Turkish Online Journal of Qualitative Inquiry (TOJQI)

Volume 12, Issue 5, June 2021: 1422 - 1442

Research Article

Roc Curve Based Muti_Objective Pso For Outlier Detection Using Hybrid Dragon Fly Optimization

[1] B.Angelin , [2] Dr.A.Geetha

ABSTRACT

Outlier detection is a fundamental advance in the information mining measure. Its principle reason to eliminate the contrary information from the first information. This cycle helps in the expulsion of information that are vital for doing to accelerate the applications like order, information annoyance, and pressure. It assumes a significant part in the climate guaging, execution examination of sportsperson, and organization interruption location frameworks. The outlier for the single variable can be effectively noticed however for the n-variable, it turns into a drawn-out measure. To improve the presentation of outlier detection in n-variable or traits a few techniques were proposed. A portion of the current strategies are factual methodologies, vicinity based measures, arrangement approaches, and record based methodologies, and optimization based methodologies. The initial four methodologies couldn't characterize the data when there is a defect in the names. However, the optimization based methodology can beat this issue even there is blemished marking. One of the current improvement approaches is Particle swarm optimization. The current strategy neglected to deal with the bigger records and more modest properties. To conquer this issue a hybrid dragon fly PSO and multi-layer feed-forward neural organization are proposed. This goal is accomplished with the assistance of the ROC bend (negative proportion) as the goal work.

Keywords: outlier data mining, multi objective PSO, ROC,hybrid dragon fly.

I. Introduction

The traditional methodology is appropriate for more modest datasets with a predetermined number of measurements. Be that as it may, for the bigger datasets or high dimensional, it deals with an issue because of the distance and contrast between the dataset is exceptionally little.

[1]Research Scholar (Ph.D), [2]Assistant Professor & Head in Computer Science,
[1, 2,] Pg&Research Department of Computer Science,
[1, 2] Chikkanna Government Arts College, Tirupur.
Mail : angelindia93@gmail.com , gee_sam@yahoo.com

This issue is overwhelmed with the assistance of sub-space learning, outfit learning, and grouping strategies . It tried on ongoing datasets like page blocks, arrhythmia, thyroid, and so forth the K-NN and Subspace-based procedures are profoundly appropriate for deciding the exceptions.[1]

Innovation headways in medical services informatics, digitalizing wellbeing records, and telemedicine has brought about fast development of medical care information. One test is instructions to viably find valuable and significant data out of such enormous measure of information through procedures like information mining. outlier location is a normal procedure utilized in numerous fields to dissect large data. Not with standing, for the enormous scope and high dimensional heath care information, the regular exception recognition techniques are not productive. This paper proposes a novel half breed exception discovery technique, in particular, Pruning-based K-Nearest Neighbor (PB-KNN), which incorporates the thickness based, cluster based strategies and KNN calculation to direct powerful outlier recognition. The proposed PB-KNN embraces the case arrangement quality character (CCQC) as the clinical quality assessment model and uses the characteristic covering rate (AOR) calculation for information grouping and dimensionality decrease. To assess the exhibition of the pruning tasks in PB-KNN, we direct broad investigations. The analysis results show that the PB-KNN technique beats the k-closest neighbor (KNN) what's more, neighborhood exception factor (LOF) regarding the exactness and productivity.[2]

In this proposed calculation to finding the reliable information focuses and get a weighted closeness score as consistency score for every data point. this theory is that the most steady data focuses will have a high consistency score. They additionally suggest that the information focuses with helpless consistency scores are bound to be exceptions. In this manner, they flag the ones with helpless score as deceitful exchanges. Nonetheless, since our calculation is an outfit of k-means for different upsides of k, it runs into same issues that k-mean calculation does. It can conceivably run into neighborhood minima and can be delicate to the request in the information is introduced. In this way, they randomize our datasets and run our calculation multiple times to get an gauge of fluctuation these issues can bring about. They report standard deviation in each trial to evaluate this difference. Then they too standardize datasets while running analyses along the component measurements as a component of the preprocessing. Finally all through their examinations and investigation they will allude exceptions as positives and inliers as negatives.[3]

Another author is build up a proposed framework The valuation of outlier recognition algorithm be a steady debate in data mining examine. Little is known in regards to strength and shortcoming of disparate customary outlier detection techniques, in this paper We play out a broad exploratory investigate the exhibition of an agent set of enhancement based exception identification across a tremendous sort of datasets arranged consequently dependent on the general execution of the exception location strategies, we likewise propose extra fitting for the assessment of exception discovery results. We utilized dragonfly and k mean calculations and advancement strategy used to identify the outlier dependent on ROC Curve technique.[4]

To improve the exhibition of outlier location in n-variable or properties a few strategies were proposed. A portion of the current techniques are factual methodologies, nearness based measures, order approaches, and list based methodologies and streamlining based methodologies. The initial four methodologies couldn't arrange the data when there is a blemish in the names. Be that as it may, the advancement based methodology can defeat this issue even there is a defective marking. One of the current improvement approach is particle swarm optimization based multi objective technique. The current technique neglected to deal with the bigger records and more modest attributes. Outlier discovery is a fundamental advance in the information mining measure. Its principle reason to eliminate the contradictory data from the first information. This cycle helps in the expulsion of information which are vital for completing to accelerate the applications like clustering, data annoyance and pressure. It assumes a significant part in the climate estimating, execution examination of sportsperson and organization interruption recognition frameworks. The outlier for the single variable can be effortlessly noticed yet for the n-variable it become a monotonous process.To beat this issue a mythical beast hybrid dragon fly based PSO and multi-layer feed forward neural organization is proposed. This goal is accomplished with the assistance of ROC bend (negative proportion) as target work. The exhibition is assessed utilizing location rate.

