Novel Fine-Tuned Attribute Weighted Naïve Bayes NLoS Classifier for UWB Positioning

Fuhu Che[*], Qasim Zeeshan Ahmed[*], Fahd Ahmed Khan[†], and Faheem A. Khan[*].

Abstract—In this paper, we propose a novel Fine-Tuned attribute Weighted Naïve Bayes (FT-WNB) classifier to identify the Line-of-Sight (LoS) and Non-Line-of-Sight (NLoS) for UltraWide Bandwidth (UWB) signals in an Indoor Positioning System (IPS). The FT-WNB classifier assigns each signal feature a specific weight and fine-tunes its probabilities to address the mismatch between the predicted and actual class. The performance of the FT-WNB classifier is compared with the state-of-the-art Machine Learning (ML) classifiers such as minimum Redundancy Maximum Relevance (mRMR)- k-Nearest Neighbour (KNN), Support Vector Machine (SVM), Decision Tree (DT), Naïve Bayes (NB), and Neural Network (NN). It is demonstrated that the proposed classifier outperforms other algorithms by achieving a high NLoS classification accuracy of 99.7% with imbalanced data and 99.8% with balanced data. The experimental results indicate that our proposed FT-WNB classifier significantly outperforms the existing state-of-the-art ML methods for LoS and NLoS signals in IPS in the considered scenario.

Index Terms—UWB, ML, WNB, mRMR, and IPS.

I. INTRODUCTION

Mobile target tracking is extensively used in commercial, civil, industrial, and military applications, such as object tracking, autonomous vehicles, forklift control, military reconnaissance, etc [1], [2]. Ultra-wideband (UWB) technology, utilizes exploratory signals with an extremely large spectrum, that provides significant advantages in tracking and positioning [3], [4]. Beside this various alternative technologies have been extensively researched and employed in indoor scenarios such as Radio Frequency Identification (RFID), Bluetooth, WiFi fingerprinting, etc. Recently, [2], [5] has shown that WiFi with channel state information can be utilized for Indoor Positioning System (IPS) and have the same positioning cost and its accuracy is not inferior to UWB. Non-line-of-sight (NLoS) conditions are a major hindrance in achieving high positioning accuracy [6], [7]. In NLoS, the direct signal path between the transmitters and the receivers is blocked by the obstacles causing estimation error [4]. To deal with the impact of NLoS, several signal-feature-based Machine Learning (ML) classification methods have been proposed [7]–[10].

Non-parametric regression technique based on Support Vector Machine (SVM) and Gaussian process to estimate the ranging error was proposed in [3]. It was shown that for signals under NLoS conditions, 28% had a ranging error of less than one meter. Deep learning methods such as Convolutional Neural Network (CNN) and Long Short-Term Memory (LSTM) were employed in [8], [9]. The highest classification accuracy of 82.14% was achieved by employing CNN and stacked LSTM together [8]. Other ML techniques such as SVM, Random Forest (RF), and Multi-Layer Perceptron (MLP) were applied in [10] and an accuracy of 82.80%, 91.9%, and 91.20% was reported for these techniques, respectively. A semi-supervised learning approach has also been adopted for NLoS identification. It was shown that NLoS identification accuracy could be improved to 94.7% with the aid of semi-supervised SVM via leveraging unlabeled data [7]. An unsupervised ML approach based on Gaussian mixture models (EM-GMM) to identify the NLoS links from the unlabelled data was proposed in [11]. [12] employed the ranging and received signal strength (RSS) to identify the NLoS condition using LoS, NLoS soft, and NLoS hard classification. Furthermore, in [13] the power delay profile was exploited to accelerate the training of NN-based classifiers. However, the dataset was perfectly balanced. Finally, in [14], a different scenario was considered for testing with respect to the training for generalization of IPS.

