Archives of Computational Methods in Engineering (2021) 28:2861–2879
https://doi.org/10.1007/s11831-020-09478-2

ORIGINAL PAPER

A Review on the Effectiveness of Machine Learning and Deep Learning Algorithms for Cyber Security

R. Geetha[1] · **T. Thilagam**[1]

Received: 13 January 2020 / Accepted: 17 August 2020 / Published online: 2 September 2020

Abstract
In recent years there exists a wide variety of cyber attacks with the drastic development of the internet technology. Detection of these attacks is of more significant in today's cyber world scenario. Machine learning (ML) and deep learning (DL) methods have been preferred by researchers across different disciplines for providing solutions to their problems. In this paper we have presented a detailed classification of various DL/ML algorithms. In addition to that a focused survey on the use of various ML/DL methods for the detection of different categories of attacks has been presented. Furthermore the various platforms and tools used for implementing DL/ML methods are explored and the security solutions for the different categories of attacks are summarized.

1 Introduction

1.1 Cyber Security

The uncommon utilization of network connected devices and vital dependence on information communication technology throughout the world. Many malicious users try to subvert credentials or simply attack host data. Over the last few years, Loukas et al. [1], there have been different examples of both proofs of concept and real-world attacks. Cyber security analysts Toch et al. [2] and experts have structured and created throughout the years various cyber defense systems to shield resources of associations from malicious attackers. These systems address cyber security threats, for example, virus, Trojans, worms, and botnets, among others Loukas et al. [3]. Existing arrangements dependent on Intrusion Detection Systems (IDS) incorporate (master) dynamic ways to deal with envision and expel vulnerabilities in processing frameworks with which to trigger responsive activities for moderation. Any assurance instrument needs to work by coordinating calculations with great and exact identification capacities, permitting fast handling of the information accumulated by the data sources. Without these capacities, IDSs can't play out their checking and examination works continuously, making it relatively difficult to identify potential cyber assaults when they are beginning to occur. This issue is because of the way that present systems give progressively high transmission rates. All the more uncommonly, the rates have expanded from 100 Mbps a couple of years prior to the present information rate of 10CGbps in wired systems. Vast volumes of data owing through systems make IDSs insufficient to assemble and dissect each system parcel. For instance, Deep Packet Inspection (DPI) instruments like Koscher et al. [4], can work appropriately on wired systems up to 1 Gbps, beginning to dispose of parcels because of overhead from 1.5 Gbps Checkoway et al. [5]. An on-going report Ward et al. [6], directed concentrated examinations to extricate a careful execution correlation by utilizing Snort and the utilization of machine learning procedures on it, assessing such IDS to process organize traffic up to 10 Gbps arrange speed. These tests show that the normal bundles drop when utilizing Snort achieves 9. 5% in 4 Gbps systems while the normal parcels drop with 10 Gbps systems ascends to 20%. However, and because of the expansion of transfer speed, IDS-based solutions making utilization of deep analysis techniques were compelled to advance towards better approaches of detection. They moved from inspecting raw network packets to analysing traffic network flows with imaginative AI-based procedures McGraw et al. [7].

✉ R. Geetha
geetha@saec.ac.in

[1] Department of Computer Science and Engineering, S.A. Engineering College, Chennai, India

1.2 Classification of Cyber Attacks

A first dimension for classifying an attack is the goal of the attack. This is often related to the wayan adversary monetizes the attack (e.g., by stealing information and selling it to advertisers orcriminals). Overall, the attack goals fall into one of the following categories Lala et al. [8] as shown in Fig. 1.

(1) Stealing information, such as data on a device, media files, and user credentials; this actionis usually performed by spyware malware; (2) tracking user information, i.e., monitoring users'sensitive data (e.g., locations, activities, or health-related data); this action is usually achieved usingmobile malware; (3) taking control of a system, as is done by Trojan, botnet, and rootkit Stefano et al. [9].

A second dimension for classifying an attack is the attack vector and it represents the vulnerability exploited by an adversary to gain access to a network or computer system to perform malicious actions. Attack vectors can be identified at three different layers: Hardware, Network and Application.

The deep learning technologies are used for cyber security analysis and intrusion detection is highly relevant. Deep learning techniques are widely used for malware analysis and in finding unforeseen threats because of malicious software [10–15].

Fig. 1 Classification of cyber attacks

1.3 Machine/Deep Learning

Artificial neural network is the basis for all the latest deep learning models. Like the nodes in deep belief network and deep Boltzmann machines these models also include formulas and latent variables which are layer-wise arranged in deep generative models.

In most of the deep learning machines the input data is made into an abstract representation. This process knows when to learn a level optimally and performs accordingly. For example, an image recognition application first layer reads the raw data as an abstract the second layer encodes arrangement of edges, then third layer encodes a part of image, the fourth layer recognizes the image. As the data passes through several layers in this process it is mentioned by word 'deep' in deep learning.

More precisely, deep learning systems have a substantial path called credit assignment path. CAP is the chain of transformation from data collection for input to output it is the casual connection between input and output. In recurrent neural networks signal propagates in a layer not lesser than one time and so the CAP depth is unlimited. Whereas in a feed forward neural network CAP depth is that of the network in which it takes place. CAP of depth 2 can emulate any function and so it is shown as a universal approximate. Layers beyond that do not add to the function approximate ability of the network. Additional layers are used for several learning features. Deep learning identifies which feature will improve the performance.

Greedy layer-by-layer method is frequently used to construct the architectures of deep learning. Feature engineering is obviated by deep learning methods, in supervised learning tasks. The data is translated into intermediate representation similar to principal components and removes redundancy in representation by deriving layered structures. In unsupervised learning tasks also deep learning algorithm are employed. This is the vital feature because the availability of unlabelled data is higher than the labelled data. The deep structures which are in unsupervised manner are the neural history compressors and deep belief network.

All the machine learning algorithmic implementations won't be considered as Deep learning. Deep learning algorithms has the ability to learn huge complex data representations, their applications are enormous. For example, singular algorithm which has statistical mechanisms such as Bayesian algorithms, function approximations like decision trees. Deep learning has the ability to learn massive data indiscriminately. It is a neural model which has the concepts of computing nodes. This model has been drawn from the complex interconnected neuronal structures of human neurons for learning process.

Machine learning has two inherited concepts training and inference. The massive volume of data in training model has been divided into several sets namely training and testing sets and also validation set. The function approximation representations of training data are learnt by a machine learning algorithm. The effectiveness of the training process is validated by the validation sets. The test set determines the final accuracy and effectiveness of the previous data. The input data is given to a trained and implemented machine learning model from which the inferred output is got this is the concept of inference.

Deep learning has various concepts like regression, classification, clustering, auto encoding and others in order to perform the learning tasks by multi-layer neural networks. In the application of multiple layers of multiple nodes, every node gets the input from the previous layers, thus the input data provides the output representation. From this it is clear that the multiple interconnected neurons are more complex.

1.4 Types of Machine Learning

A Supervised Shallow Machine Learning Algorithms (SML)

- Classification

In this the result of learning class will be a set of classes, multi-class classification which results in one class from a set of classes and multi-label classification. Every class is compared with other class in a binary way Paul et al. [16]. As shown in Fig. 2.

i Naïve Bayes (NB)

Naive Bayes is a simple but surprisingly powerful algorithm for predictive modelling and this algorithm is a classification algorithm for binary (two-class) and multi-class classification problems. The technique is easiest to understand when described using binary or categorical input values. It is called naive Bayes *or* idiot Bayes because the calculation of the probabilities for each hypothesis is simplified to make their calculation tractable.

ii. Support Vector Machines (SVM)

"Support Vector Machine" (SVM) is a supervised machine learning algorithm which can be used for both classification and regression challenges. However, it is mostly used in classification problems. In addition to performing linear classification, SVMs can efficiently perform a non-linear classification using what is called the kernel trick, implicitly mapping their inputs into high-dimensional feature spaces.

iii. K-Nearest Neighbor (KNN)

In pattern recognition, the k-nearest neighbor's algorithmic program (k-NN) could be a non-parametric methodology used for classification and regression. K-NN could be a sort of instance-based learning, or lazy learning, wherever the operate is barely approximated regionally and every one computation is delayed till classification. The k-NN algorithmic program is among the best of all machine learning

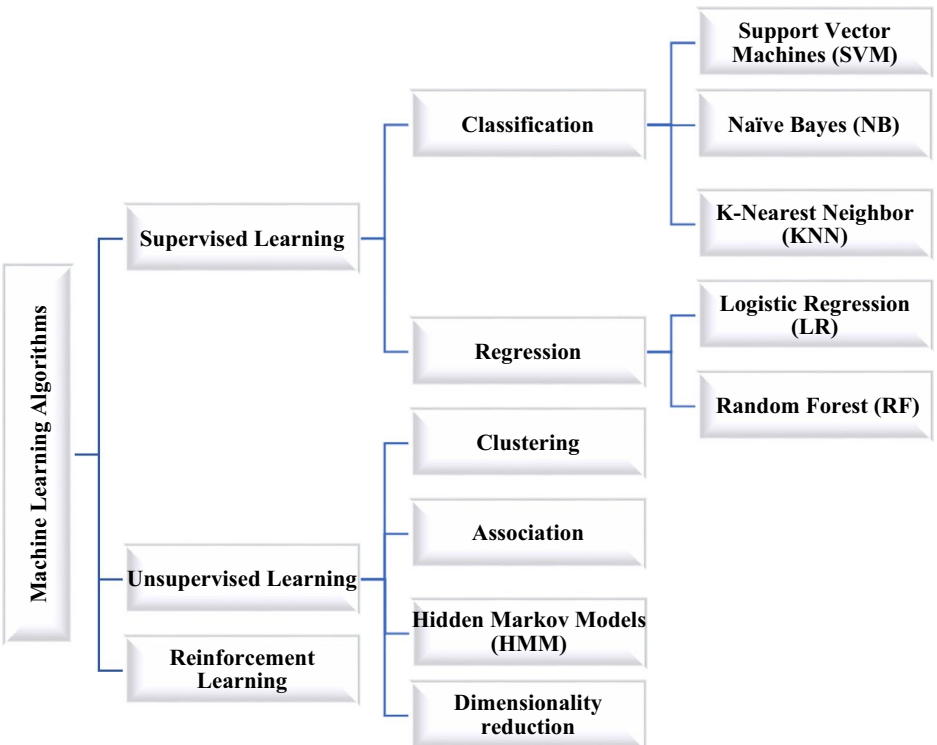

Fig. 2 Classification of machine learning algorithms

algorithms. Both for classification and regression, a helpful technique is wont to assign weight to the contributions of the neighbors, in order that the nearer neighbors contribute a lot of to the average than the more distant ones. K-Nearest Neighbors is one among the foremost basic however essential classification algorithms in Machine Learning. It belongs to the supervised learning domain and finds intense application in pattern recognition, data processing and intrusion detection.

