International Journal of Medical Toxicology & Legal Medicine Volume 27, No. 2, 2024
e-ISSN: 0974-4614
p-ISSN: 0972-0448

Detection of Unauthorized Users to Secure the Sensitive Medical Data Using Machine Learning Method

Chanumolu Kiran Kumar[1], G. Muni Nagamani[2], Jayakrishna Amathi[3], Ragini Mokkapati[4], Katakam Ranganarayana[5]

[1]Associate Professor, Department of AI & DS, Lakireddy Bali Reddy College of Engineering (A), Mylavaram, NTR(dt) Andhra Pradesh. Email:mounikakiran.138@gmail.com
[2]Asst.Professor, Department of Computer Science & Engineering,Andhra Loyola Institute of Engineering & Technology (A), Vijayawada, Andhra Pradesh. Email:dr.muninagamani@gmail.com
[3]Computer Science and Engineering, University of North Texas, 1155 Union Circle, Denton, Texas. Email:amathijayakrishna@gmail.com
[4]Department of Computer Science and Engineering, Koneru Lakshmaiah Education Foundation, Vaddeswaram, Guntur Dist., Andhra Pradesh. Email:raginimokkapati@gmail.com
[5]Senior QA Engineer, Xinthe Technologies Pvt Ltd, SRK Destiny, VIP Road, Visakhapatnam. Email:katakam916@gmail.com

Received: 17.08.2024 Revised: 12.09.2024 Accepted: 09.10.2024

ABSTRACT
A network unauthorized accessof users refers to any action that takes place without permission on a computer network. Such undesired actions consume the network resources and threaten the security of the network. An intruder is a person who gains unauthorized access to a system, to damage the system, or to disturb medical data on that system. Generally, the objective of an intruder is to gain access to a system or to increase the range of privileges accessible on a system. Unauthorized access can be caused by outsiders who access the systems from the Internet, insiders or authorized users who seek to procure additional privileges and by authorized users who misemploy the privileges granted to them. Outsiders are adversaries with no direct access to the nodes in a network, but may have access to the physical medium. Malicious activities executed on a network or a computer system, by persons with authorized system access are called insider attacks. Insiders are usually disgruntled employees who have a grudge on the company. The Machine Learning based Unauthorized Access Detection System (ML-UADS) identifies the nodes causing outliers and then remove such kind of nodes by using trained medical data.In this research work, for keen monitoring on the network to detect unauthorized users a machine learning technique is proposed to provide security to the medical data and avoids unauthorized access more accurately. The proposed method is compared with the traditional methods and the results show that the proposed method detects unauthorized users accurately.

Keywords: Unauthorized Users, Malicious Activities, Secured Medical Data, Machine Learning Method, Restricted Permissions.

1. INTRODUCTION

Mechanisms like firewall, medical data encryption and user authentication are used by organizations as a first level of defense [1] [2]. With the emergence of new varieties of attacks, these attack prevention techniques alone are not adequate in making a system completely secure because guaranteed prevention of all kinds of security breaches is impractical. A security mechanism capable of detecting unauthorized access to system resources and medical data is mandatory [3]. An Unauthorized access Detection System (IDS) can be thought of as a second level of defense as shown in Figure 1 and is not a substitute for other security services [4]. IDS operates as part of a set of system security tools to achieve a defined level of assurance for the protection of information systems.