Feature-based ML methods were well-studied in the aforementioned works, however, none of these works considered the correlation between different features of the signals. Moreover, in an imbalanced dataset, where a limited amount of NLoS samples are present, it is difficult for existing algorithms to train a robust model [8]. To address these shortcomings, we propose a novel feature-based Fine-Tuned attribute Weighted Naïve Bayes (FT-WNB) classifier. To the best of the authors' knowledge, no work has proposed signal-feature weighting to enhance NLoS classification with an imbalanced dataset. The main contributions of this work are as follows:

- A novel signal-feature-based FT-WNB algorithm for NLoS classification addresses class imbalance by assigning different weights to each feature and fine-tunes the classifier probabilities to improve classification accuracy.
- We conduct extensive experimentation to validate the superiority of the proposed FT-WNB classifier in comparison to the state-of-art algorithms. The performance is compared in terms of the confusion matrix, Receiver Operating Characteristics (ROC) curve, Area Under Curve (AUC), precision, recall, and classification accuracy.
- We show that the FT-WNB classification algorithm outperforms existing algorithms for different ratios of NLoS and LoS samples, even in the presence of imbalanced dataset.

[*]F. Che, Q. Z. Ahmed, and F. A. Khan is with the School of Computing and Engineering at the University of Huddersfield, United Kingdom. Email: {fuhu.che, q.ahmed, f.khan}@hud.ac.uk.
[†]F. A. Khan is with the School of Electrical and Computer Engineering, University of Oklahoma, Tulsa, OK, USA. Email: fahd.khan@ou.edu.
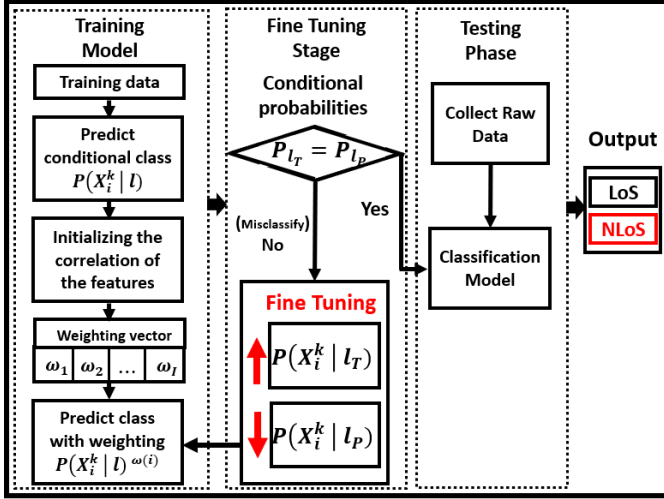
Fig. 1. Block diagram of Fine-tuned WNB algorithm.

Fig. 2. Experimental data collected in the studio.

II. FINE-TUNED WEIGHTED NAÏVE BAYES (FT-WNB) CLASSIFICATION ALGORITHM

A. UWB Range Theory

Time-of-Arrival (ToA) is used to calculate the estimated distance d between the tag and the anchor [9],

$$d = \begin{cases} c \times \text{ToA} + \epsilon, & \text{LoS} \\ c \times \text{ToA} + \epsilon + b, & \text{NLoS} \end{cases} \quad (1)$$

where, c is the speed of light in m/s, ToA is the propagation time taken from the tag to the anchor, ϵ represents the measurement noise and b is the NLoS measurement error caused by the blockages and obstacles.

B. Weighted Naïve Bayes

Naïve Bayes (NB) classifies the data with a corresponding maximum posterior probability. As features are assumed to be conditionally independent, the predicted class $\hat{l}^k$ at the k-th instance of data can be determined as

$$\hat{l}^k = \arg\max_l P(l) \prod_{i=1}^{I} P(X_i^k|l), \quad (2)$$

where $P(X_i)$ is the probability of the i-th attribute, $P(l)$ is the apriori probability of class l and I is the number of attributes. The underlying principle of WNB algorithm is that some attributes are of more importance than others [15]. This importance is specified by weighting the attributes and the predicted class, $\hat{l}^k$, in this case, is determined as

