Global Scientific JOURNALS

A Systematic literature review of Machine learning models for intrusion detection for smart utility meters; a systematic literature review

Nafuye Ivan[1]: Department of computer science, Maseno University, Kenya; inafuye@gmail.com
Dr Lillian D Wanzare[2]: Department of computer science, maseno university,kenya; ldwanzare@maseno.ac.ke
Dr. James Obuhuma[3:] jobuhuma@maseno.ac.ke, Ikwap flavia Agatha[4]; flav.agatha@gmail.com

Abstract: This systematic literature review examines the use of machine learning algorithms in smart meter applications. Smart meters are digital devices that record and transmit energy consumption data to utilities for billing, monitoring, and optimization purposes. Machine learning algorithms have been applied to smart meters to improve energy efficiency, detect anomalies, and forecast energy demand. This review analyses over 58 studies papers published between 2008 and 2023 and categorizes the machine learning algorithms based on their application and performance metrics, the features and datasets. The study finds that clustering and regression algorithms are commonly used for anomaly detection and demand forecasting, respectively with these being the support vector machines and the decision trees, while deep learning algorithms show promise for improving accuracy and interpretability. The review also identifies challenges in data and feature quality, privacy, and scalability that need to be addressed for successful deployment of machine learning in smart metering. Overall, this study provides a comprehensive overview of the state-of-the-art machine learning techniques in smart meters and suggests future research directions for advancing the field.

Introduction

The number of devices connecting to the Internet has been growing at a breath-taking pace over the past decades. From two billion in 2006, it reached 200 billion in 2020 according to a report published by Kaspersky in 2022.

Development and advancement of new technological innovation, concepts such as of "Smart utility meters" have emerged as a means to achieve more efficient and sustainable cities. Utility companies worldwide have begun deploying smart meters to service residential, commercial and industrial markets. Smart meters deliver a range of benefits including lower operational and capital expenses, support for new services, and improved operational control.

If electricity supply is controlled by devices connected to communications infrastructure, then there is the possibility that hackers could disrupt the electricity system. Anderson and Fuloria wrote an article about why we should think carefully about the security problems of smart meters. Another security expert, Krebs, reported that the FBI had seen a very large amount of electricity theft in Puerto Rico after smart meters were installed. So, there are some problems with security.

In recent years, there has been an increased interest in exploring machine learning for enhancing the detection quality of IDSs, however, ensuring security and privacy for the IoT devices is a nontrivial challenge due to their limitless computational capabilities, which is challenging for traditional security mechanisms. This has made them susceptible to wide range of cyberattacks, such as data leakage, spoofing, DoS / DDoS and with smart utility meters brought into play, it is even more certain there is a lot at stake.
To get an overview of what has been done on the application of ML in smart utility meters especially with the anomaly-based approach, a systematic literature review (SLR) will be performed. A Systematic

Literature Review shows the potential gaps in a given research area and will guide who so ever wishes to do a new research study on this problem area. By following a methodology in SLR, all relevant studies are accessed from electronic databases, synthesized, and presented.
This study presents all the available literature published so far on the application of machine learning in smart utility meters and the general advanced metering infrastructure. In this study, I present my empirical results and responses to the research questions defined which were designed to help guide and shape this review.
This paper is organized as follows: a section explaining the background, another discusses the methodology used in this review, next presents the results. A section explaining the anomaly-based machine learning approaches in smart utility meters. A section presenting challenges and way forward.

2. Related work

There have not been so many systematic literature reviews conducted on the use of machine learning algorithms in smart meters. One such review by farivar et al examined the use of machine learning algorithms for demand response in smart grid systems. Their study identified several machine learning algorithms, including decision trees, support vector machines, and neural networks, that were used for demand response in smart meters. This review also highlighted the importance of data quality and privacy concerns in smart meter data analysis. Farivar et al. (2018)

In a paper by Alejandro Hernandez et al 2022, a systematic review of machine learning techniques related to local energy communities was carried out. This paper presents the conceptualisation of a local energy community on the basis of a review of 25 energy community projects. Furthermore, an extensive literature review of machine learning algorithms for local energy community applications was conducted, and these algorithms were categorised according to forecasting, storage optimisation, energy management systems, power stability and quality, security, and energy transactions. The main algorithms reported in the literature were analysed and classified as supervised, unsupervised, and reinforcement learning algorithms. The findings demonstrate the manner in which supervised learning can provide accurate models for forecasting tasks. Similarly, reinforcement learning presents interesting capabilities in terms of control-related applications.

A systematic review of machine learning applications in the operation of smart distribution systems by Terezija Matijašević et al 2022, summarized the advantages and also the shortcomings in their comprehensive review of using machine learning based methods and algorithms in the planning and operation of smart, active distribution systems is provided. In addition to the already developed and presented applications, they identified the current research gap and proposed future research directions with machine learning applications.

Another systematic literature review published in 2019 by Zhang et al focused on the use of machine learning for energy forecasting in smart grids. The study analysed over 50 papers and identified several machine learning algorithms, including artificial neural networks, support vector machines, and random forests, being used for energy forecasting in smart meters. The review also highlighted the need for interpretability and transparency in machine learning models for energy forecasting.
However of all these, the most recent systematic literature review was by karmakar et al published in 2021 which was examining the use of machine learning algorithms for energy disaggregation in smart meters.

The study analysed over 60 papers and identified several machine learning algorithms, including convolutional neural networks, recurrent neural networks, and deep belief networks, that were being used for energy disaggregation in smart meters. The review also highlighted the need for benchmarking datasets and standardized evaluation metrics in energy disaggregation research. Karmakar et al. (2021)

All these systematic literature reviews mentioned above demonstrate the growing interest in the use of machine learning algorithms in smart meters. Further research is still needed and gap is widening each and every day in this exciting field. From these very studies there are a number of challenges still existing that need to be addressed especially relating to data quality, privacy, interpretability, and standardization.

3. Methodology

3.1. Review protocol

In order to conduct a deep analysis of software defect prediction, a Systematic Literature Review (SLR) was selected in this research. SLR collects the data from selected research studies to systematically deduce the results. SLR evaluates all the empirical research evidence to answer specific research questions (Carcary et al., 2018.)

It uses explicit criteria for deciding which studies will be included or excluded. This helps to minimize the authors' bias. SLR process consists of three phases (Boell et al., 2015): Planning, conducting and reporting the literature review. In the planning phase, a review protocol is developed. It defines the research questions, search strategy, inclusion and exclusion criteria for selecting relevant studies. Quality assessment, data extraction and data synthesis are applied to the selected primary studies in the conducting phase. The results are presented in the reporting phase of the literature review. SLR steps are shown in Figure 1.

The review has been done using the well-known review guidelines provided by (Kitchenham, 2012). Firstly, the research questions are defined. When research questions are ready, databases are used to select the relevant studies. The databases that were used in this study are Google Scholar, Semantic Scholar, ResearchGate, ScienceOpen, Academic Search, and SpringerLink.

A. phase one: planning the review

In the review protocol, the first step is to identify the purpose of the research. Then, research questions are formed to support the objective of the SLR. Then, the data are extracted based on the identified inclusion and exclusion criteria.

3.2. Research questions

This SLR aims to get insight into what studies have been published in the domain of anomaly detection, ML, intrusion detection and smart utility meters. To understand the problem, studies have been analysed from several dimensions. For this SLR study, the following four research questions (RQs) have been defined.

- RQ1- Which machine learning algorithms have been used for intrusion detection especially in smart meters?
- RQ2- Which are the major exiting features for intrusion detection especially in smart meter?
- RQ3- What datasets have been used in ML intrusion detection in smart meters?
- RQ4- How are machine learning models being evaluated?

3.3. Search strategy

The searching is done by narrowing down to the basic concepts that are relevant for the scope of this review. Machine learning and intrusion detection systems have many application fields, which means that there are a lot of published studies that are probably not in the scope of this review article. The basic searching is done by an automated search. The starting input for the search was "machine learning anomaly IDS" and "smart meters". Articles were retrieved, and abstracts were read to find the synonyms of the keywords.

The search was performed in six databases and in all these the above key word searches where performed to each of these. After the exclusion criteria were applied, and all the results were processed, and a more complex search string was built in order to avoid missing relevant studies. This final search strings were ((("machine learning) or "deep learning" AND "algorithms" AND ("smart meters" OR "intrusion detection systems"). After executing this search string, approximately 300 studies were retrieved.

A specific description of the search strings per databases are provided as follows:

From each of all the above mentioned search engines, the keys words where run in each of these and there after different extensions of .pdf, .doc where appended so as to fine tune the results to these formats. For each of the returned documents the title, abstract, keywords and algorithms used where captured so as to further extract our table of comparative study on works done so far.

3.4. Inclusion and Exclusion criteria

According to SLR, we should define a set of rules for choosing the most relevant studies. The inclusion and exclusion criteria are applied for selecting the primary studies. To exclude irrelevant studies, the studies were analysed and graded based on exclusion criteria to set the boundaries for this review. Table 3 shows the details of the inclusion and exclusion criteria.

Table 1: Inclusion and exclusion criteria

Inclusion criteria	
✓	Studies present an empirical study.
✓	Studies compare the performance of machine learning models.
✓	Studies published either in journals or conferences.
✓	studies concentrated on models in smart meters
Exclusion criteria	
☒	Publication is not related to the computing domain, nor is it related to machine learning nor intrusion detection systems nor smart utility meters.
☒	Publication is not written in English
☒	Publication that is a duplicate or already retrieved from another database
☒	Full text of the publication is not available
☒	Publication has been published before 2008

B. phase two: the second stage is conducting the review

Select Primary Studies

Figure 2 shows the overview of identified and remained studies after each step in the search process. In total, 90 primary studies are retrieved from the selected repositories based on the defined search string. After removing the duplicated studies, 44 have remained. Then the irrelevant studies are discarded based on reading the details of each study; four studies are discarded, leaving 40 primary studies to be evaluated in this SLR.

Quality Assessment

The quality assessment ensures the efficiency of the studies and is focused on extracting the studies that have sufficient information for answering the predefined research questions. We have defined a set of quality criteria to be applied to the selected primary studies. The quality criteria are shown in Table 2. Each

quality criteria must be answered using the options "Yes" or "No". The answer "Yes" represents value 1 and "No" represents value 0. We accumulate the values of all answers for each primary study. The sum is shown in Table 3. The primary study that reaches a sum lower than 75% will be excluded. After applying quality assessment on my 58 primary studies, all studies are included for the data extraction.

Data Extraction

The purpose of data extraction is to extract the data from the primary studies to answer the predefined research questions. The data extraction consists of three steps: The first step is concentrated on general information about the studies such as authors, title and publication type and publication year. The second step is focused on the implementation of machine learning algorithms, features, datasets and evaluation metrics. The third step extracts information about the empirical study and the final results of the models. Table 4 shows the characteristics that are used to answer the research questions. Table 5 shows the relationship between the primary studies and research questions, it checks if the selected studies answer the research questions or not.

When conducting the review, the publications were selected by going through all the databases. The data was extracted, which means that their information regarding authors, year of publication, type of publication, and more information regarding the research questions were stored. After all the necessary data was extracted correctly, the data was summarised and tabulated in a table to provide an overview of the related papers published.

Table 2: Quality criteria

No	Quality criteria	description
Q1	Is machine learning model clearly reported	Machine learning model must be well stated
Q2	Are features clearly mentioned	Features must be well mentioned
Q3	Is used dataset clearly mentioned	Dataset must be well mentioned
Q4	Are evaluation metrics well mentioned	Evaluation metrics is captured

Table 3: Results of quality criteria for primary studies

Study	RQ1	RQ2	RQ3	RQ4	sum
St1 Philokypros Ioulianou 2017	1	1	1	1	4
st2 Chih-Che Sun et al 2020	1	1	1	1	4
St3 Mustafa Amir Faisal et al 2016	1	1	1	1	4
St4 Dipanjan Das Roy and Dongwan Shin 2019	1	1	1	1	4
St5 Dejan Radovanovic etal 2022	1	1	1	1	4
St6 Mueen Uddin etal 2012	1	1	1	1	4
St7 Morteza Behniafar etal 2018	1	1	1	0	3
St8 Priti Prabhakar et al 2022	1	1	1	1	4
St9 Farid Molazem Tabrizi et al 2014	1	1	1	1	4
St10 Mausumi Das Nath et al 2020	1	1	1	1	4
St11 Tara Salman et al 2017	1	1	1	1	4
St12 Nebrase Elmrabit et al 2020	1	1	1	1	4
St13 Liu et. al 2014	1	1	1	1	4
St14 Lyu et. al 2017	1	0	1	1	3
St15 Summerville et. Al 2015	1	0	1	1	3
St16 Fu et al 2012	1	1	1	1	4
St17 Tsitsiroudi et al 2016	1	1	0	1	3
St18 Bostani et al	1	1	1	1	4
St19 Trilles et al 2016	1	1	1	1	4
St20 Hoang et al 2018	1	1	1	1	4
St21 Zissis et al 2017	1	1	0	0	2
St22 Sedjelmaci et al 2016	1	1	1	1	4
St23 Granjal et al 2018	1	1	0	1	3
St24 Domb et al 2017	1	0	1	1	3
St25 Sedjelmaci et. al 2017	1	0	1	1	3
St26 Rajesh Kumar Gunupudi et al 2017	1	1	1	1	4

St27 Pongle et al 2015	1	1	1	1	4
St28 Trilles et al 2015	1	1	1	1	4
St29 Moshtaghi et al 2017	1	1	1	1	4
St30 Yu et al 2017	1	1	1	1	4
St31 Zhao et al 2017	1	1	1	1	4
St32 Tama et al 2018	1	1	1	1	4
St33 Briana Arrington et al 2016	1	1	1	1	4
St34 Deris Stiawan et al 2011	1	1	1	0	3
St35 Yogitha et. al 2008	1	0	1	1	3
St36 Wenke Lee et. al 2011	1	1	1	0	3
St37 Eleazar, Matthew et. al 2008	1	1	1	1	4
St38 Archika Jain, et. al 2022	1	1	0	1	3
St39 Sushil Kumar Chaturvedi, et. al 2017	1	0	1	1	3
St40 Omprakash Chandrakar, et. al 2014	1	1	1	1	4
St41 A.R. Jakhale, et. al 2018	1	1	1	0	3
St42 R. Venkatesan, et al 2019	1	1	0	1	3
St43 Abhilasha A Sayar, et.al 2014	1	1	1	0	3
St44 S. S. Pawar et., al	1	0	1	1	3
St45 AnnkitaPatel,Risha Tiwari 2014	1	1	1	1	4
St46 Yogita B.Bhavsar& Kalyani C. Waghmare	1	1	1	1	4
St47 Rowayda A.Sadek, et al 2013	1	1	1	1	4
St48 Ahemd A Elngar,Dowalt, Fayed	1	1	1	1	4
St49 HeshamAltwaijry, Saeed Algarny 2012	1	1	1	1	4
St50 Tao Peng, WanliZuo 2009	1	1	1	1	4
St51 Wanke Lee,Salvotore J Stolfa 2008	1	1	1	1	4
St52 Renuka DeviThanasekaran 2011	1	1	1	1	4
St53 G.C. Tjhai et al 2010	1	0	1	1	3
St54 Bo Peng et al 2016	1	1	1	1	4
St 55 Shikha Singh et al 2018	1	1	1	1	4
St56 J. Viinikka et al 2009	1	1	1	1	4
St57 Revathi S et al 2013	1	1	1	1	4
St58 Ibrahim Salim.M et al 2016	1	1	1	1	4

Table 4: Data extraction characteristics mapped to research questions

Characteristic	Research question
Authors, publication, title, type, year	General
Machine learning algorithms	RQ1
Features	RQ2
Datasets	RQ3
Evaluation metrics	RQ4

