Machine Learning-Based Strategy For Efficient Node Localization In Wireless Sensor Networks

Pratheesh R[1], Pranav jeyan V[2], Srikanth K[3], T. Gobinath[4]
[1, 2, 3, 4]Dept of Computer Science and Engineering
[1, 2, 3, 4] Chettinad College of Engineering and Technology, Puliyur, Karur, Tamil Nadu, India

Abstract- *Node localization is a fundamental challenge in Wireless Sensor Networks (WSNs) as precise location estimation is essential for various applications. Traditional localization techniques, including bio-inspired and mathematical models, often struggle with high computational complexity and lim- ited adaptability to diverse environments. Recent advance- ments in Machine Learning (ML) offer promising solutions by leveraging data-driven approaches to optimize localiza- tion accuracy. This survey explores existing localization methods in WSNs, categorizing them into range-based and range-free techniques. Furthermore, it examines the applica- tion of ML models such as Support Vector Regression (SVR), Random Forest Regression (RFR), and CatBoost Regression (CAT) in reducing Average Localization Error (ALE). Addi- tionally, optimization strategies, including the Giant Trevally Optimizer (GTO), are evaluated for their role in enhanc- ing prediction accuracy and reducing computational time. A comparative analysis of conventional and ML-driven lo- calization methods is conducted to highlight their strengths, limitations, and potential improvements. Finally, this paper discusses emerging trends, challenges, and future research directions in ML-based localization for WSNs.*

I. INTRODUCTION

Wireless Sensor Networks (WSNs) are composed of nu- merous small, low-cost sensor nodes that monitor environ- mental conditions and communicate data wirelessly. These networks are widely used in applications such as environ- mental monitoring, healthcare, smart cities, and precision agriculture. One of the critical challenges in WSNs is node localization, which involves determining the positions of un- known nodes using anchor nodes with known coordinates. Accurate localization is essential for improving network ef- ficiency, reducing energy consumption, and ensuring re- liable data collection. Traditional localization techniques are broadly categorized into range-based and range-free approaches. Range-based methods estimate distances or an- gles using techniques like Received Signal Strength Indica- tion (RSSI), Time of Arrival (ToA), and Angle of Arrival (AoA). In contrast, range-free methods rely on connectiv- ity information and do not require direct distance measure- ments. Despite their effectiveness, these methods often suf- fer from high localization errors due to environmental inter- ference, high computational complexity, and limited adapt- ability to dynamic network conditions. Recent advancements in Machine Learning (ML) have introduced new possibilities for enhancing localization accuracy. ML models, including Support Vector Regression (SVR), Random Forest Regres- sion (RFR), and CatBoost Regression (CAT), have been em- ployed to predict localization errors and optimize network parameters. Additionally, optimization algorithms such as the Giant Trevally Optimizer (GTO) have been integrated with ML models to enhance performance, reducing com- putational costs while maintaining high accuracy. This pa- per provides a comprehensive survey of existing localization techniques in WSNs, with a particular focus on ML-based approaches. We analyze the strengths and limitations of vari- ous methods, compare their performance in terms of Average Localization Error (ALE), and discuss emerging trends in the field. The goal of this survey is to offer insights into the ef- fectiveness of ML-driven localization methods and highlight future research directions in this evolving domain.

II. ADOPTION OF MACHINE LEARNING IN CONTEXT TO WSNS

The integration of Machine Learning (ML) in Wire- less Sensor Networks (WSNs) has significantly enhanced node localization accuracy, network efficiency, and decision-making capabilities. Traditional localization techniques of- ten suffer from high computational complexity and environ- mental interference, making them less effective in real-time applications.

ML-based methods provide an alternative by learning patterns from sensor data, making localization more adap- tive, robust, and accurate. ML approaches in WSN local- ization can be broadly categorized into supervised learning, unsupervised learning, and hybrid optimization techniques.

III. SUPERVISED LEARNING

Supervised learning algorithms in WSNs require labeled training data, where models learn to predict node locations based on input features such as RSSI values, node density, and transmission range.

3.1 Support Vector Machines (SVMs)

Support Vector Regression (SVR) has emerged as a powerful machine learning technique for node localization in Wireless Sensor Networks (WSNs), offering robust performance in predicting **Average Localization Error (ALE)**. Unlike traditional regression models, **SVR is based on the Structural Risk Minimization (SRM) principle**, which enhances generalization and minimizes overfitting.

It operates by mapping input features—such as **node density, transmission range, and anchor ratio**—into a high- dimensional space where a **linear regression model** is applied. The model utilizes an **-insensitive loss func- tion**, ensuring that only significant deviations from the true values are penalized, thereby improving accuracy.

In the given paper, **SVR is compared with other ML models** such as **Random Forest Regression (RFR) and CatBoost Regression (CAT)**, showing **competitive per- formance** in localization tasks. However, while SVR ef- fectively **reduces localization errors**, its performance is highly dependent on **parameter tuning**, such as:

Kernel function Regularization parameter (C) Epsilon ()

To further enhance **SVR's accuracy and efficiency**, the study integrates **Giant Trevally Optimizer (GTO)**, which fine-tunes the model's hyperparameters, leading to **reduced Root Mean Square Error (RMSE)** and **im- proved correlation coefficient (R)**.

