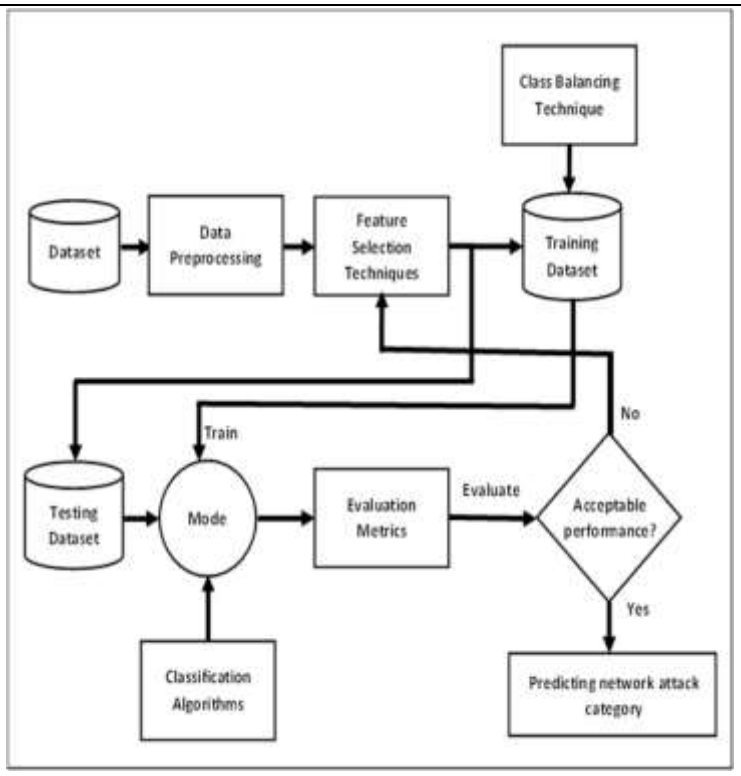


Fig.1

Technique For Model Evaluation

1) Dataset pre-processing

This stage includes the taking after steps: information standardization and information normalization.

Data standardization

As there are highlights with diverse ranges of values in the dataset we performed information standardization to change over the information from ordinary dissemination into standard typical dispersion. Hence, after rescaling, a cruel esteem of an property is break even with to 0 and the coming about dispersion is break even with to the standard deviation. The equation to calculate a standard score (z-score) is:

$z=(x-\mu)\sigma z=(x-\mu)\sigma$

where x is the information test, μ is the cruel and σ is the standard deviation.

2) Data normalization

In information normalization, the esteem of each persistent quality is scaled between 0 and 1 such that the result of traits does not rule each other.

3) Feature selection

Feature determination is a strategy that is utilized to select highlights that for the most part connect and contribute to the target variable of the dataset In this investigate, highlight choice is done utilizing Relationship Quality Assessment (CA), Data Pick up (IG) and Central Component Examination (PCA). CA measures the relationship between each include with the target variable and select as it were those relative highlights that have tolerably higher positive or negative values, i.e., closer to 1 or −1 .

4) Class imbalance

The NIDS dataset is profoundly imbalanced not as it were since the number of ordinary activity occurrences is much higher than diverse assault categories, but moreover since the distinctive categories of assault occurrences are not break even with in dispersion. This issue is known as “Class Imbalance”.

Classification algorithms

Five classification calculations, that is, RF, DT, LR, KNN and ANN were utilized to prepare the model.

Random forest

Random Timberland is an outfit classifier that is utilized for progressing classification comes about. It comprises different Choice Trees. In comparison with other classifiers, RF gives lower classification blunders. Randomization is connected for the determination of the best hubs for part when making partitioned trees in RF (Jiang et al., 2018).

www.ijprems.com
editor@ijprems.com
INTERNATIONAL JOURNAL OF PROGRESSIVE RESEARCH IN ENGINEERING MANAGEMENT AND SCIENCE (IJPREMS)
(Int Peer Reviewed Journal)
Vol. 05, Issue 01, January 2025, pp : 1691-1699
e-ISSN : 2583-1062
Impact Factor : 7.001

Decision tree

In the Choice Tree calculation, the properties are tried on inner hubs, the results of the tests are spoken to by branches, and leaf hubs hold names of the classes (Afraei, Shahriar & Madani, 2019). Quality determination strategies are utilized for recognizing hubs. Those chosen qualities minimize the data that is required for tuple classification in the brought about segment. Thus, reflecting the least instability or debasement in those allotments. In this manner, minimizing the anticipated number of tests required for tuple classification. In this investigate, ID3 calculations utilize entropy course to decide which properties ought to be questioned on, at each hub of those choice trees.

