Machine Learning-based Regression Analysis and Feature Ranking for Localization Error Prediction in Wireless Sensor Networks

Peng Wang[1*], Qiuying Han[2], Shaohui Zhang[1,3], Zhaodi Wu[4]
[1]School of Artificial Intelligence, Zhoukou Normal University, Zhoukou, 466000, Henan, China
[2]School of Computer Science and Technology, Zhoukou Normal University, Zhoukou, 466000, Henan, China
[3]School of Artificial Intelligence, Zhengzhou University, Zhengzhou, 450001, Henan, China
[4]Network and Information Center, Guizhou Normal University, Guiyang, 550000, Guizhou, China,
[*]Corresponding author
E-mail: pwang678@163.com

Keywords: artificial intelligence, data visualization, machine learning, wireless sensor network

Received: January 16, 2025

Wireless Sensor Networks (WSNs) localization is crucial for identifying the position of sensor nodes, as many applications, including environmental monitoring, target tracking, and disaster management, require accurate location information. The objective of this research is to conduct extensive data analytics using visualization techniques to explore key factors influencing localization error and to develop machine learning models for forecasting Average Localization Error (ALE) in WSNs. A dataset containing 107 records, sourced from Kaggle's online repository, was analyzed using eXtreme Gradient Boosting (XGB) for feature ranking to determine the most influential factors affecting ALE. Multiple regression models, including Support Vector Regression (SVR), Decision Tree (DT), K-Nearest Neighbors (KNN), and AdaBoost Regressor, were applied to predict ALE. The models were evaluated using R-squared (R^2), Root Mean Square Error (RMSE), and computational efficiency. The results indicate that SVR achieved the highest accuracy with $R^2 = 0.99$ and the lowest RMSE of 0.01, significantly outperforming the other models (KNN: $R^2 = 0.55$, RMSE = 0.14; DT: $R^2 = 0.41$, RMSE = 0.16; AdaBoost: $R^2 = 0.72$, RMSE = 0.16). This study demonstrates that SVR is a highly effective model for ALE prediction, reinforcing the importance of feature ranking and selection in improving localization accuracy. The findings contribute to advancing machine learning-driven localization error prediction in WSNs and provide a foundation for further exploration of hybrid and deep learning-based models.

Povzetek: Predstavljen je SVR-model za napoved lokalizacijske napake v brezžičnih senzorskih omrežjih. Kvalitetno delovanje doseže z uporabo rangiranja značilk in XGB analize.

1 Introduction

Localization turns out to be an important process in Wireless Sensor Networks (WSN) since it aids in the identification of geographic locations of the sensor nodes which are of paramount importance in most applications ranging from environmental monitoring to tracking of targets to disaster management. In this way, accurate localization is important for meaningful data collection because it links measurements with certain locations. Due to these limitations of WSNs, particularly energy, processing power and coverage area, accurate and efficient localization algorithms are highly desirable [1]. Most of these techniques involve using anchor nodes with known coordinates, distance between different nodes and other optimizing algorithms that allows an approximate estimation of the non-anchor nodes without much error [2]. The *Average Localization Error* (ALE) is known to be a primary measure to assess the efficiency of localization algorithms. It looks at the difference between the expected position and the actual position of the sensor nodes and makes conclusions on the efficiency of the localization procedure [3]. These significant factors affect ALE in WSNs including density of anchor nodes, node mobility, environmental condition, and the measurement techniques that are used. The number, distribution and density of anchor nodes have been shown to play a very important role in the accuracy of localization. The number of anchor nodes should be small to reduce the burden for the search algorithm; however, this results in a larger ALE because there are few references points [4]. The high mobility introduces dynamic changes where most of the nodes' positions undergo other changes which increases the localization error [5].

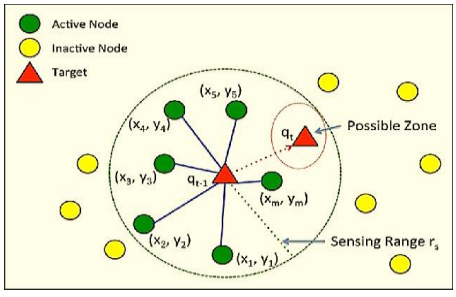


Figure 1: Localization sensor network [7]

Environmental conditions are indexes such as signal attenuation, multipath propagation or obstacles that distort distance or angle measures, and, thus, increase ALE [6]. About the measurement techniques like RSSD, TOA and AOA [7] are fundamentally different in terms of their accuracy and tolerance to noise in present network environment. Fig 1 shows the localization sensor networks are composed of active sensor nodes receiving the target's radio at current time step, with target's position at specific prediction time step.

ALE reduction provided by localization methods allow for correct positioning of sensors improving various applications, on environmental monitoring, targeting, and smart agriculture. In addition, network-based sensitive services [8] and efficient localization algorithms contribute significantly to the enhancement of network resource utilization sometimes shown from the energy consumed in the WSN hence enhancing the overall performance of WSNs in remote areas or in regions of resource constraint will be significantly influenced. Since WSNs are being used for critical applications there is more emphasis on getting the localization right for the best results for the network and to improve its performance.

The ALE Analysis is very significant as it has various applications, including

a) Disaster Management: Correct node localization also allows for effective organization of the necessary resources in the disaster area [9].

b) Smart Agriculture: Accurate location of sensors increases the efficiency of environmental condition monitoring and balancing workload [10][11].

c) Military Surveillance: Accurate positioning is vital for tracking and monitoring critical defense applications [12]. As a result of its widespread use in WSNs, there are diverse methods which can be employed to reduce ALE such as optimization, hybridization of WSN localization and machine learning mechanism. The algorithms such as Particle Swarm Optimization (PSO) and Genetic Algorithms (GA) do the optimization of the anchor node placement and reduce the error propagation as mentioned above [12]. Secondly, hybrid localization techniques are used to improve the accuracy of range-based and range-free methods while at the same time being scalable [13]. Furthermore, current ML based approaches of supervised and unsupervised learning enhance the localization estimate while learning patterns in the network data [14]. This calls for the need to increase accuracy in localization in WSN, which is the motivation for this research given the various applications of WSN in many different and complex environments. Since WSNs are used in disaster relief, military, and environmental monitoring, the localization is significant for decision making and resource control. The increase of the WSN systems' dynamic node behavior, different environmental conditions, and huge number of nodes increases demand for using more sophisticated algorithm that provide minimize ALE demand using energy consumption rules [15]. Machine learning and optimization approaches present potential solutions that are flexible and able to be constrained by the available data in real-time application that have better accuracy compared with conventional methods.

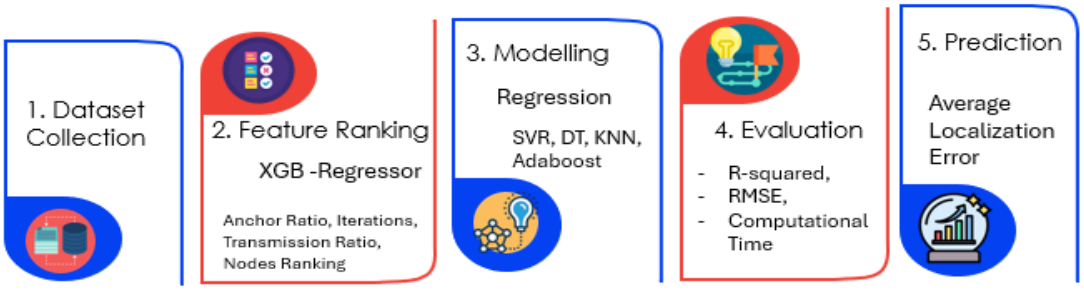


Figure 2: Steps of applied research methodology

In this research study, we plan to achieve the aim of predicting the Average Localization Error (ALE) through a big dataset collected from online repositories where all the possible network parameters affecting localization accuracy are incorporated including anchor ratio, iteration, transmission range and node density. The dataset used is then pre-processed for performing feature ranking and their importance analysis using the ensemble-based regressor model XGB before proceeding to determine parameters influencing the ALE. Subsequently, we build four regression models such as Support Vector Regression (SVR), Decision Tree Regressor (DT), K-Nearest Neighbor Regressor (KNN), and AdaBoost Regressor to predict ALE using ranked features. The accuracy of these models is determined by R-squared, and Root Mean Square Error (RMSE) from which it emerged that. Besides, the time taken in performing calculations is captured for each of them to determine their time efficiency in carrying out real-time analyses. This kind of approach, describes using fig 2, brings about comprehensive evaluation of ALE prediction, aside from the overall efficiency of the models

2 Related work

There has been various types of localization methods that have been proposed to enhance the various challenges that have developed over the years such as the energy problems, accuracy and scalability. The existing studies can further be categorised into below different categories, as shown in fig 3, describing: Range-based localization utilizes received signal strength indication, direction, angle, time-of-arrival, time-difference-of-arrival, and signal strength to estimate a node position. Some of the commonly used techniques include; a) time of arrival b)

received strength indicator. Time of Arrival is a technique that determines the time that takes for signal to be received between a transmitter and a receiver. This method is very accurate, but may be tricky because the timings of these procedures should be synchronized; not a simple feat in WSNs [16].

