International Journal of Advances in Engineering and Management (IJAEM)
Volume 4, Issue 12 Dec. 2022, pp: 754-768 www.ijaem.net ISSN: 2395-5252

Parametric and Non-parametric Analysis on RHMFO based Optimal Detection of Border Gateway Protocol Anomalies

Sunita.M, Dr.Sujata.V.Mallapur
Assistant. Professor CSE, Marri Laxman Reddy Institute of Technology and Management, Hyderabad, India
Chairperson ,Department of ISE, Sharnbasava University, Kalaburgi, India

Submitted: 18-12-2022 Accepted: 28-12-2022

ABSTRACT—For secure and effective Internet access, the Border Gateway Protocol (BGP) must be able to recognize and stop strange coincidences in real time. Regardless, over the past ten years, more study has focused on spotting BGP abnormalities; because of the rise of new bizarre behaviour from both hackers and network configuration errors, it continues to be more challenging. This work goes with the parametric and non parametric analysis of RHMFO based Optimal Detection of BGP Anomalieswith two major steps (i) Feature Extraction (ii) Anomaly Detection". Initially, extensive features such "statistical features, higher order statistical features, and correntropy features" are extracted during the feature extraction stage. For the detection process, an optimized DBN is proposed to define the presence of attack. Here, a hybrid optimization model known as RHMFO is introduced to fine-tune the weight of DBN in order to enhance the detection accuracy. The traditional Rider Optimization Method (ROA) and the MFO algorithm are conceptually combined to create the suggested RHMFO paradigm. Finally, in this paper, parametric and non-parametric analysis is performed. By varying the parameters of RHMFO, the performance of the suggested work is evaluated.

Keywords—BGP; Anomaly Detection; Multi-Features; DBN based Anomaly Detection ; RHMFO model

Nomenclature

Abbreviation	Description
BGP	Border Gateway Protocol
M-BGP	multipath BGP
IP	Internet protocol
DPPBGP	predict and prevent BGP path
NN	Neural Network
MFO	Moth-flame optimization
QSE-BGP	Quantum Security Enhanced-BGP
CAD	CUSUM anomaly detection
DBN	Deep Belief Network
RHMFO	Rider Hybridized Moth Flame Optimization

I. INTRODUCTION

BGP is an aware routing protocol whose primary goal is to communicate the data between autonomous systems via the Internet. The BGP has a key flaw in that it doesn't handle security, and path stealing is one of the most common cyber hijacks. The Border Gateway Protocol is involved in determining basic routing decisions and bases them on pathways, network regulations, or rule-sets provided by a network administrator. The safety and ease of access of some areas of the network can be impacted by various unexpected occurrences, such as power outages, configuration errors, and other sorts of attacks, which can cause complications and data loss across the Internet. Numerous autonomous systems that make up the Internet use BGP to advertise routes to one another [13]. However, these route updates aren't always reliable or accurate [7]. Anomalies are the term used to describe these strange behavior of the BGP protocol [11][12]. These anomalies can vary from a single incorrect BGP update to thousands, and they have the potential to affect the way BGP traffic behaves over time.

International Journal of Advances in Engineering and Management (IJAEM)
Volume 4, Issue 12 Dec. 2022, pp: 754-768 www.ijaem.net ISSN: 2395-5252

BGP anomaly detection [15] searches for unexpected path modifications or discrepancies in the origination of prefix published by ASes. Finding trends in information that do not match expected behavior is known as anomaly detection. In many application fields, these nonconforming patterns are repeatedly alluded to as anomalies, outliers, dissonant findings, exceptions, aberrations, surprises, oddities, or contaminants. Due to the ASes' interactions' inherent confidentiality, the discovery of route leaks is a particularly challenging issue [6].

Malicious attacks are common against BGP [8]. BGP raw data is typically used as the input for BGP anomaly detection methods [9]. BGP raw data includes both control plane information like RIB and/or BGP updates and packet forwarding information like ping and Nmap data. A BGP router can install numerous "equally-good" pathways to a destination prefix using simultaneous inter-domain border links [10]. The protocol itself need not be altered to identify the anomalies [14].

The major contribution of this research work is:

- ✓ On RHMFO-assisted Optimal Detection of BGP anomalies, Parametric and Non-parametric Analysis is performed by varying the parameters.

The work is structured as follows: Section II discusses the literature research on this issue. Section III discusses the proposed work's architecture for the BGP anomaly detection paradigm and the features extracted are explained in Section IV. Additionally, Section V discusses the optimized DBN for the BGP anamoly detection model. In Section VI, the findings from the suggested work are covered. In Section VII, the conclusion part is given.

II. LITERATURE REVIEW

A. Related works

In 2020, Pradeepa & Pushpalatha et al. [1] have developed an intelligent model in software-defined networks for DPPBGP. In software-defined networks, researchers created an intelligent model for DPPBGP. The researchers have simplified the controller workload as well as recognition duration using SFlow-integrated OpenFlow. There were three main modules in the proposed model: (a) The statistics were assessed based on the network's irregular nodes' activity. The statistic features were used in the cumulative sum abnormal detection algorithm to detect irregular behaviour and flows proficiently and perfectly with less detection time. (b) Sequence of Patterns for predicting network activity, a forecasting algorithm with an intelligent machine learning approach has been used; and (c) route hijack has been prevented by destroying the required PID based on SFlow analyzer .

In 2019, Cheng et al. [2] have proposed a new approach based on the unsupervised learning methods to detect the anomalies in the BGP. Using unsupervised learning methods, they have proposed a new approach for detecting anomalies in the BGP. After that, the various data pipeline approaches for collecting and triggering anomaly events were presented. Moreover, the authors used the DBSCAN as well as k-means to identify suspected anomalies in historical events.

In 2019, McGlynn et al. [3] have proposed a new learning-based anomaly detection mechanism for detecting the attacks in BGP. The authors have used two auto-encoders, and each of them were trained in such a way to detect the anomalies. They report the BGP update as likely-anomalous if any auto-encoder performs poorly (i.e., there are wide variations between input and output). Early findings reveal that their detector was capable of detecting anomalous MOAS disputes as well as prefix hijacking attacks.

In 2019, Elamathi et al. [4] have projected QSE-BGP, which was a modern quantum key distribution algorithm that integrates with the BGP inter-domain routing protocol. The QSE-BGP scheme had been important in BGP routing since it provided shared authentication as well as secured privacy preservation with much less computing expense, memory overhead, and authentication latency.

In 2019, Karimi et al. [5] have used the MLP Neural Networkclassifier to detect the BGP's abnormal behavior that was raised mainly due to the worm attacks. The 'data preparation, feature extraction, and classification' were the three major process followed in the proposed work. Initially, the libbgpdump tool was used to collect the raw data sets, then with the extracted features the MLP Neural Network was trained. The outcomes from MLP Neural Network exhibited the presence/ absence of attacks.

In 2019, Dai et al. [7]have introduced SVM-BGPAD an innovative technique for detecting BGP abnormalities. Initially, a feature selection algorithm based on "Fisher linear analysis and Markov random field technologies" was used. The grid scan and cross-validation approaches were used to refine the SVM parameters.

In 2020, Banu et al. [6] have introduced a new method for detecting BGP spoofing and spoofed nodes based on the BAT Optimization Algorithm. The researchers used the ECC as well as RSA cryptographic approaches to counteract the attacks. The echolocation of the Bats principle was used to identify the BGP spoofed nodes. The authors validated the path in terms of consume computing power, memory, and resources in the DAOA routing protocol.

International Journal of Advances in Engineering and Management (IJAEM)
Volume 4, Issue 12 Dec. 2022, pp: 754-768 www.ijaem.net ISSN: 2395-5252

In 2017, Schlamp et al. [8] have implemented HEAP in BGL to analyse hijacking alarms. The Internet routing registry was used to determine the commercial or corporate relationships between the event's participants. Furthermore, accidents triggered by legal operating procedure were ruled out using a topology-based reasoning algorithm. Using Internet-wide network scans, the SSL/TLS-enabled hosts were discovered. The simulation results demonstrated that the proposed work was effective in validating previously recorded warnings.

