International Journal of Emerging Technology and Advanced Engineering
Website: www.ijetae.com (E-ISSN 2250-2459, Scopus Indexed, ISO 9001:2008 Certified Journal, Volume 12, Issue 11, November 2022)
Manuscript Received: 02 July 2022, Received in Revised form: 27 August 2022, Accepted: 15 October 2022 DOI: 10.46338/ijetae1122_15

Detection of Anomaly using Machine Learning: A Comprehensive Survey

Deepak T. Mane[1], Sunil Sangve[2], Gopal Upadhye[3], Sahil Kandhare[4], SaurabhMohole[5], Sanket Sonar[6], Satej Tupare[7]
[1,2,4,5,6,7]JSPM's Rajarshi Shahu College of Engineering, Pune-411033, Maharashtra, India.
[3]Pimpri Chinchwad College of Engineering, Pune-411044, Maharashtra, India

***Abstract*--Anomaly detection is an important element in the domain of security. As a result, we undertook a literature review on ML algorithms that identify abnormalities. In this paper, we are presenting a review of the 101 research articles describing ML techniques for anomaly detection published between 2015 - 2022.The goal of this paper is to review research papers that have used machine learning to develop anomaly detection algorithmThe forms of anomaly detection examined in this study include system log anomaly detection, network anomaly detection, cloud-based anomaly detection, and anomaly detection in the medical profession. After assessing the selected research articles, we present more than 10 applications of anomaly detection. Also, we have shared a range of datasets used in anomaly detection research, in addition to revealing 30+ new ML models employed in anomaly detection. We have discovered 55 new datasets for anomaly detection. We've noticed that the majority of researchers utilize real-life datasets and an unsupervised learning technique to detect anomalies. Many ML methods may be applied in this subject, so we present a summary of all work done in the previous six years.**

***Keywords* Intrusion detection, Artificial intelligence, Anomaly detection, security, Machine learning.**

I. INTRODUCTION

The process of discovering patterns in data that do not correspond to a model of typical behavior is known as anomaly detection. Anomaly detection aims at finding or detecting unordinary events from given data. Anomaly detection is a useful approach in a variety of sectors, including transportation, public safety, and property protection. Data is currently used to make the bulk of judgments. In the security domain, detecting outliers is just the beginning. To find a proper solution, you must first assess whether the outlier is a security risk and then identify the core source of the anomaly.

Before we can discuss anomaly detection, we must first define an anomaly. In general, an anomaly is something that deviates from the norm: a departure, an exception. It is an event that does not fit into the pattern and so appears suspicious.

Some examples of anomalies are

1. Sudden rapid drop or increase in temperature.
2. Someone leaving a suspicious item in a public place.
3. An event like chain snatching, a noisy fight in a public place.
4. Fire or disastrous events.

We generally want to catch all the abnormalities, especially in public places, for a variety of reasons. As a result, we want a software program that will run smoothly, consistently, and rapidly in order to detect any outliers in the current scenario. The process of discovering and recognizing irregularities is known as anomaly or outlier detection.

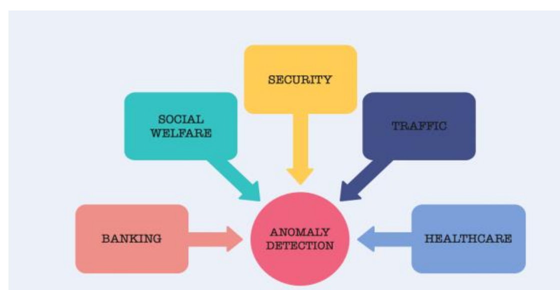

Fig.1. Applications of Anomaly Detection

Banking Sector: Anomaly detection is useful in the banking and finance industries for a variety of reasons. Identifying address fraud and identifying flaws in statistics data are only a few of them.

Social Welfare Sector: Anomaly detection may be utilized for social welfare in a variety of situations, including insurance fraud, document plagiarism, and social benefits fraud.

Security: In the subject of security, the detection of anomalies has several applications. Robberies, public attacks, knife and gun assaults, chain snatching, and other crimes are only a few of them.

Traffic: Anomaly detection may be employed in a variety of contexts, including accident detection, dangerous conditions in traffic, and so on.

International Journal of Emerging Technology and Advanced Engineering
Website: www.ijetae.com (E-ISSN 2250-2459, Scopus Indexed, ISO 9001:2008 Certified Journal, Volume 12, Issue 11, November 2022)

Accident detection has recently been a popular domain for applying computer vision to solve difficult issues such as providing timely first-aid services without the need for any human interaction.

Healthcare: A few uses in the healthcare sector include, detecting diseases using Al, avoiding fraud in drug management/prescriptions, and identifying irregularities in doctor and insurance company billing. While measures are intrinsically restricted in processing capacity and energy resources, they are also susceptible to several abnormalities, such as anomalous readings caused by erroneous calibration, electromagnetic interference, patients with perspiration, and so on, all of which can occur naturally. [2]

Recently anomaly detection using machine learning has become popular among researchers. These approaches are used to create a system that classifies normal and anomalous events into two classes.Based on various aspects anomaly detection is broadly classified into 3 major categories. Those three main categories are as follows:

1. *Supervised anomaly detection:* For supervised anomaly identification, both the normal and anomalous training datasets are including as labels. This approach aims at creating a prediction model for both anomalous and normal events and then comparing the 2 models. But there are 2 main disadvantages of this method. Firstly, the number of abnormalities in the training set is substantially lower when compared to typical occurrences. Secondly, precise and representative labels are difficult to come by, particularly for the anomaly class.
2. *Semi-supervised anomaly detection:* Only regular class cases are used in this training. As a result, everything that does not fit into the conventional category is labeled as anomalous. Semi-supervised approaches assume that just the normal class has been tagged in the training data. They are more prevalent than supervised approaches since they do not need anomalous class labels.
3. *Unsupervised anomaly detection:* The approaches do not require training datasets in this scenario. As a result, such methodologies suggest that an anomalies. However, if the assumption fails, this strategy has a significant false alarm rate. Before applying any machine learning algorithm, we must filter dattypical cases are far more prevalent in testing datasets tha first. To parse the ASCII files and generate feature statistics, we employ a database filter. More complex activities required the usage of PL/SQL code.

A variety of strategies can be used to make input data more suitable for machine learning algorithms. Feature discretization and feature selection are two of the most effective strategies for this purpose. [3]

For anomaly detection and classification dimensionality of data must be reduced. The high-dimensional dataset is a big challenge for applying an anomaly detection model. This is because of the following reasons: [8].

(i) Exponential search space – As input dimensionality increases, the number of possible feature subspaces expands exponentially, this causes exponential search space.

(ii) Data-snooping bias – When the dataset used is high-dimensional, every point looks to be an outlier. If there are enough alternative subspaces, the model can detect at least one feature subspace for each location, causing it to appear as an anomaly.

(iii) Irrelevant features — Data becomes noisy when irrelevant features are present, hiding the underlying anomalies. The main challenge then is to find a data subspace that highlights the important properties.

Intrusion detection is one of the most important applications of anomaly detection. Intrusion is a purposeful, unofficial, and unlawful attempt to gain access to, modify, or seize control of a computer system to make it unreliable or useless. The process of discovering and evaluating numerous events that occur in a system or network for the existence of intrusion is called intrusion detection. A deep learning algorithm can extract better representations from data, allowing for significantly better models to be created.[17] Intrusion Detection Systems (IDS) may be classed into three categories based on how intrusion is detected. These three groups are Signature-based, anomaly-based, and hybrid. Anomaly detection systems employ statistical approaches including clustering.[21] The problem of “Curse of dimensionality in anomaly detection”:Anomaly detection seeks to find anomalous patterns that deviate from the overall data, known as outliers. Anomaly detection is hampered by high dimensionality when a number of features are more than the data needed for the generalization of results, leading to data sparsity, wherein data points are more spread and separated. This data sparsity is caused by extraneous variables or a high noise level of several unimportant qualities, which obscure the underlying anomalies. This is commonly referred to as the "curse of dimensionality."

It is a barrier for many anomaly detection strategies that address large dimensionality and fail to sustain the effectiveness of traditional approaches in machine learning for anomaly or intrusion detection.[103]

II. LITERATURE REVIEW

Anomaly detection is a critical subject that has been studied and implemented in a variety of industries. Many anomaly detection methods have been developed expressly for certain applications, while others are more universal. Ali BouNassif [104], for example, offered a systematic assessment and review of anomaly detection approaches and applications. In which they reviewed several ML approaches such as supervised learning, semi-supervised learning, ensemble learning, deep learning, etc. Furthermore, the report discusses numerous applications of anomaly detection. Examples are Cyber security and anomaly detection, medical anomaly detection, detection fraud detection, anomaly detection using computer vision, image processing, and anomaly detection, network anomaly detection, cloud computing anomaly detection, IoT anomaly detection, and so on.

2.1 Methodology

The process we utilized to conduct this research is depicted in Figure 2. To begin, we wrote down our survey needs, including answers to questions such as: what is the need for this survey? What flow will we use to conduct the survey? How will the survey be conducted? And so on. Then we started downloading research as well as survey papers on the related topic between the years 2015-2022. The entire generated list was then uploaded to a single Google Drive account, so all coworkers could read the papers and conduct the survey analysis.

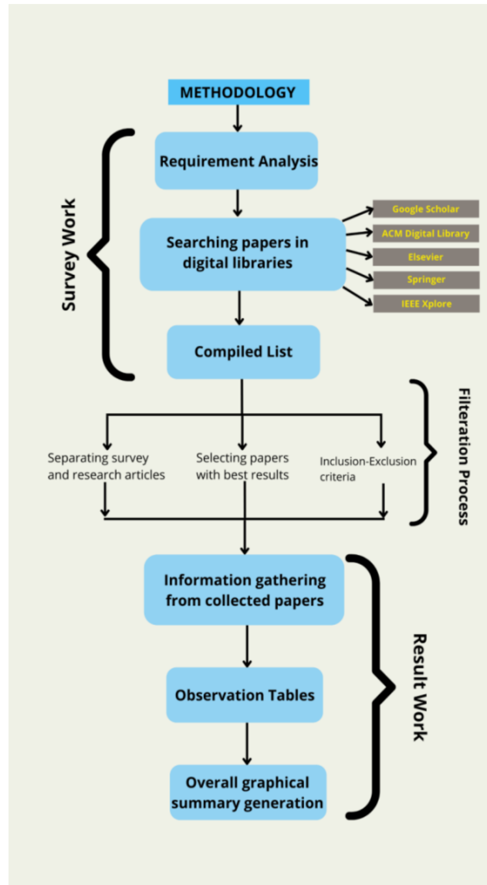

Fig. 2. The methodology used for the survey

2.1.1 Selection Criteria For Survey Papers:

On the basis of the previously given search criteria, we initially gathered 170 documents. We later screened those publications to ensure that our review contained only papers that were relevant to the subject. The following details the filtration and selection procedures:

International Journal of Emerging Technology and Advanced Engineering
Website: www.ijetae.com (E-ISSN 2250-2459, Scopus Indexed, ISO 9001:2008 Certified Journal, Volume 12, Issue 11, November 2022)

Step 1: Remove all of the duplicate articles that were gathered from the various digital libraries.

Step 2: Selecting the papers which had all the important parameters like results dataset used etc.

Step 3: Apply quality evaluation guidelines to ensure that the only papers that are eligible to provide the best response to our research objectives are included.

Step 4: Look up for more similar papers and then repeat Step 3 for adding new articles to the survey.

Finally, we created observation tables and arrived at a final conclusion.

2.2 Discussion Of Survey

2.2.1 Objectives (OBJ) of The Reveiw

OBJ-1: To specify the machine learning methods used in the anomaly detection process.

OBJ-2: To present the percentage of research papers gathered that make use of supervised, unsupervised, or semi-supervised, learning methods.

OBJ-3: To demonstrate the level of accuracy for each machine learning method used in anomaly detection.

2.2.2 Observation Tables

Table 1 displays each paper ID in the first column, followed by all relevant information such as the year of publication, method type and algorithm used, anomaly detection types such as unsupervised supervised, or semi-supervised, the dataset used, performance metrics such as AUC, precision, accuracy, F-score, recall, TPR, FPR, and various time evaluation metrics such as training, testing, execution, and computational time.

We looked at the datasets used by the ML models for anomaly detection employed in the chosen research articles as creating an ML model depends on the dataset.

ML models should be assessed with performance metrics in addition to datasets. In 101 papers, the performance metrics of the suggested models were presented in an understandable manner.

TABLE 1.
Observations on Results, methods, and Datasets.

P. ID	Year	Method Used	Dataset Used	Performance Matrices	Value
P1	2015	TYPE: Supervised ALGORITHM: Random Forest	Three types of KPI data (PV, #SR, and SR)	Precision:	89%
P2	2015	ALGORITHM: Random Forest + Linear Regression	Real medical dataset	MAE:	0.0145
				Test Time	1.43 sec
P3	2015	ALGORITHM:Support Vector Machine	Slammer, Nimda, Code Red I.	F-score:	0.88
				ROC area	0.907
P4	2015	TYPE: Supervised ALGORITHM:Fuzzyfied learning	NSL-KDD	Accuracy:	94.1%
				Error ratio:	0.059
P5	2015	TYPE:Un-Supervised	Real life dataset	Accuracy:	85%
P6	2015	ALGORITHM: Ensemble Learning	NSL-KDD	F-score:	98
				ROC area:	99.6
P7	2015	TYPE: SUPERVISED ALGORITHM: Radial Base Function	ADFA-LD	DR: FAR:	78% 21%
P8	2015	TYPE: SUPERVISED ALGORITHM:Naive Bayes	KDD"99	Accuracy: (NB)	78.941
P9	2016	ALGORITHM: ANN andFuzzy Means clustering	DARPA's KDD cup dataset 1999	Precision:	99.94
				Recall:	97.2

International Journal of Emerging Technology and Advanced Engineering
Website: www.ijetae.com (E-ISSN 2250-2459, Scopus Indexed, ISO 9001:2008 Certified Journal, Volume 12, Issue 11, November 2022)

P10	2016	TYPE:UN-SUPERVISED ALGORITHM:DBN with ISVM	Synthetic dataset: 1. Banana 2. Smiley	Area under	0.9863
				Accuracy:	0.0625
P11	2016	TYPE: SUPERVISED ALGORITHM:SVM + Random Forest	NSL-KDD99 dataset	Accuracy:	97.5
P12	2016	TYPE: SUPERVISED:SVM	Golden Dataset	Accuracy:	94% to 97%
P13	2016	TYPE: UNSUPERVISED ALGORITHM:deep belief network	DARPA KDDCUP'99 dataset	Accuracy:	97.90%
P14	2016	TYPE: UNSUPERVISED ML Method: kernel methods -EXPOSE	Smaller benchmark Datasets and KDD'99 cup	Accuracy:	1.75
P15	2016	TYPE: SUPERVISED ML Method: decision tree	KDD Cup 1999	Accuracy:	90%
P16	2016	METHOD_1:Neural network Neuro-Fuzzy method	Real time data from Denmark	Accuracy: (METHOD_1)	86.72%
P17	2017	TYPE: SUPERVISED ML Method:Ensemblelearning	MAWILab dataset	AUC:	0.999
				FPR:	5%
P18	2017	TYPE: SUPERVISED AND UNSUPERVISED Hybrid: MetaPaging, REPTree, J48, Random Tree, AdaBoostM1, Naive Bayes	NSL-KDD	Accuracy:	99.9
P19	2017	TYPE: SUPERVISED ALGORITHM:RNN	KDD dataset	Detection rate:	97.09%
				Accuracv:	81.29%
P20	2017	ALGORITHM: Convolutional neural network	Visible/Infrared Imaging Spectrometer dataset	Accuracy:	98.28
				Testing time:	483 sec
P21	2017	ALGORITHM:SVM, DNN	Secure Water Treatment data	Precision:	98.2
				F-score:	80.2
P22	2017	TYPE: SUPERVISED ALGORITHM:Logistic RegressionRF	UNSW	Accuracy:	99%
P23	2017	TYPE: UNSUPERVISED ALGORITHM:Hierarchical Temporal	NAB dataset	Latency (ms):	11.3
				Benchmark Score:	70.1
P24	2017	TYPE: SUPERVISED + UNSUPERVISED ALGORITHM:DeepLog	HDFS Log dataset.	Prec.:	95%
				F-score	96%
P25	2018	TYPE:SEMISUPERVISED	Kyoto University's 2006+	Detection Rate:	0.9336

		ALGORITHM:Adaptive Network		Accuracy:	0.9666
P26	2018	ALGORITHM: DONN + LSTM	NSLKDD	Accuracy (ACC):	89%
P27	2018	ALGORITHM: CNN, STM, and DNN	Yahoo S5 Webscope Dataset	Accuracy:	98.60%
				Recall:	89.70%
P28	2018	TYPE: SUPERVISED ALGORITHM:Extreme Learning	ISCX-IDS 2012 dataset	Detection Rate:	91%
				Misclassification	9%
P29	2018	TYPE:SEMISUPERVISED ALGORITHM:OCSVM	KDD andUNSW-NB15	Accuracy	98.59
P30	2018	ALGORITHM:RF, DNN, autoencoder	CIDDS-001	Accuracy	99.99%
P31	2018	TYPE: SUPERVISED ALGORITHM:SVM	TCP Data Set (Synthetic data set)	Accuracy	0.999 701
				F-score:	0 999 851
P32	2018	TYPE: SUPERVISED ALGORITHM:Boosted Decision Tree	Real-world Dataset	Accuracy:	0.928
P33	2018	TYPE: UNSUPERVISED ALGORITHM:RNN+ LSTM	(LANL) cyber security dataset.	Area under curve - Word:	0.984
P34	2018	TYPE: SUPERVISED & UNSUPERVISED ALGORITHM:CNN, DBN	CTU dataset and real time dataset	Precision:	0.95
				Recall:	0.38
P35	2018	TYPE:SEMISUPERVISED ALGORITHM:Gaussian model	Real life dataset of cellularnetwork	Accuracy:	92.79%
				Error Rate:	7.21%
P36	2018	TYPE: UNSUPERVISED ALGORITHM:Isolation Forest	CSIC 2010 data set	Accuracy:	88.32
				Detection Rate:	88.34
P37	2018	TYPE: UNSUPERVISED ALGORITHM: Donut(Based on VAE)	Datasets from business KPIs	F-score:	0.75 to 0.9
P38	2018	TYPE: UNSUPERVISED ALGORITHM:One class SVM	Real life dataset	Training Time:	220 sec
P39	2019	TYPE: UNSUPERVISED ALGORITHM:Autoencoder	NSL-KDD	Overall Accuracy:	92.91%
P40	2019	TYPE: SEMISUPERVISED ALGORITHM:GAN	CIFAR10 Dataset. MNIST Dataset	AUC:	0.882
P41	2019	TYPE: SUPERVISED ALGORITHM:RBM and SVM	Real-time and benchmark datasets	Acc:	99.98
P42	2019	TYPE: SUPERVISED & UNSUPERVISED ALGORITHM:DCRM and DCM	testbed	FNR:	0.91
				FPR:	0.07
P43		ALGORITHM:RF, and ANN	DS2OS traffic traces	Acc.:	99.40%

	2019			Prec.:	99%
P44	2019	TYPE: UNSUPERVISED ALGORITHM:GAN	Real-life dataset	AUC:	0.641
P45	2019	TYPE:SEMISUPERVISED ALGORITHM:Neural Network	Real-world dataset	ROC:	0.916+-0.004
				PR:	0.574+-0.008
P46	2019	TYPE:UNSUPERVISED ALGORITHM:Auto encoder ANN.	benign IoT traffic	Precision:	0 996
				Recall:	0 999
P47	2019	ALGORITHM:4 single classifiers RF, GBDTLinearRegression, kNN andgradientboosting Decision Tree.	Real life dataset from a financial company.	Precision:	0.8803
				Recall:	0.7017
				F-score:	0.8376
P48	2019	ALGORITHM:Adversarialautoencoder (AAE)	Synthetic data, cifar-10, Pixabay.	Area Under Precision Recall Curve	1
P49	2019	TYPE:UNSUPERVISED ALGORITHM:Random Forest	NSL-KDD	Precision:	0.9992
				Recall:	0.9969
P50	2019	ALGORITHM:Neural network, Analogous Particleswarm optimization	Real life dataset	Precision:	95.70%
				Svstem	5.60%
P51	2019	TYPE: SUPERVISED ALGORITHM:XGBoost	Data from real world network environment	Precision:	0.8064
				Recall:	0.7823
P52	2019	ALGORITHM:Mask R-CNN + Centroid Tracking	Vehicle collision footage compiled from YouTube	Detection Rate:	71%
				False Alarm Rate:	0.53%
P53	2019	TYPE: UNSUPERVISED ALGORITHM:PCA	Data set of benign IoT traffic that is freely available to the	F1 Score (Attack):	0.998
P54	2019	TYPE: UNSUPERVISED ALGORITHM:LOF, HBOS, KNN	PQ data (non-transformed)	Highest TPR (KNN):	60%
P55	2019	TYPE: SUPERVISED + UNSUPERVISED ALGORITHM:CNN and LSTM	KDD99	Training	0.99
				Accuracy:	0.925
P56	2019	TYPE: SEMI-SUPERVISED ALGORITHM: PU learning	Dataset from real life	Detection Rate:	85.00
				AUC:	0.8711
P57	2019	TYPE: SUPERVISED ALGORITHM:CNN-Xception	Real life (49 subjects)	Accuracy:	96.05%
				AUC:	0.99
P58	2019	TYPE: UNSUPERVISED	ImageNet dataset	AUC:	0.8067

International Journal of Emerging Technology and Advanced Engineering
Website: www.ijetae.com (E-ISSN 2250-2459, Scopus Indexed, ISO 9001:2008 Certified Journal, Volume 12, Issue 11, November 2022)

		ALGORITHM:UAD-GAN		Mis-detection:	6.23%
P59	2019	TYPE: SUPERVISED + UNSUPERVISED ALGORITHM:1. K-Means+ HMM	IoTPOT dataset.	Accuracy:	0.9209
				ROC-AUC:	0.8710
P60	2019	TYPE: SEMI-SUPERVISED ALGORITHM:AGNN,GCN	Real life dataset	Accuracy: (GCN)	84.94 ± 2.30
				Accuracy: (AGNN)	84.25 ± 3.51
P61	2019	TYPE: SUPERVISED ALGORITHM:RNN	DCASE 2019 SED dataset	F1-Score	23.79%
P62	2019	TYPE: SUPERVISED ALGORITHM:J48, NaiveBayes	Real-life dataset	Average accuracy	Above 85%
P63	2020	TYPE:UNSUPERVISED	μPMU	Accuracy:	96%
P64	2020	TYPE:SUPERVISED ALGORITHM:RF and regression	UNSW-NB15	Acc:	95.73
P65	2020	TYPE:SUPERVISED ALGORITHM:autoencoder (AE)	real life dataset	Mean Absolute Error:	2.9
				Mean Squared	15.8
P66	2020	TYPE:SUPERVISED ALGORITHM:Skip-gram and k-means	real life dataset	Accuracy:	98
P67	2020	ALGORITHM:Locally Weighted Projection Regression	Real life dataset	Accuracy:	91%
				AUC:	0.54
P68	2020	TYPE:UNSUPERVISED ALGORITHM: OCSVM And Subspace Clustering	NSL-KDD dataset	Detection rate:	90%
				False alarm rate:	9.05%
P69	2020	TYPE: SUPERVISED + NSUPERVISED ALGORITHM:Light Gradient Boosted Machine, OC SVM andIsolation Forest.	HAI dataset	Accuracy:	99%
P70	2020	ALGORITHM:SVM, DT, k-means clustering and k-nearest neighbors.	Publicly available datasets	Accuracy:	0.99
				F1-Score:	0.99
P71	2020	ALGORITHM: 1) Feature extraction: Convolutional Autoencoder and GAN	UCF crime video dataset.	Accuracy: (LR)	97
P72	2020	ALGORITHM:Logistic Regression,KNN, Decision Tree, SVM, Random Forest, XGBoost and Adaboost	History data in Distributed Control System.	Best Testing Accuracy: (SVM)	98.58%
				Precision: (RF)	99.67%

International Journal of Emerging Technology and Advanced Engineering
Website: www.ijetae.com (E-ISSN 2250-2459, Scopus Indexed, ISO 9001:2008 Certified Journal, Volume 12, Issue 11, November 2022)

P73	2020	TYPE: SUPERVISED ALGORITHM:CNN and RCNN	Real life created dataset	Best Accuracy: (RCNN)	84.6%.
P74	2020	TYPE: UNSUPERVISED ALGORITHM: GAN, ALAD	1) UNSW-NB	Prec.:	0.8743
				Recall:	0.8583
P75	2020	TYPE: UNSUPERVISED ALGORITHM:KDetect	Own Dataset	FP:	0.31
				Recall:	0.98
P76	2020	TYPE: UNSUPERVISED ALGORITHM:Clustering	JoCAD- Synthetically injected journal-level citation anomaly dataset	Precision: (%)	100
				Recall: (%)	75.45
P77	2020	TYPE:UNSUPERVISED ALGORITHM: USAD Algorithm based on auto encoder	Orange's proprietary data.	Precision:	0.7448
				Recall:	0.6428
				F1-Score:	0.6901
P78	2021	TYPE:UNSUPERVISED ALGORITHM:GANs	CUHK Avenue dataset	AUC:	89.2%
			Shanghai Tech datasets	AUC:	75.7%
P79	2021	ALGORITHM: Video Vision Transformer	CUHK Avenue dataset	AUC:	0.870
P80	2021	TYPE: UNSUPERVISED	NSL-KDD	DR:	0.90
P81	2021	TYPE:UNSUPERVISED ALGORITHM:Auto-Encoder	UNB dataset	Precision:	0.64
				Recall:	0.48
P82	2021	ALGORITHM:The Human-machine Cooperation Framework.	3 datasets for monitoring that are freely available to the public	AUC:	90.6 to 94.2%
P83	2021	ALGORITHM:DT, KNN, Random Forest, AdaBoost, K-Nearest Neighbors, SVM, Gaussian Naive Bayes , Multinomial Naive Bayes, Multi-layer Perceptron.	Subset of stored data	Prec.: (RF)	0.996
				Recall: (RF)	0.991
				F-measure: (RF)	0.994
P84	2021	TYPE: UNSUPERVISED ALGORITHM:Iterative classifier, N-over-D	CICIDS2017, CAIDA UCSD "DDoS Attack 2007" dataset.	ACC:	0.9713
				Prec.:	0.9968
P85	2021	TYPE:UNSUPERVISED ALGORITHM:Deep learning, Variational Auto encoder	real-world datasets, 1) BlogCatalog and 2) Flickr.	AUC	75.21
P86	2021	TYPE: SUPERVISED ALGORITHM:Decision Tree, AdaBoost, Artificial Neural Network, Deep learning.	China's State Grid Corporation provided the data (SGCC).	Accuracy: (DANN)	94.97%
				AUC: (DANN)	0.8703
P87	2021	TYPE:UNSUPERVISED	BGL dataset	F1-score:	92.6%

International Journal of Emerging Technology and Advanced Engineering
Website: www.ijetae.com (E-ISSN 2250-2459, Scopus Indexed, ISO 9001:2008 Certified Journal, Volume 12, Issue 11, November 2022)

		ALGORITHM:LogTAD		AUC:	96.4%
P88	2021	TYPE:SEMI-SUPERVISED ALGORITHM:Deep LearningPLELog	HDFS Dataset	Precision:	95%
				Training Time:	43m
P89	2021	TYPE: SUPERVISED+ UNSUPERVISED ALGORITHM: LSTM, CNN	InSDN dataset	Accuracy	96.32%
				AUC	0.956
P90	2021	ALGORITHM:LR, xgboost, RF, catboost.	CIDDS-002 dataset.	Accuracy:	99.8%
P91	2021	TYPE: SUPERVISED ALGORITHM:KNN, RNN, DT, RF, SVM, MLP	CICIDS 2017 dataset	F1-score:	0.87
			CICIDS 2018 dataset	F1-score:	0.72
P92	2021	TYPE: SEMI-SUPERVISED ALGORITHM: LCR-GAN	15 tabular datasets, Image dataset	Accuracy:	0.90 to 0.96 for all different datasets.
P93	2022	TYPE:UNSUPERVISED ALGORITHM:3DCAE_mse model Customized	Real Life dataset	AUC	0.754.
P94	2022	ALGORITHMS:MIDAS	DARPA	ROC-AUC	0.98
P95	2022	ALGORITHM:ADUFS Anomaly Detection Using Feature Selection.	KDD-99	ACC (%)	98.4
P96	2022	TYPE:UNSUPERVISED + SUPERVISED METHODS:ADEPTUS (RF, CatBoost and CNN)	The Raw Dataset	Recall	42.05
				Precision	20.61
P97	2022	TYPE: SUPERVISED	real-world data	ACC:	0.7990
P98	2022	TYPE:SEMI-SUPERVISED ALGORITHM:Graph neural networks (GNN)	Two benchmark datasets 1) Ground Truth	AUC:	91.67
P99	2022	TYPE:UNSUPERVISED ALGORITHM:Clustering	AIT-LDSv1.1 real-world data	TPR:	80%
				FPR:	5%
P100	2022	TYPE: SUPERVISED ALGORITHM:SVM, Naïve Bayes, Decision tree, Logistic Regression, KNN,	Intersection Dataset.	PF (%):	81.4
				CPF (%):	85.5
				FAS (%):	96.2
P101	2022	ALGORITHM: CLAD, Deep Learning.	1) CTF_dataset	Recall:	0.882
				Precision	0.805

International Journal of Emerging Technology and Advanced Engineering
Website: www.ijetae.com (E-ISSN 2250-2459, Scopus Indexed, ISO 9001:2008 Certified Journal, Volume 12, Issue 11, November 2022)

2.2.3 Most Frequently used Dataset:

Many studies have utilized the Real-Life dataset for their study, and the NSL-KDD dataset is the second most used dataset in all of the research, as seen in Fig. 3. NSL-KDD: The KDDCUP99 dataset has been changed to create the NSL-KDD dataset. This dataset addresses some of the flaws in the KDDCUP99 dataset the key advantage of the NSL-KDD dataset is that it does not contain any redundant examples, therefore the classifiers used on it are not biased against the train set's repeated records. Each record in the NSL-KDD dataset has 41 attributes, four types of attacks, and one class label.[6] Compared to the original KDD data set, the NSL-KDD data set provides the following advantages:

1. The train set does not contain duplicate records, so the classifiers will not be skewed toward more frequent records.
2. As data is less redundant the performance is not affected by approaches that give a better detection rate for more frequent records.
3. Due to a large number of entries in the dataset, the evaluation results of diverse research projects will be accurate and comparable.

2.2.4 Machine learning models:

A. Random Forest (RF):

Random Forest is a well-known ML method that uses supervised learning. RF is a classifier that combines a decision tree on different sets of data and averages them to increase the dataset's prediction performance.The bigger the volume of decision trees in the forest, the more accurate it is and the problem of errors is avoided.