$$\hat{l}^k = \arg\max_l P(l) \prod_{i=1}^{I} P(X_i^k|l)^{w(i)}, \quad (3)$$

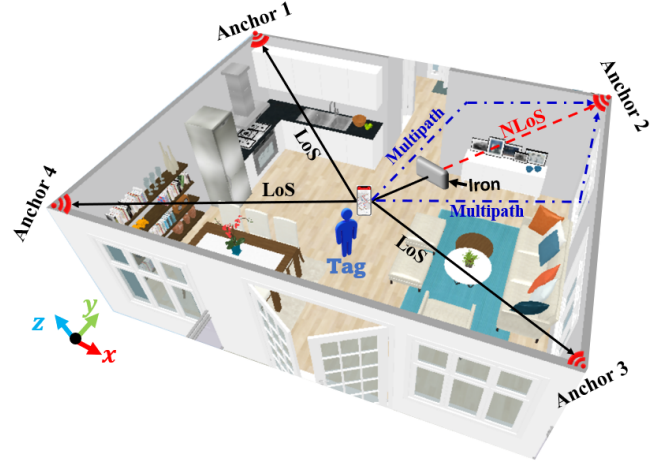where $w(i)$ is the weight associated with the i-th attribute.
.

C. FT-WNB Algorithm

The overall framework of the FT-WNB algorithm is described in Fig. 1. We first initialize the conditional probabilities as defined in (3) by applying the attribute weights to predict the class label. In the training phase, if the training instance is classified incorrectly, the predicted class l_P has a higher posterior probability than its true class l_T. Therefore, the conditional probabilities of signal features, $P(X_i^k|l_T)$ and $P(X_i^k|l_P)$ can be fine tuned in the next iteration as

$$P_{t+1}(X_i^k|l_T) = P_t(X_i^k|l_T) + \xi_{t+1}(X_i^k, l_T), \quad (4)$$
$$P_{t+1}(X_i^k|l_P) = P_t(X_i^k|l_P) - \xi_{t+1}(X_i^k, l_P), \quad (5)$$

where t represents the number of iteration and $\xi_{t+1}(\cdot)$ are the updated step size calculated as

$$\xi_{t+1}(X_i^k, l_T) = \beta\big(\alpha\big(\max P_t(X_i^k|l_T)\big) - P(X_i^k|l_T)\big)e(l_T, l_P), (6)$$
$$\xi_{t+1}(X_i^k, l_P) = \beta\big(\alpha P(X_i^k|l_P) - \min\big(P_t(X_i^k|l_P)\big)\big)e(l_T, l_P), (7)$$

where $\alpha, \beta \in (0, 1)$ controls the size of the updated step for $P_t(X_i^k|l_T)$ and $P_t(X_i^k|l_P)$, respectively. $\max(P_t(X_i^k|l_T))$ and $\min(P_t(X_i^k|l_P))$ are the maximum and the minimum value for $P_t(X_i^k|l_T)$ and $P_t(X_i^k|l_P)$, respectively. The error term, $e(l_T, l_P)$, is calculated as

$$e(l_T, l_P) = |P(l_T) - P(l_P)|. \quad (8)$$

Finally, the steps involved in the training and testing phase are summarised in Algorithms 1 and 2, respectively.

