Gharyan University Journal of Engineering Science (GUJES)
ISSN (3105-4560)
Website: http://journals.gu.edu.ly
email: gujes@gu.edu.ly

Machine Learning Approaches for Indoor Positioning: A Case Study on KNN vs WKNN-inv and WKNN-cos

Mohammed Aboughalia [1 *], Mohamed A. Elalem [2]
[1] Electrical and Computer Department, Elmergib University, Alkhoms,Libya,mohammedaboughlia@gmail.com
[2] Electrical and Computer Department, Elmergib University, Alkhoms, Libya, moh.elalem@gmail.com

[*]*Corresponding author:* mohammedaboughlia@gmail.com.

Received: 10 Feb 2026	Accepted: 26 Feb 2026	Published: 01 March 2026

Abstract

Indoor localization is of paramount importance in intelligent environments, particularly in scenarios where Global Positioning System (GPS) signals are hindered by physical barriers within indoor settings. This study investigates machine learning methodologies designed to enhance the precision of indoor positioning systems. In particular, we conduct a comparative analysis of three algorithms: the conventional K-Nearest Neighbors (KNN), a Weighted KNN utilizing inverse distance metrics (WKNN-inv), and an additional variant that employs cosine similarity measures (WKNN-cos). Each algorithm was systematically implemented and assessed concerning its positioning accuracy. The results from simulations indicate that WKNN-inv yields an accuracy enhancement of approximately 11.2% relative to KNN, whereas WKNN-cos provides a 13.0% improvement.

Keywords: Indoor localization, KNN, WKNN-inv, WKNN-cos.

1. Introduction

Indoor positioning system is one of the most important technologies associated to smart environments, thus it is used in many applications. Evaluation of algorithm system for indoor positioning involves multiple criteria. Accreditation indicates the systems reliability. Global Positioning System (GPS) serves as standard for outdoor positions, does not work well in indoor positions due to the blocking of signals by walls. The research community's pursuit of reliable indoor positioning systems has intensified over the years, fueled by the growing demand for location-based services in smart cities. These systems play a crucial role in applications such as the localization of mobile robots and asset tracking within buildings [1]. Indoor positioning plays vital in the Internet of Things (IoT) accuracy and reliability. The process of determining the indoor position is divided into two parts. The first part is determining the distance between wireless technologies and objectives using one of the following techniques: Received Signal Strength Indicator (RSSI), Time of Arrival (TOA). Angle of Arrival (AOA), and Time Difference of Arrival (TDOA). Wireless technologies used for indoor positioning WiFi, RFID, UWB, BLE, LoRA. The second part is determining the actual location using one of the following methods: Triangulation, Multilateration, and

Gharyan University Journal of Engineering Science (GUJES)
ISSN (3105-4560)
Website: http://journals.gu.edu.ly
email: gujes@gu.edu.ly

Fingerprinting [2, 3]. Among these, fingerprinting is widely adopted and involves two phases: building a signal database during a training phase and matching real-time signal measurements during a testing phase to estimate location [4] . Machine Learning (ML) has been used in indoor positioning. ML techniques are utilized to mitigate issues like Non-line-of-sight problems, device heterogeneity, and environmental variations in indoor position, offering promising solutions to improve Indoor Positioning System (IPS) performance [5]. Machine learning techniques are better than conventional techniques in terms of accuracy. Some of the commonly used ML algorithms for indoor use are the k-nearest Neighbors (kNN) algorithm, the Support Vector Machine (SVM), and the Random Forest (RF) algorithm [1, 5] .
Several prior studies have explored the effectiveness of KNN and its weighted variant WKNN in WiFi-based indoor localization. For instance, [6] evaluated both algorithms using the UJIIndoorLoc and Alcala datasets, reporting that WKNN consistently outperformed standard KNN, achieving lower positioning errors (2.27 m vs. 2.62 m on Alcala, and 18.4 m vs. 18.6 m on UJI-IndoorLoc). Similarly, [7] emphasized the simplicity and effectiveness of these algorithms, noting that WKNN generally delivers higher accuracy by assigning greater weight to closer neighbors. These findings form the foundation for the current work, which aims to further assess and compare the performance of KNN and WKNN using different weighting schemes.
This paper evaluates and compares the performance of three algorithms for indoor localization using UJIIndoorLoc dataset: the standard KNN, a Weighted KNN variant based on inverse distance, and another variant based on cosine similarity. The primary objective is to assess the positioning accuracy of each algorithm and determine which weighting strategy yields the most reliable results in a complex indoor environment.

2. System Model

We use machine learning for WiFi-based indoor positioning systems. The values of the Received Signal Strength Indicator are gathered from WiFi access points (APs), which send out signals. A machine learning algorithm uses these RSSI values as input features and examines the signal patterns to determine the user's location inside the building. The output of the algorithm is the predicted location, enabling accurate indoor positioning for navigation or location-based services [8, 9]. Figure 1 illustrates the process of using machine learning techniques for Indoor Positioning Systems based on WiFi signals.

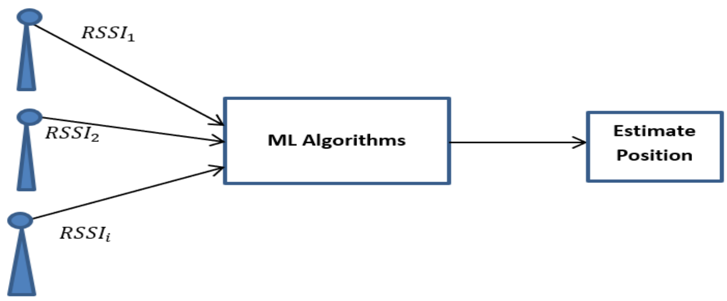

Figure1: Machine Learning for Indoor Positioning System

Gharyan University Journal of Engineering Science (GUJES)
ISSN (3105-4560)
Website: http://journals.gu.edu.ly
email: gujes@gu.edu.ly

A. Dataset

First, this work uses the UJIIndoorLoc dataset, which covers an area of over 110, 000 square meters, considered a large indoor space. It contains 19, 937 training samples and 1, 111 test samples, each containing 529 features. These features represent Wi-Fi access points, including coordinates, signal strength, floor number, building number, and other features [4]. Table 1 outlines the main components of the dataset utilized in this research.

Table 1 The General Structure of used Dataset.

AP_1	AP_2	……….	AP_{520}	(x, y)
$RSSI_{1,1}$	$RSSI_{1,2}$	……….	$RSSI_{1,520}$	(x_1, y_1)
$RSSI_{2,1}$	$RSSI_{2,2}$	……….	$RSSI_{2,520}$	(x_2, y_2)
……..	……..	………	………	………..
$RSSI_{n,1}$	$RSSI_{n,2}$	……….	$RSSI_{n,520}$	(x_n, y_n)

Each record includes 520 RSSI values representing signal strength from detected Access Points. RSSI values range from -104 dBm to 0 dBm. A value of "100" indicates that an access points was not detected.

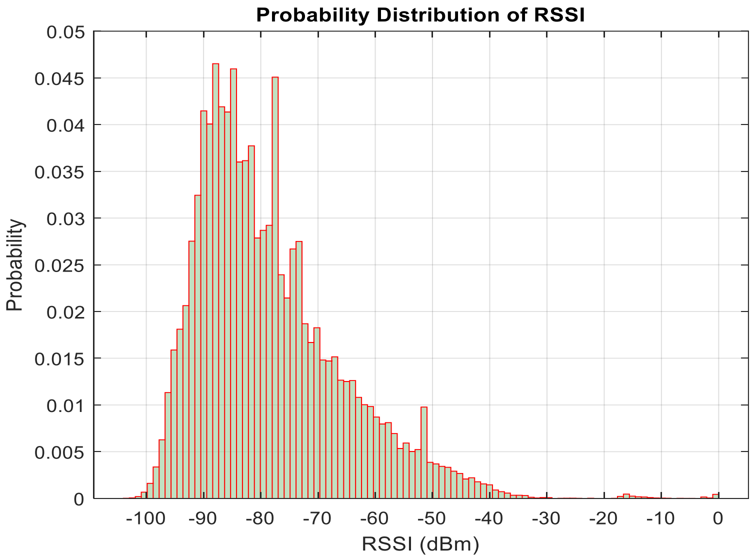

Figure 2: Probability Distribution of RSSI

Figure 2 shows the distribution of RSSI values for all APs used in the study. Values RSSI = 100 were excluded, with most signals concentrated in the range between -90 and -60 dBm. All RSSI values for all APs were adopted in all algorithms.

