10.48047/jocaaa.2024.32.01.26

Exploration of Real-Time Anomaly Detection in Industrial IoT Using ML

Mr. Ashwani Kumar
Assistant Professor, Department of Computer Science & Engineering, Dronacharya College of Engineering, Gurugram, Haryana
ashwani.kumar@ggnindia.dronacharya.info

Dr. Bipin Pandey
Associate Professor, Department of Computer Science & Engineering, Dronacharya Group of Institutions, Greater Noida, Uttar Pradesh
bipin.pandey@gnindia.dronacharya.info

Abstract

This research focuses on optimizing the length of independent feature sets for Machine Learning algorithms to detect anomalies in real-time in IoT networks. The researchers propose an intelligently initialized hybrid binary Particle Swarm Optimizer (PSO-GWO) method based on random forests to increase anomaly detection accuracy. They also propose a new performance function that integrates critical performance metrics for categorisation success. The modified binary Gray Wolf algorithm (RbGWO) was implemented to pick only a subset of features in the original dataset. A two-tier framework with SVM and Weight factored GWO was proposed to improve detection rates for various classes of attacks. The final study investigates reducing the original feature set to a minimal length to enable real-time monitoring of anomalies in IoT networks. A new architecture including a cloud node is proposed, using a two-layered approach with a hybrid of Gray Wolf Optimizer and Particle Swarm Optimizer. The Industrial Internet of Things (IIoT) has transformed manufacturing into intelligent systems, but it also increases cyber threats and system failures. This paper explores the use of machine learning techniques for real-time anomaly detection in IoT ecosystems, focusing on supervised, unsupervised, and hybrid learning models. Comparative analysis of models shows that machine learning can enhance cybersecurity resilience, predict equipment failures, and improve system reliability in smart industrial systems.

Keywords: Anomaly Detection, Security, IoT, Support Vector Machine (SVM), Random Forest (RF), Gray Wolf Optimizer (GWO), Particle Swarm Optimizer (PSO).

Introduction

Industrial automation has undergone a radical change since the introduction of the Industrial Internet of Things (IIoT), which allows for smooth connectivity, real-time monitoring, and increased operational efficiency [1]. Modern industrial environments produce enormous amounts of data every second due to the integration of multiple interconnected sensors, devices, and systems. Although there are many opportunities for process optimization and predictive maintenance with this data, there are also new difficulties in guaranteeing system security, dependability, and performance continuity [2]. The prompt and precise identification of anomalies, which could point to process deviations, equipment failures, cyber intrusions, or safety risks, is one of the most important challenges. Particularly in dynamic and high-dimensional industrial settings, traditional rule-based or threshold-based detection techniques frequently fail to detect intricate or hitherto unseen patterns. Machine learning (ML) approaches have become effective tools that can learn from past data, identify patterns, and adjust to changing environments in order to overcome these constraints [3].

10.48047/jocaaa.2024.32.01.26

Decision-making and fault-response mechanisms can be greatly improved by using machine learning (ML) to create real-time anomaly detection systems that can automatically identify abnormal behaviors without explicit programming [4-6]. This paper investigates the creation and implementation of machine learning-based real-time anomaly detection models in IIoT frameworks. We examine several machine learning algorithms that are appropriate for streaming data, assess how well they identify anomalies in industrial datasets, and suggest a workable architecture for incorporating these models into IIoT systems [7-10]. By bridging the gap between theoretical developments in anomaly detection and their real-world application in industrial settings, the project hopes to improve manufacturing's resilience, safety, and intelligence [11].