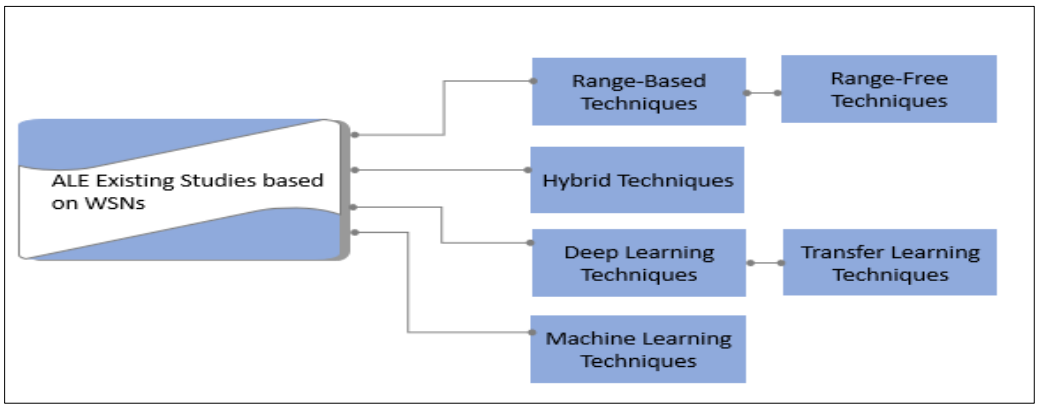


Figure 3: Taxonomy of localization studies

The Received Signal Strength Indicator (RSSI) uses signal strength signal strength to determine distance with reference to attenuation. While it is power saving and simple to set up, it is vulnerable to environment for example barriers and intensifications [17]. Angle of Arrival helps in deciding as to what direction, a signal reaches a particular node. It is effective in reducing localization error but the complication of the method demands complex hardware such as antenna arrays leading to high costs [18].

The range-free techniques do not make use of distance or angle estimation, which makes them ideal for low-cost large-scale deployment. There are main methods which include centroid localization, DV-Hop algorithm and Approximate Point in Triangle (APIT) [19]. Centroid localization technique involves using the centroid of anchor nodes within the communication range as a basis of estimating unknown node's position. Even though this method is simpler and computationally less demanding, it proved less accurate in sparse networks [20]. DV-Hop is a distributed method used in estimating node positions through hop counts of neighboring nodes and the average hop distance. It is accurate to measure occurrences of transmission but is not directly scalable to networks of irregular topologies [21].

By employing anchor nodes and triangulating the network area, a node's position is derived by its membership to particular triangular regions. While it works well in some situations, the rate at which it achieves its goal declines in networks which involve node mobility [22]. Hybrid localization is a process of combining both range-based and range-free methods their strengths while avoiding the weaknesses associated with each type. For example, in the Hybrid TOA-DV-Hop method where the TOA provides accurate distance resolution while the DV-Hop, although less accurate due to its low complexity is scalable [23]. Likewise, machine learning based hybrid methods like Artificial Neutral network and support vector machine can also deal with NLOS issues and environmental fluctuations [24]. Such use of ML techniques as deep learning allows nodes to improve estimations of location based on past data and environment [25]. In particular, algorithms such as particle swarm optimization and ant colony optimization originating from Swarm Intelligence, biological systems, effectively place anchors and localize nodes [26]. The IoT Integration techniques with 5G includes Integration of WSNs with 5G & IoT technologies enable ultra-reliable, low-latency localization which creates new opportunities of application like autonomous vehicle and smart city etc. [27].

The localization capabilities of WSNs have been improved greatly by integrating machine learning (ML) into WSNs by overcoming the issues of NLOS situations, environmental noise, and energy use [28]. Based on the above, self learning algorithms help WSN nodes forecast and respond to changes in the surroundings and enhance the degree of localization accuracy [29]. Supervised learning models of the localization algorithms cost more as they need labeled data to operate from. Some of the common used algorithms in WSN localization include; Support Vector Machines (SVM) and Neural Network (NN). SVMs are use in localization classification and regression. They are used particularly for determining LOS and NLOS situations and enhance localization in the urban areas [30]. For regression-based localization, feedforward and convolutional networks have been used. These models learn complicated relationships between certain input features such as received signal strength indication (RSSI) and geographical coordinates, for high accuracy [31]. Clustering of the sensor nodes is done without labeling them, making unsupervised learning methods very useful where we do not have the help of a learners-based approach and the environment is largely unknown. To locate other nodes and enhance the accuracy of the positioning of the anchor node, K-means and DBSCAN are incorporated to cluster the base stations/sensor nodes depending on their distance [31]. basically, reinforcement learning helps enable the sensor node to learn to optimize its positioning technique based on an interaction with the environment. Some RL based algorithms like Q-learning are used in determining the optimal positions for anchor nodes and transmission

power so as to reduce localization error [32]. Integrating the conventional localization techniques with ML produces the best results since the two work hand in hand. For example, combining Neural Networks and DV-Hop algorithms reduces computational effort without a compromise on the results [33]. Likewise, RSSI integrated with Random Forest, which belongs to hybrid models based on ensemble learning, have provided better localization accuracy when implemented in complicated terrains [34].

The two promising techniques of deep learning that are being used for localization of WSN are LSTM and Autoencoders. LSTMs deal with sequence data, therefore recommended for use in a dynamic context where node positions vary [35]. Autoencoder is applied for dimensionality reduction and feature extraction tasks, thus facilitating accurate localization in high dimensionality databases [36]. Federated learning is emerging for WSN localization because it enables distributed nodes to learn the model collectively without exchanging the raw information while protecting privacy and minimizing the exchange of data [37]. The benefits of transfer learning include an ability to fine-tune pre-learned models for WSN environments, and the ability to do so with little data, which saves time and computational power for localization [38].

Table 1: Summary of existing studies

Ref	Year	Method	Dataset	Results	Research Area	Key Findings	Limitations
[16]	2020	Range-Based	RSS Data	Mean localization error: 2.3m	Error analysis on RSS localization	Identified key errors in RSSI-based localization	Limited to RSSI-based methods, lacks generalization
[17]	2022	Range-Based	BLE RSSI	accuracy: 85%	Accuracy evaluation of BLE-based indoor localization	BLE RSSI accuracy analysis for indoor settings	Accuracy highly dependent on environmental factors
[18]	2020	Range-Based	AoA Measurements	precision: 92%	Optimized access point deployment for AoA localization	Improved AoA localization through optimal AP placement	Limited scalability in real-world applications
[19]	2024	Range-Free	UWB WSN	localization error: 1.8m	Hybrid DV-hop using PSO for range-free localization	Improved localization accuracy in UWB WSN	Computational overhead in PSO-based methods
[20]	2021	Range-Free	WSN	accuracy: 89%	Optimized DV-hop for range-free localization	Enhanced accuracy in range-free WSN localization	Accuracy drops in sparse node deployments
[21]	2023	Range-Free	3D DV-Hop	RMSE: 2.1m	3D DV-Hop localization optimization	Increased localization precision in 3D WSN	Performance depends on network density
[22]	2021	Range-Free	Convex Triangulation	localization error: 1.5m	Weighted uncertainty-based convex triangulation	More robust localization under uncertainty	Limited applicability in dynamic environments
[23]	2023	Range-Free	UWB MANET	False positive rate: 8%	Malicious node elimination in UWB MANETs	Secured localization against malicious nodes	False positives in malicious node detection
[24]	2023	ML-Based	UWB Sensor Data	accuracy: 94%	SVM-based LOS/NLOS classification	Improved LOS/NLOS classification for localization	Requires large training datasets
[25]	2024	DL-Based	Indoor Deep Learning	accuracy: 96%, F1-score: 0.93	Deep learning-based LOS/NLOS identification	High accuracy in real indoor environments	Computationally expensive
[26]	2021	ML-Based	Swarm Intelligence	precision: 90%	Swarm intelligence in WSN localization	Enhanced localization accuracy using swarm optimization	Sensitive to parameter tuning
[27]	2020	Range-Based	IoT Positioning	localization MAE: 2.5m	IoT positioning error mitigation strategies	Comprehensive error source analysis	No specific improvement strategies proposed
[28]	2023	ML-Based	WSN	accuracy: 87%	Systematic review on ML and optimization in WSN	Identified key ML-based localization techniques	Lacks empirical validation
[29]	2023	DL-Based	Target Localization	accuracy: 95%, RMSE: 1.4m	Deep learning for error-prone environments	High accuracy despite localization errors	Generalization to diverse environments is unknown
[30]	2020	ML-Based	Supervised Learning	F1-score: 0.89	Supervised learning for target localization	Improved localization accuracy	Dependent on labeled data
[31]	2019	ML-Based	Fingerprinting	MAE: 1.7m	Regression-based estimation for fingerprinting	Reduced localization errors in fingerprinting	Needs extensive fingerprint database
[32]	2023	ML-Based	5G Wireless	R2-score: 0.91	Reinforcement learning for 5G localization	Optimized power allocation for localization	Limited scalability to large networks
[33]	2022	DL-Based	ALE WSN	accuracy: 93%	DQ-learning for UWB LoS/NLoS node selection	Energy-optimized localization with deep RL	Complex training process
[34]	2024	ML-Based	5G Networks	RMSE: 1.3m	ML models for localization error prediction	High accuracy in error prediction	Performance varies across models