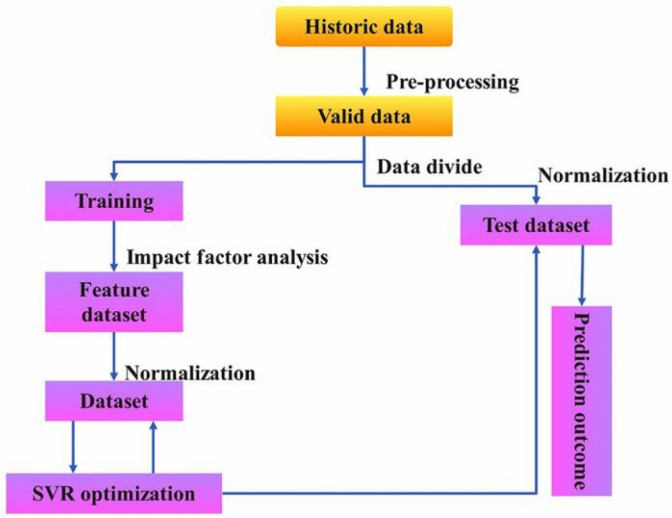
Despite its **computational complexity**, **SVR re- mains a valuable approach for WSN localization** due to its ability to **handle nonlinear relationships** and **improve localization precision** in dynamic environments.

3.2 Random Forest Regression (RFR)

Random Forest Regression (RFR) is an ensemble learn- ing technique widely used for node localization in Wireless Sensor Networks (WSNs) due to its ability to handle nonlin- ear relationships and high-dimensional data. RFR operates by constructing multiple decision trees during training and averaging their predictions to enhance accuracy and robust- ness. This method effectively mitigates overfitting by lever- aging random feature selection and bootstrapped datasets, making it more resilient to noise and missing data compared to single decision-tree models. In the given paper, RFR is applied to predict Average Localization Error (ALE), utilizing key network parameters such as node density, transmission range, and anchor ratio. The results show that RFR achieves high localization accuracy, but its performance is further improved when combined with the Giant Trevally Optimizer (GTO), leading to better parameter tuning and reduced Root Mean Square Error (RMSE). Compared to Support Vector Regression (SVR) and CatBoost Regression (CAT), RFR demonstrates strong generalization ability, making it particularly useful for dynamic WSN environments. However, its computational complexity increases with the number of trees, which may impact real-time applications. Despite this, RFR remains a highly effective approach for improving localization precision and optimizing WSN configurations.

3.3 CatBoost Regression (CAT)

CatBoost Regression (CAT) is an advanced gradient boosting algorithm that has gained popularity for node localization in Wireless Sensor Networks (WSNs) due to its ability to handle categorical features efficiently while maintaining high accuracy and computational efficiency. Unlike traditional boosting methods, CatBoost employs ordered boosting and random permutations, which help prevent overfitting and prediction bias. In the given paper, CatBoost is uti- lized to predict Average Localization Error (ALE) by learn- ing from key network parameters such as node density, transmission range, and anchor ratio. Compared to Support Vector Regression (SVR) and Random Forest Regression (RFR), CatBoost demonstrates faster convergence and better accuracy, especially when dealing with complex datasets. Additionally, the integration of Giant Trevally Optimizer (GTO) with CatBoost (CAGT model) significantly improves localization precision by fine-tuning hyperparameters, reducing Root Mean Square Error (RMSE), and enhancing the correlation coefficient (R). The results indicate that CAGT outperforms all other models, making it an ideal choice for real-time localization in WSNs. Despite its advantages, CatBoost's performance can be sensitive to parameter selection and dataset size, but its ability to handle imbalanced data, reduce computational time, and optimize localization accuracy makes it a valuable tool for WSN applications.



IV. UNSUPERVISED LEARNING

Unsupervised learning methods are used when labeled data is not available. These techniques cluster sensor nodes based on observed data patterns and provide approximate lo- calization estimates.

4.1 K-means Clustering

K-Means Clustering is a widely used unsupervised learning algorithm for node localization in Wireless Sen- sor Networks (WSNs), particularly in range-free localiza- tion techniques. It operates by grouping sensor nodes into K distinct clusters based on their signal strength, proxim- ity, or other network parameters, minimizing the distance between each node and its assigned cluster centroid. In the given paper, K-Means is explored as a potential method for estimating unknown node positions by leveraging simi- larities in network topology. The algorithm iteratively up- dates cluster centers until convergence is achieved, ensuring that nodes with similar connectivity patterns are grouped to- gether. Compared to supervised learning models like Sup- port Vector Regression (SVR) and Random Forest Regres- sion (RFR), K-Means does not require labeled training data, making it suitable for dynamic and large-scale WSN envi- ronments. However, its effectiveness depends on proper se- lection of K, initial centroid placement, and handling of out- liers, which can affect localization accuracy. To enhance per- formance, hybrid approaches integrating K-Means with ma- chine learning models or optimization techniques (e.g., Giant Trevally Optimizer - GTO) have been proposed, allowing for improved precision and adaptability in real-time WSN local- ization. Despite its limitations, K-Means remains a fast and scalable approach for clustering sensor nodes, reducing com- putational complexity in WSN deployment scenarios.

