Nanotechnology Perceptions
ISSN 1660-6795
www.nano-ntp.com

Multi Support Vector Machine to Improve Digital Image Forgery Detection

Yerragudla Vagdevi[1], J. Balaraju[2]

[1]Post Graduate Student (M.Tech.), Department of CSE, Anurag University, India
[2]Associate Professor, Department of CSE, Anurag University, India
Email: vagdeviyerragudla@gmail.com

The feature selection method proposed here is a hybrid optimization algorithm, combining the best features of the MSVM and GWO. In order to eliminate unnecessary, noisy, or irrelevant characteristics from high-dimensional datasets, this methodology primarily aims to provide a trustworthy data dimensionality reduction method. Combining GWO, Lagrange interpolation, and MSVM, our novel multi-strategy fusion algorithm is called MSVM-GWO. The idea for this mashup came from a cross between animal group cooperative conduct, the Chinese phrase "Chai Lang Hu Bao," and hybrid algorithms. In its initial implementation, the MSVM-GWO algorithm was able to handle eight challenging benchmark functions. Using MSVM-GWO, we were able to resolve ten issues with feature selection in Case 2. Whether we are testing benchmark functions or feature selection, MSVM-GWO consistently produces smaller means, lower standard deviations, improved classification accuracy, and faster execution times in our trials. Such results show that when compared to its rivals, the MSVM-GWO algorithm performs better in terms of stability, classification accuracy, and optimisation efficiency.

Keywords: Algorithm, Classification, Accuracy, GWO, MSVM.

1. Introduction

Software-Defined Networking (SDN) is a relatively new approach to network design. By taking control away from the data plane and giving it to a network controller, software-defined networking (SDN) introduces a new way of thinking about network design and improves programmability. In order to overcome the shortcomings of conventional networks and offer superior control and services, Computer Numerical Control (CNC) divisions have emerged. Its physically centralized controller, however, makes it vulnerable to issues with consistency, dependability, and scalability [1][2].

Security for software-defined networks (SDNs) is becoming increasingly important as a result of the benefits it offers to computer networks. When it comes to security in software-defined

networks, network intrusion detection is a crucial tool. Based on our analysis, the OPL-3 node clusters are the most resilient to distributed denial of service assaults, whereas the ONOS and ODL SDN controller types were shown to be less so. Following the deployment of the SDN controllers to a cloud environment, we carried out the distributed denial of service (DDoS) attacks described in [3][4].

Then, we examined the vulnerabilities. In order to identify SDN attacks, the authors of [5] used a combination of Random Forest and SDN controllers to train a model using a feature elimination pattern, and then they compared the performance of feature selection classifiers with that of machine learning classifiers. By enhancing CNN with the firefly algorithm following an increase in population diversity, the authors of [6] were able to use CNN to intrusion detection in SDN. They then suggested an SDN with loT (SD-loT) architecture that could identify DDoS attacks on SD-loT, guaranteeing the accuracy of regular traffic. By keeping an eye on and analyzing data such as system logs, network traffic, and other information, a Network Intrusion Detection System (NIDS) can identify and prevent harmful network intrusions. Researchers in the field of network security have increasingly focused on network intrusion detection systems (NIDS) due to the growing number of individuals worried about the security of their networks. The foundation of network intrusion detection systems (NIDS) is supported by both traditional NIDS and software-defined NIDS, also known as SDN-IDS[7].

Data dimension is an essential component of NIDS because high-dimensional data contains several irrelevant or redundant aspects that might impact classification speed or accuracy [8]. This is applicable to both SDN-IDS and non-SDN-IDS scenarios. Therefore, feature selection stands out as a major strategy for NIDS performance improvement; this method seeks to build a more understandable model by reducing the dimension of the feature space. Reducing the dimension usually results in greater learning performance. The authors of [9][10] created a federated learning-based and security NIDS to identify intrusions on networks that use software-defined networking (SDN) as well as those that do not.