III. EXPERIMENTAL SETUP AND DATA COLLECTION

The required dataset was collected in a studio environment $(4.8m \times 3.5m)$. The floorplan of the infrastructure is shown in Fig. 2. Decawave® MDEK-1001 UWB kits, based on DW-1000 chip, with configuration as mentioned in [16], are used to generate the dataset. During data collection for the LoS scenario, there is no obstacle between the anchor and tag. For the NLoS scenario, an iron sheet is placed between the anchors and the tag. The plates are placed very close to the tag, generally less than a meter away. In this case, the first path of the signal transmission is either completely blocked or
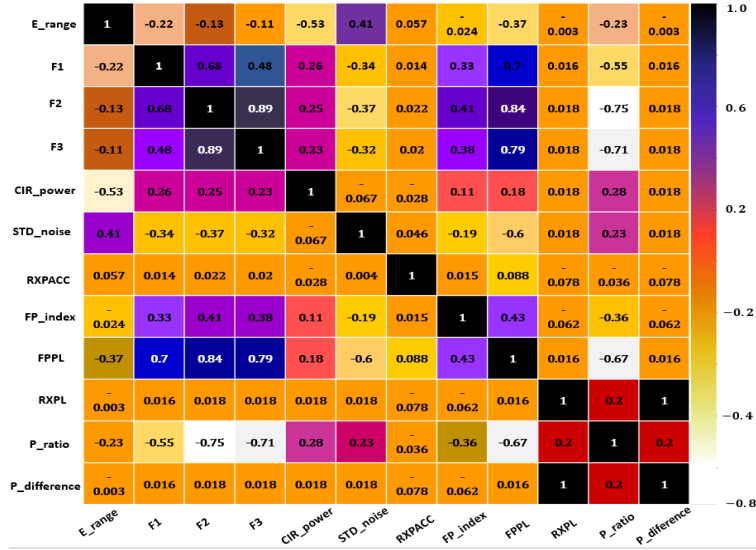
Fig. 3. Signal-features correlation matrix.

Algorithm 1 : Training Phase

Input: Training dataset consisting of signal features.
Output: The conditional probability vector
  Phase 1: Initializing conditional probabilities
  □ *Step 1:*
  for each X_i^k **do**
    calculate $\hat{l}^k$ by (2)
  end for
  for each weighted X_i^k **do**
    calculate $\hat{l}^k$ by (3)
  end for   Estimate the conditional probability vector
  Phase 2: Fine tuning conditional probabilities
  □ *Step 2:*
  if $l_T \neq l_P$ **then**
    Calculate the error by (8)
    Calculate the $\max(P_t(X_i^k|l_T))$ and $\min(P_t(X_i^k|l_P))$
    for each weighted figure X_i^k **do**
      calculate $\xi_{t+1}(X_i^k, l_T)$ by (6)
      update $P_{t+1}(X_i^k|l_T)$ by (4)
      calculate $\xi_{t+1}(X_i^k, l_P)$ by (7)
      update $P_{t+1}(X_i^k|l_P)$ by (5)
    end for
  end if
  if $l_T = l_P$ **then**
  end if
  return Classification model

Algorithm 2 : Testing Phase

Input: Test dataset with a mixture of LoS and NLoS signals
Output: Predicted class label of X_i^k
  Algorithm:
  1) Estimate the class membership probability $P(X_i^k|l)$.
  2) Predict the class label $\hat{l}^k$ of each class X_i^k.
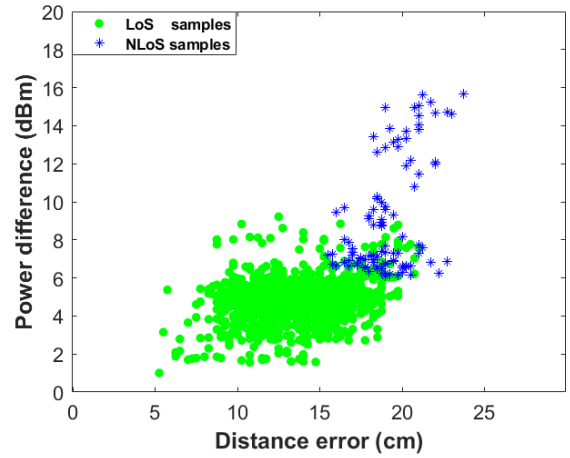 3) **Return** $\hat{l}^k$.