- Regression

i Logistic Regression (LR)

Logistic regression predicts the chance of an outcome which will solely have 2 values (i.e. a dichotomy). The prediction is predicated on the employment of 1 or many predictors (numerical and categorical). This algorithmic rule too works as similar as NB algorithmic rule, however here their performance is highly not dependent on the size of the training data.

ii Random Forest (RF)

Random forest algorithmic rule will use each for classification and therefore the regression reasonably issues. Random forest algorithmic rule may be a supervised classification algorithmic rule. Because the name recommend, this algorithmic rule creates the forest with variety of trees. Within the same approach within the random forest classifier, the upper the quantity of trees within the forest offers the high accuracy results.

Shallow Neural Network (SNN)

SNN consists of process components that square measure organized into 2 or a lot of communication layers. SNN square measure supported neural networks, with restricted variety of neurons and layers. These square measure principally used for classification activities.

B. Unsupervised Shallow Machine Learning algorithms

- Clustering

Clustering means that grouping of knowledge points, a number of the famed cluster strategies square measure hierarchical cluster and k-means. They need restricted quantifiability. These cluster strategies represent versatile answer which might be used as preliminary part before adopting a supervised algorithmic rule or for anomaly detection functions.

- Association

The unknown patterns between information square measure known and that they square measure created appropriate for prediction functions. They will manufacture AN output of not essentially valid rules, therefore they need to be given correct inspections by a personality's knowledgeable.

- Hidden Markov Models (HMM)

Hidden Markov Model (HMM) may be an applied mathematics Markov model during which the system being sculptured is assumed to be a Markov process with unobserved. The hidden mathematician model will be delineated because the simplest dynamic theorem network. Hidden mathematician models square measure particularly famed for his or her application in reinforcement learning and temporal pattern recognition like speech, handwriting, gesture recognition, part-of-speech tagging, sheet music following,Wang et al. [17] partial discharges and bioinformatics. In cyber security, HMM square measure principally used with labeled datasets.

- Dimensionality reduction

It is chiefly meted out for cupboard space reduction. It's completely different kinds of element and discriminant analysis. Auto-encoders create the input file into AN encoded output type Su et al. [18].

- Density estimation

It is a applied mathematics extraction or approximation of the info distribution; it finds the density of subgroups of knowledge to analysis correlations Marquardt et al. [19].

C. Reinforcement Learning

Reinforcement learning is a category of machine learning which is attracted by its psychological action. Its principle is analogous to infant learning new things from its past behavior to complete a new activity. This learning is entirely different from the other algorithms where it has not been given any instructions to complete the task instead it does it on its own and learns from the previous experience. Real time examples of this category are self-driving cars which reaches the right destination from the previous journey experience or a program like chess that takes the next step from the reward and punishments received from the previous moves that leads to a winning of the game. In this case the program or agent which perceives the environment and takes action tries to maximize the rewards to achieve the goal. This penalty of reward is achieved through dynamic programming.

1.5 Types of Deep Learning

DL algorithms are based on Deep Neural Networks (DNN), they are large neural networks organized in many layers which are capable of autonomous representation learning as shown in Fig. 3.

1 Supervised DL algorithms

- Fully-connected Feed forward Deep Neural Networks (FNN).

FNN will provide general purpose and flexible solutions for classification tasks, whose computational cost are little expensive. They are a variant of DNN, where the existing layer neurons will be connected to neurons in the previous layer.

- Convolutional Feed forward Deep Neural Networks (CNN).

CNN is very effective in analyzing spatial data this is because the neuron gets its input from neurons in the previous layer. CNN is not suitable for analyzing non spatial data. It has lower computation cost than FNN.

- Recurrent Deep Neural Networks (RNN).

In RNN the neurons send its output to previous layer neurons, this character makes them harder to train than FNN. They excel as sequence generators.

2. Unsupervised DL algorithms

- Deep Belief Networks (DBN).

DBN are designed using a composition of *Restricted Boltzmann Machines* (RBM), which is a class of neural networks without any output layer. DBN is good in feature extraction so that it can be used for pre-training tasks. They too need training phase, but with unlabeled datasets.

- Stacked Auto encoders (SAE).

SAE is composed of many auto encoders; it is a class of neural networks with the same number of input and output neurons. SAE performs similar to DBN; it works better for small data sets.

2 Machine Learning/Deep Learning Algorithms

Support Vector Machine

Xin et al. [20–23] discussed Support Vector Machine which belongs to the category of supervised learning that is applied for regression and classification. It is one of the most robust algorithms that solve the problems related to classification which plots the data items in n-dimensional space as points where the various features of the given data acts for the given coordinates. It separates the different data groups using the boundaries based on decisions. It supports Support Vector Classification (SVC) and Support Vector Regression (SVR).

It supports both binary and multi-class classifications. A set of instances having different class values between two groups are separated by using decision boundaries. SVC works based on these decision boundaries. The support vector, which is the closest point to the separation hyperplane, it determines the optimal separation hyper plane. The mapping input vectors located on the separation hyperplane side of the feature space fall into one class, and the positions fall into the other class on the other side of the plane during the classification process. Kernel functions are used by SVM in the case of not linearly separable data points in order to map them into higher dimensional spaces so that they become separable in those spaces Sharma et al. [24].

Saxena [25] proposed a Hybrid PSO-SVM approaches for building IDS. Information Gain and BPSO, these two feature reduction techniques were used in the study. The number of attributes reduced to 18. The classification performance was reported as 99.4% on the DoS, 99.3% on Probe or Scan, 98.7% on R2L, and 98.5% on the U2R. In the case of a Denial of Service (DoS) attack and achieves a good detection rate in the case of U2R and R2L attacks are provided with good detection rates using this method.

K-Nearest Neighbor (KNN)

The difference or similarity between two instances is measured using KNN classifier which is based on a distance function.

Rao et al. [26] used Indexed Partial Distance Search k-Nearest Neighbor (IKPDS) to experiment with various attack types and different k values (i.e., 3, 5, and 10). 12,597 samples from the NSl-KDD dataset are randomly selected to test the classification results, resulting accuracy is 99.6% and classification time is very least. Experimental results show that IKPDS, and Network Intrusion Detection Systems (NIDS), have better classification results in a short time.

Vishwakarmaet et al. [27], AkNN intrusion detection method based on the ant colony optimization algorithm (ACO), pre-training the KDD Cup 99 dataset using ACO [28], and studies on the performance of kNN-ACO, BP Neural network and support vector machine for comparative analysis with common performance measurement parameters. The accuracy rate for the method is 94.17%, and the overall FAR is 5.82%. Very small dataset is used for this method.

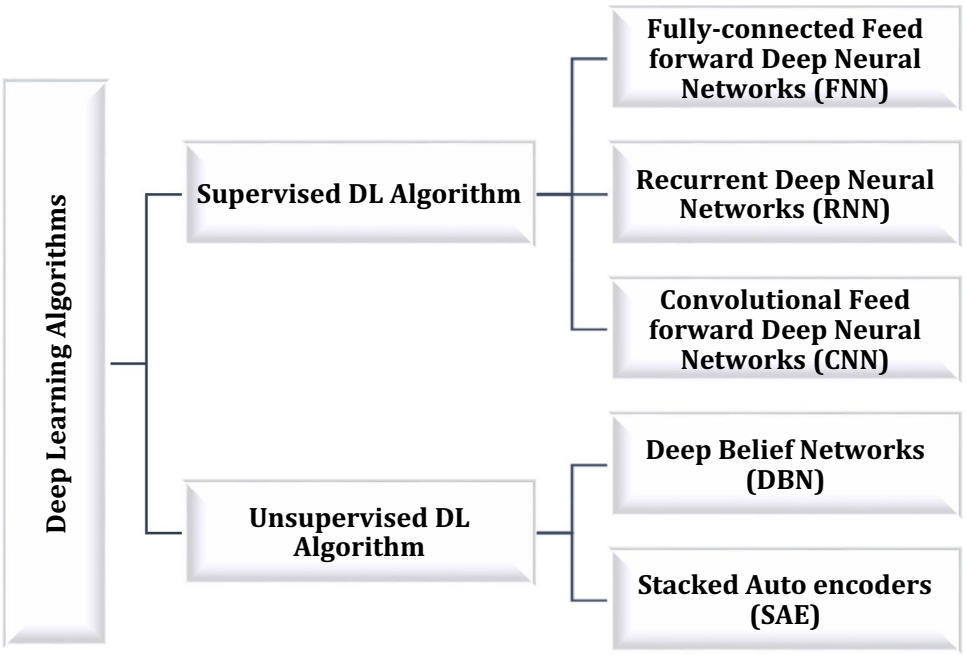

Fig. 3 Classification of deep learning algorithms

Deep Belief Network

Deep Belief Network (DBN) is a probabilistic generative model has multiple layers of stochastic and hidden variables. The Restricted Boltzmann Machine (RBM) and DBN are interrelated in order to train data efficiently through activations of one RBM for further training stages Kwon et al. [29] many hidden layers are enabled while composing and stacking a number of RBMs. Based on an energy function that can describe the high-order interactions between variables a modelling method generated from statistical physics which is the principle of Boltzmann machine (BM). RBM is a topological structure of a BM. BM is a plurality of hidden layers and a symmetric coupled random feedback binary unit neural network composed of a visible layer. The network node has two units one is the visible unit, and the other one is hidden unit which is used to express a random network and a random environment. The correlation between units is expressed using learning model by weighting the units.