III. BGP ANOMALY DETECTION MODEL: FRAMEWORK OF SUGGESTED WORK

This study follows two main phases to introduce an unique BGP anomaly detection model: (i) Phases for Feature Extraction (ii) Phase for detecting anomalies Fig. 1 provides an illustration of the planned work's architecture. First, from the BGP data gathered A^{in}, large-scale features are retrieved, including "statistical features, higher order statistical features, better holoentropy features, and correntropy features. The retrieved features are shown as F. An already trained Optimized DBN then performs the anomaly detection on the inclusion or exclusion of anomaly. Additionally, to improve the detection accuracy, the RHMFO, a conceptual fusion of the traditional ROA and MFO method, fine-tunes the weight of DBN.

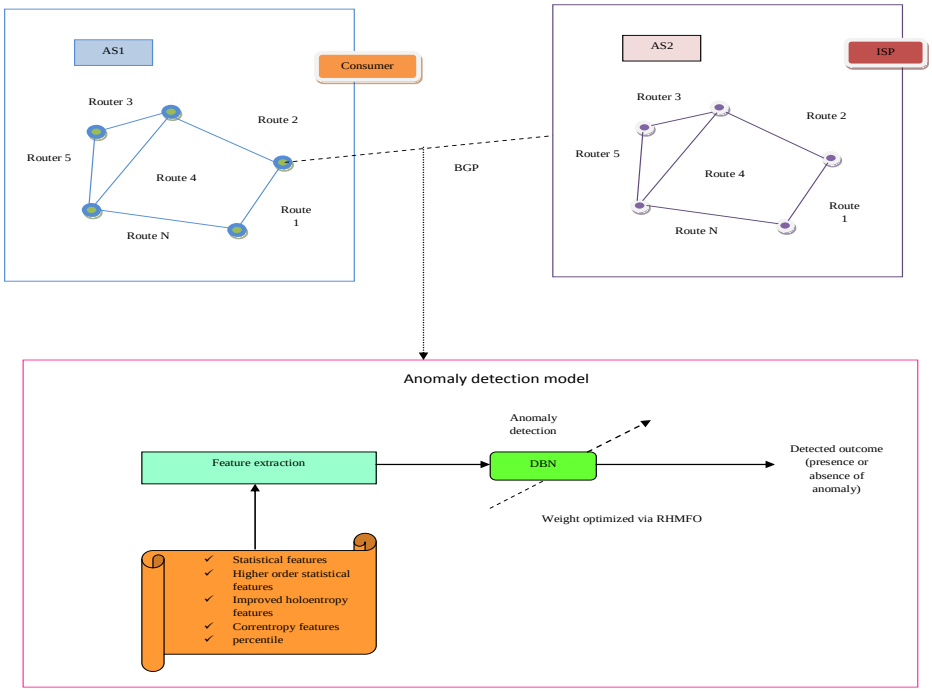


Fig. 1. Architecture of suggested RHMFO

IV. EXTENSIVE MULTI- FEATURE EXTRACTION: STATISTICAL FEATURES, HIGHER ORDER STATISTICAL FEATURES, IMPROVED HOLOENTROPY FEATURES AND CORRENTROPY FEATURES

The features are more significant in terms of data mining. The most important attributes or features of a data input must be extracted using feature extraction techniques in order to classify the data. The statistical features and higher order statistical features are the most pertinent multiple features in this research endeavour, improved holoentropy features and correntropy features were extracted from A^{in}. The section that follows provides a detailed description of these characteristics.

A. Statistical Features

In terms of how closely related the components of a distribution are to one another, the relationship between mean, median, and mode is that they have been all preferred measure of central tendencies. Standard deviation, on the other hand, is a measure of dispersion that demonstrates how

International Journal of Advances in Engineering and Management (IJAEM)
Volume 4, Issue 12 Dec. 2022, pp: 754-768 www.ijaem.net ISSN: 2395-5252

dispersed the variables are from one another.Calculating the average, median, min-max, and standard deviation are extracted from A^{in}

AM ζ : The ratio of the total number of inputs is the mean. $A^{in}=\left(A_1^{in},A_2^{in},.....,A_n^{in}\right)$ that is detected during a time interval to the sample count taken (M). It is given mathematically according to Eq. (1).

$$\xi=\sum A_n^{in}\Big/M \qquad (1)$$

SD (σ): The data is calculated in this calculation $A^{in}=\left(A_1^{in},A_2^{in},.....,A_n^{in}\right)$distribution. According to Eq. 2, SD is computed.

$$\sigma=\sqrt{\frac{\sum\left(A_n^{in}-\zeta\right)^2}{M}} \qquad (2)$$

Median:when data in $A^{in}=\left(A_1^{in},A_2^{in},.....,A_n^{in}\right)$arranged from small to high, point in centre is the median. Whether there are middle two values, the average between them is calculated, and the result is referred to as the median.

Minimal Value or Minima: Dataset $A^{in}=\left(A_1^{in},A_2^{in},.....,A_n^{in}\right)$depicts minimal value function. Minvalue and the D^{in}'s max value depicts maximal function.

Variance: The variance is the expected square deviation of a random vector from its mean in statistics. It is given mathematically according to Eq. (3).

$$Vari(A^{in})=\left[G(A^{in})-\zeta\right] \qquad (3)$$

The extracted statistical features are denoted by $F^{statistics}$.

B. Higher Order Statistical Features

Furthermore, the higher order statistical characteristics such as skewness and kurtosis [31] arecalculated by A^{in}. Skewness and Kurtosis are two metrics of the data distribution shape in general. Skewness essentially quantifies data asymmetries. Kurtosis, on the other hand, quantifies a distribution curve's bulge or peak. Kurtosis determines if the curves is almost higher than the normal curve, while skewness determines whether the supplied data collection is slanted toward that single side (mean > mode or vice versa). Skewness and Kurtosis are functions of the third and fourth moments, respectively.

Skewness: The degree to which the tailed of a distribution deviate from the ends of a normal distribution is measured statistically. It denotes the probability distribution's lack of symmetry. Moments about the mean or central moments will be used to calculate it. Skewness is mathematically expressed in the form of Eq. (4).

$$H=\frac{\sum_{i=1}^{k}\left(A^{in}-\zeta\right)^3\Big/M}{\sigma^3} \qquad (4)$$

Additionally, when calculating the skewness, the SD is calculated with M in the fraction rather than $M-1$. The skewness value for any symmetric data is close to zero, and the skewness for a normal distribution is zero.

Kurtosis: It is a measurement of the data's heavy- or light-tailedness in relation to a normal distribution.A high kurtosis value indicates that the distribution's tails are more likely to experience outliers than the tails of the normal distribution [31]. The tails of the distribution also will be shorter than the tails of a normal distribution if the value of kurtosis is very low.kurtosis Kur mathematical formula for uni-variate data's was $A^{in}=\left(A_1^{in},A_2^{in},.....,A_n^{in}\right)$, is depicted in Eq. (25).

$$Kur=\frac{\sum_{n=1}^{M}\left(A^{in}-\overline{A^{in}}\right)^4\Big/M}{\sigma^4} \qquad (5)$$

Instead of using $M-1$to determine the kurtosis, SD is calculated using the value of M that is available in the denominator. Higher order statistical features is expressed as F^{HOS}.

C. Percentile

From the "lowest to the largest value," it shows how uniformly the data elements are distributed through time [8]. D % of data values were under D percentile, while 100- D % are upon J^{th} percentile. Percentile features are depicted in notation J. The percentile features are indicated as F^{perce}.