Fig. 4. Samples of Error vs. Power Difference

distorted through multiple reflections. As a result of this, the estimated range will be biased due to the signal propagation delay. From the collected dataset, we randomly select 100 NLoS and 1000 LoS signals to generate data imbalance. We consider two more features in addition to the 10 features of [17]: i) the index of the detected first-path (FP_index) and ii) the power ratio between the estimated received power and first-path power. The features and the correlation between them are illustrated in Fig. 3. As expected, the features are clearly not equally important nor independent. It can be observed that the features amplitude of the first path ($F1$), the amplitude of the second path ($F2$), and the amplitude of the third path ($F3$), as well as first path power level ($FPPL$), have a relatively higher correlation than others.

IV. PERFORMANCE EVALUATION

Visualization of our samples is shown in Fig. 4 where the power difference versus the distance error is plotted. There are 1000 LoS samples and 100 NLoS signals, respectively. In this figure, the green samples represent the LoS signals
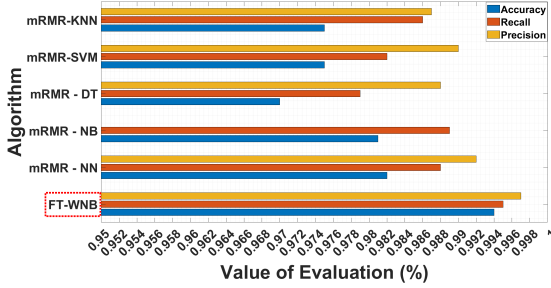
Fig. 5. Performance comparison for the proposed FT-WNB algorithm with state-of-the-art ML algorithms.

Fig. 6. Classification accuracy of the FT-WNB algorithm under imbalanced datasets having varying ratios of data.

TABLE I
RUNNING TIME AND CONFUSION MATRIX OF THE PROPOSED FT-WNB ALGORITHM AND STATE-OF-THE-ART ML ALGORITHMS

Algorithms	Running Time (s)	LoS CR	NLoS CR	TP	FN	FP	TN	Accuracy(%)
mRMR-KNN	0.049	98.6%	87%	986	14	13	87	97.5
mRMR-SVM	0.092	98.2%	90%	982	18	10	90	97.5
mRMR - DT	0.104	98.4%	89%	984	16	11	89	97.5
mRMR - NB	0.048	98.9%	90%	989	11	10	90	98.1
mRMR - NN	0.061	98.8%	92%	988	12	8	92	98.2
FT-WNB	0.041	99.5%	98%	995	5	2	98	99.4

Fig. 7. Comparisons for the existing UWB position methods.

and the blue color represents the NLoS signals. From the figure, it can be observed that few NLoS samples can be easily differentiated from LoS samples. While a large number of samples for NLoS cannot be easily differentiated from LoS samples. Therefore, for accurate classification, an appropriate algorithm is required.