They tested their system on three separate datasets: CIC-IDS2018, MQTTest, and InSDN. Using the ToN-loT and InSDN datasets, we proposed a two-stage automated NIDS in [11] that evaluates the Intrusion Detection System's accuracy and precision in SDN and no-SDN situations using long-term and short-term measures. After its proposal, the SDN-SlowRate-DDoS dataset underwent validation through 23 independent tests. Each trial began with configuring the network topology and establishing the assault flow parameters (attackers, victims, etc.) using the Long Short-Term Memory (LSTM) model. What follows is a temporary or permanent disconnect, depending on how bad the attack flows are. Much discussion has led to the conclusion that 4 is the sweet spot for SDN-IDS attack-victim ratios. In order to detect distributed denial of service (DDoS) attacks that start with VANET for software-defined networking (SDN), the authors of [12][13] introduced a Radial Basis Function (RBF) kernel for Support Vector Machine (SVM) and an exhaustive parameter search method called Grid Search Cross Verification.

To further demonstrate the model's supremacy, we tune the RBF kernel using C and V, evaluate its temporal complexity, and then compare it to other ML techniques at CICIDS 2019. This is done after proving the RBF kernel's superiority over the Poly, Linear, and Sigmod

kernels. Hadoop was proposed by the authors of [14] as a component of a hybrid feature selection method for identifying the most desirable attributes. Next, we used DARPA DDoS, CAIDA "DoS attack 2007," and CICIDS "DoS attack 2017" to evaluate the algorithm's performance compared to its predecessor.

2. Literature Survey

MSVM is a novel meta-heuristic approach that was initially suggested in 2022 by [15]. The MSVM has found extensive use in numerous domains because to its simplicity, resilience, and ease of implementation. To better estimate the performance of air cavity solar stills made of aluminum and polycarbonate, for example, In [16] suggested using the MSVM to enhance the network model. To forecast the tribological characteristics of nanocomposites of alumina-coated silver-reinforced copper, the authors of [17] integrated the LSTM model with the MSVM. Tested a binary LEO-MSVM method on the UCI dataset; it uses MSVM to identify features and incorporates LEO escape factors.

Implemented feature extraction using the MSVM and verified it using the PROMISE dataset as described in [18]. Not only that, but they contrasted various classifiers, including DT, Naive Bayes, QDA, and K-Nearest Neighbors (KNN). In addition, we used a number of algorithms to compare it to the MSVM-JOS algorithm under CEC 2017 and show that it was superior. This algorithm was a combination of the MSVM and the Joint Opposite Strategy. Further proof of the enhanced algorithm's engineering utility was provided by their Wilcoxon rank sum test of the MSVM-JOS. And they suggested QEMSVMA, an optimization algorithm for quantum evolutionary MSVM, which increased the duty cycle and yielded a sensor deployment strategy. The algorithm's complexity was decreased while its accuracy was enhanced.

Here are the four sections of our previous work [19]. We began by introducing a combined feature selection method that utilised a weighted k-nearest neighbour algorithm as part of an integrated optimisation strategy. We compared the algorithm's performance on non-SDN-IDS data to that of the KDD Cup99 dataset, which incorporates feature selection and is divided into four categories using weighted KNN. At the end of the day, we discovered that the algorithm can make non-SDN-IDS more accurate. Additionally, we tested our jumping spider optimization method on the UNSW-NB15 and KDD Cup99 datasets, and the outcomes demonstrated that it enhanced the non-SDN-IDS's accuracy and convergence speed. Combining the Harris Hawk method with small-hole imaging is the essence of this strategy. Thirdly, we suggested an optimised Harris Hawk approach using Sin chaotic mapping for the Control Placement Problems (CPP) scheme's initialization adjustments.

In experiments, the proposed technique yields a more resilient CPP, and once the format of the CPP is obtained, the diversity of the CPP is improved by finding the CPP with the Pareto front using the Cauchy variation.