II. Literature survey

B.Angelin [2021] develop a system that purpose Dragon fly K-means algorithm is to discover the optimum information with the larger records and smaller attributes with high classification rate. The high classification or detection rate is very important for good statement and intrusion detection. This objective is achieved by mistreatment the mythical monster property negative quantitative relation of the classifier as its objective or fitness operate for dragon fly k-means algorithm. In this, the planned methodology is tested on the clinical datasets which needs high classification rate for discover the wellness at associate degree early stage.[5]

Rajendra Pamula (2011) is plan a proposed framework a clustering based strategy to catch outliers . We apply K-implies clustering algorithm to isolate the data index into bunches. The focuses which are lying close to the centroid of the bunch are not plausible contender for exception and we can prune out such focuses from each group. Next we ascertain a distance based exception score for outstanding focuses. The calculations expected to figure the outlier score diminishes extensively because of the pruning of certain focuses. In light of the exception score we pronounce the top n focuses with the most noteworthy score as outliers. The exploratory outcomes utilizing genuine informational index exhibit that despite the fact that the quantity of calculations is less, the proposed technique performs better compared to the current strategy.[6]

B. Angelin (2020) is build up a proposed framework The valuation of anomaly recognition algorithm be a steady debate in information mining examine. Little is known in regards to strength and shortcoming of disparate customary outlier recognition technique s, in this paper We play out a broad exploratory investigate the exhibition of an agent set of enhancement based exception identification across a tremendous sort of datasets arranged consequently dependent on the general execution of the exception location strategies, we likewise propose extra fitting for the assessment of exception discovery results. We utilized dragonfly and k means calculations and advancement strategy used to identify the outlier dependent on ROC Curve technique.[7]

J.Antony Jeyanna , E. Indhumathi (2015) develop a proposed algorithm, that is about Outlier Detection for Network Intrusion Detection System using clustering. here two-level IDS is suggested that is equipped for identifying network assaults with a serious level of exactness. The execution of this IDS utilizes two degrees of assault recognition: a directed technique, and an outlier based strategy. The choice of managed or outlier based classifier at a specific level for a given dataset depends on the order exactness of the individual classifier for a given dataset. The main degree of order orders the test information into three classifications DoS, test, typical and

Rest (unclassified). U2R and R2L associations are delegated Rest at this stage. The primary reason at level one is to extricate however many DOS and Probe associations as could be expected under the circumstances precisely from the information utilizing a directed classifier model. At level two, the Rest classification is named U2R and R2L assaults utilizing a This IDS has a high successful generally execution for referred to just as obscure assaults. it is a blend of the an administered classifier dependent on an unmitigated clustering calculation and a regulated exception outlier identification model location calculation dependent on symmetric neighborhood relationship. The outcomes are broke down with the benchmark.[8]

Yue Fei Wang (2018) develop a proposed methodology that is OPTICS based new outlier method. Here OPTICS is a thickness based clustering strategy that can address point sets with Acknowledged manuscript different densities; nonetheless, the exception identification capacity of OPTICS is restricted by a few factors, for example, unique boundary and distinctive point set shapes. Hence, an exception location technique dependent on OPTICS was proposed, known as OD-OPTICS, which adds a pre-handling measure and alters the CD processing technique. In the first place, the Radius Filtration Strategy, which gives the key radii, is performed; and the essential distances of the point set are reflected. At that point, to filtrate invalid radii and choosing the most suitable one, the Covering Space Model was proposed. The outcomes exhibited that the recognition execution of OD-OPTICS is better than that of OPTICS.[9]

Yutao wang, His-Yung Feng (2015) develop a proposed system, that is outlier detection for scanned point cloud is used by majority voting. In that system,When filtering an article utilizing a 3D laser scanner, the gathered examined point cloud is typically contaminated by various estimation outliers. These outlier can be inadequate exceptions, disengaged or non-detached exception groups. The non-secluded anomaly bunches represent an extraordinary test to the improvement of an automatic exception location strategy since such outliers are connected to the filtered information focuses from the article surface and hard to be recognized from these legitimate surface estimation focuses. Here this system presents a viable outlier discovery technique dependent on the guideline of lion's share casting a ballot. The strategy can detect non-disengaged anomaly bunches just as different kinds of exceptions in a filtered point cloud.[10]

Jihyun Ha, Seulgi Seok (2014) are develop a proposed system that is robust Outlier detection using instability factor. Here this technique is based on the focal point of gravity, which brings about a novel measure that scores the level of outlierness. they named this new measure the insecurity factor. An article with a high shakiness factor is a promising possibility for an outlier. Dissimilar to existing distance-based and thickness based methodologies, our strategy is vigorous regarding changes in the model parameter k. Not with standing the power, it very well may be utilized deftly for both nearby and worldwide identification of exceptions by controlling the boundary. Our proposed strategy offers a helpful precariousness plot that can provide data about information, like the number and size of clusters. Utilizing five manufactured datasets, so here affirmed that the technique precisely recognizes exceptions from information with low thickness patterns just as various levels of bunch thickness. The outcomes show that the new approach beats the neighborhood thickness furthermore, low thickness designs issue that have been problematic in existing approaches. Results from the genuine datasets additionally demonstrated the adequacy of this proposed strategy.[11]

Usman Habib , Gerhard Zucker (2015) are develop a new approach that is outliers detection method in this method here using clustering in buildings data. Here to accomplish energy effectiveness in structures, a great deal of crude information is recorded, during the activity of structures. This recorded crude information is additionally utilized for the examination of the execution of structures and its various segments for example Warming, Ventilation and Air-Conditioning (HVAC). To save time furthermore, energy it is needed to guarantee flexibility of the information by distinguishing and supplanting exceptions (for example information tests that are definitely not conceivable) in the information before Nitti gritty investigation. Here talks about the means required for identifying exceptions in the information gotten from ingestion chiller utilizing their On/Off state data. It additionally proposes a technique for programmed discovery of On/Off and additionally Missing Data status of the chiller. That method utilizes two layer K-Means clustering for identifying On/Off just as Missing Data condition of the chiller. After programmed location of the chiller On/Off cycle, a strategy for exception recognition is proposed utilizing Z-Score standardization dependent on the On/Off cycle condition of chillers and bunching anomalies by Expectation Maximization bunching calculation. Besides, the aftereffects of filling the missing values with relapse and direct interjection for short and long periods are explained in this paper.[12]

Petra J. Jones , Matthew K., (2020) are develop a proposed approach for outlier detection, here they use k-means clustering of physical activity. Here, another calculation signified as FilterK is proposed for improving the immaculateness of k-implies determined actual work groups by lessening anomaly impact. Here they applied it to actual work information acquired with body-worn accelerometers and clustering utilizing k-means . Then they contrasted its exhibition and three existing anomaly location techniques: Local Outlier Factor, Isolation Forests and KNN utilizing the ground truth (class marks), normal bunch and occasion virtue (ACEP). FilterK gave practically identical gains in ACEP (0.581 → 0.596 looked at to 0.580–0.617) while eliminating a lower number of anomalies than different strategies (4% complete dataset size versus 10% to accomplish this ACEP). The primary focal point of our new anomaly identification technique is to improve the bunch purities of actual work accelerometer information, however they additionally propose it could be conceivably applied to different sorts of dataset caught by k-means clustering. So they exhibit their strategy utilizing a k-means model prepared on two free accelerometer datasets (preparing n = 90) and re-applied to an autonomous dataset (test n = 41). Marked proactive tasks incorporate resting, sitting, standing, family errands, strolling (research facility and non-laboratory based), steps and running. This kind of grouping calculation could be utilized to help with distinguishing ideal actual work designs for wellbeing.[13]

