Accredited SINTA 2 Ranking
Decree of the Director General of Higher Education, Research, and Technology, No. 158/E/KPT/2021
Validity period from Volume 5 Number 2 of 2021 to Volume 10 Number 1 of 2026

Published online at: **http://jurnal.iaii.or.id**

JURNAL RESTI
(Rekayasa Sistem dan Teknologi Informasi)
Vol. 9 No. 2 (2025) 209 - 217 e-ISSN: 2580-0760

Feature Selection Using Pearson Correlation for Ultra-Wideband Ranging Classification

Gita Indah Hapsari[1*], Rendy Munadi[2], Bayu Erfianto[3], Indrarini Dyah Irawati[4]
[1,3]Doctoral of Informatics, School of Computing, Telkom University, Bandung, Indonesia
[2]Electrical Engineering, School of Electrical Engineering, Telkom University, Bandung, Indonesia
[4]Telecomunication Technology, School of Applied Science, Telkom University, Bandung, Indonesia
[1]gitaindahhapsari@telkomuniversity.ac.id, [2]rendymunady@telkomuniversity.ac.id, [3]erfianto@telkomuniversity.ac.id, [4]indrarini@telkomuniversity.ac.id

Abstract

Indoor positioning plays a crucial role in various applications, including smart homes, healthcare, robotics, and asset tracking. However, achieving high positioning accuracy in indoor environments remains a significant challenge due to obstacles that introduce NLOS conditions and multipath effects. These conditions cause signal attenuation, reflection, and interference, leading to decreased localization precision. This research addresses these challenges by optimizing feature selection LOS, NLOS, and multipath classification within Ultra-Wideband (UWB) ranging systems. A systematic feature selection approach based on Pearson correlation is employed to identify the most relevant features from an open-source dataset, ensuring efficient classification while minimizing computational complexity. The selected features are used to train multiple machine-learning classifiers, including Random Forest, Ridge Classifier, Gradient Boosting, K-Nearest Neighbor, and Logistic Regression. Experimental results demonstrate that the proposed feature selection method significantly reduces model training and testing times without compromising accuracy. The Random Forest and Gradient Boosting models exhibit superior performance, maintaining classification accuracy above 90%. The reduction in computational overhead makes the proposed approach highly suitable for real-time applications, particularly in edge-computing environments where processing efficiency is critical. These findings highlight the effectiveness of Pearson correlation-based feature selection in improving UWB-based indoor positioning systems. The optimized feature set facilitates robust LOS, NLOS, and multipath classification while reducing resource consumption, making it a promising solution for scalable and real-time indoor localization applications.

Keywords: *Indoor Positioning; Feature selection; Pearson correlation; Machine learning; UWB Ranging*

How to Cite: G. Indah Hapsari, R. Munadi, B. Erfianto, and I. Dyah Irawati, "Feature Selection Using Pearson Correlation for Ultra-Wideband Ranging Classification", *J. RESTI (Rekayasa Sist. Teknol. Inf.)*, vol. 9, no. 2, pp. 209 - 217, Mar. 2025.
DOI: https://doi.org/10.29207/resti.v9i2.6281

1. Introduction

Positioning technology has become an indispensable tool across a broad spectrum of sectors, including logistics, the military, and various industrial applications. This technology is critical in enabling function's location-based services, such as monitoring goods within smart environments like homes and buildings [1], managing inventories in warehouses, industrial navigation robots [2], [3], asset tracking [4], and tunnel or underground positioning [5]. Positioning technology implementation is also utilized in large areas such as hospitals [6], shops [7], and museums [8]. The need for position information on the implementation of positioning and tracking in various fields is the motivation for conducting this research.

Positioning is a process for estimating the position of a tag installed on a moving object, such as a mobile phone, drone, wearable device, robot, or vehicle in an observation area. A tag is a device that emits a signal or sends a message captured by a set of anchors to estimate the tag's position. Meanwhile, an anchor is a device that is given position knowledge as a reference for estimating the tag position.

Indoor positioning systems require accuracy and precision [9] due to the relatively confined nature of the environments in which they operate. One of the primary challenges in indoor positioning is dealing with obstacles that block direct signal paths, a situation known as Non-Line of Sight (NLOS). In NLOS conditions, signals do not travel directly between the

Received: 06-01-2025 | Accepted: 26-02-2025 | Published Online: 11-03-2025

transmitter and receiver, leading to potential delays and attenuation. This issue is exacerbated by signal multipath interference caused by reflections off walls and other surfaces, which results in a complex signal pattern characterized by fading, deep shadowing effects, and delays—all of which compromise the reliability of indoor communication channels [10], as depicted in Figure 1. Such challenges necessitate innovative solutions to overcome the limitations imposed by the indoor environment and ensure the efficacy of positioning systems.

Ultra-wideband (UWB) is one of the signals used for indoor positioning. UWB has high data speeds and can reach 100 Mbps. The frequency channel used in the indoor communication system ranges from 3100 to 10600 MHz, and the operation is limited to the UWB transmitter for indoor operation only. The wide bandwidth means that UWB is also reliable against narrowband signal interference narrow-band[11]. UWB is a signal with accuracy ranging from meter to centimeter scale for use in indoor positioning and includes signals with a short range of around 10-80m [12] -Klik atau ketuk di sini untuk memasukkan teks.[14]. Although UWB is said to be reliable in terms of positioning accuracy, until now, accuracy in the NLOS environment is still a challenge to be solved by researchers.

