University of Wah
Journal of Science and Technology

www.uow.edu.pk

Computational Intelligence Approaches for Analysis of the Detection of Zero-day Attacks

Shamshair Ali, Ghazif Adeem, Saif Ur Rehman, Shujat Hussain and Syed Shaheeq Raza

***Abstract*— As more and more people are adopting internet services; the measure of cybersecurity issues is also increasing exponentially. Zero-day attacks (unknown attacks) are affecting organizations badly even large-scale organizations had become victim of zero-days. Although there are many intrusion detection systems (IDS) and intrusion prevention systems (IPS) that are being used but still most of the zero-days remain invisible from these IDS. It is because they use new vulnerabilities in the system and previously no signature is found for those specific vulnerabilities, causing them to be misclassified by the IDS. This paper aims to discuss the performance of different Machine Learning (ML) and Deep Learning (DL) algorithms used in protecting cyberspace by presenting literature on the detection of zero-days. The latest and up-to-date literature was also presented which can help readers to get the latest insights into algorithms and models. Finally, we concluded the results in terms of the highest accuracy, precision, recall, and F1-Score of the comparative research articles against various datasets.**

***Index Terms*— Zero-day Attacks, Artificial Intelligence, Machine learning, Deep learning, Cyber Security**

INTRODUCTION

IN the last decade, the trend of adopting the internet and online services has increased exponentially [1]. Technological advances have changed the way of people work, communicate, and socialize [2]. , most probably known as signatures or fingerprints. Some attacks are really dangerous as they can change. It has become an integral part of life making it easy to get access to any information, enabling global communication and a source of entertainment. The internet's main goal is to transmit data from one end of the network (node) to another end (node) over the network. The Internet can be defined as an interconnected network of hundreds of thousands of networks, computers as well as associated devices. The transformation, evolution, and invention of modern devices and gadgets like IoT devices have significantly increased internet usage throughout the world[3]. Malware can be the cause of disturbance in IoT devices [4].

With all these benefits, there is a scary thing about the internet, which is all about the privacy and security concerns that most people face over the internet. As internet usage and the number of devices are increasing by every coming day, security issues are also increasing dramatically. That is why the internet has become the playground of cyber criminals [5]. They penetration test the entire network to know security loopholes and vulnerabilities are existing in the network. Cyberspace is continuously being intruded and malicious attacks are done against any kind of systems or services [6-9]. A system can be secure if it ensures the CIA triangle, which comprises three main components: confidentiality, integrity, and availability. The security and integrity of a system are said to be compromised whenever an illegal activity, destructive program, or unauthorized entity enters a computer or network to harm [10]. Cybersecurity is a set of different tools, techniques, devices, and approaches that can be used to safeguard cyberspace against any kind of cyber-attacks and cyber threats [11].

Based on the previous history; signature-based, rule-based, and ML-supervised algorithms have produced effective results to identify previously familiar and fully functional attacks that indicate discriminate patterns [12-14] [17], active threads and opened files [6], packets routing [18]. Machine learning is the study of computer algorithms to help us to come up with accurate and somewhat precise predictions of future events and circumstances and how to act intelligently in those circumstances. In general, machine learning tackles learning on how to produce better and more

Manuscript received; Jun 28, 2022; accepted Dec 7, 2022.

S. Ali (email: shamshair.ali145@gmail.com) is affiliated with University Institute of Information Technology (UIIT).

G. Adeem (email: ghazifadeem@gmail.com), S.U. Rehman (email: saif@uaar.edu.pk, S. H.ussain (email:shujat8620@gmail.com), S. S. Raza (email: Shaheeqraza1214@gmail.com) are affiliated with Pir Mehar Ali Shah- Arid Agriculture University, Rawalpindi.

*Corresponding author email: saif@uaar.edu.pk

advantageous circumstances in the future based upon learning from past experiences [17], active threads and opened files [6], packets routing [18]. Machine learning is the study of computer algorithms to help us to come up with accurate and somewhat precise predictions of future events and circumstances and how to act intelligently in those circumstances. In general, machine learning tackles learning on how to produce better and more advantageous circumstances in the future based upon learning from past experiences. Machine learning is the creation of models that help us in analyzing data from a variety of data repositories or datasets and using that data to predict system behavior in either different or relevant or similar scenarios [19]. A real-life zero-day attack life cycle is given by Fig. 1

A serious threat is posed by the zero-day attack to the security of the internet as computer systems are exploited by zero-day vulnerabilities. The zero-day attack can be described as "a traffic pattern of interest that, in general, has no matching patterns in malware or attack detection elements in the network", as stated by authors of [20]. Zero-day attacks of unknown nature (previously not disclosed) are being taken advantage of by attackers and they use them with other complex attacks to protect themselves from getting detected by the intrusion detection techniques, thus making them harder to defend against these kinds of attacks [21]. Zero-day attacks can come in many variations such as worms (polymorphic), viruses, trojans, network attacks as well as other malware. Blended attacks are the attacks that show effectiveness is not being detected, and also the worms (polymorphic) are sometimes not detected. This comprises sophisticated mutations for evading target defenses, targeted exploitation to directly attack specific hosts, use of multiple active, passing, and scanning techniques for vulnerabilities detection, dropping shells at compromised hosts for connecting back later on, and other post-exploitation techniques [22].

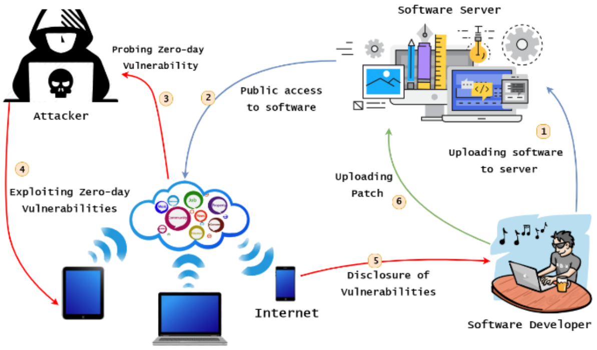

Software Server
Probing Zero-day Vulnerability
Attacker
3
2
Public access to software
1
Uploading software to server
Uploading Patch
6
4
Exploiting Zero-day Vulnerabilities
5
Disclosure of Vulnerabilities
Internet
Software Developer

Fig. 1: Zero-day Attack scenario

Fig. 1, show the real-life scenario of zero-day attack detection process, it shows that whenever a software or application comes to the internet hackers start trying to find loopholes in it to exploit the application may be for some financial benefit or for some revenge sometimes. Once they find something they may attack the application or report to the organization for the loophole [21]. Whatever the case is. Once the vulnerability becomes public or came into the knowledge of the creator, they immediately try to release the new update to deal with the problem and avoid the attack by keeping the system up to date. They fix the issue and upload the patch update to the product to keep their consumers safe from the attacks.

