Predicting of Error in Application of Wireless Sensor Networks

Kaustubh Raykar
Department of Artificial Intelligence and Machine Learning
Symbiosis Institute of Technology
Maharashtra, Pune
raykarkaustubh@gmail.com

Srinivas Kota
Department of Artificial Intelligence and Machine Learning
Symbiosis Institute of Technology
Maharashtra, Pune
srinivas.kota0334@gmail.com

Prof. Sudhanshu Gonge
*Department of Computer Science ,Symbiosis Institute of Technology,*Maharashtra Pune ,sudhanshu.gonge@sitpune.edu.in

***Abstract*—Wireless Sensor Networks (WSNs) are increasingly vital for data collection from remote areas. This research explores the use of Supervised Machine Learning (SML) algorithms for various WSN applications, including object identification, regression, and prediction. We assess the accuracy, energy efficiency, and computational complexity of common SML methods such as decision trees, support vector machines, and neural networks. Additionally, we discuss feature selection and data preprocessing techniques to enhance SML algorithm performance in WSNs. Our findings demonstrate that SML algorithms significantly enhance WSN application precision and efficiency. This research offers an in-depth analysis of the current state of SML in WSNs, along with recommendations for future investigations.**

***Index Terms*—Wireless Sensor Networks (WSNs), Supervised Machine Learning (SML), Data Processing, Data Analysis, Accuracy Assessment, WSN Applications**

I. Introduction

A WSN is made up of a collection of small and inexpensive sensors that are spatially dispersed over an area to measure physical parameters or monitor habitat conditions, and it has many practical applications such as target tracking, precision agriculture, etc. In most uses, these instruments must guess their coordinates precisely while using minimal resources. Using an integrated Global Positioning technology (GPS) technology, these sensors can rapidly find their locations.

However, due to the size and expense of GPS, it is not possible to incorporate it into all of the sensors. An alternative method is to employ the idea of localisation algorithms,in which several anchor nodes (each with incorporated GPS) aid the unknown nodes in correctly determining their locations. To address various localisation issues, a significant number of localisation methods have been developed. These algorithms are anticipated to be adaptable enough to operate in a wide range of interior and outdoor situations and topologies. There are two types of localization algorithms: range-based algorithms and range-free algorithms.

The position of unknown nodes is calculated using the distance between the anchor and unknown sensor nodes in range- based algorithms. They make use of a variety of measures, including the angle of approach, time of arrival, and the Received Signal Strength Indication. Rangebased algorithms apply trilateration techniques to pinpoint the positions of the sensor nodes and depend on distance metrics like Received Signal Strength (RSS), Time of Flight (ToF), or Angle of Arrival (AoA) to estimate the distances between sensor nodes. Rangefree algorithms determine the locations of sensor nodes by using connection data rather than distance measurements. Because of its capacity to gather data from remote and dangerous areas, Wireless Sensor Networks (WSNs) have developed as an essential technology in recent years.

Clustering is a major approach in WSNs that tries to organise sensor nodes into groups known as clusters in order to minimise energy consumption and increase network lifespan. The clustering process is especially crucial in large-scale WSNs, since each sensor node's energy consumption is a significant component affecting overall network performance. SMOTE, Random Forest, and Grid Search CV are a few examples of machine learning algorithms that have gained popularity for tackling classification and regression issues.

SMOTE is a popular approach for addressing class imbalance in datasets, an issue that arises often in machine learning. A tree based ensemble learning technique called Random Forest mixes different decision trees to increase prediction accuracy. Grid Search CV is a method for hyperparameter tuning that methodically looks for the best set of hyperparameters for a particular machine learning model.

There is no agreement on which of these strategies performs better in certain contexts, despite the fact that they have shown promising results in a variety of applications.