Ding et al. [30], apply Deep Belief Nets (DBNs) to detect malware. The PE files are used as samples which are taken from internet. DBNs are made less prone to over fitting than feed forward neural networks initialized with random weights this is done by unsupervised pre-training algorithm. DBNs produce better classification results than any other learning techniques, such as SVM, KNN, and decision tree; this is because the DBNs can learn from additional unlabelled data. The accuracy rate of the method is 96.1%.

Nadeemet et al. [31], combine neural networks with semi supervised learning to achieve better accuracy using a very small number of labelled samples. KDD Cup 99 datasets are used for tracing the non-labelled data through the Ladder Network and DBN is used to classify data of the label, the obtained accuracy is 99.18% very similar to supervised learning.

Gaoet et al. [32], used different DBN structures and compared them and adjusted the number of layers and number of hidden units in the network model, in order to obtain a four-layer DBN model. For testing KDD Cup 99 dataset was used, the accuracy, precision and FAR of the model were 93.49%, 92.33% and 0.76%.

Zhao et al. [33] aim at the problems of a large amount of redundant data, long training time, and ease of falling into a local optimum in intrusion detection. Based on deep belief network (DBN) and probabilistic neural network (PNN) an intrusion detection method is proposed. Using the DBN nonlinear learning the original data are converted to low dimensional data, n order to retain the basic attributes of the original data. Then the number of hidden nodes in each layer is optimized using the particle-swarm optimization algorithm in order to obtain the best learning performance. PNN is used to classify the low-dimensional data. For testing, KDD CUP 99 dataset was used. The accuracy, precision and FAR of the experimental results were 99.14%, 93.25% and 0.615%.

Alrawashdeh et al. [34] implemented a method for fine tuning thee deep network. The method is based on a deep belief network using Logistic Regression soft-max. To improve the overall performance of the network, the multiclass Logistic Regression layer was trained with 10 epochs on the improved pre-trained data. This method resulted a low

false negative rate of 2.47% and detection rate of 97.9% on the total 10% KDD Cup 99 test dataset.

Alom et al. [35] proposed DBN that has gone through a series of experiments in order to find its intrusion detection capabilities; this is done after training with 40% NSL-KDD datasets. The trained DBN network can effectively identify unknown attacks assigned to it up to the accuracy rate of 97.5%.

Tan et al. [36], design a DBN-based ad hoc network intrusion detection model and conduct a simulation experiment on the NS2 platform. This experiment shows that DBN can get better accuracy and applied to Ad hoc network intrusion detection technology. The accuracy and FAR were 97.6% and 0.9%.

Recurrent Neural Networks

The sequence data is processed using the recursive neural network (RNN). The data flows from the input layer to the hidden layer to the output layer, every layer is connected to each other and there is no connection between nodes in the traditional neural network, it cannot solve much problems. The strong manifestation in RNN is that the network can remember the information of the previous moment and can apply it to the calculation of the current output, this is because, the RNN relates the current output of a sequence to the previous output. Here, the nodes between the hidden layers become connected, and the input of the hidden layer includes both the output of the input layer and the previous hidden layer output. Any length of sequence data RNN can be processed theoretically, but practically in order to reduce complexity it is often assumed that the current state is only related to the previous states.

Yin et al. [37], proposed a cyclic neural network propose intrusion detection (RNN-IDS). To test the performance of the model in binary classification and multi-class classification, NSL-KDD dataset was used. The test accuracy of binary classification is 83.228% and the test accuracy of multi-classification is 81.29%.

Staudemeyer et al. [38] proposed intrusion detection that implements the LSTM recurrent neural network classifier. The LSTM classifier has certain advantages over the detection of DoS attacks than any other static classifiers in the 10% KDD Cup 99 dataset. The accuracy rate was 93.5% and FAR was 1.622%.

Convolutional Neural Networks

The recursive neural system (RNN) is utilized to process on consecutive information. In the traditional neural network model, information from the input layer to the hidden layer to the output layer; the layers are completely associated and there is no association between the hubs between each layer.

Convolutional Neural Networks (CNN) is a kind of artificial neural system that has turned into a hotspot in the field of discourse analysis and picture acknowledgment. Its weight-sharing network structure makes it increasingly like a natural neural network, thus reducing the complexity of the network model and reducing the number of loads.

This preferred standpoint is progressively evident when the network input is a multi-dimensional picture, and the picture can be straightforwardly utilized as the contribution of the system to stay away from the mind boggling highlight detachment and information reproduction in the customary acknowledgment calculation. The Convolutional Network is a multi-layered sensor explicitly intended to perceive two-dimensional shapes that are very invariant to translation, scaling, tilting, or other forms of deformation Bu et al. [39].

CNN is the primary genuinely successful learning algorithm for training multi-layer arranges structures. It decreases the quantity of parameters that must be figured out how to enhance the preparation execution of the BP calculation through spatial connections. As dl design, CNN is proposed to limit the information pre-preparing necessities. The most dominant part of CNN is the taking in highlight orders from a lot of unlabeled information. In this manner, CNN are promising for application in the system interruption identification field.

Wang et al. [40] proposed a malware traffic arrangement technique utilizing a convolutional neural system by taking traffic information as pictures.

3 Machine/Deep Learning Platforms

3.1 Machine Learning Platforms

- H2O

H2O is a completely open gracefully, disseminated in-memory AI stage with direct quantifiability. It was planned by water.ai and is composed inside the Java, Python and R programming dialects. H2O bolsters the first wide utilized applied science and machine learning calculations along with inclination helped machines, summed up direct models, profound learning and extra.

H2O conjointly has AN exchange driving Auto ML reasonableness that precisely goes through all the calculations and their hyper boundaries to gracefully a pioneer leading body of the most straightforward models. The water stage is utilized by more than 18,000 associations universally and is very fashionable in each the R and Python people group. It is accessible on Linux, MacOS and Microsoft Windows working frameworks. H2O can likewise be utilized to break

down datasets in the cloud and Apache Hadoop document frameworks.

- PMLS

Parallel ML System (PMLS) is an appropriated AI structure. It deals with the difficult framework "plumbing work", permitting you to concentrate on the ML. PMLS runs with efficiency at scale on investigation bunches and cloud reason like Amazon EC2 and Google GCE. PMLS gives basic circulated programming devices to handle the difficulties of ML at scale: Big Data (numerous information tests), and Big Models (enormous boundary and halfway variable spaces). To address these difficulties, PMLS gives 2 key stages: Bosen, a bounded-offbeat key-esteem store for Data-Parallel ML calculations Strads, a scheduler for Model-Parallel ML calculations.

- Infosys Nia

It is an information based AI stage, designed by Infosys in 2017 to accumulate and blend organizational information from people procedures and legacy frameworks into a self-learning mental article. It is intended to handle intense business errands like forecast revenues and what product should be built, understanding customer conduct and additional. Infosys Nia permits organizations to oversee customer requests essentially, with a order-to-money strategy with chance mindfulness conveyed in timeframe.

- Accord.NET Framework

It is an AI structure that is joined with sound and picture preparing libraries written in C#. The system is intended for designers to make applications like pattern acknowledgment, pc vision, pc tryout (or machine tuning in) and signal procedure for industrial use. The Accord.NET Framework is isolated into various libraries for clients to choose from. These exemplify logical registering, sign and picture procedure and backing libraries, with alternatives like common learning calculations, continuous face recognition and that just the beginning.

- IBM Watson

IBM is a major part in the field of AI, with its Watson stage lodging a variety of apparatuses designed for the two designers and business clients. Accessible as a gathering of open arthropod genus, Watson clients can approach a great deal of test code, starter units and might fabricate psychological component web indexes and virtual specialists. Watson conjointly includes a chatbot structure stage pointed toward novices, which needs almost no AI skills. Watson can considerably offer pre-prepared substance for chatbots to make instructing the hatchling a lot of quicker.

- DiffBlue

Founded by Daniel Kroening at the University of Oxford, DiffBlue is a committed code automation stage. Furthermore, it is a simple anyway accommodating one at that. Its point is to iscover bugs, refactor code, perform test composing and find and fix shortcomings in code, all done by means of automation.

- Nervana Neon

Nervana and Intel have united to fabricate the up and coming age of astute specialists and applications and Neon is its open source Python-based AI library. Established in 2014, Neon allows designers to fabricate, prepare and send profound learning advances in the cloud. Neon has loads of video instructional exercises and a 'model zoo' which houses pre-prepared calculations and example contents.

- Apache Spark MLlib

Apache Spark MLlib is an in-memory data processing framework. It options an oversized info of algorithms that specialize in classification, regression, clustering and collaborative filtering. Within the Apache setup there's conjointly AN open supply framework referred to as Singa that provides a programming tool for deep learning networks across various machines.

The Comparison of the various Machine Learning Platforms is shown in Table 1.

