ISSN 2278-3091
Volume 9, No.3, May - June 2020
International Journal of Advanced Trends in Computer Science and Engineering
Available Online at http://www.warse.org/IJATCSE/static/pdf/file/ijatcse25932020.pdf
https://doi.org/10.30534/ijatcse/2020/25932020

An Optimal Grey Wolf Optimization with Fuzzy Support Vector Machine based Intrusion Detection System in Clustered Wireless Sensor Networks

Sibi Amaran[1], Dr. R. Madhan Mohan[2]
[1]Research Scholar, Department of Computer Science and Engineering, Annamalai University, Chidambaram, India, sibi.amaran@gmail.com
[2]Associate Professor, Department of Computer Science and Engineering, Annamalai University, Chidambaram, India, madhanmohan_mithu@yahoo.com

ABSTRACT

Wireless Sensor Network (WSN) includes inexpensive, compact and battery powered sensor nodes to observe the physical parameters exist in the deployed region. Recently, cluster based WSN becomes more popular and achieves energy efficiency. As the sensor nodes undergo deployment in the harsh open environments, it is highly prone to diverse kinds of attacks. This study devises a new Intrusion Detection System (IDS) for clustered WSN (CWSN) using optimal grey wolf optimization (OGWO) and fuzzy support vector machine (FSVM) algorithm. The proposed OGWO algorithm includes a differential evolution (DE) technique for population initialization of GWO algorithm. The OGWO-FSVM algorithm operates on two-stage processes, namely feature selection and classification. The OGWO algorithm selects the optimal subset of features and then FSVM model classifies the data instances into normal and intrusion. The performance of the OGWO-FSVM model has been tested by benchmark dataset. The simulation results indicated that the OGWO-FSVM technique has shown better results over the compared methods significantly.

Key words: Intrusion detection, Clustering, Wireless Sensor Networks, Classification, Feature Selection

1. INTRODUCTION

Recently, various advancements in wireless communication as well as tiny electrical devices have activated the deployment of minimum-cost, less power sensor nodes (SNs) with the abilities of sensing, processing, and communication [1]. Hence, the complexities of WSNs are one of the well-known objectives in this study. A WSN is constructed by the integration of non-infrastructure system, and massive utilization of SNs. Therefore, due to the presence of minimum powered sensors, the SN would exchange data with one another using multi-hop to decrease the power utilization of sensors [2-4]. The main responsibility of WSN is to accumulate and observe the relevant data regarding business, armed forces, healthcare, and ecological monitoring. The function of SNs is to observe the atmosphere and provide the details to the sink or base station (BS) under the application of wireless communication. Followed by, the collected information has been examined to identify the condition of the environment or the target. Based on the hardware design, WSNs meets various resource limitations like minimum processing ability, lower storage, as well as restricted energy.

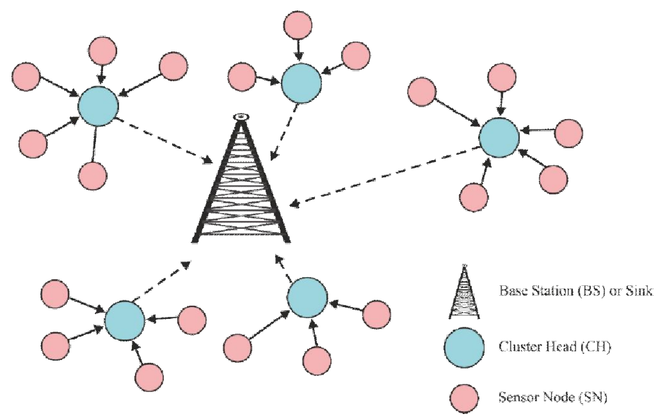

Figure 1: Architecture of CWSN

Cluster based WSN (CWSN) is one of the familiar network topologies in WSN. In CWSN, each SN is clustered, while a Cluster Head (CH) has been selected to control the task of a cluster. The operation of CH is to gather the data from every SN of particular target. Hence, CWSN effectively limits the volume of data throughout the system. The architecture of CWSN is shown in Figure 1. The benefits of CWSN are reduced power consumption, increased network scale, as well as maximum network lifespan [5]. For instance, if a WSN is used in the battlefield, SN is identified and demolished by the opponent. Specifically, the privacy of WSN needs to be important. An anticipation method has been applied to find popular attacks. Also, a corresponding prevention model is established on the basis of attacking features. Unfortunately, not all prevention models are capable of avoiding the attacks. Apart from these complications, the initial step is to detect the attack. An Intrusion Detection System (IDS) is employed prominently to predict the presence of attackers. In addition, IDS help to deploy the prevention method by applying characteristics of attacks.