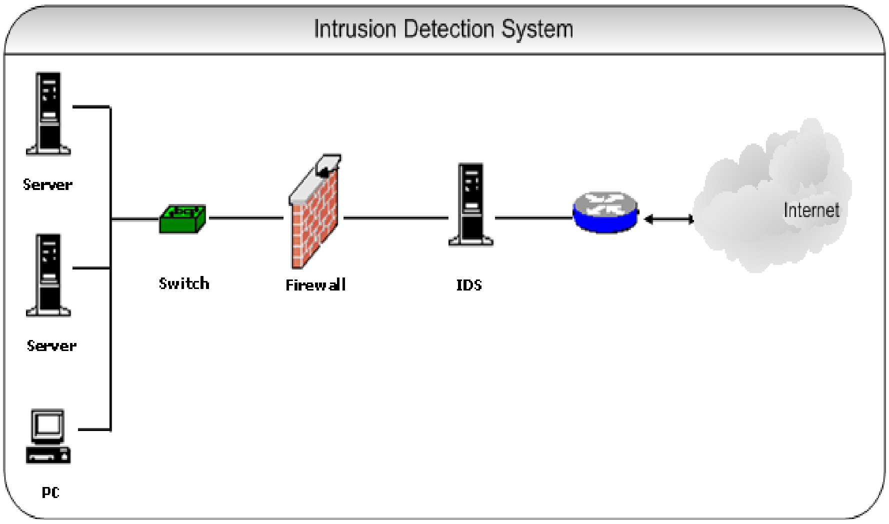

Fig 1: Unauthorized Access Detection System

A firewall looks only externally and restricts entry to networks thereby preventing outside intruders to an extent, whereas an IDS also keeps watch for attacks that emerge from within a network. IDS can sense when a system or a network is misused or is under attack [5][6]. Also, IDS automates the process of detecting malicious activities thereby helping the network administrators in monitoring the network [7] [8]. Thus, IDS extends the level of protection of the target system, resources or information and thereby plays a vital role in securing networks from intruders [24][25].

Classification of Unauthorized Access Methods

Unauthorized Access Methods (UAM) can be classified in different ways. Based on the method of deployment, UAM can be classified as Host-based IDS(HIDS) and Network-based IDS (NIDS). HIDS runs on individual devices in a network and monitors traffic to and from that particular device alone [9][10]. NIDS is placed at a point within a network from where it can monitor traffic to and from all devices in the network [11][12]. Usually it is placed at points where firewalls are placed. Hybrid approaches that use both network-based and host-based unauthorized access detection tools have also been developed [13].

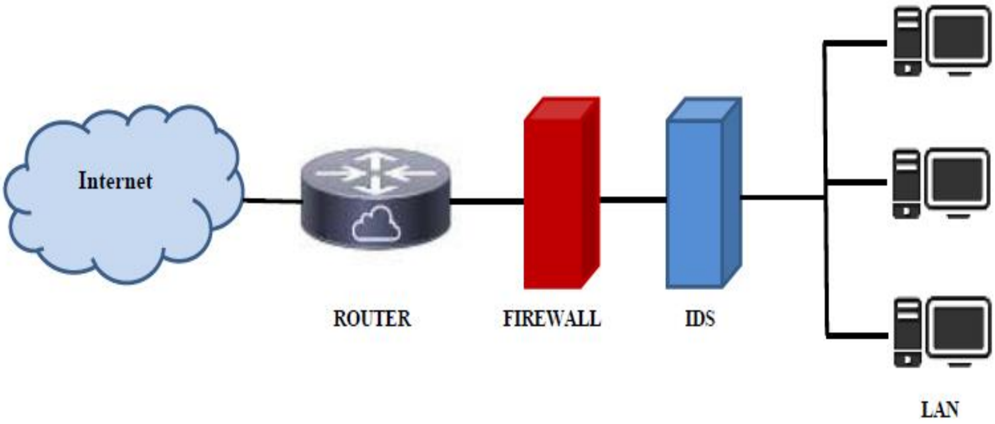

Fig. 2:A Second Level of Defense

IDS can be broadly classified as Misuse/Signature-based and Anomaly-based. Signature-based IDS maintains a database of signatures or patterns of known malicious activities and the packets on the network are compared with this database [14] [15]. If a match occurs, it will be signaled as an unauthorized access [23].

