Association Measures in Network Outlier Detection Methods

Ch.Nagamani , Suneetha Chittineni

***Abstract**: Detecting outliers before they cause any damage to the data in the network is a important constraint. Outlier detection methods need to be applied on various applications like fraud detection, network robustness analysis. This paper mainly focuses on detailed measures of both proposed intrusion and outlier detection methods with traditional methods. In the proposed work, KDD CUP data set is used. In this work, we initially divide the entire network into individual nodes for efficient monitoring. Later, the proposed methodology is applied on networks which can easily handle high / multidimensional data. While detection of outliers, the proposed method divides the entire network into sub-networks and each network is formed with density based strategy and then outlier detection is applied on them using a Efficient Crossover Design method which identifies the outliers more accurately. Finally ,the proposed method is evaluated and compared with traditional method will all possible parameters in network intrusion detection and the results prove that the performance levels of the proposed method is far better than the traditional methods.*

***Keywords** : Comparative Analysis, Data Mining, Outlier Detections, Network Data, Processing, Clustering, Classification, Feature Extraction.*

I. INTRODUCTION

With the increase in the dimensionality of the data, the techniques applied on the streaming data or large data sets cannot provide an efficient result and also takes high computation time [6]. Construction of the model is required that perfectly represents the data for the effective outlier detection. Over the years, many outlier detection techniques have been developed for anomaly detection and developing models for outliers. Various problems arise due to this increased size of data like data redundancy, missing data etc [7]. The outlier detection from the data helps to acquire the useful knowledge that will help in data analysis. This can be used for different applications in different domains that will give the efficient results.The Figure-1 defines the process of outlier detection. Initially a network is created. Network is basically the area or the region in which the dataset is present and outliers are identified in this dataset after applying the proposed irregular random forest method on it. In next step the data is taken from the network and is converted into log files. By converting the obtained data into log files the outlier can be easily detected.

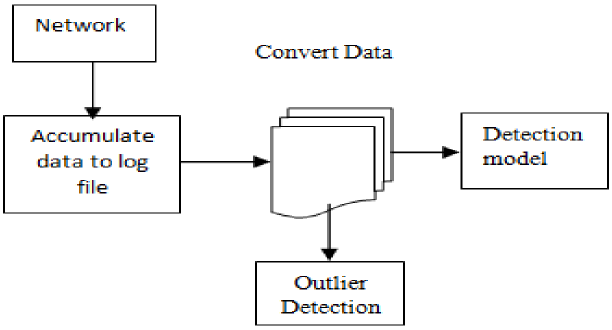


Fig-1 General Process of Detecting Outliers

The main objective of detecting outliers is to retrieve the objects from the large datasets that have different behavior than the normal object present in the data [8]. Detection of outliers is the important field of data mining for which various algorithms have been used [9].In literature, various outlier detection algorithms are used [10]. Outlier detection has been studied on a large variety of data types including high-dimensional data, uncertain data, stream data, graph data, time series data, spatial data, and spatio-temporal data.

II. ASSOCIATION MEASURES

A network which has many semi-networks are considered and the DBOD algorithm[2] for detecting outliers is applied on the sub networks. The network dataset is considered from UCI machine repository. The DBOD algorithm uses MATLAB with the end goal of information mining. The information mining is utilized to extricate the data from the extensive dataset. The Network Outlier Detection System (NODS) algorithm[1] is compared with the traditional Spatial Outlier Algorithm (SOA) and the results show that the performance levels are much better than existing methods. Below figure 2 plots the outlierness of the 1% attacks. Since the attacks are infused at the start of the dataset, the figure demonstrates the exception’s of the attacks is considerably higher than the greater part of ordinary exercises.