Harshada, C. Mandhare (2017) are develop a proposed system for study of outlier detection techniques. Here, there is an expanding request of information, outlier discovery is seeming to be a mainstream field of examination. Exception is expressed as a perception which is disparate from the other perceptions present in the informational index. It is worthwhile in the fields like clinical industry, wrongdoing recognition, deceitful exchange, public wellbeing and so forth Exception can be learnt in various fields like huge information, time arrangement information, high measurement information, organic information, unsure information and some more. More often than not 10% of the entire example informational index is inaccurate, not available or missing now and again. This paper studies and thinks about the well-known exception discovery calculations in particular, Cluster based outlier detection, Distance based exception location and Density based anomaly discovery. Near investigation of these exception identification strategies is performed to discover most effective outlier discovery strategy for estimation of the exception.[14]

III. Existing Technique

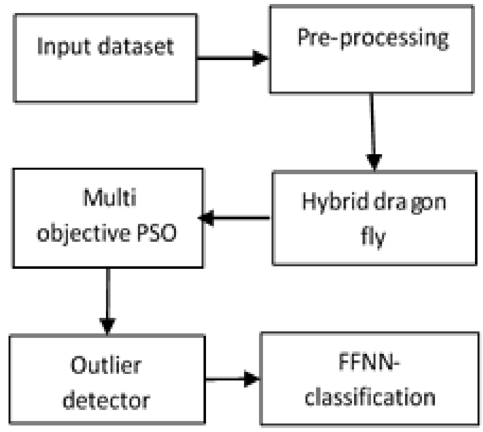
Outlier detection is a basic development in the data mining measure. Its standard motivation to dispose of the opposite data from the main data. This cycle helps in the removal of data that are indispensable for doing to speed up the applications like request, data irritation, and pressing factor. It accepts a critical part in the environment gauging, execution assessment of sportsperson, and association interference area structures. The inconsistency for the single variable can be successfully seen any way for the n-variable, it transforms into a drawn-out measure. To improve the introduction of irregularity disclosure in n-variable or attributes a couple of methods were proposed. [5][7]A part of the current systems are real procedures, area based measures, game plan approaches, and record based techniques, and smoothing out based philosophies. The underlying four approaches couldn't portray the data when there is an imperfection in the names. Nonetheless, the smoothing out based strategy can beat this issue even there is imperfect stamping. One of the current improvement approaches is Particle swarm optimization. The current methodology fails to manage the greater records and more humble properties. To overcome this problem a dragon fly based K-means clustering and multi-layer feed forward neural network is used in this existing method. This objective is refined with the help of the ROC curve (negative extent) as the objective work.[15]

The primary target of this existing Dragon fly K-mean clustering is to recognize the ideal information with the bigger records and more modest properties with high characterization rate. The high characterization or discovery rate is significant for wonderful estimating and interruption recognition. This goal is accomplished by utilizing the roc property negative proportion of the classifier as its goal or wellness work for mythical serpent fly k-mean clustering. In this, the existing technique is tried on the clinical datasets which requires high order rate for recognize the illness at an early stage. The existing strategy is tried on the clinical datasets which requires high characterization rate for identify the sickness at a beginning phase. The clinical curved seizure dataset which is downloaded from the UCI AI storehouse dataset. [16] The Iris blossom dataset is likewise utilized for assessing the proposed strategy performance. The pre-prepared information is handled utilizing mythical beast fly k-means clustering process to decide the ideal grouped data which eliminates the outlier from the dataset. The working of dragon fly calculation is propelled from However, the target capacity of the calculation is the negative proportion property of the roc-curve of the classifier.[17]

IV. Proposed System

The proposed system is hybrid dragon fly outlier detection using multi objective particle swarm optimization firstly the input data set is give to the pre-processing in this technique removed unwanted datas. Then it goes to hybrid dragon fly and multi objective PSO process this multi objective –PSO improve the ROC curve. Then outlier detector detect outlier, finally here use feed-forward neural network. The brief details about the all the blocks are discussed in below**.**

Block diagram



Input Dataset:

The proposed strategy is tried on the clinical datasets which needs high classification rate for identify the disease at an early stage. The clinical elliptical seizure dataset is downloaded from the UCI machine learning repository database. The Iris blossom dataset is additionally utilized for assessing the proposed strategy execution.

Pre-Processing

The pre-processing step to change the information into an appropriate configuration over to play out the proposed technique measure and to eliminate the insignificant data from the dataset. The interaction in the pre-preparing step is appeared in the underneath figure.

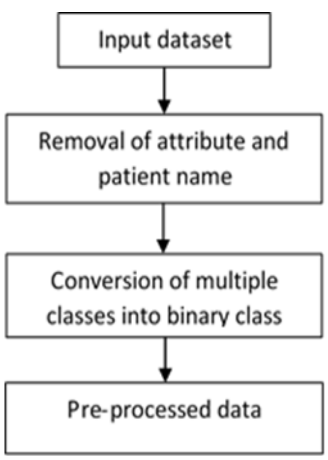


Figure 1 pre-processing process

In this, the pre-preparing steps are acted in two stages. The initial step is to eliminate the quality and patient name from the dataset. The subsequent advance is to change over the multi-classes into parallel classes for a simple anomaly recognition measure. The numerous classes are changed over into double classes as typical and unusual. This cycle happens on the dataset and the pre-prepared dataset is utilized for the dragonfly K-means clustering.

Hybrid Dragon Fly Algorithm

The main objective of any swarm is endurance. To accomplish this objectives, all the individuals should be focusing toward food sources and additionally diverted toward adversaries.. The proposed work utilizes a mixture release of the traditional dragonfly calculations. [18] The two calculations that are blended are the GA and the dragonfly calculation. The means engaged with the HDA are clarified below

Stage 1: Initializations of the general population and the progression vector and plan the targets work.