As a fourth point, we suggested a hybrid of the butterfly and black widow optimization algorithms for NIDS feature selection. For our simulation tests on binary and multi-classification, we have chosen the UNSW-NB15 dataset. The experimental findings demonstrate that the suggested technique can decrease feature dimensions and improve feature selection model performance in non-SDN-IDS.

Metaheuristic algorithms are now the de facto standard for combinatorial optimization problems [20]. Their generalizability, broad global search capabilities, and simplicity of heuristics make them useful in many domains. Metaheuristic algorithms are a great way to solve feature selection problems without using the downsides of traditional optimisation approaches. They work wonders when used as search strategies for feature subset selection. As a result, huge strides have been made in the field of feature selection thanks to the development of a plethora of metaheuristic algorithms.

The Genetic Algorithm, Ant Colony Optimisation, Grey Wolf Optim1izer, Whale Optimisation Algorithm, Multi-verse Optimizer, Salp Swarm Algorithm, Atom Search Optimisation, Harris Hawks Optimizer, Grasshopper Optimisation Algorithm, and Sooty Tern Optimisation Algorithm are all notorious examples of such algorithms.

3. Proposed Model

3.1. Proposed MSVM algorithm

Chopra et al.'s Multi Support vector machine (MSVM) algorithm mimics the natural patterns of population density and hunting tactics used by MSVM. This new metaheuristic algorithm mimics the whole hunting process of a MSVM, including finding prey, following it, and finally attacking. It does this by using mathematical modeling approaches. The MSVM algorithm treats the entire population as a set of possible first-viable solutions. The method iteratively updates the population to simulate the search, tracking, surrounding, and attacking behavior of a MSVMpopulation until the pack catches its prey, which is the halting condition.

This criterion is satisfied when there is no discernible change from one generation of the population to the next, which means that the optimal solution or collection of solutions has been found. There are four primary steps to the MSVM algorithm.

(1) Population initialization—Algorithm initialization

The MSVM algorithm begins with a randomly dispersed initial population across the search space, similar to other metaheuristic algorithms. As a definition, it is:

$$Y_0 = Y_{\min} + rand(Y_{\max} - Y_{\min})$$

In this case, Y0 stands for the original number of MSVM. As far as the search space is concerned, Ymax denotes the upper limit and Ymin its lower limit. A random integer between zero and one is represented by the symbol rand. The following is the definition of the initial matrix of prey in the MSVM:

$$Prey = \begin{bmatrix} Y_{1,1} & Y_{1,2} & \cdots & Y_{1,d} \\ Y_{2,1} & Y_{2,1} & \cdots & Y_{2,d} \\ \vdots & \vdots & \vdots & \vdots \\ Y_{n,1} & Y_{n,2} & \cdots & Y_{n,d} \end{bmatrix}$$

The variables d stands for the problem's dimensionality, n for the number of prey, yi,j for the

value of the ith prey's jth dimension, and Prey for the prey matrix. The algorithm applies an appropriate fitness function to each prey during its recurrent processing to determine its fitness value. The following method of recording the prey's fitness levels is made possible by this:

$$F_{OA} = \begin{bmatrix} f(Y_{1,1}; Y_{1,2}; \cdots; Y_{1,d}) \\ f(Y_{2,1}; Y_{2,1}; \cdots; Y_{2,d}) \\ \vdots \\ f(Y_{n,1}; Y_{n,2}; \cdots; Y_{n,d}) \end{bmatrix}$$

(2) Searching and tracking the prey—iterative search process

When they're out in nature, MSVM can see and follow their prey on their own. The male jackal takes charge when somebody in the group detects the presence of prey, and he leads the female jackal as they chase for the prey. Here is how mathematical modeling might depict this process:

$$Y_1(t) = Y_M(t) - E \cdot |Y_M(t) - rl \cdot Prey(t)|$$

$$Y_2(t) = Y_{FM}(t) - E \cdot |Y_{FM}(t) - rl \cdot Prey(t)|$$

3.2 Collaborative updating mechanism of MSVM-WOA based on Lagrange interpolation

The original MSVM algorithm's initial position update equation (Eq (6)) needs updating so that the GWO approach's hierarchical structure and the IGWO strategy's alpha wolf can be combined. Under conditions of convergence, wherein there is no change from generation to generation, the fundamental MSVM algorithm predicts that male and female jackals should be in the same position within the population.

