International Journal of Innovative Technology and Exploring Engineering (IJITEE)
ISSN: 2278-3075 (Online), Volume-9 Issue-4, February 2020

Intrusion Detection System on Big data using Deep Learning Techniques

Priyanka Dahiya, Devesh Kumar Srivastava

Abstract*: Big data is the huge amount of data with different types of V's: Velocity, Variety as well as Volume. It can be semi-structured, unstructured or structured, due to which it is not easy to analyze the data. To extract the hidden knowledge and to detect the attacks on large amount of data new architecture, techniques, algorithms, and analytics are required. Using traditional techniques to detect attacks is very difficult. In this paper, the detailed review has been done on intrusion detection on various fields using deep learning and gives an idea of applications of deep learning. The number of attacks has been increased in computer networks. A powerful Intrusion Detection System (IDS) is required to ensure the security of a network. Based on review, it is found that some studies have been done in this field, but a deep and exhaustive work has still not been done. Many researchers proposed an IDS using deep learning for unforeseen and unpredictable attacks but not for Big Data. The proposed work is based on Deep learning based intrusion detection System for big datasets named hybrid-DeepResNet-RNN run till 1,000 epochs with learning rate varying range [0.01-0.5] and three ensemble techniques, Random Forest, Decision tree regression and Gradient Boosting Tree (GBT). It is used to develop the hybrid, secure, scalable NIDS which is based on deep learning and big data techniques. The proposed classifiers produce a more reliable classification than a single classifier. The experimental results are in terms of detection rate (98.86%), false positive rate (1.110%), accuracy (99.34%) and F-Measure (97.90%). The results illuminate the better performance than existing anomaly detection techniques in the big data environment.

Keywords*: Big Data, Deep learning, Hadoop, Intrusion Detection.

I. INTRODUCTION

Anomaly detection for Big data was not handled by traditional technology. The problem of intrusion are obtained in big data were overcome by Machine Learning (ML) algorithm and this algorithm is also used to obtain valuable information from hugely available data. Now a day's big data is utilized to gather and process huge data having capacity to process and receive data rapidly and efficiently with less computational time [32].

Revised Manuscript Received on February 28, 2020.
* Correspondence Author
Priyanka Dahiya*, Manipal University Jaipur, Rajasthan. Department of School of Computing-CA, DIT University, Dehradun, India. Email: priyanka.dahiya@dituniverity.edu.in, dahiyapriyanka814@gmail.com
Devesh Kumar Srivastva, Department of School of Computing & Information Technology, Manipal University Jaipur, Rajasthan, India. Email: devesh.kumar@manipaluniversity.edu.in

Data mining and data warehousing are the most advanced techniques in data analysis and also used to determine the type of mining and recovery [33].

In ML, intelligent decisions are made spontaneously and the automatic text classifier learning is produced by an inductive process to achieve accuracy and power saving from a set of pre-classified document [9]. Anomalies are configurations obtained from the data which were not implemented by the certain idea of ordinary performance.

Various researches projected formula of machine learning for intrusion detection to cut back false positive rates as well as manufacture the correct intrusion detection system. Although to trot out big data, its various methods as well as machine learning for IDSs take an extended time in learning and classifying knowledge. Victimisation Deep learning techniques will solve several challenges like machine time , speed as well as correct IDS are developed.

Artificial neural networks' various hidden layers are contained by deep learning. In the big databases, high-level model abstractions, as well as nonlinear transformations, are applied by the deep learning methodology. Anomaly detection can also detect malformed packets. Anomaly system detects anomalous behavior. It must just be trained to recognize normal system activity. The two phases of anomaly detection system comprise of 2 phases: training as well as testing phase. The training phase consists of building a profile of normal behaviors and testing phase has the existing traffic being compared with the profile created in the training phase. [22].

II. RELATED WORK

The author defined Big Data as 3Vs, Variety, Velocity, and Volume [14]. Author defined, with 2 more Vs to the already existing Vs, adds Veracity and Value [15].

In this paper, the author believes that there is increase in the new intelligent attacks and current techniques of pattern matching are not helpful in detecting the previously unidentified attacks. The author describes existing information security also. The author suggests bigdata system model for big data analysis technology for detection of previously unknown attacks [16]. Many new techniques was introduced in big data using Hadoop and spark [36] –[40].