Gharyan University Journal of Engineering Science (GUJES)
ISSN (3105-4560)
Website: http://journals.gu.edu.ly
email: gujes@gu.edu.ly

B. K-Nearest Neighbour Algorithm

KNN algorithm is a non-parametric method that makes predictions based on the K most similar instances in the training set. For indoor localization, the algorithm identifies the K training samples whose signal strengths are most similar to those of the input (test) sample. The predicted location is then calculated by averaging the coordinates of these K nearest neighbors. The distance between two samples is typically measured using the Euclidean distance across all available RSSI features. The choice of the parameter K is crucial; it is usually selected through cross-validation to minimize prediction error [10].

$$d_i = \sqrt{\sum_{j=1}^{m} \left(f_{j_{test}} - f_{i j_{train}}\right)^2}, \tag{1}$$

where d_i is the distance between the test sample and the i^{th} training sample, $f_{i_{test}}$ is the RSSI value at the j^{th} dimension of the test point, $f_{i j_{train}}$ is the RSSI value of the i^{th} training sample at the j^{th} dimension and m is the number of dimensions WiFi access points.
The estimated location is computed as

$$(x, y) = \frac{1}{K} \sum_{i=1}^{K} (x_i, y_i) \tag{2}$$

(x_i, y_i) is the coordinates of the i^{th} nearest neighbor and K is the training samples.

C. Weighted K-Nearest Neighbor Algorithm

Weighted KNN is an extension of the standard KNN algorithm that improves estimation accuracy by assigning more influence to closer neighbors. Instead of treating all K neighbors equally, WKNN applies a weight to each one, typically based on distance or signal similarity.

- **Inverse Distance Weighting:** In this variant, the weight assigned to each neighbor is inversely proportional to its distance from the test point [10].

$$\boldsymbol{w_i} = \frac{1}{d_i}, \tag{3}$$

- **Cosine Similarity Weighting:** Another approach uses cosine similarity to measure the angular similarity between signal vectors [7].

Gharyan University Journal of Engineering Science (GUJES)
ISSN (3105-4560)
Website: http://journals.gu.edu.ly
email: gujes@gu.edu.ly

$$\boldsymbol{w_i} = similarity_i = \frac{f_{test} \cdot f_{i_{train}}}{||f_{test}||\ ||f_{i_{train}}||} \tag{4}$$

where $f_{test} \cdot f_{i_{train}}$ is the dot product of the test and training vectors and $||f_{test}||\ ||f_{i_{train}}||$ is their Euclidean norms.

Estimate positioning of target (x, y) for both weighting methods can be using following Equation

$$x = \frac{1}{\sum_{i=1}^{K} w_i} \sum_{i=1}^{K} w_i \cdot x_i \tag{5}$$

and

$$y = \frac{1}{\sum_{i=1}^{K} w_i} \sum_{i=1}^{K} w_i \cdot y_i \tag{6}$$

Figure 3 outlines a machine learning pipeline for comparing indoor localization algorithms using WiFi fingerprint data. The process begins with collecting and cleaning the dataset, followed by splitting it into training and test sets. Two models, KNN and WKNN, are trained on the training data. These models are then evaluated on the test data to assess their positioning accuracy. Finally, the accuracy of the models are compared to determine which algorithm performs best for indoor positioning. Two commonly used accuracy metrics, Mean Position Error and the Cumulative Distribution Function (CDF) of positioning errors, were used to assess the indoor positioning system's performance. These metrics offer distributional and scalar information about the accuracy of localization.

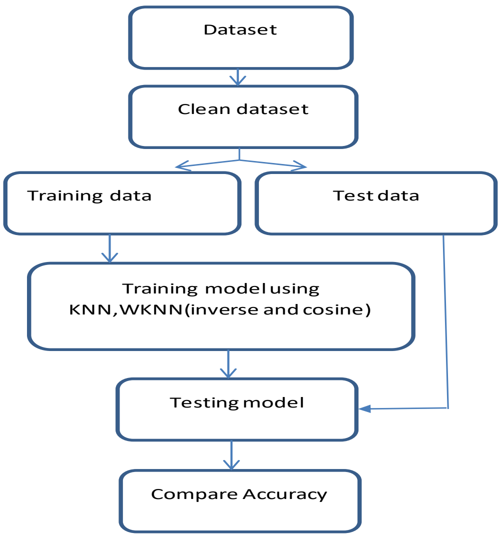

Figure 3: Indoor Localization using KNN and WKNN Algorithms

Gharyan University Journal of Engineering Science (GUJES)
ISSN (3105-4560)
Website: http://journals.gu.edu.ly
email: gujes@gu.edu.ly

3. Results and Discussion

Define The simulation evaluation is conducted using MATLAB and the UJIIndoorLoc dataset, which provides Wi-Fi signal fingerprints for indoor localization. For all models KNN and both WKNN variants the number of neighbors K was set to 5. The WKNN algorithm uses two different weighting schemes: inverse distance and cosine similarity.

