Hybridization of Machine Learning Techniques for WSN Optimal Cluster Head Selection

Praveenkumar R1*, Kirthika2, Durai Arumugam3 and Dinesh4
1Easwari Engineering College, Ramapuram, Chennai; rpjcspraveen@gmail.com
2Easwari Engineering College; Dept of ECE; jayaramankirthika8@gmail.com
3Easwari Engineering College; Dept of IT; duraiarumugam36@gmail.com
4Easwari Engineering College, Dept of ECE; dinesh.v@ee.srmrmp.edu.in

*Correspondence: Praveenkumar R; rpjcspraveen@gmail.com

ABSTRACT - Wireless sensor networks (WSN) keep developing in recent days concerning the self-covered network, self-healing network, and association of system component circuit selections that enable the implementation process. Wireless sensor network lifetime stabilization is essential to providing a higher quality experience to consumers. The wireless sensor network is associated with classifiers that keep learning the data pattern and further modify the cluster selection to produce dynamic results. The presented system is focused on creating an efficient wireless sensor network in which cluster head selection is dynamically programmed to improve the life span of the devices. The energy utilized by each node is pre-programmed with random assignments. The values are configured by the machine learning techniques to improve the hop death. The models developed using the parameters help project energy consumption. The proposed system considers a support vector machine (SVM), and the Gaussian regression process (GRP) enabled the comparative study of lifespan analysis and support systems to make cluster selection efficient. The proposed model is used to test the current selection of cluster heads using a support rectangle machine and further modify the regression process using the Gaussian regression model. Evaluation of network lifetime and flexibility obtained in cluster selection.

Keywords: Wireless sensor networks, Machine learning, Clustering Scheme, Internet of things, Network protocol.

ARTICLE INFORMATION
Author(s): Praveenkumar R, Kirthika, Durai Arumugam and Dinesh
Received: 27/03/2023; Accepted: 29/05/2023; Published: 19/06/2023; e-ISSN: 2347-470X;
Paper Id: IJEER 2703-06;
Citation: 10.37391/IJEER.110224
Webpage-link: https://ijeer.forexjournal.co.in/archive/volume-11/ijeer-110224.html
Publisher's Note: FOREX Publication stays neutral with regard to jurisdictional claims in Published maps and institutional affiliations.

1. INTRODUCTION
With the increasing usage of wireless sensor networks, the coverage of networks around the globe is around 70%, including underwater sensor networks. Software-defined networks enable the WSN models in a configurable way [1]. In existing frameworks, it is identified that high protocol overhead. The simultaneous transmission over interference of networks over the remote terminals creates various problems in the wireless sensor network. The decentralized routing networks have similar issues to other WSN systems [2]. The results are remarkable, and performance improvement is more by utilizing the adaptive software-defined network topology. The system considers dynamic validation of network data in case of any issues in the communication framework, then the data is offloaded to the cloud platform [3]. Data loss is reduced by saving it into the cluster head in case of any short-duration routing issues occurs. In order to receive the ongoing problems in wireless sensor networks, continuous monitoring and self-repairing modules are required. With the advantages of the WSN, a wide range of services are experienced by consumers [4]. The resources have various environmental constraints, but the utilization of wireless sensor networks in global applications keeps improving. Mobile applications are one of the highly applicable platforms for the wireless sensor network application. Hybrid WSN are engaged in promising solutions to mobile clients.

Figure 1: General architecture of Wireless sensor network

Figure 1 shows the generalized architecture of WSN. Wireless LAN is a prominent factor for grid computing in Environmental data. WSN lifespan needs to be analyzed using various techniques. The proposed model focuses on creating a potential architecture to provide reliable solution to any wireless sensor network environment. In a hybrid network, different devices give different responses which are not easily controlled by the
centralized mechanism. Several ways are available to enhance the lifespan of wireless sensor networks. Some of the techniques commonly utilized in wireless sensor networks are discussed below.

Routing optimization: Energy efficient routing is a technique in which routing protocols delays are minimized to provide energy consumption in each node also improve the life span of network.

Clustering: In the clustering process, the network is divided into clusters with unique cluster heads and relay control to handle the synchronous data. Optimizing the cluster heads reduce the number of transmissions. Because of altering the clustering process, the continuous connectivity of devices is reduced.