Table 5: The relationship between the primary studies and research questions

Study	RQ1	RQ2	RQ3	RQ4
St1	✓	✓	✓	✓
St2	✓	✓	✓	✓
St3	✓	✓	✓	✓
St4	✓	✓	✓	✓
St5	✓	✓	✓	✓
St6	✓	✓	✓	✓
St7	✓	✓	✓	☒
St8	✓	✓	✓	✓
St9	✓	✓	✓	✓
St10	✓	✓	✓	✓
St11	✓	✓	✓	✓
St12	✓	✓	✓	✓
St13	✓	✓	✓	✓
St14	✓	✓	✓	✓
St15	✓	✓	✓	✓
St16	✓	✓	✓	✓
St17	✓	✓	✓	✓
St18	✓	✓	✓	✓

St19	✓	✓	✓	✓
St20	✓	✓	✓	✓
St21	✓	✓	✓	☒
St22	✓	✓	✓	✓
St23	✓	✓	✓	✓
St24	✓	✓	✓	✓
St25	✓	✓	✓	✓
St26	✓	✓	✓	✓
St27	✓	✓	✓	✓
St28	✓	✓	✓	✓
St29	✓	✓	✓	☒
St30	✓	✓	✓	✓
St31	✓	✓	✓	✓
St32	✓	✓	✓	✓
St33	✓	✓	✓	✓
St34	✓	✓	✓	☒
St35	✓	✓	✓	✓
St36	✓	✓	✓	☒
St37	✓	✓	✓	✓
St38	✓	✓	✓	✓
St39	✓	✓	✓	☒
St40	✓	✓	✓	✓
St41	✓	✓	✓	☒
St42	✓	✓	✓	☒
St43	✓	✓	✓	☒
St44	✓	✓	✓	✓
St45	✓	✓	✓	✓
St46	✓	✓	✓	✓
St47	✓	✓	✓	✓
St48	✓	✓	✓	✓
St49	✓	✓	✓	✓
St50	✓	✓	✓	✓
St51	✓	✓	✓	✓
St52	✓	✓	✓	✓
St53	✓	✓	✓	✓
St54	✓	✓	✓	✓
St55	✓	✓	✓	✓
St56	✓	✓	✓	✓
St57	✓	✓	✓	✓
St58	✓	✓	✓	✓

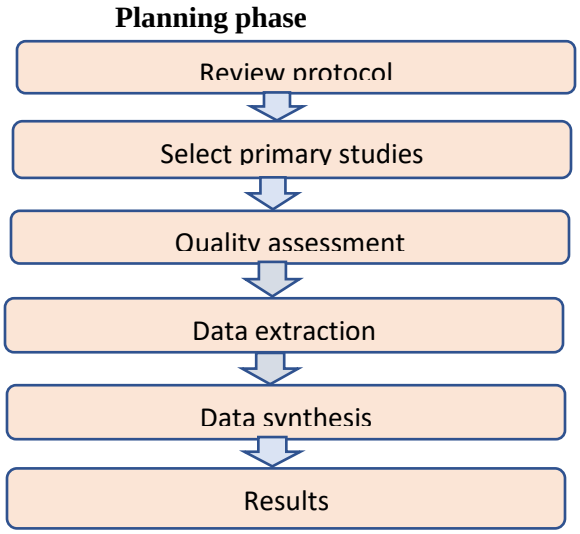

Fig. 1: Steps when conducting the SLR

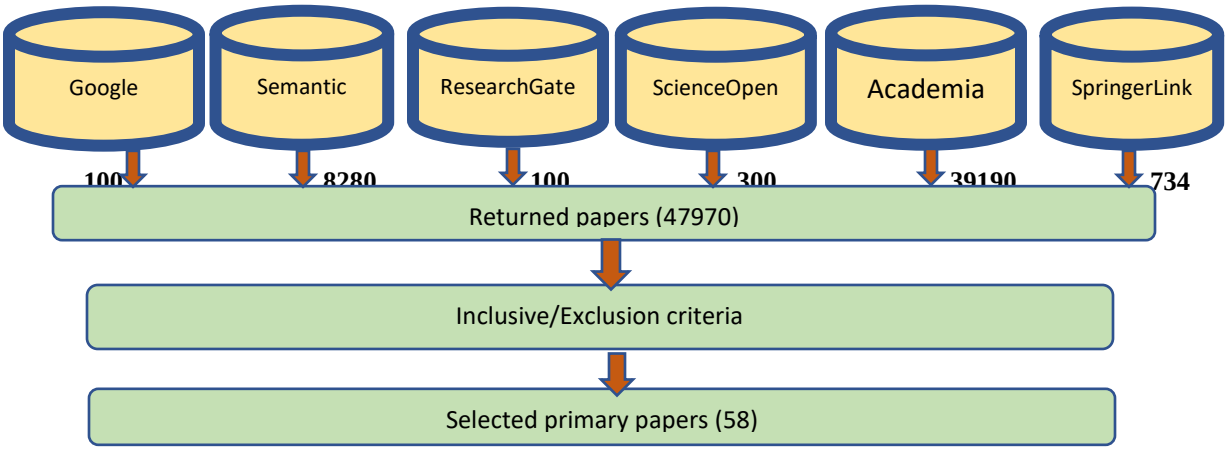

Fig. 2: search and selection processes of the SLR study.

Data Synthesis

The goal of data synthesis is to collect the data together from the selected primary studies to answer the research questions, with the aim of aggregating the evidence. There are many types of strategies to synthesize the data. In this study we used the narrative synthesis strategy. In narrative synthesis, visualization of the results is presented with the support of adding text to explain the context. In this SLR, the data are presented in a manner compatible with the research questions. The results of the primary studies are presented by using tables, bar charts, pie charts and column charts. The results are shown in section III.

Phase Three: Results

In this section, the authors present and discuss briefly the results of the SLR study. For each of the RQs, the authors instigate them with a summary of the most noteworthy results, a discussion about the most relevant facets, and, based on them, a suggestion of some explanatory hypotheses.

A. Machine learning algorithms

Security before machine learning was an uphill task that left many questions unanswered in all technology fields. Machine learning has shown great potential in intrusion detection systems, where it can learn to identify malicious network traffic patterns and anomalous behaviors that are indicative of attacks. Over time a number of machine learning algorithms have been proposed, from this study a number of these crisscross all domains and include:

Random Forests, K-Nearest Neighbors, Naive Bayes, Principal Component Analysis, Feature selection, k-means, Massive online analysis, Clustering, Bayesian, Decision tree both c 4.5 &J48, Neural networks, Convolutional neural networks (CNN), Recurrent neural networks (RNN), Median absolute deviation (MAD), Hidden markov model etc

Different machine learning algorithms work in different ways, but most involve some form of mathematical optimization or pattern recognition.

Machine learning models can play a crucial role in smart meter technology by analyzing the vast amounts of data generated by these devices. Machine learning algorithms in smart meters are typically trained using historical data on energy consumption and other relevant factors.

In this SLR, there are 58 primary studies reviewed using different machine learning models. Table 7 and figure 5 show all the machine learning algorithms that have each been used in my primary studies. Figure 4 shows the number of the studies using each machine learning algorithm captured. In addition the percentage on use of each model is shown in Fig. 6. St7 by Morteza Behniafar etal 2018 has the highest

number of eleven of algorithms cited. Furthermore most of the studies have used within a range of 1-4 algorithms on average in their studies as seen in figure 5

B. Features

Feature selection is the process of selecting a subset of relevant features (or variables) that can be used to build an effective machine learning model. In other words, feature selection involves identifying the most important features that contribute the most to the predictive power of the model while discarding the redundant or irrelevant features. Feature selection is a powerful technique in machine learning that can improve the performance, efficiency, and interpretability of models. A number of features were used in the primary studies to mention but a few these included: Connection/failed/ attempt/login/unknown/modification, Packet burst/service type, Ram, CPU, memory, disk usage/demand, Service, time/ during/connection duration/volume, Source/destination IP. However predominantly most of the smart meter models incorporated some of these with Total energy consumption, maximum energy consumption, energy consumption, hourly energy consumption.

As shown in Fig 7, thirty studies used the packet based features such as source and destination IP address among other features. Most of the studies use a combination of these features. In the smart meters the studies predominantly used more of electricity consumption features that is study St (5, 8, 9, 41, and 19).
From our primary studies, In the context of network packet analysis, some common features that are used for machine learning include packet size, source and destination IP addresses, port numbers, packet type (e.g., TCP, UDP), and payload content. However, the specific features selected will depend on the specific use case and the machine learning algorithm being used.

c. Datasets

Dataset is a data collection that is used to train and test machine learning models. A dataset typically consists of two parts: the features and the target variable. In this SLR, 58 primary studies are analyzed to find which datasets are used in them. Table 6 shows the details of the datasets used for each study. There after we proceed to calculate frequency of use of each of the cited datasets from the primary studies. Other than the generalized intrusion detection datasets and those of the IOT, a lot of attention needs to be paid to the real environment datasets of respective applicability. There is no standard dataset to be used in the models, because each organization has its own dataset.
As shown in Fig. 9, most of the studies use one dataset and then a hand full few of studies used two and three datasets, only one study St 40 cited 6 of these and others did not specify datasets used. From table 6 it can be noted that the original KDD cup of 1999 had the highest number of use with 21 studies. The KDD Cup 1999 is a well-known dataset used in the field of network intrusion detection. The KDD Cup 1999 dataset has become a benchmark dataset for intrusion detection and is often used as a standard reference in research papers and publications. The dataset includes network traffic data collected over nine weeks from a simulated network environment, including both normal and malicious traffic. The malicious traffic includes several types of attacks, such as denial-of-service attacks, probing attacks, and user-to-root attacks. Following the order from the KDD cup 1999 above, the other studies used the following datasets: NLS-KDD, IOT-23,IOT dataset-BA IOT, Real world IOT, UNSW-NB15, DARPA, Smart meter live data, Other network traffic datasets, Custom datasets, Did not use specific dataset, CICIDS 2017, Kyota 2006. Figure 9 shows the frequency of studies using a given dataset in the primary studies.

D. Evaluation Metrics

Evaluation metrics are measures used to assess the performance of machine learning models. These metrics can help determine how well a model is performing and guide further improvements. It is important to select appropriate evaluation metrics that are relevant to the specific problem being solved. Main metrics according to Li *et al*., 2018 and Tantithamthavorn *et al*., 2018 are Precision, Recall/DR, F-measure and Area under the Curve (AUC). When comparing the proposed models with the other selected baseline models, other than the above metrics studies used detection rate, accuracy, false positive, true negative, true positive, execution time, alert rates and a few did not clearly mention of how evaluation was done. Most of the studies applied Recall or Detection rate and accuracy. The detection rate or recall is an important performance metric in machine learning, especially in applications where the cost of missing a positive instance is high. For example, in this case intrusion systems, missing an attack could have serious consequences for the system and users. Therefore, a high detection rate (or recall) is desired to minimize false negatives. Accuracy on the other hand is a commonly used performance metric that measures the overall correctness of the model's predictions. The used metrics and their results are shown in Table 8. Fig. 10 also shows percentage of use of an evaluation metrics. All the used evaluation metrics are captured and shown in the above mentioned table and figure inclusive even those that did not clearly depict what was used. From table 8, the best detection rate or recall rate is 99.95 and the accuracy is 99.89.