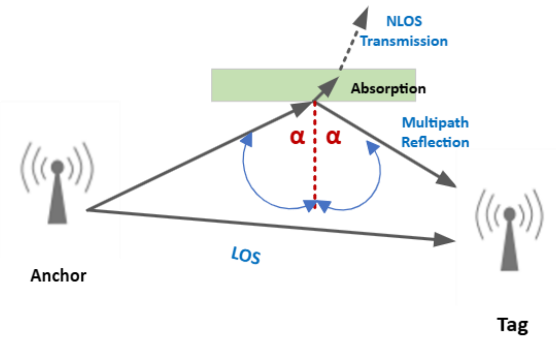


Figure 1. LOS, NLOS, and Multipath propagation signal

Many innovative solutions have been explored within the research community to address the complexities posed by NLOS conditions in positioning systems. Additionally, some of these strategies are integrated with processes designed to mitigate the impact of NLOS conditions, enhancing the overall reliability and accuracy of the positioning systems. A critical component in distinguishing Line of Sight (LOS) from NLOS conditions is the analysis of the Channel Impulse Response (CIR) of the signal. Specific characteristics of the CIR, such as the Receive Signal Strength (RSS), are instrumental in this differentiation process. Notably, researchers such as Barral [15] and Flueratoru [16] have highlighted the utility of RSS in their studies, demonstrating its effectiveness in enhancing the accuracy of NLOS/LOS classification.

Moreover, the estimation of the distance between the anchor and the tag remains a crucial variable in the effective implementation of positioning systems. Advanced ranging methods, including Time Difference of Arrival (TDoA) [3], Two-Way Ranging (TWR) [17], and Time of Flight (ToF) [18], are employed to achieve precise measurements. These techniques are essential for accurate distance estimation, playing a pivotal role in the performance and reliability of positioning systems under varying environmental conditions.

In recent years, the utilization of Ultra-Wideband (UWB) technology has become increasingly prevalent in machine learning applications aimed at enhancing positioning accuracy. The use of UWB Channel Impulse Response (UWB-CIR) is instrumental in performing essential tasks such as feature selection and feature extraction, significantly contributing to the refinement of machine learning models [19]-Klik atau ketuk di sini untuk memasukkan teks.[21].

Machine learning has been prominently featured in recent studies aiming to identify and classify NLOS propagation scenarios. Various types of machine learning methods have been proposed to perform NLOS and LOS classification, including Support Vector Machine (SVM) [22], KNN [23], CNN [24], support vector regression [25], and Random Forest (RF) [17].

Recent advances have explored the potential of machine learning to enhance the accuracy and efficiency of UWB systems under these challenging conditions. However, a more nuanced area of study involves the detailed classification of multipath effects, which has been less extensively explored. Only a select group of researchers, including Sebastian Kram[26], Jun Chang Sun [27], and Cun Liang Sang [17], have specifically addressed multipath effects. This specialized focus not only advances the theoretical understanding of multipath phenomena but also enhances practical applications in environments where signal interference is a significant challenge.

Cung Lian Sang explicitly classifies LOS, NLOS, and Multipath signals using CIR (Channel Impulse Response). This study differentiates LOS signals. NLOS and Multipath are based on CIR. A machine learning approach is used to identify LOS, NLOS, and Multipath. However, a large number of features can result in longer and make the model complex. Some implementations require models with low complexity but high accuracy. For example, communication between anchors in the self-calibration process. The speed of the anchor calibration process determines the performance of the calibration speed during the deployment process. Especially if the positioning system is implemented in an edge computing system.

This paper presents the classification of LOS, NLOS, and Multipath propagation signals by selecting features from open-source secondary datasets from previous research [17]. This study aims to explore the effectiveness of Pearson correlation for feature selection in classifying LOS, NLOS, and multipath signals in a UWB indoor positioning system. By identifying and utilizing the most relevant features. We anticipate not only an improvement in positioning accuracy but also a reduction in computational demands, enabling more scalable and real-time applications. The dataset with the best features is then used for test data and training data using machine learning classifier techniques, namely Logistic Regression (LR), K-Nearest Neighbor (KNN), Ridge Classifier (RC), Random Forest (RF) and Gradient Boosting (GB).

The contribution of this research is the selected features in the UWB ranging dataset so that the complexity of the dataset is lower but still has high accuracy. A compact dataset can reduce learning time so that positioning can be done faster, especially in real-time systems[28]. Besides, the system does not require large memory capacity if it is implemented in an edge computing system[29]. The result of this research will be applied to build a primary dataset and model for self-calibration using a hardware concept that has been published [30].

2. Research Methods

Figure 2 illustrates a structured methodological framework for optimizing machine learning-based classification in UWB-ranging systems using CIR features dataset [31]. The framework consists of five sequential stages, visually represented as a directional flowchart with interconnected blocks, emphasizing the stepwise approach to feature selection, model building, and performance evaluation. The process begins with the selection and preprocessing of a secondary dataset comprising CIR features. This dataset serves as the input for feature selection and model training, providing essential signal characteristics required for distinguishing between Line of Sight (LOS), Non-Line of Sight (NLOS), and multipath conditions.

This stage involves identifying the most informative features within the dataset using a statistical correlation-based approach, such as Pearson correlation. The objective is to eliminate redundant or irrelevant features, thereby reducing dimensionality and improving the computational efficiency of subsequent machine learning models.

After the selected features are generated, a machine learning model is built using the dataset with the selected features. Model creation uses several machine learning techniques for classification, including KNN, Random Forest, Logistic Regression, Ridge Classifier, and Gradient Boosting. The performance of each model is analyzed and compared, selecting the best model with accurate ranging.

The trained models are assessed based on various performance metrics, including accuracy, recall, precision, F1-score, and computational efficiency. A comparative analysis is conducted to determine the trade-offs between accuracy and processing time.

The stages of the method process are shown in Figure 2. The process begins by collecting a dataset of CIR features. The feature selection process is then carried out on the dataset in addition to reducing complexity as well as eliminating features that become noise in the dataset. The dataset used in this study is the UWB ranging dataset from Cung Lian Sang's research [17]. The dataset consists of 12 features that are measured using the DWM1000 UWB module. Feature selection is then carried out using Pearson. Pearson is one method to validate the correlation of the best features. Analysis is performed on the correlation value generated from each feature to determine the best feature to be used as a dataset.