[35]	2020	DL-Based	UWB Indoor Localization	accuracy: 97%, MAE: 1.2m	LSTM-based UWB localization	Significant improvement in indoor localization	Computationally intensive
[36]	2019	DL-Based	Device-Free Localization	localizatio n error: 1.8m	Convolutional autoencoder for localization	Accurate device-free localization	Limited real-world applicability
[37]	2022	ML-Based	ALE WSN	precision: 92%	Federated learning for indoor localization	Improved model reliability with dropout	Communication overhead
[38]	2024	DL-Based	ALE WSN	accuracy: 94%, RMSE: 1.5m	Deep transfer learning for URLLC	High accuracy in interference-prone networks	Requires large dataset for fine-tuning

2.1 Problem statement

For the given problem of computation of ALE in WSNs, we define the formal problem statement as follows:

Given a wireless sensor network W consisting of N nodes, where $V = \{v_1, v_2, v_3, \ldots\ldots, v_n\}$ represents the set of sensor nodes, each node v_i has a actual position $(x_i, y_i) \in \mathbb{R}^2$. The ALE is defined as in 1:

$$ALE = \frac{1}{N}\sum_{i=1}^{N}\sqrt{(x_i, \dot{x}_i)^2 + (y_i, \dot{y}_i)^2} \quad (1)$$

The objective function is to minimize ALE by optimizing the localization algorithm or strategy used for estimating $(\dot{x}_i, \dot{y}_i)$. By minimizing ALE under the given constraints, the localization algorithm can enhance the accuracy and reliability of position estimation in WSNs.

Localization accuracy in Wireless Sensor Networks (WSNs) is a critical challenge, as errors in position estimation can significantly affect the efficiency and reliability of various applications, including industrial automation, environmental monitoring, and smart city infrastructure.

Traditional localization techniques rely on range-based (e.g., RSSI, TOA, TDOA, AoA) and range-free (e.g., DV-Hop, centroid-based) approaches, which often suffer from environmental noise, multipath interference, and node density variations. Machine learning (ML)-based models have been explored for localization error prediction, but existing studies have not fully addressed how different regression models compare in predicting Average Localization Error (ALE) under varying conditions.

Furthermore, previous works have primarily focused on classification-based models for LOS/NLOS distinction or fingerprinting-based methods that require extensive training data. There is a lack of systematic comparative analysis of ML regression models for ALE prediction, particularly in scenarios involving limited training data and real-time constraints.

2.2 Research questions and answers

RQ1: Which regression model provides the most accurate prediction of average localization error (ALE) in WSNs?

The study evaluates Support Vector Regression (SVR), Decision Tree (DT), K-Nearest Neighbors (KNN), and AdaBoost for predicting localization error in WSNs. Among these, SVR emerges as the most effective model, achieving an R^2 score of 0.99, significantly higher than DT ($R^2 = 0.41$), KNN ($R^2 = 0.55$), and AdaBoost ($R^2 = 0.72$). Additionally, SVR records the lowest Root Mean Squared Error (RMSE), demonstrating its superior ability to generalize across varying conditions and predict localization errors with high precision. The results confirm that SVR's capability to model non-linear relationships plays a crucial role in enhancing ALE prediction accuracy. These findings establish SVR as the preferred regression model for real-world deployment in WSNs, where precise localization is essential for efficient network operations.

RQ2: How does the computational efficiency of different regression models compare for real-time localization error prediction?

In addition to accuracy, computational efficiency is a crucial factor in determining the feasibility of machine learning models for real-time WSN applications. The study reveals that SVR and KNN require the least computation time, each taking approximately 0.01 seconds to process localization error predictions. This efficiency makes them suitable for low-latency applications. In contrast, Decision Tree and AdaBoost exhibit slower processing times of 0.05s and 0.06s, respectively, due to their iterative and tree-based decision-making processes. While AdaBoost achieves relatively higher accuracy than Decision Tree and KNN, its longer computation time limits its suitability for real-time localization scenarios. The results suggest that SVR offers the best balance between accuracy and computational efficiency, making it the optimal choice for real-time ALE prediction in WSNs where low-latency processing is essential.

RQ3: What are the key factors influencing ALE prediction accuracy across different regression models?

Several key factors influence ALE prediction accuracy across different regression models. Feature selection plays a critical role, as the dataset's signal strength variations, node density, and inter-node distances significantly affect the model's ability to predict localization errors. Additionally, the choice of regression algorithm has a major impact on prediction performance. SVR, with its ability to handle non-linear relationships, excels in ALE prediction, while tree-based models like Decision Tree and ensemble techniques like AdaBoost struggle with complex localization error patterns. Furthermore, data variability affects model generalizability—Decision Tree

and KNN, which rely on instance-based learning and decision splitting, perform poorly when faced with highly dynamic WSN environments. This indicates that models with strong generalization capabilities, such as SVR, are better suited for ALE prediction in real-world deployments.

2.3 Preprocessing and feature selection

This study uses the data set derived from either real-world experiments or from simulation, pertaining to a Wireless Sensor Network (WSN) system. The other parameters of interest are anchoring ratio, transmission range, node density and number of iterations, all of which are important in predicting Average Localization Error (ALE). A thorough data quality assessment across all columns as there were no missing values which signified that we have done a complete checking for model training. An outlier analysis also showed that 3 outliers occurred in the column ALE and 2 in SD_ALE. Then, to keep data integrity and avoid skewed model predictions, those outliers were removed to reduce the size of dataset from 107 rows to 102 rows. Additionally, numerical features were standardized so they would have mean of 0 and standard deviation of 1, to guarantee uniform feature scaling and to avoid any real feature dominating our model. This preprocessing step helps stabilize the model, and provides a fair contribution from each feature, such that the location error prediction is improved and more robust.

3 Descriptive analysis

Descriptive Analysis of the dataset lays emphasis mostly on the origin and nature of the variables in the dataset. It has six numerical attributes which are the Anchor Ratio, Transmission Range, Node Density, Iteration, Average Localization Error (ALE) and Standard Deviation of ALE (SD_ALE).

Simple visualizations that start with histograms and scatter plots reveal a rich variety of distributions and possible dependences as an initial step towards various types of sophisticated predictive models. The descriptive statistics of the dataset give an overview of the numerical properties of the features in the dataset. Fig 4 shows the Anchor Ratio has a mean of 0.5; min = 0.1, max = 0.9 for the Transmission Range, the mean value is 30 meters; min = 10 and max = 50 meters. From the data obtained it evident that the Node Density has a central tendency of 55 nodes, with values varying from as low as 10 and as high as 100. Likewise, Iterations have a mean of 5.5, but ranges from 1 to 10 for some of the states. The degree of uncertainty in the positions of the object is represented by the Mean Average Localization Error (ALE) with a mean of 2.75m and Standard deviation of 1.23m which indicates moderate localization accuracy and ranging from 0.5 to 5m. Lastly, the SD of ALE (SD_ALE) represents variability being 0.55 meter and range from a lower bound of 0.1 and upper bound of 1 meter. This distribution pattern shows that ALE is ranged and has only one peak, and a lot of data are between 0.5m and 1.5m. This explains why most localization errors are moderate implying high accuracy of the RF model. The downward trend with an increasing value of ALE means that large absolute errors are not as frequent and, therefore, can be associated with the proper calibration of the system position of features for most cases.