3.2 Deep Learning Platforms

A Tensorflow

TensorFlow is a collection of open source programming library meant for dataflow over a range of tasks. It is an emblematic math library, and is likewise employed for AI related applications such as neural systems.

Tensorflow is a computational system for building AI models. TensorFlow provides a wide range of toolboxs that permit you to develop models at your liked level of reflection. On the other hand, you can utilize more elevated level APIs to indicate predefined structures, for example, straight repressors or neural systems.

TensorFlow is able to run over numerous GPUs and CPUs (with possible extensions such as SYCL & CUDA in the application of the computation of graphics processing units). It is present on operating systems such as macOS ,64- bit Linux, Windows, Android and iOS. Its flexible engineering

takes into consideration the simple arrangement of calculation over an assortment of platforms (TPU,GPU,CPU), and from work areas to bunches of workers to portable and edge devices.

The calculations of TensorFlow are communicated as dataflow charts with states. The word TensorFlow is coined from the activities of the performance of neural systems over multidimensional information arrays, that are referred as tensors. In May 2017, Google declared a product stack specifically for portable turn of events, TensorFlow Lite. The variant of TensorFlow is Lite which is meant for mobile and inserted AI. This variant provides an API for Android Neural Networks. Shi et al. [41] in their recent research indicated that performance of TensorFlow is the best in worker grade of multi-threaded execution environment at present with more than 8 threads.

B. Deeplearning 4 J

Deeplearning4j (DL4J) is another open-source programming delivered under Apache License 2.0. It is upheld commercially by the startup Skymind, which packs DL4J, Tensorflow, Keras and other deep learning libraries in an undertaking conveyance called the Skymind Intelligence Layer. Deeplearning4j is the principal business grade, open-source, circulated profound learning library written for Scala and Java. Incorporated with Apache Spark and Hadoop, DL4J carries AI to business situations for use on appropriated CPUs and GPUs. DL4j incorporates implementation the limited Boltzmann machine, profound conviction net, profound auto encoder, stacked denoising auto encoder and recursive neural tensor system, word2vec, doc2vec, and Glove. These algorithms all include conveyed equal versions that integrate with Apache Hadoop and Spark.

Deeplearning4j is the first commercial-grade, open-source, distributed deep-learning library written for Java and Scala. Integrated with Hadoop and Apache Spark, DL4J brings AI to business environments for use on distributed GPUs and CPUs. Deeplearning4j includes implementations of the restricted Boltzmann machine, deep belief net, deep auto encoder, stacked denoisingauto encoder and recursive neural tensor network, word2vec, doc2vec, and Glove. These algorithms all include distributed parallel versions that integrate with Apache Hadoop and Spark.

C. Theano

Theano [42] is a Python library that permits you to characterize, enhance, and assess scientific expressions including multi-dimensional clusters proficiently. In Theano, calculations are expressed utilizing a NumPy-esque sentence structure and arranged to run proficiently on either CPU or GPU models. Theano is an open source venture principally created by a Montreal Institute for Learning Algorithms at the University de Montréal. Theano too supports tensor activities, and GPU calculation, those sudden spikes in demand for Python 2 and 3, and supports parallelism through BLAS and SIMD support.

D. Torch

Torch [43] is a logical processing structure with wide support for AI based algorithms that place GPUs first. It is anything but difficult to utilize and proficient, because of a simple and quick scripting language, LuaJIT, and a basic C/CUDA usage. The objective of Torch is to have greatest adaptability and speed in building your logical calculations while making the procedure incredibly basic. Light accompanies a huge biological system of network driven packages in AI, PC vision, signal preparing, equal handling, picture, video, sound and systems administration among others, and expands on head of the Lua people group. At the heart of Torch are the well known neural system and improvement libraries which are easy to use, while having most extreme adaptability in actualizing complex neural system geographies. Construct discretionary charts of neural systems, and parallelize them over CPUs and GPUs in an effective way. Light is continually advancing: it is as of now utilized inside Facebook, Google, Twitter, NYU, IDIAP, Purdue and a few different organizations and exploration labs Shi [41].

Table 1 Comparison of machine learning platforms

Framework	License	Core language	CPU	GPU	Open source
H2O	Apache license 2.0	Java, Python, and R.	✓	✓	✓
PMLS	Apache License 2.0	C++	✓	✓	✓
Infosys Nia	Infosys	C++	✓	✓	×
Accord.NET framework	LGPLv3 and partly GPLv3	C#	✓	✓	✓
IBM Watson	IBM	Python, REST API	✓	✓	✓
DiffBlue	GNU license	Java, Python, JavaScript C# and PHP.	✓	✓	✓
Nervana Neon	Neon	Python-	✓	✓	✓
Apache Spark MLlib	Apache	R, Python	✓	✓	✓

E. Microsoft Cognitive Toolkit (CNTK)

The Microsoft Cognitive Toolkit [44] is another deep learning structure created by Microsoft Research. Microsoft Cognitive Toolkit portrays neural systems as a progression of computational steps through a coordinated diagram. It very well may be incorporated as a library in C#, Python and C++ projects, or be utilized as an independent with its own scripting language named BrainScript. It can likewise run evaluation elements of models from Javacode, and makes use of ONNX which is a open-source neural Network model that permits move amid other profound learning structures like PyTorch ,Caffe2, MXNet). CNTK has been developed as a computationally integral asset for machine learning with execution like different stages that have seen longer turn of events and more far reaching use Shi et al. [45].

F. Caffe and Caffe2

CAFFE (Convolutional Architecture for Fast Feature Embedding) [46–48] is a profound learning framework, Caffe gives a total toolbox to preparing, testing, finetuning, and sending models, with very much archived models for these undertakings and GPU support for profound learning and fundamentally picture characterization, beginning in 2014. A Caffe layer is the embodiment of a neural system layer and its Model definitions is that arrange records utilizing convention buffer language. Caffe2, as a major aspect of Facebook Source, Research and Facebook Open constructs upon the original Caffe project, actualizing an extra Python API, supports Mac OS X,Windows, Linux, iOS, Android, and other form stages.

G. MXNET

Apache MXNet [49] is an advanced open-source deep learning programming system, used to train, the deep neural network. Apache MXNet is a quick and versatile preparing and inference structure with a simple to-utilize, compact API for AI and it has been designed with a system point of view to minimize the overhead of loading of data and the Input/Output complexity. In addition it provides the interface named Gluon that permits the developers with different levels of skill set to get trained with cloud based deep learning on edge devices as well as on mobile apps. Gluon with a few lines of code permits fabrication of straight relapse, convolution systems and intermittent LSTMs for object detection, discourse acknowledgment, proposal, and personalization. It has been proved as efficient especially in the implementations of single and multi-GPU where the implementations based on CPU seems to be inadequate.

H. Keras

Keras [50] is another important API developed for the neural systems. It has been developed in Python and suitable for running over either CNTK or TensorFlow or Theano. It has been developed with the intention of empowering experimentation in a quick way. The required models are demonstrated using small and simple Python code that is easy to troubleshoot and simple to extend. A model is treated as a group of independent and complete task where the modules are configured in such a way that they can not be stopped with certain limitations as could be expected under the circumstances. To be more specific the cost capacities, neural layers, installment plans, enhancers, regularization plans ,initiation functions are the modules that are independent that you with each other and can be joined to build new models.

The Comparison of the various Deep Learning Platforms is shown in Table 2

4 Attack Categories

4.1 Hardware Attacks

Tehranipoor and Koushanfar et al. [51] at the hardware level, attacks are found that include manufacturing backdoors, gaining access to memory, and hardware tampering. The common goal of these attacks is twofold: modifying the hardware to access sensitive information and creating a backdoor (e.g., install an invisible program in the hardware circuit) that can be used to regain access to the compromised machine. Such hardware attacks can be applied to several types of devices, such as network appliances, surveillance systems, and industrial control systems.

4.2 Network Attacks

Schweitzer et al. [45] network attacks can target the network protocol or the network device software, and their goal is either the denial of service or hijacking a network connection to steal sensitive data. Specifically, frequent attacks using vectors at the network layer are Denial of Service (DoS), IP spoofing and man in the middle attacks. Desmedt et al. [15].

Network layer

Zolotukhin et al. [52] proposed the use of stacked auto encoders that recognizes the application layer DDoS attacks in the presence of encoded traffic. The proposed framework focuses on the clusters of the usual patterns of traffic observing the anamoly detection for the trivial DoS attacks without decrypting the packets flowing through the network.Moreover, the SAE to identify attacks intended to emulate run of the mill program movement dependent on their development

error of vectored discussion traffic bunches in time stretches. Kim et al. [53] developed an IDS using a LSTM recursive neural network. The model is applied over the KDDCup 1999 dataset that is able to remember around 22 attacks that falls under four categories. In correlation with other neural network systems utilizing the same training data, the authors proposed work shows improved rate of detection and precision particularly over DoS attacks. DL is additionally appropriate to ensure the security of real-world applications with respect to the fundamental circumstance of analysis. Wang et al. [54], proposed a deep convolutional framework for presenting the detection of humans from the images captured through the cameras. They have first pre-trained the network over the dataset of Image Net and further tuned it by training on the CUHK03 re-identification dataset.

There are some changes to the completely associated layers of the model in retraining on the second dataset; the creators can altogether increase the coordinating rate over the available plans. Niimi [55], explored the utilization of profound learning for credit card endorsement assurance and exchange approval. The author has validated the system in R and executed in Amazon's EC2 cloud stage. The assessment experiment proved the comparative exactness to different learning techniques with higher accuracy.

Application Attacks. Application attack is a category of phishing and client-side web attacks are common and frequent attacks, as indicated by fundamental privacy showcase players. The example attacks include email administrations and programs which is a serious threat faced by internet. Concerning through email, phishing is a type of scam where the aggressor attempts to assemble delicate data. Example of this sort includes the credit card numbers b e i n g impersonated by a reputed person t h r o u g h email and other channels of communication [56, 57]. Numerous attacks that belong to application-level category uses social designing techniques that involves users to negotiate the frameworks and manipulate them in forwarding the attack over deception [58]. A typical example of customer side web assaults includes Cross-Site Scripting that involves inserting client-side substance code such as JavaScript into the pages. The code that has been injected could be used for alternative purposes later on. For example to avoid getting the chance to control or to drive a customer to execute a couple of exercises on a distant site in light of a legitimate concern for the attacker.

A large number of application level attacks can be categorized as malware [59].

Malware is any vindictive programming that an aggressor out how to run on the objective PC. It is used to assemble delicate data, to access private PC frameworks, or to perform monstrous attack. Malware is characterized by its pernicious plan, acting in opposition to client requirements. Malware can be arranged into a few classes relying upon the plan objective. The most common malware classes are versatile malware, botnets, ransomware and banking malware. Various methods are utilized to introduce malware on an target system. For instance, portable malware is introduced by means of SMS, through unofficial application stores, or by abusing weaknesses of the OS. Once the malware is introduced, it can play out a few noxious activities like illegal access of user data or following client activities.

Application Layer

Zhu et al. [60] introduced DeepFlow which is an Android malware location gadget. The developed instrument uses FlowDroid to get the progression of information from delicate sources to sensitive sinks. The SUSI method is additionally acquired to change the information streams that include detailed arrangement of information. They have arranged the applications by means of signal conviction connect with

Table 2 Comparison of deep learning platforms

Framework	License	Core language	Bindings	CPU	GPU	Open source	Training	Pertained models	Development
Caffe	BSD	C++	Python, MATLAB	✓	✓	✓	✓	✓	Distributed
Theano	BSD	Python		✓	✓	✓	✓	✓	Distributed
Torch	BSD	Lua		✓	✓	✓	✓	✓	Distributed
Cuda-Convnet	Unspecified	C++	Python		✓	✓	✓		Discontinued
Tensor flow	Google Brain's	C++	C++ Python	✓	✓	✓	✓	✓	Distributed
Deep learning 4J	Apache2.0	JAVA, C++, Python	JAVA, C++, Python	✓	✓	✓	✓	✓	Distributed
CNTK	MIT	C++		✓	✓	✓	✓	✓	Distributed
MXNET	Apache2.0	C++ Python R, Julia	C++ Python MATLAB	✓	✓	✓	✓	✓	Distributed
Keras	MIT	Python	Python	✓	✓	✓	✓	✓	Distributed

the transformed information streams as info. Ding et al. [61], extracted opcode successions from Windows for malware order by means of DBN. Changing over the opcode highlights to n-gram portrayals, the highlights were down chosen by most extreme data gain and record recurrence. The creators exhibit both the limit of DBNs to perform classification, just as to perform auto encoding for unaided component selection to enhance the exhibition of shallow learning models like K-Nearest, Decision Tree etc.

The location of on-going assaults continuously is vital to empower ideal reaction and mitigation procedures. Uwagbole et al. [62], planned a framework to identify and forestall SQL infusion assaults by means of half and half static and unique investigation using profound learning procedures. Their intermediary based framework consolidates pattern coordinating with numerical element encoding for neural system and calculated relapse classification. The comparison of the types of Deep Learning attacks in different layers is shown in Table 3.

5 Related Research Review

The three important fields where most cyber ML algorithms are discovering application are Intrusion detection, malware analysis, and spam detection. An overview of each of these is explained below.

Pierazzi et al. [64] plans to find illegal exercises inside a computer or a network through IDS. Network IDS are widely delivered in modern enterprise networks. These frameworks were customarily founded on examples of known attacks, but yet current organizations incorporate different methodologies for anomaly detection, threat detection and order dependent on AI. Inside the raise of interruption recognition area, area, two explicit issues are applicable to our examination: the finding of botnets and of Domain Generation Algorithms. A botnet is a system that contains tainted machines controlled by aggressors and abused to do various unlawful exercises. Botnet recognition objective is to identify correspondences between tainted machines inside the botnet organize and the external order and control workers. Notwithstanding many examination proposition and business tools that address this danger, a few botnets still exist. DGA Spontaneously produces area names, and is frequently utilized by a tainted machine to speak with outside server(s) by periodically creating new hostnames. They show a genuine danger for associations on the grounds that, through DGA, which depends on language preparing strategies, it is conceivable to avoid defences dependent on static boycotts of area names. The authors have made use of ML based DGA discovery strategies.

Malware analysis is an incredibly applicable issue since current malware can automatically produce novel variations with indistinguishable malevolent impacts yet showing up from completely unique executable records. These polymorphic and transformative highlights rout the customary guideline based malware recognizable proof methodologies. ML strategies can be utilized to find malware variations and ascribing them to the right malware family.

Spam and phishing detection It has an enormous arrangement of procedures for diminishing the waste of time and potential risk brought about by undesirable messages. Phishing, speak to the route through which an attacker enters an enterprise network. Attackers are using advanced evasion strategies so that the spam and phishing detection is becoming much difficult. The spam detection process can be improved by using ML approaches.

Using supervised and unsupervised methods many approaches are introduced to tackle the above mentioned situations. The framework are made programmed with the presentation of additional limitations in neural systems such frameworks are called Bridged Multi-Layer Perceptron (BMLP) structures. Tensor auto-encoder is utilized for learning highlights from heterogeneous information and stacked to manufacture tensor profound learning model to learn various degrees of information portrayal. Tensor-based information portrayals are utilized to show nonlinear relationship of information. The exhibition of the proposed tensor profound learning model against the Stacking Auto-Encoder is analyzed by considering representative classification data sets like STL-10, CUAVE, SNAE2 and INEX 2007B. (Jan et al. [65]).

Diro et al. [66] cyber security is the serious problem in the computer sector as the demand of computer is increasing time to time. It is also a known fact thousands of zero day attacks emerging in the field of Internet of things [IoT]. For advance mechanism such as machine learning face difficulty in detecting cyber-attacks. On the other hand success of deep learning (DL) deals with the issues faced by cyber security. The application of deep learning becomes practical because of improved CPU and neural network algorithm. The use DL in cyber-attacks is succeeding in small mutation and novel attacks because of strong extracting capabilities. The self-taught and compression capabilities of deep learning architecture is the key mechanism for detecting the attacks in discriminated from benign traffic. This is compared with traditional machine learning and distributed system and evaluated against the centralized detection system.

In vehicular security in Kang et al. [67], intrusion detection using a deep learning approach has also been applied. The intrusion detection accuracy could be improved by deep belief networks (DBN) based on unsupervised pre-training. Another research of this category has been conducted by Wu et al. [68]. Authors proposed IDS;those anomalies can be detected using Auto Encoders on artificial data by IDS. Here artificial data cases are considered which will not reflect the

Table 3 Summary of the types of deep learning attacks in different layers

Layer	Types of attack	Description	References
Hardware layer	Backdoors	Main aim of this hardware attack is modifying the hardware to access sensitive information and creating a backdoor	Tehranipoor et al. [51]
Network layer	Denial of service (DoS) attack	A DoS attacks are performed by several computers at the same time. They bombard the network with so much traffic that no real networks traffic can get through and the system shuts down,	Schweitzer et al. [45]
Network layer	Distributed denial of service (DDoS) attack	A DDoS attack is also an attack on system's resources, but it is launched from a large number of other host machines that are infected by malicious software controlled by the attacker.	Zolotukhin et al. [62] and Kim et al. [52]
Application layer	Phishing	Phishing attack is the practice of sending emails that appear to be from trusted sources goal of gaining personal information. It could involve an attachment to an email that loads malware onto your computer and also be a link to an illegitimate website that can trick you into downloading malware or handing over your personal information.	Fette et al. [56]
Application layer	SQL injection attack	SQL injection is an attack method that allows hackers to run malicious code on a SQL server, allowing them to steal or delete data stored on that server.	Uwagbole et al. [61]
Application layer	Cross-site scripting (XSS) attack	Cross-site scripting (XSS) attack injects malicious code into a website which targets the visitor's Brower.	Krombholz et al. [58]
Network layer	Man in the middle attacks	Man in the middle is an attack method hackers insert themselves between your computer and the web server	Desmedt [15]
Application layer	Malware	Malware refers to several forms of destructive software, such as viruses and ransom ware. Once computer is infected, malware can take control of machine to monitor activity & send it to the attacker's home server.	Lanzi et al. [59]
Application layer	Mobile malware	Mobile Malware is specifically designed to target the operating system on mobile devices, allowing hackers to steal data on the device.	Zhu et al. [63] and Ding et al. [60]

malicious and normal behavior of real time networks. They also adopted a centralized approach which is impractical for distributed applications such as social internet of things in smart city networks.

For malicious code detection by using Auto Encoders for feature extraction and Deep Belief Networks (DBN) as a classifier for detection the deep learning approach are applied [69]. The accuracy and timing efficiency is higher in hybrid mechanism than in sing DBN. This research does not handle the distributed training and sharing of updated parameters. The other paper which has investigated deep learning scheme for malicious code detection is Wang et al. [17]. It has applied demonizing Auto Encoder for deeper features learning to identify malicious java script code from normal code. The resulting accuracy is good in the best case scenario. This approach can be hardly applied to distributedIoT/Fog systems. Using deep learning approach our model enables parallel training and parameters sharing by local

fog nodes and detects network attacks in distributed fog-to-things networks.