II. Significance

Wireless Sensor Networks (WSNs) are an essential topic of study because they have the potential to change the way we

detect and communicate with our physical environment. WSNs are made up of tiny, low-power sensors that communicate wirelessly with one another and with a central data collecting node, enabling for realtime data gathering and processing.
Work on WSNs innovation and its uses has a great possible effect and can progress a broad range of fields. By enhancing WSNs' functionality, scalability, dependability, security, and privacy, more effective and efficient WSNs may be deployed, which will enhance decision-making. Future improvements in WSN technology may result in more dependable, scalable WSNs that can function in a variety of settings and applications. We can help WSNs advance and open up new implementations and use cases for this innovation by addressing the current issues and looking into new opportunities.

III. Motivation

To solve a crucial problem that has the potential to drastically impact the functionality and dependability of wireless sensor networks, error prediction in their application is essential. Environmental monitoring, industrial automation, and healthcare are just a few of the many applications in that wireless sensor networks (WSNs) are utilized. But these networks are prone to mistakes and malfunctions, which can lead to data loss, network outages, and other issues.
By developing accurate error prediction models, one may anticipate potential issues and take preemptive steps to mitigate their effects. This can improve wireless sensor networks' overall performance and reliability, increasing their usefulness for a number of applications.

IV. Dataset Section

A	B	C	D	E	F
anchor_ratio	trans_range	node_density	iterations	ale	sd_ale
30	15	200	40	0.773545928	0.250555301
15	15	100	70	0.911940683	0.498329391
30	15	100	50	0.814867311	0.255545506
15	20	100	20	1.435332143	0.39460254
30	15	100	40	1.265909048	0.302943301
22	18	100	24	1.457227702	0.463308546
18	23	100	14	1.912673609	0.440208313
10	25	200	15	1.432541304	0.234996306
29	25	100	40	0.777448602	0.351503738
20	20	100	30	1.142195399	0.230192531

Fig. 1. Dataset

The entire dataset, which consisted of 107 occurrences and six characteristics, all of which represented quantitative data, was utilised. There are no missing values in the dataset because it only contains observational data. In order to calculate the Average Localization Error (ALE), just four factors—Anchor ratio, Transmission range, Node density, and Iterations—were used as input variables. The average localization error was then used as an output variable. While the primary goal of our research was to generate localization errors, another attribute (standard deviation value) was left out of the pre-processing stage.
The first column in this dataset is the anchor ratio (AR), which is the number of anchor nodes divided by the total number of sensors in the network. Transmission range in the collection also includes numerical figures in metres that reflect the transmission range of a sensor used to measure transmission speed. The node density characteristic displays how closely the activity nodes are connected. The fourth column indicates iteration, which is the number of times we took sensor readings. The discrepancy between the actual and projected coordinates of unknown nodes is defined as an average localization error.
To summarise our entire dataset, we employed descriptive statistics. Total count, mean, median, and mode are examples of descriptive statistics, as are standard deviation, variance, and lowest and maximum values.

V. Prior Research

Prior studies on WSNs have concentrated on using SML techniques like Regression to enhance the performance and efficiency of WSNs.
These methods have been used with WSNs to address a variety of issues including network lifetime, data reliability, and energy efficiency. To categorise sensor data and find abnormalities in WSNs, several earlier research used SML techniques including Support Vector Machine (SVM) and Naive Bayes. Other studies have grouped sensor nodes and enhanced network performance using clustering algorithms like K-Means, DBSCAN, and hierarchical clustering.
In addition, other studies have concentrated on creating novel SML algorithms that are tailored for WSNs. These algorithms consider the special features of WSNs, such as constrained resources, changing network architecture, and variable data.

VI. Procedure

In order to perform Regression tasks, we intended to compare different machine learning methods. To do this, we obtained and preprocessed data from several sources, which included characteristics and labels. In order to find any missing values, outliers, or other anomalies in the dataset, exploratory data analysis (EDA) was carried out.
We employed a number of visualisation approaches, including scatter plots, heatmaps, and histograms, to display the data. Then, we put three different algorithms into practice—SMOTE, Random Forest, and Grid Search CV—and assessed how well they performed. In order to perform regression tasks, we intended to compare different machine learning methods.
To do this, we obtained and preprocessed data from several sources, which included characteristics and labels. In order to find any missing values, outliers, or other anomalies in the dataset, exploratory data analysis (EDA) was carried out. We employed a number of visualisation approaches, including scatter plots, heatmaps, and histograms, to display the data. Then, we put three different algorithms into practice—SMOTE, Random Forest, and Grid Search CV—and assessed how well they performed.