Manuscript published on November 30, 2019.
* Correspondence Author

Ch.Nagamani *, Research Scholar, Acharya Nagarjuna University, Guntur. nagamani502@gmail.com

Dr. Suneetha Chittineni, Associate Professor, Department of Computer Applications, RVR & JC College of Engg & Tech, Chowdavaram, Guntur-522019

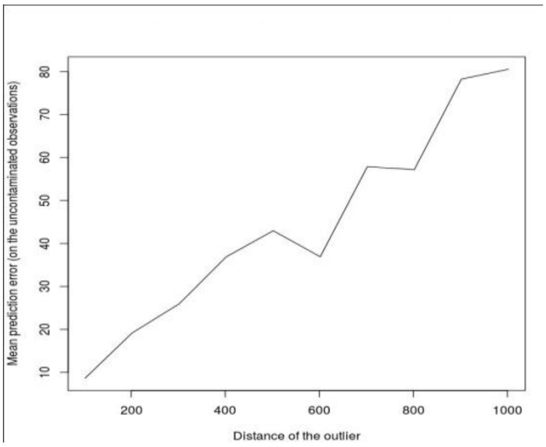


Fig 2: Attack outlierness Percentage

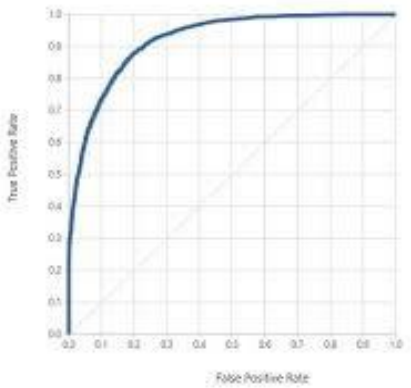


Fig 3: Attack curve level

It likewise demonstrates the different execution time of the above said calculation against the quantity of information focuses. The execution time is appeared in seconds. This diagram is drawn by taking the quantity of information focuses in X-hub and the execution time in Y-pivot.

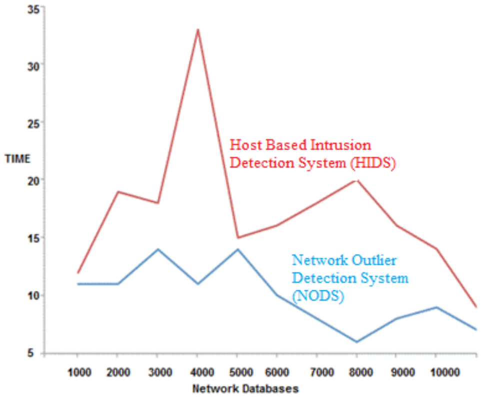


Fig 4: Association of Execution Time amongst proposed and existing calculations

The outlier detection rate analysis based on node longitude location value and the degree of the node is illustrated in the below graph.

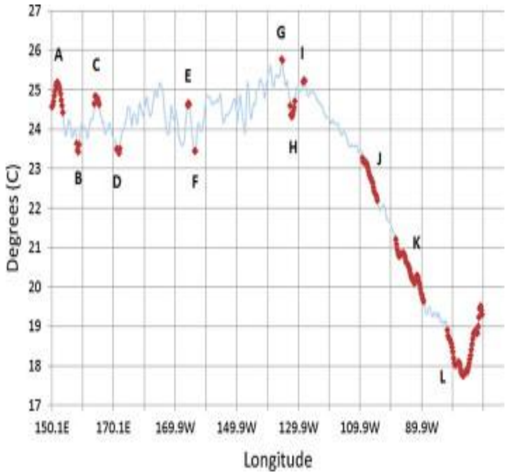
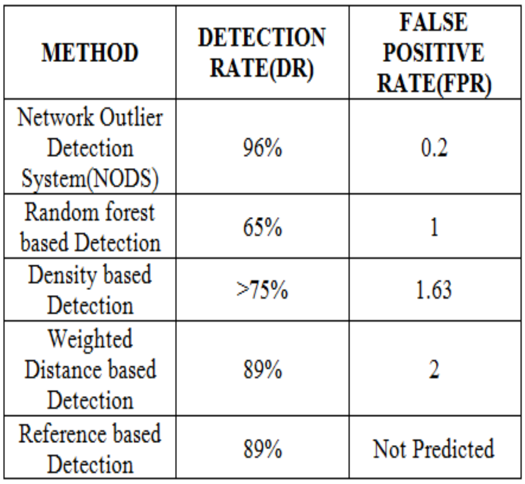


Fig 5: Outlier Detection Rate.