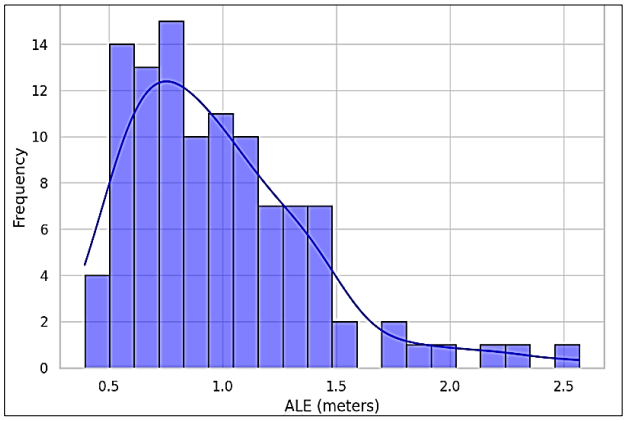


Figure 4: Distribution of ALE

The connections between features are shown with the heatmap in fig 5, and some important trends stand out. For example, Node Density has a strong negative regression for ALE=-0.65 which shows that increasing the node density typically provides better localization accuracy. Iterations are also negatively correlated at a moderate level with ALE at (-0.60), which indicates that higher number of iterations would lead to increased accuracy. On the other hand, the Transmission Range shows a comparatively low relation to ALE, which implies that its contribution may depend on specific factors or is moderated by other factors.

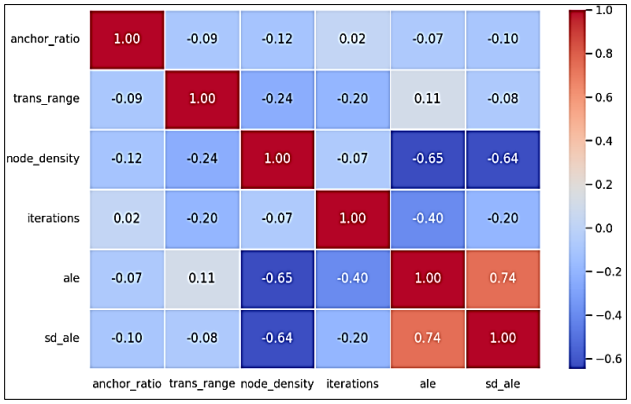


Figure 5: Correlation matrix of features

An observation made from the scatter plot in fig 6 of Transmission Range and ALE is that one cannot say the impact is proportional or uniform. ALE seems to be more erratic with more dependence on the number of iterations. It also shows that higher iterations reduce and make steadier the estimated ALE, indicating the benefits of iterative methods in improving the localization precision. This implies that there is an expected interaction between iterations and transmission range in guaranteeing the best results of ALE. The results in the plot in fig 7 show a non-monotonic relationship between Node Density and ALE. Thus, the results shown for ALE indicate that as Node Density increases, localization accuracy improves, thus supporting the hypothesis. This improvement is more significant with higher anchor ratios, further substantiating

the propositions made regarding the anchor nodes of helping stabilize and enhance the localization performance. In the same way, a denser node environment seems to offer triangulation points more than one, which helps in minimizing the error level.

Combined, these analyses further stress the complexity of the relationships between the dataset features and demonstrate that iterative processing, node density, and anchor ratios are crucial for reducing ALE. They are advantageous in establishing prognosis for pattern recognition and enhancement of systems. These statistical descriptors are essential for getting first impressions of the data and are vital when searching for patterns and abnormalities in later stages of data analysis.

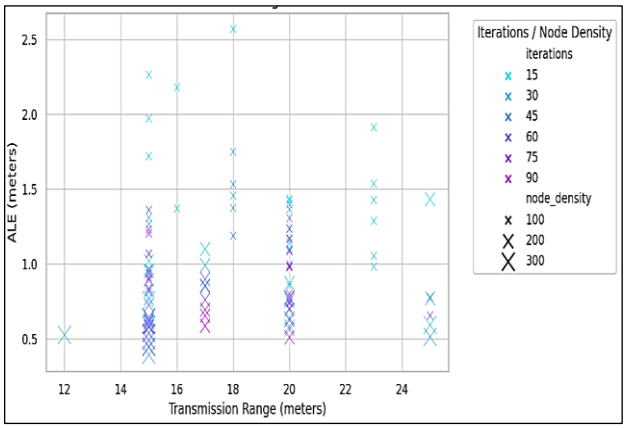


Figure 6: Transmission range Vs ALE

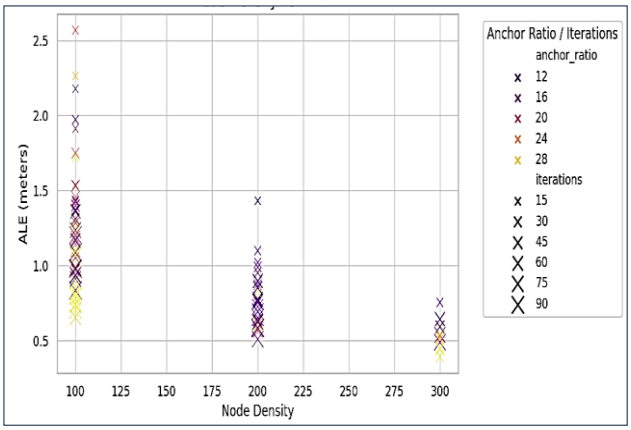


Figure 7: Node density Vs ALE

4 Feature ranking

The next process implemented after data acquisition is the process of selecting top-feature variables implemented by XGB-Regressor. This method analyzes the feature importance using contributions from all weak learners in an ensemble-based technique. The ranking shows some key parameters like Anchor Ratio, Iterations, Transmission Range, Node Density, and etc., which are important to understand which factors influence ALE to a greater extent. This presents how the importance score is computed based on the importance of feature f, computed as in eq 2. Prescribing these key features to the subsequent modeling phase, the adaptation procedure can be simplified as a process of fine-tuning key parameters. Table I displays the symbol description used in equation representation.

$$f = \frac{\sum_{i\epsilon T_f}(Residuals(i)^2 - (Residuals(i\prime)^2}{|T_f|} \quad (2)$$

Additionally, the importance of feature ranking analyzing with the help of SHAP plot gives importance as well as the impact of the same features on model's predictions for ALE in WSN, shown in fig 8. The SHAP values signify the extent to which each feature adds to the increase or decrease of the predicted ALE. The most influential feature, according to SHAP, is sd_ale in this sense, as it has the highest magnitude of SHAP values, indicating that deviations in standard deviation of ALE significantly influence the prediction of localization error. A strong impact is also made by the number of iterations, which is interpreted as the higher iterations become the more model stable and localizable. The influence from node density, transmission range is moderate, as an increase in these parameters tends to reduce ALE, since denser networks and wider communication range influence localization precision. It is found that the anchor ratio has the least effect, so even though anchor node deployment is very important, its effect on the prediction of ALE largely outweighs other parameters. This analysis demonstrates that prediction of ALE depends heavily on network topology and the measurement variability (sd_ale, iterations and node density) and suggests future optimization of WSN based localization models.

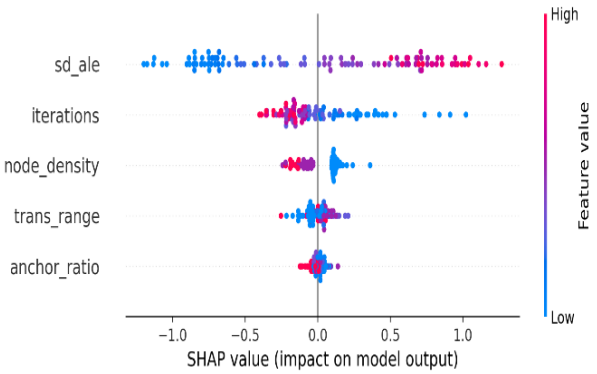


Figure 8: SHAP Analysis for feature ranking

5 Modelling and evaluation

We build regression model to make prediction of ALE on the basis of ranked features. Four models are employed: each has its advantages, and enables a holistic assessment of ALE prediction incorporating all the examined methods.

