# A Comprehensive Review and Meta-Analysis on Applications of Machine Learning Techniques in Intrusion Detection

**Manojit Chattopadhyay**
Indian Institute of Management
Raipur India
mchattopadhyay@iimraipur.ac.in

**Rinku Sen**
NSHM College of Management and Technology
Kolkata India

**Sumeet Gupta**
Indian Institute of Management
Raipur India

## Abstract

Securing a machine from various cyber-attacks has been of serious concern for researchers, statutory bodies such as governments, business organizations and users in both wired and wireless media. However, during the last decade, the amount of data handling by any device, particularly servers, has increased exponentially and hence the security of these devices has become a matter of utmost concern. This paper attempts to examine the challenges in the application of machine learning techniques to intrusion detection. We review different inherent issues in defining and applying the machine learning techniques to intrusion detection. We also attempt to identify the best technological solution for the changing usage pattern by comparing the different machine learning techniques on different datasets and summarizing their performance using various performance metrics. This paper highlights the research challenges and future trends of intrusion detection in dynamic scenarios of intrusion detection problems in diverse network technologies.

**Keywords**: Intrusion detection, machine learning, soft computing, dataset, performance metrics, cyber-infrastructure, mobile communications, mobile systems, security, wireless technology

## 1 Introduction

Intrusion detection deals with detecting any malicious activities that may jeopardise the system. The attack can be an internal attack, modification of the important system files categorized as host-based intrusion detection or attack coming from the network system commonly known as the network intrusion-based detection system. Apart from these types, any deviation of software performance from normal functioning is also termed as anomaly-based intrusion detection. The intrusion detection has been a field of study for around four decades. Since 1972, securing data has been a matter of concern for researchers. For example, Anderson *et al*., (1995) discussed about the security problems faced by United States Air Force (USAF) operations and administration. In 1980, he prepared a detection system from an audit log file. Between 1984 and 1986, Denning and Neumann (1985) developed the first real time

intrusion detection. Since then various researchers applied different techniques to mitigate the more sophisticated form of attacks increasing day by day.

The information security company – Hold Security – reported recovery of 360 million account credentials for web services from the black market (Meyer, 2017). Seals (2015) reported that the world wide distributed denial of service attack doubled from the 1st quarter of year 2014 to starting of year 2015. Companies such as eBay suffered a massive attack where 233 million users' personal databases were hacked. The heartbleed encryption bug affected about 17% of the Internet's secure web servers by making passwords vulnerable which was protected by SSL/TLS encryption (McCartney, 2014; McGregor, 2014).

The intrusion detection process encompasses various set of activities. The most difficult task is to demarcate a clear boundary between normal and malicious activities in a system. In the late 80s and early 90s, statistical approach, expert system, and time series models were capable enough of building a detection model. However, they failed to perform well in a complex highly correlated dataset and their performance degraded with noisy data. Machine learning techniques with self-learning ability can effectively overcome the above limitation. Research has been conducted on how data mining, soft computing (Fuzzy logic), and statistical (Bayesian network) approaches have been applied to solve the intrusion detection problem with discussion on the concept of honeypots to detect intrusion (Lee *et al*., 200; Kabiri *et al*., 2005). Applications of computational intelligence techniques, datasets and performance evaluation has also been surveyed in literature (Wu and Banzhaf., 2010). However, there is a dearth of literature on the detailed quantitative analysis of the vast number of machine learning algorithms successfully applied for intrusion detection.

The aim of this review is therefore to perform a comprehensive analysis of machine learning techniques, and in the process try to find out the applications of the most popular machine learning techniques on the most popular dataset. The effectiveness of these techniques on datasets in terms of performance metrics has been also evaluated. Lastly, we assess different challenges prevailing in this field and identify possible research direction.

The novelty of the present study is in performing the numerical analysis on the applied machine learning techniques to elucidate the success of intrusion detection from widely published literature and provide a trend for machine learning approaches to detect attack.

We divide our study into the following sections. In section 2, an overview of intrusion detection technology is discussed. Section 3 introduces current state-of-the-art techniques in intrusion detection. The results of the critical review are provided in section 4 and subsequently discussed in section 5. Finally, the in-depth observation on surveyed outcomes concludes in section 6.

# 2 Literature review

## 2.1 Intrusion Detection

The basic function of an intrusion detection system is to monitor an activity taking place in a system and to generate an alarm report stating whether an attack is happening or whether everything is normal. Figure 1 depicts this aspect of intrusion detection.

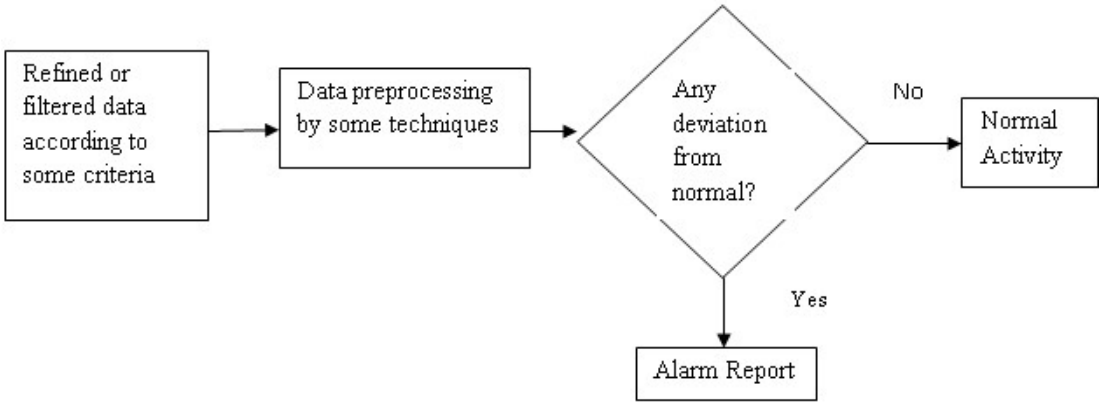


*Figure 1. Steps in intrusion detection*

Intrusion detection can be classified into two types, namely misuse and anomaly. Misuse detection identifies intruders by comparing them with predefined description of events that denote attack. It matches an event which has already occurred with these predefined events, and any case of deviation is categorized as intrusion. However, this method cannot determine new or unknown attacks. On the other hand, anomaly detection builds models of normal events, and any deviations from these normal events are categorized as attacks.

## 2.2 Technology Types

Different technologies are applied to detect intruders, but each has its own limitations as summarized in Table 1.

The description in the Table 1 reveals that a deeper analysis of the intrusion detection system is necessary. Existing threats and their solutions still need in-depth analysis from the point of view of researchers. A computing system with high computational speed, fault tolerance, and ability to deal with dynamic real time data is necessary. Several works reported successful intrusion detection system based on machine learning techniques which fit perfectly for the aspects like higher computation speed (Rathore *et al*., 2016), fault tolerance (Gao *et al*., 2015), and dealing real time big data (Singh *et al*., 2014). Different machine learning techniques bear close resemblance to computational statistics methods such as Markov Chain's Monte Carlo method and Kernel density function (Julisch & Dacier, 2002). Due to these similarities of computational statistics, machine learning has attracted researchers' attention to use them in the field of intrusion detection. On the other hand, neural networks become attractive for easy implementation, faster speed of learning, more generalizability and better dealing with non-linear activation functions and kernels. Due to increasing demand for symbolic representation of problems, symbolic artificial intelligence gains attention of the researchers.

| Techniques used to solve intrusion detection | Subsystem | Advantage | Disadvantage |
|---|---|---|---|
| **Statistical based system (Julisch & Dacier, 2002)** | Univariate | Quick and inexpensive to operate. | Extra attribute handling is difficult. |
| | Multivariate | Robust | Relationship between variables are complex, hence very difficult in handling. |
| | | Helps establish the relationship between numbers of variables and find out the relationship between them. | For giving the effective result large amount of data is required |
| **Knowledge based or expert based system (Mann & Kaur, 2013; Julisch & Dacier, 2002)** | Finite State Machine | It helps in solving the complex system using simple states. | A number of states if large, can be unmanageable. |
| | Time Series model | It forecasts by converting the nonlinear model to linear model. | Not all packet type especially real-time system packets can be converted to linear model, making the task difficult. |
| | Expert System | As databases prepared by experts, therefore, all point of view of attributes of dataset are considered. | Human interaction is needed to create the rule. |
| **Data mining technique (Sisodia et al., 2012)** | Hierarchical | Less cost to join the clusters | Once the cluster is formed joining two or more clusters, decomposing the cluster is difficult.<br>Performance degrades with noise |
| | Partitioning | Suitable for dataset where the relationship between attributes are less. | Performance degrades with noise. |
| | | Simple | Reliability on the user to create clusters |
| | Grid based method | Performance does not degrade with noise. | Clustering is done on the summary of objects and not on an individual object. If any error occurs in an individual object, then the overall result becomes inaccurate. |
| | | Efficient handling of high dimensional data<br>Takes less time | |

*Table 1. Popular intrusion detection techniques and their advantages and disadvantages*

However, the approach of combining various techniques for improving the robustness of intrusion detection is not new (Fossaceca *et al*., 2015). Mukkamala *et al*., (2005) combined artificial intelligence-based methods, such as ensemble of Neural Networks, support vector machine, and Multiple Adaptive Regressive Splines achieved better results than individual method. However, it is challenging to apply the same on a big intrusion dataset. Research (e.g. Sabhnani & Serpen, 2003) also suggest using combination of multiple classification algorithms that work efficiently against each intrusion type in future intrusion detection research

## 2.3 Detection Methodologies

To overcome problems the limitations of various techniques as mentioned in Table 1 in intrusion detection, it is important to identify the exact requirement of intrusion detection in the current environment characterized by rapidly changing data, large volume and variety. From the extensive study of articles mentioned in Table 2, we found out that the researcher has effectively solved one or two perspectives of the problem. We feel that to keep pace with the changing pattern of attack, solving one or two approaches may not be sufficient. Along with a good detection rate and accuracy, the following requirements (presented in the form of objectives – obj1, obj2 …) have to be addressed for a reliable intrusion detection.