Figure 4 shows how the optimal value of K varies with the mean positioning error for the three algorithms. The results show that K = 5 is the optimal choice with the lowest mean positioning error.

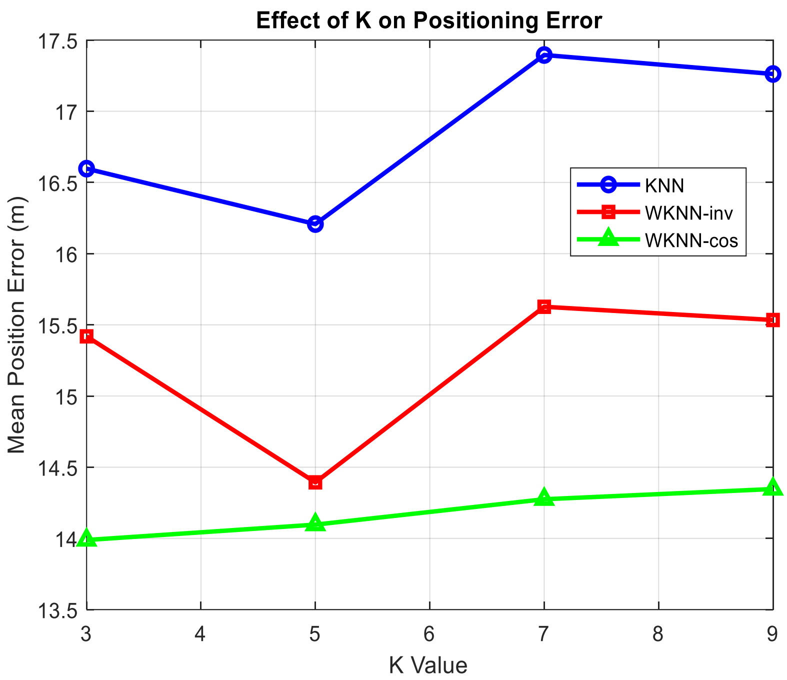

Figure 4: K Varies vs Mean Positioning Error for KNN, WKNN-inv and WKNN-cos Algorithms.

Figure 5 shows the cumulative distribution function of the error provided by each algorithm. The CDF graphic demonstrated that when it came to predicting the location of indoor objects, the WKNN-inv and WKNN-cos algorithms outperformed KNN. WKNN-cos achieves the highest accuracy. The cumulative distribution function (CDF) of positioning error is calculated.

Gharyan University Journal of Engineering Science (GUJES)
ISSN (3105-4560)
Website: http://journals.gu.edu.ly
email: gujes@gu.edu.ly