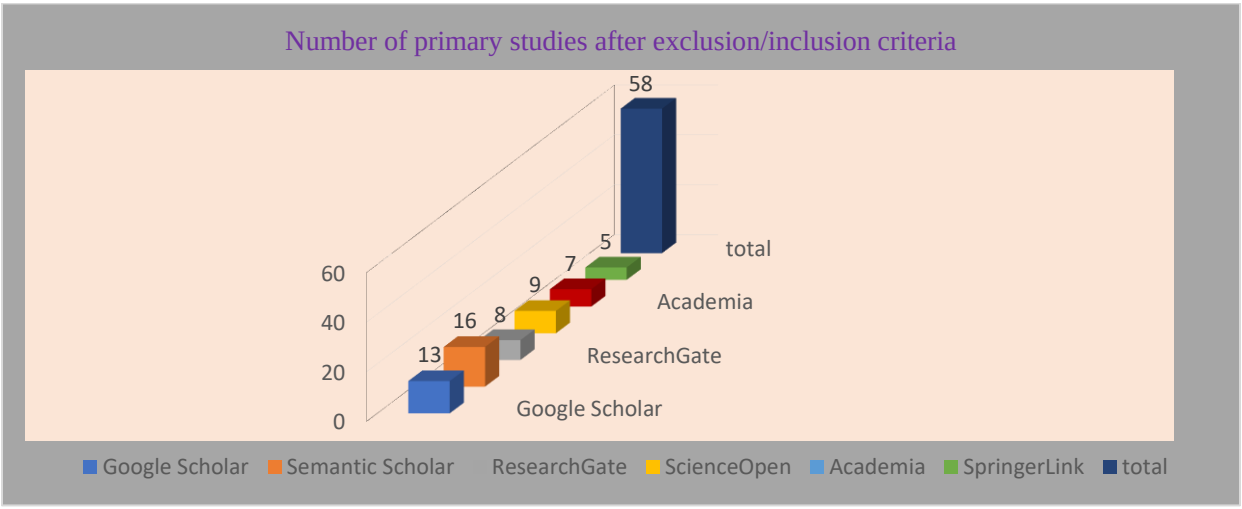

Fig 3: Number of primary studies after exclusion/inclusion criteria

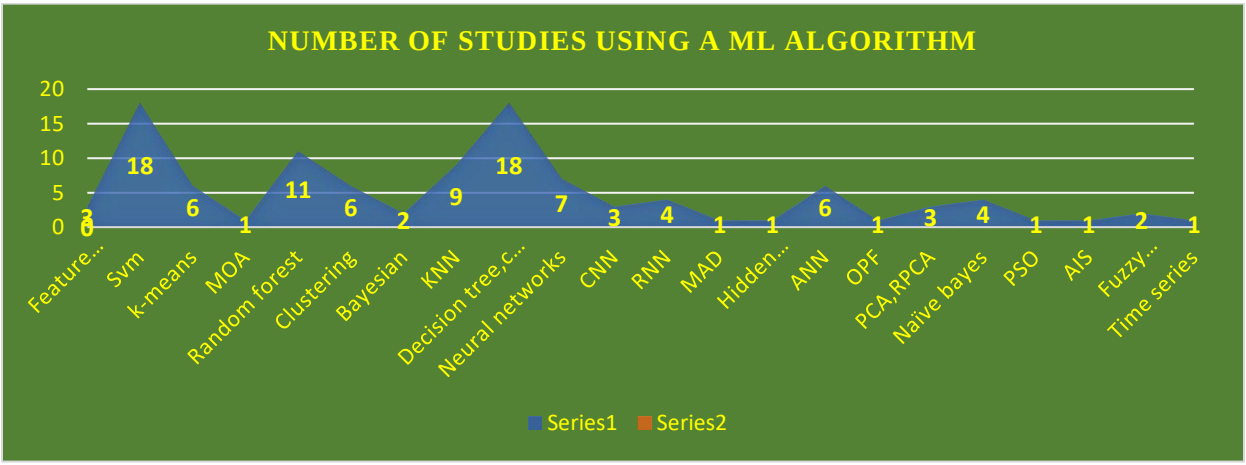

Fig 4: Number of studies using a ML algorithm

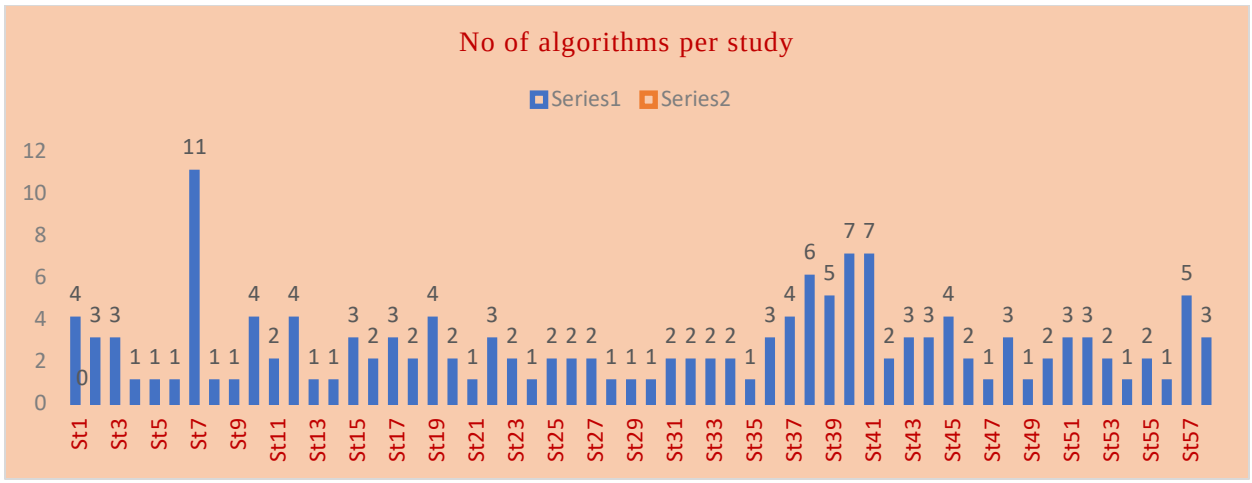

Fig 5: Number of algorithms used in each primary study

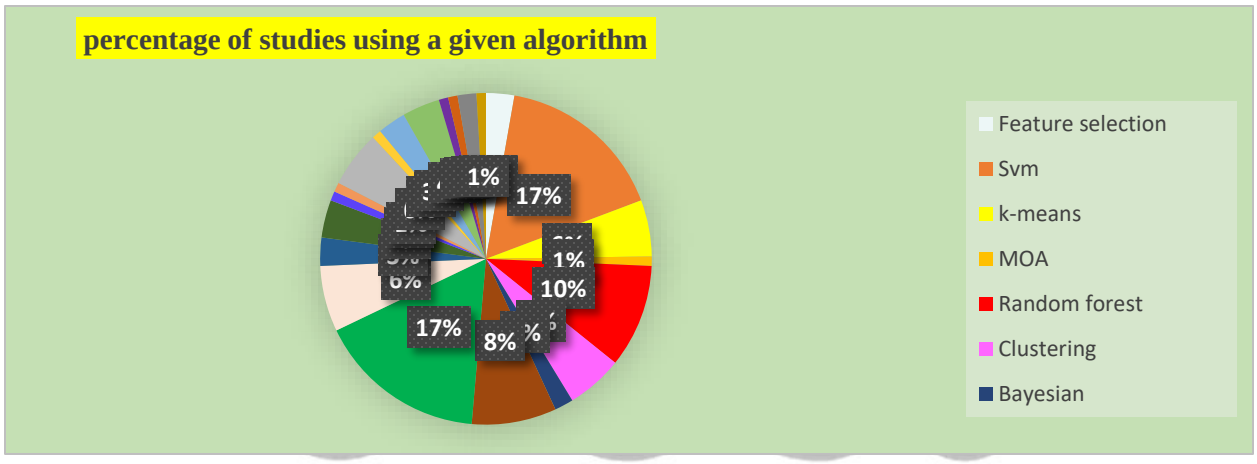

Fig 6: Percentage of studies using a given algorithm

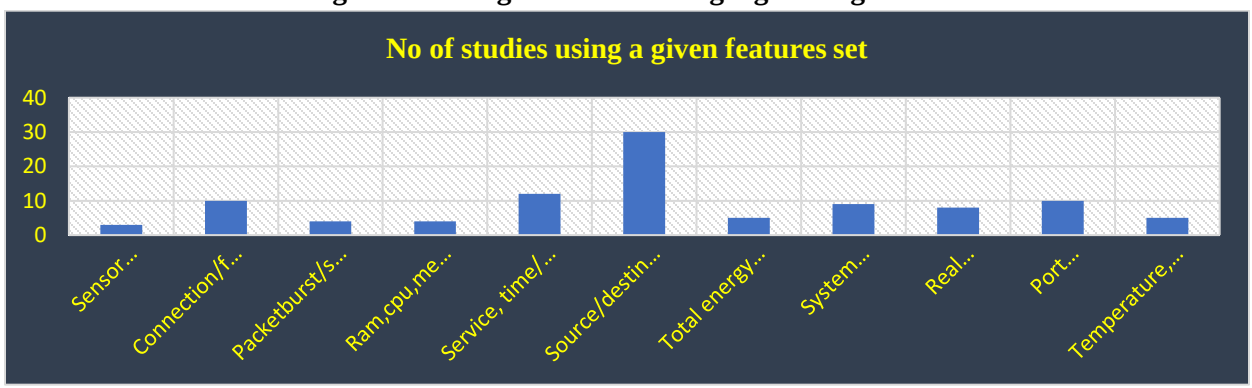

Fig 7: Number of studies using a given feature set

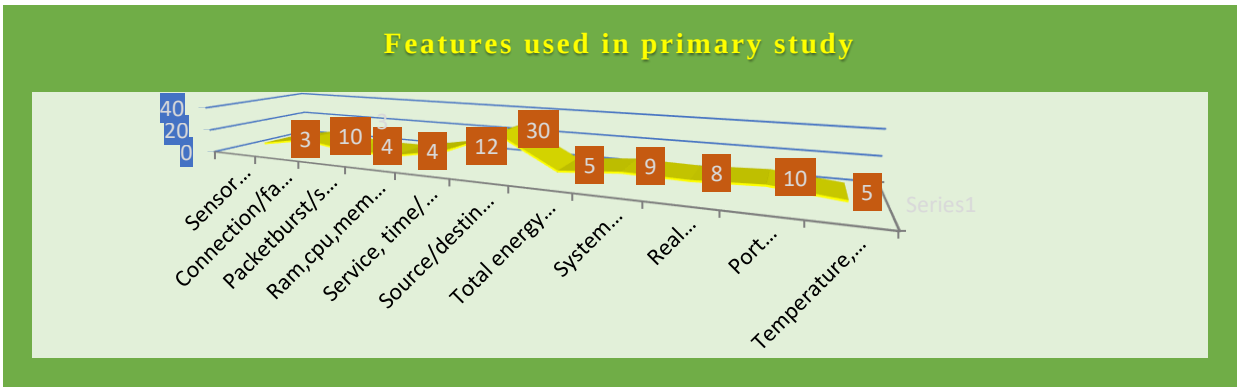

Fig 8: Line Graph for features used in the primary studies

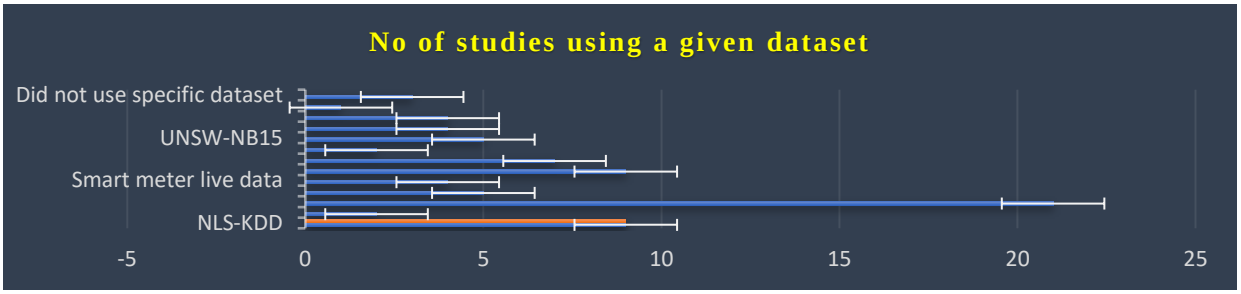

Fig 9: No of studies using a given dataset

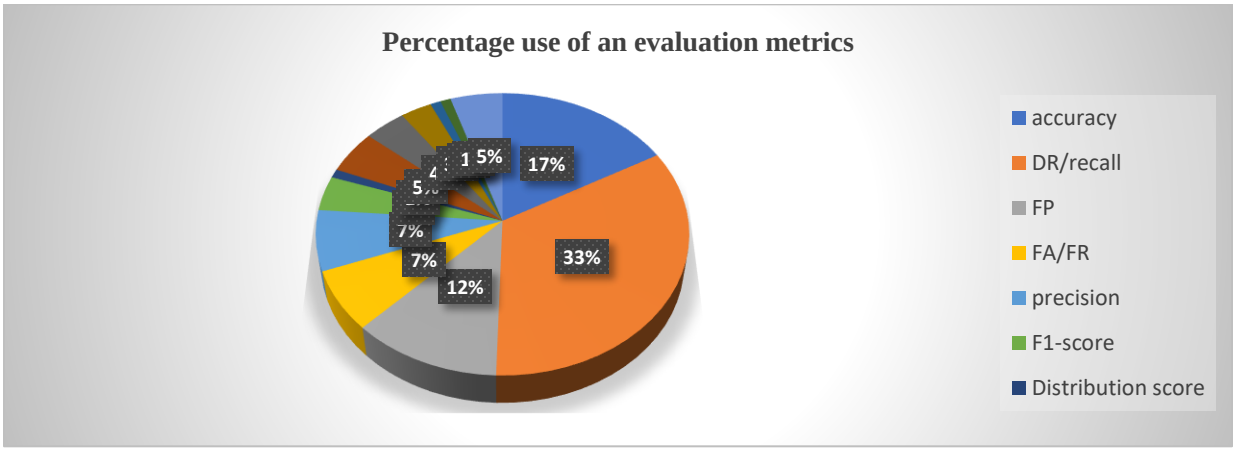

Fig 10: percentage use of an evaluation metrics

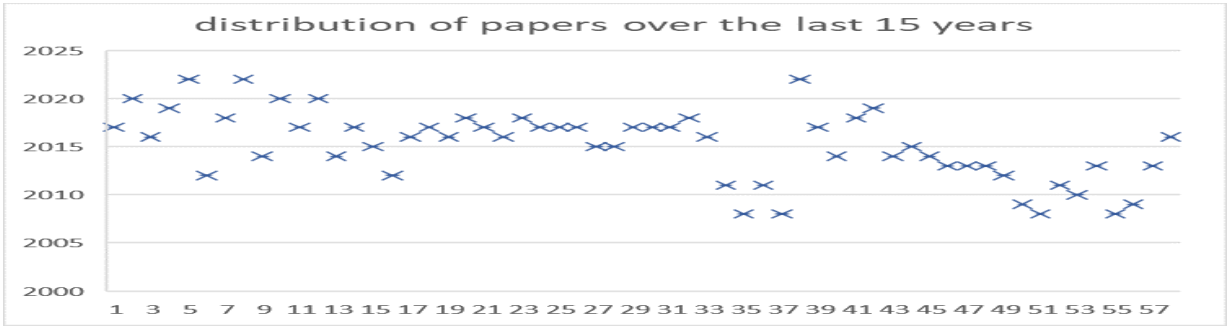

Fig 11: distribution of papers over the past 15 years

Table 6: Used datasets in primary studies

study	*Dataset*
St1 Philokypros Ioulianou 2017	*NSL-KDD Dataset*
st2 Chih-Che Sun et al 2020	*NS-3 simulation data exchanging*
St3 Mustafa Amir Faisal et al 2016	*KDD Cup 1999*
St4 Dipanjan Das Roy and Dongwan Shin 2019	*smart grid network traffic from NS-3 network simulator*
St5 Dejan Radovanovic etal 2022	*dataset of smart meter readings collected from 1,200 households in the Netherlands*
St6 Mueen Uddin etal 2012	*network traffic generated by the DARPA Intrusion Detection Evaluation (IDE) program*
St7 Morteza Behniafar etal 2018	*CICIDS2017 dataset, the IoT-23 dataset, and the N-BaIoT dataset*
St8 Priti Prabhakar et al 2022	*real-world dataset collected from a smart metering infrastructure installed*
St9 Farid Molazem Tabrizi and Karthik Pattabiraman 2014	*synthetic dataset generated using the simulator for the Open Automated Demand Response (OpenADR) standard* *real-world dataset collected from a smart metering infrastructure installed in a residential area*
St10 Mausumi Das Nath and Tapalina Bhattasali 2020	*KDD Cup 1999 dataset, the NSL-KDD dataset, and the UNSW-NB15 dataset*
St11 Tara Salman et al 2017	*synthetic dataset from CloudSim simulator, real-world dataset from a multi-cloud environment consisting of Amazon Web Services and Microsoft Azure*
St12 Nebrase Elmrabit et al 2020	*KDD Cup 1999 dataset, DARPA 1998*
St13 Liu et. al 2014	*synthetic dataset generated using the DARPA Intrusion Detection Evaluation Program (IDEVAL) dataset*
St14 Lyu et. al 2017	*KDD Cup 1999 dataset* *custom dataset collected from a real-world IoT network*
St15 Summerville et. Al 2015	*KDD Cup 1999 dataset, the UNSW-NB15 dataset, and a custom dataset collected from a real-world IoT network*
St16 Fu et al 2012	*real-world IoT network and the KDD Cup 1999 dataset*
St17 Tsitsiroudi et al 2016	*a simulation environment to generate synthetic datasets with varying numbers of nodes, network topologies, and wormhole configurations.*
St18 Bostani et al	*KDD Cup 1999 dataset and the IoT dataset*
St19 Trilles et al 2016	*(1) a smart city dataset containing air quality data, (2) a water quality dataset containing water quality data, and (3) a smart building dataset containing energy consumption data*
St20 Hoang et al 2018	*dataset of network traffic*
St21 Zissis et al 2017	*did not use a specific dataset in the paper. Instead, they proposed a framework for intelligent security on the edge of the cloud that could be applied to any network environment*
St22 Sedjelmaci et al 2016	*dataset of network traffic generated by IoT devices, which was obtained from a public dataset called the IoT-23 dataset*
St23 Granjal et al 2018	*authors did not use a specific dataset in the paper. Instead, they proposed an intrusion detection and prevention framework that could be applied to any CoAP(WSN*
St24 Domb et al 2017	*IoT-23 dataset.*
St25 Sedjelmaci et. al 2017	*did not use a specific*

St26	*Rajesh Kumar Gunupudi et al 2017*	*KDDCup99 dataset, NSL-KDD dataset, and UNSW-NB15 dataset*
St27	*Pongle et al 2015*	*network traffic data generated by simulating IoT devices and an attacker*
St28	*Trilles et al 2015*	*custom datase which consisted of simulated network traffic generated by a DDoS attack tool*
St29	*Moshtaghi et al 2017*	*Real-world datasets in including network traffic data, system call data, and sensor data.*
St30	*Yu et al 2017*	*custom dataset which consisted of simulated sensor data from a network of IoT devices*
St31	*Zhao et al 2017*	*KDDCup99 dataset*
St32	*Tama et al 2018*	*UNSW-NB15 dataset*
St33	*Briana Arrington et al 2016*	*IoT-23 dataset*
St34	*Deris Stiawan et al 2011*	*real-world NIPS system in their laboratory environment*
St35	*Yogitha et. al 2008*	*microcalcification clusters and normal tissue regions*
St36	*Wenke Lee et. al 2011*	*KDD Cup 1999 dataset*
St37	*Eleazar, Matthew et. al 2008*	*NSL-KDD dataset*
st38	*S.A. Joshi, et. al 2022*	*No specific dataset*
St39	*Sushil Kumar Chaturvedi, et. al 2017*	*KDD Cup 1999 dataset*
St40	*Omprakash Chandrakar, et. al 2014*	*NSL-KDD dataset,KDDCup'99 dataset,UNSW-NB15 dataset,CICIDS2017 dataset,DARPA1998 dataset,Kyoto 2006*
St41	*A.R. Jakhale, et. al 2018*	*No specific datasets*
St42	*R. Venkatesan, et al 2019*	*No data set*
St43	*Abhilasha A Sayar, et.al 2014*	*No data set*
St44	*S. S. Pawar et., al*	*KDD Cup 1999*
St45	*AnnkitaPatel,Risha Tiwari 2014*	*KDD Cup 1999 dataset and the NSL-KDD dataset*
St46	*Yogita B.Bhavsar& Kalyani C. Waghmare*	*KDD Cup 1999 dataset* *NSL-KDD dataset*
St47	*Rowayda A.Sadek, M. Sami Soliman& Hagar S. Elsayed 2013*	*KDD Cup 1999 dataset*
St48	*Ahemd A Elngar,Dowalt, Fayed*	*KDD Cup 1999 dataset*
St49	*HeshamAltwaijry, Saeed Algarny 2012*	*NSL-KDD dataset*
St50	*Tao Peng, WanliZuo 2009*	*KDD Cup 1999 dataset*
St51	*Wanke Lee,Salvotore J Stolfa 2008*	*dataset of UNIX process accounting data from a large commercial network*
St52	*Renuka DeviThanasekaran 2011*	*KDD Cup 1999 dataset*
St53	*G.C. Tjhai et al 2010*	*real-world industrial control system*
St54	*Yogita B et al 2013*	*KDDCup99 dataset*
St55	*WenkeLee et al 2008*	*DARPA 1998 dataset*
St56	*J. Viinikka et al 2009*	*real-world dataset collected from a large enterprise network*
St57	*Revathi S et al 2013*	*NSLKDD dataset*
St58	*Ibrahim Salim.M et al 2016*	*KDDCup99 dataset*