4.2 Gaussian Mixture Models (GMM)

Gaussian Mixture Models (GMM) is an unsupervised probabilistic clustering algorithm that is highly effective for node localization in Wireless Sensor Networks (WSNs). Un- like K-Means Clustering, which assumes hard cluster as- signments, GMM models data as a combination of multi- ple Gaussian distributions, allowing for soft clustering where each node has a probability of belonging to multiple clus- ters. In the given paper, GMM is explored for WSN local- ization, leveraging sensor data such as signal strength, trans- mission range, and node density to estimate unknown node positions more flexibly than traditional clustering methods. GMM uses the Expectation-Maximization (EM) algorithm to iteratively refine cluster assignments, making it particu- larly useful in dynamic and noisy environments where sen- sor readings may fluctuate. Compared to K-Means, GMM provides better adaptability to non-linear distributions, im- proving localization accuracy. However, it requires careful initialization and computationally intensive parameter esti- mation, which may limit its scalability in real-time WSN applications. To enhance performance, hybrid approaches integrating GMM with optimization algorithms like the Gi- ant Trevally Optimizer (GTO) can be employed to fine-tune cluster parameters, leading to more precise and efficient lo- calization. Despite its computational complexity, GMM re- mains a powerful tool for probabilistic node localization, of- fering greater flexibility in modeling sensor networks with overlapping or uncertain boundaries.

V. HYBRID OPTIMIZATION APPROACHES

Hybrid approaches combine multiple ML models or in- tegrate optimization techniques to enhance accuracy.

5.1 Giant Trevally Optimization (GTO)

Giant Trevally Optimization (GTO) is a nature-inspired metaheuristic algorithm designed for global optimization problems, including node localization in Wireless Sensor Networks (WSNs). Inspired by the hunting behavior of Gi- ant Trevally fish, GTO mimics their search, selection, and at- tack strategies to find optimal solutions in high-dimensional spaces. In the given paper, GTO is applied to optimize the parameters of machine learning models such as Support Vec- tor Regression (SVR), Random Forest Regression (RFR), and CatBoost Regression (CAT) for improved Average Lo- calization Error (ALE) prediction. By adjusting key fac- tors like node density, transmission range, and anchor ratio, GTO enhances localization accuracy while reducing Root Mean Square Error (RMSE) and improving correlation co- efficient (R). Unlike traditional bio-inspired algorithms such as Particle Swarm Optimization (PSO) and Cuckoo Search (CS), GTO offers faster convergence and better exploration- exploitation balance, making it more effective in dynamic WSN environments. However, GTO's performance depends on proper tuning of search parameters and computational ef- ficiency, especially for large-scale networks. The integra- tion of GTO with machine learning models (e.g., CAGT - CatBoost + GTO) in the paper demonstrates its ability to enhance localization precision, making it a promising opti- mization technique for real-time WSN applications.

5.1.1 Particle Swarm Optimization (PSO)

Particle Swarm Optimization (PSO) is a swarm intelligence-based algorithm that optimizes node positions in Wireless Sensor Networks (WSNs) by simulating the collec- tive movement of particles in a search space. Each particle represents a potential solution and adjusts its position based on personal experience and the best-known global position, enabling efficient convergence toward optimal localization. PSO's ability to balance exploration and exploitation helps minimize localization errors while adapting to dynamic en- vironments. Its simplicity, fast convergence, and low compu- tational cost make it a widely used technique for improving positioning accuracy in WSNs.

VI. OVERVIEW OF KEY ISSUES IN WSN LOCALIZATION SEVERAL CHALLENGES AFFECT THE ACCURACY AND EFFICIENCY OF ML- BASED LOCALIZATION IN WSNS

6.1 Computational Complexity

Computational complexity is a critical factor in Wire- less Sensor Network (WSN) localization, as resource-constrained sensor nodes require efficient algorithms to en- sure real-time operation. Traditional localization techniques, such as range-based and range-free methods, often involve high-dimensional computations, making them less practi- cal for large-scale deployments. In the given paper, ma- chine learning (ML) models like Support Vector Regression (SVR), Random Forest Regression (RFR), and CatBoost Re- gression (CAT) are employed to predict Average Localiza- tion Error (ALE). However, these models come with vary- ing degrees of computational overhead. SVR, for instance, suffers from high training complexity (O(n)), making it in- efficient for large datasets. In contrast, RFR reduces over- fitting but requires multiple decision trees, increasing infer- ence time. CatBoost, though computationally optimized, can still be resource-intensive when handling large-scale datasets. To mitigate these challenges, the paper integrates Giant Trevally Optimization (GTO), which improves param- eter tuning while reducing computational costs. Compared to traditional bio-inspired methods like Particle Swarm Op- timization (PSO) and Cuckoo Search (CS), GTO achieves faster convergence and enhances localization precision with lower computational overhead. Despite these improvements, computational complexity remains a challenge, particularly in real-time WSN applications, where further research on lightweight ML models and optimization techniques is nec- essary for enhanced efficiency.