Logistic Regression

Logistic Relapse is a probabilistic classification show. It casts the issue into a generalized straight relapse shape. It has a sigmoid bend. The condition of the sigmoid work or calculated work is:

S(x)=ex1+exS(x)=ex1+ex

This work is utilized for mapping values to probabilities. It works by mapping genuine values to other values between 0 and 1 (Kyurkchiev & Markov, 2016).

K-nearest neighbors

In K-Nearest Neighbors, a unused information point is connected with the information focuses in the preparing set and based on that connection, a esteem is alloted to that unused information point. This employments highlight similitude for expectation. In KNN Euclidean, Manhattan or Hamming separate are utilized for calculating the separate between a test information and each record of preparing information (Jain, Jain & Vishwakarma, 2020). A while later, agreeing to the esteem of remove, the lines are sorted. From those lines, K lines from the best are chosen. Based on the most visit classes of those lines, classes to the test focuses are assigned.

Artificial neural network

In the Fake Neural Organize calculation, there are three layers that comprise of computational units called neurons. These layers are input, yield and covered up layers. The number of neurons in these layers depends on the highlights of the dataset and classes which have to be identified and chosen with distinctive methods. Distinctive sorts of actuation capacities are utilized in the ANN calculation for calculating the weighted entirety of the associations between neurons. This calculation has predispositions in the covered up layer and an yield layer which are balanced to diminish blunders and progress precision in preparing and testing the show (Andropov et al., 2017).

Evaluation metrics

A perplexity lattice is utilized for the comparison of the execution of machine learning calculations. This network is utilized for the creation of distinctive measurements by the combination of the values of Genuine Negative (TN), Genuine Positive (TP), Wrong Negative (FN) and Untrue Positive (FP) (Tripathy, Agrawal & Rath, 2016). Underneath are a few of the execution measures to assess models by the utilize of the disarray matrix.

Accuracy appears the rightness or closeness of the approximated esteem to the real or genuine esteem of the demonstrate which implies a parcel of the add up to tests that are classified accurately (Lin, Ye & Xu, 2019). The taking after equation is utilized to calculate the precision of the model:

Accuracy=(TP+TN)/(TP+TN+FN+FP)

Precision appears which parcel of significant occurrences is really positive among the chosen occasions (Roy & Cheung, 2018). The taking after equation is utilized to calculate precision:

Precision=TPTP+FP

Recall or Genuine Positive Rate (TPR) calculates the division of genuine positives that are accurately recognized (Ludwig, 2017). The equation utilized to discover review is:

Recall=TP/TP+FN

F1-score is deciphered as the consonant cruel of exactness and review implies it combines the weighted normal of accuracy and review (Javaid et al., 2016). The taking after equation is utilized to calculate F1-score:

F1=2 ∗ (Accuracy ∗ Recall/Precision+Recall)

Method	Feature selection techniques	Accuracy%	Precision%	Recall%	F1-Score%
Random forest	Information gain attribute	83.5	76.8	72.3	73.7

www.ijprems.com
editor@ijprems.com

INTERNATIONAL JOURNAL OF PROGRESSIVE RESEARCH IN ENGINEERING MANAGEMENT AND SCIENCE (IJPREMS)
(Int Peer Reviewed Journal)
Vol. 05, Issue 01, January 2025, pp : 1691-1699

e-ISSN : 2583-1062
Impact Factor : 7.001

Method	Feature selection techniques	Accuracy%	Precision%	Recall%	F1-Score%
	Correlation attribute	82.3	72.2	59.6	61.9
	Principal component analysis	82.3	77.3	70.8	73.1
Decision tree	Information gain attribute	80.5	69.6	72.0	70.7
	Correlation attribute	80.8	63.3	60.9	61.9
	Principal component analysis	80.4	70.9	67.3	69.3
Logistic regression	Information gain attribute	82.2	51.2	42.3	41.9
	Correlation attribute	74.5	38.5	36.0	35.4
	Principal component analysis	80.4	51.3	40.6	40.0
K-nearest neighbor	Information gain attribute	82.7	54.0	48.1	49.6
	Correlation attribute	76.8	46.3	42.6	43.5
	Principal component analysis	81.0	57.8	51.3	53.3
Artificial neural network	Information gain attribute	81.7	60.6	54.6	54.2
	Correlation attribute	80.3	50.1	44.1	44.7
	Principal component analysis	81.2	61.2	52.4	54.4

Execution investigation of classification models

The comes about of classification models, in which Irregular Timberland accomplished the most noteworthy precision, accuracy, review, and F1-score values around 83.5%, 76.9%, 72.6% and 74.1% individually. While, LR gotten the least exactness, accuracy, review, and F1-score values of around 82.07%, 50.5%, 42.6% and 42.3%. DT accomplished an precision of 88.29% with 70.5% exactness, 72.7% review and 71.4% F1-score. KNN moreover recorded the same

INTERNATIONAL JOURNAL OF PROGRESSIVE RESEARCH IN ENGINEERING MANAGEMENT AND SCIENCE (IJPREMS)
(Int Peer Reviewed Journal)
Vol. 05, Issue 01, January 2025, pp : 1691-1699
www.ijprems.com
editor@ijprems.com
e-ISSN : 2583-1062
Impact Factor : 7.001

precision as LR but with superior accuracy, review and F1-score comes about. Though, ANN recorded normal execution with precision, accuracy, review, and F1-score values of around 81.7 0%, 61.0%, 53.2% and 54.5%.

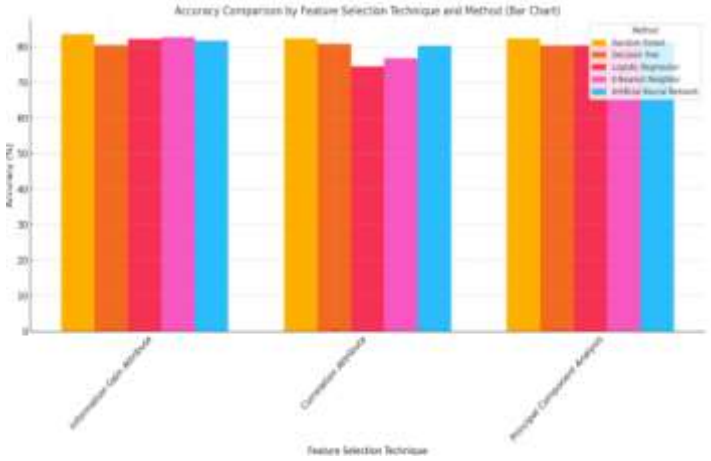


Fig.2

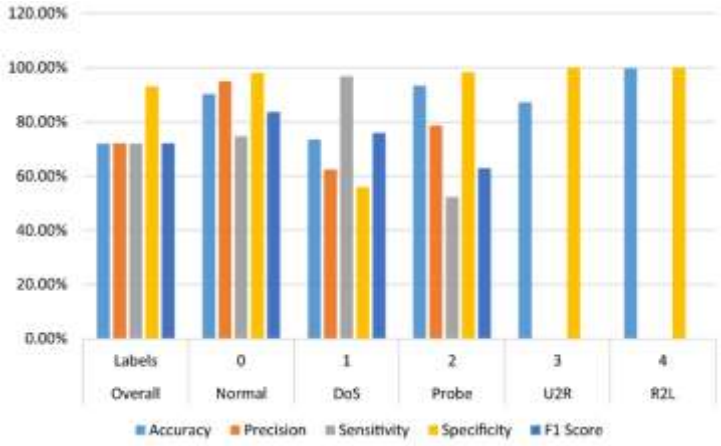


Fig.3 Random forest Accuracy for each

input

Features and Implementation

1. User Interface

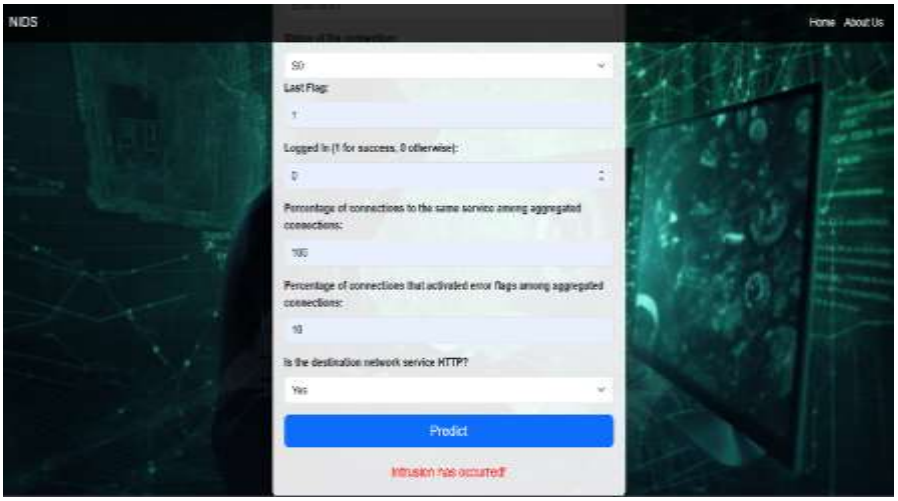
   The NIDS interface is built using HTML, CSS, etc.. It offers the following components: Home Page: Features options like login and registration

www.ijprems.com
editor@ijprems.com

INTERNATIONAL JOURNAL OF PROGRESSIVE RESEARCH IN ENGINEERING MANAGEMENT AND SCIENCE (IJPREMS)
(Int Peer Reviewed Journal)
Vol. 05, Issue 01, January 2025, pp : 1691-1699

e-ISSN : 2583-1062
Impact Factor : 7.001

After entering the packet details the the website will be able to predict whether it is a intrusion or not.



4. CONCLUSION

IDS is a crucial component of contemporary data networks. With a model that applies the RL framework to the IDS problem, this work aims to offer a fresh option. The RL algorithms have demonstrated exceptional performance in a variety of fields (such as robotics, banking, gaming, and business operations), and we can demonstrate that they may also be applicable to IDS. The suggested novel model (DQN) combines reinforcement learning with Q learning to create a simulated environment that adheres to RL environment rules. The contributions of this paper are: Introducing an innovative architecture that is based on a fusion of adversarial RL, making it an ideal replacement for prediction issues in extremely demanding networks (like IoT networks). Proving the new model's prediction performance is comparable to that of highly non- linear models while boasting the amazing benefit of requiring substantially less computing time each prediction. With an emphasis on prediction performance, prediction timeframes, and model flexibility, a detailed comparison of proposed model with a number of Machine Learning (ML) models are provided . Introducing a model.

5. REFERENCES

[1] Manuel Lopez-Martin, Belen Carro, Antonio Sanchez- Esguevillas. “Application of deep Reinforcement learning to intrusion detection for supervised problems”. Expert Systems with Applications,

[2] Ying-Feng Hsu, Morito Matsuoka. “A Deep Reinforcement Learning Approach for Anomaly Network Intrusion Detection System”. 2020 IEE 9th International Conference on Cloud- Networking (Cloudnet) |

[3] Tongtong su, Huazhi sun, jinqi zhu, sheng wang, and yablo lo “BAT: Deep Learning Methods on Network-on-Network Intrusion. Detection Using NSL-KDD Dataset”. Received January 12, 2020, accepted January 31,2020, date of publication February 10, 2020. date of current version February 18, 2020.

[4] N. Saito, T. Oda, A. Hirata, Y. Hirota, M. Hirota, and K. Katayama, “Design and implementation of a dqn based aav,” in International Conference on Broadband and Wireless Computing, Communication and Applications.

[5] K. Sethi, R. Kumar, D. Mohanty, and P. Bera, “Robust adaptive cloud intrusion detection system using advanced deep reinforcement learning,” in International Conference on Security, Privacy, and Applied Cryptography Engineering.

[6] T. T. Nguyen and V. J. Reddi, "Deep reinforcement learning for cyber security

[7] K. Jiang, W. Wang, A. Wang, and H. Wu, "Network intrusion detection combined hybrid sampling with the deep hierarchical network.

[8] S. Iannucci, O. D. Barba, V. Cardellini, and I. Banicescu, "A performance evaluation of deep reinforcement learning for modelbased intrusion response," in 2019 IEEE 4th International Workshops on Foundations and Applications of Self* Systems (FAS* W).

[9] L. M. Ibrahim, D. T. Basheer, and M. S. Mahmod,"A comparison study for intrusion database (kdd99, nsl-kdd) based on selforganization map (som) artificial neural network," Journal of Engineering Science and Technology, vol. 8, no. 1, pp.107–119, 2013

[10] Gopi, A. P., & Naik, K. J. (2021, December). A model for analysis of IoT based aquarium water quality data using CNN model. In 2021 international conference on decisionaid sciences and application (DASA) (pp. 976-980). IEEE

[11] Arepalli, P. G., Akula, M., Kalli, R. S., Kolli, A., Popuri, V. P., & Chalichama, S. (2022, September). Water Quality Prediction for Salmon Fish Using Gated Recurrent Unit (GRU) Model. In 2022 Second International Conference on Computer Science, Engineering and Applications (ICCSEA) (pp. 1-5). IEEE.

[12] Kanumalli, Satya Sandeep, et al. "Automated Irrigation Management System using IoT." 2022 6th International Conference on Intelligent Computing and Control Systems (ICICCS). IEEE, 2022.

[13] Praveena, V., Vijayaraj, A., Chinnasamy ,P., Ali, I., Alroobaea, R., Alyahyan, S.Y. and Raza, M.A., 2022. Optimal deep reinforcement learning for intrusion detection in UAVs.

[14] Tharewal, S., Ashfaque, M. W., Banu, S. S., Uma, P., Hassen, S. M., & Shabaz, M. (2022). Intrusion detection system for industrial Internet of Things based on deep reinforcement learning. Wireless Communications and Mobile Computing, 2022.