These three equations, when combined, give us Eq(6), which says that Y1(t), Y2(t), and Y3(t) have to be the same. However, because the gray wolf pack features a hierarchical structure, the influence of the α wolf cannot be considered when using Eq (6) to update locations. This has an immediate impact on the optimisation performance of the MSVM algorithm and the suggested enhancement strategy.

$$Y(t+1) = \frac{Y_1(t) + Y_2(t)}{2}$$

The current population will now have three primary members—male jackals, female jackals, and alpha wolves—after the optimisation strategy for grey wolves is considered. There are three sites where these three components are crucial: Y1(t), Y2(t), and Y3(t). Ensuring that the position update equation converges requires verifying that the well-established three-point iteration formula overlaps. The three individuals involved—the alpha wolf, the female jackal, and the male jackal—must work together and communicate well. Our new population update equation is based on this, which is computed using the Lagrange three-point interpolation formula:

$$Y(t+1) = \frac{(Y(t)-Y_2(t))(Y(t)-Y_3(t))}{(Y_1(t)-Y_2(t))(Y_1(t)-Y_3(t))} \cdot \frac{Y_1(t)}{3} + \frac{(Y(t)-Y_1(t))(Y(t)-Y_3(t))}{(Y_2(t)-Y_1(t))(Y_2(t)-Y_3(t))} \cdot \frac{Y_2(t)}{3} + \frac{(Y(t)-Y_1(t))(Y(t)-Y_2(t))}{(Y_3(t)-Y_1(t))(Y_3(t)-Y_2(t))} \cdot \frac{Y_3(t)}{3}$$

A male jackal (Y1(t)), a female jackal (Y2(t)), and an alpha wolf (Y3(t)) are the equivalent places in the tth cycle. The population's location as of the last iteration is denoted by Y(t), which is obtained at the start of this iteration. With Y(t+1) as its value, we can see where the population stands following the iterative update.

Maintaining uniform distribution weights for male and female jackals as well as alpha wolves is the primary function of the constant 3 that controls the iterative equation's convergence.

4. Results and Discussion

4.1 MSVM-GWO algorithm execution

There was cooperation between Figure 1 and the procedure for improving the algorithm. Method 1 displays the suggested MSVM-GWO method's pseudocode.

Algorithm 1: The MSVM-GWO Algorithm

Input: Maximum number of iterations T, variable dimension d, and population N.

Initializing the population

while (t<T)

if $f \leq f_M$

$f_M = f$

elseif $f_M < f \leq f_{EM}$

$f_{EM} = f$

elseif $f_M < f_{FM} < f \leq f_\alpha$

$f_\alpha = f$

end if

Calculating the random number rl associated

with the levy function

for (Iterating through each individual in the population)

Computing the energy function E for prey avoiding the jackal wolves and (8)

if |E|<1 (EXPLOITATION)

$Y1(t)=YM(t)-E\cdot|YM(t)-rl\cdot Prey(t)|$

$Y2(t)=YFM(t)-E\cdot|YFM(t)-rl\cdot Prey(t)|$

$Y3(t)=Y\alpha(t)-E\cdot|Y\alpha(t)-rl\cdot Prey(t)|$

else (EXPLORATION)