The performance of the proposed FT-WNB classifier is compared with ML classifiers in Fig. 5. For these classifier, a subset of features is selected based on minimum redundancy maximum relevance (mRMR) feature selection. mRMR is a feature selection approach that tends to select features with a high correlation with the class and a low correlation between themselves to maximize relevance toward the features and simultaneously to minimize the redundancy among features [18]. The running time, confusion matrix and the correct rate (CR) is summarised in Table I. The confusion matrix depicts metrics, True Positive (TP), False Positive (FP), False Negative (FN) and True Negative (TN), respectively. In order to compare the classification performance, we plot precision, recall and classification accuracy in Fig. 5. It can be observed that the best performing algorithm among the existing ML algorithms is the mRMR-NN, which has a precision of approximately 99.2%, recall around 98.8% and overall classification accuracy is 98.2%. mRMR-NB follows closely and has similar performance. However, it is evident from Fig. 5, that the proposed FT-WNB algorithm outperforms all the state-of-the-art ML algorithms, i.e., mRMR-KNN, -SVM, -DT, -NB and -NN by achieving a precision of 99.7%, recall of 99.5% and an overall classification accuracy of 99.4%. From Table I it can be observed that the proposed FT-WNB algorithm, TN= 98 which indicates that 98% NLoS samples are correctly classified and only 2% NLoS samples are incorrectly classified. The classification performance is significantly better than the second best case which is 92% achieved by mRMR-NN algorithm. TP= 995, refers to LoS
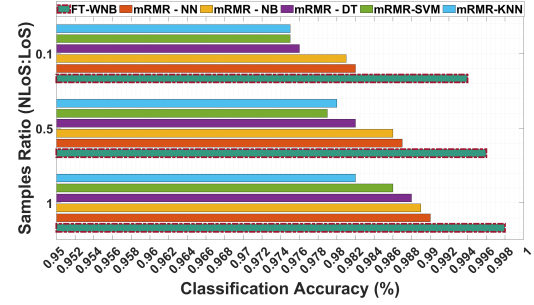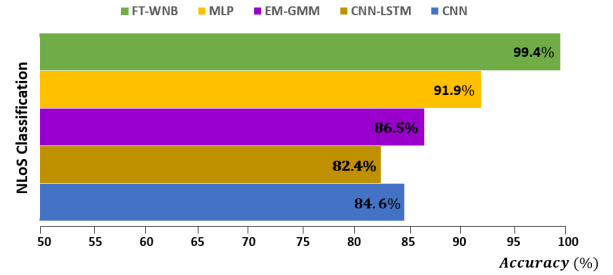
resulting in a correct rate of 99.5% which means 5 samples out of 1000 samples were inaccurately classified. Again, it is 1.3% higher than the second-best case. The average running time of FT-WNB is $0.041s$ which is better than other considered algorithms where the process of mRMR takes significant time. To sum up, in comparison to mRMR-KNN, mRMR-SVM, mRMR-DT, mRMR-NB, and mRMR-NN ML algorithms, it can be observed that the FT-WNB algorithm performs better for all performance metrics of the confusion matrix.

Fig. 6 shows the overall classification accuracy in the case of an imbalanced dataset. The ratio (NLoS: LoS) was set to 0.1, 0.5, and 1.0, respectively. It can be observed from Fig. 6 that as the ratio increases, the results of all algorithms improve because the model is trained better with equal number of LoS and NLoS samples. However, it is noticeable that the FT-WNB algorithm always maintains a higher classification accuracy for all considered ratios. This shows that the proposed FT-WNB classifier is highly robust and can assign appropriate weights to the desired features, guaranteeing better NLoS identification as compared to existing ML algorithms.

We also compares the proposed algorithm with the existing UWB NLoS identification methods as mentioned in the introduction section. The results in terms of accuracy are shown in Fig. 7. It can be clearly observed that the proposed FT-WNB algorithm performs better than MLP, EM-GMM, MLP, CNN-LSTM, and CNN methods which are some of the existing positioning methods in UWB IPS.

For the fine-tuning step, the impact of a number of selected samples is shown in Fig. 8. It can be observed that the accuracy of the algorithm improves as the number of samples increase and the proposed model benefits from the fine-tuning step. The classification accuracy improves from 98.2% to 98.6%
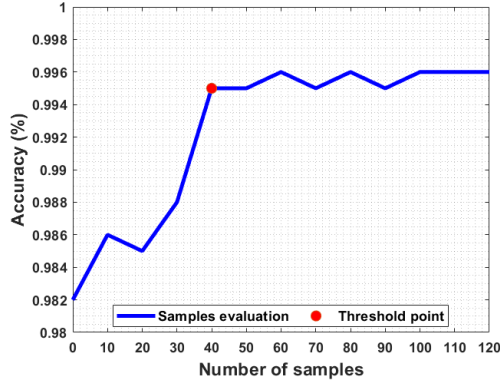
Fig. 8. Impact of number of samples evaluation

TABLE II
COMPARISON WITH [12].