Table 7: Machine learning model and its use in the primary studies

Algorithm	Study
Feature selection	St (1,16,52)
Svm	St (2,7,12,17,18,19,21,27,31,35,36,38,40,41,44,45,46,50)
k-means	St (2,11,20,26,48,53)

Massive online analysis	St (3)
Random forest	St (4,7,10,12,19,32,37,38,40,45,57)
Clustering	St (5,16,36,39,41,51)
Bayesian	St (38,7)
KNN	St (7,12,37,38,41,44,10,40,57)
Decision tree,c 4.5,J48	St (7,12,15,19,27,36,37,38,39,40,41,44,45,48,50,51,55,57)
Neural networks	St (7,19,39,43,47,51,53)
Convolutional neural networks (CNN)	St (7,15,42)
Recurrent neural networks (RNN)	St (42,7,22,30)
Median absolute deviation (MAD)	St (8)
Hidden markov model	St (9)
Artificial neural network	St (10,38,40,41,43,52)
Optimum path forest	St (18)
Principle component analysis, RPCA	St (20,31,30)
Naïve Bayes	St (37,40,45,57)
Particle swarm optimization	St (32)
Artificial immune systems	St (33)
Fuzzy logic, neural fuzzy	St (43,22)
Time series	St (56)

Table 8: Evaluation results from each primary study

study	***accuracy***	***DR***	***FPR***	***FA***	***Distribution***	***Precision***	***F1 score***	***Time***	***TN***	***MIA***	***TP***
St1 Philokypros Ioulianou 2017	☒	*99.6*	*0.3*	☒	☒	☒	☒	☒	☒	☒	☒
st2 Chih-Che Sun et al 2020	*98.4*	☒	☒	*1.6*	☒	☒	☒	☒	☒	☒	☒
St3 Mustafa Amir Faisal et al 2016	☒	*97.7*	☒	*1.06*	☒	☒	☒	☒	☒	☒	☒
St4 Dipanjan Das Roy and Dongwan Shin 2019	☒	*99.23*	☒	☒	☒	☒	☒	☒	☒	☒	☒
St5 Dejan Radovanovic etal 2022	☒	☒	☒	☒	*4% for 25 & 1% for 100*	☒	☒	☒	☒	☒	☒
St6 Mueen Uddin etal 2012	☒	*98.62*	☒	*0.42*	☒	☒	☒	☒	☒	☒	☒
St7 Morteza Behniafar etal 2018	☒	☒	☒	☒	☒	☒	☒	☒	☒	☒	☒
St8 Priti Prabhakar et al 2022	*95.5*	*97.5*	☒	☒	☒	☒	☒	☒	☒	☒	☒
St9 Farid Molazem Tabrizi and Karthik Pattabiraman 2014	*99*	☒	☒	☒	☒	☒	☒	☒	☒	☒	☒

St10 Mausumi Das Nath and Tapalina Bhattasali 2020	*97.8*	*99.5*	☒	☒	☒	☒	☒	☒	☒	☒	☒
St11 Tara Salman et al 2017	☒	*90*	☒	☒	☒	☒	☒	☒	☒	☒	☒
St12 Nebrase Elmrabit et al 2020	*99.4*	*97.4*	☒	☒	☒	☒	☒	☒	☒	☒	☒
St13 Liu et. al 2014	☒	*91.7*	☒	☒	☒	*96.8*	☒	☒	☒	☒	☒
St14 Lyu et. al 2017	☒	*94.1*	☒	☒	☒	*95.2*	☒	☒	☒	☒	☒
St15 Summerville et. Al 2015	☒	*99.64*	☒	☒	☒	*99.93*	☒	☒	☒	☒	☒
St16 Fu et al 2012	☒	*94.8*	☒	☒	☒	*96.5*	☒	☒	☒	☒	☒
St17 Tsitsiroudi et al 2016	☒	*98.6*	*0.3*	☒	☒	☒	☒	☒	☒	☒	☒
St18 Bostani et al	☒	*95.4*	☒	☒	☒	*96.2*	*95.8*	☒	☒	☒	☒
St19 Trilles et al 2016	☒	*88*	☒	☒	☒	*94*	*91*	☒	☒	☒	☒
St20 Hoang et al 2018	*98*	*98*	☒	☒	☒	*94*	*96*	☒	☒	☒	☒
St21 Zissis et al 2017	☒	☒	☒	☒	☒	☒	☒	☒	☒	☒	☒
St22 Sedjelmaci et al 2016	☒	*90*	☒	☒	☒	☒	☒	☒	☒	☒	☒
St23 Granjal et al 2018	*94*	☒	☒	☒	☒	☒	☒	☒	☒	☒	☒
St24 Domb et al 2017	*96*	☒	☒	☒	☒	☒	☒	*20-30*	☒	☒	☒
St25 Sedjelmaci et. al 2017		*97*	*3*	☒	☒	☒	☒	☒	☒	☒	☒
St26 Rajesh Kumar Gunupudi et al 2017	☒	☒	☒	☒	☒	☒	*Above 0.98*	☒	☒	☒	☒
St27 Pongle et al 2015	☒	☒	☒	☒	☒	☒	☒	☒	☒	☒	*94*
St28 Trilles et al 2015	☒	*98*	*0.5*	☒	☒	☒	☒	☒	☒	☒	☒
St29 Mo shtaghi et al 2017	☒	☒	☒	☒	☒	☒	☒	☒	☒	☒	☒
St30 Yu et al 2017	☒	☒	*0.0331*	☒	☒	☒	☒	☒	☒	☒	*1.0*
St31 Zhao et al 2017	☒	*99.51*	☒	☒	☒	☒	☒	☒	☒	☒	☒

St32 Tama et al 2018	*97.62*	*99.6*	☒	☒	☒	☒	☒	☒	☒	☒	☒
St33 Briana Arrington et al 2016	☒	*99.95*	☒	☒	☒	☒	☒	☒	☒	☒	☒
St34 Deris Stiawan et al 2011	☒	☒	☒	☒	☒	☒	☒	☒	☒	☒	☒
St35 Yogitha et. al 2008	☒	*96.25*	☒	☒	☒	☒	☒	☒	☒	☒	☒
St36 Wenke Lee et. al 2011	☒	☒	☒	☒	☒	☒	☒	☒	☒	☒	☒
St37 Eleazar, Matthew et. al 2008	☒	*0.3*	*0.0005*	☒	☒	☒	☒	☒	☒	☒	☒
st38 Archika Jain, et. al 2022	*Btn 73.82-99.74*	☒	☒	☒	☒	☒	☒	☒	☒	☒	☒
St39 Sushil Kumar Chaturvedi, et. al 2017	☒	☒	☒	☒	☒	☒	☒	☒	☒	☒	☒
St40 Omprakash Chandrakar, et. al 2014	*Btn 95.6-99.56*	☒	☒	☒	☒	☒	☒	☒	☒	☒	☒
St41 A.R. Jakhale, et. al 2018	☒	☒	☒	☒	☒	☒	☒	☒	☒	☒	☒
St42 R. Venkatesan, et al 2019	☒	☒	☒	☒	☒	☒	☒	☒	☒	☒	☒
St43 Abhilasha A Sayar, et.al 2014	☒	☒	☒	☒	☒	☒	☒	☒	☒	☒	☒
St44 S. S. Pawar et., al	☒	*99.98*	*0.007*	☒	☒	☒	☒	☒	☒	☒	☒
St45 AnnkitaPatel,Risha Tiwari 2014	*99.89*	☒	☒	☒	☒	☒	☒	*31.17 sec*	☒	☒	☒
St46 Yogita B.Bhavsar& Kalyani C. Waghmare	☒	*98.57*	☒	☒	☒	☒	☒	☒	☒	☒	☒
St47 Rowayda A.Sadek, M. Sami Soliman& Hagar S. Elsayed 2013	☒	*96.7*	☒	*3*	☒	☒	☒	☒	☒	☒	☒
St48 Ahemd A Elngar,Dowalt, Fayed	☒	*98.2*	☒	☒	☒	☒	☒	*0.18 sec*	☒	☒	☒
St49 HeshamAltwaijry, Saeed Algarny 2012	☒	*93.0*	☒	☒	☒	☒	☒	☒	*99.7*	☒	☒
St50 Tao Peng, WanliZuo 2009	☒	*97.2*	☒	*0.75*	☒	☒	☒	☒	☒	☒	☒

St51 Wanke Lee,Salvotore J Stolfa 2008	*99.1*	☒	*1*	*0.6*	☒	☒	☒	☒	☒	☒	☒
St52 Renuka DeviThanasekaran 2011	☒	☒	☒	*0.0812*	☒	☒	☒	*46 sec*	☒	☒	☒
St53 G.C. Tjhai et al 2010	☒	*96*	☒	*78.8*	☒	☒	☒	☒	☒	☒	☒
St54 Bo Peng et al 2016	☒	☒	☒	☒	☒	☒	☒	☒	☒	*0.01120*	☒
St55 Shikha Singh and Angshul Majumdar 2018	☒	☒	☒	☒	☒	☒	☒	*43.9 sec*	☒	☒	☒
St56 J. Viinikka et al 2009	☒	☒	☒	☒	*0.2-0.3*	☒	☒	☒	☒	☒	☒
St57 Revathi S et al 2013	*71.1-99.8*	☒	☒	☒	☒	☒	☒	☒	☒	☒	☒
St58 Ibrahim Salim.M et al 2016	*80.14*	*78*	*21.83*	☒	☒	☒	☒	☒	☒	☒	☒

F. Number of Publications

In this SLR, 58 primary studies published on machine learning algorithms especially in the smart utility meters were selected. After analyzing the studies, we find that all studies have published between 2010 and 2023. Also, the number of studies is increased over the years. Figure 11 above shows the distribution of the studies over the years and figure 12 below shows the map of key word searches.

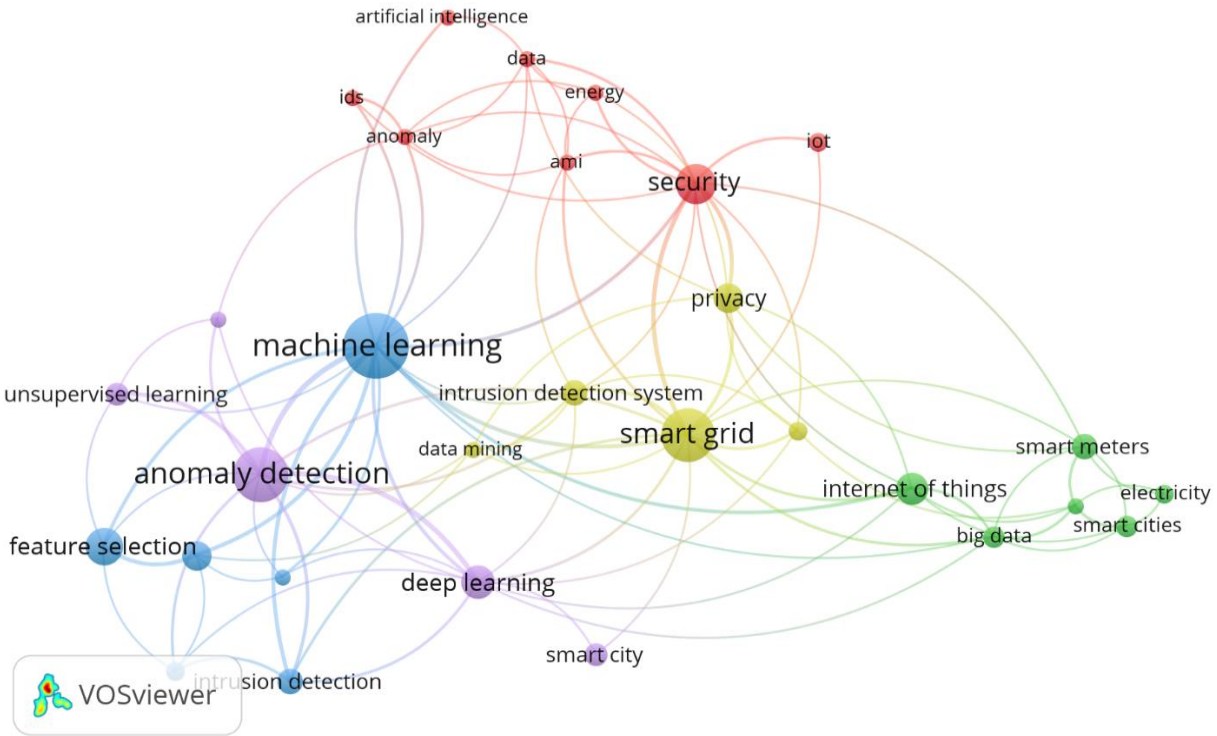

Fig 12: Key work map of the primary studies

Discussion

In this study, I have reviewed fifty eight primary studies on machine learning algorithms especially in smart utility meter intrusions that were published by 2008. This study also provides a summary of all algorithms used in machine learning, the various kinds of features more so those in smart meters, the datasets used in machine learning especially in networks and finally the findings in these studies and how the existing models are being evaluated. The primary data in this study is collected from all available research studies in the selected digital libraries. Machine learning is used to detect anomalies in energy consumption patterns that may indicate a malfunctioning meter or equipment. This can help to improve the accuracy of billing and reduce energy waste.

Following the predefined research questions in Section II, the first question is related to the machine learning algorithms used especially in smart meters. The SLR shows that though there is general use of most of the machine learning algorithms, in the smart meters the support vector machine and decision trees have more usage and this can be attributed to: known to be highly accurate and can achieve good performance in detecting attacks, both SVMs and decision trees can handle a wide range of input data types and can be applied to both supervised and unsupervised learning problems, both algorithms can be scaled up to handle large datasets and can be trained efficiently and this is the case with the enormous amount of data from these smart meters. However, use of medium absolute deviation (MAD), time series, hidden Markov and a few other machine learning models in smart meters was still challenging and this could be attributed to: these models can be more complex than simpler models like linear regression or decision trees, and may require more computational resources and expertise to implement and train, the performance of these models may not be significantly better than simpler models like decision trees or SVMs, and therefore may not be considered worth the added complexity.

The second question is related to the features that are mostly used in machine learning in the networks as well as in the smart meters. From the primary studies a number of features are available and exist when using machine learning and those that stood out were the packet based features such as source and destination among other packet features. In the smart meters it's these mentioned features alongside those that go hand in hand with the primary function of the smart which is energy consumption i.e (total, maximum, minimum, hourly and daily energy consumptions). Packet features are critical in smart meter machine learning algorithms because they provide information about the network traffic between the smart meters and the utility provider, which can be used to identify and detect anomalies or attacks.
By incorporating packet features into the analysis, machine learning algorithms can provide a more comprehensive view of the behavior of smart meter networks, which can help utilities to better manage energy supply and demand, and to detect and respond to potential security threats.

The third question is related to the datasets that are used in machine learning models. The KDD cup 1999 still remains the highly used dataset in most of the primary studies. The dataset has been widely used in the research community, which allows for more direct comparisons between different approaches and algorithms. However it does not include newer types of attacks and may not be representative of more modern network traffic patterns. Nonetheless, it remains a useful and widely used dataset for evaluating the effectiveness of machine learning algorithms for intrusion detection. There is also an emerging use of real environment data, as smart meters become more prevalent, there is a greater availability of real environment data for machine learning algorithms to train on. This makes it easier to collect and analyze large amounts

of data from a diverse range of households and businesses. Real environment data provides a more accurate representation of the actual energy usage patterns of households.

The fourth question is related to the evaluation metrics. These metrics are used to compare the model with other selected baseline models. Most of the primary studies have applied only one or two evaluation metrics. From these primary studies most of these used the Detection rate or recall metric. There was also a good use of accuracy and false positive. Detection rate or recall is a useful metric for evaluating the performance of machine learning models, particularly in scenarios where correctly identifying positive samples is crucial and the dataset is imbalanced.

In this SLR, a number of challenges related to ML algorithms in smart meters and propose the practices that can be performed to overcome these challenges.
Smart meters have the potential to revolutionize the way we consume and manage energy, but they also present several challenges when it comes to machine learning algorithms. As the number of smart meters grows, it becomes increasingly challenging to scale up machine learning algorithms to handle the increased data volume. Machine learning algorithms need to be designed to be scalable and efficient to handle the increasing amount of data generated by smart meters.
Many machine learning algorithms are complex and difficult to interpret, making it challenging to understand how they arrive at their decisions. Machine learning algorithms require access to sensitive data, raising privacy and security concerns. It is crucial to design machine learning systems that protect data privacy and prevent unauthorized access.

Machine learning algorithms can be computationally intensive and may not scale well to large datasets or distributed systems. Scalability is essential for the practical application of machine learning in real-world settings.
Machine learning algorithms require large amounts of high-quality data to train and build models. Poor quality or insufficient data can lead to inaccurate and unreliable models. Additionally, the data used for training must be representative of the real-world attacks that the IDS is expected to detect.

The selection of appropriate features is critical in intrusion detection. Machine learning algorithms require a set of relevant features to detect attacks accurately. However, the selection of the appropriate features can be challenging, as it requires an understanding of the attack patterns and the data available.

Study Limitations
This SLR has some limitations. First, I have used the predefined search string to find the relevant primary studies on machine learning in smart meters. Second, the primary studies are retrieved based on the selected digital libraries which may miss out on other library contributions. Moreover, English papers are selected only. Thirdly a time bound had been put on these papers to capture only those in the past fifteen years.

Conclusion and Future Work
This SLR was intended for identifying and analyzing the algorithms, datasets, features and evaluation metrics used in machine learning more so in the smart meters. A selection criteria was adopted for the selection of the primary studies used in this SLR. Out of all the journal, fifty eight primary studies were selected to be evaluated and extraction of information vital for this study. Extraction from the primary

studies was based on clearly stated objectives four in number. And after all the steps of conducting this SLR, below are the summarized main findings:

- The KDD cup of 1999 still remains the most used dataset in the selected primary studies of which out of the twelve used dataset it had a percentage of 27.6% which is the highest.
- Of all the twenty two algorithms under study, the SVM and decision tree each had the highest percentage of frequency of use each with a 16.5% which are the highest.
- a number of features exist for machine learning but the packet based are the highest used with the percentage of 30% while the smart meters the energy features are equally highly used with a percentage of 5%.
- The recall or detection rate and accuracy evaluation are the mostly used evaluation metrics each with a percentage of 33% and 17% respectively.
- Primary studies summarized table is shown below in Table 9. Also, the map of the primary studies that is designed by VOS viewer tool is shown in figure 12 above and this shows text mining functionality that can be used to construct and visualize co-occurrence networks of important terms extracted from a body of scientific literature (VOSviewer, 2020)"

Table 9: comparative review and summary of primary studies

study	***Retrieved***	***Reference***	***title***	***Algorithm***	***Dataset***	***features***	***Findings & resolved task***	***Detection techniques***	***Evaluation metrics***
St 1	*Institute of Electrical and Electronics Engineers*	(Ioulianou, 2017.)	*Smart Network-based Intrusion Detection System (SNIDS) for Advanced Metering Infrastructure*	*feature selection, filter, wrapper and embedded,*	*NSL-KDD Dataset, 2015*	*Network Traffic Data Meter Data Real-time Alerting*	*combine data from multiple sources, use machine learning algorithms and anomaly detection techniques, and provide real-time alerting to be effective in detecting and preventing intrusions in the AMI network*	*use threat intelligence feeds to identify any known threats or attack signatures*	*DR is 99.6 FP is o.3*
st 2	*Institute of Electrical and Electronics Engineers*	(Sun et al., 2020.)	*Intrusion Detection for Cybersecurity of Smart Meters*	*Support Vector Machine (SVM) One-Class SVM (OC-SVM) K-Means clustering algorithm*	*NS-3 simulation data exchanging*	*Sensor Report, Connection Attempt, Unknown Connection , Packet Burst, Firmware Modification, Unknown Application, RAM & cpu Demanding,*	*SVM classifier exhibits good performance with kernel functions. Compared to NN algorithms, SVM has an advantage in the shorter training time. This feature allows the proposed SVM model to be frequently updated to maintain a high level of detection accuracy.*	*Based on the TFPG technique, the pattern recognition algorithm is able to calculate the similarity index*	*detection rate of 98.4% and a false alarm rate of 1.6%*
St 3	*Institute of Electrical and Electronics Engineers*	(Faisal et al., 2014.)	*Data Stream-based Intrusion Detection System for Advanced Metering Infrastructure in Smart Grid: A Feasibility Study*	*data stream mining algorithms Massive Online Analysis (MOA) Hoeffding Tree*	*KDD Cup 1999*	*service duration time, data size to/from source/destination, various error rates*	*observed that some algorithms that uses very minimal amount of computing resources and offers moderate level of accuracy can potentially be used for the smart meter IDS. On the other hand, the algorithms that requires more computing resources and also offers higher accuracy levels can be useful for the IDSs in data concentrators and AMI headends*	*suggests using seven massive online analysis algorithms with the hoeffding tree for intrusion. This can help identify abnormal behavior in smart meters*	*detection rate of 97.7% and a false alarm rate of 1.06%*

St 4	*Institute of Electrical and Electronics Engineers*	(Roy et al., 2019.)	*Network Intrusion Detection in Smart Grids for Imbalanced Attack Types Using Machine Learning Models*	*Random Forest*	*smart grid network traffic from NS-3 network simulator*	*packet count, packet size, packet rate, and packet entropy.*	*The dataset used included normal traffic as well as several types of attacks, such as Denial-of-Service (DoS) attacks and spoofing attacks. resampling techniques such as oversampling and undersampling were used to handle imbalanced attack types in the dataset. These techniques helped balance the dataset and improve the accuracy of the RF algorithm.*	*RF algorithm to classify normal and abnormal network traffic. The RF algorithm can handle imbalanced datasets and identify anomaly*	*detection rate of 99.23%*
St 5	*Proceedings of the 11th DACH+ Conference on Energy Informatics*	(Radovanovic et al., 2022)	*How unique is weekly smart meter data*	*Clustering algorithm*	*dataset of smart meter readings collected from 1,200 households in the Netherlands*	*total energy consumption, the maximum energy consumption, and the energy consumption*	*The paper proposes a metric for quantifying the uniqueness of smart meter readings based on the clustering results. The metric is based on the probability of a random household being assigned to the same cluster as a specific household based on its energy consumption patterns.*	*K-Means algorithm to identify groups of smart meters with similar energy consumption patterns*	*Distribution of 4% and 1% for a 25 and a 100-household-sized set*
St 6	*Journal of Applied Sciences Research*	(Uddin et al., 2013)	*Signature-based Multi-Layer Distributed Intrusion Detection System using Mobile Agents.*	*signature detection algorithms*	*network traffic generated by the DARPA Intrusion Detection Evaluation (IDE) program*	*packet size, packet rate, and packet entropy*	*The paper proposes a multi-layer signature-based model that consists of several layers of signature detection algorithms. Each layer is responsible for detecting a specific type of attack, and the output of each layer is fed as input to the next layer.*	*paper proposes a method for generating signatures for each layer of the model*	*detection rate of 98.62% and a low false alarm rate of 0.42%*
St 7	*The ISC Int'l Journal of Information Security*	(Behnifar et al., 2018)	*A Survey of Anomaly Detection Approaches in Internet of Things*	*(mean/variance) Bayesian, (SVM, KNN, Decision trees, Random Forest, Neural Networks) Deep learning (autoencoders*	*CICIDS2017 dataset, the IoT-23 dataset, and the N-BaIoT dataset*	*Network-based features include packet size, packet inter-arrival time, packet length, and packet direction. Host-based features include system call traces,*	*the survey provides a comprehensive overview of the various anomaly detection approaches for IoT systems*	*Hybrid methods that combine multiple approaches were used*	*N/A*

				, Convolutional Neural Networks, Recurrent Neural Networks) Hybrid methods		*file system logs, and registry keys. Application-based features include login attempts, error messages, and HTTP status codes.*			
St 8	*Hindawi Security and Communication Networks*	(Prabhakar et al., 2022.)	*Cyber Security of Smart Metering Infrastructure Using (MAD)*	*Median Absolute Deviation (MAD)*	*real-world dataset collected from a smart metering infrastructure installed*	*hourly energy consumption data*	*The authors used the real-world dataset collected from a smart metering infrastructure installed in a residential area to develop a cyber security smart meter using MAD*	*MAD measures variability of a dataset and is useful for identifying outliers*	*Detect accuracy of 97.5% and replay attacks with an accuracy of 95.5%.*
St 9	*researchgate*	(Tabrizi et al., 2014.)	*A Model-Based Intrusion Detection System for Smart Meters*	*Hidden Markov Model (HMM)*	*synthetic dataset generated using the simulator for the Open Automated Demand Response (OpenADR) standard real-world dataset collected from a smart metering infrastructure installed in a residential area*	*energy consumption data metadata such as the time of day and day of the week*	*The authors used a model-based approach that involves building a mathematical model of the system and using it to detect anomalies in the data. The authors also considered the presence of events such as power outages or maintenance as part of the model.*	*Specifically, they used a Hidden Markov Model (HMM) to model the normal behaviour of the smart meter and detect deviations from this behaviour*	*Detection accuracy of over 99%.*
St 10	*Azerbaijan Journal of High-Performance Computing*	(Das Nath et al., 2020)	*Anomaly Detection Using Machine Learning Approaches*	*K-Nearest Neighbor, Support Vector Machines, Random*	*KDD Cup 1999 dataset, the NSL-KDD dataset, and the UNSW-NB15 dataset*	*network-based features such as packet size, protocol type, and duration,*	*the proposed machine learning-based approach showed promising results in detecting anomalies in network traffic data.*	*using ML algorithms for anomaly detection provides a new perspective in the field*	*Detection accuracy of 99.5%, probing attacks with an*

				Forest, and Artificial Neural Networks (ANNs)		*host-based features such as system call traces and file access patterns*			*accuracy of 97.8%, and U2R attacks with an accuracy of 95.3%.*
St 1 1	*Institute of Electrical and Electronics Engineers (ieee)*	(Salman et al., 2017.)	*Machine Learning for Anomaly Detection and Categorization in Multi-cloud Environments*	*One-Class Support Vector Machines (OCSVM) and Autoencoder Neural Networks (AENN) K-Means clustering algorithm*	*synthetic dataset from CloudSim simulator, real-world dataset from a multi-cloud environment consisting of Amazon Web Services and Microsoft Azure*	*CPU usage, memory usage, disk usage, and network traffic*	*The authors also considered the correlation between their features in their approach. Overall, the proposed machine learning-based approach showed promising results in detecting and categorizing anomalies in multi-cloud environments*	*Use of OCSVM,AENN & K-Means provides a new perspective in the field of anomaly detection for multi-cloud environments.*	*DR 90%.*
St 1 2	*International Conference on Cyber Security and Protection of Digital Services (Cyber Security)*	(Elmrabit et al., 2020.)	*Evaluation of Machine Learning Algorithms for Anomaly Detection*	*Decision Trees, Random Forest, Support Vector Machines (SVM), and K-Nearest Neighbor (KNN)*	*KDD Cup 1999 dataset,* *DARPA 1998*	*bytes and packets, protocol type, and duration of the communication system call dataset features number of arguments, system call type, and the process ID*	*Overall, the study found that Random Forest and SVM performed better than the other machine learning algorithms for detecting anomalies in both network traffic and system call data.*	*The study also showed that machine learning approaches outperformed the rule-based approach for anomaly detection*	*accuracy of 99.4% and a detection rate of 97.4%.*
St 1 3	*In Software Engineering and Service Science (ICSESS), 2014 5th IEEE International Conference*	(Liu et al., 2014.)	*A lightweight anomaly mining algorithm in the internet of things*	*Data aggregation anomaly detection*	*synthetic dataset generated using the DARPA Intrusion Detection Evaluation Program (IDEVAL) dataset*	*packet length, inter-arrival time, packet size, and protocol type*	*proposed algorithm is based on the concept of frequent pattern mining, which involves identifying patterns that occur frequently in the dataset. It consists of preprocessing the dataset to extract features and transform the data into a transaction database.*	*applying a set of rules to the frequent itemsets to identify anomalies.*	*precision of 96.8% and a recall of 91.7%*

St 14	*IEEE Internet of Things Journal*	(Lyu et al., 2017.)	*Fog-empowered anomaly detection in iot using hyperellipsoidal clustering*	*hyperellipsoidal clustering*	*KDD Cup 1999 dataset custom dataset collected from a real-world IoT network*	*combination of packet header information and network flow statistics.*	*algorithm can be used in Fog computing environments to efficiently detect anomalies in large datasets generated by IoT networks.*	*anomalies are detected based on the distance between the data and the hyperellipsoids.*	*precision of 95.2% and a recall of 94.1%*
St 15	*Computing and Communications Conference (IPCCC)*	(Summerville et al., 2015.)	*Ultra-lightweight deep packet anomaly detection for internet of things devices*	*Deep Packet Inspection* *Convolutional Neural Network (CNN) decision tree*	*KDD Cup 1999 dataset, the UNSW-NB15 dataset, and a custom dataset collected from a real-world IoT network*	*packet header information and network flow statistics*	*The authors also showed that the proposed algorithm is lightweight and suitable for deployment on IoT devices, as it can achieve high accuracy with a small number of parameters and low computational overhead.*	*decision tree is used to classify the extracted features as normal or anomalous.*	*precision of 99.93% and a recall of 99.64%*
St 16	*International Journal of Computer Applications*	(Fu et al., 2011)	*An intrusion detection scheme based on anomaly mining in internet of things*	*clustering algorithm feature selection algorithm*	*real-world IoT network and the KDD Cup 1999 dataset*	*packet header information, network flow statistics, and protocol-specific features such as Modbus function codes*	*The results showed that the proposed algorithm outperformed other state-of-the-art anomaly detection algorithms, including k-means clustering, one-class SVM, and Isolation Forest.*	*network traffic is compared to the model, and anomalies are detected based on the distance between the traffic and the model.*	*precision of 96.5% and a recall of 94.8%.*
St 17	*Wireless and Mobile Networking Conference (WMNC)*	(Tsitsiroudi et al., 2016.)	*A mobile application for visual-assisted wormhole attack detection in iot-enabled wsns*	*SURF (Speeded Up Robust Features) algorithm* *RANSAC (Random Sample Consensus) algorithm* *SVM*	*a simulation environment to generate synthetic datasets with varying numbers of nodes, network topologies, and wormhole configurations.*	*images captured by the mobile device.*	*The authors used SURF algorithm to extract and match features from the images captured by the mobile device.* *The extracted features were used to estimate the transformation between the two images, which was used to detect the presence of a wormhole attack.*	*authors compared their approach with existing approaches in the literature and demonstrated its superiority*	*detection accuracy of 98.6% and a false positive rate of 0.3%.*