In this paper, authors used the Hadoop HDFS storage and apache spark computing platform for proposing the distributed anomaly detection algorithm. To make the performance of their algorithm on huge datasets, they moderated their algorithm's low memory problems by computer resource scaling.

They have evaluated their algorithm by TB size datasets on 4 node cluster which better performs than traditional algorithm and 6 times faster because of in-memory capabilities of spark.

Their distributed algorithm performs with better scalability and is capable of discord discovery in multidimensional time series [17].

This work can be extended on automatic discovery of weight of each dimension.

In this paper, authors have used DistributedWekaSpark which is scalable Big Data Mining Toolkit extends basic weka and posses the power of distributed systems. Standard Weka can be used for small datasets, not large datasets because of its memory constraints (1GB) so for execution on large datasets running apache-spark out of the box, they developed DistributedWekaSpark. It is developed on the Apache Spark's top that gives rapid in-memory distributed and parallel processing. Their evaluation results show that distributed weka spark is 4 times faster than Hadoop and achieves near-linear scalability on scaling workloads. They have used classification part only using 4 types of algorithms like FP Growth, Linear regression, SVM on spark and SVM on Hadoop. They evaluated on strong and weak scaling workloads on 5GB, 20GB and 80GB datasets and 8, 32 and 128 cores systems. Furthermore, in case of large datasets, Spark approach linearity shows the strong scaling efficiencies [18].

This paper mainly contributed to briefly reviewing the present researches numerous advanced learning methods, conventional machine learning methods which are useful in solving the various issues related to big data. Also, for the data processing the combination of Signal Processing techniques with machine learning techniques is discussed in this paper. Furthermore, various research trends, as well as open issues, are provided by the authors [19].

The technical challenges like unstructured data, privacy, error handling must be addressed for efficient and fast processing of Big Data which comes under a large domain. Hadoop is an open-source software utilized for Big Data processing [11]. New intrusion detection architecture was able to correlate few data sources like IP flow, DNS, HTTP, etc. For the performance analysis of their architecture, the frameworks 5 components such as Spark, Spark, Hive, Pig, and Hadoop are focused on comparing their performance. Based on the performance, it was found that Spark and Shark were the best performers in all scenarios [12].

"Bagging and AdaBoost" implementation showed very successful results in improving the accuracy of certain classifiers for real-world as well as for artificial datasets. Tree sizes are measured. Many issues in implementing Boosting algorithms are explored. They use graphical representation to show outputs, outliers, and noise [1].

Developing a flexible and efficient NIDS for unforeseen and unpredictable attacks, deep learning-based provides more efficient results.

In this paper, authors reviewed some particular problems related to security, Intrusion Detection Architectures, Data Fusion, and Big Heterogeneous Data as well as discuss the various fields that offer new opportunities for research [2]. In this paper, Authors proposed "Extreme Learning Machine algorithm for ELM, MR ELM", to solve big data problems. By experimental results, it shows good results with good accuracy, less False Alarm rate and high Detection rate with good speed. However, there is still room for improvement [3]. The gap of this paper is that a distributed approach can be tested on other ID datasets for measuring the "MR ELM" performance scales within modern computer surroundings.

This paper introduced a novel model for intrusion detection on Apache Spark Big Data platform with the use of SVM (Support Vector Machine) classifier. KDD99 dataset is used for model training and testing. It was shown by experimental results that Spark- Chi-SVM model is more efficient for Big data and takes less time to train the data [9]. The lack of the paper is that the model can be extended as a multi-classes model for the detection of attack.

In the case of extracting the large data amount from the unsupervised data, deep learning algorithms are proved to be very valuable [4, 5].

Deep learning works on several hidden layers and learns hierarchical feature representations bypassing the "TCP/IP information" on these layers [7,8]. The authors in this paperwork on the network-based intrusion datasets to analyze the efficiency of several deep neural networks (DNNs) and traditional machine learning for NIDS [11].

This paper proposes IDS using machine learning by learning the combinations of most of the classifiers and popular feature selection techniques. Five folds cross-validation is done on NSL-KDD dataset to find results. It is finally observed that K-NN classifier produces better performance than others and, among the feature selection methods; information gain ratio based feature selection method is better [20].

This study focused on the Deep Learning algorithms' applications as well as Big Data Analytics constructions, along with the view that how Deep Learning algorithms are being used for solving the matchless experiments that are used in Big Data Analytics issues and features [10]. The authors should focus on one or more problems with Big Data.