The IDS is treated as the network monitor. An alarm signal has been raised by IDS, which prevents the network destruction. The 2 main objectives involved in IDS are Anomaly detection and Misuse detection [6]. Initially, Anomaly detection develops a model on the basis of comparing ordinary and predicted nature. This kind of detection is comprised of maximum detection value and false positive rate. Alternatively, misuse detection examines the type of an attack type by comparing the previous attacking behaviour and recent attacking behaviour. It is composed with greater accuracy; however, detection rate is minimum. In particular, misuse detection is not capable of detecting the unpopular attacks. In order to provide a solution for these issues, developers have proposed hybrid detection method which gathers the merits of anomaly detection and misuse detection. This kind of interaction could predict the unidentified attacks with best forecasting value of anomalous detection and maximum accuracy of misuse identification. The Hybrid IDS (HIDS) accomplish the main objectives of higher detection rate and lower false positive rate. This model is applied for a heterogeneous CWSN as defined in the upcoming section. Figure 2 shows the sample of attacks in WSN.

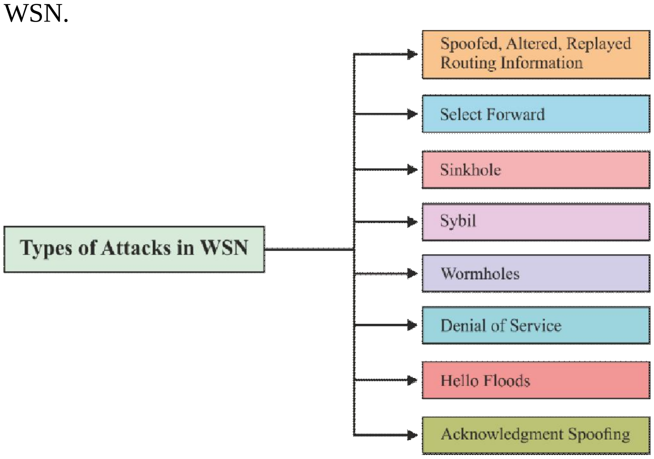

Figure 2: Types of Attacks in WSN

Kuila. P & Jana. P.K[7] projected 2 technologies on the basis of Particle Swarm Optimization (PSO) for Linear/Nonlinear Programming (LP/NLP) types of issues namely clustering and routing. They are 2 familiar optimizing issues that have been researched extensively and expand the lifetime of WSN. In this approach, routing models were introduced with effective particle encoding framework and multi-objective fitness function (FF). Gupta. V & Sharma. S.K [8] defined regarding the selection of CH in WSN with the application of Swarm Intelligence (SI). It was evolved from LEACH clustering technology. An extended edition of Ant Colony Optimization (ACO) is derived by applying Residual Energy (RE) across LEACH model to assist CH selection process. The major advantage of this approach is that, it has minimum power consumption. It is operated in 3 phases namely cluster members (CM), cluster head (CH) and leader.

Ahmad. I et al. [9] introduced a Genetic Algorithm (GA) to explore the genetic principal units which provides a feature subset with best sensitivity and optimal discriminatory power. The proper assortment of principal components becomes a challenging aspect in subset selection. In prior to precede the classification, the actual dataset undergoes pre-processing in 3 diverse modules like Neglecting symbolic measures, Feature transformation with the application of PCA and Best features subset selection by applying GA. At last, SVM model has been employed to process the classification task. Ni. Q et al.[10] described a solution for the complexities involved in CH selection that is assumed as a significant process in WSN. The attained solution is relied on fuzzy clustering pre-processing and PSO. In a deep study, fuzzy clustering model has been employed for SN clustering, in which SN comes under a cluster with appropriate possibility and number of primary clusters were examined and defined. Moreover, the FF was developed by the assumption of power utilization as well as distance factors of WSN. Consequently, the CH nodes present in a hierarchical topology was computed according to the Improved PSO (IMPSO).

Wang. G et al. [11] projected a method named FC-ANN evolved from ANN and Fuzzy clustering, that helps to resolve the issues and attains maximum prediction values, minimum false positive rate, and robust stability. Under the application of fuzzy clustering, the heterogeneous training set was portioned into various homogenous subsets. Initially, fuzzy clustering model was applied for producing diverse training subsets. Consecutively, training subsets were generated using various ANN models, which undergo training on developing diverse base models. At last, a meta-learner, fuzzy aggregation module, has been applied to perform the result aggregation.

Walid Balid et al. [12] implied that Real-time traffic observation is most required one in recent smart transportation models that is highly applicable in real-time applications. This framework states the design and execution of new intelligent wireless sensor for traffic surveillance. It is effective in processing and acts as a stable approach for transport prediction, rapid length evaluation, classification, and time-synchronization was completely formed, combined, and determined. Comprehensive system evaluation and wider data estimation was processed for tuning and verifying the network stability as well as stronger computation. Numerous works have been carried out on highway and urban roads for diverse road scenes and with several traffic constraints provided maximum accuracy, better speed estimation, and optimal length-based vehicle classification accuracy. Hence, the newly presented method is highly transferable, stable, consistent, and cost-efficient.

This paper introduces new IDS using optimal grey wolf optimization (OGWO) and fuzzy support vector machine (FSVM) algorithm in CWSN. The proposed OGWO-FSVM model involves two major processes, namely feature selection and classification. In the first stage, OGWO technique is applied to select the subset of features in an effective way. Besides, the OGWO algorithm is derived by applying differential evolution (DE) approach in the population initialization of GWO algorithm. Followed by, in the second stage, FSVM is used for classifying the data instances into appropriate set of classes. The proposed OGWO-FSVM model has the ability to predict the network data into Normal and Intrusion.

2. THE PROPOSED OGWO-FSVM MODEL

The overall working process involved in the presented OGWO-FSVM model is shown in Figure 3. As shown, the input data undergo preprocessing to normalize the data. Next, OGWO relied feature extraction is carried out to filter the feature subset from the preprocessed data. In GWO algorithm, population initialization process has been conducted by applying of DE. At last, the FSVM model is carried out to classify the data into its respective classes.

2.1 DE-GWO based Feature Selection

The GWO is defined as novel meta-heuristic approach coined by Mirjalili et al. [13], that reflects the social and hunting process of grey wolves that depends upon 3 major levels namely,

- Encircling prey
- Hunting
- Attacking prey

The mathematical function can be defined for leadership hierarchy of wolves by considering alpha, beta, delta, and omega. Initially, the best solution is alpha, the consecutive 2 solutions are beta and delta, and finally the residual participants are named as omega.

At the first stage, prey is encircled by grey wolves at the time of hunting period. The numerical expression of encircling hierarchy of grey wolves has been represented as:

$$\vec{D} = |\vec{C} \cdot \vec{Z}_{prey}(n) - \vec{Z}_{wolf}(n)|,$$
$$\vec{Z}_{wolf}(n+1) = \vec{Z}_{prey}(n) - \vec{A} \cdot \vec{D} \qquad (1)$$

where n is the current round, $\vec{A}$ and $\vec{C}$ are coefficient vectors, $\vec{Z}_{prey}$ indicates the position vector of the prey, and $\vec{Z}_{wolf}$ denotes the position vector of a grey wolf. Vectors $\vec{A}$ and $\vec{C}$ are determined by:

$$\vec{A} = 2\vec{a} \cdot \vec{r}_1 - \vec{a}, \vec{C} = 2\vec{r}_2, \qquad (2)$$

where $\vec{a}$ implies linearly decreased reduced values from 2 to 0 and $\vec{r}$ and $\vec{r}$ are random vectors within [0,1].

Basically, hunting operation is processed by alpha. Beta and delta are also considered rarely. The mathematical expression symbolizes the hunting of grey wolves, the 3 optimally derived solutions are alpha, beta, and delta were saved and remaining one is omega that is compelled to extend the positions on the basis of Eqs. (3) $-(9)$. The updated positions for grey wolves are depicted as follows.

$$\vec{D}_{alpha} = |\vec{C} \cdot \vec{Z}_{alpha} - \vec{Z}|, \qquad (3)$$
$$\vec{D}_{beta} = |\vec{C}_2 \cdot \vec{Z}_{beta} - \vec{Z}|, \qquad (4)$$
$$\vec{D} = |\vec{C}_3 \cdot \vec{Z}_{delta} - \vec{Z}|, \qquad (5)$$
$$\vec{Z}_1 = \vec{Z}_{alpha} - \vec{A}_1 \rightarrow \vec{D}_{alpha}, \qquad (6)$$
$$\vec{Z}_2 = \vec{Z}_{beta} - \vec{A}_2 \cdot \vec{D}_{beta}, \qquad (7)$$
$$\vec{Z}_3 = \vec{Z}_{delta} - \vec{A}_3 \cdot \vec{D}_{delta}, \qquad (8)$$
$$\vec{Z}(n+1) = \frac{\vec{Z}_1 + \vec{Z}_2 + \vec{Z}_3}{3}. \qquad (9)$$

The DE-GWO model presented a novel processing approach, OGWO, for the purpose of medical diagnosis [14]. OGWO is composed with 2 major stages. Initially, OGWO is mainly applied for removing the repeated data by adaptive searching for optimal feature integration in the clinical data. In the newly developed OGWO, DE has been applied for generating the recent positions of population, and GWO is used to upgrade the new positions in a discrete searching space. Secondly, the efficient FSVM classification is carried out on the basis of attained best feature subset. Additionally, OGWO is utilized for searching a feature space for optimal feature combination.