$Y1(t)=YM(t)-E\cdot|rl\cdot YM(t)-Prey(t)|$

$Y2(t)=YFM(t)-E\cdot|rl\cdot YFM(t)-Prey(t$

$Y3(t)=Y\alpha(t)-E\cdot|rl\cdot Y\alpha(t)-Prey(t)|$

end if

Updating the population positions

end for

Boundary handling

t = t+1

end while

Output: $Y_1(t)$ and f_M

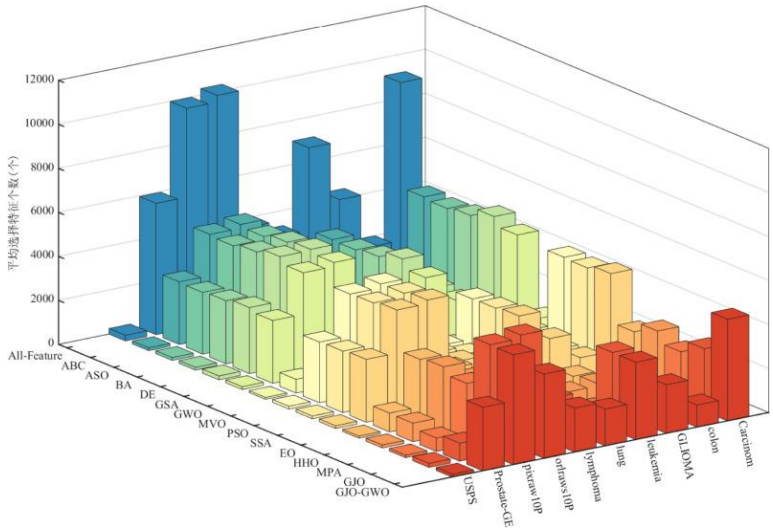

Fig 1. Average number of selected features.

The average classification accuracy and number of characteristics employed by each algorithm are displayed in Figure 1 and Figure 2, respectively.

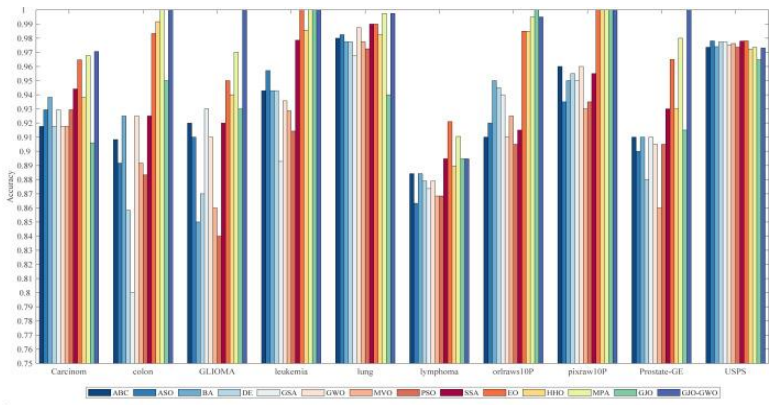

Fig 2. Various algorithms' average categorization accuracy shown as a bar chart.

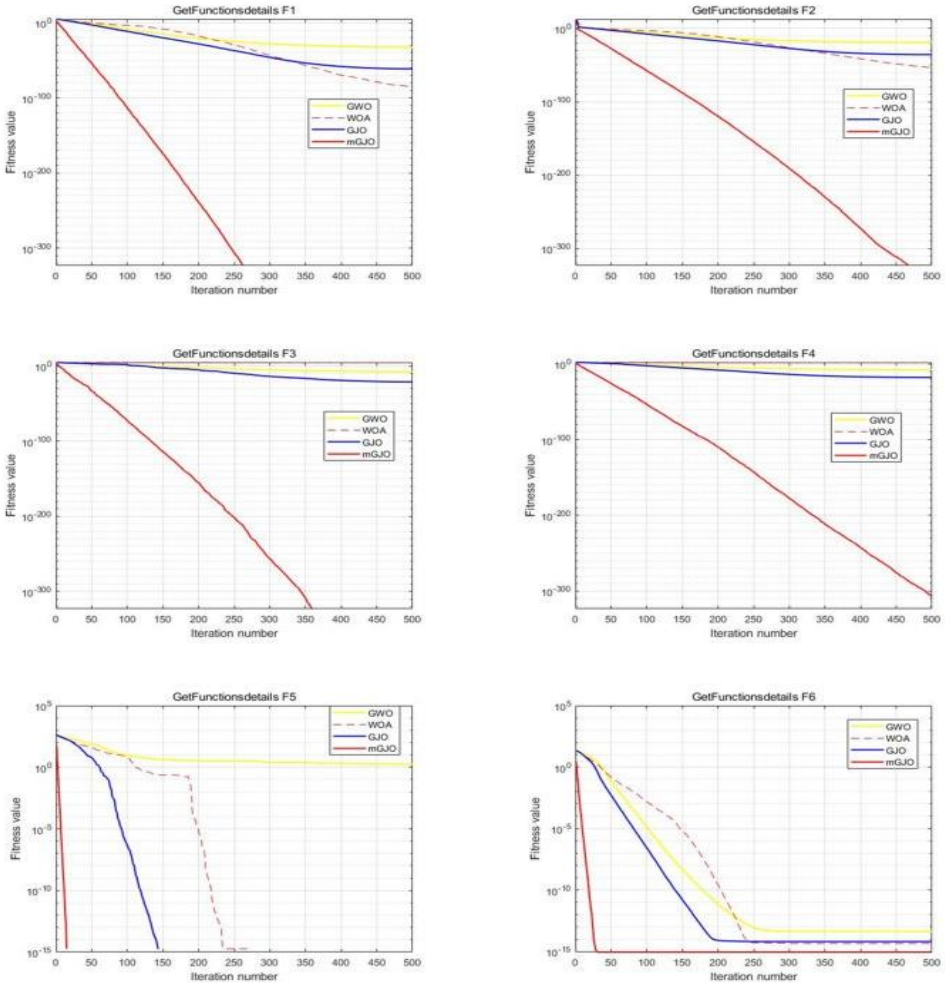

Figure 3.Test function experiment results.

Figure 3 displays the outcomes of thirty iterations of each of the four algorithms on its own test function.