Researchers also demonstrated that the zero-day attacks are more used but are not apparent, as 11 out of 18 attacks were identified as previously unknown [23]. Their investigation shows that a zero-day attack can be present in a compromised system for a long period of time (10 months on average) before even they are detected by the security guys. Authors of [24] refer to a study (statistical) showing that more than 62% of attacks are detected after the system is compromised. Furthermore, the zero-day attacks are getting more and more in the wild as their number is increasing gradually [25].

Table 1 enlists the different terms that have been used frequently in the article.

TABLE 1
ACRONYMS USED IN THE PAPER

IDS	Intrusion Detection System
IPS	Intrusion Prevention System
ML	Machine Learning
DL	Deep Learning
ZA	Zero-day Attacks
ANN	Artificial Neural Network
PCA	Principle Component Analysis
KNN	K-Nearest Neighbors
TL	Transfer Learning
GAN	Generative Adversarial Networks
ODM	Outlier Dirichlet Mixture
SCADA	Supervisory Control and Data Acquisition System
DDoS	Distributed Denial of Service
CCM	Conjunction of Combinational Motifs
DNN	Deep Neural Networks
JNNS	Java Neural Network Simulator
GRU	Gated Recurrent Unit
CNN	Convolutional Neural Networks
LSTM	Long Short-Term Memory
OTS	One Time Signature
OTP	One Time Password
RNN	Recurrent Neural Networks
SVM	Support Vector Machine

This paper focuses specifically on zero-day attacks detection techniques using different AI-based algorithms. It is a comparative study of existing techniques used by previous researcher for detecting zero-days. It gives insights into previously used zero-day attack techniques as well as techniques used in recent years. Unlike other security domains, there is not much up-to-date comparative study done in recent years in the perspective of zero-days so if someone have to work on zero-days, they may have to go through a lot of different papers to get some information about different techniques currently being used that is why we thought to work on this area to show current (latest) techniques in this area so the readers can get the information all in a single article.

This paper is structured as follows: Section 2 will be containing a literature review; Section 3 will be of comparative analysis of different techniques being used for detecting zero-days. In the end, section 4 will cover the results and conclusions that we ended up on from our comparative study in the perspective of zero-day attacks.

LITERATURE REVIEW

The most common and effective way of detecting zero-days is Intrusion Detection System (IDS). They are good at tackling the exponential rise in detecting zero-days but still, they show deficiency as compared to the previously known attacks detection [26, 27]. Such attacks either take advantage of a new vulnerability in the system or exploit some previous vulnerability in a new way that cannot be detected in any of the IDS because it does not match with the existing known signatures. The growth of internet devices often exposes the systems to brand-new attacks that cause the growth of hacking activities. In such scenarios, the chances of zero-days become higher and higher. For designing effective and efficient IDS, ML and DL techniques have been widely used for better performance of the systems [28, 29].

Transfer learning [30] for detection of time series anomalies problem employing changing the weights of the source instances to check and match against target instances was presented by the authors. In the study, nearest neighbor classification was employed for anomaly detection based upon some of the labeled instances of the source.

The authors of [31] focused on the unification of feature spaces that are homogeneous. Domains' orthogonal transformation leveraging PCA was done. Afterward, they opted for the K-nearest neighbors classification technique on that transformed space for the detection of zero-day attacks.

A novel approach known as "Transferred Deep-Convolutional Generative Adversarial Network" (tDCGAN) with the purpose of detecting real malware from the fake malware that was generated by this approach itself was proposed [32]. This approach contains a detector that learns features of real malware and generated malware by utilizing a deep autoencoder, by which GAN training is also stabilized which is then used for detecting attacks of zero-day malware. For detection of zero-days authors claimed that their model was most robust.

Authors of [33] have proposed a zero-day polymorphic worm attack framework for detecting the worms (polymorphic) that are zero-day in nature by their behavior, anomaly, and signature-based techniques. Three layers namely, analysis, detection, and resource were used in the proposed architecture. Healthy and malicious traffic was used by the detection engine for detecting zero-day attacks.

Detecting attacks through graphical models have shown improvements when compared against behavioral-based or (anomaly-based) attack detection. Different concepts for implementing graphical models have been used as stated by works [34-36].

The authors of [37] proposed a detector for anomalies using the probability of network attack occurrences. A directed graph showing nodes (nodes), edges showing their communication in the entire network can be visualized. First and foremost, a behavior model for the stochastic attacker was introduced, then afterward the detector was used for comparing to network probability of the attacker's behavior when he attacks the host under normal conditions and compromises.

For detecting variations in DDoS attacks, an architecture named DaMask was proposed by Wang et al., [38], which updates the model according to new observations based upon Bayesian network inference.

An attack graph-based zero-day attack detection using layered architecture was presented by Singh et al in [39]. The proposed zero-day attack architecture consisted of a risk analyzer, physical and path generator layers. The centralized server and database of this architecture were used for other layers. An algorithm named, AttackRank, was proposed for finding exploitation chances in the graph.

A content-based visualization framework for classifying diverse signatures of the worm by using Conjunction of Combinational Motifs (CCM) was devised by Bayoglu et al. in [35]. Vertices of the graph were taken and considered as worms' invariant parts. Signatures of unseen worms (polymorphic) were automatically generated and detected by the CCM. In Table 2, an analysis of Zero-day Attacks Detection Techniques is given.

This section touches upon some of the main approaches followed by researchers that employ Deep Neural Network (DNN) and Machine Learning (ML) techniques to come up with their framework. Cyber Resilience Recovery Model (CRRM) was proposed by Tran et al. [40], which works to handle attacks in networks that are closed.

The framework of NIST SP 800-61, which is a incident response framework for resilience and standard is joined with the Susceptible Infected-Quarantined-Recovered (SIQR) model in [41] to capture zero-day attack and recovery.

In [42], an approach of detection based upon Artificial Neural Network (ANN) was proposed by Saied et al. for unknown and known DDoS attacks depending upon particular attributes that can separate Distributed Denial of Service attacks from authentic traffic. This model was then trained with the help of Java Neural Network Simulator (JNNS) upon data that was pre-processed, and Snort AI was integrated within.

Gated Recurrent Unit (GRU) technique was employed by the authors. Its main purpose was to pick up new Distributed Denial of Service (DDoS) attacks. Authors claimed that the proposed shows better accuracy [43].

In a recent study [44], an approach was implemented to resolve issues of security in-vehicle communication that are possibly open to plenty of attacks. It is a hybrid technique based upon GRU and CNN for detecting possible attacks.

A Deep Neural Network (DNN) approach for detecting cyber-attacks was proposed by the authors. It combines techniques such as Principal Component Analysis (PCA) and GWO algorithm where the responsibility of PCA is to reduce the dimensions of the dataset and then GWO is used for optimizing the transformed dataset for the redundancy reduction in the transformed dataset. The main focus of this approach is the dimensionality reduction for making DNN-Based IDS more responsive [45, 46].

A multi-stage attention mechanism alongside CNN that is LSTM based was proposed for anomaly detection [47]. Data abnormality in data generated through various sensors in automated vehicles is specifically covered within the proposed method. An ensemble approach based upon the voting technique for deciding anomalous data from different classifiers was also proposed.