On a dataset with unbalanced classes, we used SMOTE, Random Forest, and Grid Search CV performance in this study. To analysis our model we used RMSE and MSE value. By examining the data, we hope to shed light on the relative advantages and disadvantages of various approaches and make suggestions on how best to apply them to machine learning applications.

VII. Literature Survey

A. A Machine Learning Approach to Predict the Average Localization Error With Applications to Wireless Sensor Networks

- **Authors:** Abhilash Singh, Vaibhav Kotiyal, Sandeep Sharma, Jaiprakash Nagar, Cheng-Chi Lee (2020)
- **Objective:** Proposing a machine learning-based approach for predicting the average localization error in Wireless Sensor Networks (WSNs) and exploring its applications for enhancing localization accuracy in WSNs.
- **Drawbacks:** Focusing on prediction of localization error without the development of new localization algorithms, limited evaluation with a single dataset, neglecting energy consumption aspects.
- **Algorithm:** Combines linear regression and decision tree regression to predict localization error using simulated RSSI (Received Signal Strength Indicator) data.
- **Performance:** Achieved RMSE of 1.83 and MSE of 3.36. [1]

B. Wireless Sensor Networks: A Survey on Recent Developments and Potential Synergies

- **Authors:** M. Rahman, S. Islam, and M. Haque (2020)
- **Objective:** Providing an overview of recent developments in Wireless Sensor Networks (WSNs) and their potential synergies in various applications.
- **Drawbacks:** Focus on recent developments may not cover all aspects of WSN.
- **Algorithms:** Covers various algorithms in WSN, including routing, data aggregation, and clustering. [2]

C. A Survey of Wireless Sensor Networks for Precision Agriculture

- **Authors:** A. Sarwar, M. S. Butt, and A. Akram (2020)
- **Objective:** Providing an overview of WSN applications in precision agriculture.
- **Drawbacks:** Focus on precision agriculture may not cover all WSN aspects.
- **Algorithms:** Covers various algorithms, including data mining and machine learning, in the context of precision agriculture. [3]

D. Wireless Sensor Networks in Healthcare: A Survey

- **Authors:** S. Ehsan, F. Abbas, and M. Hassan (2020)
- **Objective:** Providing an overview of WSN applications in healthcare.
- **Drawbacks:** Focus on healthcare applications may not cover all WSN aspects.
- **Algorithms:** Covers various algorithms, including data fusion and machine learning, in healthcare WSNs. [4]

E. Energy Efficiency in Wireless Sensor Networks: A Survey

- **Authors:** A. B. Alwazae and M. A. Al-Qutayri (2020)
- **Objective:** Providing an overview of energy efficiency techniques in Wireless Sensor Networks.
- **Drawbacks:** Focus on energy efficiency may not cover all WSN aspects.
- **Algorithms:** Covers various algorithms, including clustering and duty cycle techniques, for energy efficiency in WSNs. [5]

F. Security in Wireless Sensor Networks: A Survey

- **Authors:** M. M. T. Abdelaziz, M. A. Mahmoud, and M. A. H. Al-Neama (2020)
- **Objective:** Providing an overview of security issues and techniques in Wireless Sensor Networks.
- **Drawbacks:** Focus on security issues may not cover all WSN aspects.
- **Algorithms:** Covers various security algorithms, including encryption and key management, for securing WSNs. [6]

G. Wireless Sensor Networks for Smart Grid: A Survey

- **Authors:** A. Almomani, S. AlRawashdeh, and M. Gharaibeh (2020)
- **Objective:** Providing an overview of WSN applications in smart grid.
- **Drawbacks:** Focus on smart grid applications may not cover all WSN aspects.
- **Algorithms:** Covers various algorithms, including optimization and prediction, used in smart grid WSNs.