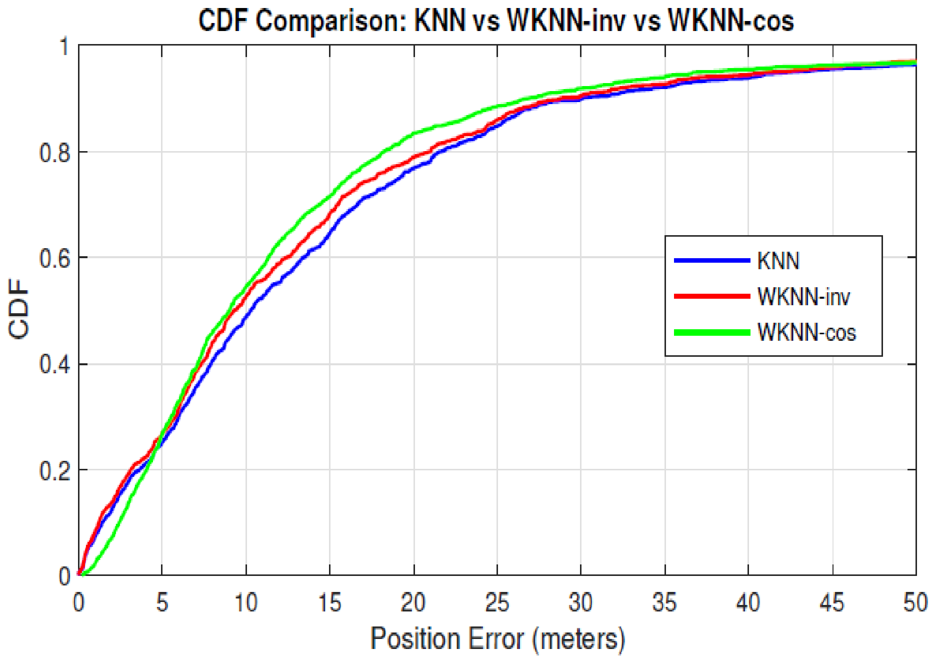

Figure 5: CDF vs Positioning Error for KNN, WKNN-inv and WKNN-cos Algorithms.

Table 2 lists the positioning errors (in meters) at the 50th, 70th, and 90th percentiles for each model. As shown, both WKNN variants outperform the standard KNN across all percentiles, with the cosine similarity variant showing the best overall accuracy.

Table 2: CDF of Positioning Error

Model	50%	70%	90%
KNN	10.44	16.68	30.23
WKNN-inv	9.27	15.49	28.9
WKNN-cos	9.02	14.43	26.86

The histogram graph is shown in Figure 6. The frequency is clearly very high when positioning error is low, but it decreases as positioning error rises. In KNNW algorithms, it is substantially lower than in KNN. Each algorithm's mean position error is determined and reported in Table II. When compared to the conventional KNN algorithm, the WKNN algorithm performs better. Moreover, the WKNN-cos has the lowest error.

Gharyan University Journal of Engineering Science (GUJES)
ISSN (3105-4560)
Website: http://journals.gu.edu.ly
email: gujes@gu.edu.ly

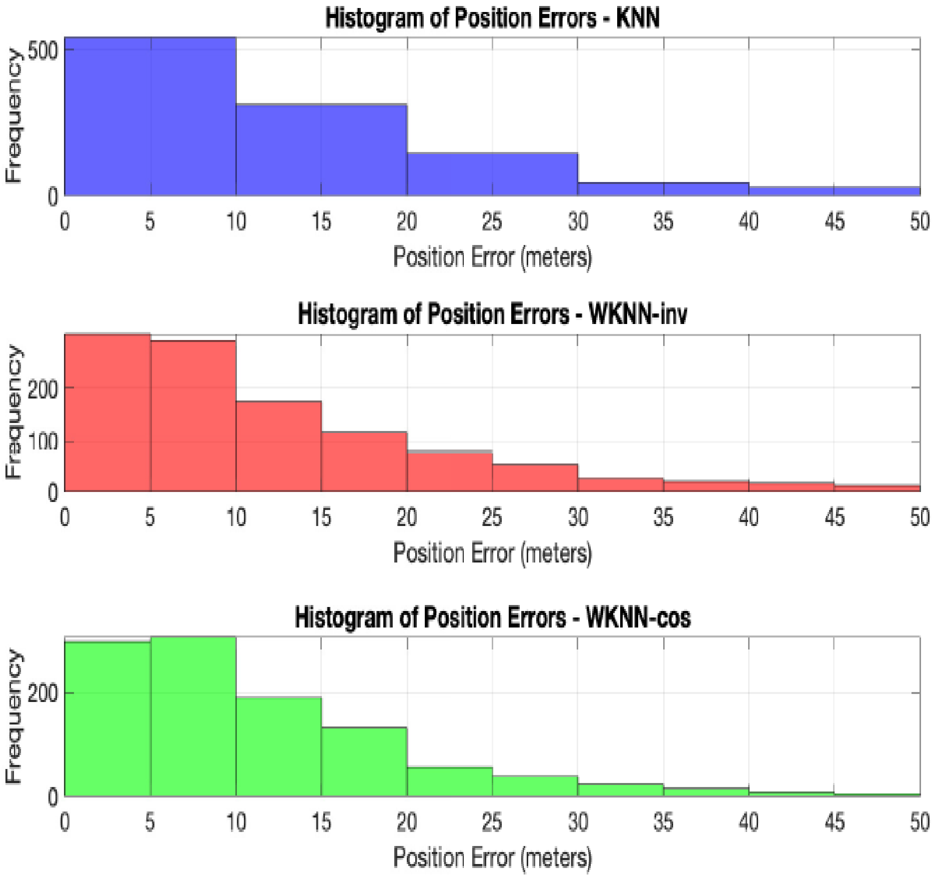

Figure 6: Histogram of Positioning Error for KNN and WKNN Algorithms.