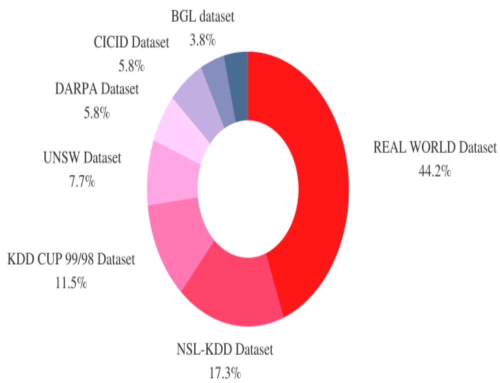

Fig. 3.Most Frequently Used Dataset in Anomaly Detection.

The RF is an ensemble learning method that employs a large number of decision trees. Its core tenet is that a set of weak learners (for example, single decision trees) can combine to become a strong learner. [1]

From Fig. 4 we can see that our task is to classify the animals. We are using a random forest for the purpose. The steps involved to give the classification result as the output are:

Step 1: In Random Forest, n random items are chosen at random from a data collection of k records.
Step 2: For each sample, separate decision trees are built.
Step 3: Every decision tree produces a result.
Step 4: For classification, the final output dependents on Majority Voting or Average.

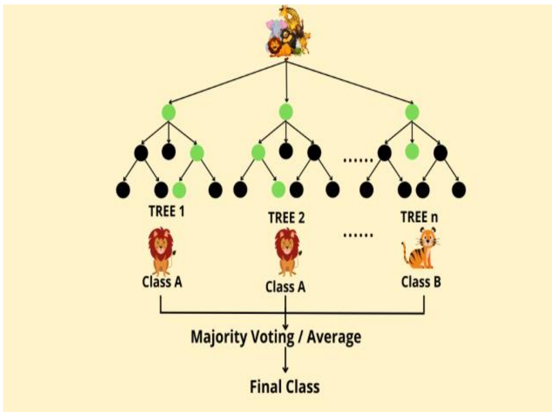

Fig. 4. Example to demonstrate how RF algorithm works

B. Support vector machine:

SVM is again an ML algorithm that is used for both regressions as well as classification. The SVM algorithm's purpose is to find the decision boundary (best-fit line) that can divide the n-dimensional area into classes so that fresh data points can be readily placed in the correct group in the future. A hyperplane denotes the ideal decision boundary.

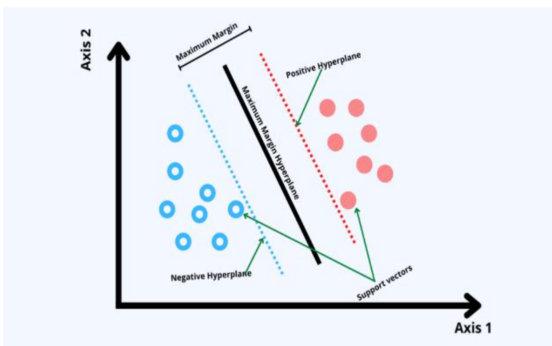

Fig. 5. SVM algorithm terminologies

Steps involved in the SVM algorithm:

- Each data item is represented as a point in an n-dimensional area (n —> number of attributes)
- Then classification is done by locating the hyperplane that best distinguishes the two classes.

C. Neural Network:

Pattern classification with neural networks entails developing an algorithm that maps input feature variables to binomial class output space.[102] A neural network is a set of optimization techniques that attempts to recognize hidden patterns or relationships in the dataset using a technique similar to how the human brain works. In this context, neural networks are systems of neurons that might be artificial or synthetic in nature.

Because NN can adjust to changing input, they can produce the best possible results without having to rethink the output criteria. The NN concept has its roots in AI and ML and is quickly gaining traction in the creation of trading systems.

D. Convolutional Neural Network

CNN has several approaches to look forward to, one of which is YOLO.This approach is highly accurate and able to outperform multiple methods, and its characteristics of real-time data analysis make it suitable for a variety of applications, including real-time vehicle detection.[105][106]

Table 2 summarizes some of the machine learning approaches covered in Table 1. It focuses on the strengths and weaknesses of respective machine learning methods.

TABLE 2.
ML Methods (Strengths or Weaknesses)

Ml Technique Used	Strength And Weakness In Anomaly Detection
Random forest	STRENGTH: The random forests approach does not require cross-validation or a test set. Because each tree is built using the bootstrap sample. WEAKNESS: RF cannot predict values that are outside the training data when used for regression, and overfitting of the dataset may occur.
LR + RF	STRENGTH: Low categorizing accuracy WEAKNESS: High detection accuracy
SVM	STRENGTH: SVM is used in classification problems as it is a supervised machine learning technique WEAKNESS: As it requires pre-acquired learning information, SVM cannot be utilized for new anomalies. As SVM has high FPR, so it is challenging to utilize it in a real-world problem.
T-SNE	STRENGTH: T-SNE data visualization is powerful, and it's simple to visualize abnormal spots with it.
Ensemble Learning	STRENGTH: It's well-known for producing more reliable outcomes. For example, bootstrap aggregating (also known as bagging) helps to avoid overfitting the training data.
DBN	STRENGTH: DBN's key benefit is its capacity to learn features, which is accomplished using layer-by-layer learning algorithms. DBN efficiently handles unlabeled data, avoiding the problems of over fitting and under fitting. WEAKNESS: Increasing the complexity of the run time
Decision Tree	STRENGTH: We can get the predicted class of an example by tracing the nodes from the tree's root based on the example's attribute values. WEAKNESS: A minor change in the data might cause a big shift in the structure of the optimal decision tree. They are usually not precise enough.
OCSVM	STRENGTH: OCSVM outperforms traditional anomaly detectors in terms of accuracy.
CNN	WEAKNESS: CNN has high time complexity so it is a time-consuming approach and it is too sluggish for patch-based solutions. Training a CNN is entirely supervised learning; hence, detecting anomalies in real-world films is hampered by the fundamental problem of training huge sets of samples from non-existent classes of anomalies.
GAN	STRENGTH: We can readily distinguish trees, streets, bicyclists, people, and parked automobiles using GANs and machines. WEAKNESS: Learning and we can even measure the distance between different items. You must supply several forms of data constantly to determine whether or not it functions correctly.
NN	STRENGTH: It's employed when a quick assessment of the supervised (labeled) target function is necessary. WEAKNESS: In the actual world, neural networks require a vast quantity of data to train, yet generating such a massive dataset is both times demanding and ineffective.
K-means	WEAKNESS: The K-means technique tightly relies on the distance between two data points, but the formula used to calculate that distance may vary, causing the results to differ. As a result, it's difficult to produce consistent results when using the K-means approach.

IJETAE
EXPLORING RESEARCH AND INNOVATIONS

International Journal of Emerging Technology and Advanced Engineering
Website: www.ijetae.com (E-ISSN 2250-2459, Scopus Indexed, ISO 9001:2008 Certified Journal, Volume 12, Issue 11, November 2022)

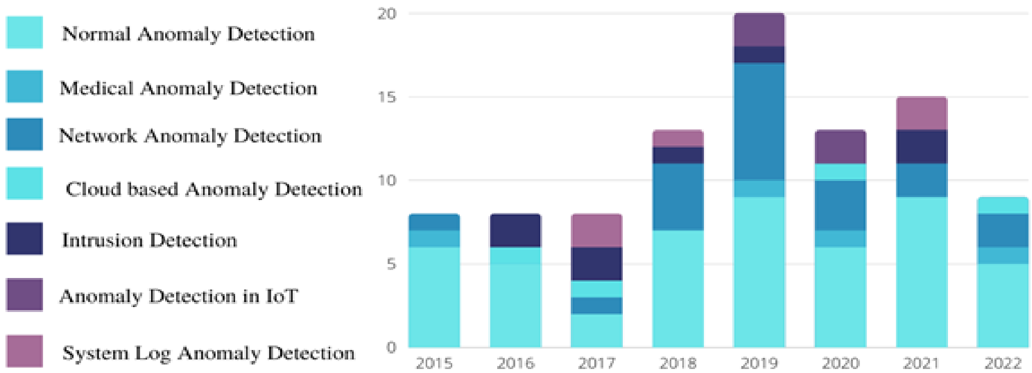

Fig. 6. Most Frequently Observed Applications of Anomaly Detectionfrom research Papers.

III. OUTCOMES OF THE SURVEY

Main objectives of this paper are to specify the machine learning methods used in the anomaly detection process and to present the percentage of research papers gathered that make use of supervised, unsupervised, or semi-supervised, learning methods. So, the outcomes of the project are shown in the form of graphical data.We gathered some data from the above observation table and created graphs and pie charts based on that data since delivering information in a visual style allows readers to get the most information from the tables 1&2.Anomaly detection has been employed in the sectors of medical applications, network anomaly, cloud computing, intrusion detection, IoT etc. which is shown in Fig. 6.Anomaly detection is frequently used for both generic anomaly detection and intrusion detection.

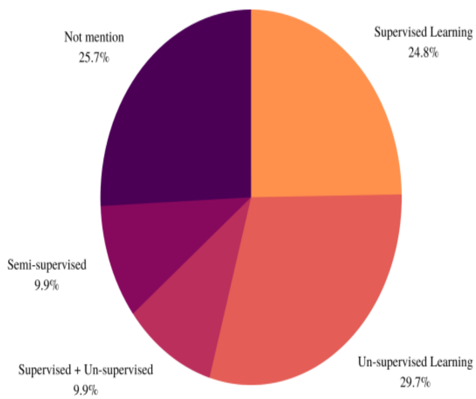

Fig. 7. Machine Learning Type Used

Fig. 7 shows that unsupervised anomaly detection was utilized in 29.7% of the articles studied, making it the most common strategy among research papers. Supervised anomaly detection was utilized by 24.8% of respondents, while supervised + unsupervised anomaly detection categorization was used by 9.9% of respondents. Semi-supervised learning, on the other hand, was mentioned in 9.9% of research papers. Surprisingly, 25.7 % of the papers in the study didn't mention what kind of machine learning for anomaly detection they have used. As can be shown, researchers were not using a combination of semi-supervised and supervised or unsupervised learning.

The most often utilized techniques from Fig. 8 are Random Forest and Support Vector Machine.

The main machine learning methods which are used in anomaly detection are classification, optimization, clustering, and regression.After analyzing various research studies in the field of anomaly detection, from the Fig. 8, we had discovered that many times on different anomaly datasets, Random forest given the best performance.