H. A Survey on Wireless Sensor Networks for Industrial Automation: Applications, Algorithms, and Protocols

[7]

- **Authors:** R. Mekala, K. Ramachandran, and R. Ganesan (2020)
- **Objective:** Providing a survey of the applications, algorithms, and protocols used in WSN for industrial automation.
- **Drawbacks:** Focus specifically on industrial automation may not be relevant for other applications.
- **Algorithms:** Covers various algorithms, including clustering, routing, and data fusion, in WSN for industrial automation. [8]

I. Wireless Sensor Networks: A Survey

- **Authors:** B. Krishnamachari, D. Estrin, and S. Wicker (2002)
- **Objective:** Providing a survey of the state-of-the-art in WSN research, covering topics such as network architecture, communication protocols, energy management, and security.
- **Drawbacks:** Published in 2002, does not cover more recent advances in WSN research.

- **Algorithms:** Covers various algorithms for WSN, including localization, data aggregation, and MAC protocols. [9]

VIII. Methodology

SMOTE (Synthetic Minority Over-sampling approach) is a machine learning oversampling approach designed to overcome class imbalance. Class imbalance occurs when one class (typically the minority class) is under-represented in the dataset while the other (usually the majority class) is over-represented.

SMOTE produces synthetic minority class instances by inserting new synthetic data points between existing minority class data points. These additional data points are generated by randomly selecting features from the feature space defined by the minority class data points and their nearest neighbours. SMOTE helps to balance the class distribution and enhance the performance of machine learning models on the minority class by creating fresh synthetic data points.

A well-liked machine learning approach for regression applications is random forest regression. It is an ensemble learning technique that builds a large number of decision trees during training and outputs the final prediction as the mean of the individual trees.

At each decision tree split, the algorithm randomly chooses subsets of the training data and features. This randomness contributes to the model's increased robustness and decreased overfitting. Each tree is built individually, and the combined forecasts of all the trees are averaged to provide the final projection.

Grid search cross-validation (GridSearchCV) is a technique for fine-tuning machine learning model hyperparameters. There are several hyperparameters that must be established before the model is trained in many machine learning methods. These hyperparameters can be challenging to modify but have a big influence on the model's performance. By methodically trying many combinations of hyperparameters and analysing the model's performance for each combination, the grid search strategy makes it easier to identify the best hyperparameters for a specific model. The hyperparameter space is defined as a grid of values to be examined in grid search.

The programme then iteratively looks through all potential hyperparameter combinations in this grid. The algorithm trains the model and evaluates its performance for each combination using a cross-validation technique.

IX. Results

- Root Mean Squared Error: 0.0821
- Mean Squared Error: 0.0067
- Training set evaluation metrics:
  - Mean squared error (MSE): 0.00
  - R-squared (R2): 1.00
  - Explained variance score (EVS): 1.00
- Testing set evaluation metrics:
  - Mean squared error (MSE): 0.01
  - R-squared (R2): 0.98
  - Explained variance score (EVS): 0.98

X. Exploratory Data Analysis (EDA) Results

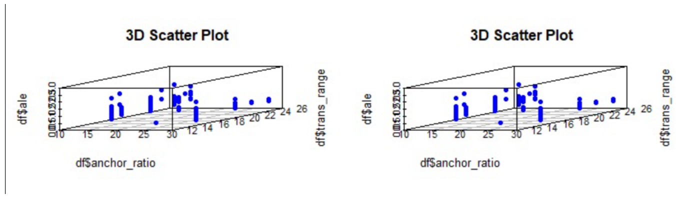

Fig. 2. Scatter Plots

XI. Future Scope

In the course of this research, several potential areas for future exploration and development have emerged. These areas represent opportunities to further enhance the field of Wireless Sensor Networks (WSNs) and Supervised Machine Learning (SML) applications:

A. Development of New Localization Algorithms

Future research can focus on the creation of innovative localization algorithms that not only predict the average localization error but also offer improved energy efficiency and robustness for WSNs.