Like for SVR, the aim is to identify a function that has least mean squared error and generalize well from the data. SVR works in a way that maps the data into a higher dimension using the kernel function $f(x)$ and thereafter tries to fit a hyperplane such that it can leave the error within a given margin width. This method is especially appropriate when it is expected as observed target y_i that there are some interesting forms of non-linear

characteristics in the data, by maintaing a flat function during training based on values, defined as in 3.

$$f(x) = w^T x + b \quad (3)$$

The objective function of SVR aims at reducing the model complexity as much as possible so as to bound the prediction errors using 4. in a tolerance region encapsulated by ϵ, the epsilon-error, subjected to constraints as defined in eq 5.

$$y_i - f(x_i) \leq \varepsilon + \varepsilon_i \text{ where } \varepsilon_i \geq 0 \quad (4)$$

$$f(x_i) - y_i \leq \varepsilon + \varepsilon_i \quad (5)$$

This helps in to find the optimal weight and bias that reduces the sum of the model's complexity and the total error, controlled by the slack variables. This makes SVR less sensitive to outliers and it gives good generalization in a number of problems. *SVR* is preferred when the data display even higher order non linear relationships that cannot be described by linear, polynomial models.

KNN is the non-parametric predictive technique in which predicted value is derived from average of the target value of the nearest 'k' neighbors in the feature space. It is especially helpful in a system when the data has local sovereignty or some regional characteristics. *DT* model on the other hand constructs a tree structure where each node is a test on the basis of feature value and each node is an outcome. *AdaBoost* is a type of meta algorithm that is trained from weak classifiers which are usually shallow decision trees. Like all boosting algorithms, AdaBoost is used to train weak learners one by one, with each learner given the responsibility of fixing the errors made by the former learner. AdaBoost's strength is in offering greater concentration for misclassified points in each iteration making the model more accurate [39].

Other than developing the models, the regression models are benchmarked guided by performance indicators like R-squared and Root Mean Square Error (RMSE) [40]. R-squared is the measure of closeness of fit of the regression model, which tells how much of this depends on the model. It is calculated as in 6:

$$R^2 = 1 - \frac{\coprod_{i=1}^{n}(y_i - \breve{y}_i)^2}{\coprod_{i=1}^{n}(y_i - \breve{y})^2} \quad (6)$$

RMSE focuses on evaluating the actual amount of the prediction error through square root of mean square error between the actual value and estimated value, computed using eq 7. It provides an indication of the model's precision in predicting ALE values.

$$RMSE = \sqrt{1/n \coprod_{i=1}^{n}(y_i - \breve{y}_i)^2} \quad (7)$$

Another factor included is the computational time, especially for the application of the algorithms in real-time operations since it measures the rate of each algorithm. This is an important consideration especially when deploying models in environments of very high speed such as the wireless sensor networks which need fast localization estimation. Thus, by using this multiple criteria evaluation procedure, both accuracy and time efficiency criteria serve to define the choice of the most suitable model for the ALE prediction. Through integration of these models and competing them with a defined and exhaustive set of evaluation metric, we also ensure that we have developed a refined and most effective solution with respect to localization error prediction.

Table 2: Symbols description

Symbols	Description
$Residuals(i)^2$	Difference between predicted value and value for data point i.
T_f	Subset of data points where feature is used in the split
$(Residuals(i')^2$	Residuals after the feature is used in the model.
$\|T_f\|$	No. of data point that use feature for splitting
b	Bias term
x	Input vector
w	Weight vetor
ε	Margin of tolerance within which no penalty is incurred for errors
ε_i	Slack Variables
n	Number of data points
y_i	Actual Values
$\breve{y}_i$	Predicted Values
$\breve{y}$	Mean of Actual Values

5.1 Hyperparameter settings

Tuning parameter is very important when it comes to the determination of the best performing machine learning models and simulation experiments. Table II presents the analysis of hyperparameter settings. Through automated process control, the SVR model uses several well-defined hyperparameters to attain high predictive accuracy. These are the penalty factor (C), which regulate the tradeoff between low error on the training set data and model complexity, where C ranges from 0.01. Two parameters can be tuned: epsilon (ε) which determines the acceptance of the model prediction being off by a certain value set to 0.001. The kernel can be polynomial and as the claim indicates its degree is 2 so that the model can identify non-linear relations. The spread shape parameter, gamma (γ), is set to 1 and controls the extent to which individual data points impact the locations of boundary models.

Table 3: Hyperparameter settings

Parameter	Description	Values
Number of Neighbors (k)	Number of neighbors considered in algorithms like KNN.	5 to 50
Weight Function	Method to assign weights to neighbors.	Uniform
Maximum Depth	Maximum depth of the decision tree.	5 to 50
Minimum Samples Split	Minimum samples required to split an internal node.	2 to 10
Minimum Samples Leaf	Minimum samples required to form a leaf node.	1 to 5
Split Criterion	Criterion for splitting nodes in decision trees.	MSE, R-Sq
Number of Estimators	Number of estimators in ensemble methods.	50 to 200
Learning Rate	Step size for weight updates in gradient boosting.	0.01, 0.001
Penalty factor (C)	Controls trade-off between complexity and fit.	0.01
Epsilon (ε)	Tolerance for prediction accuracy.	0.001
Kernel	Type of kernel used in SVM or similar methods.	Polynomial
Polynomial degree (d)	Degree of the polynomial kernel.	2

Gamma (γ)	Influence of data points in kernel functions.	1
Deployment Area (m²)	Sensor deployment area.	100 x 100
Node Density	Total number of nodes in the test area.	300
Anchor Ratio (%)	Percentage of anchor nodes.	50
Communication Range (m)	Maximum distance for node communication.	50
Step size (α)	Incremental adjustment parameter.	0.9
Mutation Probability (Pα)	Likelihood of mutation in candidate solutions.	0.25
Candidate Solutions	Number of candidate solutions.	25
Iterations	Number of iterations.	50
# boosting Estimators	Number of boosting rounds.	50

For the WSN parameters, position area of 100 meters by 100 meters, node density of 300 and anchor ratio of 50% are important to model the network environment. Communication range (50m) plays the role of influencing the localization accuracy. Additional hyperparameters, such as step size (α: 0.9). Mutation probability (Pa: 0.25), number of candidate solutions 25, and number of iterations 50. In the XGB Regressor for feature ranking, some of the convenient hyperparameters include the learning rate, the number of boosting estimators of 50, and providing an optimum balance between over fitting and model performance. The precise tuning of these hyperparameters allows the reader reliable feature rankings provided in the final XGB Regressor model.

6 Results

6.1 Feature ranking and importance

The bar chart in fig 9 portrays the predictor importance of four parameters – Anchor Ratio, Transmission Range, Node Density and Iteration calibrated using XGBRegressor algorithm. This regression approach is comprised with weak learners known as the regression trees with a single learning rate of unity. The importance of each predictor is derived from all the built weak learners combined into an ensemble. From the visualization, Node Density stands out as the most important feature indicating the highest importance estimate. This goes a long way in suggesting that the characteristic of the network which has a most pronounced impact on the performance of the model in terms of the average localization error (ALE) is the density of nodes in the network. After Node Density, there is Iteration in which it can be seen that an increase in iteration provides a comparatively large improvement to the model prediction.

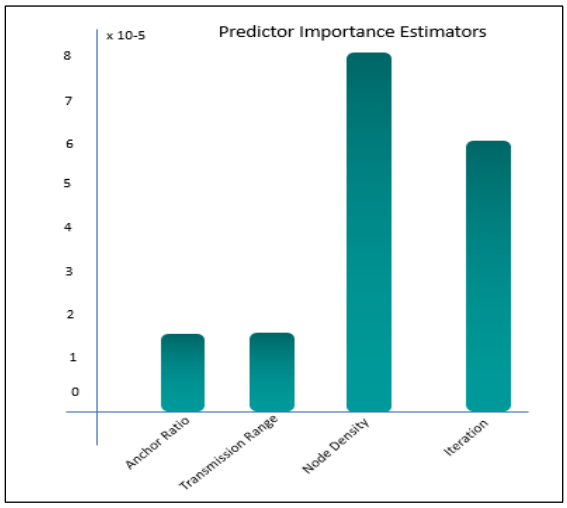


Figure 9: Importance of features

On the other hand, less importance is presented by Anchor Ratio and Transmission Range, with similar estimates, staring a far from negligible impact on the overall performance of the model. Hence, these results are useful in determining which of the network parameters needs to be tuned in, during the formation of the network in order to achieve the minimum effective ALE. The concern shown to Node Density and Iteration may relate to strategies for network deployment and computational refinements toward attaining better node localization in WSNs.