D. Standardised moment

As per probability theory and statistics, a normalized moment of a probability distribution is a standardized moment (normally a higher degree central moment). The moment scale becomes invariant since normalization is a SD term [33]. The standardised moment is measured using Eq. (6).

$$\zeta_k=E\left[\left(A^{in}-\zeta\right)^k\right]=\int_{-\infty}^{+\infty}\left(A^{in}-\zeta\right)^k.J(A^{in}) \qquad (6)$$

International Journal of Advances in Engineering and Management (IJAEM)
Volume 4, Issue 12 Dec. 2022, pp: 754-768 www.ijaem.net ISSN: 2395-5252

Here, $\zeta = C[A^{in}]$, E and J depicts probability distribution also withexpected A^{in} value. The notation ζ_k / σ_k is k^{th} mean moment, and it depicts standardized moment refers to degree k. σ_k depicts k^{th} SD power. The indicated standard moment features are those that were extracted as F^{SD}.

E. Proposed Improved Weighted Holoentropy

The "sum of entropy and relative correlations of the random vector" is how holoentropy [32] is generally defined. Utilizing attribute association to handle data is been the main goal. Holoentropy $Hol(A^{in})$ is computed only each of A^{in}'s features via expression depicts in Eq. (7).

$$Hol(A^{in}) = w \times G(A^{in}) \quad (7)$$

The weight function in the traditional holoentropy is established for each dataset at a preset value. As a result of the data being gathered from many sources, the categorization accuracy may be reduced. It is computed depending on A^{in}, as expressed in Eq. (8). Since each data feature's weight is assigned separately, the reliance on retrieved attributes grows automatically, improving detection performance. $w = \frac{1}{\sqrt{2\pi}} . d \frac{G(A^{in})^2}{2}$ (8)

In which,

$$G(A^{in}) = \sum_{t=1}^{\zeta . A^{in}} prb_t . \log(prb_t). \quad (9)$$

Here, $G(A^{in})$depicts holoentropy of A^{in}. prb_t depicts the of feature attributes probability of A^{in} .Indicating extracted fresh weight holoentropy depend features as F^{wh}.

F. Correntropy features

Since each data feature's weight is assigned separately, the reliance on retrieved attributes grows automatically, improving detection performance. Eq.(10) computes correntropy I_σ of Attack with Normal features. σ depicts kernel size, predicted features value are $G[.]$, and the Gaussian kernel function depicts $L_\sigma(.)$. Mathematically, $L_\sigma(.)$ is indicated in Eq. (11). Additionally, the correntropy is calculated using Equation (12).

$$I_\sigma(Att, Nor) = G[L_\sigma(Att - Nor)] \quad (10)$$

$$L_\sigma(.) = \frac{1}{\sqrt{2\Pi.\sigma}} \exp^{\left(-\frac{(.)}{2\sigma^2}\right)} \quad (11)$$

$$\hat{I}_{M,\sigma}(O,P) = \frac{1}{Q} \sum_{i,j=1}^{Q} [L_\sigma(Att_i - Nor_j)] \quad (12)$$

Correntropy based features is depicted as F^{corr}.

All these extracted features are together represented as $F = F^{statistics} + F^{HOS} + f^{stm} + F^{perce} + F^{wh} + F^{corr}$, these are utilized for training anomaly detection tool, the optimized DBN.

V. OPTIMIZED DEEP BELIEF NETWORK FOR BGP ANAMOLY DETECTION MODEL

A. Optimized Deep Belief Network

A stochastic neuron framework is used by DBN to provide outcomes for the sources. DBN's outcomes are depicted as a notation DBN^{out}. In actuality, DBN employs the Boltzman network to achieve probabilistic performance. The notation DBN^{out} represents the DBN result (in binary form). Furthermore, Eq. (13) and (14) denotes probability $(L_p(\delta))$, that is a sinusoidal structured function. When $S > 0$, pseudo temperature parameter S decrease probability's sound ratio. A deterministic rendition of the stochastic process of probability is also seen in Eq. (14), eq. (15). Hinton et al. found that the stochastic neural network that is Boltzmann approach covers stochastic neurons.

$$DBN^{out} = \begin{Bmatrix} 1 & with\ 1 - L_p(\delta) \\ 0 & with\ L_p(\delta) \end{Bmatrix} \quad (13)$$

$$L_p(\delta) = \frac{1}{1 + d^{\frac{-\delta}{S}}} \quad (14)$$

$$\lim_{S \to 0^+} L_p(\delta) = \lim_{S \to 0^+} \frac{1}{1 + d^{\frac{-\delta}{S}}} = \begin{cases} 0 & for \quad \delta < 0 \\ \frac{1}{2} & for \quad \delta = 0 \\ 1 & for \quad \delta > 0 \end{cases} \quad (15)$$

The Boltzmann machine's energy can be calculated using the mathematical formula Eq. (16). The energy factor is crucial for setting the neuron states in the Boltzmann system. Additionally, in DBN, the weight between both the neurons and their biases depicted by $W_{x,y}$ and η. Equations (17), (18), and (19) illustrate the mutual arrangement of visible r and hidden neurons g in terms of power.Dual states of observed and hidden states, x and y, are depicted as l_x and l_y, correspondingly.

International Journal of Advances in Engineering and Management (IJAEM)
Volume 4, Issue 12 Dec. 2022, pp: 754-768 www.ijaem.net ISSN: 2395-5252

$$Eg(l\ ) = -\sum_{x<y} W_{x,y} l_x l_y - \eta_x . l_x$$

$$\Delta Eg(l_t) = \sum_{y} {}^{l}{}_{y} W_{x,y} + \eta_x$$

(16)

$$Eg(r,g) = -\sum_{(x,y)} W_{x,y} r_x g_y - \sum_{x} r_x u_x - \sum_{y} g_x go_y$$

(17)

$$\Delta Eg\left(r_x, \vec{g}\right) = \sum_{y} W_{xy} g_y + u_x$$

(18)

$$\Delta Eg\left(\vec{r} g_y\right) = \sum_{y} W_{xy} g_x + o_y \qquad (19)$$

Weight parameter is depicted as Q. The RBM learning pattern is established using the input data's encoded probability distribution. This weight parameter $Q$is fine-tuned via hybrid approach for improving anomalies prediction accuracy. The weighted assignment made by RBM is predicated on Eq. (20), and RBM has the ability to maximize the delegated probability. The phrase u designates the input visible vector. Energy function in Eq. (21) demonstrates the capability of RBM to assign probability to each specific visible and hidden vector. Weight is depicted as E^d, and the preparation set depicted as W. Partition function U is then achieved by adding possible states overall energy of neurons, as shown Eq. (22).

$$E^{(d)} = \max_{E} \prod_{f \in S} Y\left(\vec{u}\right)$$

(20)

$$Y(u,g) = \frac{1}{U} d^{-Eg(u,g)} \qquad (21)$$

$$U = \sum_{\vec{u},\vec{g}} d^{-Eg(u,g)} \qquad (22)$$

The "difference among actual and the expected outcomes" is what the DBN refers to as the Root Mean Square (RMSE) classification error. According to Eq. (23), the classification mistakes are expressed mathematically in terms of RMSE, where $Actual$ and pre^{out} were actual and predicted results. The main goal of this research project is to reduce categorization error Err. The objective function Ob of this research study is represented in Eq. (24)

$$Err = Actual - pre^{out} \quad (23)$$

$$Ob = \min(Err) \quad (24)$$

The steps of Contrastive Divergence (CD) algorithm is listed below.

✓ By selecting the training samples u and connecting them to the observable neurons, the binary input is inferred.