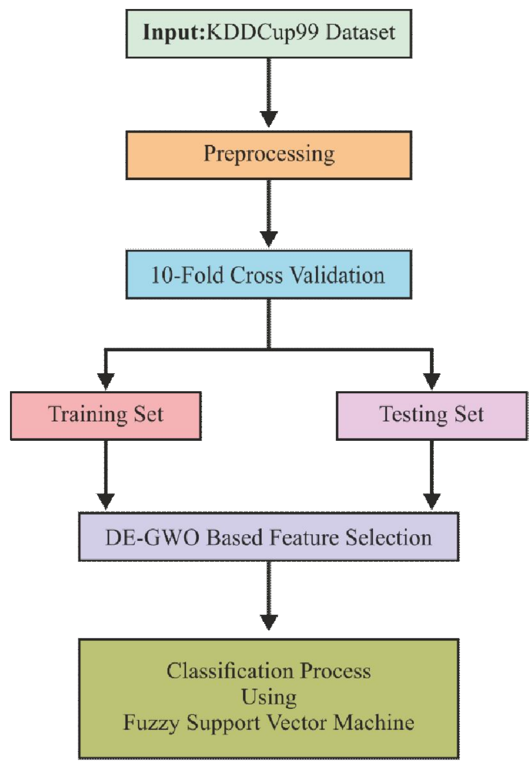

Figure 3: Workflow of OGWO-FSVM model

The FF applied in OGWO to estimate the chosen features is defined by:

$$\text{Fitness} = \alpha P + \beta \frac{N - L}{N}, \qquad (10)$$

where P denotes the classification accuracy, L defines the length of feature subset selected, Nimplies overall features, and α and β are 2 parameters corresponds to the weight of classification accuracy and the superiority of FS, $\alpha \in [0,1]$ and $\beta = 1 - \alpha$.

The DE is evolved from population-dependent direct searching model. Based on the relative studies [15], DE performs quite well than other Evolutionary Algorithms (EA) and standard functions as well as real-world optimization issues. Consider that $P_{l,Gen}, (l = 1,2, \dots, n)$ are nN_v-dimensional parameter vectors of production Gen(n is a constant). To produce a novel population of vectors, for every target vectors, 3 vectors are selected arbitrarily and weighted difference of 2 has been included with third 3rd vector.For traditional DE, the steps are provided in the following:

(a) Creating difference offspring: For every vector l from generation Gen, a mutant vector $V_{l,Gen+1}$ is represented as,

$$V_{l,Gen+1} = P_{\mathrm{r}_1,\mathrm{Gen}} + F\left(P_{\mathrm{r}_2,\mathrm{Gen}} - P_{\mathrm{r}_3,\mathrm{Gen}}\right) \quad (11)$$

where $l = \{1,2, \dots, n\}$ and r_1, r_2, and r_3 are mutually arbitrary integer indices has been selected from $\{1, 2, \dots, n\}$. Furthermore, l, r_1, r_2, and r_3 are different so $n \geq 4$. $F \in [0,2]$ is a real constant that computes amplification of additional differential vector of $(P_{r2,Gen} - P_{r3,Gen})$. Maximum values for F results in greater diversity in the produced population as well as lower values increase the convergence.DE applies crossover operation for enhancing the population diversity. It describes the following trial vector:

$$U_{l,Gen+1} = \left(U_{1l,Gen+1}, U_{2l,Gen+1}, \dots, U_{N_v l,Gen+1}\right), \quad (12)$$

where $= 1,2, \dots, N_v$ and

$$U_{ml,Gen+1} = \begin{cases} V_{ml,Gen+1} & if\ rand_m\ (0,1) \leq CR \\ P_{ml,Gen} & otherwise \end{cases} \quad (13)$$

$CR \in (0,1)$ implies the crossover constant and $rand_m(0,1) \in [0,1]$ is mth estimation of uniform random generator. The well-known measure of CRwithin (0.4, 1).

(b) Fitness evaluation of trial vector

(c) Selection: The selection model selects a vector, $U_{l,Gen+1}$ or $P_{l,Gen}$, which has to be the candidate of novel generation, $Gen + 1$. A vector can be selected with best fitter measures. Additionally, the variants of DE should provision the comparison, the conventional model of DE is chosen for depicting the convergence enhancement by opposition-based population initialization.

2.2. FSVM based Classification

The classical SVM states that, every data point is applied with similar dimension and allocated with equal penal parameter from objective function. The real-time classification domains are sample points, like noises, which is nor designated correctly and every sample point does not have similar semantics in a selection phase. In order to resolve these issues, the theory of FSVM was presented in [16]. Fuzzy membership for a sample point is deployed in which diverse sample points are capable of making various building of decision surface:

$$S = \{(p_u, q_u, s_u), u = 1, \dots, N\} \quad (14)$$

where $p_u \in R^n$ is the n-dimension sample point, $q_u \in \{-1, +1\}$ is the class label, and $s_u (u = 1, \dots, N)$ is a fuzzymembership that meets $\sigma \leq s_u \leq 1$ with an adequate minimum constant $\sigma > 0$. The quadratic optimization problem in classification can be assumed as:

$$\min_{w,s,\xi} \ \frac{1}{2} w^T w + C \sum_{u-1}^{l} s_u \, \xi_u \quad (15)$$

$$s.t.\, q_u(w^T p_u + b) \geq 1 - \xi_u, \xi_u \geq 0, u = 1, \dots, l,$$

where w implies a normal vector of dividing hyperplane, $b$shows a bias term, and Cdefines the parameter that should be computed in prior to manage the trade-off among the classification margin and expense of misclassification error. s_u is the attitude of corresponding point $p_u$to the direction of class and the slack variables ξ_u are the error values, and $s_u \xi_u$is a measure of error with various weights. It is pointed that s_u is maximum that the corresponding point becomes mandatory; as minimum the s_u is the corresponding point become lesser important and various input points make several contributions to learn the decision surface. Hence, FSVM is suitable of finding massive efficient hyperplane by enhancing the margin by affording misclassification of lower vital points. For resolving the FSM optimal issue, Eq. (16) is converted as dual problem by establishing Lagrangian multiplier,α_u:

$$\max \sum_{u=1}^{N} \alpha_u - \frac{1}{2} \sum_{u=1}^{N} \sum_{v=1}^{N} \alpha_u \, \alpha_v q_u q_v p_u p_v \quad (17)$$

$$s.t. \sum_{u=1}^{N} q_u \, \alpha_u = 0, 0 \leq \alpha_u \leq s_u C, u = 1, \dots, N.$$

When related with the remarkable SVM, the above-mentionedstatement has minimum variations that are upper bound of values of α_u. By resolving the dual problem in (3) for best α_u, w and $b$is improved as standard SVM as well.

3. PERFORMANCE VALIDATION

A series of experiments were carried out using Intel®-core ™ i7-7500 2.70-2.90 GHz CPU processor, 8 GB memory and running Windows 10 OS (64-bit). The presented model has been implemented in MATLAB R2014b version.

3.1. Dataset Description

For experimentation, KDD Cup99 dataset is the commonly utilized for the performance validation of OGWO-FSVM model. The applied dataset includes duplicate, repetitive, and imbalanced data that mainly affect the impact of the validated IDS. It comprises a set of 125,973 samples, with 41 features and 2 classes and 4 subclasses. The class distribution of the KDD Cup99 dataset is depicted in Figure 4.

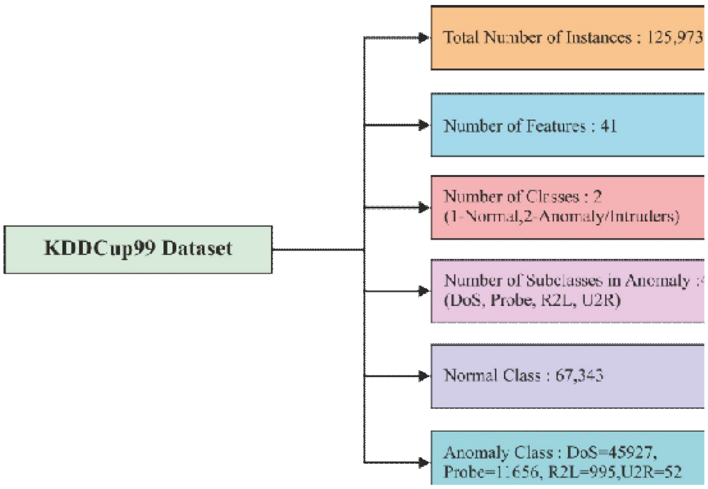

Figure 4: Class distribution of KDDCup99 Dataset

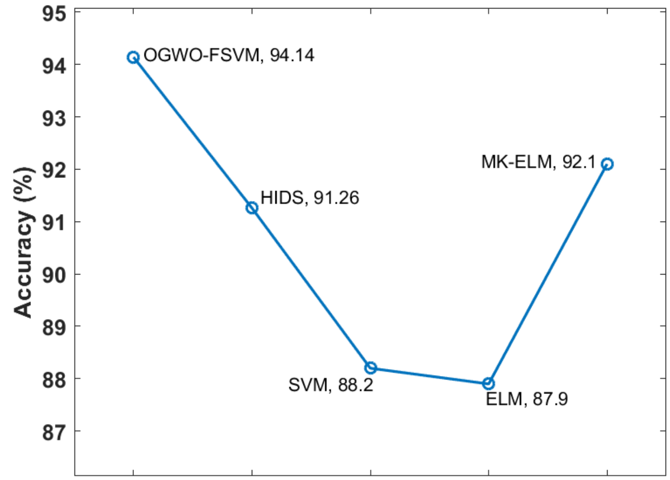

Figure 5: Accuracy analysis

3.2. Results Analysis

Table 1 illustrates the relative study of the results attained by the OGWO-FSVM model with other existing models [17, 18] on the applied KDDCup99 dataset. Figure 5 implies the accuracy analysis of the OGWO-FSVM model with existing methods. The figure stated that the ELM has exhibited ineffective classifier outcome with the least accuracy of 87.90%. At the same time, the SVM model has showed slightly better outcome with the accuracy of 88.20%. In line with, the HIDS model has demonstrated moderate results with the accuracy of 91.26%. Along with that, the MK-ELM model has resulted to even higher and closer results to OGWO-FSVM model with the accuracy of 92.10%. However, the OGWO-FSVM model has outperformed all the previous models by attaining a maximum accuracy of 94.14%.

Table 1: Classification results analysis of OGWO-FSVM model

Performance Indicators	OGWO-FSVM	HIDS	SVM	ELM	MK-ELM
Accuracy	94.14	91.26	88.20	87.90	92.10
FPR	7.59	-	2.34	3.76	2.37
TPR	95.73	-	83.73	83.84	89.42
FNR	4.26	-	16.27	16.16	10.58
Detection Rate	95.23	90.96	74.74	75.53	83.81

Figure 6 implies the FNR analysis of the OGWO-FSVM method with previous models. From the figure, it is clear that the SVM has defined inefficient classifier result with the higher FNR of 16.27. Simultaneously, the ELM approach has depicted moderate outcome with the considerable FNR of 16.16. In line with this, the MK-ELM technique has concluded with least and identical to OGWO-FSVM model with the FNR of 10.58. But, the OGWO-FSVM model performs quite-well than the existing approaches by reaching minimum FNR of 4.26.

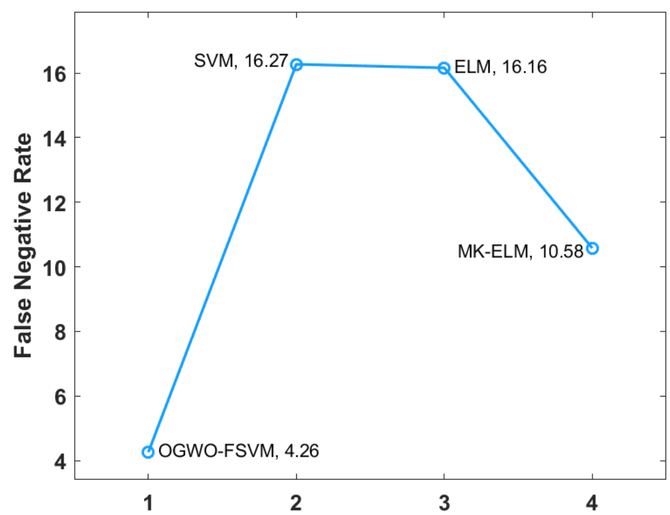

Figure 6: FNR Analysis

Figure 7 defines the FPR analysis of the OGWO-FSVM model with conventional methods. The figure portrays that, the ELM has implemented poor classification result with the lower FPR of 2.34. Concurrently, the MK-ELM approach has exhibited gradual outcome with the FPR of 2.37. Similarly, the SVM model has illustrated slightly better results with the FPR of 2.34. Hence, the OGWO-FSVM framework has outstanding performance when compared with existing models by accomplishing a greater FPR of 7.59.

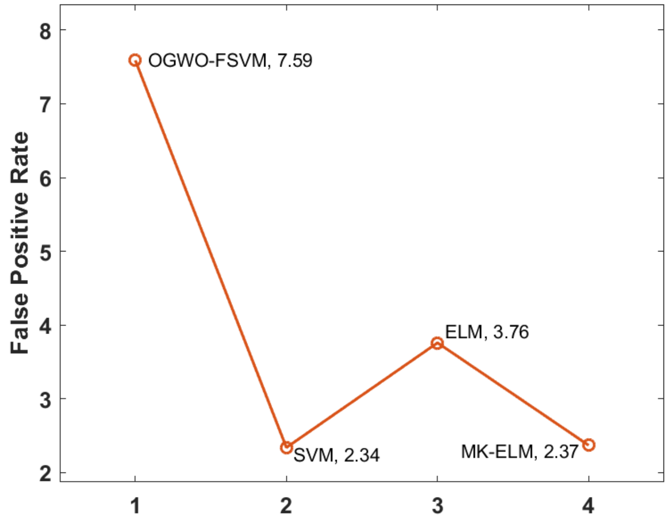

Figure 7: FPR Analysis

Figure 8 showcases the TPR analysis of the OGWO-FSVM technology with classical methods. The figure exhibited that the SVM has shown inferior classifier simulation outcome with less TPR of 83.73%. Meanwhile, the ELM method has defined manageable result with the TPR of 83.84%. Likewise, the MK-ELM approach has exhibited gradual results with the TPR of 89.42%. But, the OGWO-FSVM model is an optimal one when compared with classical approaches by reaching best TPR of 95.73%.