5. Difference between forgery image and real image:

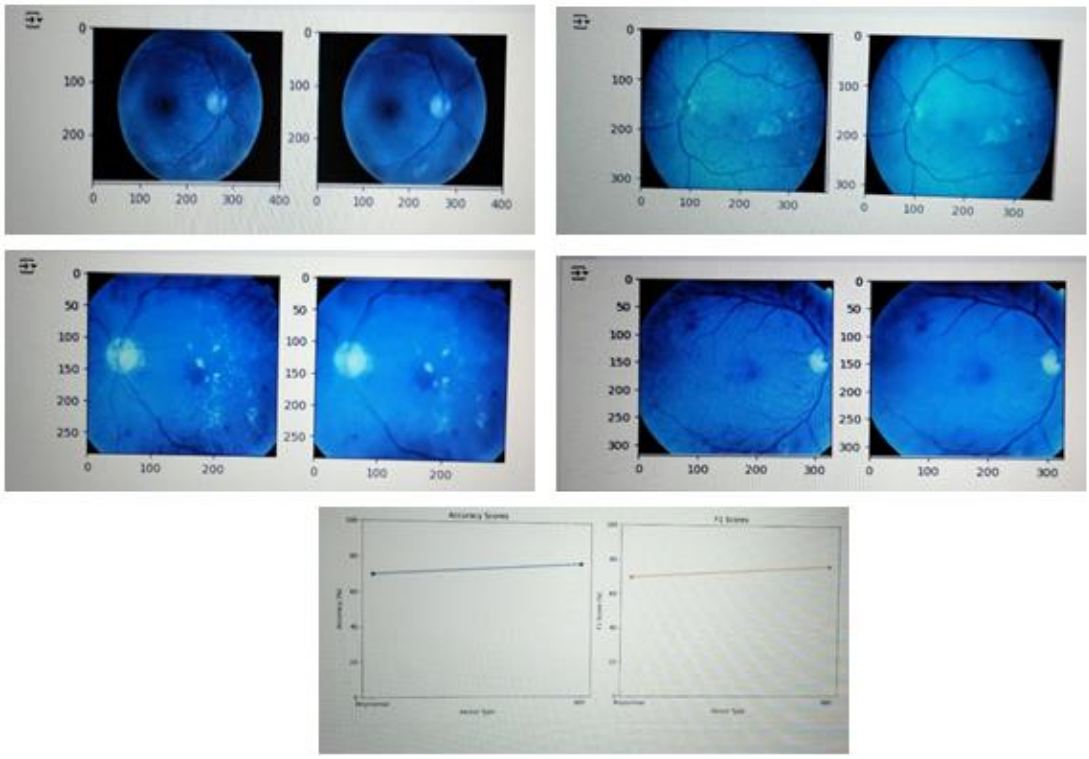

Fig 4: Graph of rbf and polynomial kernel (F1score and accuracy)

6. Conclusion and Future Directions

Resolving feature selection difficulties has been the primary focus of our research aimed at improving the optimisation performance of the MSVM method. Here are the results of the numerical experiments and mechanistic analysis:

1. A hybrid optimisation algorithm called MSVM-GWO, which stands for MSVM and Grey Wolf, has been proposed. A number of tactics are combined in it. The suggested MSVM-GWO outperformed nine individual metaheuristic algorithms in terms of convergence and stability on eight benchmark datasets. Two primary areas are where these benefits are most noticeable: a) Adding the Gray Wolf Algorithm broadens the range of possible solutions. a) The convergence performance of the algorithm is improved by using the position update strategy based on Lagrange interpolation.

2. For classification problems, we have introduced a feature selection method based on MSVM-GWO. On ten high-dimensional datasets, the proposed feature selection method beats

thirteen state-of-the-art methods in terms of accuracy, convergence speed, and runtime.

Two primary areas are where these benefits are most noticeable: a) The programming architecture of the Gray Wolf Algorithm makes it more efficient at runtime and increases solution diversity. b) The algorithm's convergence speed is effectively increased by the position update technique based on Lagrange interpolation. By deftly combining these tactics, the algorithm may dynamically change the ratio of exploration to exploitation at various points in the search process. This study presents a feature selection approach based on MSVM-GWO, and while it does a better job overall, there is room for improvement. As an example, when choosing features for classification, it frequently chooses a huge number of features, which could be problematic for future deep learning or machine learning jobs.