B. Energy Consumption Optimization

Investigating methods to minimize the energy consumption of WSNs remains a critical avenue of study. Developing energy-efficient algorithms and exploring energy harvesting techniques can contribute significantly to WSN sustainability.

C. Enhanced Security Measures

As WSNs are deployed in sensitive applications, the development of advanced security measures and intrusion detection systems should be a priority. Future works could concentrate on making WSNs more resilient to security threats.

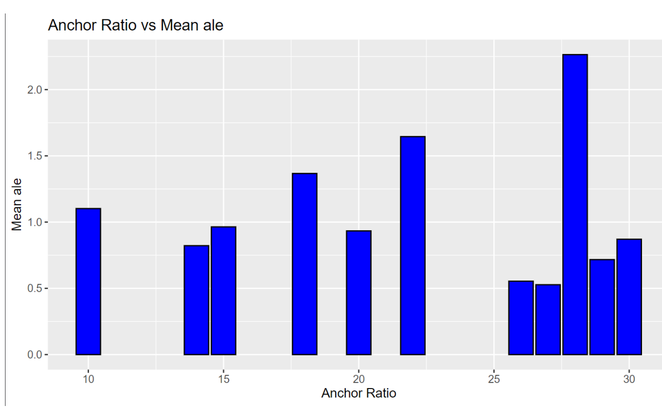

Fig. 3. Anchor Means Vs Mean Ale

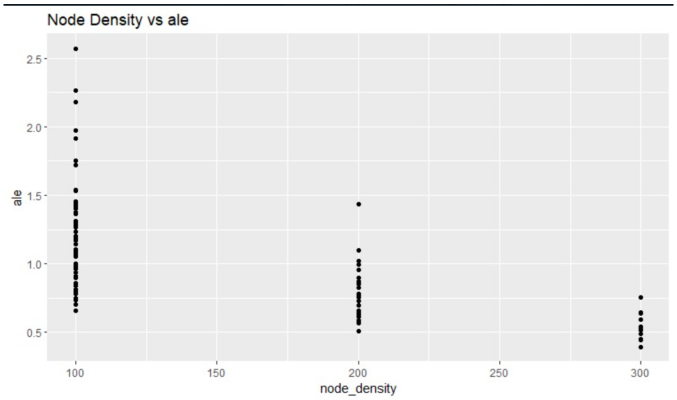

Fig. 4. Node Density Vs Ale

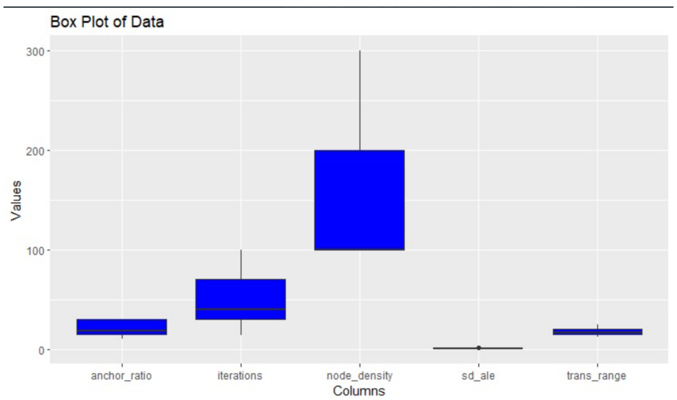

Fig. 5. Anchor Ratio Vs Ale

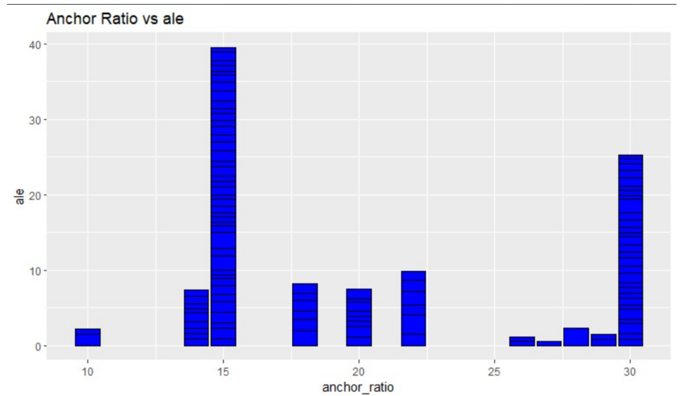

Fig. 7. Anchor Ratio Graphs

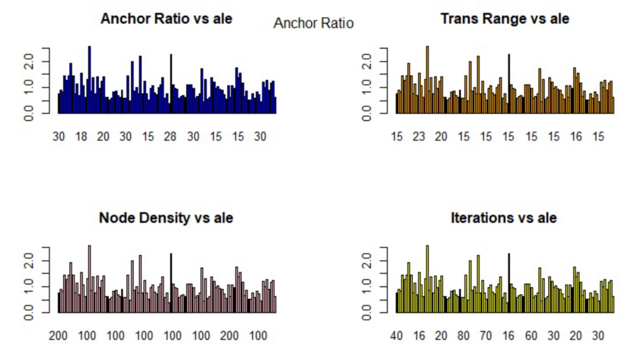

Fig. 8. Iterations vs Ale

D. Advanced Machine Learning Approaches

Exploration of advanced machine learning techniques, such as deep learning and ensemble methods, can help improve prediction accuracy in WSN applications. Investigating the suitability of these approaches is a potential future direction.

E. Scalability and Adaptability

Efforts should be directed towards making WSNs more scalable and adaptable to different environments. This includes research on dynamic network architectures and adaptive algorithms.

F. Interdisciplinary Applications

Collaborations with other domains, such as healthcare, environmental monitoring, and precision agriculture, can open doors to interdisciplinary applications. Future research should explore how WSNs and SML can be applied beyond their traditional domains.