**Obj1: Efficient clustering and classification, that is, the machine should not be over-trained and should give unbiased result**

The articles mentioned in Table 2 basically classify and cluster the attacks on a particular system. Henceforth, the techniques used by different authors is based on doing feature selection (which select the most relevant attributes contributing to the attack) and then training the system of the possible vulnerabilities. When a new set of attack is fed to the system, they can identify and classify new attack by comparing with the normal data. Any machine learning technique has a tendency of overfitting, which leads to erroneous results on a new data set. Due to overfitting, the technique fails to segregate a normal data from malicious attack and misleads the user by making proper prediction.

**Obj2: Less human interaction**

Human beings, in the form of system administrator or programmers, are usually involved in intrusion detection for configuring the environmental settings, writing requisite code and analysing the results. However, attacks evolve every fortnight and demand improvement of the human personal in detecting attacks or anomalies. Machine learning techniques make their task easier because of their self-learning capability, if these techniques are trained with certain set of patterns. The similar kind of patterns can be identified by them automatically without requiring human intervention.

**Obj3: Low computational overhead and cost**

The data required to be handled for intrusion detection is huge. Handling this huge amount of data is a herculean task for any researcher. Optimal selection of criteria which can identify the majority type of attacks is a challenging task for a researcher. Tsang (2005) tried to select an optimal feature subset for dimensionality reduction. Aslahi-Shahri *et al* (2016) and Chung *and Wahid* (2012) also tried to reduce the data, so that the execution time is reduced, and space management becomes effective. Hu *et al* (2014) also tried to reduce the communication cost of handling a distributed database. Most of the studies have strived to achieve this through the feature selection mechanism, otherwise results obtained by analysing un-filtered data by compromising space, time and system performance will have no significance later, especially in case of a real time attack.

**Obj4: Identification of the new type of attacks**

The mode of attack is constantly changing as new types of virus and worms are invented every day. The software developed today for handling the attack may not be equipped to handle these new types of attack thus resulting in loss of vital data and consequent huge economic loss. To combat this problem the researcher strives to design a system capable of identifying

new types of attack by training the system with some predefined pattern. Hu *et al*. (2014) tried to identify new types of attack by reducing the communication cost. Stopel *et al*. (2009) proposed a system to detect new types of computer worms which will give good detection rate and accuracy and can be effectively used for identifying newly discovered attacks.

**Obj5: Robustness- Capability of handling large interrelated datasets and real type packets**

Dataset used for intrusion detection is huge with interrelated data, especially when we are dealing with real time data. Different researcher like Bankovic *et al* (2007), Denning and Neumann (1985), and Creech and (2014) has proposed a system for dealing with real time attacks. To classify a real time attack is a challenging task as they are interrelated with respect to time. For example, it is really difficult to accurately structure the behaviour of the system by extracting on an exact point, when two or more consecutive attacks are executed (Chen *et al*., 2017). Moreover, the user-content analysis demands accurate result within a fraction of second. If the system is not robust to handle this situation, it can be catastrophic and can jeopardize the system.

# 3 Exploring current state–of-the-art techniques in intrusion detection

## 3.1 Research Methodology

Research articles related to several key words of the machine learning techniques and applications in intrusion detection were searched for and selected from Science Direct, Taylor & Francis, Springer, Scopus, Google Scholar, and the Emerald online portals. We obtained more than 400 articles from these sources. The sample size considered was larger than that obtained in earlier studies. Out of these 400, we considered the top 60 articles from reputed journals and analysed them. In all these 60 articles, author(s) obtained effective results by using the techniques such as SVM, ANN and fuzzy by one way or another. So we have included only these 60 articles in our paper and try to summarize the result based on them, so that from it we can give a new direction to the researcher. Surveyed articles on intrusion detection are mentioned in Table 6 in the Appendix.

| Sl. No. | Article | Ref. No. | Sl. No. | Article | Ref No. |
|---|---|---|---|---|---|
| **1** | Abraham (2005) | 1 | 35 | Mitrokotsa (2005) | 49 |
| **2** | Aburomman (2016) | 2 | 36 | Njogu (2013) | 52 |
| **3** | Aslahi (2015) | 5 | 37 | Orfila (2009) | 54 |
| **4** | Ashfaq (2016) | 4 | 38 | Ozyer (2007) | 55 |
| **5** | Bankovic (2007) | 6 | 39 | Pinzón (2013) | 57 |
| **6** | Chen (2005) | 8 | 40 | Powers (2007), | 58 |
| **7** | Chung(2012) | 10 | 41 | Peddabachigari (2007) | 56 |
| **8** | Creech (2014) | 11 | 42 | Ramasubramanian (2006) | 59 |
| **9** | Damopoulos (2012) | 12 | 43 | Revathi (2014) | 61 |
| **10** | De (2015) | 13 | 44 | Sangkatsanee (2011) | 63 |
| **11** | Elbasiony (2013) | 15 | 45 | Sani (2015) | 64 |
| **12** | Elhag (2014) | 16 | 46 | Shamshirband (2014) | 66 |
| **13** | Feng (2014) | 17 | 47 | Sheikhan (2014) | 67 |

| | | | | | |
|---|---|---|---|---|---|
| **14** | Fisch (2010) | 18 | 48 | Shon (2007) | 68 |
| **15** | Ganapathy (2012) | 20 | 49 | Sindhu (2012), | 69 |
| **16** | Grediaga (2006) | 22 | 50 | Stopel (2009) | 72 |
| **17** | Han (2006) | 23 | 51 | Subbulakshmi (2010) | 73 |
| **18** | Hoang (2009) | 24 | 52 | Tajbakhsh (2009) | 74 |
| **19** | Horng (2011) | 25 | 53 | Tong (2009) | 75 |
| **20** | Hu (2014 b) | 26 | 54 | Toosi (2007) | 76 |
| **21** | Hu (2008) | 27 | 55 | Tsang (2007) | 77 |
| **22** | Jazzar (2008) | 28 | 56 | Wang (2010) | 78 |
| **23** | Jiang (2009) | 29 | 57 | Yin (2005) | 80 |
| **24** | Jiang (2006) | 30 | 58 | Zanero (2005) | 81 |
| **25** | Khan (2006) | 33 | 59 | Zhang (2007) | 82 |
| **26** | Kim (2014) | 34 | 60 | Zhang (2005) | 83 |
| **27** | Kuang (2014) | 35 | | | |
| **28** | Kumar (2015) | 36 | | | |
| **29** | Kumar (2013) | 38 | | | |
| **30** | Lei (2012) | 40 | | | |
| **31** | Li (2009) | 41 | | | |
| **32** | Liao (2009) | 42 | | | |
| **33** | Luo (2014) | 43 | | | |
| **34** | Mabu (2011) | 44 | | | |

*Table 2: Top works in the field of intrusion detection*

## 4 Data Analysis and Results

The distribution of our studies on intrusion detection in the area of machine learning field from 2002 to 2015 is presented in Figure 2. Figure 2 reveals that the use of machine learning was at its peak in 2007, reduced in 2011, but again gained momentum during 2013-2016.

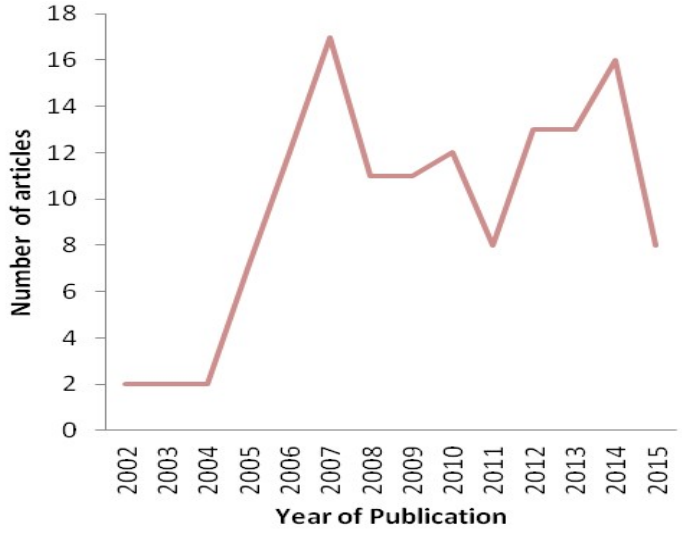


*Figure 2. Showing the distribution of research in the field of intrusion detection*

We attempted to identify the most popular techniques through our survey and summarize the result in the form of a pie chart in Figure 3.

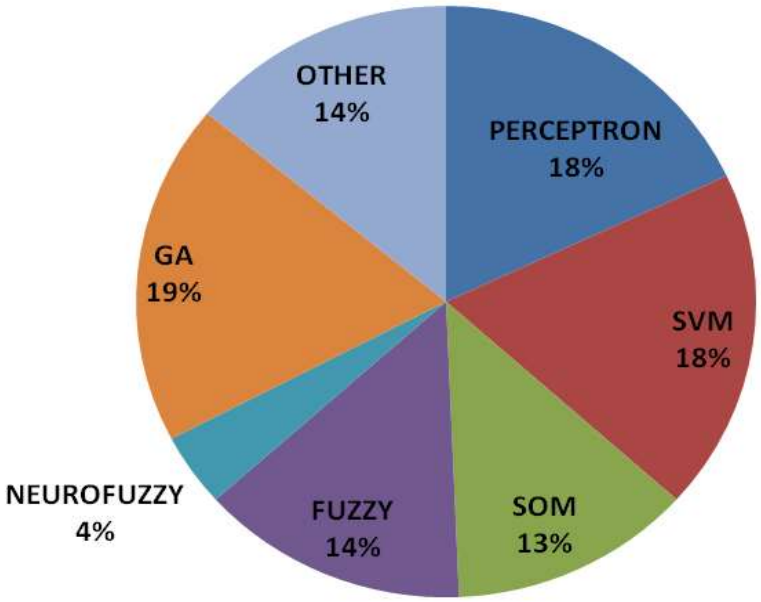


*Figure 3. Popular machine learning techniques in intrusion detection*

Figure 3 reveals that Perceptron (single layer and multi-layer) - 18%, Fuzzy logic - 14%, Genetic Algorithm - 19% and Support Vector Machine - 18% are the most popular techniques for intrusion detection. We now attempt to sub-categorize these techniques with respect to the dataset it was capable of handling effectively. We obtained these data sets from the following sources:

i. data packets from networks

ii. data collected from SNORT or TCPDump, and

iii. system call or CPU usage or memory usage.

These datasets can be summarized as follows:

- **IDS_Bag1**: This dataset comprises a collection of system call sequences collected at Massachusetts Institute of Technology Lincoln Lab for one week.
- **BSM Dataset**: This dataset is collected from the victim's Solaris machine. BSM audit logs contain system calls produced by the program running on Solaris machine. For each day, a 'BSM list file' is prepared and each line of the file indicates a session. The line contains information such as time, service, source IP, and destination IP. A '0' at the end of the line indicates the session is normal and a '1' indicates the session is intrusive.
- **KDD Dataset**: KDD training dataset comprises of single connection vectors (4,900,000). Each of connection vector has 41 features and is categorized as either normal or an attack, with just one specific type of the attack. Following are the four types of simulated attacks: viz. Denial of Service Attack (DoS), User to Root Attack (U2R), Remote to Local Attack (R2L), and Probing Attack.

- **DARPA Intrusion Detection Evaluation**: DARPA Intrusion Detection Evaluation has two parts, one operates in offline evaluation mode, and another operates in real-time evaluation mode. The offline mode normally operates in the batch mode and uses network traffic and audit logs collected on the simulated network. The system tries to identify attacks in the middle of normal activities. In the online mode, which is the real time mode, these systems are inserted into the AFRL network test bed and are used to identify attack sessions in the middle of normal activities, in real time.
- **NSL-KDD**: is a dataset which has solved many problems found in KDD dataset, though it is not suitable for real time networks because of lack of public data sets for network-based IDSs. It does not have redundant or duplicate records. The number of training and testing data is reasonable.

Effectiveness of a technique on a particular dataset can only be judged by analysing the correctness or accuracy of the result obtained. Several performance metrics or measures are taken into consideration for this purpose, namely, sensitivity or true positive rate (proportion of correctly classified positive cases), detection rate (ratio of true positive data to the total number of intrusion data present in the system), accuracy (total number of correct predictions) and ROC Curve (graph used to measure the performance of binary classifier).

From the 60 articles surveyed, we have found that GA covers 19%, SVM covers 18%, Fuzzy + neuro Fuzzy covers 18%, SOM covers 18% and Perceptron covers 18% of the total articles (Figure 3). We did not include perceptron because its effectiveness in handling real time data, as well as in other types of readily available data, is not promising as in case of other techniques. The performance of numerous techniques using these parameters on the above-mentioned dataset is summarized in Figures 4a-4d.

**Support Vector Machines (SVM)**

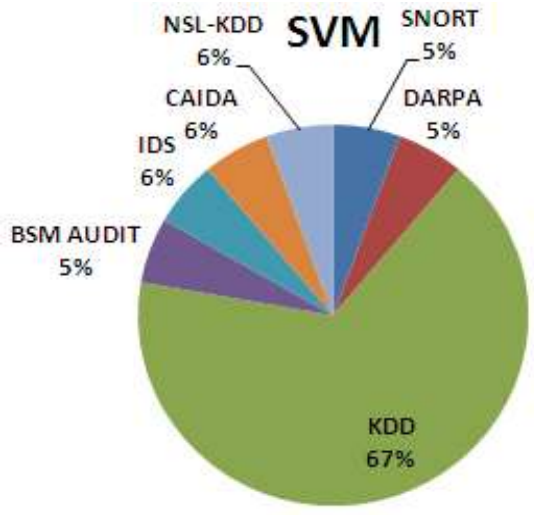


*Figure 4a. Percentage of different data set used by support vector machines (svm)*

**Self Organizing Maps (SOM)**

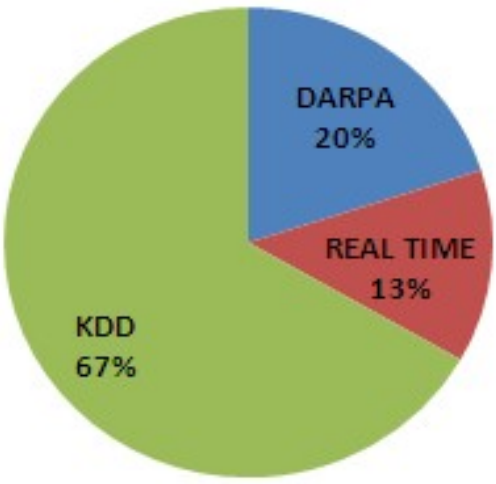


*Figure 4b. Percentage of different data set used by self-organizing maps (som)*

**Genetic Algorithms (GA)**

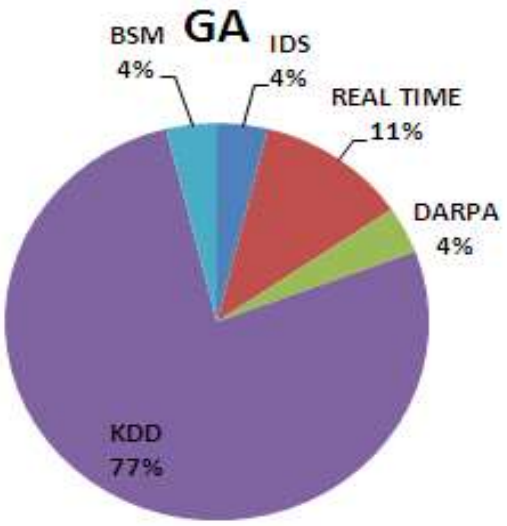


*Figure 4c. Percentage of different data set used by genetic algorithms (ga)*

**Fuzzy Logic**

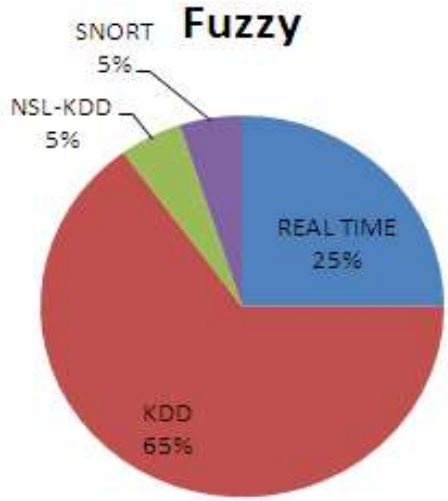


*Figure 4d. Percentage of different data set used by fuzzy logic*

We can see from Figure 4 that Fuzzy, Self-organizing maps, and genetic algorithm are used extensively in detecting real time attacks, whereas SVM is used extensively with publicly available dataset (SNORT, CAIDA, etc.). However, among all datasets, most of the experiments were carried out on KDD dataset. To further justify our claim, we take the help of the graphs in Figure 5.

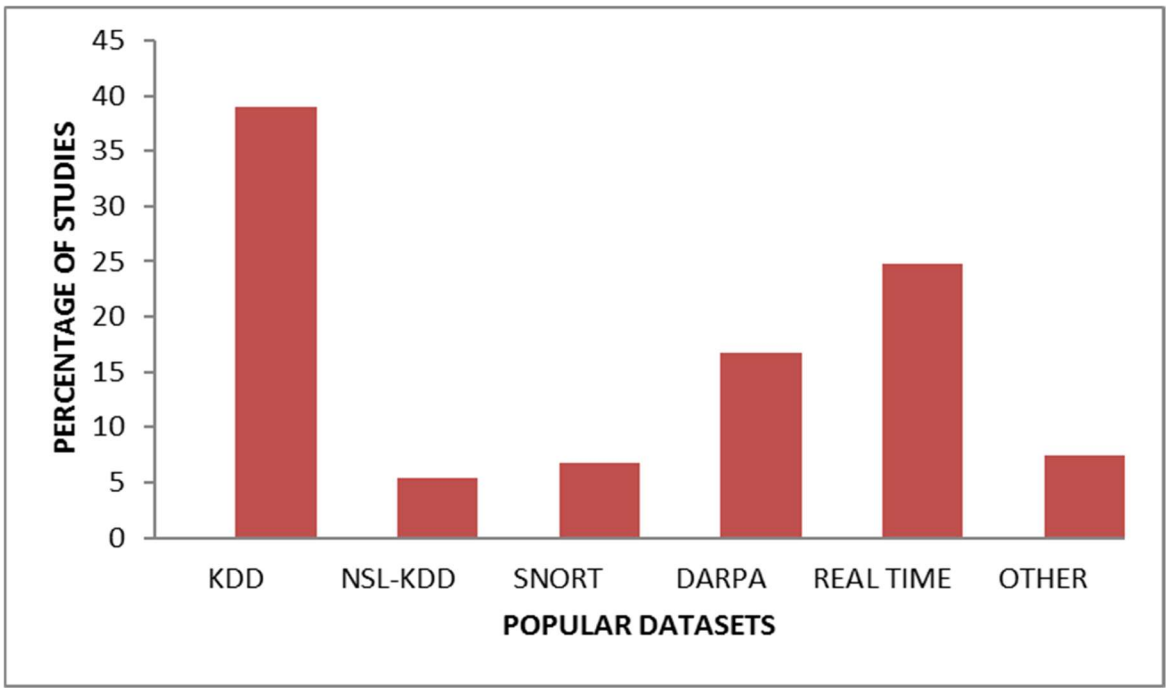


*Figure 5. Showing the popular datasets used with various techniques*

Figure 5 reveals that these techniques have used KDD dataset in nearly 40% of their intrusion detection analysis, while they have worked with real time data set for only 25% of their analysis.

After understanding the popular datasets, we now try to understand the most popular performance metrics used by the above-mentioned techniques, and we obtain Figure 6 which provides the evaluation of the different performance measures of intrusion detection.

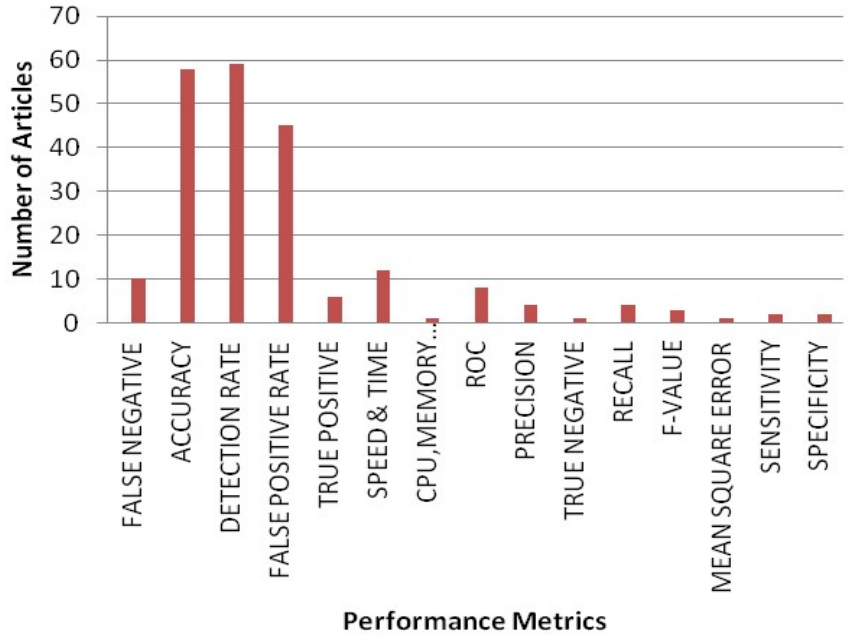


*Figure 6. Showing the different performance metrics used by our popular techniques*

From Figure 6, we can see that accuracy and detection rate and false positive rate are the most popular performance metrics used with machine learning techniques.

## 5 Discussion

The machine learning techniques and soft computing techniques are special techniques for handling huge data sets. Fuzzy logic techniques effectively cluster overlapping datasets. They are also used extensively in removing attributes which are a misfit for a particular cluster. The fuzzy rule-based system tries to match a pattern with a set of predefined patterns and thereby helps in removing misfit data of the cluster. Genetic Algorithm (GA) is also capable of handling unrelated data in clusters for using the process of selection, crossover and mutation. SOM (Self Organizing Map) reveals the most important relationships between the features and hides unwanted details. Because of abstraction, these techniques are expert in handling complex interrelated data and the suppression of unnecessary details helps in saving time and unwanted processing. SVM (Support Vector Machine) converts the highly complex data, especially the text, into a form suitable for classification. Neural network is effective by being trained on a certain pattern of data. It reports deviation from this pattern, if any, in a new dataset and thus helps in identification of dissimilarity formed by each cluster. Therefore, this technique in stand-alone mode or with other techniques can solve the security threats followed worldwide. We summarize our findings in Table 3.

| Problems in existing Technologies | Solved effectively with |
|---|---|
| Dissimilarity of the discovered pattern with existing pattern | Fuzzy clustering, BPNN |
| Dependencies among data and understanding dependency clustering | Fuzzy logic by the progressive reduction of cognitive dessonance, SOM, GA |
| Web personalization | GA, Fuzzy rule based system |
| Data summarization | Fuzzy set theory, SOM, BPNN, |
| Creation of Association rules | Fuzzy acyclic directed graph, Fuzzy rule based system |
| Difficulty in processing documents containing images | Fuzzy logic, SOM |
| Regression | Neural network, neuro fuzzy computation |
| Extra attribute handling | K-nearest neighbour, GA |
| Complexity of relationship owing to a large number of variables | SVM, Decision Tree |
| Time Consuming | SOM |
| Difficulty in managing huge dataset | Genetic algorithm, SOM, SVM |
| Reliability on the user to change cluster | Genetic algorithm |

*Table 3: Machine learning techniques and their solution as identified in Table 1*

| | Obj1 | Obj2 | Obj3 | Obj4 | Obj5 |
|---|---|---|---|---|---|
| Abrahim (2007) | √ | - | √ | √ | √ |
| Aburomman (2016) | √ | √ | - | √ | - |
| Ashfaq (2016) | √ | √ | - | √ | √ |
| Aslahi (2015) | √ | √ | √ | √ | - |
| Bankovic (2007) | √ | √ | √ | √ | - |
| Chen (2005) | √ | √ | - | √ | - |
| Chung (2012) | √ | √ | - | √ | √ |
| Creech (2014) | √ | √ | - | √ | √ |
| Damopoulos (2012) | √ | √ | - | √ | √ |
| De (2015) | √ | √ | - | √ | √ |
| Elbasiony (2013) | √ | √ | - | √ | √ |
| Elhag (2014) | √ | √ | - | √ | √ |
| Feng (2014) | √ | √ | √ | √ | √ |
| Fisch (2010) | √ | √ | - | √ | √ |
| Ganapathy (2012) | √ | √ | √ | √ | √ |
| Grediaga (2006) | √ | √ | - | √ | √ |
| Han (2005) | √ | √ | √ | √ | - |
| Hoang (2009) | √ | √ | √ | √ | - |
| Horng (2011) | √ | √ | √ | √ | - |
| Hu (2008) | √ | √ | √ | √ | - |

| | **Obj1** | **Obj2** | **Obj3** | **Obj4** | **Obj5** |
|---|---|---|---|---|---|
| Hu (2014 ) | √ | √ | - | √ | √ |
| Jazzar (2008) | √ | √ | - | √ | √ |
| Jiang (2009) | √ | √ | - | √ | - |
| Jiang (2006) | √ | √ | √ | √ | - |
| Khan (2007) | √ | √ | - | √ | - |
| Kim (2014) | √ | √ | √ | √ | - |
| Kuang (2014) | √ | √ | √ | √ | - |
| Kumar (2015) | √ | √ | - | √ | √ |
| Kumar (2013) | √ | √ | - | √ | √ |
| Lei (2012) | √ | √ | - | √ | √ |
| Li (2009) | √ | √ | √ | √ | - |
| Liao (2009) | √ | √ | - | √ | √ |
| Luo (2014) | √ | √ | - | √ | - |
| Mabu (2011) | √ | √ | - | √ | - |
| Mitrokotsa (2005) | √ | √ | - | √ | - |
| Njogu (2013) | √ | √ | √ | √ | - |
| Orfila (2009) | √ | √ | - | √ | - |
| Özyer (2007) | √ | √ | - | √ | - |
| Pinzón (2013) | √ | √ | - | √ | √ |
| Powers (2008) | √ | √ | - | √ | - |
| Ramasubramanian (2006) | √ | √ | √ | √ | √ |
| Revathi (2014) | √ | √ | - | √ | - |
| Sangkatsanee (2011) | √ | √ | - | √ | √ |
| Sani (2015) | √ | √ | - | √ | - |
| Shamshirband (2014) | √ | √ | - | √ | √ |
| Sheikhan (2014) | √ | √ | - | √ | √ |
| Shon (2007) | √ | √ | - | √ | √ |
| Sindhu (2012) | √ | √ | - | √ | - |
| Stopel (2009) | √ | √ | - | √ | - |
| Subbulakshmi (2010) | √ | √ | - | √ | √ |
| Tajbakhsh (2009) | √ | √ | √ | √ | - |
| Tong (2009) | √ | √ | - | √ | - |
| Toosi (2007) | √ | √ | - | √ | - |
| Tsang (2007) | √ | √ | - | √ | - |
| Wang (2010) | √ | √ | - | √ | - |
| Peddabachigari (2007) | √ | √ | - | √ | - |
| Yin (2005) | √ | √ | - | - | - |
| Zanero (2005) | √ | √ | - | - | - |
| Zhang (2005) | √ | √ | - | - | - |
| Zhang (2007) | √ | √ | √ | √ | - |

*Table 4: Comparison of different work according to our taxonomy*

- *Obj1: Efficient clustering and classification that is the machine should not be over trained and give bias result.*
- *Obj2: Less human interaction*
- *Obj3: Low computational and cost overhead*
- *Obj4: Identification of new type of attack*
- *Obj5: Robustness- Capability of handling large interrelated datasets and real type packets*

In Table 4 we map our objectives (Obj1 to Obj5) with the existing articles and we conclude that though most of the articles fulfil requirements of Obj1, Obj2 and Obj4, they are unable to meet the requirements of Obj3 and Obj5. As already pointed out that real time data are interrelated and demands special mechanism for segregation and this calls for special attention. We can observe that machine learning techniques with their self-learning or supervisory mode are able to detect most of the attacks and have provided very good detection and accuracy rate. However, these metrics alone do not consider the hostility of the environmental condition. For this reason, certain other metrics such as cost and sensitivity need to be taken into consideration. If the environmental condition is not taken into consideration, detection of new types of attack will be a difficult task. Moreover, reliability on KDD dataset is also not a good solution to judge the efficacy of the system regarding detection of attacks. The KDD dataset contains redundant records and has laid less stress on U2R and R2L attacks. In most of the training experiments, these attacks are likely to be missed, as the number of rows is less as compared to other types of attacks. Moreover, 75% to 78% of the records are duplicate. Based on the training of KDD dataset, the new dynamic real time attack may not be handled by the system. There is still requirement for a unified architecture or technique, which will provide a platform to identify real time attack, and also a standardized solution in handling wired and wireless attack, as shown in Table 5. It is evident from Table 5 that different environmental conditions favour different techniques.

| Area of observation | Detail observation | Concluding remark |
|---|---|---|
| Percentage of total analyzed literature | SOM and SVM covers 13% and 18% respectively. Fuzzy and GA covers 14% and 19% respectively while Perceptron covers 17% | According to our survey GA,Perceptron and SVM are most popular tool. |
| Most Common Approach | Fuzzy rule based system with GA covers nearly 11% of analyzed literature | Genetic Algorithm used on knowledge base dataset containing fuzzy rule are popular techniques used for feature selection. |
| Most common performance metrics | Accuracy and detection rate. Detection rate covers 49% of analyzed data, Accuracy covers 28% of analyzed data | Detection is given more importance than analysis during performance evaluation |

*Table 5: Observation based on machine learning techniques approach in intrusion detection*

Current trends in network involves distributed computing with increasing demand for cloud computing (i.e. more involvement of internet) (Kumar *et al*., 2010). In addition, ad-hoc and sensor networks have indicated possibility of new type of attacks. The performance of most of the above mentioned techniques in dealing with intrusion in cloud platform or determining black hole attack in sensor networks is questionable. Intrusion detection fails to determine the

attacks at different levels of architecture of a cloud. Meanwhile as stated by Modi *et al* (2013), internal attacks are also increasing. There is a lack of suitable mechanism to handle them.