Another technique for intrusion detection and prevention system using classification and one-time signature technique was proposed for the cloud in [48], by the authors. The proposed technique OTS is different from the OTP which stands for one-time password and the OTS is used for accessing the data over the cloud. Hybrid classification consisting of normalized K-Means and Recurrent Neural Network (RNN) was used.

For detecting zero-day attacks, an approach based on a deep autoencoder was proposed in [49]. Its performance was demonstrated by using two popular and well-known datasets which are NSL-KDD and CICIDS2017 and the performance was compared against the one-class SVM outlier detection.

Comparative Analysis:

In this section, the algorithms and their brief description will be provided along with the discussion regarding the results of these algorithms. A brief description of the confusion matrix and metrics such as accuracy, precision, and recall will also be given.

A. Evaluation Metrics

Before explaining the metrics, we would give a brief introduction to the confusion matrix. A confusion matrix is generally a 2x2 matrix layout that is used to visualize the performance of an algorithm. Actual performance measures that must or should be met are written vertically while the predicted ones by the algorithm are written horizontally [50]. *Fig. 2,* shows the typical confusion matrix for 2 X 2 matrix.

- True Positive (TP) is defined as the total positive instances being identified as positive.

$$\text{TPR} = \frac{\text{TP}}{\text{TP} + \text{FN}}$$

- True Negative (TN) is defined as the number of negative instances being identified as negative.

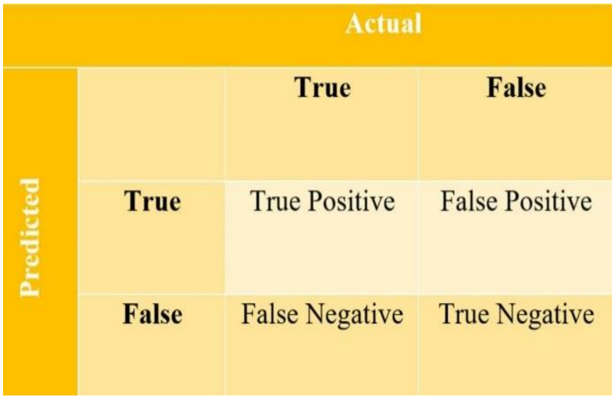
$$\text{TNR} = \frac{\text{TN}}{\text{TN} + \text{FP}}$$

- False Positive (FP) is defined as the number of negative instances being classified or predicted as positive.

$$\text{FPR} = \frac{FP}{TN + FP}$$

- False Negative (FN) is defined as the number of positive instances being classified or predicted as negative.

$$\text{FNR} = \frac{\text{FN}}{\text{TP} + \text{FN}}$$



		Actual	
		True	False
Predicted	True	True Positive	False Positive
	False	False Negative	True Negative

Fig. 2: 2x2 Confusion Matrix

Accuracy: is defined as the ratio between the number of correct predictions and a total number of predictions [51].

$$Acc = \frac{TP + TN}{TP + FP + FN + TN}$$

Precision: is defined as the ratio between TPs combined to a number of TPs and FPs. It is the percentage of correctly identified positives out of all results which were said to be positive either correctly or not [52].

$$Precision = \frac{TP}{TP + FP}$$

Recall: is defined as the ratio between TPs combined to a number of TPs and FNs. It is the percentage of correctly identified positives out of all actual positives, either correctly or not [52].

$$Recall = \frac{TP}{TP + FN}$$

F1-score: It takes both false negatives and false positives into consideration, and it is the harmonic mean of recall and precision. It performs well on datasets that are imbalanced [53].

$$F1\text{-}Score = \frac{2 * (Precision * Recall)}{(Precision + Recall)}$$

B. Algorithms and Techniques

Machine learning and deep learning approaches are widely used to deal with cyber-attack detection and prevention. In the last few years, many researchers applied ML and DL techniques to classify and detect zero-days from the systems including malware, malicious URLs and spam, etc. A Comparison of different AI based techniques for Zero-day Attack detection is given in Table 2.

Most of the zero-day malware is not detected by antiviruses that’s why they are problematic, in [32] authors proposed a zero-day malware detection framework based on Deep Autoencoder(DAE). It generates fake malware and trains the model to distinguish between real and fake malware by comparing the fake data with the real data. It learns different malware features from both real data and fake generated data using the proposed model (tDCGAN). It extracts the appropriate features from data and stabilizes the training. The trained model used transfer learning to capture malware features with an average classification accuracy of 95.74%

TABLE 2
COMPARISON OF DIFFERENT AI BASED TECHNIQUES FOR ZERO-DAY ATTACK DETECTION

ATTACKS	REF	YEAR	DATASETS	APPROACH	ACCURACY %	PRECISION %	RECALL %	F1-SCORE %
IDS	[54]	2019	ICS dataset	HML-IDS	97	98	92	95
	[54]	2019	ICS dataset	Bloom Filter	89	97	67	78
	[54]	2019	ICS dataset	RF	91	93	81	86
	[55]	2017	ICS dataset	LSTM	92	94	78	85
	[55]	2017	ICS dataset	SVDD	76	95	21	34
	[55]	2017	ICS dataset	Bloom Filter	87	97	59	73
	[56]	2021	ICS Dataset	BLOSOM	97	98	92	95
	[57]	2018	ICS dataset	MLP	95	96	90	
	[58]	2021	SWaT	CNN	97.85	98.8	83	90.2
	[59]	2020	NSL KDD, CIDD	DNN	91.83			
INSIDER THREAT DETECTION	[60]	2017	CERT	Unsupervised KNN	54	47.5	44.2	44.9
	[61]	2018	CERT	Hidden Markov Model	71.1	64.1	55.9	61.7
	[62]	2021	CERT	SVM	70	40	11	60
	[62]	2021	CERT	LSTM	75	20	59	30
	[62]	2021	CERT	DNN	86	36	73	48
	[62]	2021	CERT	MITD	92	54	54	55
	[62]	2021	CERT	HITD	97	77	92	84
	[49]	2020	NSL KDD	Autoencoder	92.96			
	[63]	2021	NSL KDD	Stacker	99.39	99.7	99	99.3
ANOMALY BASED	[62]	2021	CERT	AITD	90	49	50	49
	[56]	2021	SWaT	CNN	92	88	98	92
	[56]	2021	SWaT	DBN	80	72	72	83
	[56]	2021	SWaT	PCA+CNN	95	94	97	95
	[56]	2021	SWaT	PCA+DBN	91	88	95	91
	[56]	2021	SWaT	BLOSOM	96	96	98	96
	[47]	2020	SPMD	MSALSTM-CNN	96.56	99.06		97.37
	[47]	2020	SPMD	WAVED	94.87	98.87		95.44
	[64]	2019	SPMD	KF	97.4	94.5		91.7
	[64]	2019	SPMD	CNN	98.0	99.8		96.4
	[64]	2019	SPMD	CNN-KF	98.2	99.5		96.8

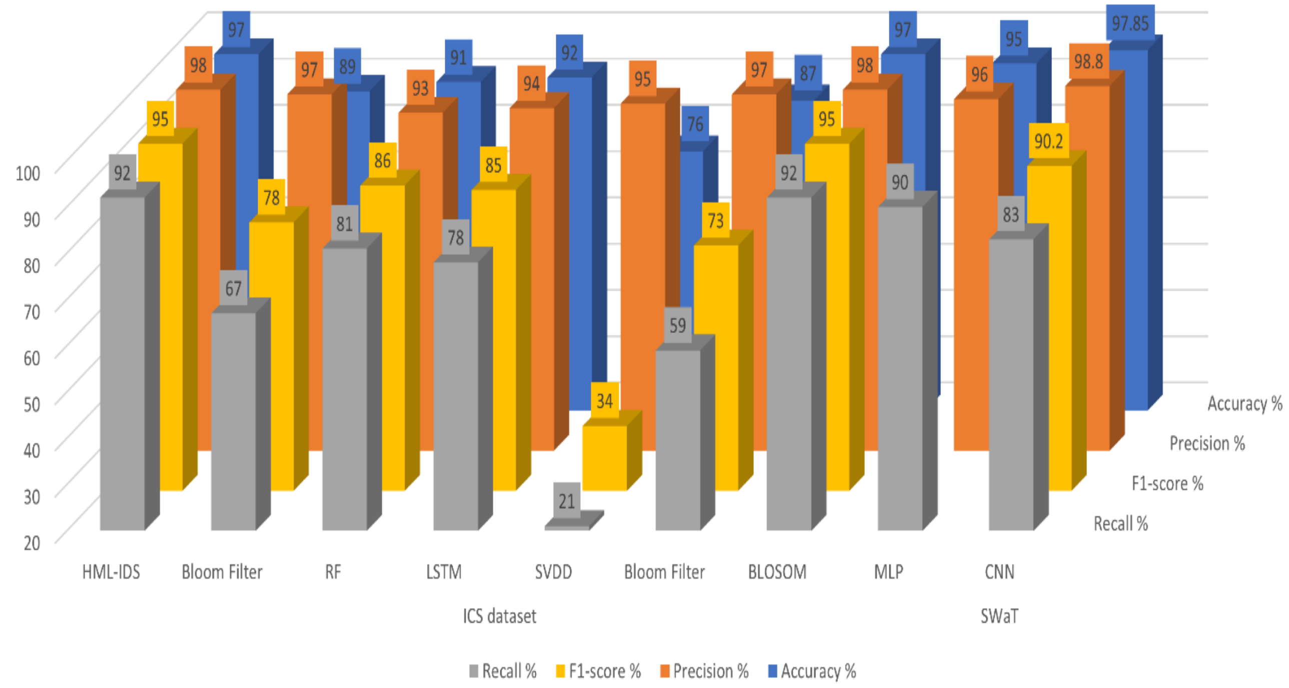


Fig. 3: Comparison of Techniques for Intrusion Detection System

A hybrid model was used that took patterns from communication that was consistent and anticipated from multiple devices of Industrial Control Systems (ICS). ICS dataset was used, which was then improved (pre-processed) by using normalization, standardization, and labeling (categorical) of the samples and feature scaling. Then features were extracted from the dataset and the feature matrix was constructed by using PCA, CCA, and ICA. The proposed technique HML-IDS was based upon instance-based k-NN learning algorithm. It was trained on the features that were reduced by reduction techniques and then the results were compared with Bloom Filter (BF), Random Forest (RF). The proposed technique showed better results as it achieved 0.97, 0.98, 0.92, 0.95 in accuracy, precision, recall, F1-score respectively. The main question to raise here is why were the deep learning techniques not used? As generally, DL techniques perform better, resulting in more improved accuracy, precision, recall, F1-score [54].

Sameera et. al [59] proposed a zero-day attack detection model for IDS systems, they applied DNN to build their model. It was difficult for them to detect and identify zero-days because there were no labels in Intrusion detection systems, so they used the Manifold alignment method to map source and targets using their mapping functions and assign soft labels, and then applied DNN to identify zero-days from the test data. They used NSL-KDD and CIDD datasets for testing and by comparing their results with other ML Models i.e., DT, RF, SVM, KNN they found that their proposed framework outperformed the other models and achieved an average accuracy of 91.83%.

Botnet detection using a reinforcement-based learning approach was discussed and presented by the authors. The whole workflow is that the network traffic captured was filtered down and reduced by control filters, feature extraction from connections and hosts was done, and then reduced by the CART algorithm. Then those features were passed into a multi-layer neural network classifier while dividing features and data into training and testing and at the end classifying them into anomalies or not. The resultant accuracy of detection is 98.3% with a lower false-positive rate of 0.012% [65].

Hidden Markov Model (HMM) is also used for insider threat detection by [61], they used the CERT r.4.2 which contains 70 malicious users ad they train users in the first week and record their activities on the different task on weekly bases and find the similarities between the activities and then classify it as malicious or normal.

Another autoencoder-based model is proposed by Kunang et. al [66]. They applied automatic feature extraction on IDS using the autoencoder approach. Their experiments showed a high accuracy of around 99.947% which is quite impressive but unfortunately, they used very old datasets which are from 1998, and hence they will be missing signatures of new attacks.

It is difficult to obtain data samples for all attacks classes in Network Intrusion Detection Systems (NIDS) to observe the traffic in production. Machine learning-based NIDSs face unknown attack traffic known as zero-day attacks that are not used in training because they were not existing at the time of training. The authors proposed a Zero-shot Learning ZSL technique to detect zero-days in NIDS to evaluate the performance of the proposed model. It maps the features of unknown attacks from the network to differentiate its attributes from known attacks. They defined a new metric named Zero-day detection rate to measure the effectiveness of the model [67].

An autoencoder-based framework for detection zero-day attacks, authors in [49] used NSL-KDD and CICIDS2017 datasets to perform the tests, they compared their model with one class SVM on both the datasets and outperformed by achieving the higher average accuracy of 92.96% for NSL-KDD and 95.19% for CICIDS2017.

Another study [60] used an ML-based insider threat detection model based on KNN, they used two approaches to detect insider threats, user-based and role base. In user-based, they calculate the abnormal score of user's session with its previous sessions based on his activities in the system, and for role-based, they apply the same technique on sessions in different roles and compare the results with the previous sessions of the same role and calculated abnormal score.

A recent study [47] for detecting anomalies in connected automated vehicles (CAVs) proposed a multistage attention mechanism with LSTM based Convolutional neural network model named MSALSTM-CNN for detecting different anomalies caused by faults, errors or may be due to cyber-attacks, whichever the cause is, it can result in accidents. Since it's a matter of human lives so this can't be compromised in any case. It converts data streams into multidimensional vectors and then processes them to detect anomalies. Another method they introduced works on the principle of average predicted probability of multiple classifiers for anomaly detection is a weight-adjusted fine-tuned ensemble: WAVED. They effectively improved the anomaly detection rate in both low and high magnitude cases of anomalous instances by gaining 2.54% in F1-score for detection of single anomaly types in the dataset. Moreover, it showed promising performance with a gain of up to 3.24% in F1-score in detecting mixed anomaly types in CAVs.