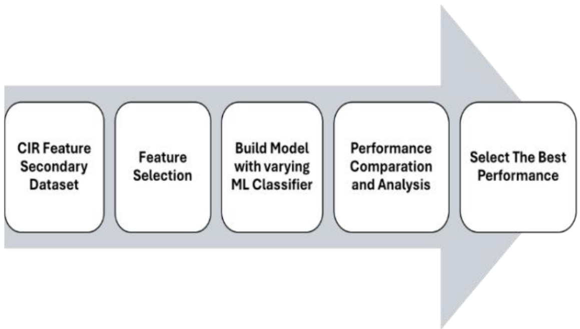


Figure 2. Research Method

2.1 Channel Impulse Response (CIR)

Channel Impulse Response (CIR) is the characteristic response of a signal on a channel. In several studies, CIR is analyzed to extract certain characteristics that will be used as input features or variables to measure environmental conditions. One of them is Receive Signal Strength CIR (RSS-CIR), which is a signal characteristic related to signal power. Other CIR variables can be related to variable time intervals or delays that occur in the signal received at the receiver.

The secondary dataset consists of a ranging dataset measured using the UWB DW1000 module. Ranging is the measured distance between UWB modules (anchors and tags). The DW1000 is a low-power, single-chip CMOS radio transceiver compliant with the IEEE 802.15.4-2011 standard. DW1000 is equipped with a variable feature to analyze the quality of the received signal based on CIR and time stamp data.

DW1000 makes it possible to calculate the estimated power of the first path signal or First Path Power Level (FPPL) using Formula 1. A is a constant value of 113.77 for PRF16 MHz or 121.74 for PRF 64MHz. Estimating Receive Signal Power symbolized by the variable RX-Power (RXP) is another status variable that can be used to detect NLOS signals. RXP can be calculated using Formula 2. In Formula 3, the difference between RXP and FPPL (Diff_P) can be used to determine the state of the LOS and NLOS channels. If the difference is < 6dB, then the channel is LOS; otherwise, if the difference is > 10 dB, then the channel is NLOS.

A graph is first plotted from the signal on LOS, NLOS, and Multipath from the dataset to prove that $Diff_P$ can be used to indicate LOS, NLOS, and Multipath propagation signals. Figure 3 shows the results of graphic plotting, which shows differences in signal characteristics, especially in the amplitude of signal fluctuations in each Difference of LOS, NLOS, and Multipath signal. Based on this, $Diff_P$ can be used to indicate the difference between LOS, NLOS, and Multipath.

Figure 3 effectively illustrates the signal variations in different UWB propagation scenarios and can significantly affect indoor positioning accuracy. The results emphasize the importance of signal classification in indoor positioning applications, particularly in mitigating NLOS errors and multipath distortions. Advanced machine learning models can be trained in these signal characteristics to improve real-time localization accuracy and enable robust UWB-based navigation systems.

$$FPPL = 10xlog_{10}\left(\frac{F_1^2+F_2^2+F_3^2}{RXPACC^2}\right) - A \quad \text{dBm} \tag{1}$$

$$RXP = 10xlog_{10}\left(\frac{Cx2^{17}}{RXPACC^2}\right) - A \quad \text{dBm} \tag{2}$$

$$Diff_P = RXP - FPPL \tag{3}$$

2.2 Dataset and Feature Selection

The dataset consists of LOS, NLOS, and Multipath datasets with the same amount of data (balance data). This data set with selected features is then used as training data for training the classification model. The dataset totals 185790 imbalanced data consisting of 61930 LOS data, 61930 NLOS data, and 61930 Multipath data.

The features in the dataset are CIR signal characteristics, which consist of 12 features, including: Distance measured between UWB modules (ranging measurement); Harmonic combining of FP1, FP2, FP3 (FP); Amplitude of the first harmonic (F1); Amplitude of the second harmonic (F2); Amplitude of the third harmonic (F3); Amplitudaof CIR (CIR-PWR); Preamble Accumulation count recorded on the DW1000 chip (RXPACC); Estimated power level of FP (FPPL); Estimation of power level from RX (RXP); Difference in value between Estimated power level from FP and Estimated power level from RX ($Diff_P$); Recorded noise standards on the DW1000 chip (SNR); Maximum noise recorded on the DW1000 chip (Pmax Noise)

Distance is the estimated distance between the anchor and the tag measured on the UWB module. The estimated distance can be formulated using equation (3). Variable $\hat{d}$ is the estimated distance between the anchor and the tag, ToA is the Time of Arrival or signal transmission time between the anchor and the tag, while c is the speed of light, namely $c = 3x10^8\ {}^{m}/_{s}$.

F1, F2, and F3 are the first path amplitudes at point 1, point 2, and point 3. CIR-PWR is a feature whose maximum CIR power value. RXPACC is the count of received preamble symbols. Pmax Noise is the maximum noise value indicated by the LED algorithm. FP is the index of the first path detected in the accumulator register.