Stage 2: Evaluate the upsides of S, A, C, E, and F from the conditions referenced here. The detachment implies the static collision avoidance of the individuals as of others around there. The partition is registered as follows.

$$S = -\sum_{i=1}^{M} \quad X - Xi \tag{1}$$

Where Y is the situation of the current individual, Xi shows the position i-th adjoining individual, and N is the quantity of adjoining people. Arrangement shows the speed comparing of people with that of the others on the area. Arrangement is computed as underneath.

$$A = \frac{\sum_{i=1}^{M} \quad Vi}{M} \tag{2}$$

Where Vi addresses the speed of the i-th adjoining person. Attachment alludes to the tendency of people toward the centre of the mass of the area. The union is registered as follows

$$C = \frac{\sum_{i=1}^{M} \quad Xi}{M} - X \tag{3}$$

The fascination for a food source is computed as underneath.

$$F = X' - X \tag{4}$$

In which, X shows the places of the current individual, X' means the situation of the food sources. The Distraction toward and foe is figured as demonstrated here.

$$F = X'' - X \tag{5}$$

Where, X is the "position" of the current individual and X" is the "position" of the enemy.

Stage 3: To up-date the dragon fly flies position on a hunt space and additionally recreate their developments, 2 vectors are required that is the progression what's more, position vector. The progression vector demonstrates the course of the dragonfly's development moreover is addressed as given here.

$$\Delta Xt+1 = (sS + aA + cC + F + eE) + w\Delta Yt \tag{6}$$

partition of the Where s is the division weight, S indicates the j-th individual, an addresses the arrangement weight, An infers the arrangement of the j-th singular, c union weight, C addresses j the union of the j-th singular, f is the food factor, F signifies the food wellspring of the j-th singular, e is the enemy factor, E addresses the situation of the enemy of the j-th individual, demonstrates the weight of dormancy, and t indicates the cycle counter.

Stage 4: The positions vector can well be registered from the progression vector esteem. It is determined as follows.

$$X_{t+1} = X_t + \Delta X_{t+1} \quad (7)$$

Where t is the present iteration

Stage 5: The places of the dragonfly is refreshed using the ensuing condition.

$$X_{t+1} = X_t + L(d_i) . X_t \quad (8)$$

In which, t infers the current emphasis, di infers the measurements of the position vector, and furthermore L shows the Levy flight.

Step 6: If the stop models are unsatisfied, at that point two hereditary administrators are added. Hybrid and change are applied when the dragonfly doesn't have as an insignificant of one adjoining dragonfly. This makes the advancement more successful. The kind of hybrid that is utilized is the 2 point hybrid. This is finished utilizing the hybrid focuses.

$$m1 = \frac{|Xt+1|}{3}$$

$$m2 = m1 + \frac{|Xt+1|}{2} \quad (9)$$

Where m1 and m2 are the two focuses that are picked as hybrid focuses.

Stage 7: Mutation is performed by supplanting a few qualities from each chromosome with new qualities. The traded qualities are self-assertively created qualities with no reiteration inside the chromosome. Here, the chromosomes are the assortment of boundaries that characterize the arrangement.[19][20] The proposed calculation's flowchart is shown in

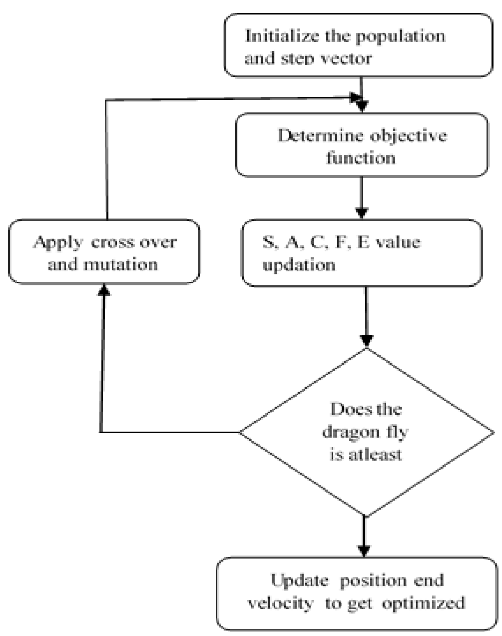


Figure 2 flowchart of hybrid dragon fly

Multi-objective Particle Swarm Optimization

Let $S \subset R$ n be a n-dimensional pursuit space, and f_i (x), i=1,… ,k, be k target capacities characterized over S. Additionally, let f be a vector work characterized as

$f_i(x) = [f_1(x), f_2(x), f_3(x)……, f_k(x)]$, (10)

and

$g_i(x) \leq 0$, i=1,……n, (11)

be m disparity imperatives. At that point, we are interested in discovering an answer, x* = (x1 *, x2 *,… , xn *), that limits f(x). The target capacities f I (x) might be clashing with one another, in this manner rendering the recognition of a solitary worldwide least at a similar point in S, incomprehensible. For this purpose, optimality of an answer in multi-objective issues should be re-imagined appropriately. Let u = (u1 ,… , uk) and v = (v1,… , vk) be two vectors of the pursuit space S. At that point, u overwhelms v, if and just if, ui ≤ vi for all i=1, 2,… , k, and ui < vi for at any rate one segment. This property is known as Pareto strength. An answer, x, of the multi-target issue is supposed to be Pareto ideal, if and just if there could be no other arrangement, y, in S with the end goal that f(y) rules f(x). [21][22]For this situation, we likewise say that x is non dominated with deference to S. The arrangement of all Pareto ideal arrangements of a issue is known as the Pareto ideal set, and it is typically meant as P*. The set

$PF = \{ f_i(x): x \in P^* \}$, (12)

This above equation is called Pareto front. This pare to front is convex if and only if, for all u, v ∈ PF* and for all λ ∈ (0, 1), there exists a w ∈ PF* such that

$\lambda ||u|| + (1-\lambda) ||v|| \geq ||w||$, (13)

while it is called concave, if and only if

$\lambda ||u|| + (1-\lambda) ||v|| \leq ||w||$ (14)

Particle Swarm Optimization

PSO abuses a populace, called a multitude, of likely arrangements, called particles, which are adjusted stochastically at every emphasis of the calculation. Be that as it may, the control of multitude contrasts fundamentally from that of developmental calculations, advancing an agreeable instead of a cutthroat model. All the more explicitly, rather than utilizing express transformation and choice administrators to adjust the populace and favour the best performing people, PSO utilizes a versatile speed vector for every molecule, which moves its position at every cycle of the calculation.[23] The particles are moving towards promising districts of the search space by misusing data springing from their own insight during the inquiry, as well as the experience of different particles. For this reason, a different memory is utilized where each molecule stores the best position it has at any point visited in the inquiry space.[24]

let us presently to put PSO all the more officially in the setting of single-target enhancement. Let S be a n-dimensional hunt space, f : S → R be the goal capacity, and N be the quantity of particles that include the multitude,[25]