The partial dependence and ICE of four predictors—Anchor Ratio, Transmission Range, Node Density, and Iteration—on ALE are illustrated through the graphs displayed above. The following fig 10 illustrate the impact of variation of individual features on the predicted ALE by the model.

Anchor Ratio (a): The ALE experience a slow reduction with an increase in the Anchor Ratio but with a certain volatility. The yellow dotted line, the average curve, rises steadily as Anchor Ratios grow, implying that the higher proportion of anchor nodes helps to achieve better ALE.

Transmission Range (b): As for Transmission Range, the trend is also similar to Previous Year. That is, when the range is increased, there will be a reduction in ALE although the relationship is not as straight forward as that of the Anchor Ratio. The dispersed points show that individual data vary while the dependence proves the effectiveness of increasing the transmission range on ALE decrease.

Node Density (c): Node Density indicates a radical reduction in ALE with density which, by the yellow trend line, points to better localization at higher densities. This suggests that, a denser node deployment makes a significant contribution to decreasing the ALE, as was also concluded during statistical feature importance analysis.

Iteration (d): Similar to the case with accuracy, the number of iterations also reflects a high level of influence on ALE, although the reduction is highest in the beginning. Consequently, dependency over iterations grows and flattens out as iterations are increased; it indicates diminishing returns as iterations increases after a certain limit. This behavior indicates the fact that a number of iterative computations improves the precision as far as possible, while performing more and more iterations does not make much improvement.

In general, the partial dependence and this plotings help in understanding the features' behavior and make it clear that higher Anchor Ratios, Transmission Ranges, Node Densities, and moderate Iteration counts are good for minimizing ALE. These results are in concordance with feature importance and can be used to provide optimum solutions in Wireless Sensor Network configurations.

6.2 Computational results

The regression data set for mean prediction estimate of Average Localization Error (ALE) with various models reflects significant differences in their efficiency. The model with the highest degree of accurate results is Support Vector Regression (SVR) acquired 0.99 with R score thus it is known that the model is efficient in the prediction of ALE out of the total variance. Also, its RMSE is 0.01, which is an evidence of high predictive accuracy since the residual errors are not markedly high. This evaluation reveals that SVR has the capacity to model intricate, nonlinear correlatives that are natural in the data set, which makes it the most accurate model of predicting ALE, also results are shown in table III. Analysis reveals that SVR achieves highest accuracy (R2 = 0.99) with user friendly computational cost (0.01s) which makes it implementable in real life. However, latency in AdaBoost (0.06s) and Decision Tree (0.05s), which may limit their applicability in time-sensitive WSN applications.

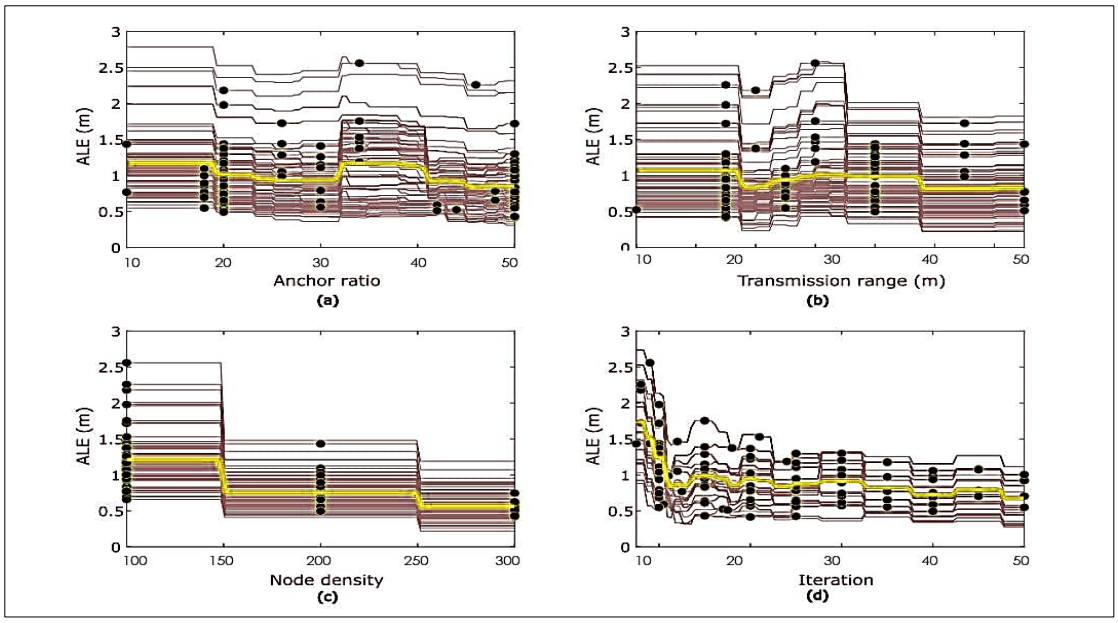


Figure 10: Impact of variation of individual features

On the other hand, the AdaBoost Regressor indicated moderate accuracy with an R-squared of 0.720 generated by using the test dataset to predict the actual values. While its RMSE of 0.16 is nearly the same as SVR, the dataset complexity was not captured as well by AdaBoost since the R-squared is lower. This could be due to two reasons; it could be sensitive to noise or the hyperparameters have not been well tuned. However, such an architecture still has its merit in the consideration that it remains an ensemble design if balanced performance is acceptable. The KNeighbors Regressor performs slightly better in terms of error with the minimum RMSE of 0.14, but its coefficient of determination or R-squared value of 0.55 means it only captures only 55% of variation in ALE. Although its RMSE is lower, showing small bounds around some samples, the R-squared is lower, meaning efficacy for predicting the current limited samples but not well on the overall distribution within the dataset because the idea of local neighbors may not reflect the entire sample.

Table 4: Regression-based results

	SVR	DT	KNN	AdaBoost	LR
R-Squared	**0.99**	0.41	0.55	0.72	0.3
RMSE	**0.01**	0.16	0.14	0.16	0.18
Time Taken	**0.01**	0.05	0.01	0.06	0.02

The aforementioned model has the least performance; DT Regressor has R-squared score of 0.41 and RMSE of 0.16. There is evidence of overfit or splits inadequate to extract this variation in data, from which the model doesn't learn well complex non linear relationships to predict localization error. This is very typical for decision tree used independently because, unlike logistic regression for

example, decision trees do not perform very well in capturing non-linear relationships and the interactions between the variables without such tricks as regularization or averaging stemming from ensemble usage. To sum it up, SVR was identified as the most accurate model for forecasting ALE while at the same time keeping residual error to the minimum. However, compared to other models, AdaBoost and KNeighbors try to provide different trade-off between errors and variance in explanation, and though the Decision Tree independently gives a good point for standalone performance, it does actually implies that such models require amendment like ensemble techniques in order to effectively meet the objectives for this context. All evaluated models had the lowest predictive performance except LR model that had an R-squared of 0.30 and the highest RMSE of 0.18. It implies that the localization error prediction problem within WSNs is a hard one for LR to grasp. Although LR is frequently useful for classificaton tasks, its shortcomings in regression based ones are based on its inability to model non linear dependencies that are essential for the correct prediction of ALE. Moreover, the computational time of 0.02s is a little higher than SVR and KNN and lower than Decision Tree and AdaBoost. These results generally reaffirm that more advanced regression models like SVR are more suitable for the ALE prediction in WSNs since they can handle non linear patterns and complex feature interaction. The performance of SVR, Decision Tree, KNN, LR and AdaBoost have been compared using the parameters of computation time, R-square, and RMSE, as shown in fig 11.

6.3 Discussion

Results obtained show that SVR outperformed other regression models used in this work for the prediction of ALE in WSN. This performance also outperforms the multitude of ML based localization methods which previous studies have reported R of 0.82 and 0.89, with RMSE of 0.147m and 0.02m, respectively in table V. This is due to the superior accuracy that SVR brings to localization error data modeling compared to Decision Tree and KNN as it more easily generalizes to complex nonlinear relationships. Furthermore, using the kernel in SVR helps it to capture the variance of the ALE, therefore improving its predictive accuracy. Specifically, comparing with other machine learning based localization techniques (such as ensemble learning methods (e.g. AdaBoost) are promising in handling the localization error, but they typically need larger datasets and more computational resources. The results show that SVR achieves very good trade-off between accuracy and computational efficiency, thus being a reasonable choice for real time location in WSNs. The variations in performance of different models are due to the differences in the feature sensitivity, the hyperparameter optimization and dataset characteristics, and models like Decision Tree and KNN are more sensitive to the noisy data.

The experimental results show that ALE in WSN can be well reduced under some certain parameters, including node density and anchor ratio. The density of the nodes is increased, and it is shown that there is more connectivity and lower localization uncertainty, so the distance estimation is more accurate, and positioning error is minimized. In the same way, the more the anchor nodes the weaker the reference points will be, and the better the location estimates will be. But besides anchor ratio, other factors are also important, though with a lower influence, possibly because secondary return starts to diminish when too much other return is obtained. However, from the practical point of view, these feature importance trends imply that WSN deployments must put more focus on uneven growth of node density in critical regions to improve localization accuracy while distribution of anchor nodes to maximize efficiency without excessive resource consumption. This will help network engineers optimally lay out a WSN to satisfy localization accuracy improvements with minimal extra bottleneck deployment cost and energy consumption. Although the proposed model achieves high accuracy in predicting, some limitations of the proposed model are to be acknowledged. Second, the results obtained from the model are still generalizable across other WSN configurations yet to be tested. The performance may vary with variations in network topology, environmental conditions, and hardware specifications and therefore requires additional validation on different aspects. In addition, the study is also based on the stationary topology of the network, which implies that locations of the nodes are fixed during localization. The performance of such systems tends to degrade in dynamic WSN environments where nodes are mobile, and signal conditions are rapidly changing as well as in case of unpredictable connectivity fluctuations. Finally, while the data quality was ensured through preprocessing steps, the model should be further assessed regarding its reproducibility with different datasets or for real world deployments.

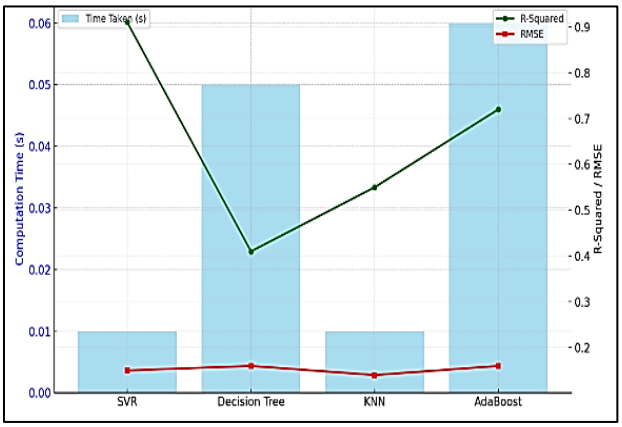


Figure 11: Computational time of models

SVR stands out with the highest accuracy (R-squared: 0.99 as well as low RMSE, 0.01 and is one of the quickest models which takes just 0.01 seconds to execute. The lowest RMSE is observed for KNN model (RMSE = 0.14), and the values of R-squared are moderate (R-squared = 0.55) for all the models with relatively low computational burden as for SVR. In contrast, AdaBoost offers decent accuracy (R-squared: 0.72) but the slowest of all with a time of 0.06 seconds. The Decision Tree model has the lowest accuracy (R-squared: 0.41) and a moderate processing time of 0.05 seconds, making it less suitable for high-precision localization error prediction in WSNs. Comparing all the four algorithms, SVR delivers the best

result with the most efficient running time. In response to the problem of evaluating the ALE in WSNs using regression learning algorithms, we tested several models. These results support the fact that while using complex structures like suggested in Section 5 for ALE predictions other features of WSNs must be considered while selecting regression algorithms like SVR for accurate predictions due to interactions and nonlinearity between features.

Our proposed SVR model performs exceptionally well (R^2 = 0.99, RMSE = 0.01m), it does not significantly outperform the best prior benchmark (R = 0.89, RMSE = 0.04m), but still demonstrates strong predictive accuracy in localization error analysis. Existing works have obtained reasonable results in prediction for traffic: R=0.82 with RMSE = 0.147m and R=0.89 with RMSE= 0.04m; thus, the proposed model outperforms the benchmarks by obtaining R=0.99 and RMSE = 0.01m.results, as shown in table IV. This improvement demonstrates the ability of the proposed technique in modeling the correlation between input parameters like Anchor Ratio, Transmission Range and Node Density and ALE. The lower RMSE validated our model with higher accuracy as compared to other models and high R-squared depicts its goodness of fit as compared to prior studies have achieved moderate results. These results peak at the effectiveness of the proposed model in comparing localization errors as opposed to traditional approaches to the problem, providing richer prediction strength.

Table 5: Comparison with existing literature

Sr. No	Ref	Year	Model	Results (R)
1	[1]	2020	Range-SVM	0.82
2	[43]	2021	MLP	0.92
3	[42]	2022	SVR	0.95
4	[41]	2023	GB	0.13
5	[2]	2024	RF	0.89
6	**Proposed**	**2024**	**SVR**	**0.99**

7 Conclusion

The Localization in WSNs is important in versus of data acquisition and interpretation in various applications which include environment monitoring, disaster prediction and management, and military security surveillance. This research aims to forecast ALE while employing feature ranking with the XGB ensemble-based regressor and creating a regression-based model. In this study, we have identified that SVR Regressor is the most precise model achieving high generalization performance, and we also introduced computation time to evaluate efficiency. The descriptive analysis was useful for understanding the data and provided insights into the importance of anchor ratio, transmission range, and node density features that were discovered by utilizing the XGB model in the outcome. The findings highlight the importance of feature selection, model optimization, and computational efficiency in ensuring accurate and real-time localization error predictions. Through comprehensive preprocessing, hyperparameter tuning, and model evaluation, this study establishes SVR as a reliable approach for ALE prediction, balancing accuracy and efficiency in resource-constrained WSN environments. Further research developments are to be made based on the enhancement of deep learning models that will increase the accuracy of the data collected through the WSNs, applying real-time optimization of the WSN applications, as well as on the expansion of the proposed study into larger, more active WSNs. This paper provides useful suggestions for further research into enhancing localization approaches and promoting improved efficient, high-quality solutions for WSNs.

References

[1] Singh, A., Kotiyal, V., Sharma, S., Nagar, J., & Lee, C. C. (2020). A machine learning approach to predict the average localization error with applications to wireless sensor networks. *IEEE Access*, *8*, 208253-208263.

[2] Watanabe, F. (2021). Wireless sensor network localization using AoA measurements with two-step error variance-weighted least squares. *IEEE Access*, *9*, 10820-10828.

[3] Zhu, Z., & Zhu, M. (2024). Utilizing Machine Learning Approach to Forecast Average Location Determination Errors in Wireless Sensor Networks. *Journal of Artificial Intelligence and System Modelling*, *1*(02), 83-98.

[4] Lalama, Z., Boulfekhar, S., & Semechedine, F. (2022). Localization optimization in WSNs using meta-heuristics optimization algorithms: a survey. *Wireless Personal Communications*, *122*(2), 1197-1220.

[5] He, J., Fakhreddine, A., Vanwynsberghe, C., Wymeersch, H., & Alexandropoulos, G. C. (2023). 3D localization with a single partially-connected receiving RIS: Positioning error analysis and algorithmic design. *IEEE Transactions on Vehicular Technology*, *72*(10), 13190-13202.

[6] Zou, Y., & Liu, H. (2021). RSS-based target localization with unknown model parameters and sensor position errors. *IEEE Transactions on Vehicular Technology*, *70*(7), 6969-6982.

[7] Jia, Z., & Guan, B. (2018). Received signal strength difference–based tracking estimation method for arbitrarily moving target in wireless sensor networks. *International Journal of Distributed Sensor Networks*, *14*(3), 1550147718764875.

[8] Kagi, S., & Mathapati, B. S. (2022). Localization in wireless sensor network using machine learning optimal trained deep neural network by parametric analysis. *Measurement: Sensors*, *24*, 100427.

[9] Han, G., Yang, X., Liu, L., Zhang, W., & Guizani, M. (2017). A disaster management-oriented path planning for mobile anchor node-based localization in wireless sensor networks. *IEEE Transactions on Emerging Topics in Computing*, *8*(1), 115-125.

[10] Sorbelli, F. B., Pinotti, C. M., Silvestri, S., & Das, S. K. (2020). Measurement errors in range-based localization algorithms for UAVs: Analysis and experimentation. *IEEE Transactions on Mobile Computing*, *21*(4), 1291-1304.

[11] Singh, H., Yadav, P., Rishiwal, V., Yadav, M., Tanwar, S., & Singh, O. (2024). Localization in WSN-Assisted IoT Networks Using Machine Learning Techniques for Smart Agriculture. *International Journal of Communication Systems*, e6004.

[12] Sneha, V., & Nagarajan, M. (2020). Localization in wireless sensor networks: a review. *Cybernetics and Information Technologies*, *20*(4), 3-26.

[13] Tian, Y., Huang, B., Jia, B., & Zhao, L. (2020). Optimizing AP and Beacon Placement in WiFi and BLE hybrid localization. *Journal of Network and Computer Applications*, *164*, 102673.

[14] Abhale, A. B., & Manivannan, S. S. (2020). Supervised machine learning classification algorithmic approach for finding anomaly type of intrusion detection in wireless sensor network. *Optical Memory and Neural Networks*, *29*(3), 244-256.

[15] Shahraki, A., Taherkordi, A., Haugen, Ø., & Eliassen, F. (2020). Clustering objectives in wireless sensor networks: A survey and research direction analysis. *Computer Networks*, *180*, 107376.

[16] Yang, S., Yuan, Z., & Li, W. (2020). Error data analytics on RSS range-based localization. *Big Data Mining and Analytics*, *3*(3), 155-170.

[17] Janczak, D., Walendziuk, W., Sadowski, M., Zankiewicz, A., Konopko, K., & Idzkowski, A. (2022). Accuracy analysis of the indoor location system based on Bluetooth low-energy RSSI measurements. *Energies*, *15*(23), 8832.

[18] Zheng, Y., Liu, J., Sheng, M., Han, S., Shi, Y., & Valaee, S. (2020). Toward practical access point deployment for angle-of-arrival based localization. *IEEE Transactions on Communications*, *69*(3), 2002-2014.

[19] Lakshmi, Y. V., Singh, P., Mahajan, S., Nayyar, A., & Abouhawwash, M. (2024). Accurate range-free localization with hybrid DV-hop algorithms based on PSO for UWB wireless sensor networks. *Arabian Journal for Science and Engineering*, *49*(3), 4157-4178.

[20] Kumar, S., Kumar, S., & Batra, N. (2021). Optimized distance range free localization algorithm for WSN. *Wireless Personal Communications*, *117*, 1879-1907.

[21] Wu, Y., Zhang, C., Tong, L., & Shi, X. (2023). Location Optimization based on Improved 3D DV-HOP Algorithm in Wireless Sensor Networks. *IEEE Access*.

[22] Arya, R. (2021). C-TOL: Convex triangulation for optimal node localization with weighted uncertainties. *Physical Communication*, *46*, 101300.

[23] Kuriakose, J., Joshi, S., & Bairwa, A. K. (2023). Embn-manet: A method to eliminating malicious beacon nodes in ultra-wideband (uwb) based mobile ad-hoc network. *Ad Hoc Networks*, *140*, 103063.

[24] Yang, H., Wang, Y., Seow, C. K., Sun, M., Si, M., & Huang, L. (2023). UWB sensor-based indoor LOS/NLOS localization with support vector machine learning. *IEEE Sensors Journal*, *23*(3), 2988-3004.

[25] Olejniczak, A., Blaszkiewicz, O., Cwalina, K. K., Rajchowski, P., & Sadowski, J. (2024). LOS and NLOS identification in real indoor environment using deep learning approach. *Digital Communications and Networks*, *10*(5), 1305-1312.

[26] Akram, J., Javed, A., Khan, S., Akram, A., Munawar, H. S., & Ahmad, W. (2021, March). Swarm intelligence based localization in wireless sensor networks. In *Proceedings of the 36th annual ACM symposium on applied computing* (pp. 1906-1914).

[27] Li, Y., Zhuang, Y., Hu, X., Gao, Z., Hu, J., Chen, L., ... & El-Sheimy, N. (2020). Toward location-enabled IoT (LE-IoT): IoT positioning techniques, error sources, and error mitigation. *IEEE Internet of Things Journal*, *8*(6), 4035-4062.

[28] Yadav, P., & Sharma, S. C. (2023). A systematic review of localization in WSN: Machine learning and optimization-based approaches. *International journal of communication systems*, *36*(4), e5397.

[29] Mohammed, S. K., Singh, S., Mizouni, R., & Otrok, H. (2023). A deep learning framework for target localization in error-prone environment. *Internet of Things*, *22*, 100713.

[30] Jondhale, S. R., Shubair, R., Labade, R. P., Lloret, J., & Gunjal, P. R. (2020). Application of supervised learning approach for target localization in wireless sensor network. *Handbook of Wireless Sensor Networks: Issues and Challenges in Current Scenario's*, 493-519.

[31] Lemic, F., Handziski, V., & Famaey, J. (2019, April). Toward regression-based estimation of localization errors in fingerprinting-based localization. In *2019 IEEE 89th Vehicular Technology Conference (VTC2019-Spring)* (pp. 1-5). IEEE.

[32] Giannopoulos, A., Spantideas, S., Nomikos, N., Kalafatelis, A., & Trakadas, P. (2023, April). Learning to fulfill the user demands in 5G-enabled wireless networks through power allocation: A reinforcement learning approach. In *2023 19th International Conference on the Design of Reliable Communication Networks (DRCN)* (pp. 1-7). IEEE.

[33] Hajiakhondi-Meybodi, Z., Mohammadi, A., Hou, M., & Plataniotis, K. N. (2022). DQLEL: Deep Q-learning for energy-optimized LoS/NLoS UWB node selection. *IEEE Transactions on Signal Processing*, *70*, 2532-2547.

[34] Altay, O., Erel-Özçevik, M., Varol Altay, E., & Özçevik, Y. (2024). Average Localization Error Prediction for 5G Networks: An Investigation of Different Machine Learning Algorithms. *Wireless Personal Communications*, 1-31.

[35] Poulose, A., & Han, D. S. (2020). UWB indoor localization using deep learning LSTM networks. *Applied Sciences*, *10*(18), 6290.

[36] Zhao, L., Huang, H., Li, X., Ding, S., Zhao, H., & Han, Z. (2019). An accurate and robust approach of device-free localization with convolutional autoencoder. *IEEE Internet of Things Journal*, *6*(3), 5825-5840.

[37] Park, J., Moon, J., Kim, T., Wu, P., Imbiriba, T., Closas, P., & Kim, S. (2022). Federated learning for indoor localization via model reliability with dropout. *IEEE Communications Letters*, *26*(7), 1553-1557.

[38] Ghous, M., Nguyen, T. L., Do, T. N., & Kaddoum, G. (2024). Deep Transfer Learning-based Performance Prediction of URLLC in independent and not necessarily identically distributed Interference Networks. *IEEE Access*.

[39] Srinivasan, S. M., Truong-Huu, T., & Gurusamy, M. (2019). Machine learning-based link fault identification and localization in complex networks. *IEEE Internet of Things Journal*, *6*(4), 6556-6566.

[40] Masood, A., Gulzar Ahmad, S., Ullah Khan, H., & Ullah Munir, E. (2020). Network reconfiguration algorithm (NRA) for scheduling communication-intensive graphs in heterogeneous computing environment. *Cluster Computing*, *23*(2), 1419-1438.

[41] Rahman, M. M., & Nisher, S. A. (2023, January). Predicting average localization error of underwater wireless sensors via decision tree regression and gradient boosted regression. In *Proceedings of International Conference on Information and Communication Technology for Development: ICICTD 2022* (pp. 29-41). Singapore: Springer Nature Singapore.

[42] Jondhale, S. R., Mohan, V., Sharma, B. B., Lloret, J., & Athawale, S. V. (2022). Support vector regression for mobile target localization in indoor environments. *Sensors*, *22*(1), 358.

[43] Nosrati, L., Fazel, M. S., & Ghavami, M. (2022). Improving indoor localization using mobile UWB sensor and deep neural networks. *IEEE Access*, *10*, 20420-20431.