Researchers developed a benign database of communication from multiple devices in ICS by observing the communication patterns of system for a period of time with the usage of Bloom Filter. Then another SCADA gas pipeline dataset was created. Afterward, researchers proposed stacked LSTM based for detecting time-series anomaly detection by combining both datasets and Bloom filter. The results were found to be much better than state-of-the-art techniques, as claimed by researchers of the research. The outcomes of the framework were then compared with SVDD, BN, and BF. The proposed work's results were 0.94, 0.78, 0.92, and 0.85 for precision, recall, accuracy, and F1-score respectively. Other deep learning approaches should have also been tried to maximize the performance. The proposed framework misclassified attack types such as MSCI, MPCI, and considered their behavior as normal instead of harmful [55].

Authors studied previous literature and noted problems with previous approaches in the detection of zero-day attacks as well as irregularities in data which results in a poor rate of attack detection. They proposed approach with steps starting from Bloom Filter payload level detection, then using Kohonen enhanced neural network by utilizing PCA and hyper-graph partitioning, and then finally using BLOSOM-based hybrid anomaly detection using datasets obtained from Singapore university (SWAT dataset) and Mississippi state university (gas-pipeline data from SCADA lab). BLOSOM model imputes packet contents for checking data's behavioral pattern in an unsupervised fashion. It helps in identification to see if the contents of the packet lie within the ANN training phase. The proposed technique using both datasets showed improved results while comparing to BLF, RF, RNN, and CNN. The results were 0.97, 0.98, 0.92, 0.95 in terms of accuracy, precision, recall and F1-score respectively [56].

The authors focused on anomaly detecting in industrial control systems by taking packet latency and jitter into consideration. An algorithm based upon Grey Wolf optimizer neural network training for anomalies detection was proposed to be applied in industrial control systems. Multiple datasets gas-pipeline, swat from different sources were used. Grey Wolf Optimizer (GWO) is the principle for imitating wolves' behavior in nature to for hunting in cooperative way. Mainly leadership hierarchy is imitated by alpha, omega, alpha, delta, beta wolves. Optimization is performed by three basic steps prey searching, prey encircling, and attacking of prey. Grey wolf algorithm's performance using gas, swat datasets were compared against PSO, BBO, ACO, ES, and PBIL optimized algorithms with ANN. The accuracy of the Grey Wolf algorithm was 98% on gas-pipeline while it achieved 96% accuracy on swat. For the concern of robustness and accuracy, GWO achieves higher. If time is of more concern then the ES algorithm takes lower time [57].

RESULTS AND DISCUSSIONS

A. Intrusion Detection Systems

In our research, we have compared the latest articles regarding zero-day attacks in multiple domains such as IDS, anomaly-based and insider threat detection etc. In the area of generics attacks detected by IDS, we have reviewed almost 10 different approaches. It can be seen that the stacker-based approach was the best in terms of accuracy as it reached up to 98.8% accuracy. In terms of precision, both HML-IDS and BLOSOM attained 98% precision. The higher recall and F1-Score (97.9%, 98.8 respectively) was achieved by the reinforcement learning approach.

Fig. 3, shows the comparison of different algorithms overall accuracy of different techniques (Bloom Filter, RF, LSTM, CNN and DNN etc.) which were trained and tested against different datasets covering a variety of attack vectors ranging from various sizes and multiple domains including industrial level systems. All the algorithms were compared using 4 main performance metrices i.e., accuracy, precision, recall and f1-score.

B. Anomaly Based

Fig. 4, lists down graphical results and comparison of different AI-based techniques used for detecting anomaly based zero-day attacks. Results were evaluated using accuracy, precision, recall and f1-score.

Discussing anomaly-based attack detection, CNN-KF based approach on the dataset of SPMD achieved 98.2% accuracy which is highest among the other compared approaches. Although the highest precision was 99.8% which was achieved CNN on the same dataset of SPMD. While the highest recall was achieved by the BLOSOM approach on the SWat dataset. MSALSTM-CNN approach outperformed others in terms of F1-score.

C. Insider Threat Detection

In case of insider threat detection, stacker showed promising performance over NSL-KDD dataset with all the four evaluation metrices having more than 99% and HITD over CERT dataset with 97% accuracy. The algorithms used for detecting insider threats including SVM, LSTM, DNN, Stacker, Autoencoder and Hidden Markov model etc. are being used for the comparison of results using, accuracy, precision, recall and f1-score. *Fig. 5*, provides the graphical representation of different AI-based techniques and their results among the used datasets in different studies by previous researchers.

CONCLUSION

In all aspects zero-day attacks are critical to any system whether they are in IDS systems or spam based, they can cause massive damage to any organization. Rival companies find zero-day vulnerabilities and take advantage of that against their competitors by gaining access to their sensitive data by compromising their systems. Attackers normally try to remain inside the system without letting anyone know about it, just to steal the confidential data instead of destroying the system. Tend to remain hidden as long as possible because once the security team gets to know about the vulnerability, they fix the issue as quickly as possible. Because of no previous records and signatures, zero-days are hard to detect by firewalls or other security measures used by the organizations to keep their systems secure from potential threats.

In the last decade, many researchers introduced AI-based approaches to detect zero-days based on their behaviors and some other factors, but they are still facing high false negative and false positive rates because most of them were using very old datasets like DARPA or KDD, although they are very large and famous datasets in cyber security obviously, they will be missing new attacks and new strategies being used nowadays in attacks.

Although some of them achieved impressive results but they used their custom datasets may be designed in a way that they produced good results on their systems when the same models were applied on other datasets they come up with many different results [68]. Moreover, everyone is using different evaluation metrics to express their results that are not even suitable for cyber security scenarios. Due to the sensitivity of the zero-days, only accuracy or precisions are not enough to measure the performance of the models they should take care of false alarms to implement their models on actual systems. Secondly, instead of using outdated datasets, new benchmarks should be used to tackle modern attacks properly.

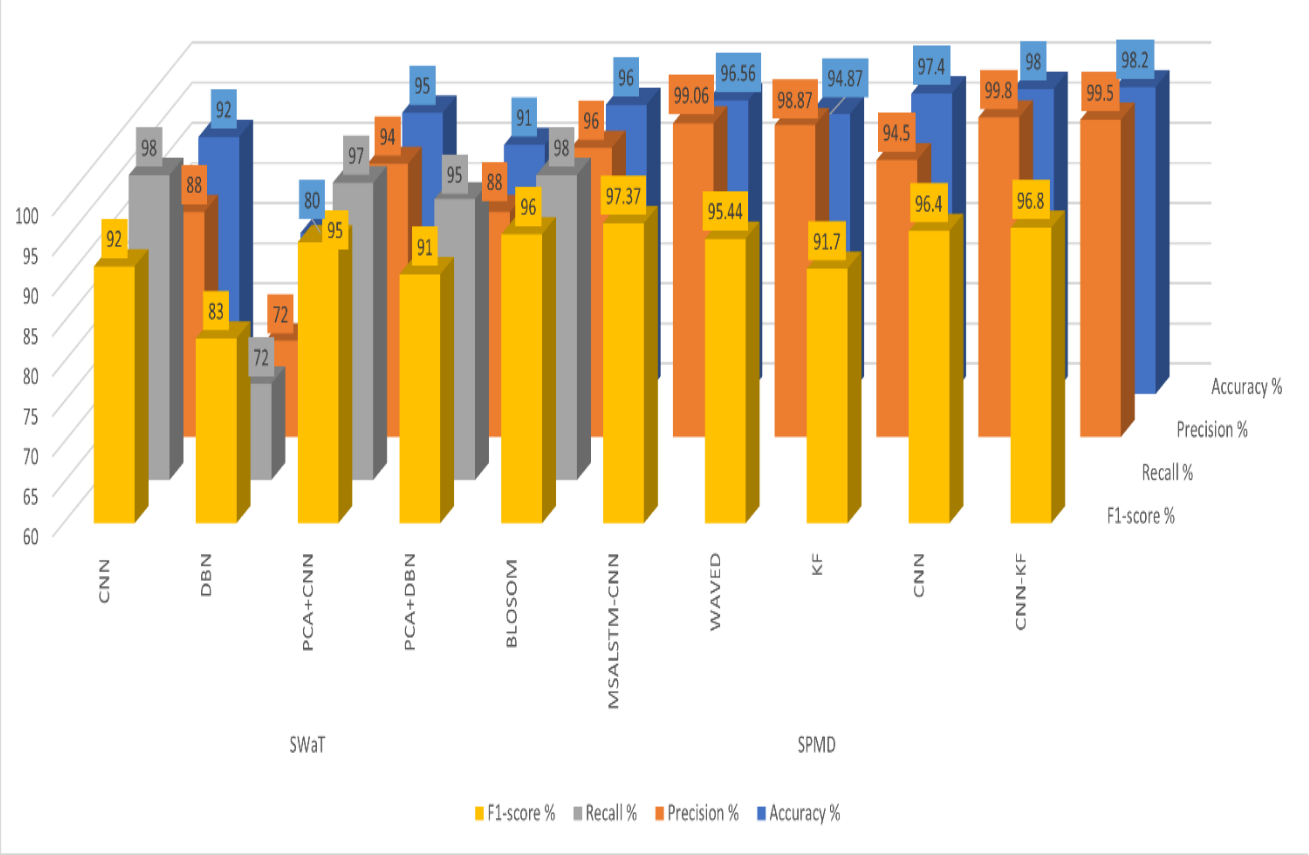


Fig. 4: Comparison of Techniques for Insider Threat Detection

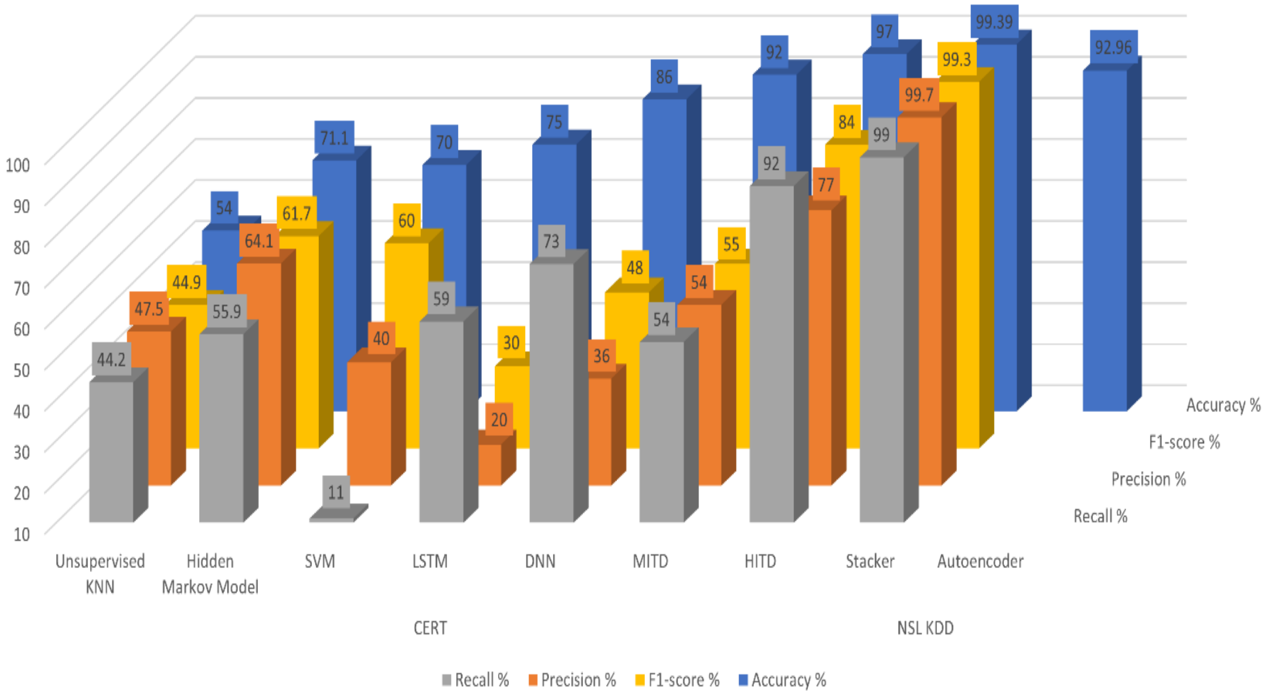


Fig. 5: Comparison of Techniques for Anomaly Based Attacks

REFERENCES

[1] H. Yoon, Y. Jang, S. Kim, A. Speasmaker, and I. Nam, "Trends in internet use among older adults in the United States, 2011–2016," *Journal of Applied Gerontology,* vol. 40, no. 5, pp. 466-470, 2021.

[2] A. Darem, "Anti-Phishing Awareness Delivery Methods," *Engineering, Technology & Applied Science Research,* vol. 11, no. 6, pp. 7944-7949, 2021.

[3] R. Malik, Y. Singh, Z. A. Sheikh, P. Anand, P. K. Singh, and T. C. Workneh, "An Improved Deep Belief Network IDS on IoT-Based Network for Traffic Systems," *Journal of Advanced Transportation,* vol. 2022, 2022.

[4] A. Al-Marghilani, "Comprehensive Analysis of IoT Malware Evasion Techniques," *Engineering, Technology & Applied Science Research,* vol. 11, no. 4, pp. 7495-7500, 2021.

[5] D. K. Bhattacharyya, and J. K. Kalita, *Network anomaly detection: A machine learning perspective*: Chapman and Hall/CRC, 2019.

[6] Y. Zeng, X. Hu, and K. G. Shin, "Detection of botnets using combined host-and network-level information." pp. 291-300.