Algorithms	Processing Time (s)	% LoS Accuracy	% NLoS Accuracy	% Total Accuracy
Range & RSS	0.0371	97.4	91	96.8
All features	0.041	99.5	98	99.4
Difference	3.9 ms	2.1 (%)	7 (%)	2.6 (%)

TABLE III
PERFORMANCE OF FT-WNB IN A DIFFERENT ENVIRONMENT.

Training	Testing Scenario	TP Scenario	FP	FN	TN	Accuracy (%)
Studio	Studio	995	5	2	98	99.4
Studio	Room	990	10	5	95	98.6

by collecting the first 10 samples and then further increases to 99.5% by just collecting the first 40 samples. However, there is no significant gain in accuracy beyond that point. Therefore, in our simulations, the number of samples required for fine-tuning is fixed at 40.

There is a tradeoff between the collected features and accuracy. In this paper, our focus was to collect a maximum number of features that yield the highest accuracy. In Table II, we present the comparison of the processing time considering only two features (Range and RSS [12]) and all the 12 gathered features. It can be observed that a difference of 3.9ms leads to an improved accuracy of 99.4% as compared to 96.8%. However, if the application does not require high accuracy, we could reduce the number of collected features and improve the processing time of the algorithm.

Our algorithm has been tested and trained in various scenarios as adopted in [14]. It can be observed from Table III that the proposed WNB has a classification accuracy of over 98%, even after being trained in a studio and tested in a room measuring $(4.8m \times 5.4m)$. This demonstrates the versatility of FT-WNB and its ability to perform in different environments, regardless of where it was trained.

V. CONCLUSION

In this work, we propose the FT-WNB algorithm which is a feature-based NLoS classification method to improve classification accuracy by automatically weighting different signal features, for imbalanced LoS and NLoS samples. The performance of the FT-WNB classifier is compared with state-of-the-art machine learning algorithms using simulation. It is shown that the proposed FT-WNB algorithm performed better in terms of confusion matrix, ROC, and AUC. In addition, the proposed FT-WNB algorithm adapts better to the different ratios of LoS and NLoS signal data. Our findings demonstrate that the proposed FT-WNB classifier efficiently alleviates the influence of the signal features in an imbalanced dataset in a harsh mixed LoS and NLoS scenario.

REFERENCES

[1] F. Zafari, A. Gkelias and K. K. Leung, "A Survey of Indoor Localization Systems and Technologies," IEEE Communications Surveys & Tutorials, vol. 21, no. 3, pp. 2568-2599, thirdquarter 2019.

[2] P. Roy and C. Chowdhury, "A survey on ubiquitous WiFi-based indoor localization system for smartphone users from implementation perspectives," CCF Trans. on Pervasive Computing and Interaction, 1-21, 2022.

[3] H. Wymeersch et.al., "A Machine Learning Approach to Ranging Error Mitigation for UWB Localization," IEEE Transactions on Communications, vol. 60, no. 6, pp. 1719-1728, June 2012.

[4] F. Che et al., "Feature-Based Generalized Gaussian Distribution Method for NLoS Detection in Ultra-Wideband (UWB) Indoor Positioning System," IEEE Sensors Journal, vol. 22, no. 19, pp. 18726-18739, 2022.

[5] Z. Zhang, H. Du, S. Choi and S. H. Cho, "TIPS: Transformer Based Indoor Positioning System Using Both CSI and DoA of WiFi Signal," IEEE Access, vol. 10, pp. 111363-111376, 2022.

[6] W. Wang, D. Marelli and M. Fu, "Multiple-Vehicle Localization Using Maximum Likelihood Kalman Filtering and Ultra-Wideband Signals," IEEE Sensors Journal, vol. 21, no. 4, pp. 4949-4956, 2021.

[7] T. Wang, K. Hu, Z. Li, K. Lin, J. Wang and Y. Shen, "A Semi-Supervised Learning Approach for UWB Ranging Error Mitigation," IEEE Wireless Communications Letters, vol. 10, no. 3, pp. 688-691, March 2021.

[8] C. Jiang, J. Shen, S. Chen, