Deep Learning in Intrusion Detection System: An Overview

Muhamad Erza Aminanto[a,*], Kwangjo Kim[b,*]

[*] School of Computing, KAIST, Korea

[a] E-mail address: aminanto@kaist.ac.kr

[b] E-mail address: kkj@kaist.ac.kr

Abstract

Identifying unknown attacks is one of the big challenges in network Intrusion Detection Systems (IDSs) research. In the past decades, researchers adopted various machine learning approaches to classify and distinguish anomaly traffic from benign traffic without prior knowledge on the attack signature. Extensive academic research on machine learning made a significant breakthrough in mimicking human brain recently. The state-of-the-art on machine learning breakthrough comes from deep learning which has been predicted to cause a powerful improvement in artificial intelligence field. Numerous complex applications have been accomplished by deep learning. One of the distinguished applications is AlphaGo from Google that uses Convolutional Neural Network. AlphaGo beat the Korean world champion in the "Go" game recently by showing superman-like capabilities in remote machine learning. The advancements on this learning algorithms may improve IDS ability to reach high detection rate and low false alarm rate. However, the deep learning implementations in intrusion detection applications may have some limitations. In this paper, we survey previous IDSs that embrace deep learning approaches. Deep learning methods such as deep belief network, restricted Boltzman machine, deep Boltzman machine, deep neural network, auto encoder, *etc.*, are commonly used in IDSs. We examine such deep learning methods with their advantages and disadvantages in order to get better understanding on how to apply deep learning. We realize that there is a confusion of how to adopt deep learning in IDS application properly. Our claim is that deep learning is useful in IDS, especially for feature extraction. In order to support our claim, we provide future challenges and directions to employ deep learning in IDS accordingly. Finally, deep learning methods can enhance future research on unknown attack detection.

Keywords: anomaly detection, neural network, deep learning, feature extraction.

1. Introduction

There are a numerous different type of attacks within cyberspace these days. Comprehensive researches have been executed in order to overcome these attacks. One common countermeasure is to use so called Intrusion Detection System (IDS). An extensive research applying machine learning methods in IDS have been done in both academia and industry. However, the security experts still desire better performance IDS which has highest detection

rate and lowest false alarm rate. In addition, overall threat analysis is expected in order to secure their networks [ZuKW15]. Improvements in IDS could be achieved by embracing the latest breakthrough in machine learning [SoPa10], called deep learning.

In recent years, deep artificial neural networks, so called deep learning, have won numerous contests in pattern recognition and machine learning [Schm15]. Deep learning belongs to a class of machine learning methods, where employs consecutive layers of information-processing stages in hierarchical manners for pattern classification and feature or representation learning [Deng14]. According to [DeYu14], there are three important reasons for the deep learning prominent recently. First, processing abilities (*e.g.*, GPU units) increased sharply. Second, computing hardware getting affordable, and last, recent breakthrough in machine learning research. Shallow and Deep Learners are distinguished by the depth of their credit assignment paths, which are chains of possibly learnable, causal links between actions and effects. Usually deep learning plays the important role in image classification results [Bene16]. In addition, deep learning is also commonly used for language, graphical modeling, pattern recognition, speech, audio, image, video, natural language and signal processing [Deng14]. There are many deep learning methods such as Deep Belief Network (DBN) [HiOT06], Boltzman Machine (BM), Restricted Boltzman Machine (RBM), Deep Boltzman Machine (DBM) [SaHi09], Deep Neural Network (DNN), Auto Encoder, Deep / stacked Auto Encoder [BeLa07], Stacked denoising Auto Encoder [VLLB10], Distributed representation [Deng14] and Convolutional Neural Network (CNN) [LBBH98]. Recently, one of the distinguished applications is AlphaGo[SHMG16] from Google that uses CNN. AlphaGo beat the Korean world champion in the “Go” game recently by showing superman-like capabilities in remote machine learning.

The advancements on learning algorithms might improve IDS ability to reach higher detection rate and lower false alarm rate. However, the deep learning implementations in intrusion detection applications have some limitations. In this paper, we survey several previous IDSs that embrace deep learning approaches. We examine such deep learning methods with their advantages and disadvantages in order to get better understanding on how to apply deep learning.

We realize that there is a confusion of how to adopt deep learning in IDS application properly since the different approaches have adopted by each previous one. Several researches use deep learning methods in partial sense only while the rest still uses conventional neural networks. The complexity of deep learning method may be one of the reasons. In addition, deep learning method requires a lot of time to train properly. However, we found that several researchers adopt deep learning method in their whole IDS. We compare the IDS performance among them. Our claim is that deep learning is very useful in IDS, especially for feature extraction. The feature extraction is a process of transforming raw data into features that better represent underlying problem to the predictive models, resulting in improved

model accuracy on unseen data. In order to support our claim, we provide future challenges and directions to employ deep learning in IDS accordingly. We conclude that deep learning method is suitable for pre-training or feature engineering/extraction, not as classifier. Finally, the deep learning methods can enhance future research on unknown attack detection.

2. Intrusion Detection System

In general, we can classify IDS into misuse detection and anomaly detection [SaBe15] . Misuse detection techniques usually utilize precise descriptions to monitor traffics. The approach often called signature-based approach. The approach is intended to identify known attack patterns only. Although misused detection techniques are most commonly used in practice [LPSR05], these methods have a significant drawback [ZaSa04]. The main drawback of misuse detection is incapable of detecting the unknown attacks since it considers known signature of attacks only. In order to maintain the performance of misuse detection, we need to keep signature of attacks updated every time which is burdensome. In addition, attackers usually combine previous attacks[ZaSa04]. This kind of attack is more difficult to develop appropriate signatures for misuse detection. On the other hand, anomaly detection focuses on detecting unusual activity patterns in the observed data[LPSR05]. Anomaly detection approach usually deals with statistical analysis and data mining techniques [TsKw06], which can detect novel attacks without prior knowledge since the classification model has the generalization ability to extract intrusion pattern and knowledge during the training phase [TsKw06].

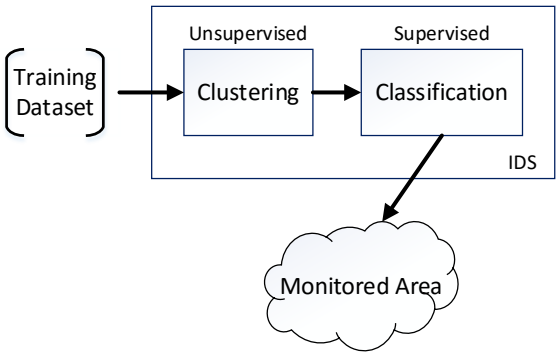

Fig. 1: Common IDS approach

Two typical methods are commonly used in IDS such as clustering and classification. It is difficult and costly to obtain bulk of labeled network connection records for supervised training in the first stage. The clustering analysis has emerged as an anomaly intrusion detection approach in recent years [TsKw06]. Clustering is an unsupervised data exploratory technique that partitions a set of unlabeled data patterns into groups or clusters such that patterns within a cluster are similar to each other but dissimilar to other clusters' pattern [TsKw06]. Meanwhile, classification is a supervised method to distinguish benign and malicious traffics based on provided data which usually comes from clustering result as

shown in Fig.1. The clustering and classification can be easily implemented by various machine learning methods.

3. Classification of Deep Learning

The term *deep learning* comes from the advancements of neural network. In deep learning, various methods have applied in order to overcome the limitations of the hidden layer. Basically, those methods employ consecutive hidden layers which hierarchically structured. Since a lot of methods belong to deep learning method, the classification of each deep learning method is essential. Deng [Deng14] differentiates deep learning into three sub-groups, generative, discriminative and hybrid. The classification is based on the intention of architectures and techniques, *e.g*., synthesis/generation or recognition/classification. The classification of the deep learning methods is shown in Fig. 2.

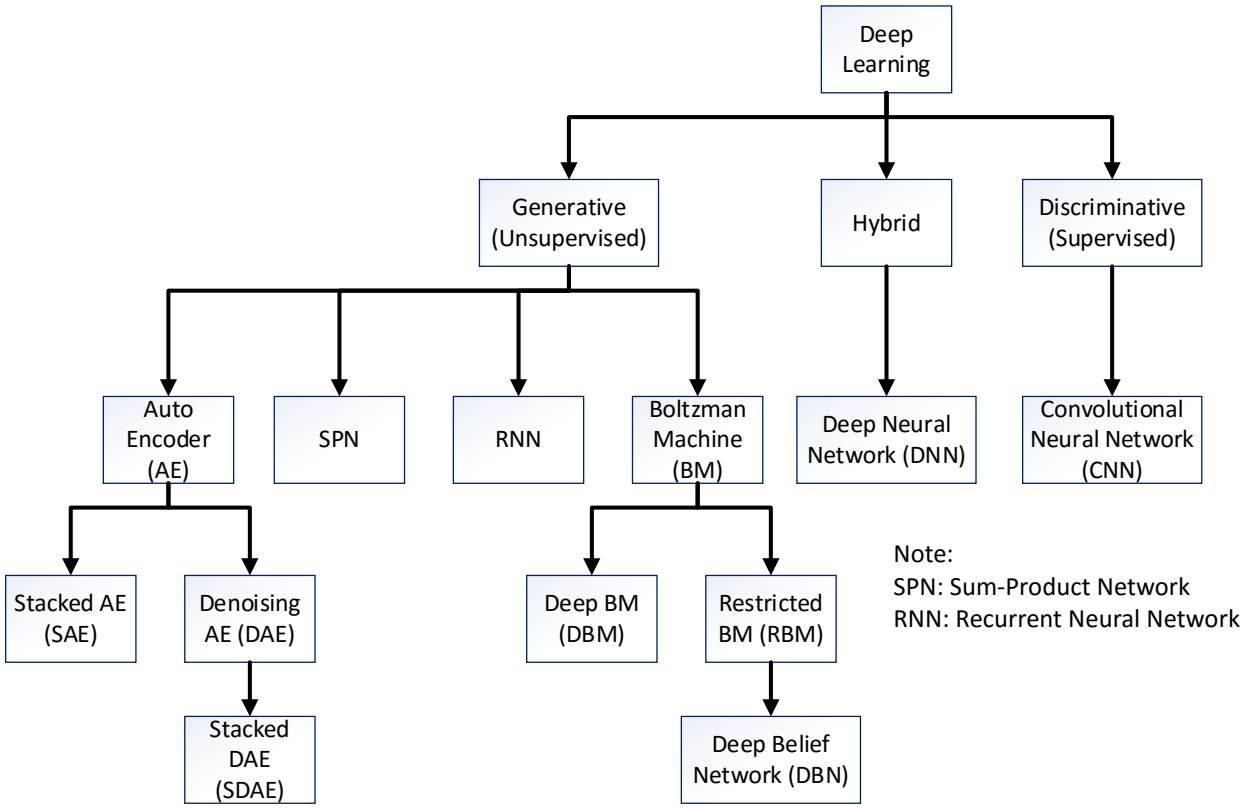

Fig. 2: Classification of deep learning methods

3.1. Unsupervised Learning

Unsupervised learning or so called generative architectures, which the data are unlabeled. The main concept of applying generative architectures to pattern recognition is unsupervised learning or pre-training [Deng14]. Since learning the lower levels of subsequent networks are difficult, deep generative architectures are needed. Thus, with limited training data, learning each lower layer in layer-by-layer approach without relying on all the layers above is important.

There are a number of methods that classified as unsupervised learning as follows:

- Auto Encoder (AE)

  Basically, AE is just similar to Artificial Neural Network (ANN) which is the number of hidden layer is three only. The main different is the nodes in the input layer are the same with the output layer. Meanwhile, the nodes in middle hidden layer are representing new features set which are low-dimensional. This architecture leads to an ability that can reconstruct the data after complicated computations. AE aims to learn a compact set of data efficiently. AE structures can be stacked to build deep networks. Each training results of the middle layer are cascaded. This structure is called Stacked Auto-Encoder (SAE) which can learn a lot of new features in different depths. In order to train more precisely, we can append an additional layer with labels once we have large amount of tagged samples [Wang15]. In addition, a Denoising Auto Encoder (DAE) is trained to reconstruct a clear correction input from a corrupted by noise input [VLLB10]. The DAE may be also stacked in order to build deep networks too.

- Boltzman Machine (BM)

  BM is a network of binary units that symmetrically paired [SaHi09]. BM has neuron units' structure that makes stochastic decisions about whether active or not [DeYu14]. If one BM result is cascaded into multiple BMs, it's called Deep BM (DBM). Meanwhile, Restricted Boltzmann machine (RBM) is a customized BM without connections among the hidden units [SaHi09]. RBM consists of visible and hidden variables such that their relations can be figured out. If multiple layers are stacked, layer-by-layer scheme, called as Deep Belief Network (DBN). DBN could be used as a feature extraction method for dimensionality reduction when unlabeled dataset and back-propagation are used (which means unsupervised training). In contrast, DBN is used for classification when appropriate labeled dataset with feature vectors are used (which means supervised training) [SERD11].

3.2. Supervised Learning

Supervised learning or discriminative deep architectures which are intended to be able to distinguish some parts of data for pattern classification [Deng14]. An example of discriminative architecture is CNN which employs a special architecture particularly suitable for image recognition. The advantage of CNN is fast to train because of its structure. CNN can train multilayer networks with gradient descent to learn complex, high-dimensional, nonlinear mappings from large collections of data[LBBH98]. CNN uses three basic concepts: local receptive fields, shared weights, and pooling[Niel15]. One extensive research that successfully deployed using CNN is AlphaGo by Google [SHMG16].

3.3. Hybrid

Hybrid deep architectures combine both generative and discriminative architectures. Basically, the hybrid architecture aims to distinguish data as well as discriminative approach. However, in the early step, it has assisted in a significant way with the generative architectures results. An example of hybrid architecture is Deep Neural Network (DNN). However, there often confusion terms between DNN and DBN. In the open literatures, DBN also uses back propagation discriminative training as a "fine-tuning". This concept of DBN is really similar to Deep Neural Network (DNN) [Deng14]. According to Deng [DeYu14], DNN is defined as a multilayer network with cascaded fully connected hidden layers, and is often use stacked RBM as a pre-training phase.

4. Applications of Deep Learning to IDS

The goal of deep learning methods is to learn feature hierarchies with features composed by lower level into higher level features [Beng09]. The methods can learn features independently at multiple levels of abstraction, and thus discover complicated functions mapping between the input to the output directly from raw data without depending on customized features by experts. In higher-level abstractions, humans often have no idea to see the relation and connection from raw sensory input. Therefore, the ability to learn complex features will become necessarily needed as the amount of data increased sharply[Beng09]. In some literatures, this capability is often called as feature extraction or feature engineering. Feature engineering is the transforming process of raw data input into features which represents the problem properly, and thus improves a model accuracy on uncovered data[YaYu14]. Since deep learning methods are impressive in feature extraction, we summarized the previous publications in Table 1 with feature extraction and classifier parameters using deep learning.

Table 1: Summary of recent IDSs using deep learning

No	Publication	Feature Extraction	Classifier
1	[SaBe15]	Normalized manually	DNN + Bayesian Calibration
2	[Wang15]	SAE	SAE/ANN
3	[YuRR15]	AE	DBN
4	[YaYu14]	SAE	ELM
5	[GGGW14]	Normalized manually	DBN
6	[JuKi15]	Normalized manually	DNN, RNN
7	[SHMG16]	MCTS	CNN
8	[FPCS13]	Normalized manually	RBM
9	[YuLX16][YLWX14]	Normalized manually	DBN
10	[WaCW16]	SDAE	Logistic Regression
11	[SERD11]	DBN	SVM
12	[SeKi16]	Normalized Manually	CNN

Note:

DNN: Deep Neural Network	DBN: Deep Belief Network	RNN: Recurrent Neural Network
SAE: Stacked Auto Encoder	AE: Auto Encoder	MCTS: Monte Carlo Tree Search
ANN: Artificial Neural Network	ELM: Extreme Learning Machine	CNN: Convolutional Neural Network
SVM: Support Vector Machine		

From Table 1, we can see that 6 out of 12 publications use deep learning methods as feature extraction method. It shows that deep learning methods are impressive in discovering sophisticated underlying structure/feature from abstract aspects. In addition, for the rest 6 publications that not use deep learning methods as a feature extraction method, they still employ deep learning to reduce the hidden layer complexity in order to get better classification results. As an example, Gao *et al.* [GGGW14] use DBN as deep neural network classifier. They show that DBN can be pre-trained in fully unsupervised learning by customizing the KDD99 dataset in order to fit as an input data for DBN method. By adding back propagation algorithm as a fine tune for DBN, they get the sophisticated result.

Wang[Wang15] shows an interesting research that employs SAE as a feature extraction for the most common dataset has been used, KDD99 dataset. He found that even though KDD99 dataset provides 41 distinct features, it's not necessarily those 41 features are important to classify benign and anomaly packets. His motivation is to find the most important features that affect the classification results significantly. If we know the important features, we can focus on this features only, also called as feature selection. Besides that, we can also use different portion/weight in different features in order to get better detection results. There are numbers of advantages from his approach. First, it reduces the researcher workload to identify the relationship in the dataset. Second, the dimensionality reduction is accomplished since features are mapped into new space. In addition, the redundant data can be filtered out.

Yan and Yu[YaYu14] leverage deep learning methods as a features generator. Actually, their research was about detecting anomaly in gas turbine combustors instead of IDS. However, they successfully show that we can use deep learning even when we just have raw data without any relationship information. They use SAE as unsupervised representation learning scheme. Then, the explicitly learned features are used as an input for classifier method, ELM. They successfully show that deep learning is able to discover sophisticated structure and features by learning raw input layer-by-layer, with higher-level features.

Fiore *et al.* [FPCS13] aim at semi-supervised anomaly detection. Semi super-vised means that they would exploit the similarity of benign traffic behavior. If one data instance doesn't belong to the benign groups, it can be suspected as an attack instance. They didn't focus on a near real-time IDS, but to reach adequate description of network traffic, and adaptive as well. They leverage discriminative RBM, which combining the descriptive power with a sharp classification ability. The discriminative RBM has the ability to express a salient input

features properly, which is generative models characteristic, with high classification accuracy. Unlike the ordinary RBM, which aims to determine a model from the given input features, the discriminative RBM seizes all possible potential variations. The main advantages of this approach are that in general this is suitable to any environment, not limited to any specific environment, and it can detects unknown anomalous traffic.

According to Wang *et al.* [WaCW16], deep models can represent the nonlinear functions; thus, deep models can achieve more accurate results on huge training data. They use denoising AE in order to cope with the common problems such as missing or noise input data. They train the AE to reconstruct the input from incomplete/corrupted input data. Basically, denoising AE would use stochastic noise in input layer. In addition, they claim that sparse random projection can reduce dimension effectively [WaCW16]. In their experiment, they are able to reduce from over 20,000 feature dimensionalities into 480 dimensionalities only. They employ logistic regression classifier as the fine tuning in the final stage. In brief, their experiment setup is Intel Xeon E5-2420 2 cores x 1.9 GHz computer with 64 GB of RAM. Their results are: 3 hidden layers took around 25 minutes, and 250 hidden layer units took around 25 minutes.

Even though it's not related with IDS, we classify the AlphaGo's [SHMG16] publication as a comparison. The reason is CNN which is used by AlphaGo, considerably as the recent and breakthrough in deep learning methods. However, to the best of our knowledge, only Seok *et al.* [SeKi16] that leverage CNN in IDS environment. AlphaGo can reduce the depth and breadth of the search tree using value network approximation and policy network sampling by CNN method. CNN adopts the convolutional approach in pixels; therefore, its suitable for image/pattern recognition. So, Seok *et al.* [SeKi16] convert a malware fingerprint into an image type, they called it "malware image", and train the CNN using that image.

5. Summary and Discussion

In summary, deep learning is a class of machine learning methods, where exploits the cascaded layers of data processing stages in hierarchical structure for unsupervised feature learning and for pattern classification. The principle of deep learning is to process hierarchical features of the provided input data, where the higher-level features are composed by lower-level features. Furthermore, the deep learning method can integrate a feature extractor and classifier into one framework which learns feature representations from unlabeled data autonomously, and thus the security experts doesn't need to craft the desired features manually[WaCW16]. Essentially, deep learning methods can discover sophisticated underlying structure/feature from abstract aspects. This abstraction ability of deep learning makes it feasible to abstract benign or malicious features among the provided data [JuKi15].

Among all deep learning methods mentioned in this paper, most of them are classified as generative architecture and CNN is the only one that belongs to discriminative architecture. It shows that deep learning methods are more suitable for feature engineering rather than classifier. Feature engineering here including feature/representation learning and feature selection[LoCL13]. The ability to model the traffic behavior from the most characterizing raw input internal dynamics is very important to shows the correlation between anomaly detection performance and the traffic model quality [PaFC13].

5.1 Methods

In dispersion through recent methods of deep learning, we highlight the most common and important properties as follows:

- SAE

  SAE composed by cascaded Auto Encoder which is the number of hidden nodes are less than the visible nodes. It encompasses that AE can compress input data into small representation and reconstruct it back. SAE is suitable for determining meta-features from complex raw input data.
- RBM

  RBM belongs to generative architecture which means it also generate new feature given raw input data. RBM is considerably flexible and salient feature generator.
- DBN

  DBN formed by subsequent of RBM processed layer-by-layer way. DBN could be used as either a feature extraction method or a classifier. In the open literatures, they composed DBN with final stage tuning such as back propagation algorithm to get better classification result. This kind of DBN scheme belongs to classifier.
- CNN

  CNN is believed to be promising deep learning methods. It has unique properties that can exploit local correlation between nearby pixels in the input data. It achieves spatially important features. Pattern and image recognition are the example of applications suitable for CNN. However, we need to figure it out how to adopt CNN in order to classify benign and anomalous traffics.

5.2 Challenges

We provide some challenges to adopt deep learning methods in IDS environment as follows:

- Training load in deep learning methods are usually huge. As an example, AlphaGo combine MCTS with deep neural network with an asynchronous multi-threaded search that executes simulations on CPU, and computes policy and value networks in parallel on GPUs. In order to do 40 search threads, they need 48 CPUs, and 8 GPUs [SHMG16].

- Incorporating deep learning methods as a real-time classifier will be really challenging. In the most previous works that leverage deep learning methods in their IDS environment, they perform the feature extraction or reducing feature dimensionalities only. However, deep learning methods still a decent method to analyze huge data.
- As mentioned before, most of deep learning methods are suitable for image and pattern recognition. Therefore, it will be a disputing argument that how to adopt deep learning methods to classify network traffic properly. One feasible way has already shown by Seok *et al.* [SeKi16] which raw malware input codes are transformed into image-type first. Afterwards, they use CNN algorithm fed by the transformed image-type malware codes.

5.3 Directions

Based on our discussion, we recommend the following points for applying deep learning in IDS environment.

- Deep learning methods are preferably used as feature extraction or reducing complex feature dimensionalities. You may use deep learning methods if you have no idea about the correlation between raw input and targeted classification output.
- You may convert your raw input into image file first before using CNN method.
- The more input data is used, the better result classification will come [Jone14].
- Combining supervised and unsupervised learning consistently provide better detection results. Therefore, leveraging SAE before using CNN may be decent choice for higher detection rate and lower false alarm rate IDS.

Acknowledgement

This work was supported by Institute for Information & communications Technology Promotion (IITP) grant funded by the Korea government(MSIP) (B0101-16-1270, Research on Communication Technology using Bio-Inspired Algorithm).

6. References

[BeLa07] BENGIO, YOSHUA ; LAMBLIN, PASCAL: Greedy layer-wise training of deep networks. In: *Advances in neural …* (2007), Nr. 1, S. 153–160 — ISBN 0262195682

[Bene16] BENENSON, RODRIGO: *Are we there yet ?* URL http://rodrigob.github.io/are_we_there_yet/build/#datasets. - abgerufen am 2016-04-21

[Beng09] BENGIO, YOSHUA: *Learning Deep Architectures for AI*. Bd. 2, 2009 — ISBN 2200000006

[Deng14] DENG, LI: A tutorial survey of architectures, algorithms, and applications for deep learning. In: *APSIPA Transactions on Signal and Information Processing* Bd. 3 (2014), Nr. January, S. e2 — ISBN 2048-7703

[DeYu14] DENG, LI ; YU, DONG: Deep Learning: Methods and Applications. In: *Foundations and Trends in Signal Processing* Bd. 7 (2014), Nr. 2013, S. 206 — ISBN 1601988141

[FPCS13] FIORE, UGO ; PALMIERI, FRANCESCO ; CASTIGLIONE, ANIELLO ; DE SANTIS, ALFREDO: Network anomaly detection with the restricted Boltzmann machine. In: *Neurocomputing* Bd. 122, Elsevier (2013), S. 13–23

[GGGW14] GAO, NI ; GAO, LING ; GAO, QUANLI ; WANG, HAI: An Intrusion Detection Model Based on Deep Belief Networks. In: *2014 Second International Conference on Advanced Cloud and Big Data* (2014), S. 247–252 — ISBN 978-1-4799-8085-7

[HiOT06] HINTON, GEOFFREY E. ; OSINDERO, SIMON ; TEH, YEE WHYE: A fast learning algorithm for deep belief nets. In: *Neural computation* Bd. 18 (2006), Nr. 7, S. 1527–54 — ISBN 0899-7667

[Jone14] JONES, NICOLA: The learning machines. In: *Nature* Bd. 505 (2014), Nr. 7482, S. 146–148 — ISBN 0028-0836

[JuKi15] JUNG, WOOKHYUN ; KIM, SANGWON: Poster : Deep Learning for Zero-day Flash Malware Detection. In: *IEEE Security* (2015), S. 2–3

[LBBH98] LECUN, YANN ; BOTTOU, LEON ; BENGIO, YOSHUA ; HAFFNER, PATRICK: Gradient Based Learning Applied to Document Recognition. In: *Proceedings of the IEEE* Bd. 86 (1998), Nr. 11, S. 2278–2324 — ISBN 0018-9219

[LoCL13] LOUVIERIS, PANOS ; CLEWLEY, NATALIE ; LIU, XIAOHUI: Effects-based feature identification for network intrusion detection. In: *Neurocomputing* Bd. 121, Elsevier (2013), S. 265–273

[LPSR05] LASKOV, PAVEL ; PATRICK, D ; SCH, CHRISTIN ; RIECK, KONRAD ; IDA, FRAUNHOFER-FIRST: Learning Intrusion Detection : Supervised or Unsupervised ? In: *Image Analysis and Processing–ICIAP* (2005), S. 50–57

[Niel15] NIELSEN, A. MICHAAEL: *Neural Networks and Deep Learning* : Determination Press, 2015

[PaFC13] PALMIERI, F. ; FIORE, U. ; CASTIGLIONE, ANIELLO: A distributed approach to network anomaly detection based on independent component analysis. In: *Concurrency Computation Practice and Experience* Bd. 26 (2013), Nr. 6, S. 1113–1129

[SaBe15] SAXE, JOSHUA ; BERLIN, KONSTANTIN: Deep Neural Network Based Malware Detection Using Two Dimensional Binary Program Features. In: *CoRR abs/1508.03096* (2015) — ISBN 9781509003198

[SaHi09] SALAKHUTDINOV, RUSLAN ; HINTON, GEOFFREY E.: Deep Boltzmann Machines. In: *Proceedings of The 12th International Conference on Artificial Intelligence and Statics* (2009), Nr. 3, S. 448–455 — ISBN 9781424439911

[Schm15] SCHMIDHUBER, JÜRGEN: Deep Learning in neural networks: An overview. In: *Neural Networks* Bd. 61, Elsevier Ltd (2015), S. 85–117 — ISBN 0893-6080

[SeKi16] SEOK, SEONHEE ; KIM, HOWON: Visualized Malware Classification Based-on Convolutional Neural Network. In: *Journal of The Korea Institute of Information Security & Cryptology* Bd. 26 (2016), Nr. 1 — ISBN 2015200000017

[SERD11] SALAMA, MOSTAFA A ; EID, HEBA F ; RAMADAN, RABIE A ; DARWISH, ASHRAF: Hybrid Intelligent Intrusion Detection Scheme. In: *Springer Berlin Heidelberg* (2011), S. 293–303

[SHMG16] SILVER, DAVID ; HUANG, AJA ; MADDISON, CHRIS J ; GUEZ, ARTHUR ; SIFRE, LAURENT ; DRIESSCHE, GEORGE VAN DEN ; SCHRITTWIESER, JULIAN ; ANTONOGLOU, IOANNIS ; U. A.: Mastering the game of Go with deep neural networks and tree search. In: *Nature* Bd. 529, Nature Publishing Group (2016), Nr. 7585, S. 484–489

[SoPa10] SOMMER, ROBIN ; PAXSON, VERN: Outside the Closed World: On Using Machine Learning for Network Intrusion Detection. In: *Security and Privacy (SP), 2010 IEEE Symposium on* (2010), S. 305–316 — ISBN 978-1-4244-6894-2

[TsKw06] TSANG, CHI-HO ; KWONG, SAM: Ant Colony Clustering and Feature Extraction for Anomaly Intrusion Detection. In: *Swarm Intelligence in Data Mining, Springer* Bd. 123 (2006), Nr. 2006, S. 101–123

[VLLB10] VINCENT, PASCAL ; LAROCHELLE, HUGO ; LAJOIE, ISABELLE ; BENGIO, YOSHUA ; MANZAGOL, PIERRE-ANTOINE: Stacked Denoising Autoencoders: Learning Useful Representations in a Deep Network with a Local Denoising Criterion. In: *Journal of Machine Learning Research* Bd. 11 (2010), Nr. 3, S. 3371–3408 — ISBN 1532-4435

[WaCW16] WANG, YAO ; CAI, WAN-DONG ; WEI, PENG-CHENG: A deep learning approach for detecting malicious JavaScript code. In: *SECURITY AND COMMUNICATION NETWORKS, WILEY* (2016)

[Wang15] WANG, ZHANYI: The Applications of Deep Learning on Traffic Identification. In: *Black Hat USA* (2015)

[YaYu14] YAN, WEIZHONG ; YU, LIJIE: On Accurate and Reliable Anomaly Detection for Gas Turbine Combustors : A Deep Learning Approach (2014), S. 1–8

[YLWX14] YUAN, ZHENLONG ; LU, YONGQIANG ; WANG, ZHAOGUO ; XUE, YIBO: Droid-Sec : Deep Learning in Android Malware Detection. In: *Sigcomm 2014* (2014), S. 371–372 — ISBN 9781450328364

[YuLX16] YUAN, ZHENLONG ; LU, YONGQIANG ; XUE, YIBO: DroidDetector : Android Malware Characterization and Detection Using Deep Learning. In: *TSINGHUA SCIENCE AND TECHNOLOGY* Bd. 21 (2016), Nr. 1, S. 114–123

[YuRR15] YUANCHENG, LI ; RONG, MA ; RUNHAI, JIAO: A Hybrid Malicious Code Detection Method based on Deep Learning. In: *International Journal of Security and Its Applications* Bd. 9 (2015), Nr. 5, S. 205–216

[ZaSa04] ZANERO, STEFANO ; SAVARESI, SERGIO M: Unsupervised learning techniques for an intrusion detection system. In: *Proceedings of the 2004 ACM symposium on Applied computing* (2004), S. 412–419 — ISBN 1581138121

[ZuKW15] ZUECH, RICHARD ; KHOSHGOFTAAR, TAGHI M ; WALD, RANDALL: Intrusion detection and Big Heterogeneous Data : a Survey. In: *Journal of Big Data* (2015)