Authors proposed a new "self-adaptive system for anomaly detection in the context of a 5G mobile network architecture" that is a deep machine learning model of level-2 for reaching the highest processing performance. Based on the number of experimental outcomes it is clear that any frameworks will be utilized by the 5G network system [21]. Limitation of this paper is that authors can extend the work from a selection for improvement to detect best deep learning model.

III. APPLICATIONS

Table 1 summarizes a few applications of deep learning with different types of methods or algorithms. These applications are based on research papers on deep learning, machine learning, and big data.

IV. RESEARCH METHODOLOGIES

A framework is proposed in which Deep Learning and Big Data Techniques are integrated to improve the performance of intrusion detection systems.

Table I: Applications of Deep Learning on Small and Big Dataset, 2003 – 2019

Year	Authors	Proposed Algorithms	Applications
2019	Manasi Gyanchandani, Nilay Khare and, Priyank Jain	Enhanced Secure Map-Reduce layer for Big Data privacy and security [32].	Privacy in Big Data
2019	Erdogan Dogdu, Osama Faker	Intrusion Detection Using Big Data and Deep Learning Techniques [31].	Big Data
2019	Shenghong Li, Jianxun Fan, Shilin Wang, and Haonan Guo,	Learning Automata Based Incremental Learning Method for Deep Neural Networks [30].	Long sequenced CNN, fully and deeply connected perception in multilayers
2019	Yanxia Sun and Sydney Kasongo Mambwe	A Deep Learning Method with Filter Based Feature Engineering [29].	Wireless Intrusion Detection System
2019	Xinheng Wang, Peisong Li, and Ying Zhang	Intrusion Detection for IoT Based on Improved Genetic Algorithm and Deep Belief Network [28].	Applied in IoT intrusion detection, recognition as well as classification
2019	Mohsen Guizani, Xiaojiang Du, Amr Mohamed, Abdulla Khalid Al-Ali, Heena Rathore	A Novel Deep Learning Strategy is proposed that uses LSTM, a type of RNN for forecasting and predicting rest tremor velocity. [27].	Classification among various patterns od attack in Brain Implants
2019	Jiawei Wang, Yao Yang, Hongmin Gao, Xiaoke Zhang, Yongchang Wang, and Chenming Li	Deep Belief Network [24].	Classification in Spatial- Spectral for Sensor Data of Hyperspectral Remote
2018	Gregorio Martínez Pérez, Manuel Gil Pérez, Félix J. García Clemente, Angel Luis Perales Gómez, and Lorenzo Fernández Maimó	A Self-Adaptive Deep Learning-Based System [21].	Detection of Anomaly in 5th generation Networks
2018	Max Welling, Taco S. Cohen	Spherical CNNS [35].	This algorithm can be used for recognition of the 3D model as well as atomization energy regression.
2018	Hayat Khaloufi, Abderrahim Beni‐ Hessane, and Karim Abouelmehdi	Big healthcare data: Preserving security and privacy	Healthcare
2017	Qian Du, Guodong Wu, and Wei Li	Transferred Deep Learning for Anomaly Detection	Hyperspectral Imagery
2017	Linkan Bian, Mohammad Marufuzzaman, Hugh Medal, Song Zhang, Fangyan Zhang, Ravi Akula, Mojtaba Khanzadeh, Sudipta Chowdhury,	Graph‐ based feature clustering	Botnet Detection
2017	Zuha Agha, Scott Purdy, Alexander Lavin, and Subutai Ahmad,	Unsupervised real-time anomaly detection .	Anomaly detection for streaming data
2017	Carlo Sansone, Giovanni Poggi, Francesco Marra	Convolutional neural networks	Identification of a model using Iris sensor
2016	Jill Slay and Nour Moustafa	Statistical analysis of the UNSW-NB15 data set and the comparison with the KDD99 data set	Dataset method to evaluate NIDS.
2016	Silver David	Supervised learning and reinforcement learning	Mastering the Game of Go with Tree Search and Deep Neural Networks
2015	Christoph Busch, Vinay Krishna Vemuri, R. Raghavendra, Kiran B. Raja	Deep sparse filtering	Using the cameras of smartphones for Iris Recognition
2012	Penn Gerald, Jiang Hui, Mohamed Abdel-rahman, Abdel-Hamid Ossama,	Local filtering and max-pooling infrequency domain	speech recognition in case of multi-speaker
2009	Mohamed, Abdel-Rahman, George Dahl, Geoffrey Hinton	Backpropagation and associative memory architecture	Phone recognition using Deep Belief Networks
2006	Yee-Whye The, Simon Osindero, Geoffrey E., and Hinton,	Complementary Priors on Belief networks	Digit Classification
2003	David Mumford and Tai Sing Lee,	Particle filtering and Bayesian belief propagation	Inference of Hierarchical Bayesian to the visual cortex