Sensor networks on the other hand are more sensitive to attack. Mobile nodes with poor inbuilt security mechanism are easy to capture via wired networks. An attacker can listen to traffic, modify the traffic or can act as one of the legitimate users. As there is no such central architecture which can help in intrusion detection, proper cryptography via public or private key is difficult to implement in mobile adhoc or sensor network. Our article has identified the impact of machine learning on intruded packets and at the same time has identified the issue of security concern that are left to be handled when dealing with real time data, sensitive data in mobile phones and sensor networks.

# 6 Conclusion

This study provides an insight into the progress of research on intrusion detection based on machine learning techniques. It has discussed the most popular machine learning techniques, and their advantages and disadvantages. As machine learning techniques are extensively used with soft computing techniques we have been analysed them separately. This review graphically provides a clear indication of the overall picture of research in the field of intrusion detection with changing usage pattern, in terms of popular performance metrics and most widely used techniques and datasets. Most of the techniques perform well with KDD. Fuzzy logic techniques perform well with real time datasets. The study also revealed that machine learnings approaches applied to intrusion detection are quite successful except in the matter of fulfilling the objectives of low computational and cost overheads and robustness (capability of handling large interrelated datasets and real type packets). Therefore, the future research may be directed towards those machine learning tools that will achieve both of these objectives (i.e. low computational cost and robustness). More obvious gap is the labelled data application like KDD dataset on which majority of the techniques are applied and it would be more worthy if intrusion data is collected and labelled partially. As the machine learning techniques require training and testing data so they can be trained using partial labelled dataset for the known attack and tested on unknown data for measuring performance. Therefore, the promising techniques may be further tested on these new data set for developing effective and efficient intrusion detection system for breakthrough performance.

Moreover since majority of the results are based on KDD dataset, approximation of the actual performance of the intrusion detection system on real time data is difficult to evaluate. The effectiveness of these techniques on real time data and effective performance metrics used for their evaluation is an open area for the future researchers.

# References

Aburomman, A. A., & Ibne Reaz, M. Bin. (2016). A novel SVM-kNN-PSO ensemble method for intrusion detection system. *Applied Soft Computing Journal*, 38, 360–372. https://doi.org/10.1016/j.asoc.2015.10.011

Anderson, D., Frivold, T., & Valdes, A. (1995). Next-generation Intrusion Detection Expert System (NIDES): A summary. *SRI International*, (May 1995), 47. https://doi.org/citeulike-article-id:7898221

Arun Raj Kumar, P., & Selvakumar, S. (2013). Detection of distributed denial of service attacks using an ensemble of adaptive and hybrid neuro-fuzzy systems. *Computer Communications*, 36(3), 303–319. https://doi.org/10.1016/j.comcom.2012.09.010

Ashfaq, R. A. R., Wang, X. Z., Huang, J. Z., Abbas, H., & He, Y. L. (2017). Fuzziness based semi-supervised learning approach for intrusion detection system. *Information Sciences*, 378, 484–497. https://doi.org/10.1016/j.ins.2016.04.019

Aslahi-Shahri, B. M., Rahmani, R., Chizari, M., Maralani, A., Eslami, M., Golkar, M. J., & Ebrahimi, A. (2016). A hybrid method consisting of GA and SVM for intrusion detection system. *Neural Computing and Applications*, 27(6), 1669–1676. https://doi.org/10.1007/s00521-015-1964-2

Banković, Z., Stepanović, D., Bojanić, S., & Nieto-Taladriz, O. (2007). Improving network security using genetic algorithm approach. *Computers and Electrical Engineering*, 33(5–6), 438–451. https://doi.org/10.1016/j.compeleceng.2007.05.010

Chen, C., Ghassami, A., Mohan, S., N. K. (2017), *A Reconnaissance Attack Mechanism for Fixed-Priority Real-Time Systems.* Arxiv.Org.

Chung, Y. Y., & Wahid, N. (2012). A hybrid network intrusion detection system using simplified swarm optimization (SSO). *Applied Soft Computing*, 12(9), 3014–3022. https://doi.org/10.1016/j.asoc.2012.04.020

Creech, G., & Hu, J. (2014). A semantic approach to host-based intrusion detection systems using contiguous and discontiguous system call patterns. *IEEE Transactions on Computers*, 63(4), 807–819. https://doi.org/10.1109/TC.2013.13

Damopoulos, D., Menesidou, S. A., Kambourakis, G., Papadaki, M., Clarke, N., & Gritzalis, S. (2012). Evaluation of anomaly-based IDS for mobile devices using machine learning classifiers. *Security and Communication Networks*, 5(1), 3–14. https://doi.org/10.1002/sec.341

De la Hoz, E., De La Hoz, E., Ortiz, A., Ortega, J., & Prieto, B. (2015). PCA filtering and probabilistic SOM for network intrusion detection. *Neurocomputing*, 164, 71–81. https://doi.org/10.1016/j.neucom.2014.09.083

Denning, D. & Neumann, P. (1985). Requirements and model for IDES—a real-time intrusion detection expert system. Document A005, *SRI International*, California. 1-74. Retrieved from faculty.nps.edu/dedennin/publications/IDESReportSRI1985.pdf (accessed on 31-05-2018)

Elbasiony, R. M., Sallam, E. A., Eltobely, T. E., & Fahmy, M. M. (2013). A hybrid network intrusion detection framework based on random forests and weighted k-means. *Ain Shams Engineering Journal*, 4(4), 753–762. https://doi.org/10.1016/j.asej.2013.01.003

Elhag, S., Fernández, A., Bawakid, A., Alshomrani, S., & Herrera, F. (2015). On the combination of genetic fuzzy systems and pairwise learning for improving detection rates on Intrusion Detection Systems. *Expert Systems with Applications*, 42(1), 193–202. https://doi.org/10.1016/j.eswa.2014.08.002

Feng, W., Zhang, Q., Hu, G., & Huang, J. X. (2014). Mining network data for intrusion detection through combining SVMs with ant colony networks. *Future Generation Computer Systems*, 37, 127–140. https://doi.org/10.1016/j.future.2013.06.027

Fisch, D., Hofmann, A., & Sick, B. (2010). On the versatility of radial basis function neural networks: A case study in the field of intrusion detection. *Information Sciences*, 180(12), 2421–2439. https://doi.org/10.1016/j.ins.2010.02.023

Fossaceca, J. M., Mazzuchi, T. A., & Sarkani, S. (2015). MARK-ELM: Application of a novel Multiple Kernel Learning framework for improving the robustness of Network Intrusion Detection. *Expert Systems with Applications*, 42(8), 4062-4080.

Ganapathy, S., Kulothungan, K., Yogesh, P., & Kannan, A. (2012). A novel weighted fuzzy C - means clustering based on immune genetic algorithm for intrusion detection. In *Procedia Engineering* (Vol. 38, pp. 1750–1757). https://doi.org/10.1016/j.proeng.2012.06.213

Gao, Z., Member, S., Cecati, C., Ieee, F., & Ding, S. X. (2015). A Survey of Fault Diagnosis and Fault - Tolerant Techniques Part II : Fault Diagnosis with Knowledge - Based and Hybrid / Active Approaches. *IEEE Transactions on Industrial Electronics*, 62(6), 3768–3774. https://doi.org/10.1109/TIE.2015.2419013

Grediaga, Á., Ibarra, F., García, F., Ledesma, B., & Brotóns, F. (2006). Application of Neural Networks in Network Control and Information Security. In *Advances in Neural Networks* ISNN 2006 (pp. 208–213). https://doi.org/10.1007/11760191_31

Han, S.-J., & Cho, S.-B. (2006). Evolutionary neural networks for anomaly detection based on the behavior of a program. *IEEE Transactions on Systems, Man, and Cybernetics. Part B, Cybernetics : A Publication of the IEEE Systems, Man, and Cybernetics Society*, 36(3), 559–570. https://doi.org/10.1109/TSMCB.2005.860136

Hoang, X. D., Hu, J., & Bertok, P. (2009). A program-based anomaly intrusion detection scheme using multiple detection engines and fuzzy inference. *Journal of Network and Computer Applications*, 32(6), 1219–1228. https://doi.org/10.1016/j.jnca.2009.05.004

Horng, S. J., Su, M. Y., Chen, Y. H., Kao, T. W., Chen, R. J., Lai, J. L., & Perkasa, C. D. (2011). A novel intrusion detection system based on hierarchical clustering and support vector machines. *Expert Systems with Applications*, 38(1), 306–313. https://doi.org/10.1016/j.eswa.2010.06.066

Hu, W., Gao, J., Wang, Y., Wu, O., & Maybank, S. (2014). Online ada boost-based parameterized methods for dynamic distributed network intrusion detection. *IEEE Transactions on Cybernetics*, 44(1), 66–82. https://doi.org/10.1109/TCYB.2013.2247592

Hu, W., Hu, W., & Maybank, S. (2008). AdaBoost-based algorithm for network intrusion detection. *IEEE Transactions on Systems, Man, and Cybernetics*, 38(2), 577–583. https://doi.org/10.1109/TSMCB.2007.914695

Jazzar, M., & Jantan, A. (2008). An approach for anomaly intrusion detection based on causal knowledge-driven diagnosis and direction. *Studies in Computational Intelligence*, 149, 39–48. https://doi.org/10.1007/978-3-540-70560-4_4

Jiang, H., & Ruan, J. (2009). The application of genetic neural network in network intrusion detection. *Journal of Computers*, 4(12), 1223–1230. https://doi.org/10.4304/jcp.4.12.1223-1230

Jiang, S., Song, X., Wang, H., Han, J. J., & Li, Q. H. (2006). A clustering-based method for unsupervised intrusion detections. *Pattern Recognition Letters*, 27(7), 802–810. https://doi.org/10.1016/j.patrec.2005.11.007

Julisch, K., & Dacier, M. (2002). Mining intrusion detection alarms for actionable knowledge. In *Proceedings of the eighth ACM SIGKDD international conference on Knowledge discovery and data mining - KDD '02* (p. 366). https://doi.org/10.1145/775047.775101