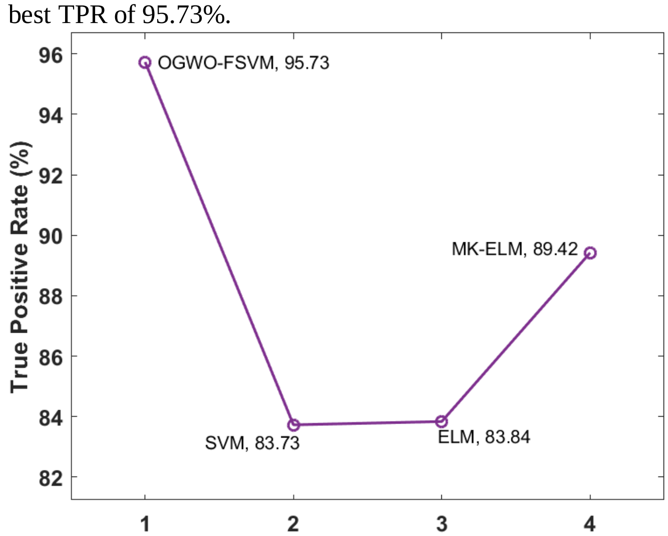

Figure 8: TPR Analysis

Figure 9 depicts the detection rate analysis of the OGWO-FSVM model with previous technologies. From the figure, it is evident that the SVM has implied unproductive classifier outcome with the lessen detection rate of 74.74%. Meantime, the ELM model has defined considerable outcome with the detection rate of 75.53%. On the same way, the MK-ELM model has shown better results with the detection rate of 83.81%. In line with this, the HIDS approach has exhibited even better and identical to OGWO-FSVM model with the detection rate of 90.93%.

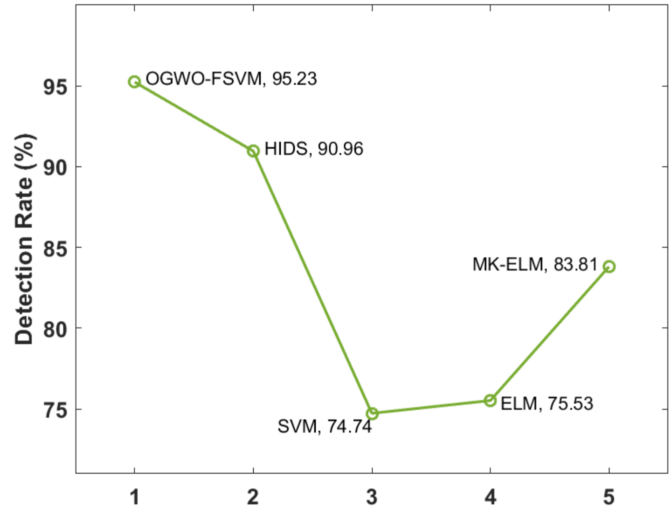

Figure 9: Detection Rate Analysis

Therefore, the OGWO-FSVM model has shown effective results than existing models by achieving optimal detection rate of 95.23%. From the above-mentioned results, it is evident that the OGWO-FSVM model has found to be an effective model over the compared methods.

4. CONCLUSION

This paper has developed effective IDS using OGWO and FSVM algorithm in CWSN. The proposed OGWO-FSVM model involves two major processes, namely feature selection and classification. Initially, the input data undergo preprocessing to normalize the data. Next, OGWO based feature extraction task is carried out for extracting the feature subset from the preprocessed data. At last, the FSVM model is carried out to classify the data into its respective classes. For experimentation, KDD Cup99 dataset is the commonly applied for ensuring the results of the OGWO-FSVM model. The simulation outcome indicated that the OGWO-SVM model has outperformed the compared methods with the higher accuracy of 94.14%, detection rate of 95.23% and TPR of 95.73%. Therefore, the OGWO-FSVM model can be employed as an appropriate tool for IDS in CWSN.