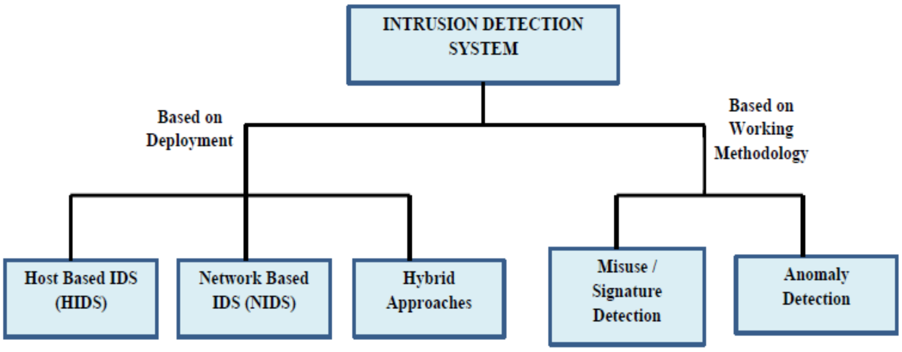

Fig. 3: Classifications of IDS

Anomaly based IDS first establishes a baseline of normal behavior of the network and if the traffic is different from the baseline, alarm will be raised [16]. Any unusual conduct will be detected by anomaly based IDS and thus has the ability to detect signs of attacks without definite comprehension of details [22]. The key advantage of such techniques is high Detection Rate (DR) because of the ability to detect previously unseen unauthorized access [17][18]. They can even detect insiders who misuse their privileges [19] [20]. But the False Alarm Rate (FAR) of such systems is high because anomalous behavior does not always signal that an attack is taking place. So, anomaly detection systems need to be carefully tuned to avoid high FAR [21].

Challenges in Unauthorized access Detection System (IDS)

- To identify unauthorized access occurring at various locations of the network.
- To monitor all the systems in the network for secure data communication.
- To detect outliers in an entire network.
- To understand the type of attack and handle different kinds of attack.
- To manage systems involved in large networks.
- To handle high dimensional medical data in a network and to provide anenvironment for data transfer in the network.

2. LITERATURE SURVEY

R S M Lakshmi Patibandla Tet al. [26] introduced an unauthorized access detection mechanism based on K-Nearest-Neighbor(K-NN) and Particle Swarm Optimization(PSO) methods to detect unauthorized access. In this, comparisons made in between simple KNN and PSO based KNN andshown accurate results [27][28].

Patibandla R S M Lakshmiet al. [29] have proposed an online oversampling principal component analysis algorithm and it aims at detecting the presence of outliers from a large amount of data via an online updating technique [30]. In this oversampling scheme, the target instance is duplicated multiple times so as to magnify the impact of outliers. The variation of principal directions on adding or removing a record is studied and this characteristic is used to decide the unauthorized access of a data point.

Hawkins et al. [31][32] have proposed a four layered framework for finding outliers from high dimensional datasets. In the training phase, the important features are selected using Entropy and for each selected attribute, z-score is computed [33]. Each training data object is classified as normal or outlier based on majority vote. The outliers are further classified into different attack classes using a Bayesian network classifier [34]. In the testing phase, the testing records are read one by one and the distance between the testing record and the previously separated class objects is calculated [35][36].

An unauthorized access recognition framework is utilized to identify a few kinds of malevolent practices that can bargain the security and trust of a PC framework [37][38]. This incorporates system attacks against defenseless administrations, information driven attacks on applications [39].There are a few different ways to order an IDS relying upon the sort also, area of the sensors and the procedure utilized by the motor to create cautions [40]. In numerous basic IDS usage every one of the three parts are consolidated in a solitary gadget or apparatus [41][42].

K.Santhi Sriet al. [43] proposed an auto-associative kernel regression(AAKR) model to detect cyber-attacks. It is a multi-layer data driven cyber-attack detection system for ICS(Industrial Control Systems).This handled five attacks including MITM(Man In The Middle), DOS(Denial Of Service), data tampering, false data injection and Data exfiltration.

K.Santhi Sriet al. [44] outlined how the standards of human immunology could be fused into a PC unauthorized access recognition system. A database of typical arrangements of framework calls was constructed and those successions which were not found in the database were considered as peculiarities.

K.Santhi Sriet al. [45] demonstrated that Unauthorized access location utilizing Elman systems gave a superior execution contrasted with profile based methods. The framework related information must be mapped into feature space and the decision of feature space depended on the application. As the framework labeled information is fine grained, overhead expands in this manner diminishing the framework execution [49].

S.Sasikalaet al. [46] proposed a Statistical Component of the Next Generation Unauthorized access Detection Expert System (NIDES) in identifying irregular program conduct. Yang Z et al. [47] developed a Non-Deterministic Push Down Automata (NDPDA) to speak to program performed and assessed the proposed structure for host-based IDS against imitation attacks. M.Venkata Raoet al. [48] structured and actualized a worldly arrangement clustering based unauthorized access recognition which is a client profile subordinate [50].

3. PROPOSED METHOD

Database malicious detection system is a log based component for the identification of malicious exchanges in the database. Malicious database exchanges are identified with security attacks completed either remotely or inside the network[51] [52]. The review log files are utilized by malicious detection system to get the succession of directions executed by every client, which is then contrasted and the profile of the approved exchanges to distinguish potential information exchanges.

The current database security interruption locations strategies are frequently in view of inconsistency recognition approach and oddity based techniques are generally inclined to creating a moderately huge number of fakes alert. Host based Outlier Detection(HODS) is a traditional method for detecting outliers in database, but it failed to detect them automatically as it needs an administrator and false positive rate is also high. To reduce false positive rate, Machine Learning based Unauthorized Access Detection System (ML-UADS) is proposed with an intellectual access supervision for network database security. This method uses an identification strategy for detecting outliers while exchanging of data inside database applications in the network.

Methods for Identification of Outlier Location

During detection process, location of outlier is needed. Commonly used outlier location identification methods are Z-Score, DBSCAN and Isolated forests.

Z-Score Method

The z-score is a metric that demonstrates what number of standard deviations an information point is from the example's mean, accepting a Gaussian appropriation. This makes z-score a parametric strategy. Frequently information indicates are not portrayed by a Gaussian dispersion, this issue can be settled by applying changes to information i.e scaling it.

Density based Spatial Clustering of Applications with Noise (DBSCAN)

DBSCAN is a density based grouping method, it is centered around discovering neighbors by density on a 'n-dimensional circle' with range ε. A group can be characterized as the maximal arrangement of 'density associated focuses' in the component space.

Isolated Forests

Its fundamental rule is that anomalies are not many and a long way from the remainder of the perceptions. To construct a tree, the calculation randomly picks an element from the component space and a split worth going between the maximums and essentials. This is mentioned for all the objective facts in the preparation set. To fabricate the forest a tree outfit is made averaging every one of the trees in the forest.

Kinds of Anomalies

There are three kinds of anomalies, such as node, linkage and sub graph exceptions.

Node exceptions are the vertices with unordinary attributes in a graph. Here, a graph is denoted as a network. They could be defined in different ways: node exceptions might be fundamentally insignificant, by being segregated from whatever is left of the vertices, or by being in the focal point of a star molded example.

Linkage anomalies are the edges with unexpected attributes in the graph. These are by and large characterized as edges that interface two divergent, yet each thickly associated segments or communities of the system.

Sub graph anomalies are characterized as parts of the network, which shows unexpected attributes in subnet.

This proposed algorithm will effectively identifies the anomalies in the network database system and enhances the security of the network database. In this, let “x” bethe initial record from dataset, “R” is the Record set, “K” is the distance from one node in network to other node and “A” is the cluster set created from “K” with all records. The proposed algorithm is used for identification is unauthorized users to provide security to the data.

Algorithm:ML-UADS

```
Step-1:  Start with the full set of attributesand null selected feature set.
Step-2:  Choose an attribute from the total set with the highest in-formation gain ratio.
Step-3:  Split the dataset into sub datasets depending on the at-tribute values.
Step-4:  Read input Matrix M
For each i in M
for each j in (M-1)
```

```
         $Kij = [ Wj* Pij + c + £err]$
    M[I,J]= Kij + Ri
    end for
end for
Step-5:  Return M[I,J];
Step-6:  Repeat step 2 to 5 for each of the sub-datasets with theset of attributes, if instance in a sub-dataset belongs to morethan one class.
Step-7:  for each test instance Ri in DS
    yi = Ŷb* $Xi$
    R(i)←yi * Kij
 End for
Step-8:  yi =$argmaxy$∈{$c1$,$c2$,…,$cn$ } $b$={1,2,..$n$} $C$b X W
Step-9:  return (yi)
Step-10: Output the selected feature set
```

4. RESULTS

Proposed Machine Learning based Unauthorized Access Detection System (ML-UADS) method is implemented in ANACONDA SPYDER and KDD CUP data set is considered for identification of outliers in the network based on intellectual access supervision.

In the proposed ML-UADS method, the parameters considered are comparison of execution time, outlier detection rate, data security level in outlier detection. Every parameter is clearly illustrated. The productivity of the proposed ML-UADS calculation is likewise contrasted with Nested Looping calculation utilizing the informational indexes. The new proposed calculations are nearly tried utilizing different informational collections. The execution of the above said calculations are pictorially represented.

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 network information focuses in X-pivot and the execution time in Y-pivot.

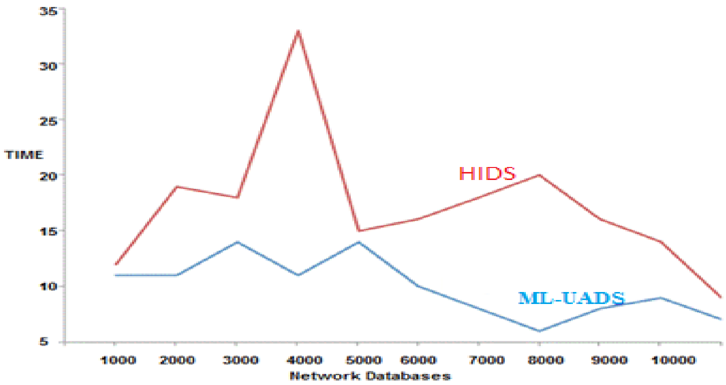

Fig. 4: Comparison of Execution Time of ML-UADS over HIDS

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

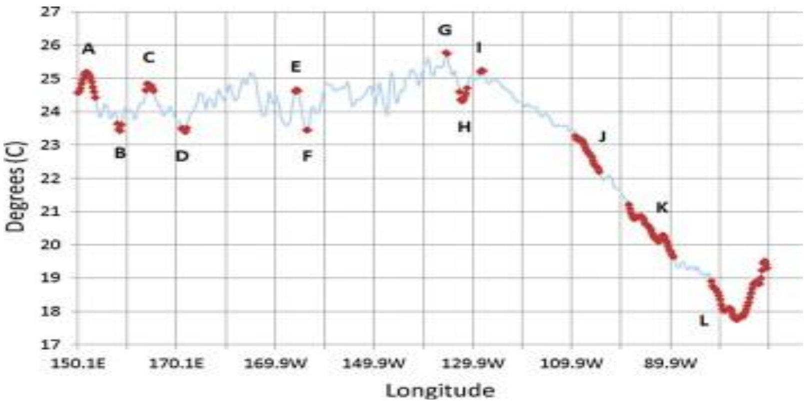

Fig. 5: Outlier Detection Rate in ML-UADS over NIDS

The Fig5 depicts the outlier detection rate of the ML-UADS method and the Host based unauthorized access detection method based on number of nodes in network. The outlier detection time in the proposed method is compared with the existing method and the results show that the proposed method is better in performance.

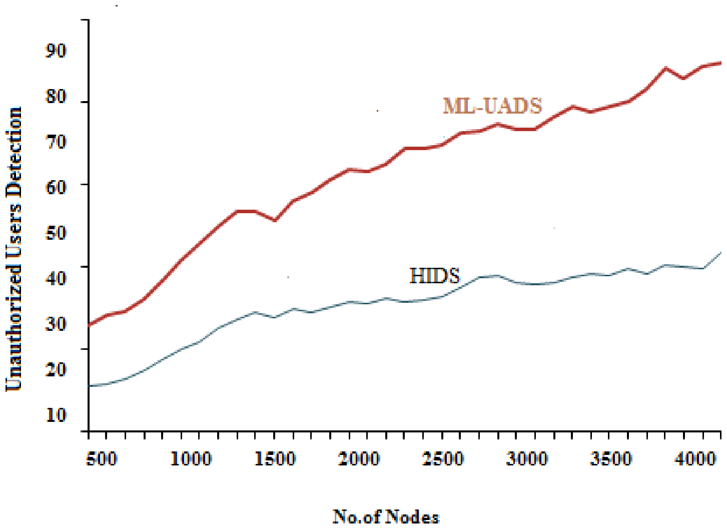

Fig. 6: Detection Rate in ML-UADS over HIDS based on no.of Nodes

Table 1: Execution Tabulation in ML-UADS with Traditional Methods

METHOD	DETECTION RATE(DR)	FALSE POSITIVE RATE (FPR)
Machine Learning based Unauthorized Access Detection System (ML-UADS)	95%	0.3
Random Forest Based Detection	67%	1.1
Density Based Detection	>70%	1.72
Weighted Distance Based Detection	88%	2.31
Reference Based Detection	87%	Not Predicted

The security for the data in the proposed ML-UADS method is better when compared to existing host based IDS. The Figure 7 illustrates the security levels of the high dimensional data. This method considers the log files as input and then calculates the distance of each node to the origin and then the unauthorized users are identified. The nodes list is separately maintained and then the normal nodes are listed out.

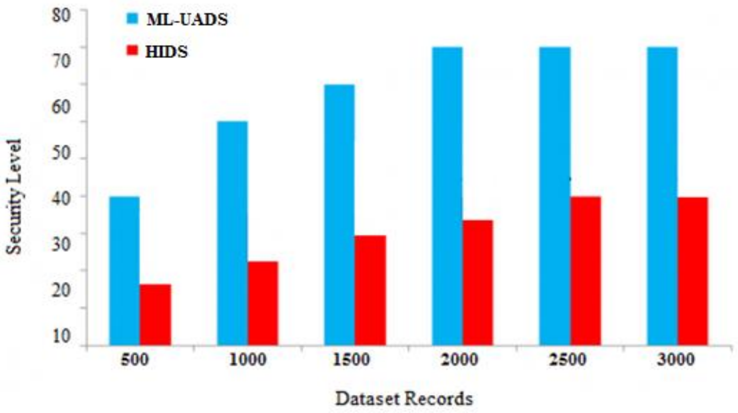

Fig. 7: Data Security Levels of ML-UADS over HIDS

In the ML-UADS method, when a dataset of network log files is given as input, the unauthorized user detection rate is high in the ML-UADS method.

5. CONCLUSION

The proposed ML-UADS method is designed speciallyto handle datasets having mixed attributes effectively and easily. It is alsopresented in a generalized manner so that it can be employed for identifying unauthorized users in any application domain. This method employs clustering in a newfashion by first grouping the data points into clusters based on the values ofcategorical attributes and then fixing the boundaries of each cluster based onthe values of numerical attributes.In this research work, asystem that comprises of Machine Learning basedUnauthorized Access Detection subsystem for securing the data has been proposed and actualized.In this, performance of ML-UADS is compared with HIDS in terms of execution time, detection rate and data security level in network.In future, the process of improving the performance of the system can be simplified by reducing the feature set considered by avoiding irrelevant features in the dataset.