Al-Qurishi et al. [70] Sybil attacks are increasing in social networks, Sybil accounts are emerging a lot. The operators of these accounts are working with new techniques in order to avoid detection of their Sybil attacks. The existing Sybil detection techniques are not much useful in preventing and controlling attacks. This must be overcome by updating the existing detection techniques with new data and well developed strategies. A prediction system with the implementation of deep learning solution model has three integrated modules, they are, a data harvesting module, a feature extracting mechanism and a deep regression model, this system can solve problems of Sybil attacks. This model evaluates the data fed to it, which is optimized.

Chen et al. [71] smartphone security has been threatened due to the arrival of mobile malware. Attackers are polluting the training data which shows the ineffectiveness of existing machine learning malware detection tools. This problem can be solved by KUAFUDET learns mobile malware using adversarial detection and it is scalable. KUAFUDET has two phases, an offline training phase and online detection phase. The offline training phase gets data from training set. The online phase uses the classifier trained by the offline training phase. These two fields are interlinked together via a self-adaptive learning scheme to address the adversarial environment. An automated camouflage detector is used to remove the false negatives and send them back to offline training phase.

Pachauria et al. [72] medical wireless sensor networks suffer a lot from a wide range of faults and anomalies. To avoid this, many technologies have been developed but they are not up to the level, so that experiments are being made using machine learning algorithms. It experiments on real time medical data and detects sensor faults in a fast and accurate manner.

Rehman et al. [73] The key factor in security of smart phones is the detection of malware detection . However signature based method used now, that not provide accurate information of zero day attacks and many dimensional viruses. For this hybrid framework work is given has the solution of malware detection. Upcoming method uses the both signature and heuristic based detection. The authors have used machine learning algorithm for extracting files and binaries, for this huge amount of research has been done and finally they that found signature and heuristic has improved accuracy.

Hai et al. [74], industrial anti-virus tools are detecting malware existence using signature based techniques. Malware like metamorphic or polymorphic virus uses some techniques such as mutation and dynamically executed contents (DEC) methods in order to escape from detection tools. Packing or calling external code is a DEC method used by malware programs. New techniques are arising to detect suspicious behavior from various mutated samples of virus; one of them is Control Flow graph (CFG). CFG form such as IDA pro do not reflect the behaviors of DEC methods, they are generated by binary analysis tools. This approach is costly to analyses CFG's from binaries. This disadvantage makes this approach not suitable for real-world applications. An improvised form of CFG which reflects the DEC behaviors is named as lazy-binding CFG. Deep learning technology well plays with image recognition on huge dataset. In collaboration with deep learning technology, the lazy-binding CFG performs image-based representation. This technique is applied for malware detection on real-world applications with higher accuracy.

Pajouh et al. [75] Internet of things devices are being employed in different fields, for different needs. IoT devices have high capabilities and improved applications so that they are a great target for attacks and malware. IoT malware can be detected using Recurrent Neural network (RNN) deep learning. RNN can be used to analyze ARM-based IoT application's opcodes. With three various long short- term memory (LSTM) configurations the trained model is evaluated and the LSTM approach gives the best result.

Rav et al. [76] *demonstrated* the gaining familiarity of wearable devices in today's world in fields such as sports, wellbeing and healthcare. Since these applications require large analysis and classification, deep learning techniques are preferred. Since these techniques require high computing environment they show low efficiency in wearable devices. So the authors have combined the inertial sensor and data together for accurate and real time classification. To overcome the limitation in deep learning the authors proposed on-node computation. To optimize the upcoming method with on node sensor spectral domain pre-processing is required for data before sending it. The accuracy in classification of upcoming deep learning method is against the state of art methods that using both real time and laboratory. They also proved that the method with on-node sensor is consistent for the smart phones and wearable devices.

He et al. [77], the application of computer and intelligence of communication has shown that increase in monitoring and control of smart grids. The dependence in the field of IT increases the vulnerability of attacks. A severe threat to supervisory control and data acquisition (SCADA) is due to attack of integrity of data in false data injection (FDI). In this paper deep learning technique is used to recognize behavior of FDI attacks with measurements of data and capture the FDI attacks. By doing so, our upcoming detection mechanism effectively relaxes the potential attacks and increases the accuracy. The performance of the upcoming strategy through IEEE 118-bus test system has been illustrated. The scalability by using IEEE-300 bus test system has also been measured.

Hasana et al. [78] the Optical Burst Switching (OBS) network is majorly affected by the Denial of service attack which is also known as Burst header packet flooding attack. In this attack, without any prior acknowledgement about Data Burst the malicious BHP is flood into the network, which in turn leads to poor network performance, data loss and DOS, low utilization of bandwidth. The machine predicted analysis works effectively identifying these attacks in OBS network, but this is not suitable for traditional machine learning approaches like Naïve Bayes, K-Nearest Neighbor's. To overcome this Deep Convolution Neural Network Model is introduced which detects the attack at very early stage.

Liu et al. [79] deep learning has been widely used in network attack detection problems. Deep learning models are used to analyses the payload in turn gives a convolutional neural network based payload classification approach (PL-CNN) and also recurrent neural network based payload classification approach (PL-RNN). These approaches are efficient and practically possible in attack detection.

Dong et al. [80] many traditional machine learning methods are employed in the intrusion detection, but the traditional learning methods are poor in detection performance and accuracy. To get situation assessment of network the intrusion detection got data from monitoring security events. By introducing deep learning models the performance and accuracy rate can be improved.

Loukas et al. [3, 81] One of the growing detections is detection of cyber-attacks in vehicles. Vehicles support limited and light weight processing resources. At the same time, attacks against vehicles are difficult, often making intrusion detection systems less practical than behaviour-based ones, which is opposite to the conventional computing systems. This technique can be improved with computational offloading which is used resource constrained mobile devices. The real time data is taken and which is given as input to cyber and physical processes, which send data as time series to neural network architecture. This is more reliable and the detection is accurate.

Al-Hawawreh et al. [21] internet industrial control systems, they connect the technical part of applications with physical processors, now they are being threatened by cyber-attacks. This can be overcome by using deep learning models; it has deep auto-encoder and deep feed forward neural network architecture, which are derived from datasets like NSL-KDD AND UNSW-NB15. These models learn and validate data collected from TCP/IP packets. It is a powerful technique.

Yin et al. [37] intrusion detection plays a vital role in information security and to identify various attacks in the network. The Recurrent neural networks (RNN-IDS) are well suited for making a classification model with great accuracy and improved performing ability. By using RNN-IDS the accuracy and performance rate of intrusion detection is improved.

Tang et al. [82], software defined networking (SDN) improves the network security with logical centralized controllers and global network overview. In flow-based anomaly detection SDN environment a deep learning approach is applied. The Deep Neural Network model for intrusion detection is trained with NSL-KDD dataset for improved performance.

Feng et al. [83], proposed the use of an inbuilt play gadget to identify Denial of Service and security attacks. Stay away from the identified assault to spreading out in bigger scope. In the examination, profound LSTM and neural system (DNN) discovery model are used to identify DoS assaults. Later the authors used CNN discovery model to identify XSS and SQL assaults. They have demonstrated that these discovery models accomplished extremely high accuracy, recall, precision and F1−score. Furthermore efficiency with respect to time factor among LSTM,CNN and DNN is in undesirable range. They have concluded that their proposed technique can further be applied for the detection of attacks in ad-hoc networks with little bit modification.

Shenfield et al. [84] this methodology gives to distinguish malignant system traffic utilizing artificial neural systems reasonable for use in profound bundle investigation based interruption identification systems. Down to earth results utilizing a scope of run of the mill arrange traffic information (pictures, dynamic connection library documents, and a determination of different records, for example, logs, music records, and word preparing documents) and malevolent shell code records sourced from the online endeavor and weakness repository endeavor dbB.

Jan [65], has demonstrated that the proposed the ANN design can distinguish among kind and noxious system traffic accurately. The proposed ANN architecture gets a medium exactness of 98%, a normal region under the recipient administrator characteristic bend of 0.98, and a normal bogus positive pace of under 2% in rehashed 10- fold cross-approval. This shows the grouping strategy is powerful, exact, and precise. The tale way to deal with hazardous system traffic location proposed in this paper has the efficiency to enhance the intrusion detection which improves both conventional and cyber physical attacks.

Li et al. [85] Hybrid malicious conspire is dependent on Auto Encoder and DBN (Deep Belief Networks). The Auto Encoder is utilized in profound learning technique to decrease the information dimensionality. This could change over exceptionally troublesome high-dimensional information into low dimensional codes with the nonlinear planning, DBN is composed of multilayer RBM and a layer of BP neural system. Each layer of RBM is based on unaided preparing where the yield vector is set. In the wake of giving