Data aggregation: Data aggregation is one of the techniques in which data optimization is applied in middle nodes before the data is transmitted to the destination over the sink nodes. The lifetime of the data transmission improved according to the intermediate nodes and their connectivity stamps. Sleep scheduling is utilized in many sensor nodes. To improve the lifetime of the sensor networks, it should be allowed to go sleep mode significantly to save the energy consumption utilized by the ideal nodes. The nodes adjust the network if it is not utilized in the current connectivity. The network is connected with unique power mechanism that continuously gives power to the node. Centralized network controls associated with the power line to reduce the transmission power. The utilization of energy-harvesting techniques is considered. Various algorithms are implemented to make the energy-efficient wireless sensor network. In some instances, energy harvesting is implemented to collect the environment data and analyze the virtual machine that could help the lifetime analysis of wireless sensor network

- The presented system is focused on creating machine learning enabled wireless sensor network mechanism that adopts the changing environment and analyses the lifespan of network.
- Utilizing self-repairing models the presented approach considers continuous support vector machine (CSVM) algorithm and Gaussian regression process utilized in achieving score towards lifespan saving.

The journal is constructed by explaining the detailed literature background about existing works in Section 2 and followed with system tool selection and problem behind the wireless sensor network is discussed in section 3 provided with the system architecture with details system designs steps in Section 4. The Conclusion and research scope is provided in the end.

2. BACKGROUND STUDY
S. Mahmoudzadeh et al (2019) the author presented a system in which the deep evolutionary technique is employed to make the resilient changes in the environmental conditions affecting the WSN. In order to provide an optimized system that adopts the deploying models and integrity associated with the sensor network, genetic algorithm is utilized. The complex constraints in Environmental changes are considered as one parameter to impact the research in the wireless sensor network. The presented study is helpful in analyzing the changing environment. The limited dataset is the drawback.

N. Saeed et al (2018) the author presented a wireless sensor network in underwater communication in which Multi hop communication network is discussed. Underwater communication consists of various optical cables to make the data transfer. The randomly scaled node connectivity from source to destination employs different environments and changes in underwater network that directly impacts the data efficiency. The probability of connection in WSN diversity the multi-hop functioning between sensors and server. Efficient communication between sensor nodes is stimulated in the proposed approach using a randomly scaled sector graph method.

N. Saeed et al (2019) the proposed framework discuss on efficient battery harvesting affected by the different ambient systems. The author presented a system where sufficient storage of network localization is carried out. The presented system considers underwater optical communication for analysis. Block kernel matrices are utilized in received signal strength-based measurement for making a path approach. The proposed system considers the estimation of shortest path technique to make the optical sensor network flexible towards the goal.

Jawahar et al. (2018) the author presented a system in which linear sensor networks to monitor the underwater processing of pipelines are considered here. The presented approach utilized autonomous underwater vehicles to monitor the pipeline arrangements and collect data from each sensor node connected over the pipelines. Vehicles passing through the pipelines cover information about environmental changes, leakages surface problems and transfers the data immediately to the communication Centre using Wimax or cellular communication. Most of the communication satellites have the required data, which is helpful to make the sink node placed in the sense network in case of any issues available. A considerable amount of energy saving is influenced by the network, and continuous monitoring is helpful in making the collision free network.

George et al. (2022) the author presented a predictive analysis model utilized to evaluate an efficient system to analyze network life time and systematic selection of cluster head. The dynamic parameters for wireless sensor networks are connected with various sensor nodes that collect the environmental data. The dynamic parameters support machine learning models to be applied in most of the wireless sensor networks to adjust the tuning process based on the demand. The dynamic changes occurring in the nodes are routed appropriately through downstream framework. To make the routing process more efficient solving the intermediate problems at that instants are important.

Y. Zhou et al. (2020) the author presented a cross-layer approach provided with joint optimization and link layer network flow with physical layer transmission. The system considered energy optimized underwater sensor network. The
primary goal of the system is to maximize the reliability of the network by improving the lifetime. The energy efficient acoustic network is created with optimization principle that solves the network issues immediately. The optimal network flow is created with each link and network lifetime.