St 18	*Computer Communications journal*	(Bostani et al., 2017.)	*Hybrid of anomaly-based and specification-based ids for internet of things using unsupervised opf based on mapreduce approach*	*One-class Support Vector Machine (SVM) classifier based on the Optimum-Path Forest (OPF) algorithm*	*KDD Cup 1999 dataset and the IoT dataset*	*number of packets, packet size, source IP address, destination IP address, protocol type, and time duration*	*authors used the OPF algorithm to cluster the data into normal and abnormal groups, and then applied the One-class SVM algorithm to classify the test data. authors report that their approach achieved high detection rates for both the KDD Cup 1999 dataset and the IoT dataset, with low false positive rates*	*The approach achieves high detection rates with low false positives and shows promise for detecting intrusions in real-world IoT networks.*	*precision of 96.2%, a recall of 95.4%, and an F1 score of 95.8%.*
St 19	*Telecommunication International Symposium*	(Trilles et al., 2016)	*A domain-independent methodology to analyze iot data streams in real-time. a proof of concept implementation for anomaly detection from environmental data*	*decision trees, random forests, support vector machines (SVMs), and neural networks*	*(1) a smart city dataset containing air quality data, (2) a water quality dataset containing water quality data, and (3) a smart building dataset containing energy consumption data*	*statistical features, frequency domain features, and time-domain*	*authors report that their methodology can be applied to various domains, such as smart cities, smart buildings, and environmental monitoring. They also show that their methodology can be used to identify patterns and anomalies in data streams and to make predictions about future values of the data.*	*The methodology is evaluated using real-world datasets and various modeling techniques, and the results show promise for identifying patterns and anomalies in data streams*	*precision of 94%, a recall of 88%, and an F1 score of 91%.*
St 20	*Advanced Communication Technology (ICACT) 8 20th International Conference*	(Hoang et al., 2018.)	*A pcabased method for iot network traffic anomaly detection*	*Principal Component Analysis (PCA) K-means clustering*	*dataset of network traffic*	*Statistical moments of packet counts and sizes, inter-arrival times, and durations.*	*method was able to effectively detect anomalous traffic patterns in the IoT network. The method was able to identify various types of anomalies, including TCP SYN floods, UDP floods, and botnet traffic.*	*method outperformed traditional statistical anomaly detection methods, such as the Mahalanobis distance method*	*accuracy of 98%, a precision of 94%, a recall of 98%, and an F1 score of 96%.*
St 21	*Engineering, Technology and Innovation (ICE/ITMC),*	(Zissis et al 2017.)	*Intelligent security on the edge of the cloud*	*SVM*	*did not use a specific dataset in the paper. Instead, they proposed a*	*packet header information, flow duration, and payload size*	*framework includes a multi-layer security model that can detect and respond to network attacks in real-time.*	*Blockchain technology: used to ensure secure communication and data sharing*	*N/A*

	International Conference				*framework for intelligent security on the edge of the cloud that could be applied to any network environment*		*use of machine learning algorithms at the edge of the network can reduce the need for centralized processing and improve the scalability and efficiency of the network.*	*between edge devices.*	
St 2 2	*Communications (ICC), 2016 IEEE International Conference*	(Sedjelmaci et al., 2016.)	*A lightweight anomaly detection technique for low-resource iot devices: A game-theoretic methodology*	*Neuro-fuzzy* *Recurrent Neural Network (RNN)* *Variational Autoencoder (VAE)*	*dataset of network traffic generated by IoT devices, which was obtained from a public dataset called the IoT-23 dataset*	*packet header information, including source and destination IP addresses, ports, and protocol types.*	*proposed technique was found to be effective in detecting various types of anomalies, including botnet traffic, DDoS attacks, and port scanning.* *proposed technique is designed to be lightweight and can be implemented on low-resource IoT devices, making it suitable for deployment in IoT environments with limited computational resources.*	*features were preprocessed and transformed using VAE, the RNN was then used to model the temporal behavior of the devices and detect anomalies based on the transformed features.*	*DR 90%*
St 2 3	*Security and Communication Networks journal*	(Granjal et al., 2018.)	*An intrusion detection and prevention framework for internet integrated coap wsn*	*Threshold based method* *reputation-based mechanism*	*authors did not use a specific dataset in the paper. Instead, they proposed an intrusion detection and prevention framework that could be applied to any CoAP(WSN*	*packet header information, including source and destination IP addresses, ports, and protocol types*	*framework integrates machine learning and reputation-based mechanisms.* *framework can improve the security and reliability of CoAP-based WSNs, which are becoming increasingly important in the context of the Internet of Things.*	*Reputation-based mechanism: used to identify suspicious nodes and block them from participating in the network.*	*Accuracy 94%*

St 24	*Journal of Information Security and Applications*	(Dombet al., 2017.)	*Lightweight adaptive randomforest for iot rule generation and execution*	*Adaptive Random Forest (ARF)algorithm*	*IoT-23 dataset.*	*packet header information*	*Authors proposed a lightweight adaptive random forest algorithm for rule generation and execution in IoT devices.* *features were used to train the ARF model to generate and execute rules for the IoT devices.* *found to be effective in detecting anomalous behavior and generating rules to mitigate the risk of security breaches.*	*Anomaly detection algorithm was used to identify anomalous behavior in the data.*	*accuracy rate of 96% and an execution time of 20-30 milliseconds*
St 25	*IEEE Transactions on Vehicular Technology*	(Sedjelmaci et al., 2017.)	*An accurate security game for lowresource iot devices*	*Game theory* *iterative algorithm*	*did not use a specific*	*number of vulnerabilities, the severity of the vulnerabilities, and the number of successful attacks*	*authors proposed a game-theoretic approach to optimize the security of IoT devices with limited resources.* *approach was found to be effective in finding the optimal strategy for the defender, which maximizes the security of the IoT device while minimizing the cost of protection*	*A two-player game between an attacker and a defender, where the attacker's goal is to compromise and the defender's goal is to protect*	*DR 97%, FPR 3%*
St 26	*Future Generation Computer Systems journal*	(Gunupudi et al., 2017.)	*A self-constructing feature clustering approach for anomaly detection*	*K-means clustering algorithm and a distance-based outlier detection algorithm*	*KDDCup99 dataset, NSL-KDD dataset, and UNSW-NB15 dataset*	*packet header information, including source and destination IP addresses, ports*	*The approach involves clustering similar features together to create a set of clusters, and then identifying anomalies based on their distance from the centroid of the cluster.* *authors argue that the proposed approach is suitable for real-time anomaly detection in high-speed networks, as it is designed to be scalable and efficient.*	*proposed approach is also self-adaptive, meaning that it can adapt to changes in the network traffic patterns over time*	*F1-score of over 0.98*

St 27	*International Journal of Computer Applications*	(Pongle et al., 2015)	*Real time intrusion and wormhole attack detection in internet of things*	*C4.5 decision tree algorithm and the Support Vector Machine (SVM) algorithm*	*network traffic data generated by simulating IoT devices and an attacker*	*packet size, packet rate, and packet content.*	*Approach involves monitoring network traffic in real-time and identifying anomalous behavior using a set of predefined rules and machine learning algorithms.*	*Blackhole and wormhole detection*	*True Positive Detection 94%*
St 28	*Springer ,International Conference on Context-Aware Systems and Applications*	(Machaka et al., 2016)	*Using the cumulative sum algorithm against distributed denial of service attacks in internet of things*	*CUSUM algorithm*	*custom datase which consisted of simulated network traffic generated by a DDoS attack tool*	*packet rate and packet size*	*Proposed a method for detecting Distributed Denial of Service (DDoS) attacks in IoT networks using the Cumulative Sum (CUSUM) algorithm. Approach was found to be more effective than several other state-of-the-art DDoS detection techniques, including threshold-based and statistical-based methods.*	*Involves monitoring network traffic in real-time and applying the CUSUM algorithm to identify sudden increases in traffic volume*	*DR 98% FPR 0.5%*
St 29	*International Journal of Intelligent Systems*	(Moshtaghi et al., 2017)	*Exponentially weighted ellipsoidal model for anomaly detection*	*Exponentially weighted ellipsoidal*	*Real-world datasets in including network traffic data, system call data, and sensor data.*	*packet size and protocol type, temperature and humidity readings*	*Authors proposed an anomaly detection approach based on the Exponentially Weighted Ellipsoidal Model (EWEM). The EWEM is a statistical model that represents the normal behavior of a system as an ellipsoid in a high-dimensional space.*	*It is designed to be adaptable and robust to changes in the normal behavior*	*N/A*
St 30	*IEEE Internet of Things Journal*	(Yu et al., 2017.)	*Recursive principal component analysis-based data outlier detection and sensor data aggregation in iot systems*	*Recursive Principal Component Analysis (PCA)*	*custom dataset which consisted of simulated sensor data from a network of IoT devices*	*temperature, humidity, and light level readings.*	*Proposed a recursive approach to update the PCA model as new data is collected, which allows the approach to adapt to changes in the sensor data over time.*	*Involves applying PCA to sensor data to identify outliers and detect abnormal sensor readings.*	*TPR 1.0000, FPR 0.0331*

St 3 1	*Dependable, Autonomic and Secure Computing, Intl Conf on Pervasive Intelligence & Computing*	(Zhao et al., 2017)	*A dimension reduction model and classifier for anomaly-based intrusion detection in internet of things*	*Principal Component Analysis (PCA)* *Support Vector Machine (SVM)*	*KDDCup99 dataset*	*Source and destination IP addresses, protocol types, and service types.* *number of bytes sent and received, and the duration of the connection*	*Authors proposed an approach based on dimension reduction and classification for anomaly-based intrusion detection in IoT systems. Approach can help improve the security and reliability of IoT systems by detecting and preventing malicious network traffic.*	*Features were used to reduce dimensionality and improve the accuracy of the anomaly detection.*	*DR 99.51%*
St 3 2	*MATEC Web of Conferences*	(Tama et al., 2018.)	*An integration of pso-based feature selection and random forest for anomaly detection in iot network*	*particle swarm optimization (PSO)* *random forest classification*	*UNSW-NB15 dataset*	*set of 49 network traffic features, including source and destination IP addresses, packet length, and port numbers*	*outperformed several other state-of-the-art anomaly detection techniques, including k-nearest neighbor, decision tree, and support vector machine classifiers. authors argue that the PSO-based feature selection method helps to reduce the dimensionality of the network traffic data, while the random forest classifier helps to improve the accuracy and reliability of the anomaly detection.*	*dataset includes both normal and anomalous network traffic, and is designed to simulate realistic attack scenarios in a variety of network environments.*	*DR accuracy of 99.6% on the UNSW-NB15 dataset and 97.62% on the real-world IoT dataset*
St 3 3	*IEEE Computer Communication and Networks (ICCCN),International Conference*	(Arrington et al., 2016.)	*Behavioral modeling intrusion detection system (bmids) using internet of things (iot) behavior-based anomaly detection via immunity-inspired algorithms*	*negative selection algorithms and artificial immune system (AIS)*	*IoT-23 dataset*	*18 network traffic features, including source and destination IP addresses, packet length, and protocol type*	*Authors argue that the BMIDS approach is effective for detecting both known and unknown threats in IoT networks, and can be used in a variety of IoT applications, such as smart homes and healthcare systems.*	*approach is based on the immune system's ability to recognize and respond to foreign antigens*	*DR accuracy 99.95%*

St 3 4	*International Journal of Computer Applications*	(Stiawan et al., 2011)	*Characterizing Network Intrusion Prevention System*	*statistical methods and data analysis techniques, such as descriptive statistics and hypothesis testing*	*real-world NIPS system in their laboratory environment*	*packet rate and packet size*	*authors found that the NIPS system had a high detection rate for known attacks and intrusions, but was less effective in detecting unknown or novel attacks.*	*Analyzed the NIPS system's performance in terms of its ability to accurately detect and prevent network intrusions*	*N/A*
St 3 5	*IEEE transactions on medical imaging*	(El-Naqa et al., 2008.)	*A Support Vector Machine Approach for Detection of Microcalcifications*	*Support Vector Machine (SVM)*	*microcalcification clusters and normal tissue regions*	*Noise, co-occurrence matrix and Gabor filters*	*study shows that the SVM approach can be an effective tool for the detection of microcalcifications in mammograms, which can aid in the early diagnosis of breast cancer*	*applied a pre-processing step to enhance the contrast and remove noise from the mammograms.*	*DR 96.25%*
St 3 6	*International Symposium on Distributed Computing and Applications to Business, Engineering and Science*	(Distributed et al., 2011.)	*Study on Multi-Grade Intrusion Detection Model Based on Data Mining Technology*	*decision trees, clustering algorithms, and support vector machines (SVMs)*	*KDD Cup 1999 dataset*	*connection duration, service type, number of failed login attempts, and number of outbound bytes*	*find that the SVM algorithm performs particularly well in detecting certain types of intrusions, such as DoS attacks.*	*Identify algorithms and features most effective in detecting different types of intrusions*	*NA*
St 3 7	*Researchgate*	(Eskin et al., 2008)	*Adaptive Model Generation for Intrusion Detection Systems*	*decision trees, random forests, Naive Bayes, and k-nearest neighbors (KNN).*	*NSL-KDD dataset*	*Connection duration, number of failed login attempts, and number of outbound packets.*	*results show that the adaptive models generated by their approach outperformed the static models that use a fixed set of features and machine learning algorithm*	*approach involves three main stages: feature selection, model generation, and model adaptation.*	*DR 0.3, FPR 0.0005*
st 3 8	*ResearchGate International Conference on Intelligent Engineering and Management (ICIEM)*	(Jain et al., 2022.)	*A Review: Data Mining Classification Techniques*	*decision trees, Bayesian networks, artificial neural networks, k-nearest neighbors (KNN), support vector machines*	*No specific dataset*	*wrapper, filter*	*They conclude that there is no single best algorithm that works well for all types of datasets and classification tasks.* *However, they note that decision trees and SVMs are among the most widely used and effective classification algorithms.*	*discuss the importance of feature selection and extraction in classification tasks.*	*73.82-99.74*

				(SVM), and random forests.					
St 39	*International journal of advanced research in science & engineering*	(Professor, 2015)	*A Survey-Comparative Study on Intrusion Detection System*	*decision trees, support vector machines, clustering, neural networks, and genetic algorithms*	*KDD Cup 1999 dataset*	*network traffic data, system call data, and application data.*	*the paper suggests that future research should focus on developing more robust and scalable intrusion detection systems that can handle the increasing volume and complexity of network traffic data.*	*statistical-based techniques, machine learning-based techniques, and hybrid techniques that combine multiple methods*	*NA*
St 40	*International Journal of Advanced Research in Computer and Communication Engineering*	(Kalarani & Selva Brunda, 2014)	*A Survey on Efficient Data Mining Techniques for Network Intrusion Detection System (IDS)*	*Decision Trees Naive Bayes k-Nearest Neighbors (k-NN) Support Vector Machines (SVM) Artificial Neural Networks (ANN) Random Forest Gradient Boosting*	*NSL-KDD dataset,KDD Cup'99 dataset,UNSW-NB15 dataset,CICIDS2017 dataset,DARPA 1998 dataset,Kyoto 2006+ dataset*	*source IP, destination IP, protocol type, source port, destination port,number of packets, number of bytes, duration,payload content, keywords, mean, standard deviation, entropy*	*Decision Trees, Naive Bayes, and k-NN are suitable for real-time intrusion detection due to their simplicity and low computational requirements. SVM and ANN are more suitable for detecting complex and non-linear intrusion patterns. Random Forest and Gradient Boosting can improve the accuracy of intrusion detection, but at the cost of increased computational requirements. NSL-KDD dataset is widely used as a benchmark dataset for evaluating intrusion detection techniques.*	*Ensemble techniques, such as bagging and boosting, can improve the performance of intrusion detection systems.*	*Accuracy 95.6-99.56%*
St 41	*Researchgate International Journal of Scientific Research in Computer Science, Engineering and Information Technology*	(Hussain & Priscilla, 2018)	*A Survey on Various Kinds of Anomalies Detection Techniques in the Mobile Adhoc Network Environment*	*Decision Tree, Artificial Neural Network (ANN), Support Vector Machine (SVM), and K-Nearest*	*No specific datasets*	*traffic flow, packet size, hop count, inter-packet delay, and energy consumption*	*the authors note the importance of considering resource constraints in MANETs when selecting and designing anomaly detection techniques.*	*suggest that a combination of different techniques, including signature-based, anomaly-based, and specification-based approaches*	*NA*