✓ Hidden neurons probability Y_g is obtained by multiplying u observed vector via weight matrix E, and it is depicted as $\left(Y_g = \psi(E.u)\right)$ in Eq. (25)

$$p\left(g_y \to 1 \mid f\right) = \psi\left(o_y + \sum_{x} u_x W_{x,y}\right)$$

(25)

✓ "Hidden states" g were sampled from probability Y_d

✓ Positive gradient υ^+ was computed as $\upsilon^+ = u.Y_g^A$, that is u *and* Y_g outline product

✓ Eq. (26) depicts how to create again the observable state $u*$ from latent state $g*$. Hidden states $d*$ re-sampling gives permission restoration of visible state $u*$

$$p\left(u_x \to 1 \mid g\right) = \psi\left(u_x + \sum_{y} g_y W_{x,y}\right) \quad (26)$$

✓ Negative gradient, $u*$ vectors outline product and $g*$ is υ^-. $\upsilon^- = u*.g^{*A}$ is used to measure the negative gradient.

✓ "Weight updates are produced by removing the negative gradient." υ^- from +ve gradient υ^+".

The updating of the weight is calculated in conjunction with Eq. (27).

$$\Delta E = \tau\left(\upsilon^+ - \upsilon^-\right) \qquad (27)$$

✓ The weights are changed based on the recently gathered values shown in eq. (28).

$$W_{x,y}^* = \Delta Y_{x,y} + Y_{x,y} \qquad (28)$$

B. Proposed RHMFO Model

For fine-tuning of DBN weight W, this research study introduces a brand-new hybrid optimization model known as RHMFO. Here, W is updated utilizing ROA with MFO and At last, the geometric mean is calculated for the best ROA and MFO results. It is claimed that the last acquired solution is the best one. As a result, a best global solution is possible with less convergence. Figure 2 depicts the solution fed to the RHMFO as its input.

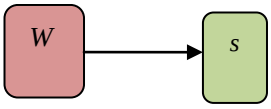


Fig. 2. Solution Encoding

The steps followed in the proposed work are depicted below:

International Journal of Advances in Engineering and Management (IJAEM)
Volume 4, Issue 12 Dec. 2022, pp: 754-768 www.ijaem.net ISSN: 2395-5252

Step 1- Initialize search agent population $Popn$. Set present iteration $iter = 1$ and the max iteration $\max^{iter}$.

Step 2- Give Opposition assisted learning to $Popn$. Opposition-based learning (OBL)[40] is a well-known method for improving the first answer by contrasting the real population of solutions and its opposite solution simultaneously and choosing the solution that is better adapted for the problem as the initial solution. As a result, in order to proceed with the best one, the scores and their opposing points are computed simultaneously. Better solutions are possible thanks to the OBL-based initialization's assurance of a quicker convergence rate.

Step 3- Using the conventional ROA approach, the population (the weight of the DBN) is optimized. The solutions' position is then modified. "ROA [37] is founded on an idea of a number of cyclists" traveling in that direction to succeed. Each rider category has an equal number of participants, which is then categorized into 4 clusters. The four sorts of riders are "bypass rider, aggressor, follower, and overtaker."

a) Initialization: Riders is depicted as v with arbitrary initialize field and start initialization process with 4 groups. Cluster initiate is shown by Eq. (29). Rider's count depicted by b and equal to v, axis co-ordinates is depicted by c, and the h^{th} direction of rider at specific time is represented by $s^t(h,a)$. According to Eq. (30), the assessment of riders is reviewed for any cluster. $s^t = \{s^t(h,a)\}; \quad 1 \le h \le b; \quad 1 \le s \le c$

(29)

$$b = R + v + T + q \qquad (30)$$

$R, v, Tand\, q$ depicts count of bypass riders, followers, overtaker, and attackers. Eq. (31) demonstrates the link between these factors as a result.

$$c = v = T = z = b/4$$

(31)

Bypass passenger limits are follower, overtaker, and attacker is $[s_1, s_{b/4}]$, $[Q_{b/4+1}, Q_{b/2}]\left[s_{b/2+1}, s_{3b/4}\right]$ and s respectively.

Then, parameters including steering, gear, the accelerator, and the brake are initialized. The steering angle is shown in Eq. (32) at time $(time)$, where $Steer_{h,a}^{time}$ depicts steering angle of the h^{th} rider vehicle. Eq. (33) represents the steering angle at the beginning position and initial time.

$$S^{time} = \{S_{h,a}^{time}\}; \quad 1 \le h \le b; \quad 1 \le a \le c$$

(32)

$$Steer_{h,a}^{time} = \begin{cases} \theta_i; & if\ a = 1 \\ Steer_{h,a}^{time} + \varphi; & if\ a \ne 1 \& \\ & St_{r,s}^{tim} + \varphi \le 360 \\ Steer_{h,a}^{time} + \varphi - 360; & otherwise \end{cases}$$

(33)

Direction angle of h^{th} rider vehicle was depicted as θ_i, and coordinate angle was depicted as φ. While Eq. (35) assesses the coordinate angle, Eq. (34) examines the recognition of the h^{th} rider's location angle.

$$\theta_{rider} = h * \frac{360^{\circ}}{b}$$

(34)

$$\varphi = \frac{360}{B}$$

(35)

Equations (36), (37), and (38) establish the rider's vehicle's gear, accelerator, and brake.

$$Gear = \{Gear_h\}; \quad 1 \le i \le b$$

(36)

$$accel = \{accel_h\}; 1 \le h \le b$$

(37)

$$A = \{A_h\}; \qquad 1 \le h \le b$$

(38)

Gear, accelerator, and brake of h^{th} rider vehicle were depicted as $Gear_h$, $accel_h$ and A_h. The top speed is expressed in Eq. (39) because the space limit value controls how fast the vehicle travels. s_v and s_R depicts max/minvalue of h^{th} the position of rider, and R_{OFF} depicts offtime.

$$V_{\max} = \frac{Q_v - Q_R}{R_{OFF}} \qquad (39)$$

b) Success rate: The computation of the success rate for each rider is handled following the completion of the initialization phase. The distance related evaluation of the performance is determined by Equation (40); h^{th} rider location depicts s_h, and aim location depicts R_p. A high - performing rate must be attained while minimizing the differences amongst riders.

$$S_i = \frac{1}{s_h - R_p}$$

(40)

c) Leading rider: The leading rider is decided by the rate of success; the rider who is closer to the goal is thought to have a better success rate. Before the finish of the race, there won't be any established leading riders because the cyclists' placements will shift.

International Journal of Advances in Engineering and Management (IJAEM)
Volume 4, Issue 12 Dec. 2022, pp: 754-768 www.ijaem.net ISSN: 2395-5252

Rider's location update: Each set handles the identification of the leading rider through the rider's update position.

Update technique for bye pass riders: Since bye pass riders do not follow the prime riders as described in Eq. (41), they bypass the regular route and their orientation is modified at random. δ *and* β depicts random numbers in limit 0 to 1, while ∂ *and* ℓ depicts random numbers in limit [1, *RN*].

$$s_{time+1}^{c}(h,a)=\delta\left[a_{time}(\varpi,h)+\beta(h)+s_{time}(\ell,h)*[1-\beta(h)]\right] \quad (41)$$

Update protocol for follower: By shifting the follower's position to match the positioning of the leading leader, the motorcyclists can follow this procedure and reach their goals safely and conveniently. The coordinate selector is used to calculate the follower's location update according to Eq. (42) for the chosen values in J .

$$s_{time+1}^{V}(h,x)=B^{G}(G,x)+\left[\cos(Steer_{h,a}^{time})*s^{G}(G,x)*dt_{hi}^{time}\right] \quad (42)$$

Leading location of rider is expressed as Q^{S} , the h^{th} rider steering angle at x^{th} coordinate is depicted as $W_{h,a}^{time}$, and distance needed for covering by h^{th} rider is depicted as dt_{h}^{time} .