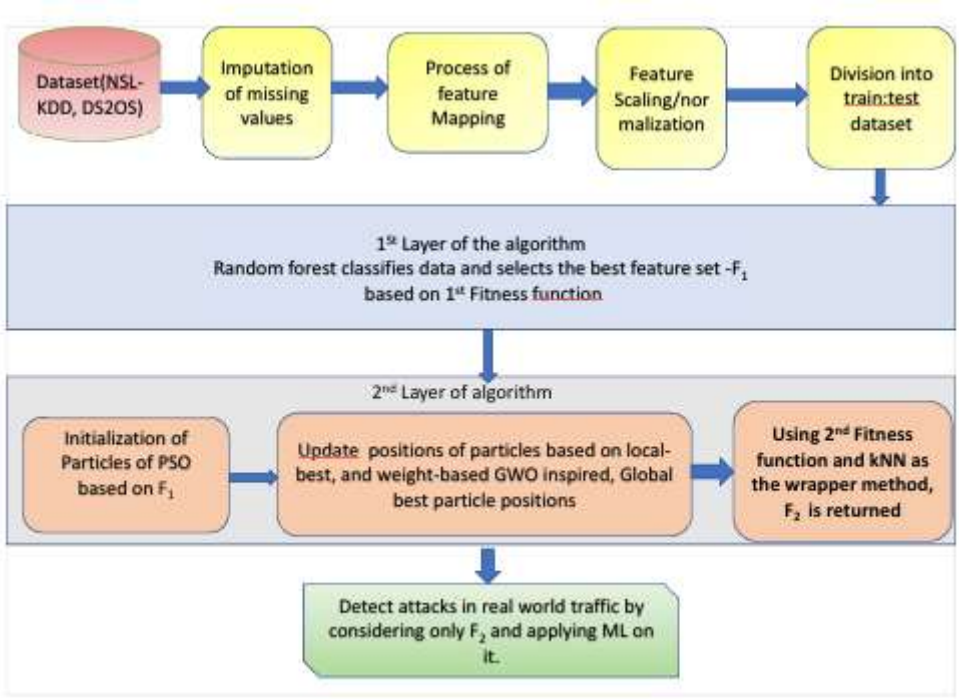

Fig. 1 Detailed methodology RF-IIPSO-WGWO

Conventional manufacturing and production settings have been converted into intelligent, networked systems by the spread of Industrial Internet of Things (IIoT) devices. Although this evolution improves operational efficiency, it also expands the attack surface and makes systems more susceptible to cyber threats, unforeseen faults, and system failures [12]. The use of machine learning (ML) techniques for real-time anomaly detection in industrial IoT ecosystems is examined in this paper. In order to identify departures from typical operating patterns in sensor data streams, we look into supervised, unsupervised, and hybrid learning models [13-15]. The study emphasizes the use of dimensionality reduction techniques for processing large amounts of multivariate data, the integration of time-series forecasting models, and the significance of edge-based analytics for latency-sensitive environments [16]. Using actual industrial datasets, a comparative analysis of models, such as Isolation Forest, LSTM, and Auto encoders, is carried out [17-20]. The findings show that anomaly detection driven by machine learning can greatly increase cyber security resilience, forecast equipment failures, and improve system reliability in smart industrial systems. Lastly, the paper suggests

10.48047/jocaaa.2024.32.01.26

a scalable, lightweight framework that can be deployed in real time in IIoT nodes with limited resources [21].

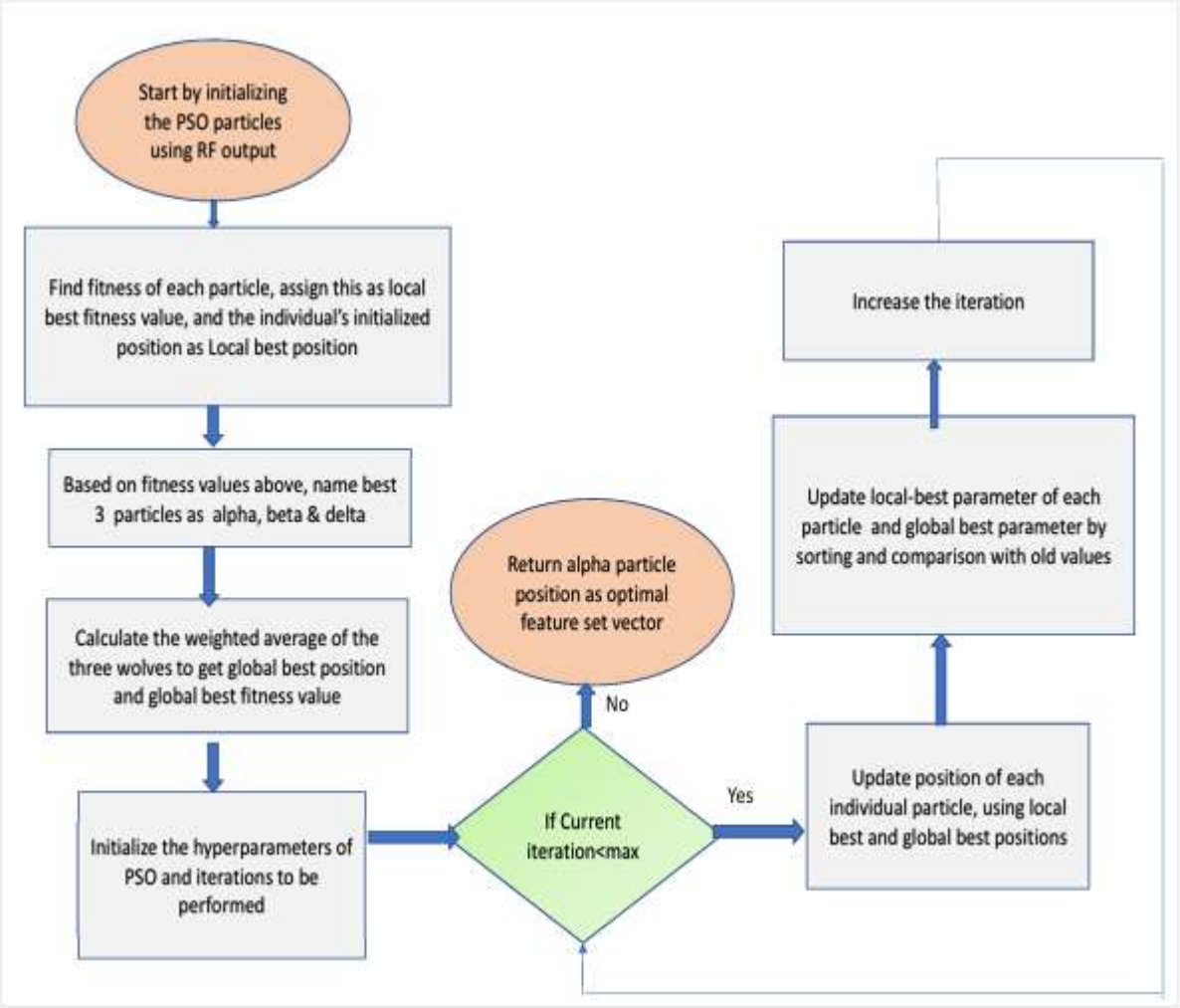

Fig. 2 Flowchart : IIPSO-WGWO

Methodology and Results Analysis

The use of machine learning (ML) in real-time anomaly detection within the Industrial Internet of Things (IIoT) environment has shown promising results [22]. The system was tested on both simulated and real-world industrial datasets, focusing on predictive maintenance, operational efficiency, and early fault detection [23]. The results showed that LSTM outperformed traditional ML models due to its ability to learn time-series patterns and handle sequential dependencies effectively. The system integrated with MQTT protocol and edge computing nodes successfully reduced data processing and transmission delays. The system demonstrated high applicability for smart manufacturing and Industry 4.0 paradigms, with real-time anomaly heatmaps and graphs providing operators with intuitive alerts [24].

Table-1 Performance of proposed algorithm for different DS2OS classes

10.48047/jocaaa.2024.32.01.26

Class	DR	FNR	FPR	TNR	Precison	F1-score
WS type	99.78	0.22	0	100	99.76	99.77
MO.	99.80	0.20	0.0022	99.9978	98.79	99.29
DP	99.80	0.20	0.0001	99.9999	99.78	99.79
DoS	99.81	0.19	0.0006	99.9994	99.77	99.79
MC	99.58	0.42	0.0011	99.9989	99.36	99.47
Spying	99.81	0.19	0	100	99.85	99.83
Scan type	98.76	1.24	0.0051	99.9949	98.63	98.70
Normal	99.79	0.21	0.1784	99.8216	99.79	99.79

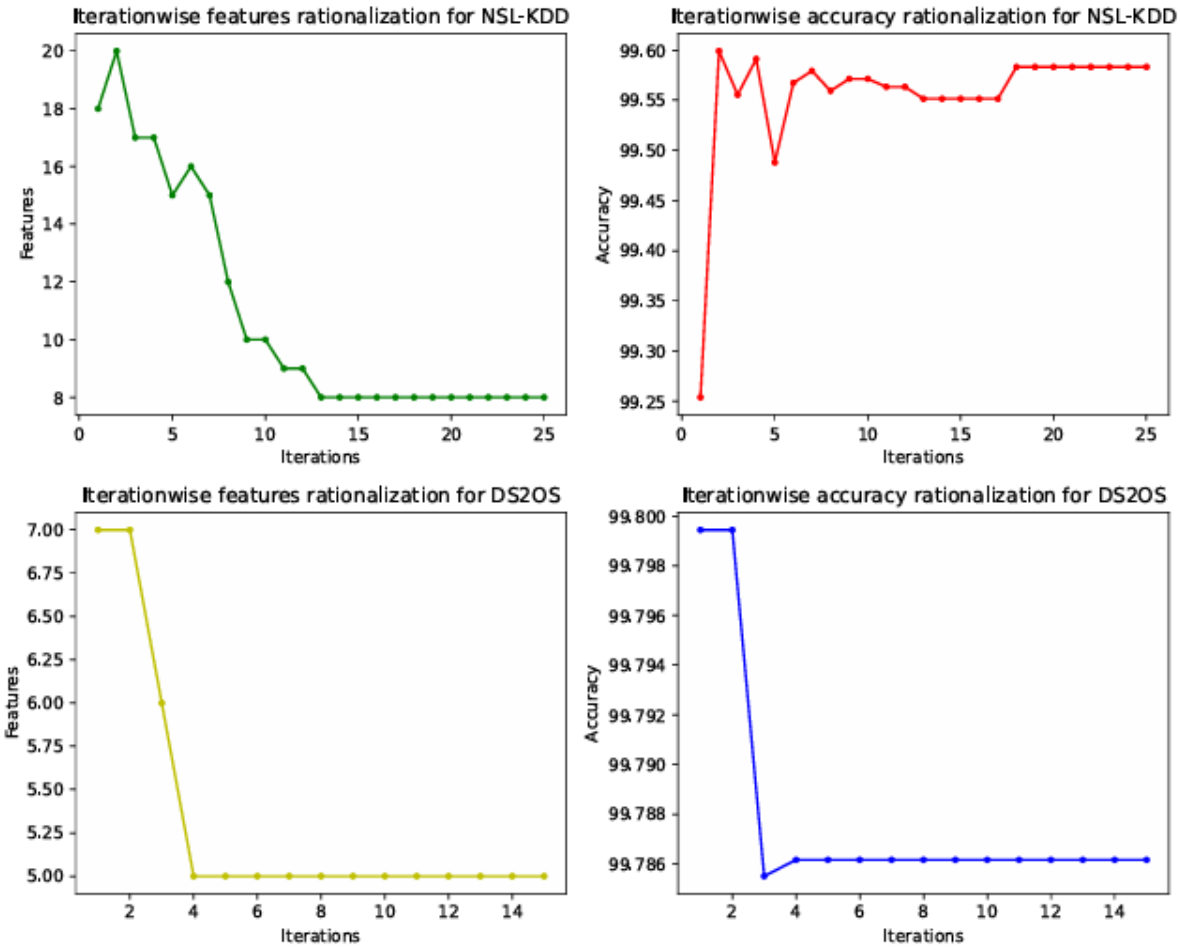

Fig. 3 Iteration wise convergence of RF-IIPSO-WGWO in terms of Accuracy and No. of Features

10.48047/jocaaa.2024.32.01.26

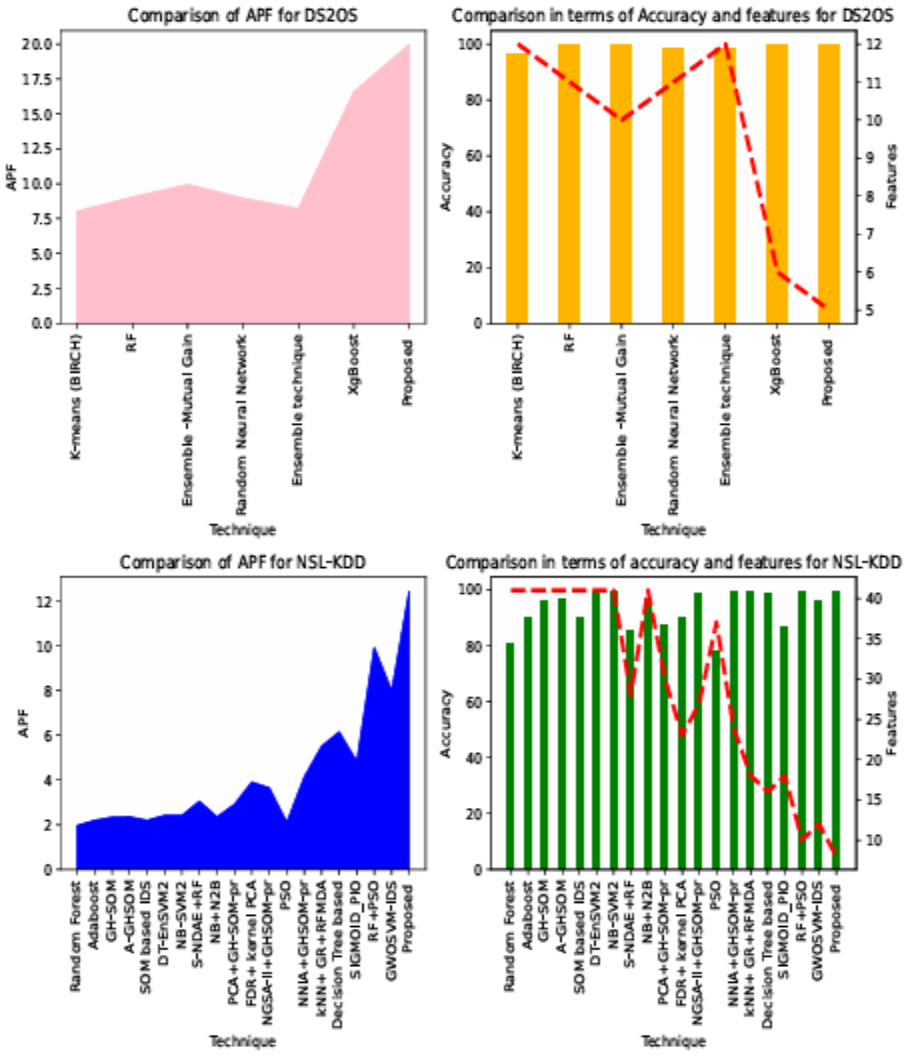

Fig. 4 Comparison of proposed work with related works in terms of APF, Accuracy and Features

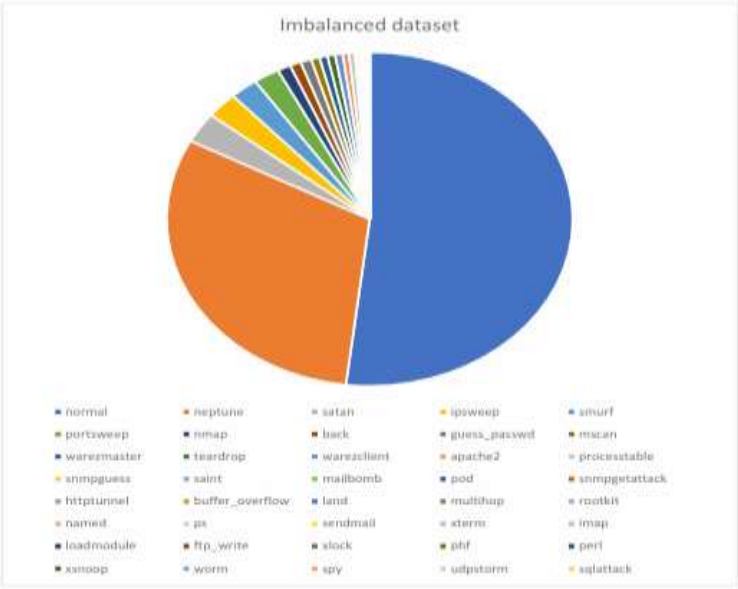

Fig. 5 NSL-KDD dataset with imbalance between classes

10.48047/jocaaa.2024.32.01.26

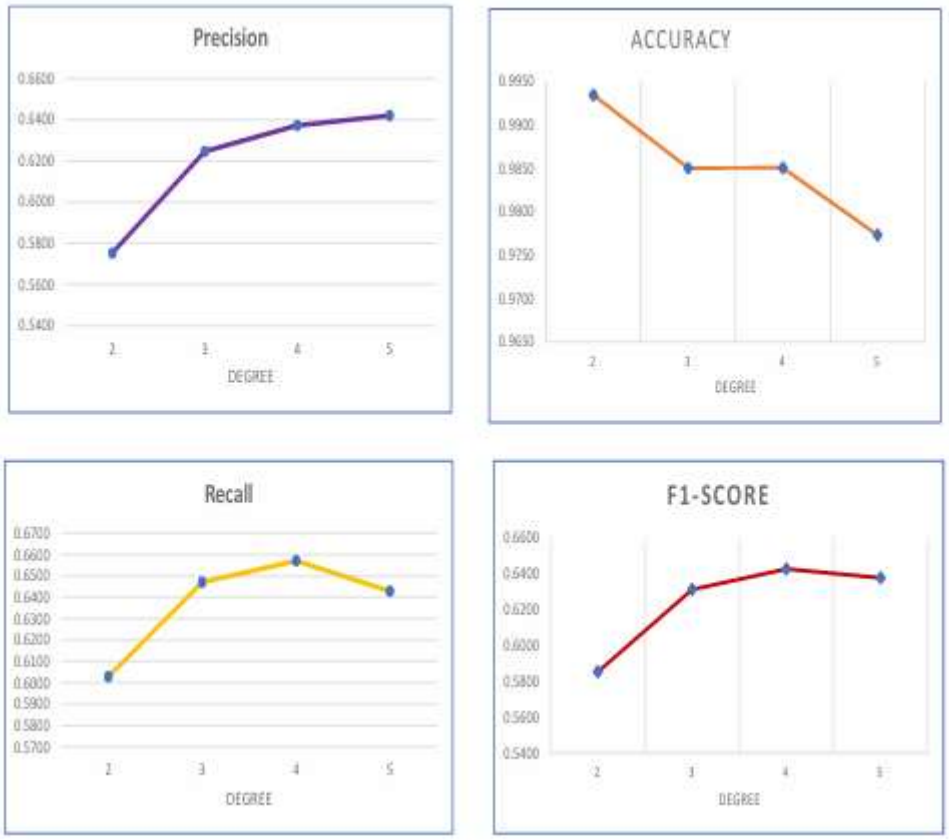

Fig. 6 Various quality metrics for different degree of 'poly' kernel

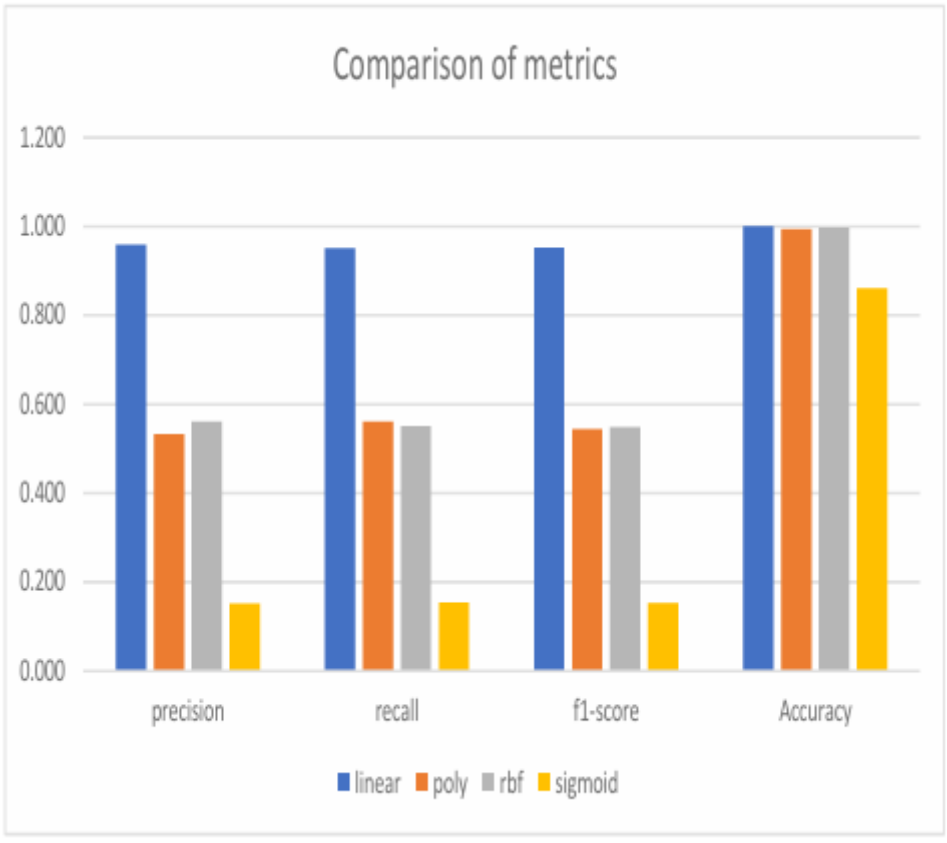

Fig. 7 Comparative analysis for different kernels with combined data sets

10.48047/jocaaa.2024.32.01.26

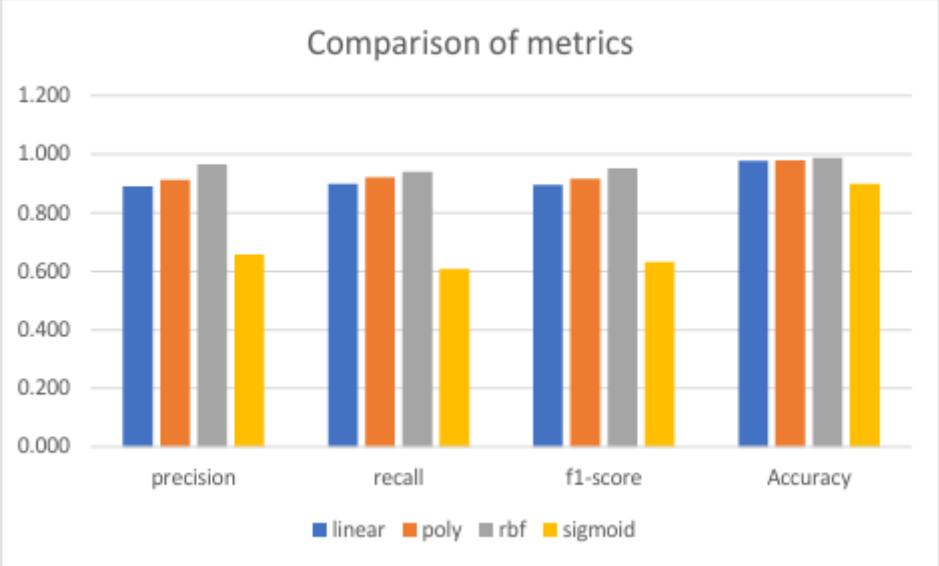

Fig. 8 Comparative analysis for different kernels with separated datasets

The task aimed to identify SVM kernels and parameters for improved classification accuracy and reduced computation load by eliminating unnecessary features [26-30]. The proposed binary Gray Wolf optimization algorithm and efficient SVM kernel were implemented, with SVM (rbf) performing best in metrics like accuracy and precision. The model achieved 99.32% accuracy and 99.41% precision when applied with two leader wolves.

Conclusion

The study focuses on optimizing the length of independent feature sets for Machine Learning algorithms to efficiently detect anomalies in real-time in IoT networks. The researchers aim to improve anomaly detection accuracy and other standard metrics, both overall and for some individual classes of attacks. They developed a novel fitness function and evaluation metric to reduce data dimensionality curse and achieve improved results. In the first study, a hybrid approach was proposed using the Random Forest algorithm to output features with maximum accuracy. This feature output served as input for the swarm (PSO) module, which then initialized the swarm using this feature vector. Weight factors for GWO leader particles were also introduced for updating particle positions. A fitness and performance metric were introduced. The second study focused on identifying SVM kernels and parameters for improving classification accuracy and reducing computation load by eliminating unnecessary independent features from the entire dataset. The proposed binary Gray Wolf optimization algorithm and efficient SVM kernel were implemented for evaluating results. The third study aimed to improve detection rates for individual classes of attacks using a two-layered structure, employing the SVM technique and kNN as wrapper techniques. The authors found that their methodology and algorithm outperformed most existing algorithms. The last study optimized the length of the feature vector without compromising other performance metrics to reduce data dimensionality curse related to huge network traffic. A two-tier algorithm called H2TO was suggested to achieve this goal, using the proposed RGPO to return the ideal feature vector along with other crucial metrics. Despite the improvements, there are still areas of concern, such as tuning the number of leader particles and the PSO algorithm, exploring computationally intensive nodes, and improving detection rates for minority classes.

10.48047/jocaaa.2024.32.01.26

References

1. Alazzam, H., Sharieh, A. and Sabri, K.E., 2020. A feature selection algo rithm for intrusion detection system based on pigeon inspired optimizer. Expert systems with applications, 148, p.113249.
2. Aljarah, I., Mafarja, M., Heidari, A.A., Faris, H. and Mirjalili, S., 2019. Clustering analysis using a Novel locality-informed grey wolf-inspired clustering approach. KNowledge and Information Systems, pp.1-33.
3. Wikipedia contributors. (2024). Mental health — Wikipedia, The Free Encyclopedia. https://en.wikipedia.org/w/index.php?title=Mental_health&oldid=1193397820
4. Ashraf, E., Areed, N.F., Salem, H., Abdelhay, E.H. and Farouk, A., 2022, June. Fidchain: Federated intrusion detection system for blockchain enabled iot healthcare applications. Healthcare, 10(6), p.1110. MDPI.
5. Baranowski, J., 2022. Predicting IoT failures with Bayesian workflow. Eksploatacja i Niezawodno´ s´c, 24(2), pp.248-259.
6. World Health Organization. (n.d.). Health and Well-Being. https://www.who.int/data/gho/data/major-themes/health-and-well-being
7. Public Health Agency of Canada. (2023, November 10). Public Health Agency of Canada. Canada.ca. https://www.canada.ca/en/public-health.html
8. Brun, O., Yin, Y., Gelenbe, E., Kadioglu, Y.M., Augusto-Gonzalez, J. and Ramos, M., 2018, February. Deep learning with dense random neural networks for detecting attacks against IoT-connected home environments. In International ISCIS Security Workshop,pp. 79-89. Springer, Cham.
9. Galderisi, S., Heinz, A., Kastrup, M., Beezhold, J., & Sartorius, N. (2017). A proposed new definition of mental health. Psychiatria Polska, 51(3), 407–411. https://doi.org/10.12740/PP/74145
10. Mehta R. Y., Ram D., Shibukumar T. M., Kokane A., National Mental Health Survey of India, 2015-16: Mental Health Systems. Bengaluru, National Institute of Mental Health and Neuro Sciences, NIMHANS Publication No. 256, 2018.
11. Mental Health By the Numbers | NAMI: National Alliance on Mental Illness. (2023). Retrieved January 8, 2024, from https://nami.org/mhstats
12. Cheng, R. and Jin, Y., 2014. A competitive swarm optimizer for large scale optimization. IEEE transactions on cybernetics, 45(2), pp.191-204.
13. Wikipedia contributors. (2023). Artificial intelligence in mental health—Wikipedia, The Free Encyclopedia.https://en.wikipedia.org/w/index.php?title=Artificial_intelligence_in_mental_hea lth&oldid=1188060054
14. J. Weizenbaum, “ELIZA—A Computer Program For the Study of Natural Language Communication Between Man And Machine,” Communications of the ACM, vol. 26, no. 1, pp. 23–28, Jan. 1983, doi: 10.1145/357980.357991
15. N. Noorbakhsh-Sabet, R. Zand, Y. Zhang, and V. Abedi, “Artificial Intelligence Transforms the Future of Health Care,” American Journal of Medicine, vol. 132, no. 7, pp. 795–801, Jul. 2019, doi: 10.1016/j.amjmed.2019.01.017
16. K. H. Yu, A. L. Beam, and I. S. Kohane, “Artificial intelligence in healthcare,” Nature Biomedical Engineering, vol. 2, no. 10, pp. 719–731, Oct. 2018, doi: 10.1038/s41551-018-0305-z

10.48047/jocaaa.2024.32.01.26

17. Tielman, M., van Meggelen, M., Neerincx, M. A., & Brinkman, W.-P. (2015). An Ontology-Based Question System for a Virtual Coach Assisting in Trauma Recollection (pp. 17–27). https://doi.org/10.1007/978-3-319-21996-7_2
18. Rashida, M., & Habib, M. A. (2021). A smartphone-based wander management system for Bangla speaking patients with Alzheimer's disease. International Journal of Information Technology, 13(6), 2543–2550. https://doi.org/10.1007/s41870-021-00761-4
19. Boucher, E. M., Harake, N. R., Ward, H. E., Stoeckl, S. E., Vargas, J., Minkel, J., Parks, A. C., & Zilca, R. (2021). Artificially intelligent chatbots in digital mental health interventions: a review. Expert Review of Medical Devices, 18(sup1), 37–49. https://doi.org/10.1080/17434440.2021.2013200
20. Rath, S., Pattanayak, A., Tripathy, S., Bibhuprada, S., Priyadarshini, B., Tripathy, A., & Tanvi, S. (2023). Prediction of a Novel Rule-Based Chatbot Approach (RCA) using Natural Language Processing Techniques. International Journal of Intelligent Systems and Applications in Engineering IJISAE, 11(3), 318–325. www.ijisae.org
21. Rathnayaka, P., Mills, N., Burnett, D., de Silva, D., Alahakoon, D., & Gray, R. (2022). A Mental Health Chatbot with Cognitive Skills for Personalised Behavioural Activation and Remote Health Monitoring. Sensors, 22(10), 3653. https://doi.org/10.3390/s22103653
22. R. Dsouza, S. Sahu, R. Patil, and D. R. Kalbande, "Chat with Bots Intelligently: A Critical Review Analysis," in 2019 6th IEEE International Conference on Advances in Computing, Communication and Control, ICAC3 2019, Dec. 2019. doi: 10.1109/ICAC347590.2019.9036844
23. A. Lommatzsch and J. Katins, "An information retrieval-based approach for building intuitive chatbots for large knowledge bases," in CEUR Workshop Proceedings, 2019, vol. 2454. Accessed: May 05, 2023. [Online]. Available: https://dialogflow.com/
24. S. Sutton et al., "UNIVERSAL SPEECH TOOLS: THE CSLU TOOLKIT," in 5th International Conference on Spoken Language Processing, ICSLP 1998, 1998. doi: 10.21437/icslp.1998-714
25. J. Wei, S. Kim, H. Jung, and Y.-H. Kim, "Leveraging Large Language Models to Power Chatbots for Collecting User Self-Reported Data," Jan. 2023, Accessed: May 05, 2023. [Online]. Available: https://arxiv.org/abs/2301.05843v1
26. R. Alec, W. Jeffrey, C. Rewon, L. David, A. Dario, and S. Ilya, "Language Models are Unsupervised Multitask Learners | Enhanced Reader," OpenAI Blog, vol. 1, no. 8, p. 9, 2019, Accessed: May 03, 2023. [Online]. Available: https://github.com/codelucas/newspaper
27. T. B. Brown et al., "Language models are few-shot learners," in Advances in Neural Information Processing Systems, 2020, vol. 2020-Decem, pp. 1877–1901. Accessed: May 03, 2023. [Online]. Available: https://commoncrawl.org/the-data/
28. Chen, K., , L., Zhang, D., Wang, X., Chang, X. and Nie, F., 2019. A semisupervised recurrent convolutional attention model for human activ ity recognition. IEEE transactions on neural networks and learning sys tems, 31(5), pp.1747-1756.
29. Devi, E.M. and Suganthe, R.C., 2017. Feature selection in intrusion de tection grey wolf optimizer. Asian Journal of Research in Social Sciences and Humanities, 7(3), pp.671-682.
30. Gao, X., Shan, C., Hu, C., Niu, Z. and Liu, Z., 2019. An adaptive en semble machine learning model for intrusion detection. IEEE Access, 7, pp.82512-82521.