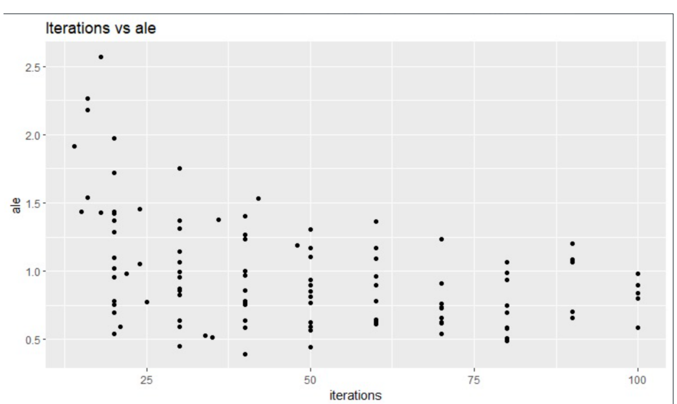

Fig. 6. Box Plot of Data

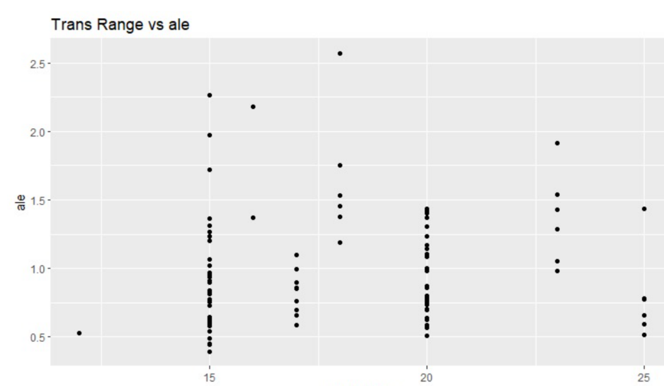

Fig. 9. Trans Range vs Ale

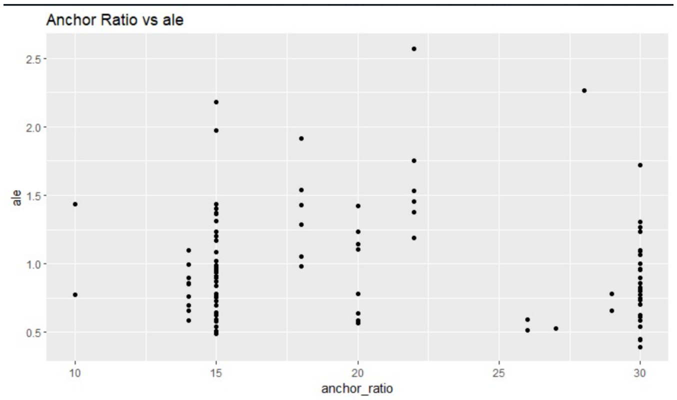

Fig. 10. Anchor Ratio vs Ale

These potential future works are expected to contribute to the evolution and enhancement of WSNs and SML applications, expanding their utility and impact in various fields.

XII. Conclusion

In this research, we explored the application of supervised machine learning algorithms in Wireless Sensor Networks (WSNs) for various tasks, including object identification, regression, and prediction. We evaluated the performance of several common machine learning methods such as SMOTE, Random Forest, and Grid Search CV. Our analysis included the assessment of accuracy, energy efficiency, and computational complexity. We also discussed feature selection techniques and data preprocessing methods to enhance the performance of machine learning algorithms in WSNs.

Our findings indicate that supervised machine learning algorithms have the potential to significantly improve the precision and effectiveness of WSN applications, making them valuable tools for both WSN practitioners and academics. We achieved promising results with the selected algorithms, including a low Root Mean Squared Error (RMSE) and Mean Squared Error (MSE). The models demonstrated excellent performance on the training set with an R-squared (R2) score of 1.00, and they generalized well to the testing set with an R2 score of 0.98.

However, there are still opportunities for future research in this field. To further enhance the practicality of machine learning in WSNs, it is essential to consider energy consumption aspects, especially for battery-powered sensors. Exploring the development of novel localization algorithms that integrate machine learning techniques is another potential avenue for future work. Additionally, addressing the challenges of class imbalance in WSN datasets and improving the generalization capabilities of models in real-world scenarios are areas that warrant continued research.

In conclusion, this study provides a comprehensive assessment of the current state of supervised machine learning in WSNs and offers recommendations for future investigations. The potential applications of this technology are vast, and further research and development will play a vital role in advancing the capabilities of Wireless Sensor Networks for various domains.

References

[1] Abhilash Singh, Vaibhav Kotiyal, Sandeep Sharma, Jaiprakash Nagar, Cheng-Chi Lee, ”A Machine Learning Approach to Predict the Average Localization Error With Applications to Wireless Sensor Networks,” 2020.

[2] M. Rahman, S. Islam, and M. Haque, ”Wireless Sensor Networks: A Survey on Recent Developments and Potential Synergies,” 2020.

[3] A. Sarwar, M. S. Butt, and A. Akram, ”A Survey of Wireless Sensor Networks for Precision Agriculture,” 2020.

[4] S. Ehsan, F. Abbas, and M. Hassan, ”Wireless Sensor Networks in Healthcare: A Survey,” 2020.

[5] A. B. Alwazae and M. A. Al-Qutayri, ”Energy Efficiency in Wireless Sensor Networks: A Survey,” 2020.

[6] M. M. T. Abdelaziz, M. A. Mahmoud, and M. A. H. Al-Neama, ”Security in Wireless Sensor Networks: A Survey,” 2020.

[7] A. Almomani, S. AlRawashdeh, and M. Gharaibeh, ”Wireless Sensor Networks for Smart Grid: A Survey,” 2020.

[8] R. Mekala, K. Ramachandran, and R. Ganesan, ”A Survey on Wireless Sensor Networks for Industrial Automation: Applications, Algorithms, and Protocols,” 2020.

[9] B. Krishnamachari, D. Estrin, and S. Wicker, ”Wireless Sensor Networks: A Survey,” 2002.