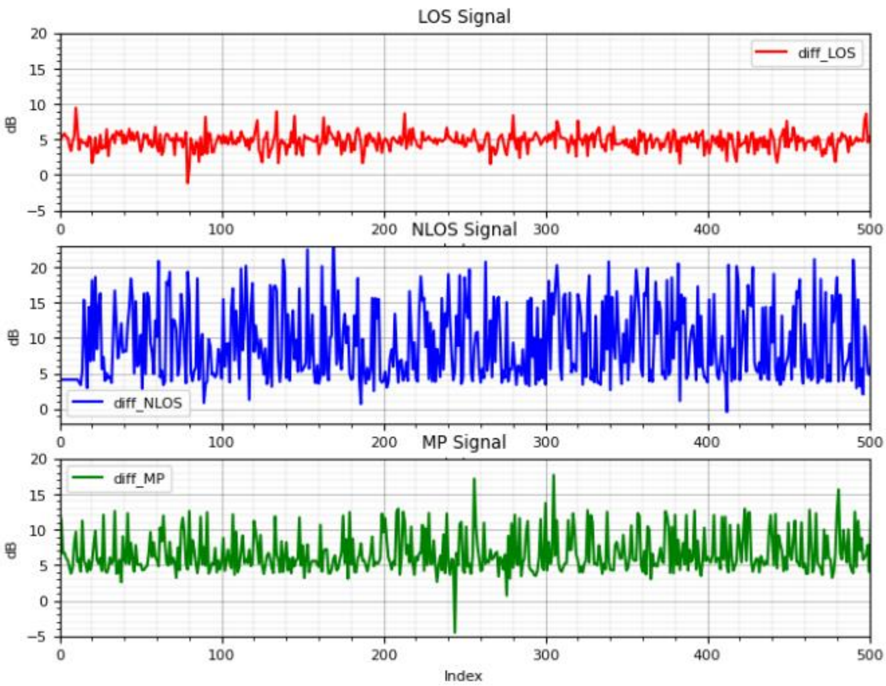


Figure 3. $Diff_P$ Signal characteristics in LOS, NLOS, and Multipath

One of the quality-related status variables on the DW1000 that can be used to test the quality of the received signal is The Standard Deviation of Channel Impulse Response Estimate (CIRE) Noise value, which is contained in the SNR variable in the Register file: 0x12-Rx Frame Quality Information. This variable can be used to measure noise and measure the timestamp of the received frame. SNR can be used as an absolute value, or this value can be compared with the First Path Amplitude Value. A higher absolute value of CIRE means the quality of the received timestamp is poor. High noise means the actual first path is buried in noise. The F1, F2, F3, CIR, RCPACC, Pmax Noise, FP, and SNR features are obtained from the diagnostic register on the DW1000 UWB module.

The selection of features in the dataset does not include features that have been part of the FPPL and RXP formula calculations, namely features F1, F2, F3, RXPACC, and $Diff_P$). Meanwhile, CIR is still included in correlation validation because CIR is a feature that represents signal amplitude power which, if it decreases, can represent NLOS propagation. Therefore, the features validated in the feature selection process are FP, Distance, CIR, FPPL, RXP, and STDNoise.

After obtaining the selected features, then prepare a dataset with these selected features for training and model testing. Several combinations of comparisons between training data and test data are carried out to get the best performance on the model being tested. A comparison of test and training data that shows the best performance will be used for performance analysis based on the confusion matrix.

The model was built using Logistic Regression (LR), K-Nearest Neighbor (KNN), Ridge Classifier (RC), Random Forest (RF) and Gradient Boosting (GB). Performance analysis to determine the best classifier is carried out by analyzing the Accuracy, Recall, Confusion Matrix, Precision, and f1-score parameters. At this stage, comparisons are made between each machine learning technique to determine the classifier with the best accuracy for LOS, NLOS, and multipath classification.

3. Results and Discussions

This section describes the result and analysis of feature selection. An analysis was also carried out on the model built using features selected based on accuracy and learning time.

The CIR dataset uses the Cung Lian Sang data set by using the FP, Distance, CIR, FPPL, RXP, and SNR features. Then, correlation testing is carried out on the dataset to carry out feature selection. Figure 4 shows a correlation heatmap representing the Pearson correlation coefficients between various features in a dataset.

The correlation matrix quantifies the linear relationships between different variables, with values ranging from -1 to 1, where 1 represents a perfect positive correlation, -1 indicates a perfect negative correlation, and values close to 0 suggest no correlation. The color gradient from dark blue (negative correlation) to bright orange (positive correlation) visually represents the strength and direction of relationships between the features.

3.1 Feature Selection

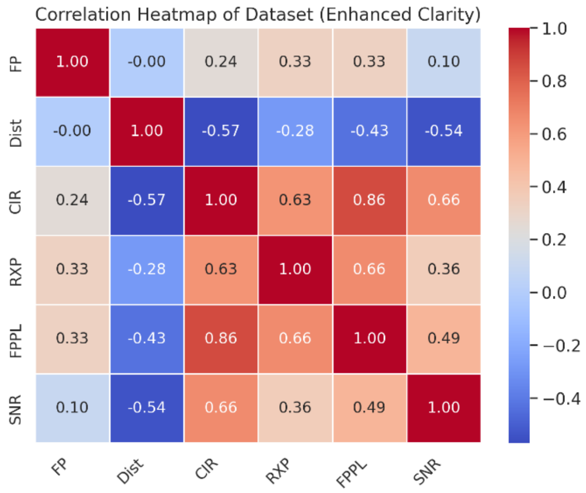


Figure 4. Correlation Matrix

As shown in Figure 4, CIR and FPPL exhibit the highest positive correlation (0.86), suggesting that the First Path Power Level strongly influences the Channel Impulse Response amplitude. CIR and Distance show a strong negative correlation (-0.57), implying that as the measured distance increases, the CIR amplitude tends to decrease due to signal attenuation and path loss. Distance is negatively correlated with FPPL (-0.43) and RXP (-0.28), reflecting the expected trend where signal strength diminishes as the distance between transmitter and receiver increases. SNR exhibits a moderate correlation with CIR (0.66) and FPPL (0.49), suggesting that stronger received signals generally lead to higher signal-to-noise ratios. FP (First Path index) has relatively weak correlations with other features, indicating that it may be less influential in defining the overall positioning accuracy, so we remove FP from the feature.