IJETAE
EXPLORING RESEARCH AND INNOVATIONS

International Journal of Emerging Technology and Advanced Engineering
Website: www.ijetae.com (E-ISSN 2250-2459, Scopus Indexed, ISO 9001:2008 Certified Journal, Volume 12, Issue 11, November 2022)

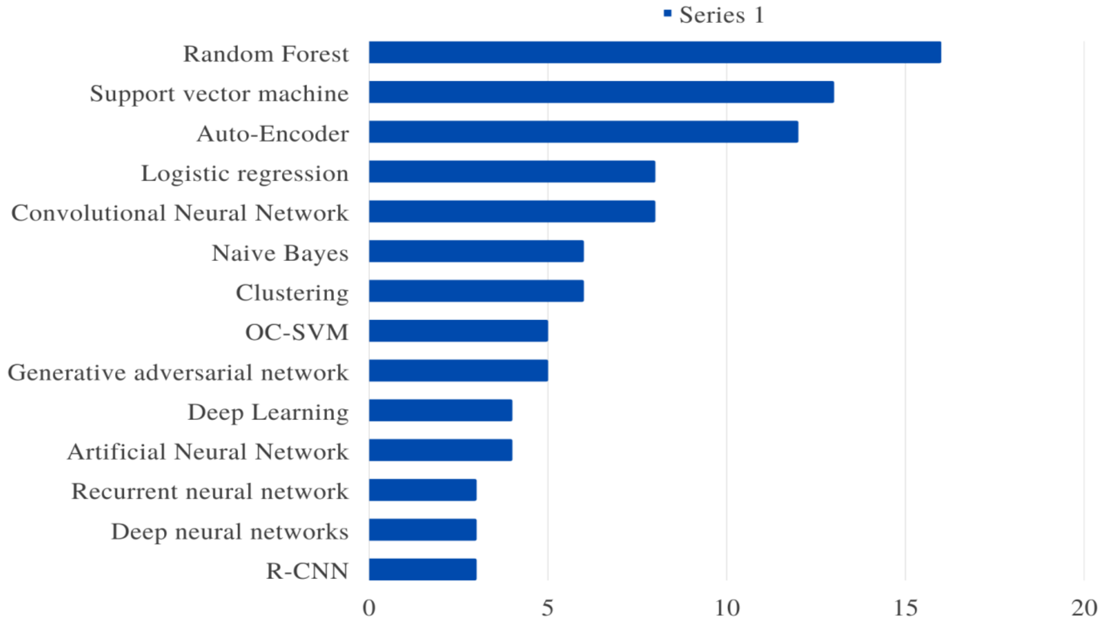

Fig. 8. Machine Learning Techniques Used for anomaly detection

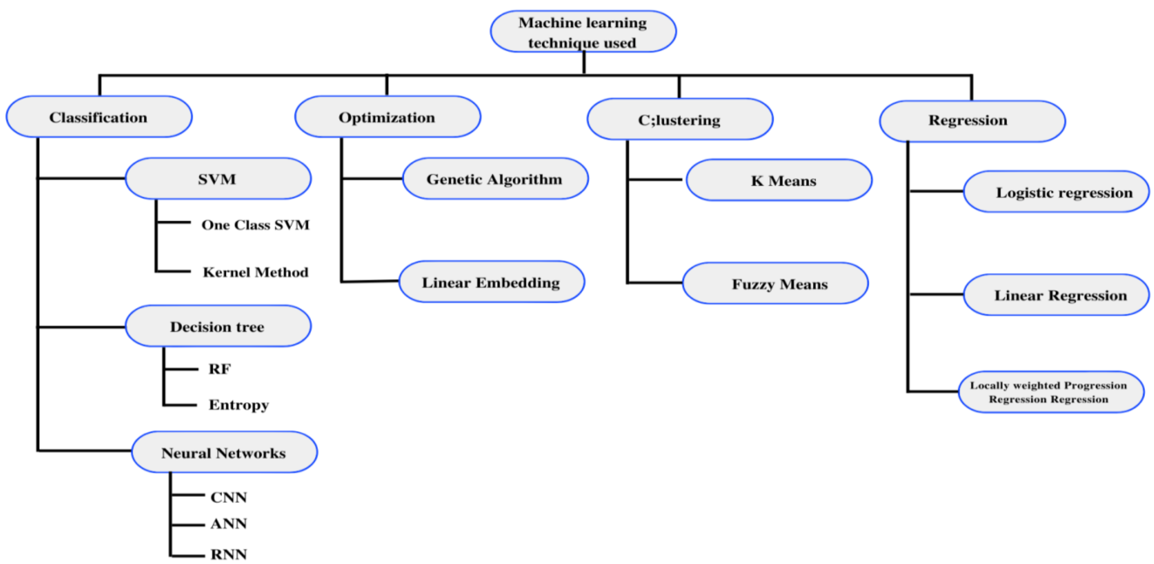

Fig. 9. All Machine Learning Techniques Used (Tree Diagram)

IV. LIMITATIONS OF SURVEY

The new researcher can save time and effort by using our survey report instead of reading irrelevant research articles. As a result, researchers will be able to eliminate mistakes in the early stages of their study and will have a clear route for conducting a good systematic survey on the issue of anomaly detection.

However, in our study, we used a restricted number of studies that we thought were important. This review is limited from the perspective of the time frame considered as well as several research papers reviewed. However, we believe that depending on more sources would have enhanced this evaluation.

International Journal of Emerging Technology and Advanced Engineering
Website: www.ijetae.com (E-ISSN 2250-2459, Scopus Indexed, ISO 9001:2008 Certified Journal, Volume 12, Issue 11, November 2022)

V. CONCLUSION

In this extensive literature analysis, anomalies were discovered using machine learning approaches (ML). The research focused on publications published between 2015 and 2022. For this study, we looked at 101 papers from various journals and conferences. In this research paper, we have discovered that anomaly detection is primarily used in major applications: medical, network anomalies, cloud computing, intrusion detection, Internet of Things, and traffic domain. The majority of researchers employ real-life datasets, with NSL-KDD being the most commonly used pre-existing dataset. Furthermore, most researchers utilized a technique of supervised anomaly detection. The most often used algorithms for anomaly identification are RF and SVM. Moreover, we discovered that 44.2 % of researchers employed real-world datasets in their studies for their models. Finally, we noticed that unsupervised anomaly detection was used in 29.7% of the journal articles we examined, making it the most prevalent strategy among research papers. The second most commonly used strategy was supervised anomaly detection, which was used in 24.8 percent of the publications. We have also discovered that most researchers are more interested in building their own hybrid anomaly detection model than using traditional ones.Based on the findings of this study, we suggest to new researchers that they should undertake further research on machine learning studies of anomaly detection to learn more about ML model efficacy and performance under various circumstances such as varying related factors, using novel and own datasets, etc. Researchers can also use ML models to provide a common structure for addressing this challenge. We have seen that this sector needs improvement. We also noticed that several researchers used outdated datasets in their studies so a new study should be conducted using newly created datasets. From our research, we conclude that this subject requires research on an integrated model to handle the problem of several anomalous events occurring at the same time, and in the future, we can expect a lot more progress in the field.

REFERENCES

[1] D. Liu, Y. Zhao, H. Xu, Y. Sun, D. Pei, J. Luo, X. Jing, and M. Feng, ''Opprentice: Towards practical and automatic anomaly detection through machine learning,'' in Proc. Internet Meas. Conf., Oct. 2015, pp. 51–78, doi: 10.1145/2815675.2815679

[2] G. Pachauri and S. Sharma, ''Anomaly detection in medical wireless sensor networks using machine learning algorithms,'' ProcediaComput. Sci., vol. 70, pp. 325–333, Jan. 2015, doi: 10.1016/j.procs. 2015.10.026

[3] M. Cosovic, S. Obradovic, and L. Trajkovic, ''Performance evaluation of BGP anomaly classifiers,'' in Proc. 3rd Int. Conf. Digit. Inf., Netw., Wireless Commun. (DINWC), Feb. 2015, pp. 115–120.

[4] G. D'angelo, F. Palmieri, M. Ficco, and S. Rampone, ''An uncertainty managing batch relevance-based approach to network anomaly detecion,'' Appl. Soft Comput., vol. 36, pp.408–418, Nov. 2015, doi: 10.1016/j.asoc.2015.07.029

[5] J. Lundstrom, W. O. De Morais, and M. Cooney, ''A holistic smart home demonstrator for anomaly detection and response,'' in Proc. IEEE Int. Conf. Pervas.Comput.Commun. Workshops (PerCom Workshops), Mar. 2015, pp. 330–335, doi: 10.1109/PERCOMW.2015. 7134058

[6] N. F. Haq, A. R. Onik, and F. M. Shah, ''An ensemble framework of anomaly detection using hybridized feature selection approach (HFSA),'' in Proc. SAI Intell. Syst. Conf. (IntelliSys), Nov. 2015, pp. 989–995, doi: 10.1109/IntelliSys.2015.7361264

[7] WaqasHaider, Jiankun Hu, Miao Xie," Towards Reliable Data Feature Retrieval and Decision Engine in Host-Based Anomaly Detection Systems",2015,DOI:10.1109/ICIEA.2015.7334166

[8] BhanuVrat,Nikhil Aggarwal,S. Venkatesan,"Anomaly Detection in IPv4 and IPv6 networks using machine learning"**,** Published 1 December 2015, 2015 Annual IEEE India Conference (INDICON) DOI:10.1109/INDICON.2015.7443752

[9] N. Pandeeswari and G. Kumar, ''Anomaly detection system in cloud environment using fuzzy clustering-based ANN,'' Mobile Netw. Appl., vol. 21, no. 3, pp. 494–505, Jun. 2016, doi: 10.1007/s11036- 015-0644-x

[10] S. M. Erfani, S. Rajasegarar, S. Karunasekera, and C. Leckie, ''High- dimensional and large-scale anomaly detection using a linear one-class SVM with deep learning,'' Pattern Recognit., vol. 58, pp. 121–134, Oct. 2016, doi: 10.1016/j.patcog.2016.03.028.