Table 1: Execution Tabulation



METHOD	DETECTION RATE(DR)	FALSE POSITIVE RATE(FPR)
Network Outlier Detection System(NODS)	96%	0.2
Random forest based Detection	65%	1
Density based Detection	>75%	1.63
Weighted Distance based Detection	89%	2
Reference based Detection	89%	Not Predicted

The networks range considered in the NODS method are compared with existing method in terms of data processing.

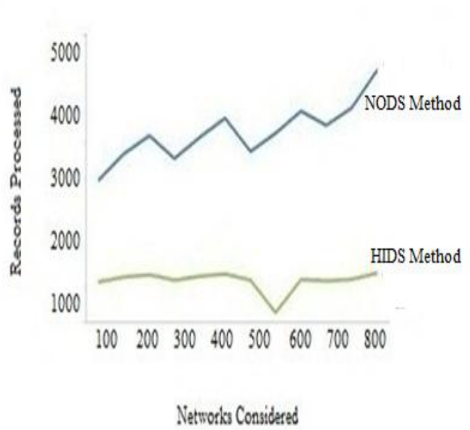


Fig 6: Data Processing Levels

The Outlier detection rate is much better and quicker than the traditional methods. The outliers are effectively identified and can be resolved. The outlier detection rate is depicted as below.

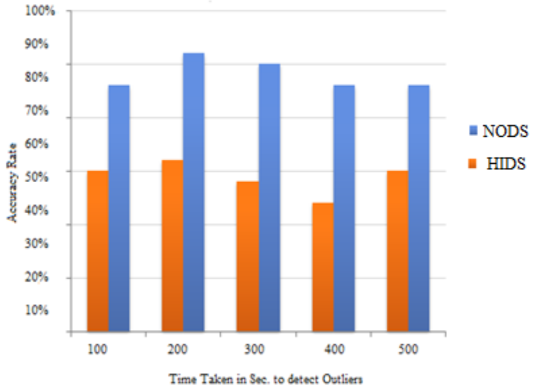


Fig 7: Accuracy in detecting Outliers.

When there are semi-networks in a network then the outlier detection among the semi-networks are identified accurately and displayed. The identified outliers in a semi-network are depicted in below figure.

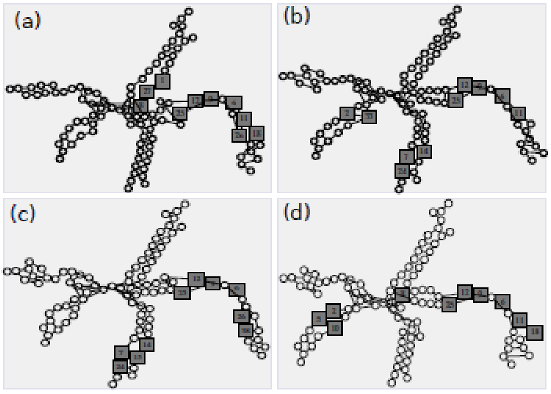


Fig 8: Outliers identified in semi-networks of a network.

The "KDD 99" dataset is accessible from DARPA's intrusion dataset assessment program. This dataset has been broadly utilized in both intrusion identification also, for Intrusion detection, and information have a place with four principle attack classes. The Efficient crossover design method[4] effectively identifies the intrusions with 97% accuracy when compared to traditional methods which Efficiently identifies the intrusions. The Processing time and the payload levels are depicted in figure 9 which shows that the processing time for outlier detection is low in the Efficient Cross over method and figure 10 shows the detection rate.

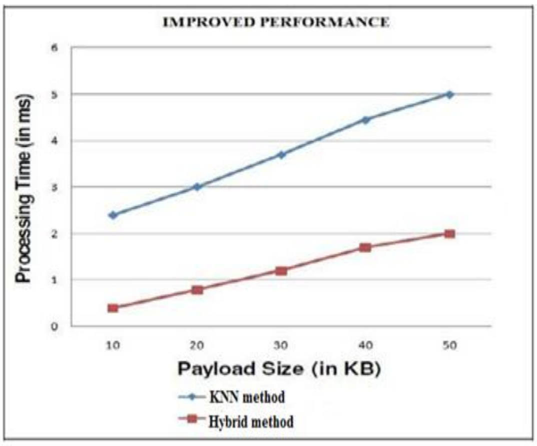


Fig 9: Processing Time Vs Payload

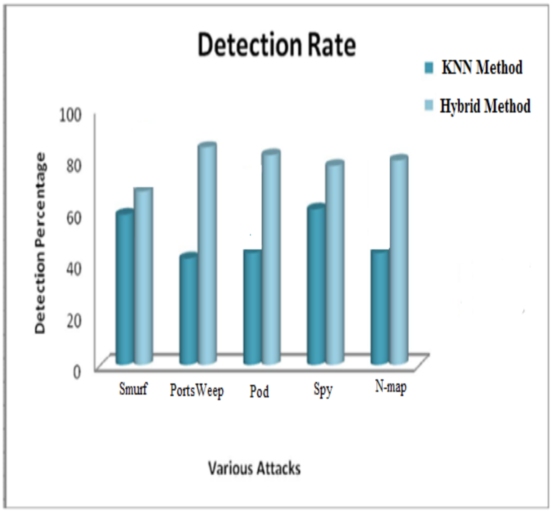
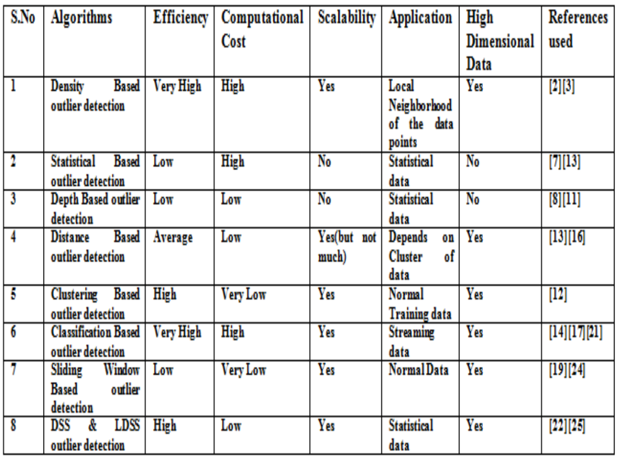


Fig 10: Rate of Intrusion Detection

2.1 Efficiency

It is the evaluation of the average execution time required for an algorithm to complete work on a given data set. Efficiency of an algorithm is measured by its order. It is helpful for quantifying implementation difficulties of certain problems. Based on the references identified, the association of different outliers are represented in Table-2.

Table-2 Association of Outlier Detection Techniques



S.No	Algorithms	Efficiency	Computational Cost	Scalability	Application	High Dimensional Data	References used
1	Density Based outlier detection	Very High	High	Yes	Local Neighborhood of the data points	Yes	[2][3]
2	Statistical Based outlier detection	Low	High	No	Statistical data	No	[7][13]
3	Depth Based outlier detection	Low	Low	No	Statistical data	No	[8][11]
4	Distance Based outlier detection	Average	Low	Yes(but not much)	Depends on Cluster of data	Yes	[13][16]
5	Clustering Based outlier detection	High	Very Low	Yes	Normal Training data	Yes	[12]
6	Classification Based outlier detection	Very High	High	Yes	Streaming data	Yes	[14][17][21]
7	Sliding Window Based outlier detection	Low	Very Low	Yes	Normal Data	Yes	[19][24]
8	DSS & LDSS outlier detection	High	Low	Yes	Statistical data	Yes	[22][25]

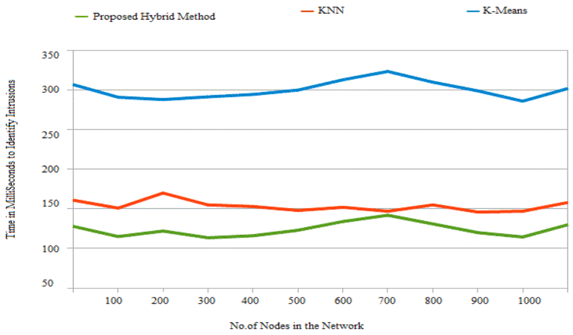


Fig 11: Time taken for identification of intrusions

2.2 Computational Cost

It is directly proportional to the computational complexity of the algorithm. It is the evaluation of the number of steps required by the algorithm related for input of an instance or a given size in the worst case. Function of size is measured by the number of steps.

Retrieval Number: D8545118419/2019©BEIESP
DOI:10.35940/ijrte.D8545.118419
Journal Website: www.ijrte.org

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication

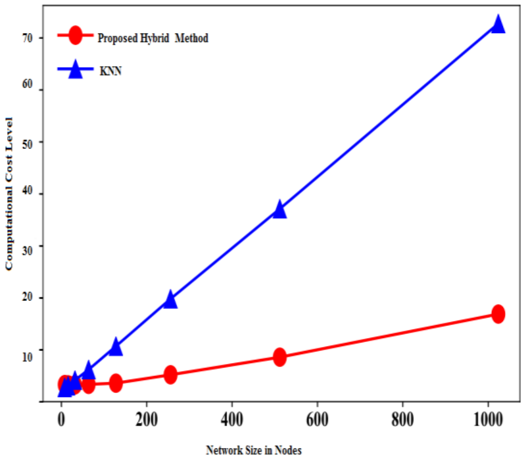


Fig 12: Computational Cost Levels

2.3 Scalability

It is defined as the capability of the product or a computer application to continue to function well even when it is changed in size or volume, as per the user requirements. It is basically a rescaling like expandability of an application program which can be used on larger operating systems for handling large number of users and also for better performance.

2.4 Applicability

As each algorithm has its boundaries and limits set for being applicable on any given set of data. Depending upon the data set i.e. whether it's a statistical data or large dataset, various algorithms are applied on the datasets to detect outliers. All the above stated outlier detection algorithms are compared in table-1 with respect to certain parameters like efficiency, computational cost, scalability, applicability etc.

III.OVERALL ASSOCIATION OF PROPOSED & EXISTING METHODS

Intrusion detection system (IDS) has been attempted early with various data mining techniques. The dataset used for these are KDD-CUP 99, NSL-KDD, UNSW-NB 15. The performance of algorithm on KDD dataset was plotted and the results are shown in fig.13.

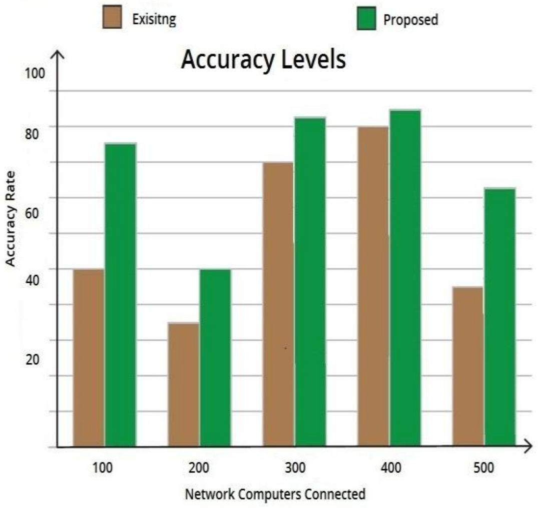


Fig-13: Accuracy Levels

The time for identification of outliers are illustrated in Figure-14.

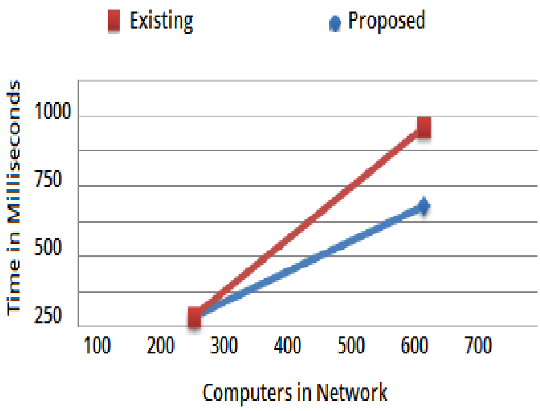


Fig-14: Identification Time

The Hybrid Framework[4] introduced a two-level half and half interruption location technique in light of directed and anomaly strategies. This technique shows extraordinary execution in perceiving remarkable characterization ambushes and likewise tremendous scale attacks new and presented strikes when attempted with a NSL KDD datasets. The system setup time for creating sub-networks from a large network is represented in Figure 15.

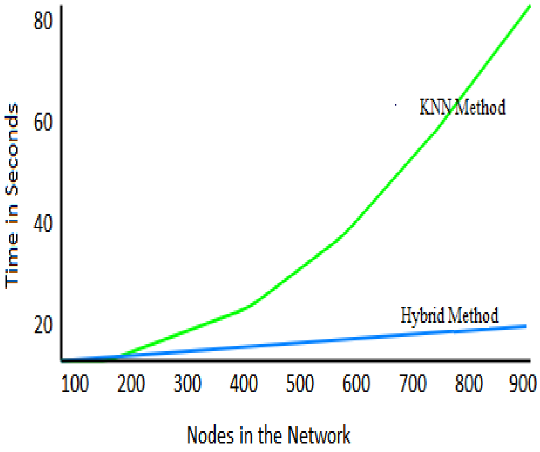


Fig 15: System Setup Time

The proposed method detects more attacks when compared to traditional method. The results show that the proposed method identifies more attacks than the traditional methods.

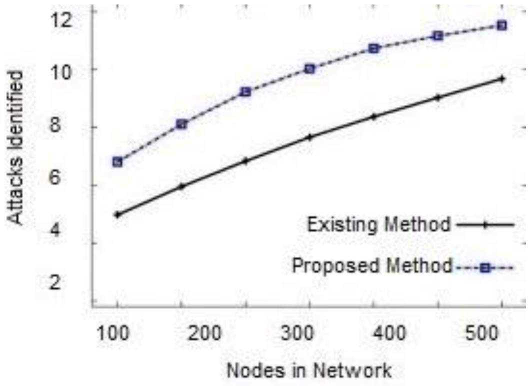


Fig 16: Identified Attacks

The Efficient crossover design effectively identifies the intrusions by inspecting every parameter for secured data transmission. In future, we might want to investigate a versatile execution of our calculation. The Proposed Hybrid method performs intrusion detection and identifies the errors at 97% accuracy whereas the traditional methods exhibits 87% accuracy rate.

VI. CONCLUSION

The speed of processing the data is to be increased that helps in the reduction of processing cost of data. There is no single universally applicable outlier detection approach of the current techniques. This paper presents the study of different existing outlier detection techniques and the way in which they are categorized. It is found that efficiency and computational complexity depends upon the data distribution and type of data. It is also observed that no individual algorithm is much suited for the high dimensional data. There is need of developing some new algorithms or improvement in the existing one is required. In future work, discussed algorithms will be explored using different parameters which are not included in this paper. What's more, we re-port on and investigate different Intrusion location systems under directed and unsupervised methodologies. In light of our audit, we watch that the idea of Intrusion is distinctive for various application spaces. Hence, advancement of a successful exception discovery method for blended kind and developing system movement information, particularly within the sight of clamor, is a testing undertaking. This system will attempt to influence a more suitable social affair to approach in light of speedier and profitable classifiers with a specific end goal to make a basic duty in the examination of the outlier acknowledgment. ECD-IDS Method registered the highest accuracy rate 97%, with the smallest false positive rate. It seems that the ECD method presents acceptable performance parameters except the false negative parameter. In this research work, an information mining based Efficient intrusion detection framework has been intended for recognizing typical and meddlesome occasions. The significant commitments of this work are the proposition of a Cross over structure for powerful Intrusion Detection, the identification methods for system as well as host Intrusion Detection frameworks that utilization arrangement and grouping calculations to upgrade the exhibition of the intrusion identification system.

REFERENCES

1. Ch.Nagamani , Dr.Suneetha Chittineni , “Network Intrusion Detection Mechanisms using Outlier Detection”, ICICCT-2018 IEEE conference, pp-1468-1473,2018.
2. Ch.Nagamani, Dr.Suneetha Chittineni,” Efficient Neighborhood Density Based Outlier Detection Inside a Sub Network with High Dimensional Data”,””pp-107-111, 2019.
3. Ch.Nagamani, Dr.Suneetha Chittineni,” Network Database Security with Intellectual Access Supervision using Outlier Detection Techniques”, International Journal of Advanced Intelligence Paradigms/Forthcoming Articles.
4. Ch.Nagamani, Dr.Suneetha Chittineni,” Intrusion Detection Methods for Secure Data Communication using Efficient Cross Over Design Technique ”, Journal of Advanced Research in Dynamical and Control Systems, Vol. 11, No. 9, 2019.
5. C. C. Aggarwal, “Outlier ensembles,” *ACM SIGKDD Explorations Newsletter*, vol. 14, no. 2, pp. 49–80, 2017.
6. M. Gupta, J. Gao, C. C. Aggarwal, and J. Han, “Outlier detection for temporal data: a survey,” *IEEE Transactions on Knowledge and Data Engineering*, vol. 26, no. 9, pp. 2250–2267, 2014.
7. A. Zimek, E. Schubert, and H.-P. Kriegel, “A survey on unsupervised outlier detection in high-dimensional numerical data,” *Statistical Analysis and Data Mining*, vol. 5, no. 5, pp. 363–387, 2012.
8. P. Gogoi, D. K. Bhattacharyya, B. Borah, and J. K. Kalita, “A survey of outlier detection methods in network anomaly identification,” *The Computer Journal*, vol. 54, no. 4, pp. 570– 588, 2011.
9. H. Huang, K. Mehrotra, and C. K. Mohan, “Rank-based outlier detection,” *Journal of Statistical Computation and Simulation*, vol. 83, no. 3, pp. 518–531, 2013.
10. H. P. Kriegel, P. Kroger, E. Schubert, and A. Zimek, “Outlier Detection in Axis-Parallel Subspaces of High Dimensional Data,” in *Proceedings of the Pacific-Asia Conference on Advances in KnowledgeDiscovery andDataMining*, pp. 831–838, Springer- Verlag, 2009.
11. F. Keller, E. M¨uller, and K. B¨ohm, “HiCS: High contrast subspaces for density-based outlier ranking,” in *Proceedings of the IEEE 28th International Conference on Data Engineering, ICDE 2012*, pp. 1037–1048, USA, April 2012.
12. S. Ramaswamy, R. Rastogi, and K. Shim, “Efficient algorithms for mining outliers from large data sets,” in *Proceedings of the ACM SIGMOD International Conference on Management of Data*, pp. 427–438, 2000.
13. F. Angiulli and C. Pizzuti, “Fast Outlier Detection in High Dimensional Spaces,” in *Proceedings of the European Conference on Principles of Data Mining and Knowledge Discovery*, pp. 15– 26, Springer-Verlag, Heidelberg, Berlin, Germany, 2002.
14. M. M. Breunig, H.-P. Kriegel, R. T. Ng, and J. Sander, “LOF: identifying density-based local outliers,”*ACMSIGMODRecord*, vol. 29, no. 2, pp. 93–104, 2000.
15. H.-P. Kriegel, P. Kr¨oger, E. Schubert, andA. Zimek, “LoOP: local outlier probabilities,” in *Proceedings of the ACM 18th International Conference on Information and Knowledge Management (CIKM ’09)*, pp. 1649–1652, ACM Press, November 2009.
16. H. Ville, I. Karkkainen, and P. Franti, “Outlier Detection Using k-Nearest Neighbour Graph,” in *Proccedings of the IEEE International Conference on Pattern Recognition*, vol. 3, pp. 330– 433, 2004.
17. J. Zhang, Y. Jiang, K. H. Chang, S. Zhang, J. Cai, and L. Hu, “A concept lattice based outlier mining method in lowdimensional subspaces,” *Pattern Recognition Letters*, vol. 30, no. 15, pp. 1434–1439, 2009.
18. J. Zhang, X. Yu, Y. Li, S. Zhang, Y. Xun, and X. Qin, “A relevant subspace based contextual outlier mining algorithm,” *Knowledge-Based Systems*, vol. 99, no. 72, pp. 1–9, 2016.
19. J. K. Dutta, B. Banerjee, and C. K. Reddy, “RODS: Rarity based Outlier Detection in a Sparse Coding Framework,” *IEEE Transactions on Knowledge and Data Engineering*, vol. 28, no. 2, pp. 483–495, 2016.
20. E. M¨uller, I. Assent, U. Steinhausen, and T. Seidl, “OutRank: Ranking outliers in high dimensional data,” in *Proceedings of the 2008 - IEEE 24th International Conference on Data Engineering Workshop, ICDE’08*, pp. 600–603, Mexico, April 2008.
21. E. M¨uller, M. Schiffer, and T. Seidl, “Adaptive outlierness for subspace outlier ranking,” in *Proceedings of the 19th International Conference on Information and Knowledge Management and Co-located Workshops, CIKM’10*, pp. 1629–1632, Canada, October 2010.
22. A. Lazarevic and V. Kumar, “Feature bagging for outlier detection,” in *Proceedings of the KDD-2005: 11th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 157–166, USA, August 2005.
23. B. Van Stein, M. Van Leeuwen, and T. Back, “Local subspacebased outlier detection using global neighbourhoods,” in *Proceedings of the 4th IEEE International Conference on Big Data, Big Data 2016*, pp. 1136–1142, USA, December 2016.
24. A. Zimek, M. Gaudet, R. J. G. B. Campello, and J. Sander, “Subsampling for efficient and effective unsupervised outlier detection ensembles,” in *Proceedings of the 19th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD 2013*, pp. 428–436, USA, August 2013.
25. J. R. Pasillas-Diaz and S. Ratte, “Bagged subspaces for unsupervised outlier detection,” *International Journal of Computational Intelligence*, vol. 33, no. 3, pp. 507–523, 2017.

26. A. Ghoting, M. E. Otey, and S. Parthasarathy, “LOADED: Linkbased outlier and anomaly detection in evolving data sets,” in *Proceedings of the Fourth IEEE International Conference onData Mining, ICDM 2004*, pp. 387–390, UK, November 2004.
27. A. Koufakou and M. Georgiopoulos, “A fast outlier detection strategy for distributed high-dimensional data sets with mixed attributes,” *Data Mining and Knowledge Discovery*, vol. 20, no. 2, pp. 259–289, 2010.
28. K. Zhang and H. Jin, “An effective pattern based outlier detection approach for mixed attribute data,” in *AI 2010: Advances in Artificial Intelligence*, vol. 6464 of *Lecture Notes in Computer Science*, pp. 122–131, Springer, Berlin, Germany, 2010.
29. Y.-C. Lu, F. Chen, Y. Wang, and C.-T. Lu, “Discovering anomalies on mixed-type data using a generalized Student-t based approach,” *Expert Systems with Applications*, vol. 28, no. 10, pp. 1–10, 2016.
30. K. Do, T. Tran,D. Phung, and S. Venkatesh, “Outlier detection on mixed-type data: an energy-based approach,” in *Advanced Data Mining and Applications*, pp. 111–125, Springer International Publishing, Cham, switzerland, 2016.
31. H. Huang, K. Mehrotra, and C. K. Mohan, “Outlier detection using modified-ranks and other variants,” Electrical Engineering and Computer Science 72, 2011, https://surface.syr.edu/eecs techreports/72/.
32. M. Radovanovi´c, A. Nanopoulos, and M. Ivanovi´c, “Reverse nearest neighbors in unsupervised distance-based outlier detection,” *IEEE Transactions on Knowledge and Data Engineering*, vol. 27, no. 5, pp. 1369–1382, 2015.
33. G. Bhattacharya, K. Ghosh, and A. S. Chowdhury, “Outlier detection using neighborhood rank difference,” *Pattern Recognition Letters*, vol. 60, pp. 24–31, 2015.
34. B. Tang and H. He, “A local density-based approach for outlier detection,” *Neurocomputing*, vol. 241, pp. 171–180, 2017.