The results on LOS, NLOS, and Multipath correlation maps show high correlation values for several features including Distance, CIR, RXP, FPPL, and SNR. Based on the Pearson correlation results, these features will be used next to conduct model training and model testing.

3.2 Classification of LOS, NLOS, and Multipath Methods

Model training uses machine learning techniques for classification, including LR, KNN, RC RF, and GB. The total amount of data is divided into training data and test data. Accuracy observations are carried out by testing several comparisons between the training data and the data set (Test Size) as in Formula 4.

$$Test\ Size = \frac{Data\ Test}{Data\ Training} \quad (4)$$

Figure 5 shows the results of the accuracy of each Test Size value. The best accuracy results are obtained on Test Size 0.1, especially on KNN, RF, and GB. There is an increase in accuracy as the amount of training data increases in KNN, RF, and GB. Meanwhile, in LR and RC, increasing the amount of training data results in accuracy that sometimes decreases or increases even though the value is not large. From the results of this test, further training and testing of the model will use a test size of 0.1.

The next test is to test the performance of the classification results of each model. Figure 5 shows the results of the normalized confusion matrix for each classifier using the selected features. RF and GB show almost the same results, namely the results of LOS, NLOS, and Multipath classification above 90%. Meanwhile, the KNN algorithm shows lower LOS classification results than NLOS and Multipath. Low performance in LOS, NLOS, and Multipath classification results in the LR and RC algorithms, especially in the results of multipath classification. From the results of this test, RF shows the best performance in accuracy, followed by GB and KNN.

Next, a comparison of accuracy was carried out between before and after feature selection. This is done to determine whether the classification performance is still reliable after feature selection. The accuracy comparison is shown in Figure 5, where RF is the best classifier both before and after feature selection. There is a 1% decrease in the F1-Score value and 0.52% in the accuracy value. The GB classifier produces the same F1_Score value, but there is a decrease in accuracy of 0.28%. The same thing happened to the LR and RC classifiers, which also experienced a decrease in accuracy of 0.39% after feature selection. Unlike the case with KNN, which actually experienced an increase in the value of the F1-Score and its accuracy after feature selection was carried out.

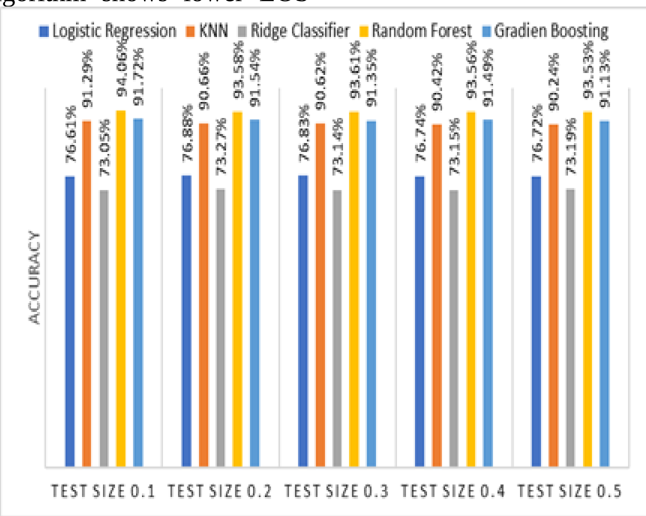


Figure 5. Accuracy based on test size

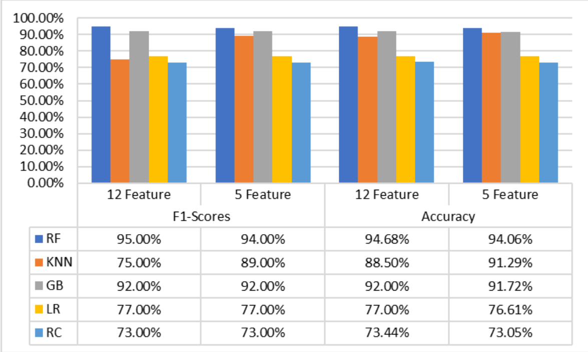
Complexity testing is carried out by calculating the length of execution time when machine learning trains the LOS, NLOS, and Multipath classification models. Testing was carried out by comparing the execution time for the classification process of 12 features (before feature selection) and 5 features (after feature selection). The test results of the training time are shown in Table 1.

Table 1. Execution Time of Classification

Classifier	Training Time of 5 feature (second)	Training Time using 12 feature (second)	Test Time using 5 feature (second)	Test Time using 12 feature (second)
KNN	0.24168 ± 0.01956	0.41377 ± 0.03785	0.42218 ± 0.03411	1.01603 ± 0.09205
RF	15.13889 ± 0.64458	26.36599 ± 1.02798	0.29320 ± 0.01788	0.31537 ± 0.02689
GB	46.57136 ± 2.90652	97.06870 ±2.73955	0.07777 ± 0.00568	0.08546 ± 0.01136
LR	2.29269 ± 0.11382	4.79339 ±0.12986	0.00300 ± 0.00148	0.00385 ± 0.00042
RC	0.35484 ± 0.02866	0.38696 ± 0.02821	0.00262 ± 0.00060	0.00397 ± 0.00024

From Table 1, feature selection significantly affects training and testing times. Gradient Boosting (GB) has the highest computational cost, with training times increasing from 46.57s (5 features) to 97.06s (12 features). Random Forest (RF) also sees a rise from 15.13s to 26.36s, though test times remain relatively small. Logistic Regression (LR) and Ridge Classifier (RC) require significantly less time, with LR training in just 2.29s (5 features) to 4.79s (12 features) and minimal test times (0.00262s–0.00397s). KNN, though computationally simpler, sees its training time rise from 0.24s to 0.41s and test time from 0.42s to 1.01s, making it inefficient for large datasets. Overall, feature selection reduces computational overhead while maintaining efficiency, benefiting tree-based and regression models.