Various existing frameworks are considered here for the purpose of study of wireless sensor network and dynamic routing protocols [11]-[27].

3. SYSTEM DESIGN
A machine learning algorithm plays a significant role in the development of customized designs in wireless sensor networks. The network model is customized towards the need of energy efficient transmission and clustering. The presented system considers each nodes with unique (random) distance, cost assigned to it. The created nodes are deployed into random communication and further the nodes behavior is monitored. The node interruptions are evaluated through machine learning algorithms. The presented system comparatively validates support vector model and Gaussian regression process on detection of abnormality. The malicious nodes are identified and removed from the network to improve the life span.

4. METHODOLOGY
In this work, a machine learning based predictor is proposed for evaluating lifetime and Cluster Head (CH) selection for a Wireless Sensor Network (WSN). The proposed method is used to Predict CHs and optimum number of nodes in a network and optimum number of nodes in a wireless sensor network. Forecast the energy consumption of nodes.

For building model, the dynamic parameters of wireless network is collected from performance calculator. These collected parameter are utilized for building model. The hop depth, advertising, number of motes, backbone, routing, reporting interval and path stability are the predictors and current consumption, data latency and build time are the response variable to establish the model for estimating power and performance of the network. The remaining energy in each node, distance from base station and data transmission rate are predictors and the priority of the cluster head is the response variable to creating model for achieving routing path in WSN. The proposed predictor model for life time estimation is analyzed and compared with support vector machine model. For cluster head selection, the proposed model is investigated with existing state of art method.

Figure 2 shows the system architecture of proposed network evaluation model. The initial phase of the network model contains the wireless sensor network (WSN). The model configures number of 1000 nodes in the initial stage. These nodes are assigned with random cost and distance. The nodes are deployed into communication with each other using cluster head. Each group of nodes have the specific cluster head selection. The node deployment is done with the help of cluster head selection. Specific numbers of nodes are connected below the cluster head. The cluster head is capable of sharing the communication data to each nodes effectively. The presented system is configured with kernel based support vector machine and neural network enhanced Gaussian regression model.

Support Vector Machine Model (SVMM)
The support vector machine technique is a supervised mechanism that can be highly utilized for the classification and regression-based problems of various random data. SVM is a popular machine learning algorithm used for classification and regression tasks. SVM is effective in handling both linear and non-linear relationships between variables. It aims to find an optimal hyperplane that maximally separates different classes or predicts the target variable accurately. SVMs have been widely applied in various domains due to their ability to handle high-dimensional data and their robustness against overfitting. The technique is carried out by identifying the best boundary representing the hyperplane that separates the random data into different classes. These data points, on another side helpful to classify the unexpected issues into the class of vectors that support the analysis process and focus on the point that highly co-ordinates with the training data. The system's ultimate goal is to find new closely correlated data. The expression for the linear support vector machine is given by the equation below.

\[ y_a = w_a x + b \]  

Where,

- \( y_a \) act as the predicted output with class label
- \( x \) is the input feature vector with data points
- \( w_a \) is the Bias weight in the vector
- \( b \) act as the bias term

The proposed goal of the support vector machine is to find the weight vectors associated with making the pattern match for the prediction of a different class. This is implemented by maximizing the boundary of the data that could be separated between the hyperplane.

In case of nonlinear classification, the third parameter called kernel is included in the data map to identify the relativity extraction easier. The equation for kernel-based support that our machine is given below.

\[ y = \text{sum} (\alpha_i y_i K(x_i, x)) + b \]  

The proposed approach considered support our machine as an existing system for making the selection of cluster head to reduce the overhead problems.

Steps involved
Step 1: Load the required libraries for making the SVM analysis
Step 2: Read the dataset, pre-process it by resizing the vector size, extract the features X-variable and Y-variable in the free space
Step 3: formulate the training data 70% and testing data 30%
Step 4: Configure SVM classifier with the training data and testing data.
Step 5: test the SVM classifier with the given data and fit.
Step 6: predict the model and fit the design
Step 7: Formulate the performance evaluation using accuracy calculation.

Gaussian regression model

The Gaussian regression mechanism is a supervised learning technique utilized for regression-based data analysis. GPR is a probabilistic regression method that models the relationship between input variables and output variables based on Gaussian processes. GPR provides a flexible and non-parametric approach to regression, allowing for modelling complex relationships and capturing uncertainty in predictions. GPR is often used when the data distribution is unknown or when the relationship between variables is non-linear and requires a more flexible modelling approach. It works with the probabilistic approach in which the data points are not estimated simply by the uncertainty of the prediction but also by the connectivity of the data points. In simple words, Gaussian regression process is a method to make Gaussian distribution of data and analyse the relationship between input and output to predict core connection. The regression process handled by the equation given below.

\[ p(y|x,X,Y) = N(y|m(x),K(x,x')) \]  

Where:

- \( p(y|x,X,Y) \) is the probability function
- \( y \)-act as output variable
- \( x \)- is the location
- \( (X,Y) \)- training data
- \( N(y|m(x),K(x,x')) \)-multivariate normal distribution
- \( m(x) \)-denotes Mean
- \( K(x,x') \)-Covariance

The probability of observation time and connectivity associated with the training data provides multivariate typical distribution results and is associated with the Cove variance of the system. The parameters associated with the Gaussian regression process are given below. \( m(x) \) is the primary function of given input data mapped with the expected results named \( y \) using the Gaussian regression process to the input \( x \). The covariance function \( K \) of the \( x \) ‘is best suited for the training data to estimate the probability of uncertainty between the training data and the testing data. The Gaussian distribution model helps make the relationship between the input and output data and formulate the central covariance present.

Probability of uncertainty present between the training data and the testing data is considered. The Gaussian distribution model supports the relationship between the input and output data and formulate its main covariance.

SVM and GPR are suitable when the relationships between variables are non-linear or when there is no prior knowledge about the underlying data distribution. These algorithms can capture intricate patterns and provide accurate predictions even in complex scenarios. SVM and GPR may have been used to evaluate the performance of the proposed system or method in the paper. By using these algorithms, the authors can assess the effectiveness and accuracy of their approach in addressing the system impairments or solving the research problem. SVM and GPR are well-established machine learning techniques with known performance characteristics. By using these algorithms, the authors can compare their results with existing studies or establish benchmarks for future research in the field.

The machine learning algorithm discussed above, such as a support vector machine algorithm and Gaussian regression process, is comprehensively studied here. These two methods are applied to the creator sensor network using random nodes, and the source-to-destination arrival time is calculated. During the anomaly situations, the node drops the energy consumption. The proposed system focused on remodelling cluster notes based on the associated energy consumption. For the simulation purpose, the energy data is being degraded by the software programs and further involvement of support in terms of calculating the lifespan ability to handle the energy degradation situation and continue the network communication without affecting the data.

5. RESULTS AND DISCUSSION

![Figure 3: Configuration of network](image)

Figure 3 Shows the configuration of network with 30 motes and 3 hops. The number of hops connected together and deployed in the common network.

![Figure 4: Current consumption of HOPS](image)
Figure 4 Shows the current consumption of different HOPS connected in the given network. The above figure shows the relationship between the Hop depths vs. Current consumption. As the depth of the system gets increased the consumption also get increased.

Figure 5 Shows the adjusted bias points towards the connected network. The figure shows the dynamic adjustments happening in the whole model, where the proposed approach focused on 95% confidence level on bonding.

Figure 6: Shows the functional evaluation with respect to minimum objective

The relationship between the functional evaluations and objective score is done here, as the objective score increased for the observed readings, the proposed model is dynamically updated.

Figure 7 shows the SVM regression performance depicting that true response rate with respect to adjusted bias values of SVM. The conventional responses are correlated with the predicted results.

Figure 8: Objective function model

Figure 8 shows the objective function model of proposed SVM algorithm. The estimated values are correlated with the observed values.

Figure 9: Current consumption

Figure 9 shows the current consumptions of various HOPS. The representation shows the performance of different hops with respect to the current consumption.

Figure 7: SVM regression

Figure 10: Performance measure – GRP
Figure 10 shows the performance measure of Gaussian regression process method. Comparing the SVM score, the GRP achieved with maximum correlation with the expected response and predicted response.

Figure 11: Objective function of GRP

Figure 11 shows the objective function of GRP model obtained after the multiple HOPS.

Figure 12: SVM performance graph

Figure 12 shows the SVM performance graph in which true response with respect to predicted response is shown. It shows the observed objective function value and the estimated objective function value for the 30 function evaluations. It also gives the total time and the evaluation time of the performance.

Figure 13: GPR performance graph

Figure 13 shows the GPR performance graph in which true response with respect to predicted response is shown. It shows the observed objective function value and the estimated objective function value for the 30 function evaluations. It also gives the total time and the evaluation time of the performance.

Figure 14: Number of Packets vs. Rounds
Figure 14 shows number of packets vs. rounds obtained during the dynamic protocol. The packet size increases with respect to the number of rounds.

Figure 15: Number of Dead nodes vs. Rounds

Figure 15 shows number of dead nodes vs. rounds obtained during the dynamic protocol. As the number of rounds get increases, then dead nodes may obtain until all the nodes are covered.

Figure 16: Sum of energy of Nodes vs. Rounds

Figure 16 shows sum of energy of nodes vs. rounds obtained during the dynamic protocol. The prime challenges faced by the proposed model are handling the dynamic changes synchronously during the protocol operations. The allocation of dynamic values needs to be properly maintained to secure the data. In case of frequent data loss occurs, the system need to download the data into the cloud.

In this study, we conducted a comparative analysis to assess the impact of system impairments on wireless sensor networks (WSNs) before and after implementing our proposed machine learning-enabled system. Before implementing our system, WSNs often faced several system impairments that affected their overall performance and lifespan. These impairments included inefficient cluster head selection, high energy consumption, limited network lifetime, and decreased reliability. These issues were primarily caused by static programming and lack of adaptability to changing environments.

After implementing our system, which utilized machine learning algorithms for cluster head selection and lifespan analysis, we observed significant improvements in mitigating system impairments. The dynamic programming approach enabled more efficient cluster head selection, resulting in improved energy utilization and extended network lifetime. The comparative analysis revealed that the implementation of our system led to a notable reduction in energy consumption. By leveraging machine learning techniques, the system optimized energy assignments, thereby minimizing wasteful energy expenditure and addressing the issue of hop death. This improvement not only extended the lifespan of individual sensor nodes but also enhanced the overall network reliability. Moreover, the system's ability to adapt to changing environments and analyze network lifespan contributed to reducing system impairments. By employing support vector machine (SVM) and Gaussian regression process (GRP) models, we achieved more accurate predictions for cluster head selection. This dynamic approach allowed the system to adjust its operation based on real-time data, improving overall efficiency and performance.

6. CONCLUSION

Wireless sensor networks are almost applied in every field towards the development of 5G communications and effective internet of things environment. Wireless sensor networks need to be improved in near future to handle all the issues coming over the network in automated way without any manual interruptions. The presented system is focused on taking and important problem called lifetime benefits of sensor notes through cluster head selection technique. Machine learning algorithm for employed over here such as support vector machine as an existing system and Gaussian regression process (GRP) proposed system. By applying the algorithms to the cluster head selection mechanism during the face of energy degradation process the presented system perform civil during the critical time and handle the lifespan problem of wireless sensor network. The propagation time for dynamically controlling the 100 nodes took 23 seconds as per the proposed SVM-GRP method. The correlation is 98% in GRP method where SVM achieved 88% on correlation with predicted and observed values. For the system need to be improved by applying multiple machine learning algorithms hybrid together
to handle the situation in the more efficient way and employee
the error free communication model.

REFERENCES

credit value based AODV routing algorithms for selfish node behavior
detection. Cluster Computing. doi:10.1007/s10586-018-1851-

(2019). Selfish node detection based on hierarchical game theory in IoT
EURASIP Journal on Wireless Communications and Networking,

"Mitigating packet dropping problem in mobile ad hoc networks:
proposals and challenges." IEEE communications surveys & tutorials
13, no. 4 (2011): 658-672. DOI=https://doi.org/10.1109/SURV.2011.072210.00026

"Mitigating routing misbehavior in mobile ad hoc networks." In Proceedings of the
6th annual international conference on Mobile computing and networking.