Table 3 Mean Position Error

Model	Mean position error
KNN	16.2
WKNN-inv	14.4
WKNN-cos	14.1

The following formula calculates the percentage improvement of KNN and WKNN.

$$improvement\% = \frac{mean\ error_{KNN} - mean\ error_{WKNN}}{mean\ error_{KNN}} * 100$$

For WKNN-inv:

$$\frac{16.2 - 14.4}{16.2} * 100 = 11.2\%$$

Gharyan University Journal of Engineering Science (GUJES)
ISSN (3105-4560)
Website: http://journals.gu.edu.ly
email: gujes@gu.edu.ly

For WKNN-cos:

$$\frac{16.2 - 14.1}{16.2} * 100 = 13\%$$

4. Conclusion

This study presents a comparative analysis of two machine learning approaches for indoor localization using Wi-Fi fingerprinting: the standard KNN algorithm and two weighted variants, WKNN incorporating inverse distance and cosine similarity. The experiments were conducted using the UJIIndoorLoc dataset. The results demonstrated that both WKNN-inv and WKNN-cos outperform the traditional KNN algorithm across multiple evaluation metrics. In particular, the cosine similarity-based WKNN achieved the highest accuracy, yielding the lowest mean positioning error and the most favorable error distribution across different percentiles. As a future work, it might deal with exploring additional machine learning models such as decision trees, random forests, and deep learning techniques. Furthermore, combining these approaches through hybrid or ensemble methods may lead to improved accuracy in the indoor positioning systems.

References

[1] A. Nessa, B. AdhIkari, F. Hussain and X. N. Fernando, "A Survey of Machine Learning for Indoor Positioning," *IEEE Access,vol 8,* pp 214945–214965, 2020, doi: 10.1109/ACCESS.2020.3039271.

[2] A. Gadhgadhi, Y. Hacha¨ıchi and H. Zairi, "A Machine Learning based Indoor Localization," *2020 4th International Conference on Advanced Systems and Emergent Technologies (IC_ASET),* pp. 33-38,2020, doi: 10.1109/ICASET49463.2020.9318284.

[3] Dissanayake, R. M. M. R. Rathnayake, M. W. P. Maduranga, V. Tilwari and M. B, "RSSI and Machine Learning-Based Indoor Localization Systems for Smart Cities,". *MDPI/Eng,* pp 1468–1494, 2023, https://doi.org/10.3390/eng4020085.

[4] N. Singh, S. Choe and R. Punmiya, "Machine Learning Based Indoor Localization Using Wi-Fi RSSI Fingerprints:AnOverview,"*IEEEAccess*,pp.127150–127174, 2021, doi:10.1109/ACCESS.2021.311108.

[5] B. Adhikari, "Neural Network Based Recursive Least SquareTechnique for Indoor Wireless Positioning," Ryerson University, Canda Torinu , 2021.

[6] F. Qin, T. Zuo and a. X. Wang, "CCpos: WiFi Fingerprint Indoor Positioning System Based on CDAE-CNN," *Sensors* ,pp 1-17, 2021,http://doi.org/10.3390/s21041114.

[7] S. Liu, R. d. Lacerda and J. Fiorina, "Performance Analysis of Adaptive K for Weighted K-Nearest Neighbor based Indoor Positioning," Vehicular Technology Conference (VTC2022-Spring), Helsinki, Finland, 2022, doi:10.1109/vtc2022-spring54318.2022.9860699.

[8] E. I. A. Costa and J. A. Carro, "Development of AI/ML Methods for Advanced Device Localization in Beyond 5G Systems," Lund University, Sweden, 2023.

[9] R. Uttarwar and J. Valentín, "Indoor Positioning and Machine Learning Algorithms," *Lund University*, Sweden , 2021.

[10] H. Lu, L. Zhang, H. Chen, S. Zhang, S. Wang, H. Peng and J. Zou, "An Indoor Fingerprint Positioning Algorithm Based on WKNN and Improved XGBoost," *Sensors, 2023,* https://doi.org/10.3390/s23083952.