Figure 6 presents a comparative analysis of F1-scores and accuracy for five different machine learning classifiers: RF, KNN, GB, LR, and RC. RF and GB, being tree-based models, demonstrate robustness to feature reduction with minimal accuracy degradation. However, LR and RC show relatively lower scores, suggesting that these models may not be as effective in handling the complexity of UWB indoor positioning data. The improvement in KNN's accuracy and F1-score suggests that reducing the feature set eliminates redundant information, leading to better distance-based classification.



	F1-Scores 12 Feature	F1-Scores 5 Feature	Accuracy 12 Feature	Accuracy 5 Feature
RF	95.00%	94.00%	94.68%	94.06%
KNN	75.00%	89.00%	88.50%	91.29%
GB	92.00%	92.00%	92.00%	91.72%
LR	77.00%	77.00%	77.00%	76.61%
RC	73.00%	73.00%	73.44%	73.05%

Figure 6. Accuracy and F1-Scores of LOS, NLOS and Multipath classification

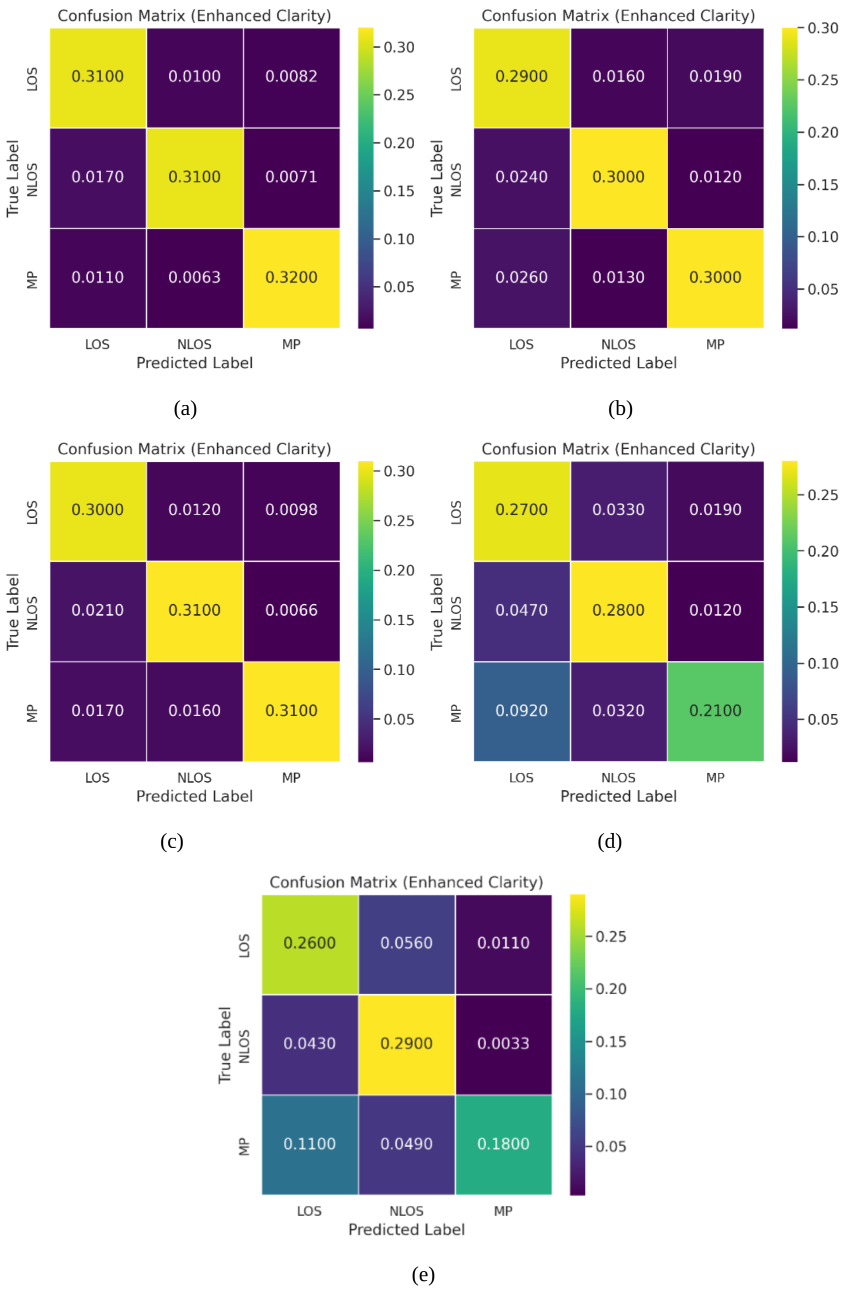


Figure 7. Confusion Matrix of LOS, NLOS, and Multipath : (a) RF, (b) KNN, (c) GB, (d) LR, and (e) RC

The detail of the confusion matrix is present in Figure 7. Figures 7(a) and 7(c) show the highest classification accuracy of RF and GB, with high values along the diagonal (above 0.30), indicating their strong capability in distinguishing LOS, NLOS, and MP. KNN (b) and LR (d) exhibit moderate classification performance, with slightly higher misclassification rates, particularly in distinguishing MP from LOS and NLOS. RC (e) has the lowest performance, with noticeable misclassifications, particularly between MP and LOS. These results suggest that tree-based models (RF and GB) are the most effective for UWB-based signal classification, while linear models (LR and RC) struggle with distinguishing multipath conditions.