Overtaker update protocol: Comparative success rate, pathway predictor, and location selector make up the bulk of this updating mechanism. The overtaker's location change equation can be seen in Eq. (43). In time *time* , rider pathway predictor was depicted as as $Y_{t}\ (E)$

$$s_{time+1}^{T}(h,x)=s_{time}(h,x)+\left[Y_{time}(h)*s^{G}(G,x)\right] \quad (43)$$

Attacker update procedure: Since the attacker updated position takes the place of the lead rider, it is used in the same manner as a follower. The method of updating the attacker's position is described by Equation (44).

$$s_{time+1}^{z}(h,a)=Q^{G}(G,x)+\left[\cos(W_{h,a}^{time})*s^{G}(G,x)\right]+dt_{h}^{time} \quad (44)$$

e) Calculating the success rate: The rate of success for each traveller is calculated using the location update technique. The position of the present rider is changed to that of the leading rider because that rider's overall performance is highest.

f) Rider parameter update: Rider characteristics including gear, the accelerator, the brake, and the steering must be changed in order to identify the most effective and optimal response.

Step 4: The MFO model is used to process the input population. The MFO is predicated on how the moths migrate in the direction of the flames.

(a) Moth's Position update [38]s: The universal optimal of optimization problems as shown in Eq. (45) is determined via the MFO algorithm, which is a three-tuple.

$$MFO=Spr(\ddot{J},R) \quad (45)$$

Starting random placement of moth is indicated as $Spr:\phi\rightarrow\{V,OR\}$, search space moth motion is denoted by $\ddot{J}:V\rightarrow V$, and end search is denoted as $R:V\rightarrow\{true,\ false\}$ in above formula.

The exponential spirals function determines the moth's motion. In Eq. (46) eq. (47), eq. (48), this is demonstrated mathematically.

$$Spr(\ddot{J},R)=dis_{i}\cdot d^{ht}\cdot\cos(2\pi.k)+R_{h} \quad (46)$$

Here, $\dot{t}=(D-1)*ran+1$ (47)

$$h=-1+time*\left(-\frac{1}{\max^{time}}\right) \quad (48)$$

Here, $dist_{h}=\left|E_{a}-Y_{h}\right|$ is distance among the h^{th} moth and a^{th} flame, k depicts spiral shape, *ran* is a random number range limit $[-1,1]$, solution update with MFO is expressed as s^{x} .

Update the flames count: This portion reduces the number of flames to enhance the MFO algorithm's exploitation process. It is computed by Eq. (49), where G depicts overall amount of flame.

$$flaco=round\left(G-it*\frac{G-time}{\max_{time}}\right) \quad (49)$$

Step 5-The geometric mean of the weight function updated using ROA and MFO is computed in order to achieve the maximum solutions to global with faster convergence. The final result reveals the search agent's ideal position.

Step 6- Return $o*$

Step 7- Terminate

Algorithm 1: Pseudo code of RHMFO	
Initialize: $Popn, \max^{itr}$ Set current iteration $time=1$ Max^{time}	
	Apply OBL to *Popn*
	while $time < Max^{time}$

International Journal of Advances in Engineering and Management (IJAEM)
Volume 4, Issue 12 Dec. 2022, pp: 754-768 www.ijaem.net ISSN: 2395-5252

	Compute the fitness function utilizing Eq. (24).
	set rider constraints
	Determine the success rate
	While $time < R_{off}$
	for $u = 1$ to s
	Update bypass position rider as per Eq. (41)
	Update follower position as per Eq. (42)
	Update over taker position as per Eq. (43)
	Update attacker position as per Eq. (44)
	Rank riders depending on success rate
	Choose rider with higher success rate as leading one.
	Update the rider constraints
	Return s^S
	$time = time + 1$
	end for
	end while
	For $i = 1 : size(Moth\_pos, 1)$
	for $j = 1 : size(Moth\_pos, 2)$
	Update g_i as per Eq. ()
	Update $W(\ddot{J}, R)$ as per Eq. (45)
	Update the position of search agent using Eq. (46)
	end for j
end for i	
Return s^x	
Compute the geometric mean between s^x and s^S. The outcome is the best solution $o*$	
Return $o*$	
End while	
Terminate"	

VI. RESULTS AND DISCUSSION

A. Simulation Procedure

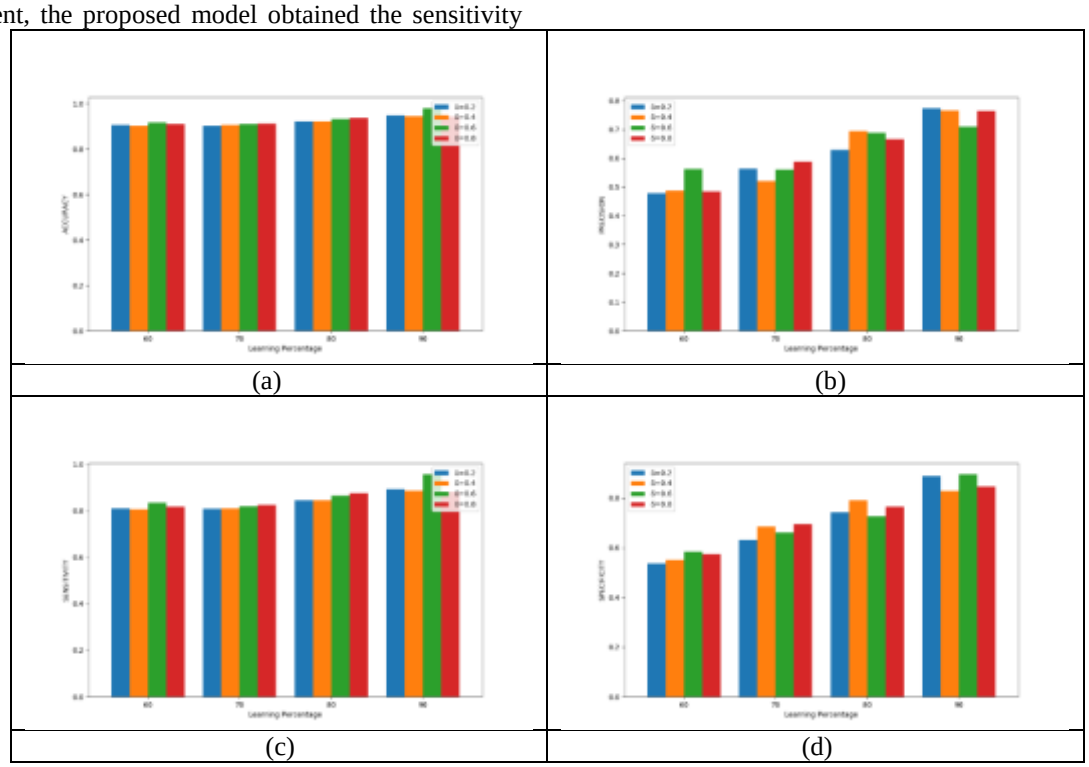
The proposed DBN+RHMFO method based BGP anomaly detection was implemented in python and the obtained results were analyzed. Here, the DBN+RHMFO method was evaluated by varying the parameters like δ under the values 0.2, 0.4, 0.6 and 0.8 as well as β variations under the values such as 0.2, 0.4, 0.6 and 0.8 respectively. Also, the performance of proposed model was evaluated in terms of positive measures (Accuracy, Precision, Sensitivity, Specificity), negative measure (FDR, FNR, FPR) and other measures (F-measure, MCC, NPV). The performance analysis was measured with different learning percentage ranges from 60, 70, 80 and 90.