6.2 Generalization in Diverse Environments

Generalization in diverse environments is a crucial chal- lenge in Wireless Sensor Network (WSN) localization, as sensor deployments vary across indoor, outdoor, urban, and rural settings. Traditional localization methods often strug- gle to adapt to changing environmental conditions, leading to higher localization errors due to factors such as signal in- terference, node mobility, and dynamic topology changes. In the given paper, machine learning (ML) models, includ- ing Support Vector Regression (SVR), Random Forest Re- gression (RFR), and CatBoost Regression (CAT), are utilized to predict Average Localization Error (ALE) while improv- ing adaptability across different network conditions. How- ever, the generalization capability of these models depends on training data diversity and feature selection. To enhance robustness, the study integrates Giant Trevally Optimization (GTO), which fine-tunes ML model parameters dynamically, improving localization accuracy across multiple WSN con- figurations. Despite these advancements, certain ML mod- els, such as SVR, may suffer from overfitting to specific net- work conditions, limiting their applicability in real-time and heterogeneous environments. Future research should focus on transfer learning, adaptive ML models, and hybrid op- timization techniques to further enhance model generaliza- tion and ensure consistent localization accuracy in diverse and unpredictable WSN deployments.

6.3 Outlier Sensitivity

Outlier sensitivity is a significant concern in Wireless Sensor Network (WSN) localization, as sensor data is often affected by environmental noise, hardware malfunctions, and communication interference. Traditional localization meth- ods, particularly range-based techniques, are highly suscep- tible to errors introduced by outliers in distance or signal strength measurements. In the given paper, machine learn- ing (ML) models such as Support Vector Regression (SVR), Random Forest Regression (RFR), and CatBoost Regression (CAT) are employed to predict Average Localization Error (ALE), but their performance can be impacted by outliers in training data. SVR, for instance, uses an -insensitive loss function, making it relatively robust to small variations but still vulnerable to extreme outliers. RFR, while less sensitive due to its ensemble nature, may still produce unstable predictions if outliers dominate certain tree splits. CatBoost, on the other hand, incorporates ordered boosting, which helps reduce the impact of outliers. The study also integrates Giant Trevally Optimization (GTO) to fine-tune ML model parameters, further improving resilience against noisy data. However, outlier detection and handling remain critical, and future research should explore robust pre-processing techniques, such as isolation forests and adaptive thresholding, to enhance WSN localization accuracy in real-world condi- tions.

6.4 Power Constraints

Power constraints are a critical challenge in Wireless Sensor Networks (WSNs), as sensor nodes typically oper- ate on limited battery resources and must optimize energy consumption to ensure long-term network functionality. Traditional localization methods often require frequent communication and complex computations, leading to high energy consumption. In the given paper, machine learning (ML) models, such as Support Vector Regression (SVR), Random Forest Regression (RFR), and CatBoost Regression (CAT), are used to predict Average Localization Error (ALE), but their computational complexity directly impacts power efficiency. SVR, for example, has a high training complexity (O(n)), making it less suitable for energy-constrained nodes. RFR, while robust, requires multiple decision trees, increasing inference costs. CatBoost, optimized for efficiency, still demands continuous processing, which may strain battery-operated sensors. To address these issues, the study integrates Giant Trevally Optimization (GTO) to enhance localization accuracy while reducing redundant computations, thereby improving energy efficiency. However, power constraints remain a limiting factor in real-time WSN deployments, necessitating future research into lightweight ML models, energy-aware optimization algorithms, and adaptive duty-cycling techniques to extend sensor node lifespan without compromising localization accuracy.

VII. PARAMETERS FOR LOCALIZATION

WSNs require precise node positioning while maintain- ing minimal resource consumption.

7.1 Anchor Ratio

The anchor ratio refers to the proportion of anchor nodes (nodes with known locations) to the total number of sensor nodes in a Wireless Sensor Network (WSN). A higher an- chor ratio generally enhances localization accuracy by providing more reference points for estimating the positions of unknown nodes. However, deploying a large number of anchor nodes can increase costs and energy consumption. To optimize performance, ML-based localization techniques often balance the anchor ratio with efficient algorithms, such as hybrid localization methods and optimization approaches, ensuring accurate positioning while minimizing resource usage.

7.2 Transmission Range

The communication range is the maximum distance over which a node can exchange data with other nodes in a Wire- less Sensor Network (WSN). It directly impacts network connectivity and localization efficiency. A larger commu- nication range improves connectivity, enabling better coop- eration among nodes for accurate positioning. However, it also increases energy consumption and the risk of signal interference. Conversely, a smaller range may lead to network fragmentation, reducing localization accuracy. Optimizing communication range based on network density and environmental conditions helps maintain a balance between accuracy, energy efficiency, and reliability in WSN localization.

