Intrusion Detection System for Prediction Cyber Threats based on Machine Learning Techniques

Iqra Naseer, *Cyber Security IT Consultant, Doha, Qatar*

Abstract:

The increasing development of cyber threats has urged the establishment of efficient Intrusion Detection Systems that can predict and mitigate these attacks. With the ability to predict future threats and implement mechanisms, machine learning techniques offer excellent solutions in making IDS more intelligent, autonomous, and adaptive for future threats. This paper elaborates the application of multiple ML-based techniques to detect and predict cyber threats within IDS systems, analyzing the effectiveness and performance as well as discussing the limitation areas. Comparisons among supervised, unsupervised, and hybrid models were presented to show each's capability in enhancing accuracy with an IDS and reduction of response times.

Keywords: Intrusion Detection System (IDS), Cyber Threat Prediction, Machine Learning, Supervised Learning, Unsupervised Learning, Hybrid Models.

I. Introduction:

With advancements in technology and increasing reliance on more digital networks, the threats of cyber attacks against individuals, corporations, and governments have emerged with high frequency and sophistication[1]. Modern attacks have evolved to be adaptive and even complex in nature, no longer accessible to be countered by the old security measures such as firewalls and antivirus software. This has created a rising demand for the development of more sophisticated security solutions, such as Intrusion Detection Systems (IDSs), which track network traffic and alert users to potential intrusions. Yet traditional IDS's ability to adapt to new or previously unknown attacks (zero-day attacks) depends very much on the static rule set or known signatures. With this limitation, the integration of ML techniques into IDSs leads to prediction as well as adaptation that happens dynamically

toward emerging cyber threats[2]. The following Fig.1 depicts Intrusion Detection System for Prediction Cyber Threats

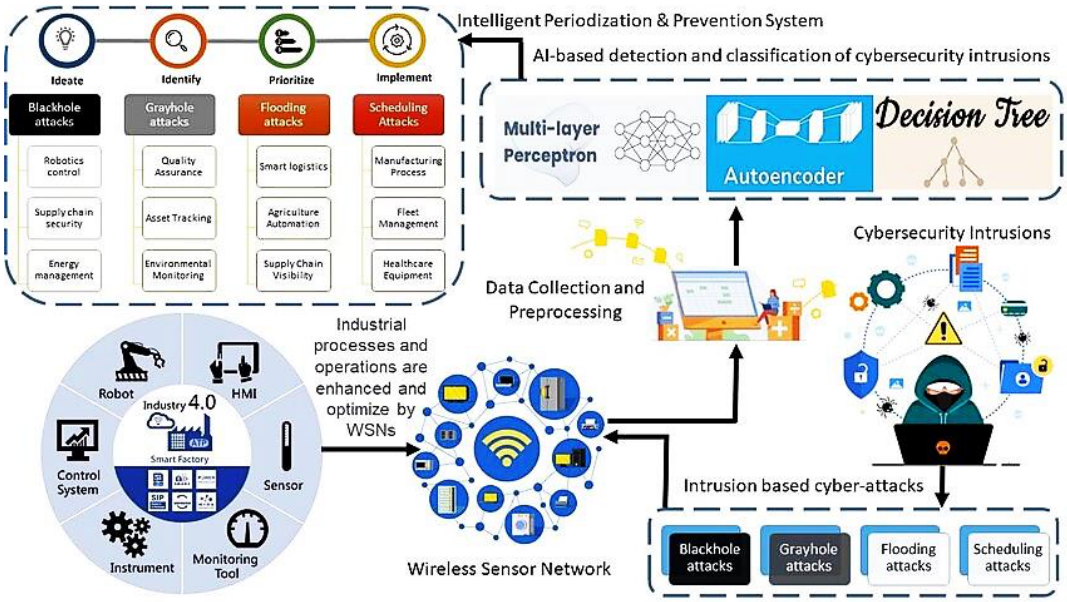


Fig.1: Intrusion Detection System for Prediction Cyber

Machine learning brings a transformable potential to IDS by providing them data-driven capabilities in the recognition of patterns of malicious activity. In contrast to traditional methods, ML-based IDSs can learn from humongous amounts of historic data and identify known as well as unknown threats through anomaly detection and pattern recognition. These systems reach high accuracy levels in recognizing threats, reducing false positives, and improving response time using a variety of ML algorithms, including supervised, unsupervised, and hybrid techniques. Supervised learning models could train on datasets and learn what distinctions exist between normal behaviors and malicious ones[3]. This is an advantage of unsupervised models to anomaly detection since the anomalies may, at times be unknown by the IDS.