References

1. Savaliya, R. H. Jhaveri, Q. Xin, S. Alqithami, S. Ramani, T. A. Ahanger, Securing industrial communication with software-defined networking, Math. Biosci. Eng., 18 (2021), 8298–8314. https://doi.org/10.3934/mbe.2021411.
2. Balaraju, J., and PVRD Prasada Rao. "A Novel Node Management in Hadoop Cluster by Using DNA." International Journal of Information Technology Project Management (IJITPM) 12.4 (2021): 38-46.
3. N. T. Hoang, H. N. Nguyen, H. A. Tran, S. Souihi, A novel adaptive east–west interface for a heterogeneous and distributed SDN network, Electronics, 11 (2022), 975. https://doi.org/10.3390/electronics11070975.
4. Balaraju, J., et al. "Dynamic password to enforce secure authentication using DNA." Int. J. Intell. Syst. Appl. Eng 12.1 (2023): 55-61.
5. P. Wu, Y. Shang, S. Bai, L. Cheng, H. Tang, A lightweight path consistency verification based on INT in SDN, Math. Biosci. Eng., 20 (2023), 19468–19484. https://doi.org/10.3934/mbe.2023862
6. Yazdinejadna, R. M. Parizi, A. Dehghantanha, M. S. Khan, A kangaroo-based intrusion detection system on software-defined networks, Comput. Networks, 184 (2021), 107688. https://doi.org/10.1016/j.comnet.2020.107688
7. Balaraju, J., and P. V. R. D. Prasada Rao. "Dynamic Node Identification Management in Hadoop Cluster Using DNA." Smart Computing Techniques and Applications: Proceedings of the Fourth International Conference on Smart Computing and Informatics, Volume 2. Springer Singapore, 2021.
8. S. Badotra, S. Tanwar, S. Bharany, A. U. Rehman, E. T. Eldin, N. A. Ghamry, et al., A DDoS vulnerability analysis system against distributed SDN controllers in a cloud computing environment, Electronics, 11 (2022), 3120. https://doi.org/10.3390/electronics11193120
9. M. W. Nadeem, H. G. Goh, V. Ponnusamy, Y. Aun, DDoS detection in SDN using machine learning techniques, Comput. Mater. Continua, 71 (2022), 771–789. https://doi.org/10.32604/cmc.2022.021669
10. Balaraju, J., Prasada Rao. PVRD, ―Designing authentication for Hadoop cluster using DNA algorithm‖. Int. J. Recent. Technol. Eng. (IJRTE) ,8(3), 2019. ISSN: 2277-3878. https://doi.org/ 10.35940/ijrte.C5895.0983.
11. J. Wang, Y. Liu, H. Feng, IFACNN: Efficient DDoS attack detection based on improved firefly algorithm to optimize Convolutional Neural Networks, Math. Biosci. Eng., 19 (2022), 1280–1303. https://doi.org/10.3934/mbe.2022059
12. F. Zhang, Z. Gao, K. Niu, Network intrusion detection model based on BiGRU system (in Chinese), Comput. Technol. Dev., 33 (2023), 144–149. https://doi.org/10.3969/j.issn.1673-

629X.2023.01.022
13. Balaraju, J., and P. V. R. D. Prasada Rao. "Innovative secure authentication interface for Hadoop cluster using DNA cryptography: A practical study." Soft Computing and Signal Processing: Proceedings of 2nd ICSCSP 2019 2. Springer Singapore, 2020.
14. J. Liu, Y. Yan, Artificial fish feature selection network intrusion detection system (in Chinese), J. Xidian Univ., 50 (2023), 132–138. https://doi.org/10.19665/j.issn1001-2400.2023.04.013
15. J . Li, K. Cheng, S. Wang, F. Morstatter, R. P. Trevino, J. Tang, et al., Feature selection: A data perspective, ACM Comput. Surv., 50 (2017), 1–45. https://doi.org/10.1145/3136625
16. O. Friha, M. Ferrag, S. Lei, M. Leandros, C. Kim-Kwang, M. Nafaa, FELIDS: Federated learningbased intrusion detection system for agricultural Internet of Things, J. Parallel Distrib. Comput., 165 (2022), 17–31. https://doi.org/10.1016/j.jpdc. 2022.03.003
17. R. A. Elsayed, R. A. Hamada, M. I. Abdalla, S. A. Elsaid, Securing IoT and SDN systems using deep-learning based automatic intrusion detection, Ain Shams Eng. J., 14 (2023), 102211. https://doi.org/10.1016/j.asej.2023.102211
18. N. M. Yungaicela-Naula, C. V. Rosales, J. A. Perez, E. Jacob, C. M. Cagnazzo, Physical assessment of an SDN-based security framework for DDoS attack mitigation: Introducing the SDN-SlowRate-DDoS dataset, IEEE Access, 11 (2023), 46820–46831. https://doi.org/10.1109/ACCESS.2023.3274577
19. G. O. Anyanwu, C. I. Nwakanma, J. M. Lee, D. S. Kim, Optimization of RBF-SVM kernel using grid search algorithm for DDoS attack detection in SDN-based VANET, IEEE Internet Things J., 10 (2022), 8477–8490. https://doi.org/10.1109/JIOT.2022.3199712
20. Zhiqing GUO. Research on Feature SelectionMethod Based on Improved Whale Optimization Algorithm. Master's degree, Liaoning Technical University.2022. 10.27210/d.cnki.glnju. 2022.000421