REFERENCES

[1] T. Rupa Rani, G. Lavanya, Dr. H. Shaheen, R. Anusha, and K. Nartkannai, **Open IOT Service Platform Technology with Semantic Web,** International Journal of Advanced Trends in Computer Science and Engineering, Vol. 9, pp.17-21, 2020. https://doi.org/10.30534/ijatcse/2020/0491.12020

[2] Gleen A. Dalaorao, Ariel M. Sison, Ruji P. Medina, **Applying Modified TF-IDF with Collocation in Classifying Disaster-Related Tweets,** International Journal of Advanced Trends in Computer Science and Engineering, Vo. 9, pp. 28-33, 2020. https://doi.org/10.30534/ijatcse/2020/0691.12020

[3] Shuhaida Mohamed Shuhidan, Nur Aini Shuhada Fadzil and Saidatul Rahah Hamidi, **The Visualization of Sentiment Analysis for Malaysia Financial News Headlines,** International Journal of Advanced Trends in Computer Science and Engineering, Vol. 9, pp. 589-594, 2020.

[4] J. Uthayakumar, T. Vengattaraman, and P. Dhavachelvan. **A new lossless neighborhood indexing sequence (NIS) algorithm for data compression in wireless sensor networks,** Ad Hoc Networks, Vol. 83, 2019, pp. 149-157. https://doi.org/10.1016/j.adhoc.2018.09.009

[5] K. Akkaya, and M. Youngish. **A survey on routing protocols for wireless sensor networks,** Ad Hoc Network, Vol. 3, no. 3, 2005, pp. 325–349.

[6] R. A. Kemmerer, and G. Vigna. **Intrusion detection a brief history and overview,** Computer, Vol. 35, no. 4, 2002, pp. 27–30. https://doi.org/10.1109/MC.2002.1012428

[7] P. Kuila, and P.K. Jana. **Energy efficient clustering and routing algorithms for wireless sensor networks: particle swarm optimization approach,** Eng. Appl. Artif. Intell. Vol. 33, 2014, pp. 127–140.

[8] V. Gupta, and S.K. Sharma. **Cluster head selection using modified ACO,** Proceedings of Fourth International Conference on Soft Computing for Problem Solving, 2015, pp. 11–20. https://doi.org/10.1007/978-81-322-2217-0_2

[9] I. Ahmad, M. Hussain, A. Alghamdi, and A. Alelaiwi. **Enhancing SVM performance in intrusion detection using optimal feature subset selection based on genetic principal components,** Neural Comput. Appl. Vol. 24, no. 7-8, 2014, pp. 1671–1682.

[10] Q. Ni, Q. Pan, H. Du, C. Cao, and Y. Zhai. **A novel cluster head selection algorithm based on fuzzy clustering and particle swarm optimization,** IEEEACM Trans. Comput. Biol. Bioinform. Vol. 14, no. 1, 2017, pp. 76–84.

[11] G. Wang, J. Hao, J. Ma, and L. Huang. **A new approach to intrusion detection using Artificial Neural Networks and fuzzy clustering,** Expert Syst. Appl. Vol. 37, no. 9, 2010, pp. 6225–6232.

[12] W. Balid, H. Tafish, and H. H. Refai. **Intelligent vehicle counting and classification sensor for real-time traffic surveillance,** *IEEE Transactions on Intelligent Transportation Systems*, Vol. 19, no. 6, 2017, pp.1784-1794. https://doi.org/10.1109/TITS.2017.2741507

[13] S. Mirjalili, S. M. Mirjalili, and A. Lewis. **Grey Wolf Optimizer,** Advances in Engineering Software, Vol. 69, 2014, pp. 46–61.

[14] Q. Li, H. Chen, H. Huang, X. Zhao, Z. Cai, C. Tong, W. Liu, and X. Tian. **An enhanced grey wolf optimization based feature selection wrapped kernel extreme learning machine for medical diagnosis,** *Computational and mathematical methods in medicine*, 2017. https://doi.org/10.1155/2017/9512741

[15] S. Rahnamayan, H. R. Tizhoosh, and M. M. Salama. **A novel population initialization method for accelerating evolutionary algorithms,** *Computers & Mathematics with Applications*, Vol. 53, no. 10, 2007, pp. 1605-1614.

[16] X. Gu, T. Ni, and H. Wang. **New fuzzy support vector machine for the class imbalance problem in medical datasets classification,** The scientific world journal, 2014. https://doi.org/10.1155/2014/536434

[17] S. S. Wang, K. Q. Yan, S. C. Wang, and C. W. Liu. **An integrated intrusion detection system for cluster-based wireless sensor networks,** Expert Systems with Applications, Vol. 38, no. 12, 2011, pp. 15234-15243. https://doi.org/10.1016/j.eswa.2011.05.076

[18] W. Zhang, D. Han, K. C. Li, and F. I. Massetto. **Wireless sensor network intrusion detection system based on MK-ELM,** *Soft Computing*, 2020, pp. 1-14. https://doi.org/10.1007/s00500-020-04678-1