[11] N. Chand, P. Mishra et al., ''A comparative analysis of SVM and its stacking with other classification algorithm for intrusion detection,'' in Proc. Int. Conf. Adv. Comput., Commun.,Automat. (ICACCA), Apr. 2016, pp.1–6, doi: 10.1109/ICACCA.2016.7578859

[12] A. Kulkarni, Y. Pino, M. French, and T. Mohsenin, ''Real-time anomaly detection framework for many-core router through machine-learning techniques,'' ACM J. Emerg. Technol. Comput. Syst., vol. 13, no. 1, pp. 1–22, Dec. 2016, doi: 10.1145/2827699

[13] K. Alrawashdeh and C. Purdy, ''Toward an online anomaly intrusion detection system based on deep learning,'' in Proc. 15th IEEE Int. Conf. Mach. Learn. Appl. (ICMLA), Dec. 2016, pp. 195–200, doi: 10.1109/ICMLA.2016.0040

[14] M. Schneider, W. Ertel, and F. Ramos, ''Expected similarity estimation for large-scale batch and streaming anomaly detection,'' Mach. Learn., vol. 105, no. 3, pp. 305–333, Dec. 2016, doi: 10.1007/s10994-016-5567- 7

[15] B. Cui and S.He, ''Anomaly detection model based on Hadoop platform and weka interface,'' in Proc. 10th Int. Conf. Innov. Mobile Inter- net Services Ubiquitous Comput. (IMIS), Jul. 2016, pp. 84–89, doi: 10.1109/IMIS.2016.50

[16] R. Jain and H. Shah, ''An anomaly detection in smart cities modeled as wireless sensor network,'' in Proc. Int. Conf. Signal Inf. Process. (IConSIP), Oct. 2016, pp. 1–5, doi: 10.1109/ICONSIP.2016.7857445

International Journal of Emerging Technology and Advanced Engineering
Website: www.ijetae.com (E-ISSN 2250-2459, Scopus Indexed, ISO 9001:2008 Certified Journal, Volume 12, Issue 11, November 2022)

[17] J. Vanerio and P. Casas, ''Ensemble-learning approaches for network security and anomaly detection,'' in Proc. Workshop Big Data Anal. Mach. Learn. Data Commun.Netw., Aug. 2017, pp. 1–6, doi: 10.1145/3098593.3098594

[18] ShadiAljawarneh, MontherAldwairi, MuneerBaniYasin, Anomaly-based intrusion detection system through feature selection analysis and building hybrid efficient model, Journal ofComputational Sciencehttp://dx.doi.org/0.1016/j.jocs.2017.03.006

[19] C. Yin, Y. Zhu, J. Fei, and X. He, ''A deep learning approach for intrusion detection using recurrent neural networks,'' IEEE Access, vol. 5, pp. 21954–21961, 2017, doi: 10.1109/ACCESS.2017.2762418

[20] W. Li, G. Wu, and Q. Du, ''Transferred deep learning for anomaly detecion in hyperspectral imagery,'' IEEE Geosci. Remote Sens. Lett., vol. 14, no. 5, pp. 597–601, May 2017, doi: 10.1109/LGRS.2017.2657818

[21] J. Inoue, Y. Yamagata, Y. Chen, C. M. Poskitt, and J. Sun, ''Anomaly detection for a water treatment system using unsupervised machine learning,'' in Proc. IEEE Int. Conf. Data Mining Workshops (ICDMW), Nov. 2017, pp. 1058–1065.

[22] T. Salman, D. Bhamare, A. Erbad, R. Jain, and M. Samaka, ''Machine learning for anomaly detection and categorization in multi-cloud environments,'' in Proc. IEEE 4th Int. Conf. Cyber Secur. Cloud Comput. (CSCloud), 3rd IEEE Int. Conf. Scalable Smart Cloud (SSC), Jun. 2017, pp. 97–103, doi: 10.1109/CSCloud.2017.15

[23] Subutai Ahmad, Alexander Lavin, Scott Purdy, Zuha Agha, "Un-supervised real-time anomaly detection for streaming data," Neurocomputing (2017), doi: 10.1016/j.neucom.2017.04.070

[24] Min Du, Feifei Li, GuinengZheng, VivekSrikumar, "DeepLog: Anomaly Detection and Diagnosis from System Logs", 2017, DOI:http://dx.doi.org/10.1145/3133956.3134015

[25] D. A. Kumar and S. R. Venugopalan, ''A novel algorithm for network anomaly detection using adaptive machine learning,'' in Progress in Advanced Computing and Intelligent Engineering (Advances in Intelligent Systems and Computing), vol. 564. 2018, pp. 59–69, doi: 10.1007/978-981-10-6875-1_7

[26] S. Naseer, Y. Saleem, S. Khalid, M. K. Bashir, J. Han, M. M. Iqbal, and K. Han, ''Enhanced network anomaly detection based on deep neural networks,'' IEEE Access, vol. 6, pp. 48231–48246, 2018, doi: 10.1109/ACCESS.2018.2863036

[27] T.-Y.KimandS.-B.Cho,''Web traffic anomaly detection using C-LSTM neural networks,'' Expert Syst. Appl., vol. 106, pp. 66–76, Sep. 2018, doi: 10.1016/j.eswa.2018.04.004

[28] B. G. Atli, Y. Miche, A. Kalliola, I. Oliver, S. Holtmanns, and A. Lendasse, ''Anomaly-based intrusion detection using extreme learning machine and aggregation of network traffic statistics in probability space,'' Cognit. Comput., vol. 10, no. 5, pp. 848–863, Oct. 2018, doi: 10.1007/s12559-018-9564-y

[29] Y. Tian, M. Mirzabagheri, S. M. H. Bamakan, H. Wang, and Q. Qu, ''Ramp loss one-class support vector machine; a robust and effective approach to anomaly detection problems,'' Neurocomputing, vol. 310, pp. 223–235, Oct. 2018, doi: 10.1016/j.neucom.2018.05.027

[30] R. Abdulhammed, M. Faezipour, A. Abuzneid, and A. AbuMallouh, ''Deep and machine learning approaches for anomaly-based intrusion detection of imbalanced network traffic,'' IEEE Sensors Lett., vol. 3, no. 1, pp. 1–4, Jan. 2019, doi: 10.1109/LSENS.2018.2879990

[31] S. D. Anton, S. Kanoor, D. Fraunholz, and H. D. Schotten, ''Evaluation of machine learning-based anomaly detection algorithms on an industrial modbus/TCP data set,'' in Proc.13th Int. Conf. Availability, Rel. Secur., Aug. 2018, vol. 41, no. 9, pp. 1–41, doi: 10.1145/3230833. 3232818

[32] J. Zhang, R. Gardner, and I. Vukotic, ''Anomaly detection in wide area network meshes using two machine learning algorithms,'' Future Gener. Comput. Syst., vol. 93, pp. 418–426, Apr. 2019, doi: 10.1016/j.future.2018.07.023

[33] A. Brown, A. Tuor, B. Hutchinson, and N. Nichols, ''Recurrent neural network attention mechanisms for interpretable system log anomaly detection,'' in Proc. 1st Workshop Mach. Learn. Comput. Syst., Jun. 2018, pp. 1–8, doi: 10.1145/3217871.3217872

[34] L. F. Maimo, A. L. P. Gomez, F. J. G. Clemente, M. G. Perez, and G. M. Perez, ''A self-adaptive deep learning-based system for anomaly detection in 5G networks,'' IEEE Access, vol. 6, pp. 7700–7712, 2018, doi: 10.1109/ACCESS.2018.2803446

[35] B. Hussain, Q. Du, and P. Ren, ''Semi-supervised learning based big data- driven anomaly detection in mobile wireless networks,'' China Commun., vol. 15, no. 4, pp. 41–57, Apr. 2018, doi: 10.1109/CC.2018.8357700

[36] A.M. Vartouni, S.S. Kashi,andM.Teshnehlab,''Ananomaly detection method to detect Web attacks using stacked auto-encoder,'' in Proc. 6th Iranian Joint Congr. Fuzzy Intell. Syst. (CFIS), 2018, pp. 131–134.

[37] HaowenXuet. Al.. "Unsupervised Anomaly Detection via Variational Auto-Encoder for Seasonal KPIs in Web Applications." In WWW 2018: The 2018 Web Conference, April 23-27, 2018, Lyon, France. ACM, New York, NY, USA, 10 pages.https://doi.org/10.1145/3178876.3185996

[38] MatteoChiniet. Al. " Unsupervised Smoke Detection in Normally Smoking Environments." In International Conference on Distributed Smart Cameras (ICDSC '18), September 3–4, 2018, Eindhoven, Netherlands.ACM, New York, NY, USA, 6 pages.https://doi.org/10.1145/3243394.3243699

[39] Md. ShahanurAlamet. Al. "Memristor Based Autoencoder for Unsupervised Real-Time Network Intrusion and Anomaly Detection",2019,https://doi.org /10.1145/3354265.3354267

[40] S. Akcay, A. Atapour-Abarghouei, and T. P. Breckon, ''GANomaly: Semi-supervised anomaly detection via adversarial training,'' in Proc. Asian Conf. Comput. Vis., in Lecture Notes in Computer Science: Includ- ing Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics, vol. 11363, 2019, pp. 622–637, doi: 10.1007/978-3- 030-20893-6_39

[41] S. Garg, K. Kaur, N. Kumar, and J. J. P. C. Rodrigues, ''Hybrid deep- learning-based anomaly detection scheme for suspicious flow detection in SDN: A social multimedia perspective,'' IEEE Trans. Multimedia, vol. 21, no. 3, pp. 566–578, Mar. 2019, doi: 10.1109/TMM.2019.2893549

[42] X. Chen, B. Li, R. Proietti, Z. Zhu, and S. J. B. Yoo, ''Self-taught anomaly detection with hybrid unsupervised/supervised machine learning in optical networks,'' J. Lightw. Technol., vol. 37, no. 7, pp. 1742–1749, Apr. 1, 2019.