				Neighbor (KNN) classifiers, clustering,statistical techniques					
St 4 2	*Google scholar Journal of service research*	(Venkatesan et al., 2019.)	*A Modular Accelerator Generator for Neural Networks*	*convolutional neural networks, and recurrent neural networks*	*No data set*	*The modular accelerator generator presented in the paper has several features. It can generate hardware accelerators for different neural network architectures*	*The paper shows that the generated accelerators achieve high performance and energy efficiency compared to software-based implementations of the same neural network architectures.*	*accelerator is used to customize the design for different neural network architectures, allowing users to optimize performance and energy efficiency for specific applications*	*NA*
St 4 3	*International Journal of Computer Science and Mobile Computing*	(Sayar et al., 2014)	*A Review of Intrusion Detection Systems in Computer Network*	*artificial intelligence, fuzzy logic and neural network*	*No data set*	*network traffic analysis, protocol analysis, and host-based monitoring. machine learning, such as clustering and classification algorithms*	*provides a comprehensive review of various IDS algorithms and techniques. It also highlights the importance of selecting the appropriate IDS approach*	*authors recommend future research in developing IDSs that can detect unknown attacks and in enhancing IDS performance while reducing false positives*	*NA*
St 4 4	*Research gates*	(Perez et al., 2017.)	*Intrusion detection in computer networks using hybrid machine learning techniques*	*J48 decision tree, K-Nearest Neighbor (KNN),Support Vector Machine (SVM)*	*KDD Cup 1999 dataset*	*failed login attempts, the duration of the connection, and the protocol type*	*outperforms individual algorithms in terms of detection rate and false positive rate. achieves an overall detection rate of 99.98% and a false positive rate of 0.007%.*	*demonstrates the effectiveness of using a combination of machine learning algorithms in improving the accuracy of IDSs.*	*detection rate of 99.98% and a false positive rate of 0.007%.*
St 4 5	*International Journal of Computer Applications*	(Patel & Tiwari, 2014)	*Bagging Ensemble Technique for intrusion Detection System*	*Decision Tree (DT), Naive Bayes (NB), Random Forest (RF), and Support Vector Machine (SVM)*	*KDD Cup 1999 dataset and the NSL-KDD dataset*	*protocol type, service, and flag, count, duration, error rate.*	*The BET model achieved better accuracy and detection rate compared to the individual base classifiers and other ensemble models such as AdaBoost and Random Subspace*	*proposed BET model is effective in detecting network intrusion and can improve the performance of intrusion*	*Accuracy 99.89%, 31.17sec*

St 46	*Semantic scholar*	(Bhavsar et al., 2008)	*Intrusion Detection System Using Data Mining Technique Support Vector Machine*	*Data Mining, SVM*	*KDD Cup 1999 dataset NSL-KDD dataset*	*Preprocessing steps included normalization, feature scaling, and one-hot encoding*	*The SVM model achieved an accuracy of 98.81% on the NSL-KDD dataset and 97.51% on the KDD Cup 1999 dataset*	*The SVM model was able to accurately detect various types of attacks, including DoS, Probe, and R2L, with low false positive rates.*	*DR 98.57 %*
St 47	*International Journal of Computer Science Issues*	(Sadek et al., 2013.)	*Effective Anomaly Intrusion Detection System based on Neural Network with Indicator Variable and Rough set Reduction*	*neural network*	*NLS KDD Cup 1999 dataset*	*protocol type, service, flag, and number of failed login attempts.*	*achieved an accuracy of 96.79% on the NLSKDD Cup 1999 dataset, which is higher than the accuracy achieved by other state-of-the-art anomaly detection systems.*	*use of indicator variable and rough set reduction techniques led to a significant reduction in the number of false alarms*	*detection accuracy about 96.7% with a false alarm rate of 3%.*
St 48	*Arab Journal Platform*	(Elngar et al., 2013)	*A Real Time Anomaly Intrusion Detection System with High Accuracy*	*K-means C4.5 decision tree and the Chi-square*	*KDD Cup 1999 dataset*	*source IP address, destination IP address, protocol type, number of failed logins, and number of file creations.*	*achieves high accuracy in detecting anomalies, with a detection rate of 99.76% and a false positive rate of 0.001%. The system also has a low detection time and can detect attacks in real-time*	*uses statistical techniques such as the mean and standard deviation to normalize the data.*	*DR accuracy (98.2%) and improving the speed to 0.18 sec*
St 49	*Elsevier Journal of King Saud University – Computer and Information Sciences*	(Altwaijry, 2013)	*Bayesian based Intrusion Detection System*	*Naïve Bayes algorithm*	*NSL-KDD dataset*	*protocol type, source and destination addresses, and service type,the number of connections, the number of failed logins, and the duration of the connection, the time of day and the day of the week*	*compared their approach with other classification algorithms, such as k-Nearest Neighbor (k-NN) and Support Vector Machines (SVM), and found that the proposed system outperformed these*	*Bayesian-based intrusion detection system is effective in detecting various types of attacks in computer networks.*	*TN = 99.7%, DR= 93.0%*
St 50	*International Journal of Computer Science and*	(T. Peng &	*Data mining Intrusion Detection*	*Decision tree and SVM*	*KDD Cup 1999 dataset*	*protocol type, source and destination addresses, source*	*proposed intrusion detection system achieved high detection rates with low false positive rates in real-time*	*authors attributed the system's performance to the use of data mining*	*DR= 97.2%, FR= 0.75%*

	Network Security	Zuo, 2006)	*System in real time*			*and destination ports, duration of the connection*	*monitoring of network traffic. The system also outperformed other traditional intrusion detection systems, such as Snort*	*algorithms and the combination of frequent itemsets and classification rules*	
St 5 1	*International journal of computer applications*	(Li et al., 2019)	*Advanced Data Mining and Applications*	*clustering, decision trees, and neural networks*	*dataset of UNIX process accounting data from a large commercial network*	*user ID, process ID, CPU usage, disk activity, and network traffic*	*the adaptive nature of their approach allowed it to adapt to changes in system behavior over time, making it more robust to new types of attacks.*	*approach uses a combination of statistical and machine learning algorithms*	*accuracy = 99 1%, FP=1%,FN=0.6%*
St 5 2	*International Journal of Information Technology Convergence and Services (IJITCS)*	(Renuka DeviThanasekaran, 2011)	*A Robust & Efficient Real Time Network Intrusion Detection System Using Artificial Neural Network in Data Mining*	*Artificial Neural Network (ANN) backpropagation algorithm feature selection algorithm*	*KDD Cup 1999 dataset*	*source and destination IP addresses, source and destination ports, and packet payload to train the ANN model*	*system achieved an accuracy of 97.5% in detecting network intrusions on the KDD Cup 1999 dataset, outperforming other machine learning and data mining based intrusion detection systems*	*feature selection algorithm was found to improve the detection accuracy of the system by reducing the number of features used in training the ANN model*	*FA= 0.0812, time training 46 sec*
St 5 3	*Computers & Security journal*	(Tjhai et al., 2010.)	*A preliminary two-stage alarm correlation and filtering system using SOM neural network and K-means algorithm*	*neural network and K-means algorithm*	*real-world industrial control system*	*Two-stage alarm correlation Self-Organizing Maps*	*proposed system was able to effectively cluster and filter alarms from a real-world ICS, reducing the number of alarms to a manageable amount for operators to handle*	*SOM proved to be effective in reducing the number of alarms and identifying related alarms*	*78.8% of false alarms, 96% DR*
St 5 4	*IEEE International MultiDisciplinary Conference on Cognitive Methods in Situation Awareness and Decision Support*	(Peng et al., 2016.)	*A Two-stage Pattern Recognition Method for Electric Customer Classification in Smart Grid*	*clustering algorithm, Density-Based Spatial Clustering, k-Nearest Neighbor (k-NN)*	*real-world smart grid system*	*electricity usage of each customer over a period of time*	*first stage of the method using DBSCAN is able to group customers into meaningful clusters, while the second stage using k-NN is able to accurately assign customers to different classes*	*Density-Based Spatial Clustering of Applications with Noise (DBSCAN) is used to group customers based on their consumption patterns*	*MIA= 0.01120*

St55	IEEE transactions on smart grid	(Singh et al., 2017.)	Deep Sparse Coding for Non-Intrusive Load Monitoring	deep sparse coding architecture	REDD dataset	electricity consumption data	quality of the clustering depends on the initial centroid positions, with better initial positions leading to better clustering results	approach is able to accurately disaggregate the overall electricity consumption into individual appliance-level consumption for most of the appliances in the dataset	Run times=43.9 secs
St56	Information Fusion Journal	(Viinikka et al., 2009)	Processing intrusion detection alert aggregates with time series modeling	Time series	real-world dataset collected from a large enterprise network	Alert count, Alert types,Time of day,Day of the week,Source and destination IP addresses,Network traffic volume,	use of time series modeling, trend analysis approach, and evaluation metrics helps to improve the accuracy and efficiency of the method	trend analysis approach helps to identify patterns and trends in the alert data and detect abnormal behaviors	0.2–0.3% of intervals
St57	International Journal Engineering Research and Technology (IJERT)	(Revathi et al., 2013.)	A detailed analysis on NSLKDD dataset using various machine learning techniques for intrusion detection	Decision Tree, Random Forest, Naive Bayes, K-Nearest Neighbor, and Support Vector Machine	NSLKDD dataset	Source and destination IP addresses, Packet size and duration,Port number,Number of failed login attempts, Number of outbound packets	Random Forest and Support Vector Machine perform better than other techniques in terms of detection rate and false alarm rate	feature selection and oversampling techniques help to improve the performance of the technique	Accuranc e range of 71.1-99.8
St58	IEEE International Conference on Engineering and Technology (ICETECH)	Ibrahim Salim. M et al 2016.)	A Study on IDS for preventing Denial of service attack using Outliers techniques	Outliers techniques Local Outlier Factor (LOF) and Isolation Forest	KDDCup99 dataset	Time of day,Packet direction,Packet type,Port numbers,Source and destination IP addresses,Packet rate,Packet size	combination of LOF with other machine learning algorithms, such as Decision Trees and Random Forests, can further improve the accuracy of the IDS system.	system was able to achieve high detection rates for DoS attacks while maintaining a low false positive rate.	Accuracy =80.14%, FPR=21.83%,DR=78%

References

1. R. T.-J. of I. T. C. and, & 2011, undefined. (2011). A Robust and Efficient Real Time Network Intrusion Detection System Using Artificial Neural Network In Data Mining. *Citeseer*, *1*(4). https://doi.org/10.5121/ijitcs.2011.1402
2. Altwaijry, H. (2013). Bayesian based intrusion detection system. *Lecture Notes in Electrical Engineering*, *170 LNEE*, 29–44. https://doi.org/10.1007/978-94-007-4786-9_3
3. Arrington, B., Rufus, R., Esterline, A., Arrignton, B., & Barnett, L. (n.d.). Behavioral modeling intrusion detection system (BMIDS) using internet of things (IoT) behavior-based anomaly detection via immunity-inspired algorithms. *Ieeexplore.Ieee.Org*. https://doi.org/10.1109/ICCCN.2016.7568495
4. Behniafar, M., Nowroozi, A., ISeCure, H. S.-, & 2018, undefined. (2018). A survey of anomaly detection approaches in internet of things. *Researchgate.Net*. https://www.researchgate.net/profile/Alireza-Nowroozi-2/publication/327417983_A_Survey_of_Anomaly_Detection_Approaches_in_Internet_of_Things/links/5ebe8ad192851c11a867f565/A-Survey-of-Anomaly-Detection-Approaches-in-Internet-of-Things.pdf
5. Bhavsar, Y., Emerging, K. W.-I. J. of, & 2013, undefined. (2008). Intrusion detection system using data mining technique: Support vector machine. *Academia.Edu*, *9001*(3). https://www.academia.edu/download/44581943/project1.pdf
6. Boell, S., Information, D. C.-K.-J. of, & 2015, undefined. (2015). On being 'systematic'in literature reviews in IS. *Journals.Sagepub.Com*, *30*(2), 161–173. https://doi.org/10.1057/jit.2014.26
7. Bostani, H., Communications, M. S.-C., & 2017, undefined. (n.d.). Hybrid of anomaly-based and specification-based IDS for Internet of Things using unsupervised OPF based on MapReduce approach. *Elsevier*. Retrieved April 15, 2023, from https://www.sciencedirect.com/science/article/pii/S0140366416306387
8. Carcary, M., Doherty, E., European, G. C.-E. 2018 17th, & 2018, undefined. (n.d.). An analysis of the systematic literature review (SLR) approach in the field of IS management. *Books.Google.Com*. Retrieved April 15, 2023, from https://books.google.com/books?hl=en&lr=&id=gU9mDwAAQBAJ&oi=fnd&pg=PA78&dq=SLR+(Okoli,+2015):++&ots=Euj-e0i0VZ&sig=6XrBhegDjMdnpt499Ai2BVgIDQs
9. Das Nath, M., Bhattasali, T., Das, M., & Xavier's College, S. (2020). *Anomaly Detection Using Machine Learning Approaches*. https://doi.org/10.32010/26166127.2020.3.2.196.206
10. Distributed, H. A.-2011 10th I. S. on, & 2011, undefined. (n.d.). Study on Multi-grade Intrusion Detection Model Based on Data Mining Technology. *Ieeexplore.Ieee.Org*. Retrieved April 15, 2023, from https://ieeexplore.ieee.org/abstract/document/6118730/
11. Domb, M., Bonchek-Dokow, E., Security, G. L.-J. of I., & 2017, undefined. (n.d.). Lightweight adaptive Random-Forest for IoT rule generation and execution. *Elsevier*. Retrieved April 15, 2023, from https://www.sciencedirect.com/science/article/pii/S2214212616302332
12. Elmrabit, N., Zhou, F., Li, F., … H. Z. conference on, & 2020, undefined. (n.d.). Evaluation of machine learning algorithms for anomaly detection. *Ieeexplore.Ieee.Org*. Retrieved April 15, 2023, from https://ieeexplore.ieee.org/abstract/document/9138871/
13. El-Naqa, I., Yang, Y., … M. W.-I. transactions on, & 2002, undefined. (n.d.). A support vector machine approach for detection of microcalcifications. *Ieeexplore.Ieee.Org*. Retrieved April 15, 2023, from https://ieeexplore.ieee.org/abstract/document/1176643/
14. Elngar, A. A., El A Mohamed, D. A., M Ghaleb, F. F., Elngar, A., El Mohamed, A. A., & Ghaleb, F. M. (2013). A real-time anomaly network intrusion detection system with high accuracy. *Digitalcommons.Aaru.Edu.Jo*, *2*(2), 49. https://doi.org/10.12785/isl/020201
15. Engineering, D. Z., (ICE, T. and I., & 2017, undefined. (n.d.). Intelligent security on the edge of the cloud. *Ieeexplore.Ieee.Org*. Retrieved April 15, 2023, from https://ieeexplore.ieee.org/abstract/document/8279999/
16. Eskin, E., Miller, M., Zhong, Z.-D., Yi, G., Lee, W.-A., & Stolfo, S. (2000). *Adaptive model generation for intrusion detection systems*. https://academiccommons.columbia.edu/doi/10.7916/D8GX4J9V
17. Faisal, M., Aung, Z., … J. W.-I. S., & 2014, undefined. (n.d.). Data-stream-based intrusion detection system for advanced metering infrastructure in smart grid: A feasibility study. *Ieeexplore.Ieee.Org*. Retrieved April 15, 2023, from https://ieeexplore.ieee.org/abstract/document/6720175/?casa_token=RtVwPRVhiw0AAAAA:6KF4Aw3ZmkyXbNOQkCFnyBUdVbsSJLi6sb_E0UNzwxft5lHtxbeLbiLcQda53G4nDVbC_wMUF8FC2w
18. Fu, R., Zheng, K., Zhang, D., & Yang, Y. (2011). An intrusion detection scheme based on anomaly mining in internet of things. *IET Conference Publications*, *2011*(591 CP), 315–320. https://doi.org/10.1049/CP.2011.1014
19. Granjal, J., Networks, A. P.-S. and C., & 2018, undefined. (n.d.). An intrusion detection and prevention framework for internet-integrated CoAP WSN. *Hindawi.Com*. Retrieved April 15, 2023, from https://www.hindawi.com/journals/scn/2018/1753897/
20. Gunupudi, R., Nimmala, M., … N. G.-F. G., & 2017, undefined. (n.d.). CLAPP: A self constructing feature clustering approach for anomaly detection. *Elsevier*. Retrieved April 15, 2023, from https://www.sciencedirect.com/science/article/pii/S0167739X16308718
21. Hoang, D., conference, H. N.-2018 20th I., & 2018, undefined. (n.d.). A PCA-based method for IoT network traffic anomaly detection. *Ieeexplore.Ieee.Org*. Retrieved April 15, 2023, from https://ieeexplore.ieee.org/abstract/document/8323766/
22. Hussain, N., & Priscilla, M. (2018). *A Survey on Various Kinds of Anomalies Detection Techniques in the Mobile Adhoc Network Environment.* https://www.researchgate.net/profile/A-M-J-Niyaz-

Hussain/publication/324721255_A_Survey_on_Various_Kinds_of_Anomalies_Detection_Techniques_in_the_Mobile_Adhoc_Network_Environment/links/5adf044b0f7e9b285943ad3a/A-Survey-on-Various-Kinds-of-Anomalies-Detection-Techniques-in-the-Mobile-Adhoc-Network-Environment.pdf

23. Ioulianou, P. (n.d.). Smart Network-based Intrusion Detection System (SNIDS) for Advanced Metering Infrastructure. *Studentnet.Cs.Manchester.Ac.Uk*. Retrieved April 15, 2023, from http://studentnet.cs.manchester.ac.uk/resources/library/thesis_abstracts/MSc17/FullText/Ioulianou-Philokypros-diss.pdf
24. Jain, A., Somwanshi, D., … K. J.-2022 3rd I., & 2022, undefined. (n.d.). A Review: Data Mining Classification Techniques. *Ieeexplore.Ieee.Org*. Retrieved April 15, 2023, from https://ieeexplore.ieee.org/abstract/document/9853036/
25. Kalarani, P., & Selva Brunda, S. (2014). Survey on Efficient Data Mining Techniques for Network Intrusion Detection System (IDS). In *International Journal of Advanced Research in Computer and Communication Engineering* (Vol. 3, Issue 9). www.ijarcce.com
26. Kitchenham, B. A. (2012). *Systematic review in software engineering*. 1–2. https://doi.org/10.1145/2372233.2372235
27. Li, J., Wang, S., Qin, S., Li, X., & Wang, S. (2019). *Advanced Data Mining and Applications: 15th International Conference, ADMA 2019, Dalian, China, November 21–23, 2019, Proceedings*. https://books.google.com/books?hl=en&lr=&id=56O-DwAAQBAJ&oi=fnd&pg=PR5&dq=Data+mining+Intrusion+Detection+System+in+real+time+by+Tao+Peng,+Wanli Zuo+&ots=k1_Vssjvwl&sig=d56cArFfhfHyUn43XLor-BILQbw
28. Liu, Y., on, Q. W.-2014 I. 5th I. C., & 2014, undefined. (n.d.). A lightweight anomaly mining algorithm in the Internet of Things. *Ieeexplore.Ieee.Org*. Retrieved April 15, 2023, from https://ieeexplore.ieee.org/abstract/document/6933768/?casa_token=gp2ksoWCsVsAAAAA:uH7gi_fwB-vKzC85Lqz9b1pOz3A9hlvU3r7LW_CpZr6vRo6E76iHIIauyaZfp8_QKYVyaoYHv3oDuw
29. Lyu, L., Jin, J., Rajasegarar, S., … X. H.-I. I. of, & 2017, undefined. (n.d.). Fog-empowered anomaly detection in IoT using hyperellipsoidal clustering. *Ieeexplore.Ieee.Org*. Retrieved April 15, 2023, from https://ieeexplore.ieee.org/abstract/document/7936471/
30. Machaka, P., McDonald, A., … F. N.-C.-A. S., & 2016, undefined. (2016). Using the cumulative sum algorithm against distributed denial of service attacks in internet of things. *Springer*, *165*, 62–72. https://doi.org/10.1007/978-3-319-29236-6_7
31. Moshtaghi, M., Erfani, S., … C. L.-I. J. of, & 2017, undefined. (2017). Exponentially weighted ellipsoidal model for anomaly detection. *Wiley Online Library*, *32*(9), 881–899. https://doi.org/10.1002/int.21875
32. on, T. R.-2016 I. I. C., & 2016, undefined. (n.d.). A study on IDS for preventing denial of service attack using outliers techniques. *Ieeexplore.Ieee.Org*. Retrieved April 15, 2023, from https://ieeexplore.ieee.org/abstract/document/7569352/?casa_token=5c1QxmryQuoAAAAA:gAA0SPJkNglfGzKM90C3ZWGb_yik5e93MK7hrGsgDoiJani1E3RfR5sdN3aTy80w0nokeyeV8fY2wQ
33. Patel, A., & Tiwari, R. (2014). BAGGING ENSEMBLE TECHNIQUE FOR INTRUSION DETECTION SYSTEM. In *International Journal For Technological Research In Engineering* (Vol. 2, Issue 4). www.ijtre.com
34. Peng, B., Wan, C., Dong, S., Lin, J., … Y. S.-… on S. G., & 2016, undefined. (n.d.). A two-stage pattern recognition method for electric customer classification in smart grid. *Ieeexplore.Ieee.Org*. Retrieved April 15, 2023, from https://ieeexplore.ieee.org/abstract/document/7778853/?casa_token=oJsEujGLjKkAAAAA:g2nGekjvrahiiBzz4Whsfvfy8w6GwPgJspfEW2M5J6UlvS3FOsb-0tm2FN284q9-N_WGrPDSl94D1w
35. Peng, T., & Zuo, W. (2006). Data Mining for Network Intrusion Detection System in Real Time. In *IJCSNS International Journal of Computer Science and Network Security* (Vol. 6, Issue 2).
36. Perez, D., Astor, M., … D. A.-2017 X. L. A., & 2017, undefined. (n.d.). Intrusion detection in computer networks using hybrid machine learning techniques. *Ieeexplore.Ieee.Org*. Retrieved April 15, 2023, from https://ieeexplore.ieee.org/abstract/document/8226392/
37. Pongle, P., Applications, G. C.-I. J. of C., & 2015, undefined. (2015). Real time intrusion and wormhole attack detection in internet of things. *Citeseer*, *121*(9), 975–8887. https://citeseerx.ist.psu.edu/document?repid=rep1&type=pdf&doi=3d237886c50e4663311e59012593fe5253eb387a
38. Prabhakar, P., Arora, S., Khosla, A., … R. B.-S. and, & 2022, undefined. (n.d.). Cyber Security of Smart Metering Infrastructure Using Median Absolute Deviation Methodology. *Hindawi.Com*. Retrieved April 15, 2023, from https://www.hindawi.com/journals/scn/2022/6200121/
39. Professor, A. (2015). A Survey-Comparative Study on Intrusion Detection System. *International Journal of Advanced Research in Computer and Communication Engineering*, *4*. https://doi.org/10.17148/IJARCCE.2015.4794
40. Radovanovic, D., Unterweger, A., Eibl, G., Engel, D., & Reichl, J. (2022). How unique is weekly smart meter data? *Energy Informatics*, *5*. https://doi.org/10.1186/S42162-022-00205-8
41. Revathi, S., … A. M. E. R. & T. (IJERT, & 2013, undefined. (n.d.). A detailed analysis on NSL-KDD dataset using various machine learning techniques for intrusion detection. *Academia.Edu*. Retrieved April 15, 2023, from https://www.academia.edu/download/65451210/a_detailed_analysis_on_nsl_kdd_dataset_using_IJERTV2IS120804.pdf
42. Roy, D., Information, D. S.-2019 I. C. on, & 2019, undefined. (n.d.). Network intrusion detection in smart grids for imbalanced attack types using machine learning models. *Ieeexplore.Ieee.Org*. Retrieved April 15, 2023, from https://ieeexplore.ieee.org/abstract/document/8939744/?casa_token=oTk-jaHPRFcAAAAA:NV78nzBs9q-vHW1YCe5-MM_i1dzkNTHF9qKNe-_blepmlTR33KWnuDohbXhn6mVosoZJr68Z0uI71w

43. Sadek, R., Soliman, M., Computer, H. E.-I. J. of, & 2013, undefined. (n.d.). Effective anomaly intrusion detection system based on neural network with indicator variable and rough set reduction. *Citeseer*. Retrieved April 15, 2023, from https://citeseerx.ist.psu.edu/document?repid=rep1&type=pdf&doi=529308f1120942793939dfe2b146fdc151cabb66
44. Salman, T., Bhamare, D., Erbad, A., … R. J.-2017 I. 4th, & 2017, undefined. (n.d.). Machine learning for anomaly detection and categorization in multi-cloud environments. *Ieeexplore.Ieee.Org*. Retrieved April 15, 2023, from https://ieeexplore.ieee.org/abstract/document/7987183/?casa_token=-lPj14T1XNAAAAAA:qxndiVLd78-ArmCJXxBVkBtr9kgAnPSy7yw_7NmecO6Z-ebPiU2G2RlJSTa77LZP8QsGZ2rKlx8QdQ
45. Sayar, A., Pawar, S., Computer, V. M.-I. J. of, & 2014, undefined. (2014). A review of intrusion detection system in computer network. *Academia.Edu*, *3*(2), 700–703. https://www.academia.edu/download/33100399/V3I2201499a46.pdf
46. Sedjelmaci, H., … S. S.-2016 I. international, & 2016, undefined. (n.d.). A lightweight anomaly detection technique for low-resource IoT devices: A game-theoretic methodology. *Ieeexplore.Ieee.Org*. Retrieved April 15, 2023, from https://ieeexplore.ieee.org/abstract/document/7510811/
47. Sedjelmaci, H., … S. S.-I. T. on, & 2017, undefined. (n.d.). An accurate security game for low-resource IoT devices. *Ieeexplore.Ieee.Org*. Retrieved April 15, 2023, from https://ieeexplore.ieee.org/abstract/document/7920414/
48. Singh, S., Grid, A. M.-I. T. on S., & 2017, undefined. (n.d.). Deep sparse coding for non–intrusive load monitoring. *Ieeexplore.Ieee.Org*. Retrieved April 15, 2023, from https://ieeexplore.ieee.org/abstract/document/7847445/?casa_token=rCJQRDYDXY8AAAAA:F5JACJ9N2LK6xW-nU0Yj5Qx68mFr27uw2olgobxbnaOD4IgDM6XyLCF-YA2huK-DS8oSgEQ3FfL6Ig
49. Stiawan, D., Idris, M., of, A. A.-I. J., & 2011, undefined. (2011). Characterizing network intrusion prevention system. *Repository.Unsri.Ac.Id*, *14*(1), 975–8887. https://repository.unsri.ac.id/8337/
50. Summerville, D., … K. Z.-2015 I. 34th, & 2015, undefined. (n.d.). Ultra-lightweight deep packet anomaly detection for Internet of Things devices. *Ieeexplore.Ieee.Org*. Retrieved April 15, 2023, from https://ieeexplore.ieee.org/abstract/document/7410342/
51. Sun, C., Cardenas, D., … A. H.-I. T. on, & 2020, undefined. (n.d.). Intrusion detection for cybersecurity of smart meters. *Ieeexplore.Ieee.Org*. Retrieved April 15, 2023, from https://ieeexplore.ieee.org/abstract/document/9144275/?casa_token=aIszSZKCPxUAAAAA:Mi_bdm3BkZJEJh5ipBTXWkrnwCEAs6DsFTTD2ovMNxBJ-HdHsmJ0Mqf9OOrnEqvyXeuG2yAr2xfTxQ
52. Tabrizi, F., International, K. P.-2014 I. 15th, & 2014, undefined. (n.d.). A model-based intrusion detection system for smart meters. *Ieeexplore.Ieee.Org*. Retrieved April 15, 2023, from https://ieeexplore.ieee.org/abstract/document/6754583/?casa_token=SJsuD8V_UXwAAAAA:LnigfrkAipvrhAv3zaUMZC8eGHhxuL-I0FWZpMux3RR9v4rmwGtnZlNy7mep3Cz7kB5llFbtvgophg
53. Tama, B., Conferences, K. R.-M. W. of, & 2018, undefined. (n.d.). An integration of pso-based feature selection and random forest for anomaly detection in iot network. *Matec-Conferences.Org*. https://doi.org/10.1051/matecconf/201815901053
54. Tjhai, G., Furnell, S., Papadaki, M., Security, N. C.-C. &, & 2010, undefined. (n.d.). A preliminary two-stage alarm correlation and filtering system using SOM neural network and K-means algorithm. *Elsevier*. https://doi.org/10.1016/j.cose.2010.02.001
55. Trilles, S., Belmonte, Ò., … S. S.-I. J. of, & 2017, undefined. (2016). A domain-independent methodology to analyze IoT data streams in real-time. A proof of concept implementation for anomaly detection from environmental data. *Taylor & Francis*, *10*(1), 103–120. https://doi.org/10.1080/17538947.2016.1209583
56. Tsitsiroudi, N., … P. S.-2016 9th I., & 2016, undefined. (n.d.). EyeSim: A mobile application for visual-assisted wormhole attack detection in IoT-enabled WSNs. *Ieeexplore.Ieee.Org*. Retrieved April 15, 2023, from https://ieeexplore.ieee.org/abstract/document/7543976/
57. Uddin, M., Rehman, A. A., Uddin, N., Memon, J., Alsaqour, R., & Kazi, S. (2013). Signature-based Multi-Layer Distributed Intrusion Detection System using Mobile Agents. *Academia.Edu*, *15*(1), 79–87. https://www.academia.edu/download/69250334/Signature-based_Multi-Layer_Distributed_20210909-2089-hf0acs.pdf
58. Venkatesan, R., Shao, Y., … M. W.-2019 I., & 2019, undefined. (n.d.). Magnet: A modular accelerator generator for neural networks. *Ieeexplore.Ieee.Org*. Retrieved April 15, 2023, from https://ieeexplore.ieee.org/abstract/document/8942127/?casa_token=gUVeH3po2VIAAAAA:l9F5TKk9mifP8TmmjYuwpzWBpawhCBBcemI6v3POUeP5D5AvRplsdNGEsto7ClCA1pXi-Jdp8bav-Q
59. Viinikka, J., Debar, H., Mé, L., Lehikoinen, A., Fusion, M. T.-I., & 2009, undefined. (2009). Processing intrusion detection alert aggregates with time series modeling. *Elsevier*. https://doi.org/10.1016/j.inffus.2009.01.003
60. Yu, T., Wang, X., Journal, A. S.-I. I. of T., & 2017, undefined. (n.d.). Recursive principal component analysis-based data outlier detection and sensor data aggregation in IoT systems. *Ieeexplore.Ieee.Org*. Retrieved April 15, 2023, from https://ieeexplore.ieee.org/abstract/document/8048463/
61. Zhao, S., Li, W., Zia, T., on, A. Z.-2017 I. 15th I. C., & 2017, undefined. (2017). A dimension reduction model and classifier for anomaly-based intrusion detection in internet of things. *Ieeexplore.Ieee.Org*. https://doi.org/10.1109/DASC-PICom-DataCom-CyberSciTec.2017.141