B. Performance Analysis on proposed DBN+RHMFO model by varying parameter δ

The performance of the proposed DBN+RHMFO model is evaluated by varying parameter δ variation under the values 0.2, 0.4, 0.6 and 0.8 is shown in fig 3, fig 4, fig 5. Also, it estimates the performance with respect to positive, negative and other measures. While varying the parameter δ to 0.6, the proposed DBN+RHMFO model attains the maximum accuracy as 97% (approximately) in the 90% learning percent. Next, in the variation 0.4 and 0.8, approximately the proposed work maintains 93% and 95% of accuracy in the learning percent 80. Further the analysis has been taken for the negative measures, in the 90% learning percent, by varying the parameter δ under 0.6, the proposed model gets the minimum value as 0.0028. When comparing all the variations in the FPR measure, particularly in the 90% learning rate, δ

International Journal of Advances in Engineering and Management (IJAEM)
Volume 4, Issue 12 Dec. 2022, pp: 754-768 www.ijaem.net ISSN: 2395-5252

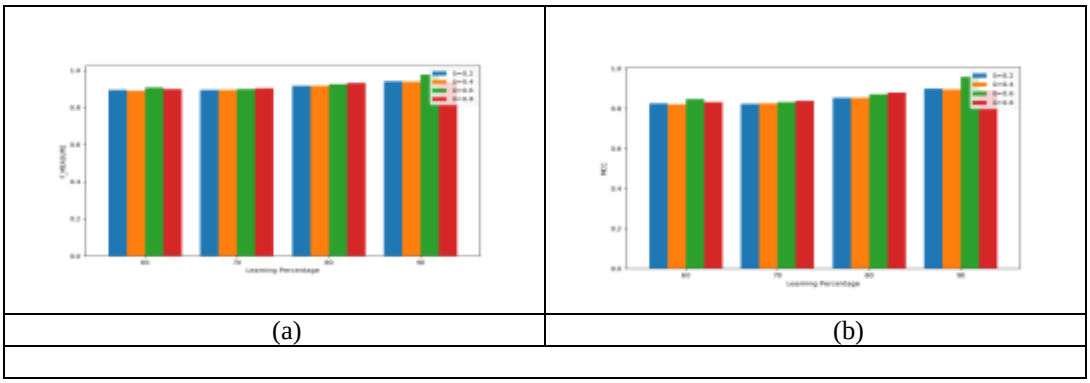
under the variation 0.2 acquired the minimum value as 0.12. It is observed that, (~) 90% of F-measure is finally maintained by the proposed DBN+RHMFO model under the variation δ =0.6 in the 90% learning percent. The MCC and NPV attains the maximum value as (~) 95% under the variation 0.6 in the 90% learning rate. From the fig 3(c), in the 60% learning percent, the proposed model obtained the sensitivity values as 80%, 79%, 83% and 82% (approximately) under the variation δ =0.2, δ =0.4, δ =0.6 and δ =0.8. According to the analysis, high accuracy rate was obtained for the proposed DBN+RHMFO method, which shows the efficiency for BGP anomaly detection.



(a) (b)

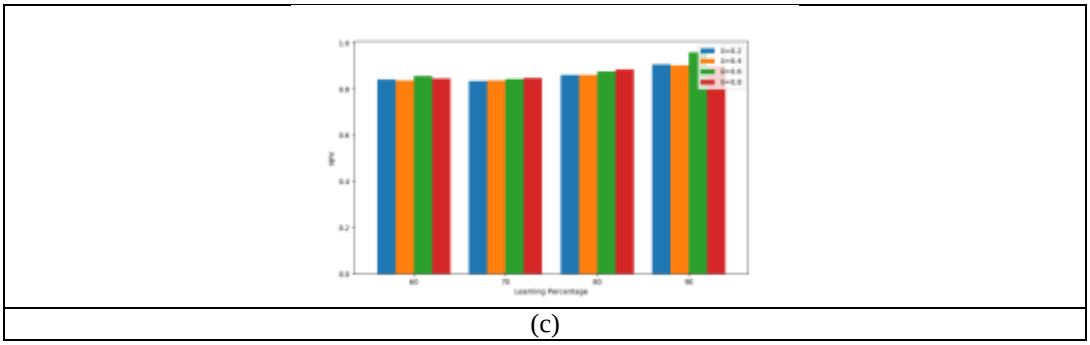
(c) (d)

Fig. 3. Performance Analysis of proposed DBN+RHMFO by varying parameter δ with respect to positive measure a) Accuracy b) Precision c) Sensitivity d) Specificity



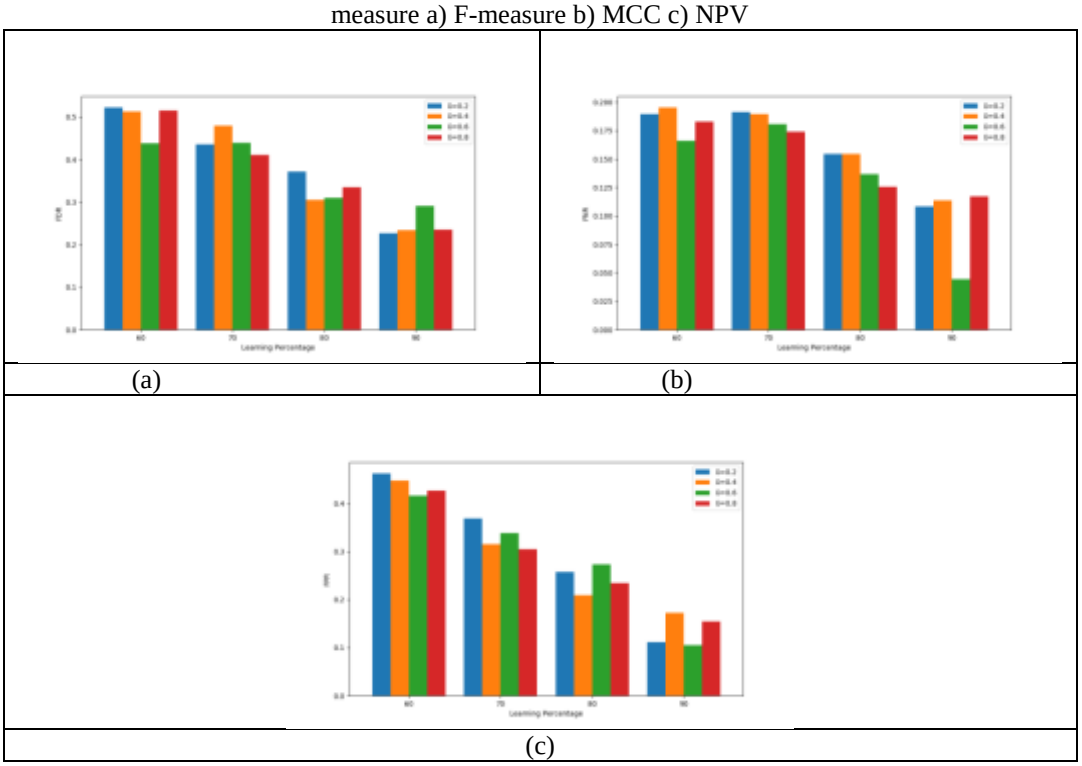
(a) (b)

International Journal of Advances in Engineering and Management (IJAEM)
Volume 4, Issue 12 Dec. 2022, pp: 754-768 www.ijaem.net ISSN: 2395-5252



(c)

Fig. 4. Performance Analysis of proposed DBN+RHMFO by varying parameter δ with respect to positive measure a) F-measure b) MCC c) NPV



(a) (b)