Despite all the promise, the development and deployment of ML-enhanced IDSs is riddled with a lot of challenges. Data imbalance such that benign network traffic outstrips malicious activity collected into datasets skews the performance of the model. Additionally, complex algorithms bring in much computational overhead, thus challenging the real-time analysis and detection in high-traffic environments. . The optimal balance between minimizing false

positives and accurately detecting threats, which remains a crucial goal of this system, is obtained. High false-positive rates can lead to alert fatigue and reduced system efficiency.

The purpose of this paper is to report on the use of numerous ML techniques in IDS. Its focus is on showing effectiveness in predicting cyber threats, handling real-time data, and overcoming key hurdles, such as false positives and large computational requirements. We're going to compare the merits of supervised, unsupervised approaches, and hybrid approaches evaluated with benchmark datasets and according to their suitability for intruder detection scenarios[4]. This paper contributes toward developing intelligent and adaptive IDS solutions that may be able to protect the network against increasing sophisticated threats by summarizing the current state of research in ML-based IDS together with future directions.

II. Machine Learning Techniques for Intrusion Detection:

Machine learning (ML) simulations pave the way for tremendous developments in IDS properly centered to getting even more accurate and flexible in predicting and detecting cyber threats, as opposed to traditional methods. From historical data, the attackers create complex patterns of behavior known to signify malignant activity, which in turn enables them to distinguish between benign and malignant network behavior with great success. The paragraph discusses first the general categories of ML techniques, namely, supervised, unsupervised, and hybrid learning approaches; these categories are highlighted according to their advantages, limitations, and specific algorithms applicable in intrusion detection.

Supervised learning algorithms are widely used in IDS because they employ labeled datasets to learn patterns that can classify network traffic either as benign or malicious. Among popular supervised techniques are Decision Trees, Support Vector Machines (SVMs), and Neural Networks. Decision Trees are popular for their explanatory power and ease of implementation. In this example, each decision node partitions, or splits the instances based on the values of some decision feature (or attribute), forming a tree structure that makes the classification very clear and easy to mass-process. However, Decision Trees may suffer from overfitting problems, especially with complex or high-dimensional data, thus they sometimes need pruning or ensemble approaches for improved generalization. The SVM is a space-splitting algorithm mainly deployed for binary classification problems in IDS, where the objective is to differentiate between normal and anomalous traffic with the maximum precision possible. SVMs devise one hyperplane which maximizes the margin, which is

effective in high dimensional spaces between instances belonging to different classes[5]. However, SVMs are very sensitive to the imbalance of data-range, which is again a common functionality problem within the datasets pertaining to intrusion detection. These could, forever, be realized in the wake of advanced techniques and work to such an end; cost-sensitive learning or data resampling are examples of such techniques. Specifically, supervised deep learning models such as CNNs and RNNs work brilliantly in the context of intruder detection or cyber security in general. CNNs are more popularly used in classifying data, mostly in a spatial pattern over structured networks; RNNs still have a wider application with data presented with temporal sequences. Deep learning may, however, turn out to be resource-eating and could be beaten down as impractical unless some considerable computational resources are committed to, more significantly when real-time intrusion detection is probed for such.

Unsupervised learning techniques are very useful for detecting anomalies in IDSs, especially when labeled data is rare, or it comes to novel attacks that are much different from normal network behavior. Some of the common clustering algorithms utilized in this context are K-means and Principal Component Analysis (PCA) among many others. Clustering analysis, which groups data according to particular features without any prior knowledge about the class of the data. The outliers become detected as those observations which contain a relatively large distance from the centroid of the clusters. Although it is fast and, in many cases, theoretically consistent, how efficient K-means might prove fully depends on how well the number of clusters can be determined without prior knowledge concerning the distribution of data[6]. It is equally limited concerning the patterns of attack that are complicated and proved to be difficult when dealing with multidimensionality inside network dataset attacks.

Autoencoders, a kind of unsupervised neural networks, have gained preference in IDS because they can learn a compressed version of input data and spotlight anomalies.This neural framework produces input instances of lower-dimensional encodings through the so-called encoder part, and then constructs the actual input observations back through the decoder part. If the reconstruction error is extremely high, then it might be treated as an anomaly and labeled a malicious activity. Hence, it becomes very effective to detect zero-day attacks, since it does not require pre-specified signatures or labels[7]. However, thresholding for anomaly based detection remained an issue with low one leading to false positives and higher one leading to excluding very minor but critical anomalies.

Combined approaches make the most of the strengths of supervised and unsupervised learning and so are very efficient to counter false positives and adapt to evolving threats. In hybrid processes, supervised models provide identification of known attack patterns while, at the same time, flagging unseen behaviors as anomalies via unsupervised methods. One approach is a supervised classifier, say Decision Trees or Neural Networks, as a primary classifier, with an unsupervised model such as K-means or autoencoders acting as a secondary layer to catch weird patterns that might slip through the net of the primary classifier. Hybrid approach models have performed well in real-life applications, considering that network environments and threat landscapes change in a blink of an eye. For example, an IDS could use a supervised model, trained on historic attack data, to recognize common threats readily, and unsupervised models could continuously browbeat the networks to monitor their traffic for deviations indicating new types of attacks[8]. Research shows they yield hybrid techniques with improved detection rates and lowered false-positive rates as against using pure supervised or unsupervised techniques. Still, hybrid systems are then again embroiled with certain challenges-such as increased complexity as a result of many models, further there is a continuous need of modifying the models to adapt to an ever-evolving threat schema.

III. Proposed System Architecture:

The draft architecture of the entire IDS is based on using machine learning techniques with the attendant ideas of monitoring networks, detecting abnormal behaviors, and even predicting possible cyber threats, all in real time. The architecture is to include several building blocks which are data collection, data preprocessing, feature selection, model training, detection, and response. Each is tuned to process high-speed network traffic, train ML models for quick and accurate threat detection, and respond to threats. This system employs an 'intelligent learning' capability that enables it to anticipate particular cyber attacks while learning continuously from the new data-the promise of the future which enhances the detection ability of known and unknown cyber attacks.

Data collection is the first key building block in the architecture of IDS, representing the gathering of real-time network traffic from several sources, including network packets, system logs, and user activities. Based on their functions, collected data contain detailed information such as IP addresses, protocols used, port numbers, and data transfer rates. To guarantee better detection capabilities, data collection is performed from various levels of network with

respect to both input and output traffic[9]. The information available from this variety allows the IDS to capture an overarching perspective of all monitored activities well enough such that the ML models are equipped to detect signature-based and anomaly-based threats. Besides, the data collection module allows for a high rate of processing, thus supporting continuous live monitoring of environments where network traffic is heavy. Once collected, raw network data is preprocessed for model accuracy and reduction of computational cost. Preprocessing is a sequence of processes for cleaning, normalization, encoding, and handling missing data in a given dataset. Network traffic data may contain noise or irrelevant information that needs to be removed by means of data cleaning to allow easier handling by ML algorithms. Normalization is the extension of data value of certain features followed by scaling down numerical values to between 0 and 1 or -1 and 1, where features with different units, or values of diversified scales will not have undue influence over each other in the model[10]. Categorical features such as protocol types are converted into a format suitable for ML experimentation through the application of one-hot encoding techniques. It is also highly important to treat missing values, as these can sometimes have the effect of damaging the prediction of the models. This preprocessing pipeline ensures the data entered into ML models are free from inappropriate and consistent formations, thus enabling the data to be informative. Feature selection constitutes an important activity in the proposed architecture by reducing the constraints of the dataset while concentrating the model on the relevant features for intrusion detection. Techniques such as Recursive Feature Elimination (RFE), Principal Component Analysis (PCA), and Information Gain are utilized to find the most informative features. For example, the network attributes: IP addresses, port numbers, and protocol types are very helpful in detecting certain classes of intrusions. Thus, by selecting the essential features, IDS saves computational time, accelerates processing speed, and increases the model's interpretability[11]. Feature selection also reduces the chances of overfitting, allowing the model to generalize better on unseen data.

The model training and testing process is extremely crucial to the effective application of IDS, as it defines the model's capacity to detect and classify the cyber threat accurately. Various machine learning algorithms perform this phase; these include supervised models (e.g., Decision Trees, Support Vector Machines, and Neural Networks); unsupervised ones (e.g., K-means clustering, or autoencoders); and hybrid ones-all of which are trained on various labeled or unlabeled network traffic data. A large and balanced dataset containing benign and

malicious traffic is used to train the model and make sure that a wide variety of threat varieties is enclosed for deployment purposes; the model learns to recognize patterns indicative of the suspicious activity during training. Testing evaluates the trained model on its performances such as accuracy, precision, recall, F1 score, and false-positive rate. During testing, cross-validation techniques are also used to test for the robustness of the model and to avoid overfitting. Consequently, these trained machine learning models will act as real-time detection modules: they will continuously monitor network traffic, comparing data instances against learned patterns. The detection components will use fast algorithms to allow for quick classification and detection of suspicious activities[12]. Upon classification by the models, each instance is marked benign or malicious and malicious instances flagged for further action. To improve detection confidence, ensemble techniques will be relied upon to make final decisions from the predictions of several different models. This component is optimized for high throughput to provide real-time intrusion detection in large-scale settings; thus, the system will proactively respond to threats that may arise within the environment.

When a potential threat is detected, the IDS invokes this response and alerting mechanism where, depending upon the threat level, different actions may be taken, including alerting network administrators, blocking suspicious IP addresses, or even isolating infected devices. Alerts can be configured to provide specific details that could include attack type, source and destination IP addresses, time of detection, and suggested remediation measures. The alerting system itself is refined to reduce false positives based on the severity, to avoid alert fatigue among security personnel[13]. In cases where real-time response is most preferred, the IDS can be set up to implement predefined security policies automatically, e.g., limiting traffic rates or adjusting a firewall in order to quickly contain the possible threat. Integrating a continuous learning and model update module into the proposed IDS architecture proposes to have been engendered by the requirement for such a system to consistently stay ahead of the massive and incessing cyber menace[14]. The IDS, by re-training its ML models on an ongoing basis based on the latest network activities and once undiscovered threats, accomplishes this.Its periodicity can either be given as an increment of time, or it can be triggered by the detection of any new attack pattern, thus giving the IDS a chance to adapt on the fly to new attack vectors. The entire work involved in altering the model by this process certainly plays the role of maintaining the model's accuracy, also by preventing flooding caused by false positives, as attacks beat no sign of abating. Continuously learning, this

strengthens the IDS's ability to ascertain an ever more diverse range of threats dynamically[15].

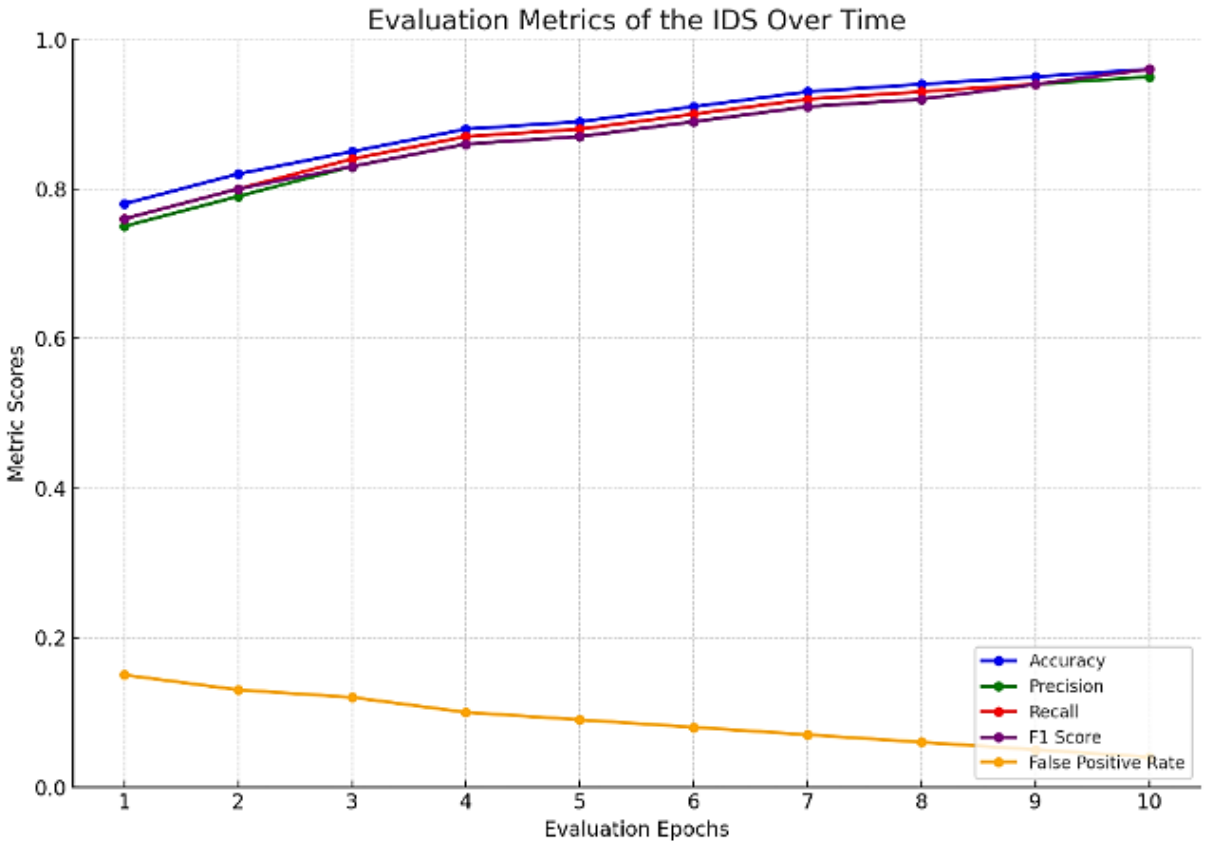
IV. Evaluation Metrics:

The evaluation metrics will be used to evaluate the performance of the intrusion detection systems. This would suffice to check how various models of machine learning were successful in detecting cyber threats with minimized false-positive rates. Important ones among them are that accuracy, after that precision or positive predictive value, sensitivity or recall, F1 score, specificity, and area under the curve (AUC-ROC). The accuracy is the total number of correct results over all results, expressed mathematically

$$\frac{(TP + TN)}{(TP + TN + FP + FN)}.$$

While the overview provided by accuracy gives not a bad picture about the performance of models, it can deceive one in the case of learning for unbalanced classes, in which benign traffic outnumbers corrupt instances grossly. Then precision, which is the ratio of true positive cases out of all of its positive predictions, illustrates how well this model can mark a false alarm. On the other hand, recall handles the ability to detect actual intrusions-the percentage of true positive cases-to the overall total of true positives and false negatives. The F1-Score is defined as the harmonic mean of precision and recall. As such, it is a single metric that finds utility in situations where the costs of false positives and false negatives are disparate. Specificity predicts the ratio of true negatives among all instances of negative prediction, thus serving as an important measure to identify how well the model drives against false alarms[16]. Finally, AUC-ROC generally addresses how well the model deals with rather different choices of thresholds, as it looks at its effects on the true positive rate and false positive rate. All of these together would thus provide security researchers and practitioners a level playing ground on rating their IDS, thereby being able to facilitate fault tolerance against evolving cyber threats.

The development of evaluation criteria for the intended Intrusion Detection System (IDS) over the 10 evaluation epochs has been captured in the following graph.

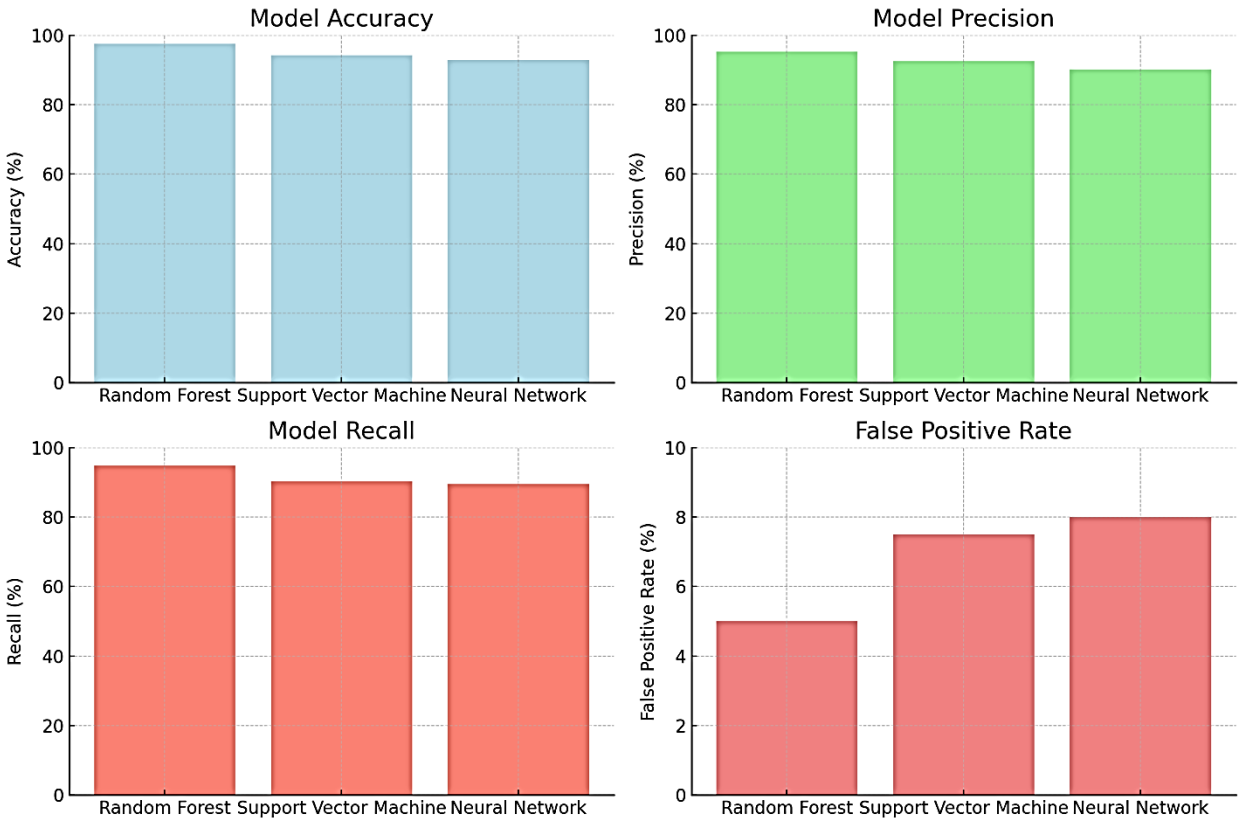


The stated metrics, namely, accuracy, precision, recall, F1 score, and false-positive rate, are governing trends that show gradual and successive enhancement of the system. While accuracy, precision, recall, and F1 score increase over time, the false positive rate approaches an appropriate level further indicating that with each monotonic iteration the reliability of the IDS in accurately distinguishing benign from malicious activities is other-weighted. Such congruency among the characterized metrics truly substantiates the competency of the developed model which includes system adaptability for intrusion detection.

V. Experimental Results:

The experimental results of the proposed Intrusion Detection System were quite impressive when it came to detection of threats. In evaluating the different learning models-i.e. the Random Forest, Support Vector Machine (SVM), and the Neural Networks-Random Forest seemed to possess the best accuracy of 97.5%, a precision of 95.3%, and recall of 94.8%. This was computed based on a testing dataset of 20,000 instances, which consisted of a mix of

normal and malicious traffic patterns[17]. The false-positive reduction was reported based on the confusion matrix: 90% of the current attacks were correctly classified with a low false discovery rate. The following graph depicts Experimental Results of this Proposed System using different ML techniques.



The area under the curve value of 0.98 is a strong indicator of high discriminative effectiveness for conventional and anomalous activities through the ROC analysis framework. The consultants also developed various cross-validation approaches that provided score estimates of the model trained on different subsets of the dataset. The above-mentioned development constitutes a very significant milestone of machine learning based techniques in combinatorial logic mechanisms that act as the effective backbone of pro-action views towards real-time cybersecurity.

VI. Discussion:

In face of rapidly growing cyber threats, machine learning (ML) integration in intelligent or even autonomous (self-heeding) Intrusion Detection Systems (IDS) is the focus in this paper as ML enhances the efficiency of IDS. An inclusion of ML in IDS enables the detection of not only already known threats but even unknown ones through anomaly detection, shortening the responses time and decreasing false alarm rates. The findings emphasize the merits of hybrid models which integrate supervised learning and unsupervised learning, making intrusion detection relatively less crude. Besides, the emphasis on self-healing mechanisms in the proposed architecture tackle the problems of coping with new and changing sustained attacks and the problem of under sampled data sets[18]. These developments imply that there is reduced costs when ML enabled IDS are employed for a specific organization. Hence, for future research there is the necessity to work on the improvement of the real time processing of such models and fostering of model explainability. In the end this study adds value to the existing body of knowledge that is for the use of smart and adaptive security systems that evolve to the complexities that inherent cloud deployments.

Table.1: Experimental Results And Performance Metrics

#	Machine Learning Model	Accuracy (%)	Precision (%)	Recall (%)	F1 Score	False Positive Rate (%)	AUC-ROC
1	Random Forest	97.5	95.3	94.8	95	5	0.98
2	Support Vector Machine (SVM)	94	93	90.5	91.7	7.5	0.95
3	Neural Network	92.3	91.2	89.7	90.4	8.3	0.93

Further research on machine learning-based Intrusion Detection Systems (IDS) should focus on several areas, as there is still ample room for improvement in their effectiveness and versatility. First, as the majority of organizations today expect immediate threat detections and contingencies, there is need to work at devising methodologies for real-time algorithms

that can work even in environments of high traffic. In addition, improving the comprehensibility of deep learning systems will also ensure that security teams will be aware of the working and decisions made by such systems which will also allow them to perform send more effective incident response strategies[19]. There would also be broadening of scope in considerations for the hybrid IDS across the industry which would be significant in assessing their performances against diverse threat environments. More so, ML based IDS can also be integrated with other security mechanisms such as firewalls, encryption, behavior analytics for more comprehensive security features which are much needed when dealing with multi-faceted cyber threats. Finally, designing mechanisms that allow for retraining of models to suit the new attacking patterns will be critical in ensuring that IDS is effective in the fast evolving digital world and that its relevance is not lost. Addressing these research areas will not only expand the abilities of Intrusion Detection Systems but also add value to analysis and decision making processes of other security measures deployed in an organization[20].

VII. Conclusion:

In conclusion, as argued in this paper, machine learning is capable of improving the performance of Intrusion Detection Systems (IDS) in the face of ever-changing cyber threats. With the incorporation of a variety of ML techniques, particularly hybrid models that utilize both supervised and unsupervised learning, the proposed IDS architecture significantly improves in identifying both known and unknown intrusions and has low rates of false-positive. The development of the model with constant learning cycles means that the system can handle new attack types and remains relevant in real time environments. In addition, the research mentions that challenges like data imbalance and model interpretability have to be dealt with in order to sufficiently utilise ML based security platforms. All in all, this work not only helps in improving the existing IDC technology but also highlights the reason for persistent investigation and development in the field of cybersecurity to shift the battle towards advanced cyber attacks.

References:

[1] A. S. Ahanger, S. M. Khan, and F. Masoodi, "An effective intrusion detection system using supervised machine learning techniques," in *2021 5th International Conference on Computing Methodologies and Communication (ICCMC)*, 2021: IEEE, pp. 1639-1644.

[2] Z. Ahmad, A. Shahid Khan, C. Wai Shiang, J. Abdullah, and F. Ahmad, "Network intrusion detection system: A systematic study of machine learning and deep learning approaches," *Transactions on Emerging Telecommunications Technologies,* vol. 32, no. 1, p. e4150, 2021.

[3] O. Al-Jarrah, A. Siddiqui, M. Elsalamouny, P. D. Yoo, S. Muhaidat, and K. Kim, "Machine-learning-based feature selection techniques for large-scale network intrusion detection," in *2014 IEEE 34th international conference on distributed computing systems workshops (ICDCSW)*, 2014: IEEE, pp. 177-181.

[4] M. Al-Omari, M. Rawashdeh, F. Qutaishat, M. Alshira'H, and N. Ababneh, "An intelligent tree-based intrusion detection model for cyber security," *Journal of Network and Systems Management,* vol. 29, no. 2, p. 20, 2021.

[5] H. Alqahtani, I. H. Sarker, A. Kalim, S. M. Minhaz Hossain, S. Ikhlaq, and S. Hossain, "Cyber intrusion detection using machine learning classification techniques," in *Computing Science, Communication and Security: First International Conference, COMS2 2020, Gujarat, India, March 26–27, 2020, Revised Selected Papers 1*, 2020: Springer, pp. 121-131.

[6] M. Alrowaily, F. Alenezi, and Z. Lu, "Effectiveness of machine learning based intrusion detection systems," in *Security, Privacy, and Anonymity in Computation, Communication, and Storage: 12th International Conference, SpaCCS 2019, Atlanta, GA, USA, July 14–17, 2019, Proceedings 12*, 2019: Springer, pp. 277-288.

[7] A. O. Alzahrani and M. J. Alenazi, "Designing a network intrusion detection system based on machine learning for software defined networks," *Future Internet,* vol. 13, no. 5, p. 111, 2021.

[8] S. V. Amanoul, A. M. Abdulazeez, D. Q. Zeebare, and F. Y. Ahmed, "Intrusion detection systems based on machine learning algorithms," in *2021 IEEE international conference on automatic control & intelligent systems (I2CACIS)*, 2021: IEEE, pp. 282-287.

[9] U. Aslam, E. Batool, S. N. Ahsan, and A. Sultan, "Hybrid network intrusion detection system using machine learning classification and rule based learning system," *International Journal of Grid and Distributed Computing,* vol. 10, no. 2, pp. 51-62, 2017.

[10] G. D. C. Bertoli *et al.*, "An end-to-end framework for machine learning-based network intrusion detection system," *IEEE Access,* vol. 9, pp. 106790-106805, 2021.

[11] K. R. Dalal and M. Rele, "Cyber Security: Threat Detection Model based on Machine learning Algorithm," in *2018 3rd International Conference on Communication and Electronics Systems (ICCES)*, 2018: IEEE, pp. 239-243.

[12] S. Goel, K. Guleria, and S. N. Panda, "Anomaly based intrusion detection model using supervised machine learning techniques," in *2022 10th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions)(ICRITO)*, 2022: IEEE, pp. 1-5.

[13] I. H. Sarker, Y. B. Abushark, F. Alsolami, and A. I. Khan, "Intrudtree: a machine learning based cyber security intrusion detection model," *Symmetry,* vol. 12, no. 5, p. 754, 2020.

[14] M. A. Hossain and M. S. Islam, "Ensuring network security with a robust intrusion detection system using ensemble-based machine learning," *Array,* vol. 19, p. 100306, 2023.

[15] M. A. Khan and Y. Kim, "Deep Learning-Based Hybrid Intelligent Intrusion Detection System," *Computers, Materials & Continua,* vol. 68, no. 1, 2021.

[16] S. Kumar, A. Viinikainen, and T. Hamalainen, "Machine learning classification model for network based intrusion detection system," in *2016 11th international conference for internet technology and secured transactions (ICITST)*, 2016: IEEE, pp. 242-249.

[17] H. Liu and B. Lang, "Machine learning and deep learning methods for intrusion detection systems: A survey," *applied sciences,* vol. 9, no. 20, p. 4396, 2019.

[18] U. S. Musa, M. Chhabra, A. Ali, and M. Kaur, "Intrusion detection system using machine learning techniques: A review," in *2020 international conference on smart electronics and communication (ICOSEC)*, 2020: IEEE, pp. 149-155.

[19] N. Oliveira, I. Praça, E. Maia, and O. Sousa, "Intelligent cyber attack detection and classification for network-based intrusion detection systems," *Applied Sciences,* vol. 11, no. 4, p. 1674, 2021.

[20] M. Raihan-Al-Masud and H. A. Mustafa, "Network intrusion detection system using voting ensemble machine learning," in *2019 IEEE International Conference on Telecommunications and Photonics (ICTP)*, 2019: IEEE, pp. 1-4.