Table 4 Summary of the various attacks, algorithms and solutions

References	Problem/attack	Algorithm/method	Security solution
Muhammad Al-Qurishi et al. [70]	Sybil attack	DL solution model	Solves the problem of Sybil attacks on Twitter
Sen Chen et al. [71]	Poisoning attack	KUAFUDET-2 phase learning approach	Remove the false negatives and send them back to offline training phase
Pachauria et al. [72]	Fault detection mechanisms	J48, Random Forests, kNN, Additive Regression	Detecting sensor faults quickly, accurately with a low false alarm ratio.
Rehman et al. [73]	Detection of zero-day attacks and polymorphic viruses.	SVM, Decision Tree, W-J48 and KNN.	Helps in improving accuracy in malware detection
Hai et al. [74]	Detecting malware using DL	CNN, lazy-binding strategy for the CFG generation	Malware detection on real-world computer programs with very high accuracy
Pajouh et al. [75]	ARM-based IoT applications' execution operation Codes	RNN	High accuracy in the detection of new malware samples
Ravi et al. [76]	accurate and real-time activity classification	Classification	Providing shallow and learnt features from a deep learning approach for time-series data classification
He et al. [77]	False Data Injection	State Vector Estimator (SVE), Deep-Learning Based Identification	Overcoming FDI attack that compromises the limited number of state measurements of the power system for electricity theft
Hasana et al. [78]	DOS attack	Deep CNN model	Automatic detection of the edge nodes at an early stage
Liu et al. [79]	Payload attacks	CNN-based payload classification approach and RNN-based payload classification approach	Helps the network analysts to understand the abnormal traffic packets
Dong et al. [80]	IDS	DL	Helps in classifying the network traffic
Loukas et al. [1]	DoS, command injection and malware	Deep multilayer perceptron and RNN	Reduction in detection latency achieved through offloading
Muna Al-Hawawreh et al. [21]	NIDS	Deepfeed forward neural network	Provides high detection rate and lower false positive rate
Yin et al. [37]	IDS	RNN	Provides a higher accuracy rate and detection rate with a low false positive rate
Tang et al. [82]	IDS	DNN	Provides flow-based anomaly detection in SDN environments.
Feng et al. [83]	DoS and privacy attacks	DNN, CNN LSTM, XSS and SQL attacks	Helps in providing the attack Detection models with achieve very high *Accuracy, Precision, Recall*
Shenfield et al. [84]	IDS	Artificial neural networks	Provides an efficient IDS for cyber–physical systems such as smart-grids
Li et al. [85]	Malicious Code Detection	Boltzmann Machines, Neural Network	Helps in reducing the time complexity and has better detection performance
Shone et al. [87]	NIDS	DL classification model	Helps in evaluating the benchmark KDD Cup '99 and NSL-KDD datasets

the input tests into the half breed model, the pragmatic outcomes show that the location accuracy got by the half and half location technique proposed is higher than that of single DBN. This proposed strategy decreases the time intricacy and has better discovery execution.

Niyaz et al. [86] The network IDS helps the admin to distinguish security penetrates in their associations. There exists numerous difficulties while building up a flexible and potential NIDS for unanticipated and eccentric attacks. The authors gave a profound learning based technique for growing such an efficient and flexible NIDS. They have utilized Self-showed Learning which is a profound learning based procedure, for NSL- KDD - a benchmark dataset for arrange interruption.

The summary of the various attacks, algorithms and solutions is given in Table 4.

6 Conclusion

Machine Learning and Deep learning techniques are inspired by the ability of human brain in learning from previous experience instantaneously. These techniques have been widely adopted by various areas of research for providing solutions for their problems. Cyber security is gaining insight nowadays with the increased internet usage and wide variety of network applications. In this paper we have presented a literature review of the different ML/DL methods for cyber security attacks. Recent tools and platforms for DL and ML are focused and the security solutions to the different categories of attacks are discussed and summarized.

Compliance with Ethical Standards

Conflict of interest On behalf of all authors, the corresponding author states that there is no conflict of interest.

References

1. Loukas G, Vuong T, Heartfield R, Sakellari G, Yoon Y, Gan D (2018) Cloud-based cyber-physical intrusion detection for vehicles using deep learning, security analytics and intelligence for cyber physical systems. IEEE Access 6:3491–3508. https://doi.org/10.1109/ACCESS.2017.2782159
2. Toch E, Bettini C, Shmueli E, Radaelli L (2018) The privacy implications of cyber security systems: a technological survey. ACM Comput Surv. https://doi.org/10.1145/3172869
3. Loukas G, Vuong T, Heartfield R, Sakellari G, Yoon Y, Gan D (2018) Cloud-based cyber-physical intrusion detection for vehicles using deep learning. Spec Sect Secur Anal Intell Cyber Phys Syst 6:2169–3536
4. Koscher K (2010) Experimental security analysis of a modern automobile. In: Proceedings of IEEE Security Privacy, May 2010, pp 447–462
5. Checkowayet S (2011) Comprehensive experimental analyses of automotive attack surfaces. In: Proceedings of Usenix security symposium, p 6
6. Ward D, Ibarra I, Ruddle A (2013) Threat analysis and risk assessment in automotive cyber security. Int. J. Passeng Cars 6(2):507–513
7. McGraw G (2013) Cyber war is inevitable (unless we build security in). J Strateg Stud 36(1):109–119
8. Lala C, Panda B (2001) Evaluating damage from cyber attacks: a model and analysis. IEEE Trans Syst Man Cybern Part A Syst Hum 31:300–310
9. Cristalli S, Pagnozzi M, Graziano M, Lanzi A, Balzarotti D (2016) Micro-virtualization memory tracing to detect and prevent spraying attacks. In: Proceedings of the 25th USENIX security symposium, pp 431–446
10. Hatcher WG, Yu W (2018) Survey of deep learning: platforms. Appl Emerg Res Trends 6:2169–3536
11. Bonarini A, Lazaric A, Montrone F, Restelli M (2009) Reinforcement distribution in fuzzy Q-learning. Fuzzy Sets Syst Spec Issue Fuzzy Sets Interdiscip Percept Intell 160(10):1420–1443
12. Ge L, Zhang H, Xu G, Yu W, Chen C, Blasch EP (2015) Towards map reduce based machine learning techniques for processing massive network threat monitoring data. Networking for Big Data, published by CRC Press & Francis Group, USA
13. Huang HH, Liu H (2014) Big data machine learning and graph analytics: Current state and future challenges. In: 2014 IEEE international conference on big data (Big Data), pp 16–17
14. Yu W, Ge L, Xu GG, Fu X (2014) Towards neural network based malware detection on android mobile devices. In: Pino R, Kott A, Shevenell M (eds) Cybersecurity systems for human cognition augmentation, vol 61. Advances in information security. Springer, Cham. https://doi.org/10.1007/978-3-319-10374-7_7
15. Desmedt Y (2011) Man-in-the-middle attack. In: van Tilborg HCA, Jajodia S (eds) Encyclopedia of Cryptography and Security. Springer, Boston, MA. https://doi.org/10.1007/978-1-4419-5906-5_324
16. Paul M (2017) Multiclass and Multi-Label Classi_cation. [Online]. http://cmci.colorado.edu/classes/INFO-4604/_les/slides-7_multi.pdf
17. Wang Y, Cai W, Wei P (2016) A deep learning approach for detecting malicious JavaScript code. Secur Commun Netw 9:1520–1534. https://doi.org/10.1002/sec.1441
18. Su B, Ding X, Wang H, Wu Y (2018) Discriminative dimensionality reduction for multi-dimensional sequences. IEEE Trans Pattern Anal Mach Intell 40(1):77–91
19. Marquardt D, Doclo S (2017) Noise power spectral density estimation for binaural noise reduction exploiting direction of arrival estimates. In: Proceedings of IEEE workshop on applications of signal processing to audio and acoustics, pp 234–238
20. Xin Y, Kong L, Liu Z (2018) Machine learning and deep learning methods for cyber security. IEEE 6:2169–3536
21. Al-Hawawreh M, Moustafa N, Sitnikova E (2018) Identification of malicious activities in industrial internet of things based on deep learning models. J Inf Secur Appl 41:1–11. https://doi.org/10.1016/j.jisa.2018.05.002
22. Yang Q, An D, Min R, Yu W, Yang X, Zhao W (2017) Optimal PMU placement based defense against data integrity attacks in smart grid. IEEE Trans Forens Inf Secur (T-IFS) 12(7):1735–1750
23. Yang X, Ren X, Lin J, Yu W (2016) On binary decomposition based privacy-preserving aggregation schemes in

real-time monitoring systems. IEEE Trans Parallel Distrib Syst 27(10):2967–2983
24. Sharma RK, Kalita HK, Borah P (2016) Analysis of machine learning techniques based intrusion detection systems. In: Proceedings of international conference on advanced computing networking and informatics, pp 485–493
25. Saxena H, Richariya V (2014) Intrusion detection in KDD99 dataset using SVM-PSO and feature reduction with information gain. Int J Comput Appl 98(6):25–29
26. Rao KS (2017) Fast kNN classifiers for network intrusion detection system. Indian J Sci Technol 10(14):1–10
27. Vishwakarma S, Sharma V, Tiwari A (2017) An intrusion detection system using KNN-ACO algorithm. Int J Comput Appl 171(10):18–23
28. Umarani Srikanth G, Geetha R (2018) Task scheduling using Ant Colony Optimization in multicore architectures: a survey. Soft Computing. 22:5179–5196
29. Kwon D, Kim H, Kim J, Suh SC, Kim I, Kim KJ (2017) A survey of deep learning-based network anomaly detection. Clust Comput 4(3):1–13
30. Ding Y, Chen S, Xu J (2016) Application of deep belief networks for opcode based malware detection. In: Proceedings of international joint conference on neural networks, pp 3901–3908
31. Nadeem M, Marshall O, Singh S, Fang X, Yuan X (2016) Semi supervised deep neural network for network intrusion detection. In: Proceedings of the KSU conference on cybersecurity, education, research and practice, pp 1–13
32. Gao N, Gao L, Gao Q, Wang H (2014) An intrusion detection model based on deep belief networks. In: Proceedings of 2nd international conference on advanced cloud big data, pp. 247–252
33. Zhao G, Zhang C, Zheng L (2017) Intrusion detection using deep belief network and probabilistic neural network. In: Proceedings of IEEE international conference on computer science and engineering, vol 1, pp 639–642
34. Alrawashdeh K, Purdy C (2017) Toward an online anomaly intrusion detection system based on deep learning. In: Proceedings IEEE international conference on machine learning and applications, pp 95–200
35. Alom MZ, Bontupalli VR, Taha TM (2016) Intrusion detection using deep belief networks. In: Proceedings of national aerospace and electronics conference, pp 339–344
36. Tan Q, Huang W, Li Q (2016) An intrusion detection method based on DBN in ad hoc networks. In: Proceedings of the international conference on wireless communication and sensor network, pp. 477–485
37. Yin CL, Zhu YF, Fei JL, He XZ (2017) A deep learning approach for intrusion detection using recurrent neural networks. IEEE Access 5:21954–21961
38. Staudemeyer RC (2015) Applying long short-term memory recurrent neural networks to intrusion detection. S Afr Comput J 56(1):136–154
39. Bu SJ, Cho BS (2017) A hybrid system of deep learning and learning classifier system for database intrusion detection. In: Hybrid artificial intelligent systems, pp. 615–625
40. Wang W, Zhu M, Zeng X, Ye X, Sheng Y (2017) Malware traffic classification using convolutional neural network for representation learning. In: Proceedings of the international conference on information networking, pp 712–717
41. Shi S, Wang Q, Xu P, Chu X (2016) Benchmarking state-of-the-art deep learning software tools. [Online]. https://arxiv.org/abs/1608.07249
42. (2017) Theano. [Online]. http://deeplearning.net/software/theano/
43. (2017) Torch: a scientific computing framework for LuaJIT. [Online]. http://torch.ch/
44. (2017) The Microsoft cognitive toolkit. [Online]. Available:https://docs.microsoft.com/en-us/cognitive-toolkit/
45. Schweitzer N, Stulman A, Shabtai A, Margalit RD (2016) Mitigating denial of service attacks in OLSR protocol using fictitious nodes. IEEE Trans Mob Comput 15:163–172
46. (2017) Caffe2: a new lightweight, modular, and scalable deep learning framework. [Online]. https://caffe2.ai/
47. Jia Y et al. (2014). Caffe: convolutional architecture for fast feature embedding. [Online]. https://arxiv.org/abs/1408.5093
48. (2017) Caffe. [Online]. http://caffe.berkeleyvision.org/
49. (2017) Apache MXNet: a flexible and efficient library for deep learning. [Online]. https://mxnet.apache.org/
50. (2017) Keras: the Python deep learning library. [Online]. https://keras.io/
51. Tehranipoor M, Koushanfar F (2010) A survey of hardware Trojan taxonomy and detection. IEEE Des Test Comput 27:10–25
52. Zolotukhin M, Hämäläinen T, Kokkonen T, Siltanen J (2016) Increasing web service availability by detecting application-layer DDoSattacks in encrypted traffic. In: 2016 23rd International conference on telecommunications (ICT), pp 1–6
53. Kim J, Kim J, Thu T, Kim H (2016) Long short term memory recurrent neural network classifier for intrusion detection. In: Proceedings of international conference on platform technology and service (PlatCon), pp 1–5
54. Wang S, Shang Y, Wang J, Mei L, Hu C (2015) Deep features for person re-identification. In: 2015 11th International conference on semantics, knowledge and grids (SKG), pp 244–247
55. Niimi A (2015) Deep learning for credit card data analysis. In: 2015 World congress on internet security (WorldCIS), pp 73–77
56. Fette I, Sadeh N, Tomasic A (2007) Learning to detect phishing emails. In: Proceedings of the 16th international conference on world wide web. ACM, pp 649–656
57. Ma J, Saul LK, Savage S, Voelker GM (2009) Beyond blacklists: learning to detect malicious web sites from suspicious URLs. In: Proceedings of the 15th ACMSIGKDD international conference on knowledge discovery and data mining. ACM, pp 1245–1254
58. Krombholz K, Hobel H, Huber M, Weippl. E (2015) Advanced social engineering attacks. J Inf Secur Appl 22:113–122
59. Lanzi A, Balzarotti D, Kruegel C, Christodorescu M, Kirda E (2010) AccessMiner: using system-centric models for malware protection. In: Proceedings of the 17th ACM conference on computer and communications security, pp 399–412
60. Zhu D, Jin H, Yang Y, Wu D, Chen W (2017) Deep flow: deep learning-based malware detection by mining Android application for abnormal usage of sensitive data. In: Proceedings of IEEE symposium on computers and communications (ISCC), pp 438–443
61. Ding Y, Chen S, Xu J (2016) Application of deep belief networks for opcode based malware detection. In: Proceedings of international joint conference on neural networks (IJCNN), pp 3901–3908
62. Uwagbole SO, Buchanan WJ, Fan L (2016) Numerical encoding to tame SQL injection attacks. In: Proceedings of NOMS 2016—2016 IEEE/IFIP network operations and management symposium, pp 1253–1256
63. Yu W, Zhang H, Ge L, Hardy R (2013) On behavior-based detection of malware on android platform. In: 2013 IEEE global communications conference (GLOBECOM), pp 814–819
64. Pierazzi F, Apruzzese G, Colajanni M, Guido A, Marchetti M (2017) Scalable architecture for online prioritization of cyber threats. In: International conference on cyber conflict (CyCon)
65. Jan CB (2017) Deep learning in big data analytics: a comparative study. Comput Electr Eng. https://doi.org/10.1016/j.compeleceng.2017.12.009
66. Chilamkurti N, Diro AA (2017) Distributed attack detection scheme using deep learning approach for Internet of Things.

Future Gener Comput Syst. https://doi.org/10.1016/j.future.2017.08.043
67. Kang M-J, Kang J-W (2016) Intrusion detection system using deep neural network for in-vehicle network security. PLoS ONE 11(6):e0155781. https://doi.org/10.1371/journal.pone.0155781
68. Wu C, Guo Y, Ma Y (2015) Adaptive anomalies detection with deep network. In: The seventh international conference on advanced 2015 cognitive technologies and applications, IARIA, pp 181–186
69. Li Y, Maand R, Jiao R (2015) A hybrid malicious code detection method based on deep learning. SERSC Int J Secur Appl 9:205–216. https://doi.org/10.14257/ijsia.2015.9.5.21
70. Al-Qurishi M, Alrubaian M, Rahman SMM, Alamri A, Hassan MM (2017) A prediction system of Sybil attack in social network using deep-regression model. Future Gener Comput Syst 87:743–753. https://doi.org/10.1016/j.future.2017.08.030
71. Chen S, Xue M, Fan L, Hao S, Xu L, Zhu H, Li B (2017) Automated poisoning attacks and defenses in malware detection systems: an adversarial machine learning approach. Comput Secur 73:326–344. https://doi.org/10.1016/j.cose.2017.11.007
72. Pachauria G, Sharma S (2015) Anomaly detection in medical wireless sensor networks using machine learning algorithms. In: Proceedings of 4th international conference on eco-friendly computing and communication systems, Published by Elsevier B.V. Peer-review under responsibility of organizing committee
73. Rehman UZ (2017) Machine learning-assisted signature and heuristic-based detection of malwares in Android devices. Comput Electr Eng 69:828–841. https://doi.org/10.1016/j.compeleceng.2017.11.028
74. Hai NM, Dung LN, Mao NX, Tho QT (2018) Auto-detection of sophisticated malware using lazy-binding control flow graph and deep learning. Comput Secur. https://doi.org/10.1016/j.cose.2018.02.006
75. HaddadPajouh H, Dehghantanha A, Khayami R, Choo KR (2018) A deep recurrent neural network based approach for internet of things malware threat hunting. Future Gener Comput Syst 85:88–96. https://doi.org/10.1016/j.future.2018.03.007
76. Rav D, Wong C, Lo B, Yang G-Z (2017) A deep learning approach to on-node sensor data analytics for mobile or wearable devices. IEEE J Biomed Health Inform 21(1):56–64
77. He Y, Mendis GJ, Wei J (2016) Real-time detection of false data injection attacks in smart grid: a deep learning-based intelligent mechanism. IEEE Trans Smart Grid. https://doi.org/10.1109/tsg.2017.270384
78. Hasana MZ, Hasanb KMZ, Sattar A (2018) Burst header packet flood detection in optical burst switching network using deep learning model. Procedia Comput Sci 143:970–977. https://doi.org/10.1016/j.procs.2018.10.337
79. Liu H, Lang B, Liu M, Yan H (2018) CNN and RNN based payload classification methods for attack detection. Knowl Based Syst 163:332–341. https://doi.org/10.1016/j.knosys.2018.08.036
80. Dong B, Wang X, (2016).Comparison deep learning method to traditional methods using for network intrusion detection. In: 8th IEEE international conference on communication software and networks. https://doi.org/10.1109/iccsn.2016.7586590
81. Loukas G (2015) Cyber-physical attacks: a growing invisible threat. Butterworth-Heinemann, Oxford
82. Tang TA, LotfiMhamdi DM, Raza Zaidi SA, Ghogho, M (2016) Deep learning approach for network intrusion detection in software defined networking. Int Conf Wirel Netw Mob Commun. https://doi.org/10.1109/WINCOM.2016.7777224
83. Feng F, Liu X, Yong B, Zhou R, Zhou Q (2018) Anomaly detection in ad-hoc networks based on deep learning model: a plug and play device. J LATEX Templates Ad Hoc Netw 84:82–89. https://doi.org/10.1016/j.adhoc.2018.09.014
84. Shenfield A, Day D, Ayesh A (2018) Intelligent intrusion detection systems using artificial neural networks. Korean Inst Commun Inf Sci 2:95–99. https://doi.org/10.1016/j.icte.2018.04.003
85. Li Y, Ma R, Jiao R (2015) A hybrid malicious code detection method based on deep learning. Int J Secur Appl 9:205–216
86. Niyaz Q, Sun W, Javaid AY, Alam M (2015) A deep learning approach for network intrusion detection system. BICT 2015:03–05
87. Shone N, Ngoc TN, Phai VD, Shi Q (2018) A deep learning approach to network intrusion detection. IEEE Trans Emerg Top Comput Intell 2:41–50. https://doi.org/10.1109/TETCI.2017.2772792
88. Hatcher WG, Yu W (2018) A survey of deep learning: platforms, applications and emerging research trends. IEEE Access 6:2169–3536