[7] I. Studnia, V. Nicomette, E. Alata, Y. Deswarte, M. Kaâniche, and Y. Laarouchi, "Survey on security threats and protection mechanisms in embedded automotive networks." pp. 1-12.

[8] J. Meakins, "A zero-sum game: the zero-day market in 2018," *Journal of Cyber Policy,* vol. 4, no. 1, pp. 60-71, 2019.

[9] B. Fang, Q. Lu, K. Pattabiraman, M. Ripeanu, and S. Gurumurthi, "ePVF: An enhanced program vulnerability factor methodology for cross-layer resilience analysis." pp. 168-179.

[10] V. Ambalavanan, "Cyber threats detection and mitigation using machine learning," *Handbook of Research on Machine and Deep Learning Applications for Cyber Security*, pp. 132-149: IGI Global, 2020.

[11] D. Craigen, N. Diakun-Thibault, and R. Purse, "Defining cybersecurity," *Technology Innovation Management Review,* vol. 4, no. 10, 2014.

[12] S. He, J. Zhu, P. He, and M. R. Lyu, "Experience report: System log analysis for anomaly detection." pp. 207-218.

[13] M. Al-Qatf, Y. Lasheng, M. Al-Habib, and K. Al-Sabahi, "Deep learning approach combining sparse autoencoder with SVM for network intrusion detection," *IEEE Access,* vol. 6, pp. 52843-52856, 2018.

[14] H. Hindy, D. Brosset, E. Bayne, A. K. Seeam, C. Tachtatzis, R. Atkinson, and X. Bellekens, "A taxonomy of network threats and the effect of current datasets on intrusion detection systems," *IEEE Access,* vol. 8, pp. 104650-104675, 2020.

[15] K. Pan, E. Rakhshani, and P. Palensky, "False Data Injection Attacks on Hybrid AC/HVDC Interconnected Systems With Virtual Inertia Vulnerability, Impact and Detection," *IEEE Access,* vol. 8, pp. 141932-141945, 2020.

[16] T. Zoppi, A. Ceccarelli, L. Salani, and A. Bondavalli, "On the educated selection of unsupervised algorithms via attacks and anomaly classes," *Journal of Information Security and Applications,* vol. 52, pp. 102474, 2020.

[17] M. Hanselmann, T. Strauss, K. Dormann, and H. Ulmer, "CANet: An unsupervised intrusion detection system for high dimensional CAN bus data," *IEEE Access,* vol. 8, pp. 58194-58205, 2020.

[18] Z. Shu, J. Wan, J. Lin, S. Wang, D. Li, S. Rho, and C. Yang, "Traffic engineering in software-defined networking: Measurement and management," *IEEE access,* vol. 4, pp. 3246-3256, 2016.

[19] A. Subasi, *Practical Machine Learning for Data Analysis Using Python*: Academic Press, 2020.

[20] C. Chapman, "Chapter 1 - Introduction to practical security and performance testing," *Network Performance and Security, Syngress, ISBN 9780128035849*, pp. 1-14, 2016.

[21] M. S. Alzahrani, and F. W. Alsaade, "Computational Intelligence Approaches in Developing Cyberattack Detection System," *Computational Intelligence and Neuroscience,* vol. 2022, 2022.

[22] A. P. Singh, "A study on zero day malware attack," *International Journal of Advanced Research in Computer and Communication Engineering,* vol. 6, no. 1, pp. 391-392, 2017.

[23] L. Bilge, and T. Dumitraş, "Before we knew it: an empirical study of zero-day attacks in the real world." pp. 833-844.

[24] T. T. Nguyen, and V. J. Reddi, "Deep reinforcement learning for cyber security," *arXiv preprint arXiv:1906.05799,* 2019.

[25] K. Metrick, P. Najafi, and J. Semrau, *Zero-Day Exploitation Increasingly Demonstrates Access to Money, Rather than Skill—Intelligence for Vulnerability Management,* Technical Report, FireEye Technical Report., 2020.

[26] N. Kaloudi, and J. Li, "The ai-based cyber threat landscape: A survey," *ACM Computing Surveys (CSUR),* vol. 53, no. 1, pp. 1-34, 2020.

[27] H. Hindy, E. Hodo, E. Bayne, A. Seeam, R. Atkinson, and X. Bellekens, "A taxonomy of malicious traffic for intrusion detection systems." pp. 1-4.

[28] A. Khraisat, I. Gondal, P. Vamplew, and J. Kamruzzaman, "Survey of intrusion detection systems: techniques, datasets and challenges," *Cybersecurity,* vol. 2, no. 1, pp. 1-22, 2019.

[29] H. Hindy, D. Brosset, E. Bayne, A. Seeam, C. Tachtatzis, R. Atkinson, and X. Bellekens, "A taxonomy and survey of intrusion detection system design techniques, network threats and datasets," 2018.

[30] V. Vercruyssen, W. Meert, and J. Davis, "Transfer learning for time series anomaly detection." pp. 27-37.

[31] N. Sameera, and M. Shashi, "Transfer Learning Based Prototype for Zero-Day Attack Detection," *International Journal of Engineering and Advanced Technology (IJEAT),* vol. 8, no. 4, 2019.

[32] J.-Y. Kim, S.-J. Bu, and S.-B. Cho, "Zero-day malware detection using transferred generative adversarial networks based on deep autoencoders," *Information Sciences,* vol. 460, pp. 83-102, 2018.

[33] R. Kaur, and M. Singh, "A hybrid real-time zero-day attack detection and analysis system," *International Journal of Computer Network and Information Security,* vol. 7, no. 9, pp. 19-31, 2015.

[34] X. Sun, J. Dai, P. Liu, A. Singhal, and J. Yen, "Using Bayesian networks for probabilistic identification of zero-day attack paths," *IEEE Transactions on Information Forensics and Security,* vol. 13, no. 10, pp. 2506-2521, 2018.

[35] B. Bayoğlu, and İ. Soğukpınar, "Graph based signature classes for detecting polymorphic worms via content analysis," *Computer Networks,* vol. 56, no. 2, pp. 832-844, 2012.

[36] Z. Yichao, Z. Tianyang, G. Xiaoyue, and W. Qingxian, "An improved attack path discovery algorithm through compact graph planning," *IEEE Access,* vol. 7, pp. 59346-59356, 2019.

[37] J. Grana, D. Wolpert, J. Neil, D. Xie, T. Bhattacharya, and R. Bent, "A likelihood ratio anomaly detector for identifying within-perimeter computer network attacks," *Journal of Network and Computer Applications,* vol. 66, pp. 166-179, 2016.

[38] B. Wang, Y. Zheng, W. Lou, and Y. T. Hou, "DDoS attack protection in the era of cloud computing and software-defined networking," *Computer Networks,* vol. 81, pp. 308-319, 2015.

[39] U. K. Singh, C. Joshi, and D. Kanellopoulos, "A framework for zero-day vulnerabilities detection and prioritization," *Journal of Information Security and Applications,* vol. 46, pp. 164-172, 2019.