The proposed study demonstrates the effectiveness of Pearson correlation-based feature selection in classifying LOS, NLOS, and multipath (MP) signals in Ultra-Wideband (UWB) indoor positioning while reducing computational complexity. Compared to paper [17], which used all 12 features, the proposed method selects the most relevant five features (Distance, CIR, RXP, FPPL, and SNR), maintaining high classification accuracy while significantly reducing training and testing times. Random Forest (RF) and Gradient Boosting (GB) maintain over 90% accuracy, similar to the full-feature model, while training time for GB decreases from 97.06s to 46.57s and for RF from 26.36s to 15.13s. KNN benefits the most, improving accuracy after feature selection. The findings highlight that feature selection optimizes performance, making UWB-based indoor positioning more efficient for real-time applications, particularly in edge computing environments

4. Conclusions

This study demonstrates the effectiveness of Pearson correlation-based feature selection in improving the classification of LOS, NLOS, and Multipath signals in Ultra-Wideband (UWB) ranging systems. The selected features Distance, CIR, RXP, FPPL, and SNR exhibit strong correlations, contributing to more efficient and accurate machine learning classification. Experimental results show that RF and GB outperform other classifiers, achieving over 90% classification accuracy, making them the most suitable for real-time UWB positioning applications. KNN benefits significantly from feature selection, improving its accuracy and F1-score. However, LR and RC show relatively lower performance, highlighting their limitations in complex indoor environments. Feature selection also reduces computational complexity, leading to faster model training and testing, particularly benefiting real-time and edge computing applications. Despite RF and GB having high accuracy, their training times remain longer than other models, presenting a trade-off between accuracy and efficiency. Future research can focus on developing a primary dataset using selected features, optimizing classification models for self-calibration in real-time indoor positioning systems, and improving execution speed while maintaining high accuracy.

Acknowledgements

This work was supported by the Research and Community Service Unit of Telkom University for providing Doctoral research grants and the Informatics Doctoral Program at Telkom University for providing facilities for research. The author declares no conflict of interest. This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

References

[1] M. Rosiak, M. Kawulok, and M. Maćkowski, "The Effectiveness of UWB-Based Indoor Positioning Systems for the Navigation of Visually Impaired Individuals," *Applied Sciences (Switzerland)*, vol. 14, no. 13, Jul. 2024, doi: 10.3390/app14135646.

[2] X. Yu, Q. Li, J. P. Queralta, J. Heikkonen, and T. Westerlund, "Applications of UWB Networks and Positioning to Autonomous Robots and Industrial Systems," Mar. 2021, doi: 10.1109/MECO52532.2021.9460266.

[3] W. Zhao, A. Goudar, X. Qiao, and A. P. Schoellig, "UTIL: An ultra-wideband time-difference-of-arrival indoor localization dataset," *International Journal of Robotics Research*, Sep. 2024, doi: 10.1177/02783649241230640.

[4] W. He, M. Xi, H. Gardner, B. Swift, and M. Adcock, "Spatial anchor based indoor asset tracking," in *Proceedings - 2021 IEEE Conference on Virtual Reality and 3D User Interfaces, VR 2021*, Institute of Electrical and Electronics Engineers Inc., Mar. 2021, pp. 255–259. doi: 10.1109/VR50410.2021.00047.

[5] X. You, D. Tian, C. Liu, X. Yu, and L. Song, "Vehicles positioning in tunnel: A real-time localization system using DL-TDOA technology," *Journal of Internet Technology*, vol. 22, no. 5, pp. 967–978, 2021, doi: 10.53106/160792642021092205003.

[6] J. Kolakowski, V. Djaja-Josko, M. Kolakowski, and K. Broczek, "UWB/BLE tracking system for elderly people monitoring," *Sensors (Switzerland)*, vol. 20, no. 6, Mar. 2020, doi: 10.3390/s20061574.

[7] P. Wang, B. Guo, Z. Wang, and Z. Yu, "ShopSense:Customer Localization in Multi-Person Scenario with Passive RFID Tags," *IEEE Trans Mob Comput*, vol. 21, no. 5, pp. 1812–1828, May 2022, doi: 10.1109/TMC.2020.3029833.

[8] Y. X. Zhang and Y. Zi, "Mixed reality annotations system for museum space based on the uwb positioning and mobile device," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, Springer Science and Business Media Deutschland GmbH, 2020, pp. 328–342. doi: 10.1007/978-3-030-58465-8_25.

[9] H. Fei, F. Xiao, B. Sheng, H. Huang, and L. Sun, "Motion path reconstruction in indoor environment using commodity Wi-Fi," *IEEE Trans Veh Technol*, vol. 68, no. 8, pp. 7668–7678, Aug. 2019, doi: 10.1109/TVT.2019.2921408.

[10] S. M. Asaad and H. S. Maghdid, "A Comprehensive Review of Indoor/Outdoor Localization Solutions in IoT era: Research Challenges and Future Perspectives," Jul. 20, 2022, *Elsevier B.V.* doi: 10.1016/j.comnet.2022.109041.

[11] J. Ninnemann, P. Schwarzbach, and O. Michler, "Toward UWB Impulse Radio Sensing: Fundamentals, Potentials, and Challenges," in *UWB Technology - New Insights and Developments*, IntechOpen, 2023. doi: 10.5772/intechopen.110040.

[12] N. El-Sheimy and Y. Li, "Indoor navigation: state of the art and future trends," Dec. 01, 2021, *Springer*. doi: 10.1186/s43020-021-00041-3.

[13] Y. Li *et al.*, "Location-Enabled IoT (LE-IoT): A Survey of Positioning Techniques, Error Sources, and Mitigation," Apr. 2020, doi: 10.1109/JIOT.2020.3019199.

[14] C. Lian, S. Referent, I. Ulrich, R. Korreferent, and I. U. Witkowski, "Bidirectional UWB Localization with Rigorous Sectoral Evaluations." Accessed: Feb. 18, 2025. [Online].