$S = \{x_1, x_2, x_3 \ldots, x_N\}.$ (15)

i is represent parcel in the search space

$x_i = (x_{i1}, x_{i2}, \ldots, x_{in}) \in S,$ (16)

P represent best position

$p_i = (p_{i1}, p_{i2}, \ldots, p_{in}) \in S$ (17)

v is velocity of the particle

$v_i = (v_{i1}, v_{i2}, \ldots, v_{in}).$ (18)

$v_{ij}(t+1) = w\, v_{ij}(t) + c1\, r1\, (p_{ij}(t) - x_{ij}(t)) + c2\, r2\, (p_{gj}(t) - x_{ij}(t)),$ (19)

$x_{ij}(t+1) = x_{ij}(t) + v_{ij}(t+1),$ (20)

where $i = 1, 2, \ldots n$; $j = 1, 2, \ldots N$;

w is a positive boundary called latency weight; c1 and c2 are two positive constants called intellectual and social boundary, individually; and r1, r2, are acknowledge of two free arbitrary factors that accept the uniform appropriation in the reach (0, 1). The best situation of every molecule is refreshed at every emphasis by setting.[26]

$p_i\,(t+1) = x_i\,(t+1)$, if $f(x_i) < f(p_i)$ (21)

Otherwise it remains unchanged. Obviously, an update of the index g is also required at each iteration.

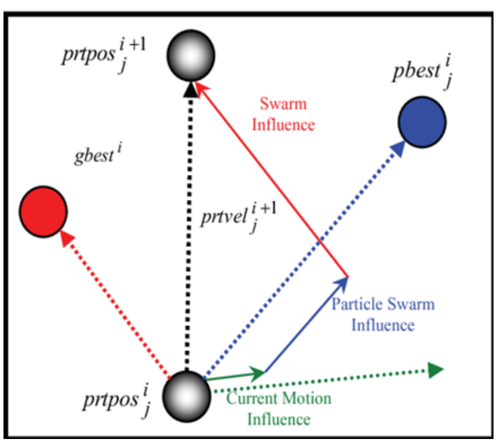


Figure 3 multi-objective PSO

Feed forward neural network (ML-FFNN)

The idea data from the hybrid dragon fly based multiple objective particle swarm optimization utilized for the characterization and furthermore to assess the exception discovery execution. The particle interaction is performed utilizing the multi-facet feed forward neural organization. The multi-layer feed forward neural organization comprises of three layers specifically Input layer, covered up layer and yield layer. The Multi-layer feed forward neural network utilizes the back engendering calculation which is valuable for preparing by decreasing the mean square mistake between the yields and targets which is proliferated in reverse from the yield layer to the input layer. By diminishing the mean square mistake, the identification pace of the order is expanded.[27][28]

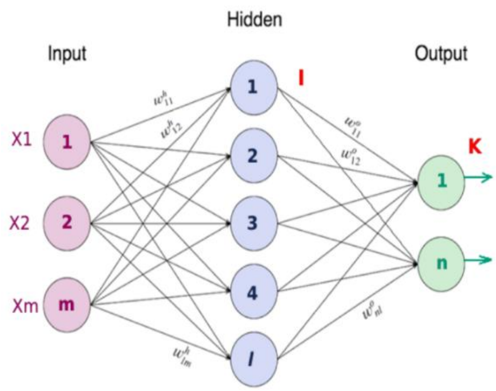


Figure 4 feed forward neural network

The input layer comprises of the information loads dependent on the traits which is meant as far as i neurons. The hidden layer neurons are meant by j neurons. The association between the info and covered up layer is made through the weight work W_{ij} which is given by the accompanying condition

$$wjk^{(n+1)} = W_{ij}{}^{n} - \lambda(\frac{\partial E}{\partial W_{jk}})^{n}$$
(22)

Where λ is the learning coefficient which is consistently more prominent than nothing and Y is demonstrates the output layer. The output K is gotten from the secret layer utilizing the accompanying conditions.[29]

$$K = f(\mathcal{E}_j) \qquad (23)$$

Where $f(\mathcal{E}_i)$ the exchange work for proliferating the sign from the jth neuron to ith neuron. The $(\mathcal{E}_i)$ is the possible capacity of the info neuron which is given by

$$\mathcal{E}_j \qquad (24)$$

$$= \vartheta_j \sum_{k\epsilon} w_{jk}.x_k$$

Where ϑ_i the edge coefficient and W_{ij} is the weight coefficient which changes till it limit the mean square blunder between the necessary output and anticipated output through back engendering measure. The contribution of the feed forward neural organization is the information from the ideal bunch and the hidden layer is 10 and the output are the objectives of the information. The organization is prepared with th3 70% of ideal information through cross-hold approval measure and tried with the 30% of information. The ordered output is assessed utilizing recognition rate.[30][31]

Outlier Detection:

In this, the outlier data detection is straightforwardly by applying in below the condition to the segment of the optimized data. The recipe used to recognize the outlier is as per the following: [32][33]

$$outlier = 1.4826 * median\big(abs(D) - median(D)\big) \quad (25)$$

Evaluation Metrics

The outlier execution is assessed utilizing characterization of datasets as ordinary and strange. The multi-objective PSO execution is assessed utilizing the identification or exactness pace of the classifier as follows.

$$detection\ rate = \frac{correctly\ identified\ classes}{total\ number\ of\ classes} \quad (26)$$

In light of this, the proposed strategy execution is assessed on three datasets and it is contrasted and the current strategies. The consequence of the proposed strategy is clarified intricately in the accompanying segment.

V. Experimental Results and discussion:

The proposed technique is recreated utilizing Matrix Laboratory programming R2018a variant in windows 10 climate.

The proposed philosophy is tried on the three kinds of straightforwardly accessible datasets specifically Arrhythmia, diabetes and curved seizure datasets from the UCI AI safe site. The depiction of the dataset is given beneath.

Tab1. Input dataset characteristics

Parameter	**Elliptical seizure**	**IRIS flower**
Number of records	11500	150
Number of attributes	178	4
Actual number of classes	5	3
Converted number of classes	2	3
Existing techniques	NO	No

applied		

At first, the information is isolated as data sources and focuses for handling. At that point the info information is multi-objective based PSO process and from the group the ideal bunch is chosen with the assistance of wellness work that is the Region of curve of the classifier.