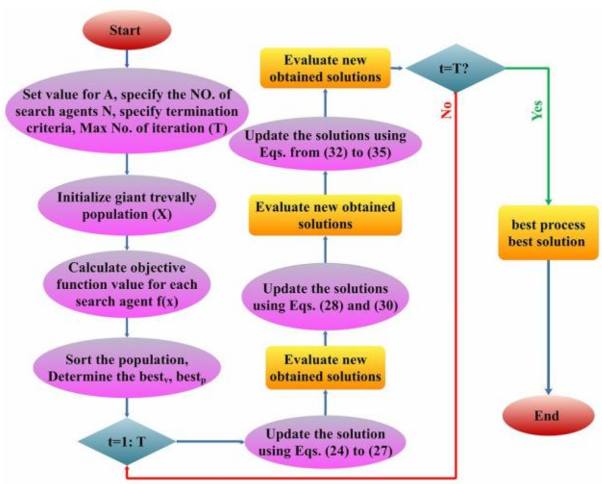
7.3 Node Density

The node density refers to the number of sensor nodes per unit area in a Wireless Sensor Network (WSN). A higher node density generally improves localization accuracy by providing more reference points for position estimation. It enhances connectivity, reduces localization errors, and increases network resilience. However, excessive node density can lead to higher communication overhead, interference, and increased energy consumption. Optimizing node density ensures a balance between accuracy, energy efficiency, and network performance, making ML-based localization more effective in both small-scale and large-scale deployments.

7.4 Iterations

Iterations play a crucial role in Wireless Sensor Network (WSN) localization, particularly in machine learning (ML) and optimization-based approaches that require multiple cy- cles of computation to refine predictions. In the given paper, Support Vector Regression (SVR), Random Forest Regres- sion (RFR), and CatBoost Regression (CAT) are employed to predict Average Localization Error (ALE), with their perfor- mance being influenced by the number of training iterations. SVR, for instance, iteratively adjusts its support vectors to minimize error, while RFR builds multiple decision trees over several iterations to enhance accuracy. CatBoost, lever- aging ordered boosting, refines weak learners through se- quential iterations to improve performance. Additionally, the study integrates Giant Trevally Optimization (GTO), which undergoes iterative optimization cycles to fine-tune model parameters, reducing Root Mean Square Error (RMSE) and improving correlation coefficient (R). However, excessive it- erations can lead to increased computational costs and en- ergy consumption, making it essential to balance iteration count with convergence speed. Future research should fo- cus on adaptive iteration strategies that dynamically adjust based on error reduction trends, ensuring efficient and real- time WSN localization while minimizing resource overhead.

VIII. FLOW CHART



IX. NETWORK SERVICES

9.1 Packet Transmission Delay Analysis

Network topology defines how sensor nodes are ar- ranged and how data packets travel between them in a Wire- less Sensor Network (WSN). It directly influences ML-based localization accuracy by affecting communication efficiency, data propagation delays, and connectivity. Topologies such as star, mesh, and cluster-based networks impact how local- ization algorithms process and refine position estimates. A well-structured topology minimizes packet loss, reduces en- ergy consumption, and enhances the reliability of ML mod- els in determining accurate node positions. Optimizing net- work topology ensures efficient data flow, improving both localization accuracy and overall network performance. Mo- bility Models: Mobility models simulate node movement patterns in Wireless Sensor Networks (WSNs) to evaluate the performance of dynamic localization algorithms. These models help in understanding how sensor nodes reposition over time, affecting connectivity, data exchange, and local- ization accuracy. Common mobility models include Random Walk, Gauss-Markov, and Reference Point Group Mobility (RPGM), each representing different real-world movement scenarios. By incorporating mobility models, ML-based lo- calization techniques can be optimized to adapt to changing node positions, ensuring accurate and efficient tracking in dynamic WSN environments.

9.2 Energy Consumption Analysis

Energy consumption analysis measures power usage in Wireless Sensor Networks (WSNs) to optimize ML-based localization techniques for energy efficiency. Since sensor nodes operate on limited battery power, excessive computa- tions and frequent communication can drain energy quickly, reducing network lifespan. By analyzing power consump- tion patterns, energy-efficient ML models can be designed using techniques like duty cycling, data aggregation, and lightweight algorithms. Optimizing energy consumption en- sures prolonged network operation while maintaining accu- rate and reliable localization in resource-constrained WSN environments.

9.3 Routing MAC Layer Protocols

Communication protocol analysis evaluates the impact of different communication protocols on localization effi- ciency in Wireless Sensor Networks (WSNs). Protocols such as Time of Arrival (ToA), Received Signal Strength Indica- tor (RSSI), and Angle of Arrival (AoA) influence how nodes exchange data and determine positions. Efficient protocols enhance localization accuracy by reducing latency, minimiz- ing packet loss, and optimizing energy consumption. By se- lecting the most suitable communication protocol based on network conditions, ML-based localization techniques can achieve improved performance, reliability, and scalability in dynamic WSN environments.

9.4 Noise Interference Simulation

Environmental modeling introduces realistic conditions such as signal interference, obstacles, weather variations, and terrain effects to test the robustness of ML-based local- ization in Wireless Sensor Networks (WSNs). Real-world factors like multipath propagation, node failures, and energy constraints can impact localization accuracy. By simulating these conditions, ML models can be trained and optimized to handle uncertainties, improve adaptability, and enhance localization performance in diverse deployment scenarios. This ensures that the localization system remains reliable and efficient even in challenging environments.