"UUV’s hierarchical DE-based motion planning in a semi
dynamic underwater wireless sensor network." IEEE transactions on cybernets

"Outlier detection and optimal anchor placement for 3-D underwater optical
wireless sensor network localization." IEEE Transactions on Communications

Performance analysis of connectivity and localization in multi-hop
underwater optical wireless sensor networks. IEEE Transactions on
Mobile Computing, 18(11), 2604-2615.

vehicles in wireless sensor networks for underwater pipeline monitoring."

Regression Models for Evaluation of Network Lifetime and Cluster-Head

communication based on cluster states. Modern Physics Letters A, 35(21),
2050178.

multiple mobile sinks for WSNs." Computers, Materials & Continua 62.2

[12] C. Tankard, "Advanced persistent threats and how to monitor and deter

Monitoring Based IoT Security Models & Deployments From a Pragmatic
Perspective," in IEEE Access, vol. 11, pp. 2621-2651, 2023, doi:
10.1109/ACCESS.2023.3234632.

Based Relay Selection in Cooperative NOMA,” in IEEE Access, vol. 11,
pp. 1578-1577, 2023, doi: 10.1109/ACCESS.2022.3233443

Monitoring of Aquaculture Ponds With Wireless Sensor Networks," in
IEEE Sensors Journal, vol. 23, no. 1, pp. 6-20, 1 Jan.1, 2023, doi:
10.1109/JSEN.2022.3222510.

Edge Intelligence,” in IEEE Transactions on Mobile Computing, doi:
10.1109/TMC.2023.3237374

Transactions on Instrumentation and Measurement, vol. 72, pp. 1-11,
2023, Art no. 5501011, doi: 10.1109/TIM.2022.3224513.

[18] P. S. Bouzinis, N. A. Mitsou, P. D. Diamantoulakis, D. Tyrovolas and G.
K. Karagiannidis, “Intelligent Over-the-Air Computing Environment,” in
2023, doi: 10.1109/LWC.2022.3219250.

Network Image Compression via Distributed Dictionary Learning," in

Analysis of Hybrid Optimization Algorithm for Virtual Head Selection
in Wireless Sensor Networks", Wireless Personal Communications,

topology wireless sensor network for the internet of things’, International

Coding Protocols in WSNs," 2022 IEEE/ACCS 19th International
Conference on Computer Systems and Applications (AICKSA), Abu
Dhabi, United Arab Emirates, 2022, pp. 1-5, doi:
10.1109/ARCS56895.2022.10017702.

[23] A. Budholiya and A. B. Manwar, "Machine Learning based Analysis of
6th International Conference on Electronics, Communication and
Aerospace Technology, Coimbatore, India, 2022, pp. 181-184, doi:
10.1109/ICECT5336.2022.10009170.

"Intrusion Detection System in IoT Network by using Metaheuristic
Algorithm with Machine Learning Dimensional Reduction Technique," 2022
3rd International Conference on Computing, Analytics and
Networks (ICAN), Rajpura, Punjab, India, 2022, pp. 1-6, doi:
10.1109/ICAN56228.2022.10007341.

IoT Applications," 2022 International Conference on Electrical,
Computer, Communications and Mechatronics Engineering (ICECCME),
Maldives, 2022, pp. 1-6, doi:
10.1109/ICECCME55909.2022.9988338.

CSI-Based Human Action Perception,” in IEEE Transactions on Human-

[27] Chandan, S. Kumar and S. Sinha, “Machine Learning based Robust
Techniques to Detect DDoS Attacks in WSN,” 2022 International
Conference on Intelligent Innovations in Engineering and Technology
(ICIEET), Coimbatore, India, 2022, pp. 294-300, doi:
10.1109/ICIEET55458.2022.9967543.

In-Sensors: Computation-enabled Intelligent Sensors for Next Generation
of IoT," 2022 IEEE Sensors, Dallas, TX, USA, 2022, pp. 01-04, doi:
10.1109/SENSORS52175.2022.9967240.

© 2023 by the Praveenkumar R, Kirithka, Durai Arumugam and Dinesh. Submitted for possible open access publication under the terms and conditions of the Creative Commons Attribution (CC BY) license (http://creativecommons.org/licenses/by/4.0/).