system framework

Five classifiers are used to classify network traffic datasets, and these are Deep Residual Network (Deep ResNet), Recurrent Neural Network (RNN) and three ensemble techniques, Random Forest, decision tree regression and Gradient Boosting Tree (GBT). The dataset used for experimental work is UNSW-NB15. The proposed method used the distributed computing environment Apache Spark integrated with Keras Deep Learning Library (Tensorflow) to implement the deep learning technique while the ensemble. techniques are implemented using Apache Spark Machine Learning.

• Compare preprocessing techniques to remove the missing and null value in UNSW-NB15 dataset.

• Handle categorical data without label and compare this encoder using BackwardDifferenceEncoder, binary encoder and One hot encoder in Tensorflow. • Divide the data into training and testing sets.

• Apply the pre-processing techniques and k-fold cross validation is used in this work to evaluate the machine learning models on training datasets.

• Using the different function for comparing different approaches are logistic regression, RandomForestRegressor, decision tree regression and mean_absolute_error to fit the model.

• Compare the results of different datasets using the proposed new Deep learning and big data techniques.

• To enhance the classification accuracy and reduce the false alarm rate.

V. RESULT AND DISCUSSION

The classification accuracy, classification error rate and detection rate of anomaly detection is identified by using the following formula:

- Accuracy: Defined as the percentage of correctly classified records over the total number of records.

$$Accuracy = \frac{TP+TN}{TP+FP+TN+FN} \quad (1)$$

- Precision (P): Defined as the % ratio of the number of true positives (TP) records divided by the number of true positives (TP) and false positives (FP) classified records.

$$P = \frac{TP}{TP+FP} \times 100\% \quad (2)$$

- Recall (R): Defined as the % ratio of number of true positives records divided by the number of true positives and false negatives (FN) classified records.

$$R = \frac{TP}{TP+FN} \times 100\% \quad (3)$$

- F-Measure (F): Defined as the harmonic mean of precision and recall and represents a balance between them.

$$F = \frac{2.P.R}{(P+R)} \times 100\% \quad (4)$$

Table II: General behavior anomaly detection

Actual case	Normal prediction	Attack prediction
Normal behavior	TN	FP
Anomalies	FN	TP

• True negative Rate (TNR): the rate between the normal labels with the packets and the same label with the packets identified by the system.

• True positive Rate (TPR): the rate between the attack labels with the packets and the same label with the packets identified by the system.

• False positive Rate (FPR): the rate between the packets documented as normal and the packets identified by the attack class of the system.

• False negative Rate (FPR): the rate between the packets recognized as attack and the packets noticed by the normal class member of the system.

The "Fig. 1", shows the proposed method results with other methods using UNSW 15 datasets. The given datasets were separated into train and test datasets, and normalized using L2 normalization. In terms of accuracy noted that the DT classifiers performed better than the other classifiers namely GBT and RF. Additionally, the performance of DT, GBT and RF classifiers remains the same range across different datasets. The performance obtained in terms of FPR is less comparatively to other classical machine learning classifiers in all the datasets.

Table III: Configuration of proposed hybrid- DeepREstNet-RNN model

Layers	Type	Output shapes	Number of units	Activation Function	Parameters
0-7	Fully connected	(None, 512)	512	Leaky ReLU	4,83,728
7-8	Batch Normalization	(None, 512)	512		2048
8-9	Dropouts (0.01)	(None, 512)	512		0
9-13	Fully connected	(None, 256)	256	Leaky ReLU	2,35,678
13-14	Batch Normalization	(None, 256)	256		1024
14-15	Dropouts (0.01)	(None, 256)	256		0
15-16	Fully connected	UNB ISCX 2012, CICIDS2017, UNB ISCX 2018, NSL-KDD and UNSW-NB15.	Binary, Multiclass		4,83,728

Table IV: Performance of proposed method and other methods using UNSW-NB15

dataset

Algorithm	Recall (TPR)	Precision	Accuracy
Decision tree regression (DTR)	97.1	67.5	98.72
Gradient Boosting Tree (GBT).	97.5	70.5	98.85
Random Forest (RF)	97.9	70.6	98.93
Hybrid-DeepResNet-RNN Layer 1	98	71.9	99.3

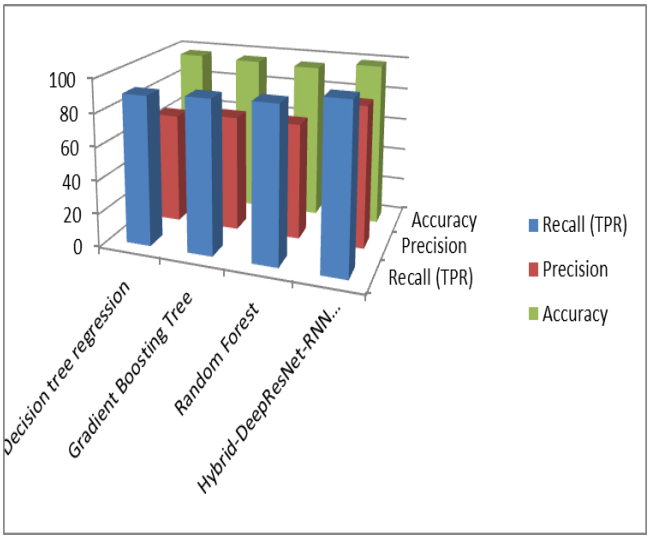

Fig.1. Performance of proposed method and other methods using UNSW-NB15 dataset

VI. CONCLUSION AND FUTURE WORK

In this paper, hybrid intrusion detection hybrid-DeepResNet-RNN method is proposed and this is compared with the three ensemble methods: Random Forest, Decision tree regression and Gradient Boosting Tree (GBT). The framework employed distributed deep learning model with RNNs for handling and analyzing very large scale data sets. The RNN model was chosen by comprehensively evaluating their performance in comparison to classical machine learning classifiers on various benchmark IDS datasets. Experimental results show that the proposed method hybrid-DeepResNet-RNN is more accurate and efficient than the traditional machine learning techniques. The Random forest gives better results than other 2 methods. The experiments show the results in the form of detection rate (98.86%), false positive rate (1.110%), accuracy (99.34%) and F-Measure (97.90%).

The performance can be further improved by training complex RNNs architectures with the addition of more nodes through distributed approach .elimination of human intervention is of vital importance. Machine, Deep Belief Network, and similar algorithm have made intrusion detection performance more efficient. Application shows in table give an idea that deep learning has many applications and used in many fields but less work is on intrusion detection on big data using deep learning.

REFERENCES

1. Nassar Nassar M, al Bouna B, Malluhi , “Secure outsourcing of network flow data analysis. In: Big Data (BigData Congress)”, IEEE International Congress On. IEEE, Santa Clara, CA, USA. 2013, pp. 431–432
2. Richard Zuech*, Taghi M Khoshgoftaar and Randall Wald, “Intrusion detection and Big Heterogeneous Data: a survey”, Zuech et al. Journal of Big Data (2015) 2:3, DOI 10.1186/s40537-015-0013-4
3. Junlong Xiang, Magnus Westerlund, Dušan Sovilj, Using Extreme Learning Machine for Intrusion Detection in a Big Data Environment Sciences ,2014, pp. 73-82
4. Hinton GE, Osindero S, Teh Y-W, “A fast learning algorithm for deep belief nets. Neural Comput 18(7)”, 2006, pp.1527–1554
5. Bengio Y, Lamblin P, Popovici D, Larochelle H, “Greedy layer-wise training of deep networks”, Vol. 19, 2007.
6. PekkaPääkkönen∗, DanielPakkala1, “Reference Architecture and Classification of Technologies, Products and Services for Big Data Systems”, Available online 2 February 2015.
7. Y. LeCun, Y. Bengio, and G. Hinton, “Deep learning”, Nature, vol. 521, no. 7553, 2015, pp. 436.
8. Y. Xin et al., “Machine learning and deep learning methods for cybersecurity," IEEE Access, vol. 6, 2018, pp. 35365-35381.
9. Suad Mohammed Othman1, Fadl Mutaher Ba‑Alwi1, Nabeel T. Alsohybe1 and Amal Y. Al‑Hashida2, “Intrusion detection model using machine learning algorithm on Big Data environment”, Othman et al. J Big Data 2018 5:34 , https://doi.org/10.1186/s40537-018-0145-4
10. Maryam M Najafabadi1, Flavio Villanustre, Taghi M Khoshgoftaar1, Naeem Seliya1,Randall Wald1 and Edin Muharemagic, “Deep learning applications and challenges in big data analytics”, Najafabadi et al. Journal of Big Data (2015), DOI 0.1186/s40537-014-0007-7 pp. 1-21
11. R. Vinayakumar , Mamoun Alazab, (Senior Member, Ieee), K. P. Soman1,Prabaharan Poornachandran, Ameer Al-Nemrat, And Sitalakshmi Venkatraman, “Deep Learning Approach for Intelligent Intrusion Detection System”, January 3, 2019, DO I 10.1109/ACCESS.2019.2895334
12. Harshawardhan S. Bhosale1, Prof. Devendra P. Gadekar, “A Review Paper on Big Data and Hadoop”, International Journal of Scientific and Research Publications, Volume 4, Issue 10, October 2014 , pp.2250-3153
13. Marchal_y, Xiuyan Jiangz, Radu State, Thomas Engel, “A Big Data Architecture for Large Scale Security Monitoring”, Samuel, Springer, 2014.
14. Laney D 3d data management: Controlling data volume, velocity and variety. Technical Report 949, META Group (now Gartner). http://blogs.gartner.com/doug-laney/files/2012/01/ad949-3D-Data-Management-Controlling-Data-Volume-Velocity-and-Variety.pdf
15. Sourcefire (2015) Snort, Home Page. http://www.snort.org/. Accessed 2015-1-10
16. Sung-Hwan Ahn, Nam-Uk Kim, Tai-Myoung Chung, “Big Data Analysis System Concept for Detecting Unknown Attacks”, IEEE, 2014.
17. Yafei Wu, Yongxin Zhu, Tian Huang, “Distributed Discord Discovery: Spark Based Anomaly Detection in Time Series”, IEEE, 2015.
18. Aris-Kyriakos Koliopoulos, Paraskevas Yiapanis, Firat Tekiner, Goran Nenadic, John Keane, “A Parallel Distributed Weka Framework for Big Data Mining using Spark” , IEEE 2015.
19. Junfei Qiu, Qihui Wu, Guoru Ding, Yuhua Xu and Shuo Feng, “A survey of machine learning for big data processing”, Springer Open 2016
20. Saroj Kr. Biswas1, “Intrusion Detection Using Machine Learning: A Comparison Study”, February 1, 2018, International Journal of Pure and Applied Mathematics Special Issue .
21. Lorenzo Fernández Maimó, Ángel Luis Perales Gómez, Félix J. García Clemente, Manuel Gil Pérez, and Gregorio Martínez Pérez ,“A Self-Adaptive Deep Learning-Based System for Anomaly 1Detection in 5G 1Networks”,IEEE explore.
22. Kamaldeep Singh, Sharath Chandra Guntuku, Abhishek Thakur, Chittaranjan Hota, “Big Data Analytics framework for Peer-to-Peer Botnet detection using Random Forests”, Springer, September 2014.
23. R.Vani, “Towards Efficient Intrusion Detection using Deep Learning Techniques: A Review”, International Journal of Advanced Research in Computer and Communication Engineering , vol 6, issue 10, 2017, pp. 375-384.

24. Chenming Li 1, Yongchang Wang 1, Xiaoke Zhang , Hongmin Gao, Yao Yang 1 and JiaweiWang,"Deep Belief Network for Spectral–Spatial Classification of Hyperspectral Remote Sensor", Data, January 2019
25. Introducing Deep Learning Self-Adaptive Misuse Network Intrusion Detection Systems, 2019. Digital Object Identifier 10.1109/ACCESS.2019.2893871.
26. Farrukh Aslam Khan, Abdu Gumaei, Abdelouahid Derhab, Amir Hussain, "A Novel Two-Stage Deep Learning Model for Efficient Network Intrusion Detection", Special Section On Artificial Intelligence And Cognitive Computing For Communication And Network, February 15, 2019, Digital Object Identifier 10.1109/ACCESS.2019.2899721.
27. Heena Rathore 1, Abdulla Khalid Al-Ali 1,Amr Mohamed 1, Xiaojiang Du,Mohsen Guizani 1, " A Novel Deep Learning Strategy for Classifying Different Attack Patterns for Deep Brain Implants".
28. Ying Zhang1 , Peisong Li1, And Xinheng Wang, " Intrusion Detection for IoT Based on Improved Genetic Algorithm and Deep Belief Network", Digital Object Identifier 10.1109/ACCESS.2019.2903723, March 25, 2019
29. Sydney Mambwe Kasongo And Yanxia Sun, "A Deep Learning Method With Filter Based Feature Engineering for Wireless Intrusion Detection System", March 18, 2019, Digital Object Identifier 10.1109/ACCESS.2019.2905633
30. Haonan Guo , Shilin Wang , Jianxun Fan5, And Shenghong Li, "Learning Automata Based Incremental Learning , Method for Deep Neural Networks" April 11, 2019. Digital Object Identifier 10.1109/ACCESS.2019.2907645
31. Osama Faker, Erdogan Dogdu, , "Intrusion Detection Using Big Data and Deep Learning Techniques", April 2019.
32. Priyank Jain , Manasi Gyanchandani and Nilay Khare, "Enhanced Secured Map Reduce layer for Big Data privacy and security", Jain et al. J Big Data (2019).
33. Y. Wang, X. Li and X. Ding, "Probabilistic framework of visual anomaly detection for unbalanced data", Elsevier, Neuro computing, vol. 201, 2016, pp.12-18. https://doi.org/10.1186/s40537-019-0193-4 pp- 1-17
34. S. Suthaharan, "Big data classification: Problems and challenges in network intrusion prediction with machine learning", ACM SIGMETRICS, Performance Evaluation Review, vol.41, no. (4), 2014, pp.70-73
35. Taco S. Cohen, Mario Geiger, Jonas Koehler, Max Welling, "Spherical CNNs", Proceedings of the International Conference on Learning Representations, 2018, pp. 1-15
36. Devesh Kumar Srivastava "Big challenges in big data research", Data mining and knowledge engineering, vol 6, issue 7, 2014, pp. 282-286
37. Priyanka Dahiya, Devesh Kumar Srivastava, "Network intrusion detection in big dataset using Spark", International Conference on Computational Intelligence and Data Science (ICCIDS 2018),Procedia computer science, vol 132, 2018, pp. 253-262.
38. Priyanka Dahiya, Devesh Kumar Srivastava," A Comparative Evolution of Unsupervised Techniques for Effective Network Intrusion Detection in Hadoop", International Conference on Advances in Computing and Data Sciences,2018, pp. 279-287
39. Devesh Kumar Srivastava, Ravinder Yadav, Gaurav Agrwal, "Map reduce programming model for parallel K-mediod algorithm on hadoop cluster", 7th international conference on communication systems and network technologies, 2017, pp. 74-78
40. Ravinder Yadav, Aravind Kilaru, Devesh Kumar Srivastava, Priyanka Dahiya, "Performance Evaluation of Word Count Program Using C#, Java and Hadoop", International Conference on Smart Trends for Information Technology and Computer Communications, 2016, pp. 299-307

AUTHOR PROFILE

Priyanka Dahiya received the Master"s degree from Sikkim Manipal Institute of Technology, Sikkim, India, in 2011 in Computer Science and Engineering. She is currently pursuing the Ph.D. degree with Manipal University Jaipur, Rajasthan, India. She is currently working in DIT University of School of computing. Her current research interests include data mining, deep learning and big data.

Dr. Devesh Kumar Srivastava received the Master's degree from AAI, Allahabad in 2005; and the Ph.D. degree from UTU, Dehradun in 2012.He was with Software Services Pvt Ltd Gurgaon as Senior Software Engineer for two year and nine months. He is currently a Professor with the department of School of Computing & Information Technology, Manipal University Jaipur, Rajasthan, India. His current research interests include Software Engineering, Operating System, Data Mining, big data, DBMS, Web Tech, Computer Architecture, Computer Network, Oops Technology..