Available: https://www.researchgate.net/publication/367157814_Bidirectional_UWB_Localization_with_Rigorous_Sectoral_Evaluations

[15] V. Barral, C. J. Escudero, J. A. García-Naya, and R. Maneiro-Catoira, "NLOS identification and mitigation using low-cost UWB devices," *Sensors (Switzerland)*, vol. 19, no. 16, Aug. 2019, doi: 10.3390/s19163464.

[16] L. Flueratoru, E. S. Lohan, and D. Niculescu, "Self-Learning Detection and Mitigation of Non-Line-of-Sight Measurements in Ultra-Wideband Localization," in *2021 International Conference on Indoor Positioning and Indoor Navigation, IPIN 2021*, Institute of Electrical and Electronics Engineers Inc., 2021. doi: 10.1109/IPIN51156.2021.9662532.

[17] C. L. Sang, B. Steinhagen, J. D. Homburg, M. Adams, M. Hesse, and U. Rückert, "Identification of NLOS and multi-path conditions in UWB localization using machine learning methods," *Applied Sciences (Switzerland)*, vol. 10, no. 11, Jun. 2020, doi: 10.3390/app10113980.

[18] Z. Cui, T. Liu, S. Tian, R. Xu, and J. Cheng, "Non-Line-of-Sight Identification for UWB Positioning Using Capsule Networks," *IEEE Communications Letters*, vol. 24, no. 10, pp. 2187–2190, Oct. 2020, doi: 10.1109/LCOMM.2020.3003688.

[19] A. G. Ferreira, D. Fernandes, S. Branco, A. P. Catarino, and J. L. Monteiro, "Feature selection for real-time nlos identification and mitigation for body-mounted uwb transceivers," *IEEE Trans Instrum Meas*, vol. 70, 2021, doi: 10.1109/TIM.2021.3070619.

[20] F. Che, Q. Z. Ahmed, F. A. Khan, and P. I. Lazaridis, "Feature-Based Generalized GaussianDistribution Method for NLoS Detectionin Ultra-Wideband (UWB) IndoorPositioning System," *IEEE SENSOR JOURNAL*, vol. 22, no. 19, pp. 18726–18739, 2022, doi: 10.23919/ICAC50006.2021.9594226.

[21] F. Che, W. Bin Abbas, Q. Z. Ahmed, B. Amjad, F. A. Khan, and P. I. Lazaridis, "Weighted Naive Bayes Approach for Imbalanced Indoor Positioning System Using UWB," in *2022 IEEE International Black Sea Conference on Communications and Networking, BlackSeaCom 2022*, Institute of Electrical and Electronics Engineers Inc., 2022, pp. 72–76. doi: 10.1109/BlackSeaCom54372.2022.9858257.

[22] M. Stocker, M. Gallacher, C. A. Boano, and K. Römer, "Performance of support vector regression in correcting UWB ranging measurements under LOS/NLOS conditions," in *CPS-IoTBench 2021 - Proceedings of the 2021 Benchmarking Cyber-Physical Systems and Internet of Things*, Association for Computing Machinery, Inc, May 2021, pp. 6–11. doi: 10.1145/3458473.3458820.

[23] V. Barral, C. J. Escudero, J. A. García-Naya, and P. Suárez-Casal, "Environmental cross-validation of NLOS machine learning classification/mitigation with low-cost uwb positioning systems," *Sensors (Switzerland)*, vol. 19, no. 24, Dec. 2019, doi: 10.3390/s19245438.

[24] M. Si, Y. Wang, H. Siljak, C. K. Seow, and H. Yang, "A lightweight CIR-based CNN with MLP for NLOS/LOS identification in a UWB positioning system," *IEEE Communications Letters*, 2023, doi: 10.1109/LCOMM.2023.3260953.

[25] A. Sitti Syathirah, R. Munadi, G. Indah Hapsari, and J. Telekomunikasi, "Enhancement of Accuracy For UWB Indoor Positioning Using Support Vector Regression."

[26] S. Kram, M. Stahlke, T. Feigl, J. Seitz, and J. Thielecke, "UWB channel impulse responses for positioning in complex environments: A detailed feature analysis," *Sensors (Switzerland)*, vol. 19, no. 24, Dec. 2019, doi: 10.3390/s19245547.

[27] J. Sun, S. Ma, H. Zhou, C. Du, and S. Li, "Measurement and Statistical Analysis of Distinguishable Multipaths in Underground Tunnels," *Int J Antennas Propag*, vol. 2020, 2020, doi: 10.1155/2020/2501832.

[28] X. Zheng *et al.*, "Full Parameter Time Complexity (FPTC): A Method to Evaluate the Running Time of Machine Learning Classifiers for Land Use/Land Cover Classification," *IEEE J Sel Top Appl Earth Obs Remote Sens*, vol. 14, pp. 2222–2235, 2021, doi: 10.1109/JSTARS.2021.3050166.

[29] A. G. Putrada, M. Abdurohman, D. Perdana, and H. H. Nuha, "CIMA: A Novel Classification-Integrated Moving Average Model for Smart Lighting Intelligent Control Based on Human Presence," *Complexity*, vol. 2022, 2022, doi: 10.1155/2022/4989344.

[30] G. I. Hapsari, R. Munadi, B. Erfianto, and I. D. Irawati, "Accuracy Improvement for Indoor Positioning Using Decawave on ESP32 UWB Pro with Display and Regression," *Journal of Robotics and Control (JRC)*, vol. 5, no. 3, pp. 851–862, 2024, doi: 10.18196/jrc.v5i3.20825.

[31] Cung Lian Sang, "Identification-of-NLOS-and-MP-in-UWB-using-ML (Dataset)," Identification of NLOS and Multi-Path Conditions in UWB Localization Using Machine Learning Methods, 2020. Accessed: Feb. 18, 2025. [Online]. Available: https://github.com/cliansang/Identification-of-NLOS-and-MP-in-UWB-using-ML