The convergence curve for selecting the particle swarm optimization for the classification for each disease is shown in the following figures

The above figure 5 shows that the proposed DF-PSO based on multi-objective for curved seizure datasets is met after cycle 4 and the worth is bit by bit diminishing from 0.015 to 0.0065 which is best for eliminating the peculiarity information and produce the best characterization result. The improvement time is high when contrasted with the other datasets because of the huge number of records.

Sepal flower dataset convergence curve:

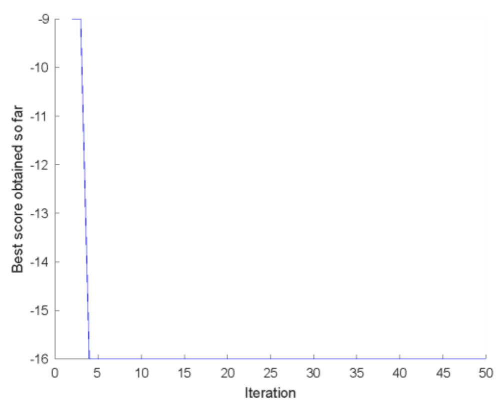


Figure 5. dragonfly convergence curve

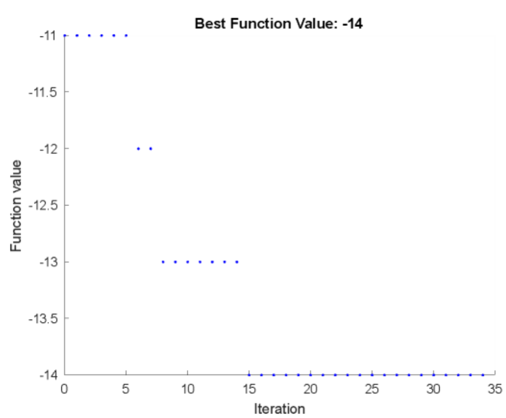


Figure 5.a PSO convergence curve

From the above figure 6, the ROC worth of the -9 is dropped to -16 after 5 cycles. From the figure, it is seen that the calculation begins to combine after 3 and it keep up its steady situation after fourth cycle.

Petal dataset convergence curve:

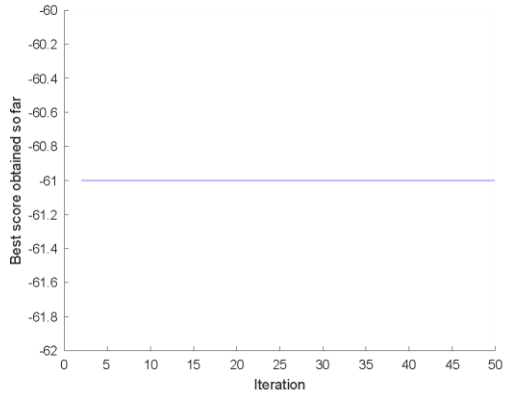


Figure 6. dragonfly convergence curve

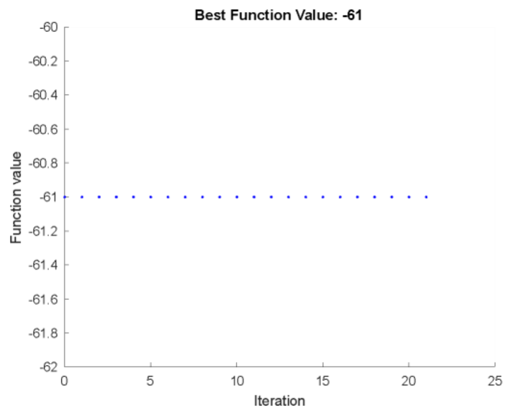
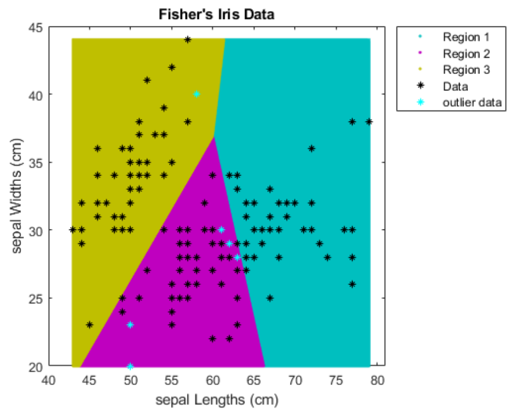


Figure 6.a PSO convergence curve

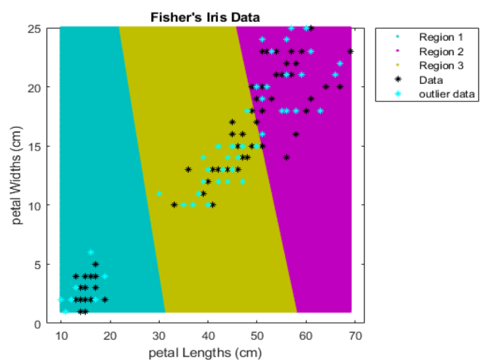
From the above figure 6, the ROC worth of the -61 for all cycles. From the figure, it is seen that the calculation begins to unite after 3 and it keep up its steady situation after fourth emphasis.

Fig 7. Outlier detection of IRIS for sepal dataset



The figure 7 shows the three classes of the iris blossom dataset in three tones and the information focuses are demonstrated as far as the dark tone. The ideal bunch focuses from winged serpent fly calculation is referenced in blue tone. The outlier focuses dependent on the proposed distance equation is notice in cyan tone. It additionally saw that the group focuses are more in the bunch region1 when contrasted with different bunches.

Fig 8. Outlier detection of IRIS for sepal dataset



The figure 8 shows the petal information conveyance of three classes of Iris bloom dataset in three locales and dark tone is utilized for sign of the dataset appropriation. The information focuses are grouped utilizing Dragon fly PSO to decide the ideal bunch and the ideal group focuses are appeared in blue tone. The anomaly information is shown in cyan tone and it is determined utilizing the proposed distance equation.