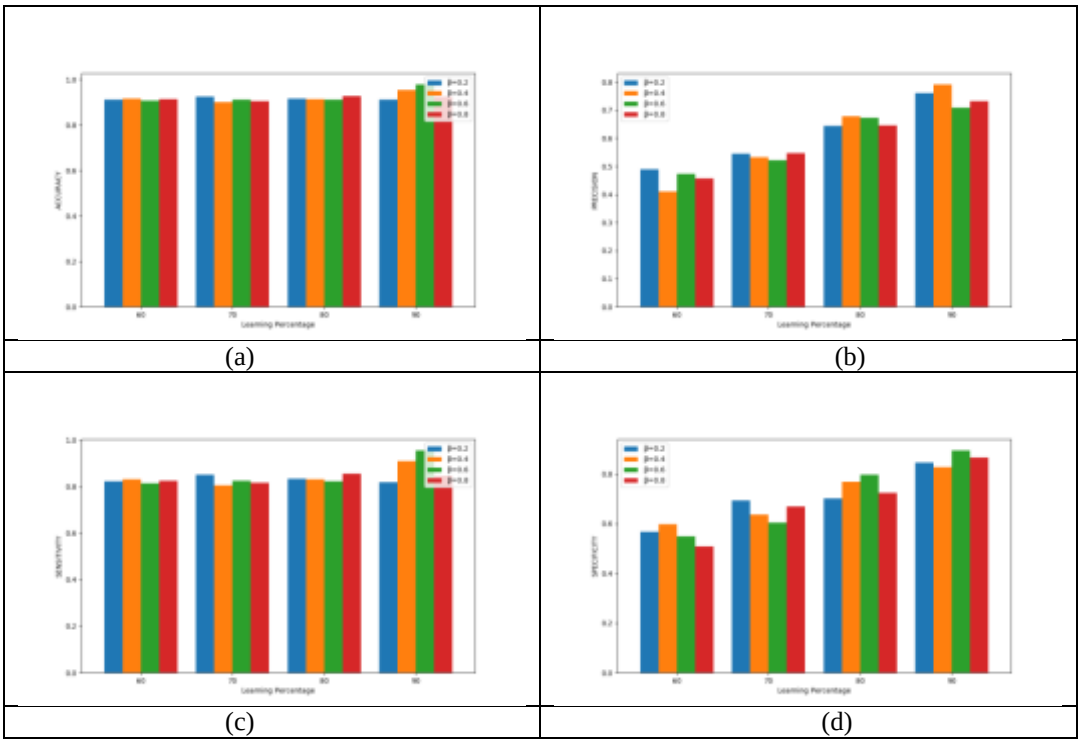
(c)

Fig. 5. Performance Analysis of proposed DBN+RHMFO by varying parameter δ with respect to positive measure a) FDR b) FNR c) FPR

C. Performance Analysis on proposed DBN+RHMFO model by varying parameter β

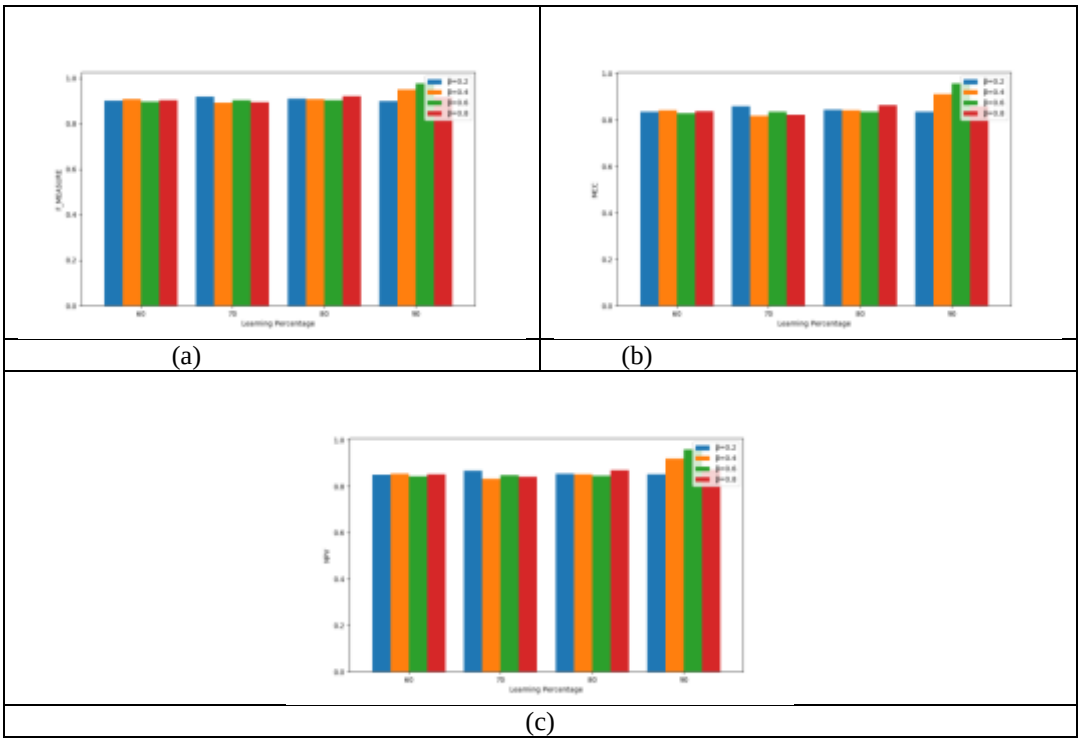
The performance of the suggested DBN+RHMFO model is analyzed by varying parameter β variation under the values 0.2, 0.4, 0.6 and 0.8 and its visual illustration is shown in fig 6, fig 7 and fig 8. Additionally, it evaluate the performance with respect to accuracy, sensitivity, precision, F-measure, MCC, FNR and so on. The outcomes show that the proposed DBN+RHMFO model is efficient and suitable for anomaly detection. Subsequently, the proposed DBN+RHMFO model has the highest accuracy as (~) 96% under the variation β =0.6 in the 90% learning rate. Next, highest accuracy is obtained in the same learning percent under the variation β =0.4 as (~) 90%. Moreover, 75% of precision is obtained under the variation β = 0.4 in the 90% learning percent. The F-measure, MCC and NPV has attained the highest value for the suggested model as (~)90%, 92% and 91% respectively under the variation β =0.6. Subsequently, under the variation of β in 0.8 the obtained value for the FPR is 0.14. Additionally, the proposed DBN+RHMFO model acquired the minimum error value in the FDR, FNR as 0.028 and 0.12 under the variation β =0.6 in the 90% learning rate. Hence the performance of the DBN+RHMFO model is proved under different parametric variations with maximum accuracy and minimum error.

International Journal of Advances in Engineering and Management (IJAEM)
Volume 4, Issue 12 Dec. 2022, pp: 754-768 www.ijaem.net ISSN: 2395-5252



(a) (b) (c) (d)

Fig. 6. Performance Analysis of proposed DBN+RHMFO by varying parameter β with respect to positive measure a) Accuracy b) Precision c) Sensitivity d) Specificity



(a) (b) (c)

Fig. 7. Performance Analysis of proposed DBN+RHMFO by varying parameter β with respect to positive measure a) F-measure b) MCC c) NPV

International Journal of Advances in Engineering and Management (IJAEM)
Volume 4, Issue 12 Dec. 2022, pp: 754-768 www.ijaem.net ISSN: 2395-5252

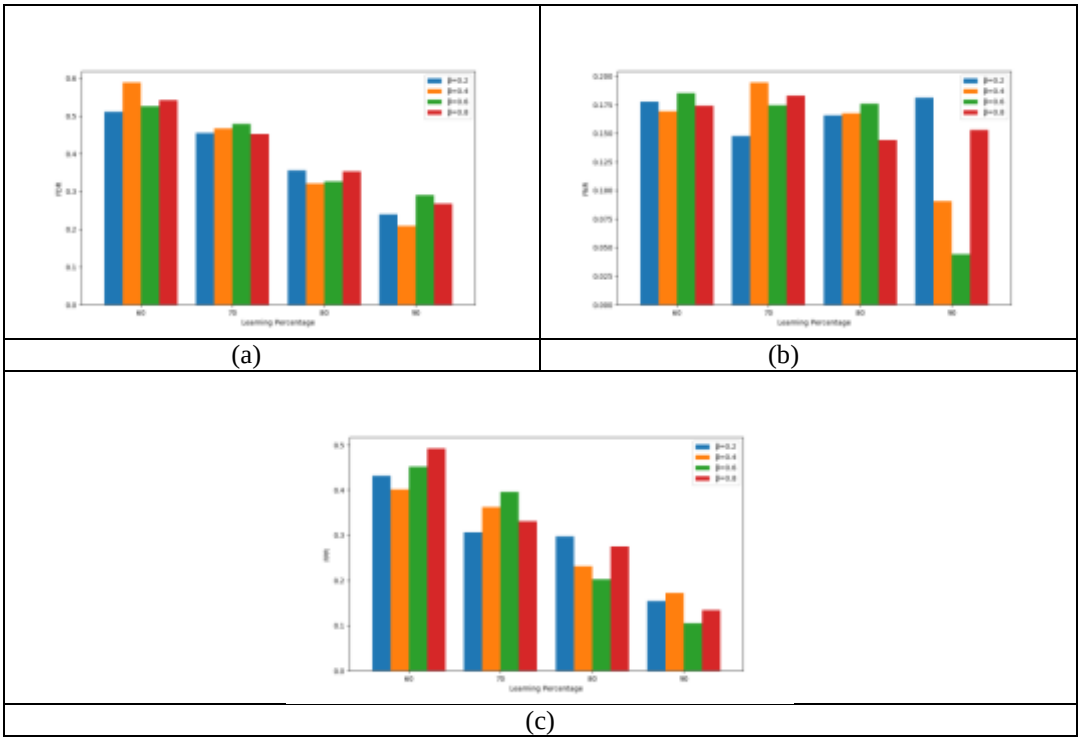


Fig. 8. Performance Analysis of proposed DBN+RHMFO by varying parameter β with respect to positive measure a) FDR b) FNR c) FPR