[40] H. Tran, E. Campos-Nanez, P. Fomin, and J. Wasek, "Cyber resilience recovery model to combat zero-day malware attacks," *computers & security,* vol. 61, pp. 19-31, 2016.

[41] J. Sterman, "Business dynamics(p. c2000)," Irwin/McGraw-Hill, 2010.

[42] A. Saied, R. E. Overill, and T. Radzik, "Detection of known and unknown DDoS attacks using Artificial Neural Networks," *Neurocomputing,* vol. 172, pp. 385-393, 2016.

[43] S. ur Rehman, M. Khaliq, S. I. Imtiaz, A. Rasool, M. Shafiq, A. R. Javed, Z. Jalil, and A. K. Bashir, "DIDDOS: An approach for detection and identification of Distributed Denial of Service (DDoS) cyberattacks using Gated Recurrent Units (GRU)," *Future Generation Computer Systems,* vol. 118, pp. 453-466, 2021.

[44] A. Rehman, S. U. Rehman, M. Khan, M. Alazab, and T. Reddy, "CANintelliIDS: detecting in-vehicle intrusion attacks on a controller area network using CNN and attention-based GRU," *IEEE Transactions on Network Science and Engineering,* 2021.

[45] S. P. RM, P. K. R. Maddikunta, M. Parimala, S. Koppu, T. R. Gadekallu, C. L. Chowdhary, and M. Alazab, "An effective feature engineering for DNN using hybrid PCA-GWO for intrusion detection in IoMT architecture," *Computer Communications,* vol. 160, pp. 139-149, 2020.

[46] P. More, and P. Mishra, "Enhanced-PCA based Dimensionality Reduction and Feature Selection for Real-Time Network Threat Detection," *Engineering, Technology & Applied Science Research,* vol. 10, no. 5, pp. 6270-6275, 2020.

[47] A. R. Javed, M. Usman, S. U. Rehman, M. U. Khan, and M. S. Haghighi, "Anomaly detection in automated vehicles using

multistage attention-based convolutional neural network," *IEEE Transactions on Intelligent Transportation Systems*, 2020.

[48] V. Balamurugan, and R. Saravanan, "Enhanced intrusion detection and prevention system on cloud environment using hybrid classification and OTS generation," *Cluster Computing,* vol. 22, no. 6, pp. 13027-13039, 2019.

[49] H. Hindy, R. Atkinson, C. Tachtatzis, J.-N. Colin, E. Bayne, and X. Bellekens, "Utilising deep learning techniques for effective zero-day attack detection," *Electronics,* vol. 9, no. 10, pp. 1684, 2020.

[50] I. Markoulidakis, I. Rallis, I. Georgoulas, G. Kopsiaftis, A. Doulamis, and N. Doulamis, "Multiclass Confusion Matrix Reduction Method and Its Application on Net Promoter Score Classification Problem," *Technologies,* vol. 9, no. 4, pp. 81, 2021.

[51] Adeem, G., ur Rehman, S. and Ahmad, S., 2022. Classification of Citrus Canker and Black Spot Diseases using a Deep Learning based Approach. [52] J. Davis, and M. Goadrich, "The relationship between Precision-Recall and ROC curves." pp. 233-240.

[53] Arooj, S., Rehman, S.U., Imran, A., Almuhaimeed, A., Alzahrani, A.K. and Alzahrani, A., 2022. A Deep Convolutional Neural Network for the Early Detection of Heart Disease. Biomedicines, 10(11), p.2796.

[54] Asad, R., Arooj, S. and Rehman, S.U., 2022. Study of Educational Data Mining Approaches for Student Performance Analysis. Technical Journal, 27(01), pp.68-81.

[55] C. Feng, T. Li, and D. Chana, "Multi-level anomaly detection in industrial control systems via package signatures and LSTM networks." pp. 261-272.

[56] S. S. Jagtap, S. S. VS, and V. Subramaniyaswamy, "A hypergraph based Kohonen map for detecting intrusions over cyber–physical systems traffic," *Future Generation Computer Systems,* vol. 119, pp. 84-109, 2021.

[57] A. Mansouri, B. Majidi, and A. Shamisa, "Metaheuristic neural networks for anomaly recognition in industrial sensor networks with packet latency and jitter for smart infrastructures," *International Journal of Computers and Applications,* vol. 43, no. 3, pp. 257-266, 2021.

[58] D. Nedeljkovic, and Z. Jakovljevic, "CNN based Method for the Development of Cyber-Attacks Detection Algorithms in Industrial Control Systems," *Computers & Security*, pp. 102585, 2021.

[59] N. Sameera, and M. Shashi, "Deep transductive transfer learning framework for zero-day attack detection," *ICT Express,* vol. 6, no. 4, pp. 361-367, 2020.

[60] B. Böse, B. Avasarala, S. Tirthapura, Y.-Y. Chung, and D. Steiner, "Detecting insider threats using radish: A system for real-time anomaly detection in heterogeneous data streams," *IEEE Systems Journal,* vol. 11, no. 2, pp. 471-482, 2017.

[61] O. Lo, W. J. Buchanan, P. Griffiths, and R. Macfarlane, "Distance measurement methods for improved insider threat detection," *Security and Communication Networks,* vol. 2018, 2018.

[62] M. N. Al-Mhiqani, R. Ahmad, Z. Z. Abidin, K. H. Abdulkareem, M. A. Mohammed, D. Gupta, and K. Shankar, "A new intelligent multilayer framework for insider threat detection," *Computers & Electrical Engineering*, pp. 107597, 2021.

[63] T. Zoppi, and A. Ceccarelli, "Prepare for trouble and make it double! Supervised–Unsupervised stacking for anomaly-based intrusion detection," *Journal of Network and Computer Applications,* vol. 189, pp. 103106, 2021.

[64] F. Van Wyk, Y. Wang, A. Khojandi, and N. Masoud, "Real-time sensor anomaly detection and identification in automated vehicles," *IEEE Transactions on Intelligent Transportation Systems,* vol. 21, no. 3, pp. 1264-1276, 2019.

[65] M. Alauthman, N. Aslam, M. Al-Kasassbeh, S. Khan, A. Al-Qerem, and K.-K. R. Choo, "An efficient reinforcement learning-based Botnet detection approach," *Journal of Network and Computer Applications,* vol. 150, pp. 102479, 2020.

[66] Y. N. Kunang, S. Nurmaini, D. Stiawan, and A. Zarkasi, "Automatic features extraction using autoencoder in intrusion detection system." pp. 219-224.

[67] M. Sarhan, S. Layeghy, M. Gallagher, and M. Portmann, "From Zero-Shot Machine Learning to Zero-Day Attack Detection," *arXiv preprint arXiv:2109.14868,* 2021.

[68] K. Shaukat, S. Luo, V. Varadharajan, I. A. Hameed, and M. Xu, "A survey on machine learning techniques for cyber security in the last decade," *IEEE Access,* vol. 8, pp. 222310-222354, 2020.