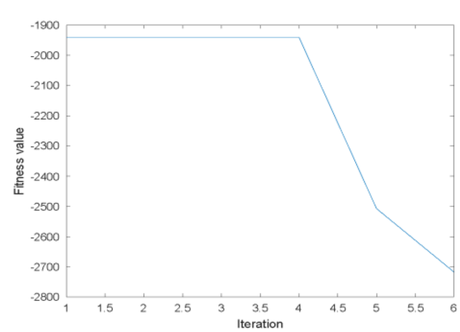


Figure 9. dragonfly convergence curve for epileptic seizure

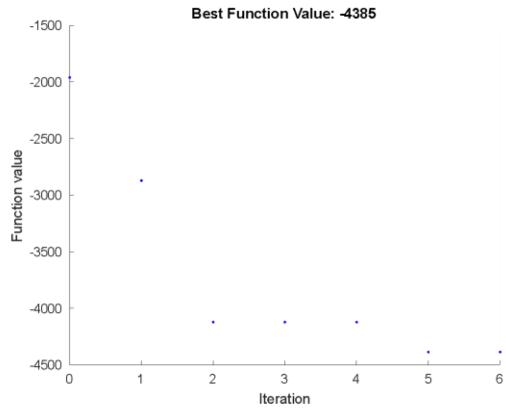


Figure 9.a PSO convergence curve for epileptic seizure

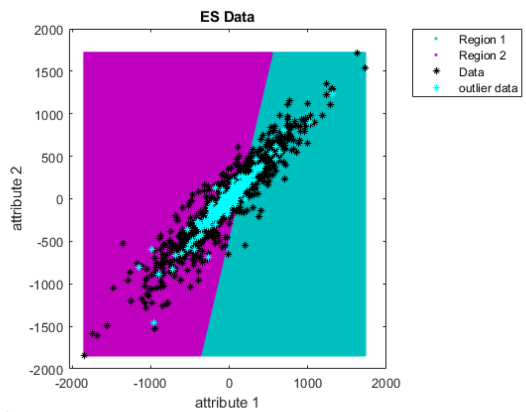


Figure 10. outlier detection in epileptic seizure

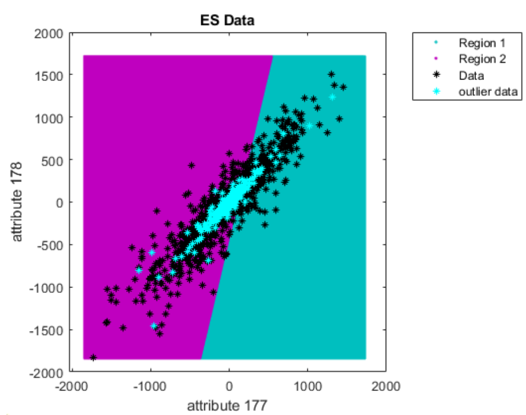
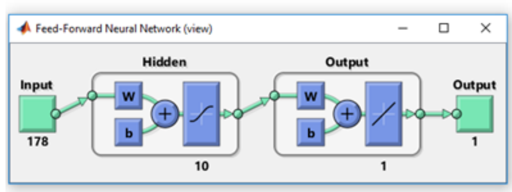


Figure 10.a. outlier detection in epileptic seizure

The figure 10 shows the petal information conveyance of three classes of Iris bloom dataset in two locales and dark tone is utilized for sign of the dataset appropriation. The information focuses are grouped utilizing Dragon fly PSO to decide the ideal bunch and the ideal group focuses are appeared in blue tone. The anomaly information is shown in cyan tone and it is determined utilizing the proposed distance equation.

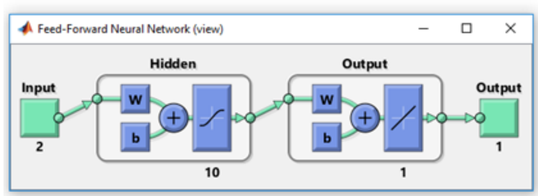
From the Dragonfly based multi-objective PSO clustering the ideal group for the order is resolved for individual datasets and it is prepared utilizing feed forward neural organization. The ideal dataset is characterized into preparing and testing utilizing hold out approach at 0.3%. The preparation information is utilized for process the organization and testing information is utilized for testing and assessment measure.

Fig 11. Neural network diagram of Elliptical seizure



The above figure 11 is the preparation network for the circular seizure informational collection with 178 qualities as info and 10 secret layers and an output layer.

Fig 12. Neural network for Iris data classification



The figure 12 shows the neural organization chart of the iris data characterization. This chart is normal for both the petal and sepal grouping. The information variable will be fluctuating as petal or sepal dependent on the grouping. In view of the information, the covered up and output neuron weight additionally fluctuated.

The proposed DF-PSO algorithm for anomaly identification is tried on two datasets and the after effects of two datasets specifically circular seizure and Iris dataset are determined utilizing the assessment metric in above condition. The outcomes are arranged and it appeared in the beneath table 2.

Tab 2. Comparison of outlier detection approaches

Dataset	**Existing**	**Proposed**
Elliptical seizure	0.9649	0.9749
Iris dataset	0.9759	0.9859

The above table 2 shows that the proposed strategy ready to recognize the outlier alongside the characterization of information with high identification pace of 0.9859 for the Iris data index and 0.9749 for the curved seizure dataset. It likewise demonstrates that the proposed DF-PSO able to measure bigger datasets like Elliptical seizure with 11500 records and more modest datasets like Iris blossom with 150 records.

VI. Conclusion

Outlier detection is a significant interaction in data mining cycle to eliminate the data which are created because of the manual blunder or aggravation. It helps in different applications like climate determining, execution investigation and interruption discovery. A few methods were proposed to play outlier detection. However, they deal with an issue in decide the ideal data with high identification rate and low calculation time. A standout amongst other methodology for outlier detection is muti-objective based particle swarm optimization. Be that as it may, it fizzled for the bigger datasets because of the rehashed distance estimations. In this, it is overwhelmed by the proposed hybrid dragon particle swarm optimization alongside the proposed distance estimation to play out the outlier location. The proposed method ready to measure the more dataset like 11,500 record with 178 traits of Elliptical seizure dataset and 4 credits of the Iris dataset. The hybrid dragon fly based PSO ready to characterize the elliptical seizure dataset with 0.9749 location rate and 0.9859 for the Iris data collection. In view of the exhibition assessment it is seen that the proposed exception recognition can handle a wide range of datasets independent of sizes.

VII. Future work

In future the proposed strategy can be reached out by differing the enhancement procedure to decrease the computational time for the more datasets and to. handle wide range of datasets with independent of size and handle more records.

VIII. References

[1]. K. Yan, X. You, X. Ji, G. Yin and F. Yang, "A Hybrid Outlier Detection Method for Health Care Big Data," 2016 IEEE International Conferences on Big Data and Cloud Computing (BDCloud), Social Computing and Networking (SocialCom), Sustainable Computing and Communications (SustainCom) (BDCloud-SocialCom-SustainCom), 2016, pp. 157-162, doi: 10.1109/BDCloud-SocialCom-SustainCom.2016.34.

[2]. Zhao, Handong, and Yun Fu. "Dual-regularized multi-view outlier detection." *Twenty-Fourth International Joint Conference on Artificial Intelligence*. 2015.

[3]. Xu, X., Liu, H., Li, L., & Yao, M. (2018). A comparison of outlier detection techniques for high-dimensional data. International Journal of Computational Intelligence Systems, 11(1), 652-662.

[4]. Perozzi, Bryan, et al. "Focused clustering and outlier detection in large attributed graphs." *Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining*. 2014.

[5]. Angelin, B. "A Roc Curve Based K-Means Clustering for Outlier Detection Using Dragon Fly Optimization." *Turkish Journal of Computer and Mathematics Education (TURCOMAT)* 12.9 (2021): 467-476.

[6]. Pamula, Rajendra, Jatindra Kumar Deka, and Sukumar Nandi. "An outlier detection method based on clustering." *2011 Second International Conference on Emerging Applications of Information Technology*. IEEE, 2011.

[7]. Angelin, B., and A. Geetha. "DRAGONFLY K-MEANS FOR OUTLIER DETECTION USING ROC CURVE METHOD."
International Research Journal of Modernization in Engineering Technology and Science-1070-1076.

[8]. Jeyanna, J. Antony, E. Indumathi, and D. Shalini Punithavathani. "A network intrusion detection system using clustering and outlier detection." *Int. J. Innov. Res. Comput. Commun. Eng.(IJIRCCE)* 3.2 (2015): 975-982.

[9]. Wang, Yue Fei, et al. "A new outlier detection method based on OPTICS." *Sustainable cities and society* 45 (2019): 197-212.

[10].Ha, Jihyun, Seulgi Seok, and Jong-Seok Lee. "Robust outlier detection using the instability factor." *Knowledge-Based Systems* 63 (2014): 15-23.

[11]. Habib, Usman, et al. "Outliers detection method using clustering in buildings data." *IECON 2015-41st Annual Conference of the IEEE Industrial Electronics Society*. IEEE, 2015.

[12]. Jones, Petra J., et al. "FilterK: A new outlier detection method for k-means clustering of physical activity." *Journal of biomedical informatics* 104 (2020): 103397.

[13]. Wang, Yutao, and Hsi-Yung Feng. "Outlier detection for scanned point clouds using majority voting." *Computer-Aided Design* 62 (2015): 31-43.

[14]. Mandhare, Harshada C., and S. R. Idate. "A comparative study of cluster based outlier detection, distance based outlier detection and density based outlier detection techniques." *2017 International Conference on Intelligent Computing and Control Systems (ICICCS)*. IEEE, 2017.

[15]. Souza, Alberto MC, and Joseé RA Amazonas. "An outlier detect algorithm using big data processing and internet of things architecture." *Procedia Computer Science* 52 (2015): 1010-1015.

[16]. Chen, Chuan, et al. "Robust multi-view k-means clustering with outlier removal." *Knowledge-Based Systems* 210 (2020): 106518.

[17]. Porwal, Utkarsh, and Smruthi Mukund. "Credit card fraud detection in e-commerce: An outlier detection approach." *arXiv preprint arXiv:1811.02196* (2018).

[18]. Paulheim, Heiko. "Identifying wrong links between datasets by multi-dimensional outlier detection." *WoDOOM*. 2014.

[19]. Liu, Bo, et al. "An efficient approach for outlier detection with imperfect data labels." *IEEE transactions on knowledge and data engineering* 26.7 (2013): 1602-1616.

[20]. Jiang, Sheng-yi, and Qing-bo An. "Clustering-based outlier detection method." *2008 Fifth International Conference on Fuzzy Systems and Knowledge Discovery*. Vol. 2. IEEE, 2008.

[21]. Yan, Huanqian, Lei Wang, and Yonggang Lu. "Identifying cluster centroids from decision graph automatically using a statistical outlier detection method." *Neurocomputing* 329 (2019): 348-358.

[22]. Ghorbel, Oussama, et al. "Fast and efficient outlier detection method in wireless sensor networks." *IEEE sensors journal* 15.6 (2015): 3403-3411.

[23]. Marghny, M. H., and Ahmed I. Taloba. "Outlier detection using improved genetic k-means." *arXiv preprint arXiv:1402.6859* (2014).

[24]. Ott, Lionel, et al. "On integrated clustering and outlier detection." *Advances in neural information processing systems* 27 (2014): 1359-1367.

[25]. Shou, Zhao-yu, Meng-ya Li, and Si-min Li. "Outlier detection based on multi-dimensional clustering and local density." *Journal of Central South University* 24.6 (2017): 1299-1306.

[26]. Wang, Xiaochun, et al. "A fast MST-inspired kNN-based outlier detection method." *Information Systems* 48 (2015): 89-112.

[27]. Christy, A., G. Meera Gandhi, and S. Vaithyasubramanian. "Cluster based outlier detection algorithm for healthcare data." *Procedia Computer Science* 50 (2015): 209-215.

[28]. Campello, Ricardo JGB, et al. "Hierarchical density estimates for data clustering, visualization, and outlier detection." *ACM Transactions on Knowledge Discovery from Data (TKDD)* 10.1 (2015): 1-51.

[29]. Moradi Koupaie, Hossein, Suhaimi Ibrahim, and Javad Hosseinkhani. "Outlier detection in stream data by clustering method." *International Journal of Advanced Computer Science and Information Technology (IJACSIT) Vol* 2 (2014): 25-34.

[30]. Jin, Lizhong, Junjie Chen, and Xiaobo Zhang. "An outlier fuzzy detection method using fuzzy set theory." *IEEE Access* 7 (2019): 59321-59332.

[31]. Wang, Yuan, Xiaochun Wang, and Xia Li Wang. "A spectral clustering based outlier detection technique." *International Conference on Machine Learning and Data Mining in Pattern Recognition*. Springer, Cham, 2016.

[32]. Bhattacharya, Gautam, Koushik Ghosh, and Ananda S. Chowdhury. "Outlier detection using neighborhood rank difference." *Pattern Recognition Letters* 60 (2015): 24-31.

[33].Manandhar, Prajowal, and Zeyar Aung. "Intrusion detection based on outlier detection method." *International Conference on Intelligent Systems, Data Mining and Information Technology, ICIDIT*. 2014.