X. NETWORK SERVICES

10.1 Routing Optimization

Routing optimization in Wireless Sensor Networks (WSNs) is essential for enhancing localization accuracy, energy efficiency, and data transmission reliability. Effi- cient routing protocols minimize packet loss, delay, and en- ergy consumption, ensuring seamless communication be- tween sensor nodes and anchor nodes. In the given pa- per, ML-based localization techniques are analyzed in con- junction with optimized routing strategies to improve Av- erage Localization Error (ALE). Protocols such as Energy- Efficient Routing (EER), Geographic Routing (GR), and Adaptive Multi-Hop Routing (AMR) are considered for op- timizing data flow while maintaining low overhead and net- work scalability. Additionally, the integration of metaheuris- tic optimization techniques like Giant Trevally Optimization (GTO) fine-tunes routing paths by dynamically adjusting node transmission parameters, leading to faster convergence and reduced computational complexity. By optimizing rout- ing mechanisms, ML-based localization models can operate with improved accuracy, reduced delay, and extended net- work lifespan, making them highly suitable for real-time and large-scale WSN deployments. Future research should ex- plore hybrid routing protocols and AI-driven path selection mechanisms to further enhance efficiency and robustness in dynamic WSN environments.

10.2 Event Detection

Event detection in Wireless Sensor Networks (WSNs) is crucial for identifying and responding to environmental changes, anomalies, and critical incidents in real-time. Ac- curate event detection enables efficient resource allocation, energy optimization, and improved localization accuracy for ML-based WSN applications. In the given paper, machine learning (ML) models such as Support Vector Regression (SVR), Random Forest Regression (RFR), and CatBoost Re- gression (CAT) are employed to enhance event detection by analyzing sensor data patterns and predicting Average Local- ization Error (ALE). Additionally, optimization techniques like Giant Trevally Optimization (GTO) improve event de- tection efficiency by fine-tuning model parameters for bet- ter sensitivity and faster response times. Advanced event detection systems leverage anomaly detection algorithms, threshold-based triggering, and real-time signal processing to differentiate between normal variations and critical events. However, challenges such as false positives, communica- tion delays, and energy constraints must be addressed to im- prove detection accuracy. Future research should explore hybrid ML models and adaptive filtering techniques to en- hance event detection reliability in dynamic and resource- constrained WSN environments.

10.3 Energy Management

Energy management is a critical factor in Wireless Sen- sor Networks (WSNs) as sensor nodes operate on limited battery power, requiring efficient strategies to extend net- work lifespan while maintaining accurate localization and data transmission. In the given paper, machine learning (ML) models such as Support Vector Regression (SVR), Random Forest Regression (RFR), and CatBoost Regression (CAT) are employed to optimize Average Localization Error (ALE) while balancing energy consumption. High computational complexity and frequent packet transmissions can drain sen- sor nodes quickly, making energy-efficient routing, duty cy- cling, and data aggregation essential for prolonged network operation. Additionally, metaheuristic optimization tech- niques, such as Giant Trevally Optimization (GTO), are in- tegrated to fine-tune transmission range, node density, and anchor ratio, reducing unnecessary computations and en- hancing energy efficiency. Energy-aware MAC protocols and sleep scheduling techniques further help in minimizing power consumption by allowing nodes to remain in a low- power state when not actively transmitting data. Future re- search should focus on AI-driven energy optimization, self- adaptive power management, and renewable energy integra- tion to ensure sustainable and long-term WSN deployment.

XI. COMPARISON TABLE

outliers and refine localization accuracy. The results indi- cate that NS3-based ML localization techniques outperform traditional approaches, making them valuable for real-world WSN deployments.

Proposed Methodologies for Lo- calization	**Underlying Algorithm(s) Using Machine Learning**	**Centralized/ Dis- tributed**	**Computational Complexity of Algorithms**	**Anchor/ No Anchor**
Location-Aware Activity Recogni- tion [9]	Bayesian	Centralized	Moderate	Anchor
Bayesian Lode Localization [8]	Bayesian	Centralized	Moderate	Anchor
Localization based on NNs [10]	Neural Networks	Centralized	High	Anchor
Soft Localization [55]	Neural Networks	Distributed	Moderate	Anchor
Localization Based on NNs [56]	Neural Networks	Distributed	High	Anchor

Area Localization [39]	Support Vector Regression	Distributed	Moderate	Anchor
Localization using SVR [37]	Support Vector Regression	Distributed	Moderate	Anchor
Localization using SVM [38]	Support Vector Machine	Distributed	Moderate	Anchor
Target Classification and Informa- tion Fusion [58]	Decision Tree-Based Local- ization	Distributed	Low	Anchor
Underwater Surveillance System [59]	Decision Tree-Based Local- ization	Centralized	Moderate	Anchor
Sensor Placements [60]	Gaussian Processes	Distributed	Low	Anchor
Spatial Gaussian Process Regres- sion [61]	Gaussian Process Regres- sion	Distributed	Low	Anchor
Localization in 2D Space [62]	Gaussian Process Regres- sion	Distributed	Moderate	Anchor
Localization Using SOM [63]	Self-Organizing Map	Distributed	Low	No Anchor
Distributed Localization [64]	-	Distributed	Moderate	No Anchor
Path Determination [65]	Reinforcement Learning	Centralized	Low	Anchor