[43] M. Hasanet. Al. ''Attack and anomaly detection in IoT sensors in IoT sites using machine learning approaches,'' Internet Things, vol. 7, Sep. 2019, Art. no. 100059, doi: 10.1016/j.iot.2019.100059

[44] L. Deecke, Ret.al.''Image anomaly detection with generative adversarial networks,'' in Proc. Joint Eur. Conf. Mach. Learn.Knowl. Discovery Databases, in Lecture Notes in Computer Science: Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics, vol. 11051, 2019, pp. 3–17, doi: 10.1007/978-3-030-10925-7_1

[45] G. Pang, C. Shen, and A. van den Hengel, ''Deep anomaly detection with deviation networks,'' in Proc. 25th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining, Jul. 2019, pp. 353–362, doi: 10.1145/3292500.3330871

[46] R.Bhatia, S.Benno,J. Esteban,T.V. Lakshman, and J.Grogan, ''Unsupervised machine learning for network-centric anomaly detection in IoT,'' in Proc. 3rd ACM CoNEXT Workshop Big DAta, Mach. Learn. Artif.Intell.Data Commun.Netw. (Big-DAMA), 2019, pp. 42–48, doi: 10.1145/3359992.3366641

[47] J. Wang, J. Liu, J. Pu, Q. Yang, Z. Miao, J. Gao, and Y. Song, ''An anomaly prediction framework for financial IT systems using hybrid machine learning methods,'' J. Ambient Intell. Humanized Comput., vol. 2019, Dec. 2019, doi: 10.1007/s12652-019-01645-z

[48] R. Chalapathy, E. Toth, and S. Chawla, ''Group anomaly detection using deep generative models,'' in Proc. Joint Eur. Conf. Mach. Learn.Knowl. Discovery Databases, in Lecture Notes in Computer Science: Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics, vol. 11051, 2019, pp. 173–189, doi: 10.1007/978-3-030- 10925-7_11

[49] F. Yihunie, E. Abdelfattah, and A. Regmi, ''Applying machine learning to anomaly-based intrusion detection systems,'' in Proc. IEEE Long Island Syst., Appl. Technol. Conf. (LISAT), May 2019, pp. 1–5, doi: 10.1109/LISAT.2019.8817340

[50] A.Selvaraj, R.Patan, A.H.Gandomi, G.G.Deverajan, and M.Pushparaj, ''Optimal virtual machine selection for anomaly detection using a swarm intelligence approach,'' Appl. Soft Comput., vol. 84, Nov. 2019, Art. no. 105686, doi: 10.1016/j.asoc.2019.105686

[51] YujunShi,Kehua Miao "Detecting Anomalies in Application Performance Management System with Machine Learning Algorithms"2019 3rd International Conference on Electronic Information Technology and Computer Engineering (EITCE),doi:10.1109/EITCE47263.2019.9094916

[52] Earnest Paul Ijjina,Dhananjai Chand,Savyasachi Gupta,K. Goutham " Computer Vision- based Accident Detection in Traffic Surveillance" 2019 10th International Conference on Computing, Communication and Networking Technologies (ICCCNT), doi:10.1109/ICCCNT45670.2019.8944469

[53] Randeep Bhatia, Steven Benno, Jairo Esteban, T V Lakshman, and John Grogan. 2019." Unsupervised machine learning for network-centric anomaly detection in IoT." In 3rd ACM CoNEXT Workshop on Big DAta, Machine.https://doi.org/10.1145/3359992.3366641

[54] Rajiv Punmiya,Olga Zyabkina,SanghoChoe,Jan Meyer, "Anomaly Detection in Power Quality Measurements Using Proximity-Based Unsupervised Machine Learning Techniques" DOI:10.1109/PQ.2019.8818236.

[55] Xianglin Lu, Pengju Liu, Jiayi Lin," Network Traffic Anomaly Detection Based on Information Gain and Deep Learning", 2019,https://doi.org/10.1145/3325917.3325946.

[56] Shuning Wu, Joel Fulton, Ningwei Liu, Charles Feng, Ligang Zhang, "Risky Host Detection with Bias Reduced Semi-Supervised Learning" https://doi.org/10.1145/3349341.3349365

[57] Klang E, Barash Y, Yehuda Margalit R, Soffer S, Shimon O, Albshesh A, BenHorin S, Michal Amitai M, Eliakim R, Kopylov U, "Deep learning algorithms for automated detection of Crohn's disease ulcers by video capsule endoscopy, Gastrointestinal Endoscopy" (2019), doi:https://doi.org/10.1016/j.gie.2019.11.012.

[58] Hanling Wang, Mingyang Li, Fei Ma, Shao-Lun Huang, and Lin Zhang. 2019. Poster Abstract: "Unsupervised Anomaly Detection via Generative Adversarial Networks." In The 18th International Conference on Information Processing in Sensor Networks (co-located with CPS-IoT Week 2019) (IPSN '19), April 16–18, 2019, Montreal, QC, Canada.ACM, New York, NY, USA, 2 pages.https://doi.org/10.1145/3302506.3312605

[59] SulaimanAlhaidari, Mohamed Zohdy, "Hybrid Learning Approach of Combining Cluster-Based Partitioning and Hidden Markov Model for IoT Intrusion Detection", 2019,https://doi.org/10.1145/3325917.3325939.

[60] AdrienBenamira, Benjamin Devillers, Etienne Lesot, Ayush K. Ray, ManalSaadi, Fragkiskos D. MalliarosCentraleSupelec, University of Paris-Saclay, France "Semi-Supervised Learning and Graph Neural Networks for Fake News Detection", 2019,http://dx.doi.org/10.1145/3341161.33429588

[61] Dezhi Wang, Lilun Zhang, Xinghua Cheng, Dan Zou, Wenbin Xiao, "An Improved Weakly Supervised Learning System for Detection of Sound Events in Domestic Environments", © 2019 Association for Computing Machinery,https://doi.org/10.1145/3369985.3370007.

[62] GeorgiosTertytchny, Nicolas Nicolaou, Maria K. Michael. 2019. "Differentiating Attacks and Faults in Energy Aware Smart Home System using Supervised Machine Learning." In INTERNATIONAL CONFERENCE ON OMNI-LAYER INTELLIGENT SYSTEMS (COINS), May 5–7, 2019, Crete, Greece., 6 pages.https://doi.org/10.1145/3312614.3312642

[63] A. Barua, D. Muthirayan, P. P. Khargonekar, and M. A. Al Faruque, ''Hierarchical temporal memory-based machine learning for real-time, unsupervised anomaly detection in smart grid: WiP abstract,'' in Proc. ACM/IEEE 11th Int. Conf. Cyber-Phys. Syst. (ICCPS), Apr. 2020, pp. 188–189, doi: 10.1109/ICCPS48487.2020.00027

[64] Z. Chkirbene, S. Eltanbouly, M. Bashendy, N. AlNaimi, and A. Erbad, ''Hybrid machine learning for network anomaly intrusion detection,'' in Proc. IEEE Int. Conf. Informat., IoT, Enabling Technol. (ICIoT), Feb. 2020, pp. 163–170, doi: 10.1109/ICIoT48696.2020.9089575

[65] E. Quatrini, F. Costantino, G. Di Gravio, and R. Patriarca, ''Machine learning for anomaly detection and process phase classification to improve safety and maintenance activities,'' J. Manuf. Syst., vol. 56, pp. 117–132, Jul. 2020, doi: 10.1016/j.jmsy.2020.05.013

International Journal of Emerging Technology and Advanced Engineering
Website: www.ijetae.com (E-ISSN 2250-2459, Scopus Indexed, ISO 9001:2008 Certified Journal, Volume 12, Issue 11, November 2022)

[66] Y. Liu, Z. Pang, M. Karlsson, and S. Gong, ‘‘Anomaly detection based on machine learning in IoT-based vertical plant wall for indoor climate control,’’ Building Environ., vol. 183, Oct. 2020, Art. no. 107212, doi: 10.1016/j.buildenv.2020.107212

[67] P.Tang, W.Qiu, Z.Huang, S.Chen, M.Yan, H.Lian,andZ.Li,‘‘Anomaly detection in electronic invoice systems based on machine learning,’’ Inf. Sci., vol. 535, pp. 172–186, Oct. 2020, doi:10.1016/j.ins.2020.03.089

[68] I.G.A.Poornimaand, B.Paramasivan, ‘‘Anomaly detection in wireless sensor network using machine learning algorithm,’’ Comput. Commun., vol. 151, pp. 331–337, Feb. 2020, doi: 10.1016/j.comcom.2020.01.005

[69] Janghoon Kim, Hyunpyo Choi, Jiho Shin, Jung Take Seo."Study on Anomaly Detection Technique in an Industrial Control System Based on Machine Learning." International Conference on Intelligent computing and its Emerging Applications (ACM ICEA '20), December 12-15, 2020, Gang Won, Republic of Korea. ACM, New York, NY, USA, 5 pages,https://doi.org/10.1145/3440943.3444743

[70] Brandon Phillips, Eric Gamess, and Sri Krishnaprasad. 2020. “An Evaluation of Machine Learning-based Anomaly Detection in a SCADA System Using the Modbus Protocol.” In 2020 ACM Southeast Conference (ACMSE 2020), April 2–4, 2020, Tampa, FL, USA.ACM, New York, NY, USA, 9 pages.https://doi.org/10.1145/3374135.3385282

[71] Muhammad Aqeel, Khan Bahadar Khan, Muhammad AdeelAzam, Muhammad HamzaGhouri, Hammad–ur-RehmanKhalid, Fawwad Hassan Jaskani, “Detection of Anomaly in Videos Using Convolutional Autoencoder and Generative Adversarial Network model”, 2020,DOI:10.1109/INMIC50486.2020.9318140

[72] HelmiQosim and Zulkarnain. 2020. “Fault Detection System Using Machine Learning on Synthesis Loop Ammonia Plant.” In Proceedings of ACM APCORISE’20, June, 2020, Depok, West Java, Indonesia. ACM, New York, NY, USA, 7 pages.https://doi.org/10.1145/3400934.3400950

[73] Harsh Jain, AdityaVikram, Mohana, AnkitKashyap, Ayush Jain, “Weapon Detection using Artificial Intelligence and Deep Learning for Security Applications”, 2020, IEEE, DOI:10.1109/ICESC48915.2020.9155832

[74] Tram Truong-Huu, NidhyaDheenadhayalan, ParthaPratimKundu, VasudhaRamnath, Jingyi Liao, Sin G. Teo, and Sai Praveen Kadiyala. 2020. “An Empirical Study on Unsupervised Network Anomaly Detection using Generative Adversarial Networks.” In Proceedings of the 1st Security and Privacy on Artificial Intelligent Workshop (SPAI ’20), October 6, 2020, Taipei, Taiwan. ACM, New York, NY, USA, 10 pages.https://doi.org/10.1145/3385003.3410924

[75] Swati Sharma, AmadouDiarra, FredericoAlvares, and Thomas Ropars. 2020 “KDetect: Unsupervised Anomaly Detection for Cloud Systems Based on Time Series Clustering.” In 3rd International Workshop on Systems and Network Telemetry and Analytics (SNTA '20), June 23, 2020, Stockholm, Sweden.ACM, New York, NY, USA, 8 pages.https://doi.org/10.1145/3391812.3396271

[76] BaaniLeenKaur Jolly, Lavina Jain, DebajyotiBera, and TanmoyChakraborty. 2020. “Unsupervised Anomaly Detection in Journal-Level Citation Networks.” In ACM/IEEE Joint Conference on Digital Libraries in 2020 (JCDL '20), August 1-5, 2020, Virtual Event, China.ACM, New York, NY, USA, 10 pages.https://doi.org/10.1145/3383583.3398531

[77] JulienAudibert, PietroMichiardi, FrédéricGuyard, Sébastien Marti, and Maria A. Zuluaga. 2020. “USAD: UnSupervised Anomaly Detection on Multivariate Time Series.” In Proceedings of the 26th ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD'20), August 23-27, 2020, Virtual Event, CA, USA.ACM, New York, NY, USA, 10 pages.https://doi.org/10.1145/3394486.3403392

[78] Weichao Zhang,GuanjunWang;Mengxing Huang,Hongyu Wang,Shaoping Wen “Generative Adversarial Networks for Abnormal Event Detection in Videos Based on Self-Attention Mechanism”,doi:10.1109/ACCESS.2021.3110798

[79] Hongchun Yuan,ZhenyuCai,Hui Zhou,Yue Wang,Xiangzhi Chen, “TransAnomaly: Video Anomaly Detection Using Video Vision Transformer”, date of current version September 14, 2021, supported in part by the National Natural Science Foundation of China under Grant 41776142, doi:10.1109/ACCESS.2021.3109102

[80] GuoPu, Lijuan Wang, Jun Shen, and Fang Dong, “A Hybrid Unsupervised Clustering-Based Anomaly Detection Method”, Volume 26, Number 2, April 2021, DOI: 10.26599/TST.2019.9010051.

[81] Aviv Yehezkel, EyalElyashiv, and Or Soffer. 2021. “Network Anomaly Detectin Using Transfer Learning Based on Auto-Encoders Loss Normalization.” In Proceedings of the 14th ACM Workshop on Artificial Intelligence and Security (AISec '21), November 15, 2021, Virtual Event, Republic of Korea. ACM, New York, NY, USA, 11 pages.https://doi.org/ 10.1145/3474369.3486869

[82] “Fan Yang, Zhiwen Yu, Liming Chen, JiaxiGu, Qingyang Li, and Bin Guo. 2020. “Human-Machine Cooperative Video Anomaly Detection.” Proc. ACM Hum.-Comput. Interact. 4, CSCW3, Article 274 (December 2020), 18 pages.https://doi.org/10.1145/3434183

[83] Marcos A. O. Cavalcanti, Pedro R. M. Inácio, and Mário M. Freire. 2021. “Performance Evaluation of Container-Level Anomaly-Based Intrusion Detection Systems for Multi-Tenant Applications Using Machine Learning Algorithms”. In The 16th International Conference on Availability, Reliability and Security (ARES 2021), August 17–20, 2021, Vienna, Austria.ACM, New York, NY, USA, 9 pages.https://doi.org/10.1145/3465481.3470066

[84] Wesley Joon-WieTannet. al. “Filtering DoS Attacks from Unlabeled Network Traffic Data Using Online Deep Learning.”In Proceedings of the 2021 ACM Asia Conference on Computer and Communications Security (ASIA CCS '21), June 7-11, 2021, Hong Kong, Hong Kong.ACM, New York, NY, USA, 15 pages.https://doi.org/10.1145/3433210.3453083

[85] ParsaKavehzadeh, MohammadrezaSamadi, and Maryam Amir Haeri. 2021. “Unsupervised Anomaly Detection on Node Attributed Networks: A Deep Learning Approach.” In 2021 The 4th International Conference on Information Science and Systems (ICISS 2021), March 17–19, 2021, Edinburgh, United King-dom. ACM, New York, NY, USA, 6 pages.https://doi.org/10.1145/3459955.3460597

[86] MulyanaSaripuddinet. al. “Random Undersampling on Imbalance Time Series Data for Anomaly Detection.”In 2021 The 4th International Conference on Machine Learning and Machine Intelligence (MLMI’21), September 17–19, 2021, Hangzhou, China.ACM, NewYork, NY, USA, 6 pages.https://doi.org/10.1145/3490725.3490748

[87] Xiao Han and Shuhan Yuan. 2021. “Unsupervised Cross-system Log Anomaly Detection via Domain Adaptation.” In Proceedings of the 30th ACM Internatonal Conference on Information and Knowledge Management (CIKM'21), November 1-5, 2021, Virtual Event, QLD, Australia.ACM, New York, NY, USA, 5 pages.https://doi.org/10.1145/3459637.3482209

[88] L. Yang et al., "Semi-Supervised Log-Based Anomaly Detection via Probabilistic Label Estimation," 2021 IEEE/ACM 43rd International Conference on Software Engineering (ICSE), 2021, pp. 1448-1460,

[89] Mahmoud Said Elsayed, Nhien-An Le-Khac, Hamed Z. Jahromi, and Anca Delia Jurcut. 2021. “A Hybrid CNN-LSTM Based Approach for Anomaly Detection Systems in SDNs.” In The 16th International Conference on Availability, Reliability and Security (ARES 2021), August 17–20, 2021, Vienna, Austria.ACM, New York, NY, USA, 7 pages.https://doi.org/10.1145/3465481.3469190

[90] Quang-Vinh Dang “Evaluating machine learning algorithms for intrusion detection systems using the dataset CIDDS-002.” In 2021 4th International Conference on Computer Science and Software Engineering (CSSE 2021) (CSSE2021), October 22–24, 2021, Singapore, Singapore.ACM, New York, NY, USA, 7 pages.https://doi.org/10.1145/3494885.3494906

[91] HatityeChindove and Dane Brown. 2021. “Adaptive Machine Learning Based Network Intrusion Detection.” In International Conference on Artificial Intelligence and its Applications (icARTi ’21), December 9–10, 2021, Virtual Event, Mauritius.ACM, New York, NY, USA, 6 pages.https://doi.org/10.1145/3487923.3487938

[92] Shuo Liu et. al. “LCR-GAN: Learning Crucial Representation for Anomaly Detection.” In 2021 5th International Conference on Computer Science and Artificial Intelligence (CSAI 2021), December 04–06, 2021, Beijing, China.ACM, New York, NY, USA, 6 pages.https://doi.org/10.1145/3507548.3508229

[93] Shehroz S. Khan;Pratik K. Mishra;NizwaJaved;BingYe;KristineNewman;AlexMihailidis;AndreaIaboni, “Unsupervised Deep Learning to Detect Agitation From Videos in People With Dementia, date of current version” January 28, 2022., doi: 10.1109/ACCESS.2022.3143990

[94] Siddharth Bhatia at.Al. “Real-Time Anomaly Detection in Edge Streams.”ACM Trans. Knowl.Discov.Data.16, 4, Article 75 (January 2022),https://doi.org/10.1145/3494564.

[95] A. N. M. Bazlur Rashid, Mohiuddin Ahmed, Leslie F. Sikos, and Paul Haskell-Dowland. 2022. “Anomaly Detection in Cybersecurity Datasets via Cooperative Co-evolution-based Feature Selection”. ACM Trans. Manag.Inf. Syst. 13, 3, Article 29 (February 2022), 39 pages.https://doi.org/10.1145/3495165

[96] David Ohana, Bruno Wassermann, Nicolas Dupuis, Elliot Kolodner, EranRaichstein, and Michal Malka. 2022. “Hybrid Anomaly Detection and Prioritization for Network Logs at Cloud Scale”. In Seventeenth European Conference on Computer Systems (EuroSys’22), April 5–8, 2022, RENNES, France.ACM, New York, NY, USA, 15 pages.https://doi.org/10.1145/3492321.3519566.

[97] Shenglin Zhang et. Al.“Robust System Instance Clustering for Large-Scale Web Services.”In Proceedings of the ACM Web Conference 2022 (WWW ’22), April 25–29, 2022, Virtual Event, Lyon, France.ACM, New York, NY, USA, 12 pages.https://doi.org/10.1145/3485447.3511983

[98] XuexiongLuo, et. Al. “ComGA: Community-Aware Attributed Graph Anomaly Detection.” In Proceedings of the Fifteenth ACM International Conference on Web Search and Data Mining (WSDM '22), February 21-25, 2022, Tempe, AZ, USA.ACM, New York, NY, USA, 9 pages.https://doi.org/10.1145/3488560.3498389

[99] Max Landauer, Florian Skopik, Markus Wurzenberger, and Andreas Rauber. 2022. “Dealing with Security Alert Flooding: Using Machine Learning for Domain-independent Alert Aggregation.” ACM Trans. Priv. Secur. 25, 3, Article 18 (April 2022), 36 pages.https://doi.org/10.1145/3510581

[100] FeiGao at. Al.“A Trajectory Evaluator by Subtracks for Detecting VOT-based Anomalous Trajectory.” ACM Trans. Knowl.Discov.Data 16, 4, Article 67 (January 2022), 19 pages.https://doi.org/10.1145/3490032

[101] Yulu, Cao* and Honglu, Gan. 2022.” CLAD: A Deep Learning Framework for Continually Learning in Anomaly Detection.” In 2022 The 5th International Conference on Software Engineering and Information Management (ICSIM) (ICSIM 2022), January 21–23, 2022, Yokohama, Japan. ACM, New York, NY, USA, 6 pages.https://doi.org/10.1145/3520084.3520109

[102] D.T. Mane, U.V. Kulkarni, “Modified Fuzzy Hypersphere Neural Network for Pattern Classification using Supervised Clustering,” Procedia Computer Science, Volume 143, 2018, Pages 295-302, ISSN 1877-0509,https://doi.org/10.1016/j.procs.2018.10.399.

[103] Thudumu, S., Branch, P., Jin, J. et al.A comprehensive survey of anomaly detection techniques for high dimensional big data.J Big Data**7,** 42 (2020).https://doi.org/10.1186/s40537-020-00320-x

[104] A. B. Nassif, M. A. Talib, Q. Nasir and F. M. Dakalbab, "Machine Learning for Anomaly Detection: A Systematic Review," in IEEE Access, vol. 9, pp. 78658-78700, 2021, doi: 10.1109/ACCESS.2021.3083060.

[105] Mane, D.T. and Kulkarni, U.V., 2020.A survey on supervised convolutional neural network and its major applications. In Deep Learning and Neural Networks: Concepts, Methodologies, Tools, and Applications (pp. 1058-1071). IGI Global.

[106] Upadhye, GopalDadarao et al. “Improved Model Configuration Strategies for Kannada Handwritten Numeral Recognition.” Image Analysis & Stereology, vol 40, issue 3 (2021),Doi: 10.5566/ias.2586