D. Analysis on Wilcoxon Test, T-test and P-test

TABLE I. WILCOXON TEST

δ =0.2, β =0.2	0.067889
δ =0.4, β =0.4	0.067889
δ =0.6, β =0.6	0.065600
δ =0.8, β =0.8	0.067889

Table I shows the data about Wilcoxon test of the proposed DBN+RHMFO method for the variation under δ =0.2, 0.4, 0.6, 0.8 and β =0.2, 0.4, 0.6, 0.8. The β and δ under the variation 0.2, 0.4 and 0.8 obtained the same value as 0.067889. Next in the variation of β, δ as 0.6 gained the value as 0.065600.

TABLE II. T TEST

δ =0.2, β =0.2	129.593786
δ =0.4, β =0.4	201.795746
δ =0.6, β =0.6	117.647623
δ =0.8, β =0.8	53.697565

Table II illustrates the T test of the proposed DBN+RHMFO model under the variation β =0.2, 0.4, 0.6, 0.8 and δ =0.2, 0.4, 0.6, 0.8. When comparing all the variations, the δ =0.4 and β =0.4 obtained the maximum value as 201.7957. Next, the

International Journal of Advances in Engineering and Management (IJAEM)
Volume 4, Issue 12 Dec. 2022, pp: 754-768 www.ijaem.net ISSN: 2395-5252

variation of β, δ under 0.2 also attain the highest value as 129.5937.

TABLE III. P TEST

δ =0.2, β =0.2	$1.423611e^{-11}$
δ =0.4, β =0.4	$9.992261e^{-13}$
δ =06, β =0.6	$2.542791e^{-11}$
δ =0.8, β =0.8	$2.800312e^{-09}$

Table III represents the P test for the proposed DBN+RHMFO model for the variation under δ =0.2, 0.4, 0.6, 0.8 and β =0.2, 0.4, 0.6, 0.8. The variation β =0.4 and δ =0.4 has the value as $9.992261e^{-13}$. Similarly, δ, β under the variation as 0.8 as $2.800312e^{-09}$.

VII. CONCLUSION

This work has done a parametric and non parametric analysis on the proposed BGP anomaly detection model. the two main steps (i) Feature Extraction (ii) Anomaly Detection" are followed to introduce a novel BGP anomaly detection model. Optimized DBN, was used to determine the presence of attack. Furthermore, a hybrid optimization model known as RHMFO was used to fine-tune the weight of DBN. Finally, The performance of proposed work was evaluated by varying the parameters.

REFERENCES

[1] Elamathi, N., Jayashri, S. & Pitchai, R. Enhanced secure communication over inter-domain routing in heterogeneous wireless networks based on analysis of BGP anomalies using soft computing techniques. Soft Comput 23, 2735–2746 (2019). https://doi.org/10.1007/s00500-019-03836-4

[2] Pradeepa, R., Pushpalatha, M. A hybrid OpenFlow with intelligent detection and prediction models for preventing BGP path hijack on SDN. Soft Comput 24, 10205–10214 (2020). https://doi.org/10.1007/s00500-019-04534-x

[3] M. Karimi, A. Jahanshahi, A. Mazloumi and H. Z. Sabzi, "Border Gateway Protocol Anomaly Detection Using Neural Network," 2019 IEEE International Conference on Big Data (Big Data), 2019, pp. 6092-6094, doi: 10.1109/BigData47090.2019.9006201.

[4] Pablo MorianoRaquel HillL. Jean Camp, "Using bursty announcements for detecting BGP routing anomalies", Computer Networks, 2021.

[5] P. Fonseca, E. S. Mota, R. Bennesby and A. Passito, "BGP Dataset Generation and Feature Extraction for Anomaly Detection," 2019 IEEE Symposium on Computers and Communications (ISCC), 2019, pp. 1-6, doi: 10.1109/ISCC47284.2019.8969619.

[6] S. Abd El Monem, A. Khalafallah and S. I. Shaheen, "BGP Route Leaks Detection Using Supervised Machine Learning Technique," 2020 2nd Novel Intelligent and Leading Emerging Sciences Conference (NILES), 2020, pp. 15-20, doi: 10.1109/NILES50944.2020.9257981.

[7] K. McGlynn, H. B. Acharya and M. Kwon, "Detecting BGP Route Anomalies with Deep Learning," IEEE INFOCOM 2019 - IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS), 2019, pp. 1039-1040, doi: 10.1109/INFCOMW.2019.8845138.

[8] Z. Li, A. L. G. Rios and L. Trajković, "Machine Learning for Detecting Anomalies and Intrusions in Communication Networks," in IEEE Journal on Selected Areas in Communications, vol. 39, no. 7, pp. 2254-2264, July 2021, doi: 10.1109/JSAC.2021.3078497.

[9] N. H. Hammood and B. Al-Musawi, "Using BGP Features Towards Identifying Type of BGP Anomaly," 2021 International Congress of Advanced Technology and Engineering (ICOTEN), 2021, pp. 1-10, doi: 10.1109/ICOTEN52080.2021.9493491.

[10] J. Li, S. Zhou and V. Giotsas, "Performance Analysis of Multipath BGP," IEEE INFOCOM 2021 - IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS), 2021, pp. 1-6, doi: 10.1109/INFOCOMWKSHPS51825.2021.9484523.

[11] T. Matcharashvili and A. Prangishvili, "Quantifying regularity of the Internet Interdomain Routing based on Border Gateway Protocol (BGP) data bases," 2020 International Conference on Electrical, Communication, and

Computer Engineering (ICECCE), 2020, pp. 1-5, doi: 10.1109/ICECCE49384.2020.9179264.

[12] M. Milani, M. Nesler, M. Segata, L. Baldesi, L. Maccari and R. L. Cigno, "Improving BGP Convergence with Fed4FIRE+ Experiments," IEEE INFOCOM 2020 - IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS), 2020, pp. 816-823, doi: 10.1109/INFOCOMWKSHPS50562.2020.9162970.

[13] A. Fiade, M. A. Agustian and S. U. Masruroh, "Analysis of Failover Link System Performance in OSPF, EIGRP, RIPV2 Routing Protocol with BGP," 2019 7th International Conference on Cyber and IT Service Management (CITSM), 2019, pp. 1-7, doi: 10.1109/CITSM47753.2019.8965373.

[14] T. Arai, K. Nakano and B. Chakraborty, "Selection of Effective Features for BGP Anomaly Detection," 2019 IEEE 10th International Conference on Awareness Science and Technology (iCAST), 2019, pp. 1-6, doi: 10.1109/ICAwST.2019.8923583.

[15] Massimo Candela, Giuseppe Di Battista and Luca Marzialetti, "Multi-view routing visualization for the identification of BGP issues", Journal of Computer Languages,2020.