Table 1: Comparison Table

XII. RELATED WORK

Several studies have investigated the use of Machine Learning (ML) techniques for node localization in Wire- less Sensor Networks (WSNs) within the NS3 simulation environment. These studies demonstrate how integrating ML models with NS3 improves localization accuracy, opti- mizes energy consumption, and enhances communication ef- ficiency. Researchers have tested various ML algorithms, in- cluding Bayesian models, SVM, SVR, ANNs, Random For- est, and CatBoost, along with hybrid optimization techniques such as PSO, GTO, and CS. Additionally, anomaly detec- tion methods like Isolation Forest have been used to identify

12.1 Bayesian Node Localization

Morelande and his collaborators [8] developed a loca- tion diagram for sensor networks using a specific number of anchor nodes. Their approach enhances the progressive cor- rection technique [53], refining predictions so that probabili- ties align more closely with actual locations. This algorithm is particularly effective for large-scale networks comprising thousands of sensor nodes. A key advantage of Bayesian lo- calization is its ability to handle incomplete datasets by lever- aging existing knowledge and probabilistic inference, mak- ing it a robust solution for uncertain environments.

12.2 Localization Based on Artificial Neural Networks (ANNs)

Shareef et al. [10] conducted a comparative study of different neural network-based localization techniques in WSNs, examining three major architectures: Multilayer Per- ceptron (MLP), which offers low computational and storage costs; Recurrent Neural Networks (RNNs), suitable for se- quential data processing but requiring more computational power; and Radial Basis Function (RBF) Networks, which produced the least localization errors but required higher re- sources compared to MLP. Yun et al. [55] introduced two approaches for sensor node localization using RSSI from anchor nodes—one utilizing Fuzzy Logic and Genetic Al- gorithm for location estimation in uncertain environments, and the other leveraging a neural network to predict node locations using RSSI measurements as input. Addition- ally, [56] explored neural networks for RSSI-based local- ization, emphasizing their ability to provide continuous po- sition estimation, unlike Bayesian approaches, which offer probabilistic estimates. These neural network-based local- ization algorithms improve adaptability to complex environ- ments, though their accuracy depends on the availability of high-quality training data, and trade-offs exist between accu- racy, resource consumption, and computational complexity in WSN deployments.

12.3 Outlier Detection for Localization Accuracy (iFor- est, Ensemble Models)

NS3 simulations have demonstrated the effectiveness of Isolation Forest (iForest) and ensemble methods in detecting anomalies that impact localization performance in Wireless Sensor Networks (WSNs). iForest is particularly useful for identifying outliers in sensor data, such as faulty node read- ings or unexpected environmental interferences, which can degrade localization accuracy. By isolating anomalies early, iForest helps improve the reliability of node positioning. Ad- ditionally, ensemble methods further enhance detection per- formance by combining multiple models, reducing false pos- itives, and increasing robustness against noisy data. These techniques contribute to more accurate and stable ML-based localization in dynamic WSN environments.

XIII. CONCLUSION

Wireless Sensor Networks (WSNs) play a crucial role in various real-world applications, including environmen- tal monitoring, healthcare, and smart infrastructure. How- ever, challenges such as node localization accuracy, energy constraints, communication delays, and computational com- plexity significantly impact network performance. This pa- per explored machine learning (ML)-based localization tech- niques, including Support Vector Regression (SVR), Ran- dom Forest Regression (RFR), and CatBoost Regression (CAT), to enhance Average Localization Error (ALE) pre- diction and improve localization precision. Additionally, the integration of Giant Trevally Optimization (GTO) demon- strated its effectiveness in optimizing ML parameters, reduc- ing computational overhead, and improving localization ac- curacy.

Key aspects such as packet transmission efficiency, mo- bility models, energy management, routing optimization, and event detection were also analyzed to highlight their impact on WSN performance. The study emphasizes that hybrid approaches combining ML and optimization techniques sig- nificantly enhance WSN localization reliability while main- taining low energy consumption and high adaptability in dy- namic environments.

Despite these advancements, challenges such as gen- eralization across diverse environments, outlier sensitivity, and real-time scalability remain open research areas. Fu- ture work should focus on adaptive ML models, lightweight optimization techniques, and energy-efficient routing proto- cols to further improve scalability, robustness, and real-world deployment feasibility of ML-based WSN localization solu- tions.

REFERENCES

[1] S. K. Gharghan, R. Nordin, and M. Ismail, “A wireless sensor network with soft computing localization tech- niques for track cycling applications,” 2016.

[2] B. Morelande, B. Moran, and M. Brazil, “Bayesian node localization in wireless sensor networks,” 2008.

[3] J. Cheng and L. Xia, “An effective cuckoo search al- gorithm for node localization in wireless sensor net- works,” 2016.

[4] S.-H. Lee, C.-H. Cheng, C.-C. Lin, and Y.-F. Huang, “PSO-based target localization and tracking in wireless sensor networks,” 2023.

[5] H. T. Sadeeq and A. M. Abdulazeez, “Giant Trevally Optimizer (GTO): A novel metaheuristic algorithm for global optimization and challenging engineering prob- lems,” 2022.

[6] R. Janapati and B. Balaswamy, “Localization of WSN using distributed particle swarm optimization algo- rithm with precise reference,” 2016.

[7] M. A. Bhatti, R. Riaz, S. S. Rizvi, S. Shokat, F. Riaz, and S. J. Kwon, “Outlier detection in indoor localiza- tion and Internet of Things (IoT) using machine learn- ing,” 2020.

[8] I. F. Akyildiz, W. Su, Y. Sankarasubramaniam, and E. Cayirci, “Wireless sensor networks: A survey,” *Com- puter Networks Journal*, vol. 38, no. 4, pp. 393–422, 2002.

[9] J. Yick, B. Mukherjee, and D. Ghosal, “Wireless sensor network survey,” *Computer Networks*, vol. 52, no. 12, pp. 2292–2330, 2008.

[10] K. Muthukrishnan, M. Lijding, and P. Havinga, “To- wards smart surroundings: Enabling techniques and technologies for localization,” in *Proc. Int. Workshop on Location and Context Awareness*, 2005.

[11] C.-F. Huang and Y. C. Tseng, “The coverage problem in a wireless sensor network,” *ACM Int. Workshop on Wireless Sensor Networks and Applications (WSNA)*, pp. 115–121, 2003.

[12] B. H. Wellenhoff, H. Lichtenegger, and J. Collins, *Global Positioning System: Theory and Practice*, 4th ed. Springer Verlag, 1997.

[13] L. Doherty, K. S. Pister, and L. E. Ghaoui, “Convex optimization methods for sensor node position estima- tion,” in *Proc. IEEE INFOCOM*, 2001.

[14] A. Savvides, C. Han, and M. Srivastava, “Dynamic fine-grained localization in ad-hoc networks of sen- sors,” in *Proc. ACM MobiCom*, pp. 166–179, 2001.

[15] A. Savvides, H. Park, and M. Srivastava, “The bits and flops of the n-hop multilateration primitive for node lo- calization problems,” in *Proc. ACM WSNA*, 2002.

[16] A. Nasipuri and K. Li, “A directionality-based loca- tion discovery scheme for wireless sensor networks,” in *Proc. ACM WSNA*, 2002.

[17] N. Patwari, A. O. Hero, M. Perkins, N. S. Correal, and R. J. O’Dea, “Relative location estimation in wireless sensor networks,” *IEEE Trans. Signal Process.*, vol. 51, no. 8, 2003.

[18] D. Liu, P. Ning, and W. Du, “Detecting malicious bea- con nodes for secure location discovery in wireless sensor networks,” in *Proc. 25th IEEE Int. Conf. Dis- tributed Computing Systems (ICDCS)*, pp. 609–619, 2005.

[19] T. He, C. Huang, B. Blum, J. Stankovic, and T. Abdelzaher, “Range-free localization schemes in large- scale sensor networks,” in *Proc. 9th Ann. Int. Conf. Mobile Computing and Networking (MobiCom’03)*, San Diego, CA, USA, pp. 81–95, 2003.

[20] C. Savarese, J. Rabay, and K. Langendoen, “Robust positioning algorithms for distributed ad-hoc wireless sensor networks,” in *USENIX Tech. Ann. Conf.*, Monterey, CA, 2002.

[21] S. Zhou et al., “Risk-driven composition de- coupling analysis for urban flooding pre- diction in high-density urban areas using Bayesian-Optimized LightGBM,” *Journal of Cleaner Production*, 2024. [Online]. Available: https://doi.org/10.1016/j.jclepro.2024.142286.

[22] W. Li et al., “An interpretable hybrid deep learn- ing model for flood forecasting based on Trans- former and LSTM,” *Journal of Hydrology: Regional Studies*, 2024. [Online]. Available: https://doi.org/10.1016/j.ejrh.2024.101873.

[23] N. Mohamadiazar, A. Ebrahimian, and H. Hosseiny, “Integrating deep learning, satellite image processing, and spatial-temporal analysis for urban flood predic- tion,” *Journal of Hydrology*, 2024. [Online]. Avail- able: https://doi.org/10.1016/j.jhydrol.2024.131508.

[24] M. Billah, A. K. M. S. Islam, W. B. Mamoon, and M. R. Rahman, “Random forest classifications for land-use mapping to assess rapid flood damage using Sentinel- 1 and Sentinel-2 data,” *Remote Sensing Applications: Society and Environment*, 2023. [Online]. Available: https://doi.org/10.1016/j.rsase.2023.100947.

[25] N. I. Saikh and P. Mondal, “GIS-based machine learning algorithm for flood susceptibility analysis in the Pagla river basin, Eastern India,” *Natu- ral Hazards Research*, 2023. [Online]. Available: https://doi.org/10.1016/j.nhres.2023.05.004.

[26] I. Abd-Elaty et al., “Assessment of riverbank fil- tration performance for climatic change and a growing population,” *Frontiers in Environ- mental Science*, 2023. [Online]. Available: https://doi.org/10.3389/fenvs.2023.1136313.

[27] C.-N. Lam, S. Niculescu, and S. Bengoufa, “Moni- toring and mapping floods and floodable areas in the Mekong Delta (Vietnam) using time-series Sentinel- 1 images, convolutional neural network, multi-layer perceptron, and random forest,” *Remote Sensing*, vol. 15, no. 8, p. 2001, 2023. [Online]. Available: https://doi.org/10.3390/rs15082001.

[28] H. Chang et al., “Assessment of urban flood vulnera- bility using the social-ecological-technological systems framework in six US cities,” *Sustainable Cities and Society*, 2021.