Kabiri, P., & Ghorbani, A. A. (2005). Research on intrusion detection and response: A survey. *International Journal of Network Security*, 1(2), 84–102. https://doi.org/10.1.1.129.698

Khan, L., Awad, M., & Thuraisingham, B. (2007). A new intrusion detection system using support vector machines and hierarchical clustering. *The VLDB Journal*, 16(4), 507–521. https://doi.org/10.1007/s00778-006-0002-5

Kim, G., Lee, S., & Kim, S. (2014). A novel hybrid intrusion detection method integrating anomaly detection with misuse detection. *Expert Systems with Applications*, 41(4 PART 2), 1690–1700. https://doi.org/10.1016/j.eswa.2013.08.066

Kuang, F., Xu, W., & Zhang, S. (2014). A novel hybrid KPCA and SVM with GA model for intrusion detection. *Applied Soft Computing Journal*, 18, 178–184. https://doi.org/10.1016/j.asoc.2014.01.028

Kumar, G., & Kumar, K. (2015). A Multi-objective Genetic Algorithm Based Approach for Effective Intrusion Detection Using Neural Networks. In *Intelligent Methods for Cyber Warfare* (pp. 173–200). https://doi.org/10.1007/978-3-319-08624-8_8

Kumar, G., Kumar, K., & Sachdeva, M. (2010). The use of artificial intelligence based techniques for intrusion detection: A review. *Artificial Intelligence Review*, 34(4), 369–387. https://doi.org/10.1007/s10462-010-9179-5

Lee, W., Stolfo, S., & Mok, K. (2000). Adaptive Intrusion Detection: A Data Mining Approach. *Artificial Intelligence Review* (Vol. 14). https://doi.org/10.1023/A:1006624031083

Lei, J. Z., & Ghorbani, A. A. (2012). Improved competitive learning neural networks for network intrusion and fraud detection. *Neurocomputing*, 75, 135–145. https://doi.org/10.1016/j.neucom.2011.02.021

Li, Y., Wang, J. L., Tian, Z. H., Lu, T. B., & Young, C. (2009). Building lightweight intrusion detection system using wrapper-based feature selection mechanisms. *Computers and Security*, 28(6), 466–475. https://doi.org/10.1016/j.cose.2009.01.001

Liao, N., Tian, S., & Wang, T. (2009). Network forensics based on fuzzy logic and expert system. *Computer Communications*, 32(17), 1881–1892. https://doi.org/10.1016/j.comcom.2009.07.013

Luo, B., & Xia, J. (2014). A novel intrusion detection system based on feature generation with visualization strategy. *Expert Systems with Applications*, 41(9), 4139–4147. https://doi.org/10.1016/j.eswa.2013.12.048

Mabu, S., Chen, C., Lu, N., Shimada, K., & Hirasawa, K. (2011). An Intrusion-Detection Model Based on Fuzzy Class-Association-Rule Mining Using Genetic Network Programming. *IEEE Transactions on Systems, Man, and Cybernetics, Part C: Applications and Reviews*, 41(1), 130–139. https://doi.org/10.1109/TSMCC.2010.2050685

McCartney, J. (2014). *The top 10 data breaches of the past 12 months*, http://www.in.techradar.com/news/software/security-software/The-top-

10-data-breaches-of-the-past-12-months/articleshow/38472866.(Accessed on 12-05-2016).

McGregor, J. (2014). *The Top 5 Most Brutal Cyber Attacks Of 2014 So Far*, http://www.forbes.com/sites/jaymcgregor/2014/07/28/the-top-5-most-brutal-cyber-attacks-of-2014-so-far/#5722d80721a6.(Accessed on 12-05-2016).

Meyer, C., (2017). *Data Theft*, http://www.infosecwriters.com/Papers/CMeyer_DataTheft.pdf (accessed on 27-03-2018)

Mitrokotsa, A., & Douligeris, C. (2005). Detecting denial of service attacks using emergent self-organizing maps. *Proceedings of the Fifth IEEE International Symposium on Signal Processing and Information Technology*, 2005., 375–380. https://doi.org/10.1109/ISSPIT.2005.1577126

Modi, C., Patel, D., Borisaniya, B., Patel, H., Patel, A., & Rajarajan, M. (2013). A survey of intrusion detection techniques in Cloud. *Journal of Network and Computer Applications*. https://doi.org/10.1016/j.jnca.2012.05.003

Mukkamala, S., Sung, A. H., & Abraham, A. (2005). Intrusion detection using an ensemble of intelligent paradigms. *Journal of Network and Computer Applications*, 28(2), 167–182. https://doi.org/10.1016/j.jnca.2004.01.003

Njogu, H. W., Jiawei, L., & Kiere, J. N. (2013). Network specific vulnerability based alert reduction approach. *Security and Communication Networks*, 6(1), 15–27. https://doi.org/10.1002/sec.520

Orfila, A., Estevez-Tapiador, J. M., & Ribagorda, A. (2009). Evolving high-speed, easy-to-understand network intrusion detection rules with genetic programming. *In Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* (Vol. 5484 LNCS, pp. 93–98). https://doi.org/10.1007/978-3-642-01129-0_11

Özyer, T., Alhajj, R., & Barker, K. (2007). Intrusion detection by integrating boosting genetic fuzzy classifier and data mining criteria for rule pre-screening. *Journal of Network and Computer Applications*, 30(1), 99–113. https://doi.org/10.1016/j.jnca.2005.06.002

Peddabachigari, S., Abraham, A., Grosan, C., & Thomas, J. (2007). Modeling intrusion detection system using hybrid intelligent systems. *Journal of Network and Computer Applications*, 30(1), 114–132. https://doi.org/10.1016/j.jnca.2005.06.003

Pinzón, C. I., De Paz, J. F., Herrero, Á., Corchado, E., Bajo, J., & Corchado, J. M. (2013). IdMAS-SQL: Intrusion Detection Based on MAS to Detect and Block SQL injection through data mining. *Information Sciences*, 231, 15–31. https://doi.org/10.1016/j.ins.2011.06.020

Powers, S. T., & He, J. (2008). A hybrid artificial immune system and Self Organising Map for network intrusion detection. *Information Sciences*, 178(15), 3024–3042. https://doi.org/10.1016/j.ins.2007.11.028

Ramasubramanian, P., & Kannan, A. (2006). A genetic-algorithm based neural network short-term forecasting framework for database intrusion prediction system. *Soft Computing*, 10(8), 699–714. https://doi.org/10.1007/s00500-005-0513-9

Rathore, M. M., Ahmad, A., & Paul, A. (2016). Real time intrusion detection system for ultra-high-speed big data environments. *Journal of Supercomputing*, 72(9), 3489–3510. https://doi.org/10.1007/s11227-015-1615-5

Revathi, S., & Malathi, A. (2014). Multi-tier framework using sugeno fuzzy inference system with swarm intelligence techniques for intrusion detection. *Indian Journal of Science and Technology*, 7(9), 1437–1443.

Sabhnani, M., Serpen, G., & More, K. K. (2003). Application of Machine Learning Algorithms to KDD Intrusion Detection Dataset within Misuse Detection Context. *Proceedings of International Conference on Machine Learning: Models, Technologies, and Applications (MLMTA),* 209–215. https://doi.org/citeulike-article-id:9827151

Sangkatsanee, P., Wattanapongsakorn, N., & Charnsripinyo, C. (2011). Practical real-time intrusion detection using machine learning approaches. *Computer Communications*, 34(18), 2227–2235. https://doi.org/10.1016/j.comcom.2011.07.001

Sani, R. A., & Ghasemi, A. (2015). Learning a new distance metric to improve an SVM-clustering based intrusion detection system. In *Proceedings of the International Symposium on Artificial Intelligence and Signal Processing, AISP 2015* (pp. 284–289). https://doi.org/10.1109/AISP.2015.7123497

Seals, T. (2015). DDoS attacks more than double in 12 months. *Infosecurity*. Retrieved from http://www.infosecurity-magazine.com/news/ddos-attacks-more-than-double-in/ (accessed on 28-03-2018)

Shamshirband, S., Patel, A., Anuar, N. B., Kiah, M. L. M., & Abraham, A. (2014). Cooperative game theoretic approach using fuzzy Q-learning for detecting and preventing intrusions in wireless sensor networks. Engineering *Applications of Artificial Intelligence*, 32, 228–241. https://doi.org/10.1016/j.engappai.2014.02.001

Sheikhan, M., & Jadidi, Z. (2014). Flow-based anomaly detection in high-speed links using modified GSA-optimized neural network. *Neural Computing and Applications*, 24(3–4), 599–611. https://doi.org/10.1007/s00521-012-1263-0

Shon, T., & Moon, J. (2007). A hybrid machine learning approach to network anomaly detection. *Information Sciences*, 177(18), 3799–3821. https://doi.org/10.1016/j.ins.2007.03.025

Singh, K., Guntuku, S. C., Thakur, A., & Hota, C. (2014). Big Data Analytics framework for Peer-to-Peer Botnet detection using Random Forests. *Information Sciences*, 278, 488–497. https://doi.org/10.1016/j.ins.2014.03.066

Sisodia, D., Singh, L., Sisodia, S., & Saxena, K. (2012). Clustering Techniques: A Brief Survey of Different Clustering Algorithms. *International Journal of Latest Trends in Engineering and Technology (IJLTET),* 1(3), 82–87.

Sivatha Sindhu, S. S., Geetha, S., & Kannan, A. (2012). Decision tree based light weight intrusion detection using a wrapper approach. *Expert Systems with Applications*, 39(1), 129–141. https://doi.org/10.1016/j.eswa.2011.06.013

Stopel, D., Moskovitch, R., Boger, Z., Shahar, Y., & Elovici, Y. (2009). Using artificial neural networks to detect unknown computer worms. *Neural Computing and Applications*, 18(7), 663–674. https://doi.org/10.1007/s00521-009-0238-2

Subbulakshmi, T., Mercy Shalinie, S., Suneel Reddy, C., & Ramamoorthi, A. (2010). Detection and classification of DDoS attacks using fuzzy inference system. In *Communications in Computer and Information Science* (Vol. 89 CCIS, pp. 242–252). https://doi.org/10.1007/978-3-642-14478-3_25

Tajbakhsh, A., Rahmati, M., & Mirzaei, A. (2009). Intrusion detection using fuzzy association rules. *Applied Soft Computing*, 9(2), 462–469. https://doi.org/10.1016/j.asoc.2008.06.001

Toosi, A. N., & Kahani, M. (2007). A new approach to intrusion detection based on an evolutionary soft computing model using neuro-fuzzy classifiers. *Computer Communications*, 30(10), 2201–2212. https://doi.org/10.1016/j.comcom.2007.05.002

Tsang, C. H., Kwong, S., & Wang, H. (2007). Genetic-fuzzy rule mining approach and evaluation of feature selection techniques for anomaly intrusion detection. Pat*tern Recognition*, 40(9), 2373–2391. https://doi.org/10.1016/j.patcog.2006.12.009

Wang, G., Hao, J., Ma, J., & Huang, L. (2010). A new approach to intrusion detection using Artificial Neural Networks and fuzzy clustering. *Expert Systems with Applications*, 37(9), 6225–6232. https://doi.org/10.1016/j.eswa.2010.02.102

Wu, S. X., & Banzhaf, W. (2010). The use of computational intelligence in intrusion detection systems: A review. *Applied Soft Computing Journal*, 10, 1–35. https://doi.org/10.1016/j.asoc.2009.06.019

Yin, C., Tian, S., Huang, H., & He, J. (2005). Applying Genetic Programming to Evolve Learned Rules for Network Anomaly Detection. *Advances in Natural Computation, First International Conference, ICNC 2005, Proceedings, Part III*, 3612, 323–331. https://doi.org/doi:10.1007/11539902_38

Zanero, S. (2005). Analyzing TCP traffic patterns using Self Organizing Maps. *Image Analysis and Processing - Iciap 2005, Proceedings*, 3617, 83–90. https://doi.org/10.1007/11553595_10

Zhang, C., Jiang, J., & Kamel, M. (2005). Intrusion detection using hierarchical neural networks. *Pattern Recognition Letters*, 26(6), 779–791. https://doi.org/10.1016/j.patrec.2004.09.045

Zhang, J.-R., Zhang, J., Lok, T.-M., & Lyu, M. R. (2007). A hybrid particle swarm optimization–back-propagation algorithm for feedforward neural network training. *Applied Mathematics and Computation*, 185(2), 1026–1037. https://doi.org/10.1016/j.amc.2006.07.025

# Appendix

| Reference. No. | Reference details (selected sixty articles studied for meta-analysis in the research work) |
|---|---|
| **1** | Abraham A. (2005). Evolutionary Computation in Intelligent Network Management. In: Ghosh A., Jain L.C. (eds) *Evolutionary Computation in Data Mining. Studies in Fuzziness and Soft Computing*, vol 163. Springer, Berlin, Heidelberg, https://doi.org/10.1007/3-540-32358-9_9 |
| **2** | Aburomman, A. A., & Ibne Reaz, M. Bin. (2016). A novel SVM-kNN-PSO ensemble method for intrusion detection system. *Applied Soft Computing Journal*, *38*, 360–372. https://doi.org/10.1016/j.asoc.2015.10.011 |
| **3** | Ashfaq, R. A. R., Wang, X. Z., Huang, J. Z., Abbas, H., & He, Y. L. (2017). Fuzziness based semi-supervised learning approach for intrusion detection system. *Information Sciences*, |
| **4** | Aslahi-Shahri, B. M., Rahmani, R., Chizari, M., Maralani, A., Eslami, M., Golkar, M. J., & Ebrahimi, A. (2016). A hybrid method consisting of GA and SVM for intrusion detection system. *Neural Computing and Applications*, *27*(6), 1669–1676. https://doi.org/10.1007/s00521-015-1964-2 |
| **5** | Banković, Z., Stepanović, D., Bojanić, S., & Nieto-Taladriz, O. (2007). Improving network security using genetic algorithm approach. *Computers and Electrical Engineering*, *33*(5–6), 438–451. https://doi.org/10.1016/j.compeleceng.2007.05.010 |
| **6** | Chen, W. H., Hsu, S. H., & Shen, H. P. (2005). Application of SVM and ANN for intrusion detection. *Computers & Operations Research*, 32(10), 2617-2634. doi:10.1016/j.cor.2004.03.019. |
| **7** | Chung, Y. Y., & Wahid, N. (2012). A hybrid network intrusion detection system using simplified swarm optimization (SSO). *Applied Soft Computing*, *12*(9), 3014–3022. https://doi.org/10.1016/j.asoc.2012.04.020 |
| **8** | Creech, G., & Hu, J. (2014). A semantic approach to host-based intrusion detection systems using contiguousand discontiguous system call patterns. *IEEE Transactions on Computers*, *63*(4), 807–819. https://doi.org/10.1109/TC.2013.13 |
| **9** | Damopoulos, D., Menesidou, S. A., Kambourakis, G., Papadaki, M., Clarke, N., & Gritzalis, S. (2012). Evaluation of anomaly-based IDS for mobile devices using machine learning classifiers. *Security and Communication Networks*, *5*(1), 3–14. https://doi.org/10.1002/sec.341 |
| **10** | De la Hoz, E., De La Hoz, E., Ortiz, A., Ortega, J., & Prieto, B. (2015). PCA filtering and probabilistic SOM for network intrusion detection. *Neurocomputing*, 164, 71–81. https://doi.org/10.1016/j.neucom.2014.09.083 |
| **11** | Elbasiony, R. M., Sallam, E. A., Eltobely, T. E., & Fahmy, M. M. (2013). A hybrid network intrusion detection framework based on random forests and weighted k-means. *Ain Shams Engineering Journal*, 4(4), 753–762. https://doi.org/10.1016/j.asej.2013.01.003 |
| **12** | Elhag, S., Fernández, A., Bawakid, A., Alshomrani, S., & Herrera, F. (2015). On the combination of genetic fuzzy systems and pairwise learning for improving detection rates on Intrusion Detection Systems. *Expert Systems with Applications*, 42(1), 193–202. https://doi.org/10.1016/j.eswa.2014.08.002 |

| Reference. No. | Reference details (selected sixty articles studied for meta-analysis in the research work) |
|---|---|
| **13** | Feng, W., Zhang, Q., Hu, G., & Huang, J. X. (2014). Mining network data for intrusion detection through combining SVMs with ant colony networks. *Future Generation Computer Systems*, 37, 127–140. https://doi.org/10.1016/j.future.2013.06.027 |
| **14** | Fisch, D., Hofmann, A., & Sick, B. (2010). On the versatility of radial basis function neural networks: A case study in the field of intrusion detection. *Information Sciences*, 180(12), 2421–2439. https://doi.org/10.1016/j.ins.2010.02.023 |
| **15** | Ganapathy, S., Kulothungan, K., Yogesh, P., & Kannan, A. (2012). A novel weighted fuzzy C -means clustering based on immune genetic algorithm for intrusion detection. In *Procedia Engineering* (Vol. 38, pp. 1750–1757). https://doi.org/10.1016/j.proeng.2012.06.213 |
| **16** | Grediaga, Á., Ibarra, F., García, F., Ledesma, B., & Brotóns, F. (2006). Application of Neural Networks in Network Control and Information Security. In *Advances in Neural Networks* ISNN 2006 (pp. 208–213). https://doi.org/10.1007/11760191_31 |
| **17** | Han, S.-J., & Cho, S.-B. (2006). Evolutionary neural networks for anomaly detection based on the behavior of a program. *IEEE Transactions on Systems, Man, and Cybernetics. Part B, Cybernetics : A Publication of the IEEE Systems, Man, and Cybernetics Society*, 36(3), 559–570. https://doi.org/10.1109/TSMCB.2005.860136 |
| **18** | Hoang, X. D., Hu, J., & Bertok, P. (2009). A program-based anomaly intrusion detection scheme using multiple detection engines and fuzzy inference. *Journal of Network and Computer Applications*, 32(6), 1219–1228. https://doi.org/10.1016/j.jnca.2009.05.004 |
| **19** | Horng, S. J., Su, M. Y., Chen, Y. H., Kao, T. W., Chen, R. J., Lai, J. L., & Perkasa, C. D. (2011). A novel intrusion detection system based on hierarchical clustering and support vector machines. *Expert Systems with Applications*, 38(1), 306–313. https://doi.org/10.1016/j.eswa.2010.06.066 |
| **20** | Hu, W., Hu, W., & Maybank, S. (2008). AdaBoost-based algorithm for network intrusion detection. *IEEE Transactions on Systems, Man, and Cybernetics*, *38*(2), 577–583. https://doi.org/10.1109/TSMCB.2007.914695 |
| **21** | Hu, W., Hu, W., & Maybank, S. (2008). AdaBoost-based algorithm for network intrusion detection. *IEEE Transactions on Systems, Man, and Cybernetics*, 38(2), 577–583. https://doi.org/10.1109/TSMCB.2007.914695 |
| **22** | Jazzar, M., & Jantan, A. (2008). An approach for anomaly intrusion detection based on causal knowledge-driven diagnosis and direction. *Studies in Computational Intelligence*, 149, 39–48. https://doi.org/10.1007/978-3-540-70560-4_4 |
| **23** | Jiang, H., & Ruan, J. (2009). The application of genetic neural network in network intrusion detection. *Journal of Computers*, 4(12), 1223–1230. https://doi.org/10.4304/jcp.4.12.1223-1230 |
| **24** | Jiang, S., Song, X., Wang, H., Han, J. J., & Li, Q. H. (2006). A clustering-based method for unsupervised intrusion detections. *Pattern Recognition Letters*, 27(7), 802–810. https://doi.org/10.1016/j.patrec.2005.11.007 |

| Reference. No. | Reference details (selected sixty articles studied for meta-analysis in the research work) |
|---|---|
| **25** | Khan, L., Awad, M., & Thuraisingham, B. (2007). A new intrusion detection system using support vector machines and hierarchical clustering. *The VLDB Journal*, 16(4), 507–521. https://doi.org/10.1007/s00778-006-0002-5 |
| **26** | Kim, G., Lee, S., & Kim, S. (2014). A novel hybrid intrusion detection method integrating anomaly detection with misuse detection. *Expert Systems with Applications*, 41(4 PART 2), 1690–1700. https://doi.org/10.1016/j.eswa.2013.08.066 |
| **27** | Kuang, F., Xu, W., & Zhang, S. (2014). A novel hybrid KPCA and SVM with GA model for intrusion detection. *Applied Soft Computing Journal*, 18, 178–184. https://doi.org/10.1016/j.asoc.2014.01.028 |
| **28** | Kumar, G., & Kumar, K. (2015). A Multi-objective Genetic Algorithm Based Approach for Effective Intrusion Detection Using Neural Networks. In *Intelligent Methods for Cyber Warfare* (pp. 173–200). https://doi.org/10.1007/978-3-319-08624-8_8 |
| **29** | Arun Raj Kumar, P., & Selvakumar, S. (2013). Detection of distributed denial of service attacks using an ensemble of adaptive and hybrid neuro-fuzzy systems. *Computer Communications*, 36(3), 303–319. https://doi.org/10.1016/j.comcom.2012.09.010 |
| **30** | Lei, J. Z., & Ghorbani, A. A. (2012). Improved competitive learning neural networks for network intrusion and fraud detection. *Neurocomputing*, 75, 135–145. https://doi.org/10.1016/j.neucom.2011.02.021 |
| **31** | Li, Y., Wang, J. L., Tian, Z. H., Lu, T. B., & Young, C. (2009). Building lightweight intrusion detection system using wrapper-based feature selection mechanisms. *Computers and Security*, 28(6), 466–475. https://doi.org/10.1016/j.cose.2009.01.001 |
| **32** | Liao, N., Tian, S., & Wang, T. (2009). Network forensics based on fuzzy logic and expert system. *Computer Communications*, 32(17), 1881–1892. https://doi.org/10.1016/j.comcom.2009.07.013 |
| **33** | Luo, B., & Xia, J. (2014). A novel intrusion detection system based on feature generation with visualization strategy. *Expert Systems with Applications*, 41(9), 4139–4147. https://doi.org/10.1016/j.eswa.2013.12.048 |
| **34** | Mabu, S., Chen, C., Lu, N., Shimada, K., & Hirasawa, K. (2011). An Intrusion-Detection Model Based on Fuzzy Class-Association-Rule Mining Using Genetic Network Programming. *IEEE Transactions on Systems, Man, and Cybernetics, Part C: Applications and Reviews*, 41(1), 130–139. https://doi.org/10.1109/TSMCC.2010.2050685 |
| **35** | Mitrokotsa, A., & Douligeris, C. (2005). Detecting denial of service attacks using emergent self-organizing maps. *Proceedings of the Fifth IEEE International Symposium on Signal Processing and Information Technology*, 2005., 375–380. https://doi.org/10.1109/ISSPIT.2005.1577126 |
| **36** | Njogu, H. W., Jiawei, L., & Kiere, J. N. (2013). Network specific vulnerability based alert reduction approach. *Security and Communication Networks*, 6(1), 15–27. https://doi.org/10.1002/sec.520 |

| Reference. No. | Reference details (selected sixty articles studied for meta-analysis in the research work) |
|---|---|
| **37** | Orfila, A., Estevez-Tapiador, J. M., & Ribagorda, A. (2009). Evolving high-speed, easy-to-understand network intrusion detection rules with genetic programming. In *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* (Vol. 5484 LNCS, pp. 93–98). https://doi.org/10.1007/978-3-642-01129-0_11 |
| **38** | Özyer, T., Alhajj, R., & Barker, K. (2007). Intrusion detection by integrating boosting genetic fuzzy classifier and data mining criteria for rule pre-screening. *Journal of Network and Computer Applications*, 30(1), 99–113. https://doi.org/10.1016/j.jnca.2005.06.002 |
| **39** | Pinzón, C. I., De Paz, J. F., Herrero, Á., Corchado, E., Bajo, J., & Corchado, J. M. (2013). IdMAS-SQL: Intrusion Detection Based on MAS to Detect and Block SQL injection through data mining. *Information Sciences*, 231, 15–31. https://doi.org/10.1016/j.ins.2011.06.020 |
| **40** | Powers, S. T., & He, J. (2008). A hybrid artificial immune system and Self Organising Map for network intrusion detection. *Information Sciences*, 178(15), 3024–3042. https://doi.org/10.1016/j.ins.2007.11.028 |
| **41** | Peddabachigari, S., Abraham, A., Grosan, C., & Thomas, J. (2007). Modeling intrusion detection system using hybrid intelligent systems. *Journal of Network and Computer Applications*, 30(1), 114–132. https://doi.org/10.1016/j.jnca.2005.06.003 |
| **42** | Ramasubramanian, P., & Kannan, A. (2006). A genetic-algorithm based neural network short-term forecasting framework for database intrusion prediction system. *Soft Computing*, 10(8), 699–714. https://doi.org/10.1007/s00500-005-0513-9 |
| **43** | Revathi, S., & Malathi, A. (2014). Multi-tier framework using sugeno fuzzy inference system with swarm intelligence techniques for intrusion detection. *Indian Journal of Science and Technology*, 7(9), 1437–1443. |
| **44** | Sangkatsanee, P., Wattanapongsakorn, N., & Charnsripinyo, C. (2011). Practical real-time intrusion detection using machine learning approaches. *Computer Communications*, 34(18), 2227–2235. https://doi.org/10.1016/j.comcom.2011.07.001 |
| **45** | Sani, R. A., & Ghasemi, A. (2015). Learning a new distance metric to improve an SVM-clustering based intrusion detection system. In *Proceedings of the International Symposium on Artificial Intelligence and Signal Processing, AISP 2015* (pp. 284–289). https://doi.org/10.1109/AISP.2015.7123497 |
| **46** | Shamshirband, S., Patel, A., Anuar, N. B., Kiah, M. L. M., & Abraham, A. (2014). Cooperative game theoretic approach using fuzzy Q-learning for detecting and preventing intrusions in wireless sensor networks. *Engineering Applications of Artificial Intelligence*, 32, 228–241. https://doi.org/10.1016/j.engappai.2014.02.001 |
| **47** | Sheikhan, M., & Jadidi, Z. (2014). Flow-based anomaly detection in high-speed links using modified GSA-optimized neural network. *Neural Computing and Applications*, 24(3–4), 599–611. https://doi.org/10.1007/s00521-012-1263-0 |
| **48** | Shon, T., & Moon, J. (2007). A hybrid machine learning approach to network anomaly detection. *Information Sciences*, 177(18), 3799–3821. https://doi.org/10.1016/j.ins.2007.03.025 |

| Reference. No. | Reference details (selected sixty articles studied for meta-analysis in the research work) |
|---|---|
| **49** | Sivatha Sindhu, S. S., Geetha, S., & Kannan, A. (2012). Decision tree based light weight intrusion detection using a wrapper approach. *Expert Systems with Applications*, 39(1), 129–141. https://doi.org/10.1016/j.eswa.2011.06.013 |
| **50** | Stopel, D., Moskovitch, R., Boger, Z., Shahar, Y., & Elovici, Y. (2009). Using artificial neural networks to detect unknown computer worms. *Neural Computing and Applications*, *18*(7), 663–674. https://doi.org/10.1007/s00521-009-0238-2 |
| **51** | Subbulakshmi, T., Mercy Shalinie, S., Suneel Reddy, C., & Ramamoorthi, A. (2010). Detection and classification of DDoS attacks using fuzzy inference system. In *Communications in Computer and Information Science (Vol. 89 CCIS, pp. 242–252).* https://doi.org/10.1007/978-3-642-14478-3_25 |
| **52** | Tajbakhsh, A., Rahmati, M., & Mirzaei, A. (2009). Intrusion detection using fuzzy association rules. *Applied Soft Computing*, 9(2), 462–469. https://doi.org/10.1016/j.asoc.2008.06.001 |
| **53** | Tong, X., Wang, Z., & Yu, H. (2009). A research using hybrid RBF/Elman neural networks for intrusion detection system secure model. *Computer Physics Communications*, 180(10), 1795-1801.doi:10.1016/j.cpc.2009.05.004. |
| **54** | Toosi, A. N., & Kahani, M. (2007). A new approach to intrusion detection based on an evolutionary soft computing model using neuro-fuzzy classifiers. *Computer Communications*, 30(10), 2201–2212. https://doi.org/10.1016/j.comcom.2007.05.002 |
| **55** | Tsang, C. H., Kwong, S., & Wang, H. (2007). Genetic-fuzzy rule mining approach and evaluation of feature selection techniques for anomaly intrusion detection. *Pattern Recognition*, *40*(9), 2373–2391. https://doi.org/10.1016/j.patcog.2006.12.009 |
| **56** | Wang, G., Hao, J., Ma, J., & Huang, L. (2010). A new approach to intrusion detection using Artificial Neural Networks and fuzzy clustering. *Expert Systems with Applications*, 37(9), 6225–6232. https://doi.org/10.1016/j.eswa.2010.02.102 |
| **57** | Yin, C., Tian, S., Huang, H., & He, J. (2005). Applying Genetic Programming to Evolve Learned Rules for Network Anomaly Detection. *Advances in Natural Computation, First International Conference, ICNC 2005, Proceedings, Part III*, 3612, 323–331. https://doi.org/doi:10.1007/11539902_38 |
| **58** | Zanero, S. (2005). Analyzing TCP traffic patterns using Self Organizing Maps. Image Analysis and Processing - *Iciap 2005, Proceedings*, 3617, 83–90. https://doi.org/10.1007/11553595_10 |
| **59** | Zhang, C., Jiang, J., & Kamel, M. (2005). Intrusion detection using hierarchical neural networks. *Pattern Recognition Letters*, 26(6), 779–791. https://doi.org/10.1016/j.patrec.2004.09.045 |
| **60** | Zhang, J.-R., Zhang, J., Lok, T.-M., & Lyu, M. R. (2007). A hybrid particle swarm optimization–back-propagation algorithm for feedforward neural network training. *Applied Mathematics and Computation*, 185(2), 1026–1037. https://doi.org/10.1016/j.amc.2006.07.025 |

*Table 6: Studied top works in the field of intrusion detection*