REFERENCES

1. Schlegl T, Seeböck P, Waldstein SM, Langs G, Schmidt-Erfurth U. f-AnoGAN: fast unsupervised anomaly detection with generative adversarial networks. Med Image Anal. (2019) 54:30–44. doi: 10.1016/j.media.2019.01.010
2. Chandola V, Banerjee A, Kumar V. Anomaly detection for discrete sequences: a survey. IEEE TransnKnowl Data Eng. (2010) 24:823–39. doi: 10.1109/TKDE.2010.235
3. Hawkins S, He H, Williams G, Baxter R. Outlier detection using replicator neural networks. In: International Conference on Data Warehousing and Knowledge Discovery. Aix-en-Provence: Springer (2002). p. 170–80.
4. Chen J, Sathe S, Aggarwal C, Turaga D. Outlier detection with autoencoder ensembles. In: Proceedings of the 2017 SIAM International Conference on Data Mining. SIAM (2017). p. 90–8.
5. Sabokrou M, Fayyaz M, Fathy M, Moayed Z, Klette R. Deep-anomaly: fully convolutional neural network for fast anomaly detection in crowded scenes. Comput Vision Image Understand. (2018) 172:88–97. doi: 10.1016/j.cviu.2018.02.006
6. Zimek A, Filzmoser P. There and back again: outlier detection between statistical reasoning and data mining algorithms. Wiley Interdiscipl Rev Data Min KnowlDiscov. (2018) 8:e1280. doi: 10.1002/widm.1280
7. Knox EM, Ng RT. Algorithms for mining distancebased outliers in large datasets. In: Proceedings of the International Conference on Very Large Data Bases. Citeseer (1998). p. 392–403.
8. Ramaswamy S, Rastogi R, Shim K. Efficient algorithms for mining outliers from large data sets. In: ACM Sigmod Record. Vol. 29. ACM (2000). p. 427–38.
9. Angiulli F, Pizzuti C. Outlier mining in large high-dimensional data sets. IEEE Trans Knowl Data Eng. (2005) 17:203–15. doi: 10.1109/TKDE.2005.31
10. Angiulli F, Fassetti F. Dolphin: an efficient algorithm for mining distance-based outliers in very large datasets. ACM Trans KnowlDiscov. Data. (2009) 3:4. doi: 10.1145/1497577.1497581
11. Yang Z, Wu C, Chen T, Zhao Y, Gong W, Liu Y. Detecting outlier measurements based on graph rigidity for wireless sensor network localization. IEEE Trans Vehicul Technol. (2012) 62:374–83. doi: 10.1109/tvt.2012.2220790
12. Abukhalaf H, Wang J, Zhang S. Mobile-assisted anchor outlier detection for localization in wireless sensor networks. Int J Future Gen CommunNetw. (2016) 9: 63–76. doi: 10.14257/ijfgcn.2016.9.7.07
13. Abukhalaf H, Wang J, Zhang S. Outlier detection techniques for localization in wireless sensor networks: a survey. Int J Future Gen CommunNetw. (2015) 8:99–114. doi: 10.14257/ijfgcn.2015.8.6.10
14. Tenenbaum JB, De Silva V, Langford JC. A global geometric framework for nonlinear dimensionality reduction. Science. (2000) 290:2319–23. doi: 10.1126/science.290.5500.2319
15. Pang Y, Yuan Y. Outlier-resisting graph embedding. Neurocomputing. (2010) 73:968–74. doi: 10.1016/j.neucom.2009.08.020
16. Schubert E, Gertz M. Intrinsic t-stochastic neighbor embedding for visualization and outlier detection. In: International Conference on Similarity Search and Applications. Springer (2017). p. 188–203.
17. Lakshman Narayana Vejendla and A Peda Gopi, (2019),” Avoiding Interoperability and Delay in Healthcare Monitoring System Using Block Chain Technology”, Revue d'IntelligenceArtificielle, Vol. 33,No. 1, 2019,pp.45-48.
18. A Peda Gopi and Lakshman Narayana Vejendla, (2019),” Certified Node Frequency in Social Network Using Parallel Diffusion Methods”, Ingénierie des Systèmes d' Information, Vol. 24,No. 1, 2019,pp.113-117. DOI: 10.18280/isi.240117
19. Lakshman Narayana Vejendla and Bharathi C R,(2018),“Multi-mode Routing Algorithm with Cryptographic Techniques and Reduction of Packet Drop using 2ACK scheme in MANETs”, Smart Intelligent Computing and Applications, Vo1.1, pp.649-658.DOI: 10.1007/978-981-13-1921-1_63DOI: 10.1007/978-981-13-1921-1_63

20. Lakshman Narayana Vejendla and Bharathi C R, (2018), “Effective multi-mode routing mechanism with master-slave technique and reduction of packet droppings using 2-ACK scheme in MANETS”, Modelling, Measurement and Control A, Vol.91, Issue.2, pp.73-76.DOI: 10.18280/mmc_a.910207
21. Lakshman Narayana Vejendla and Bharathi C R,(2017),“Using customized Active Resource Routing and Tenable Association using Licentious Method Algorithm for secured mobile ad hoc network Management”, Advances in Modeling and Analysis B,Vol.60, Issue.1, pp.270-282. DOI: 10.18280/ama_b.600117
22. Madabhushi A, Shi J, Rosen M, Tomaszeweski JE, Feldman MD. Graph embedding to improve supervised classification and novel class detection: application to prostate cancer. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. Berlin, Heidelberg: Springer (2005). p. 729–37.
23. Cook DJ, Holder LB. Graph-based data mining. IEEE Intell Syst Appl. (2000) 15:32–41. doi: 10.1109/5254.850825
24. Eberle W, Holder L. Anomaly detection in data represented as graphs. Intell Data Anal. (2007) 11:663–89. doi: 10.3233/IDA-2007-11606
25. Rahmani A, Afra S, Zarour O, Addam O, Koochakzadeh N, Kianmehr K, et al. Graph-based approach for outlier detection in sequential data and its application on stock market and weather data. Knowl Based Syst. (2014) 61:89–97. doi: 10.1016/j.knosys.2014.02.008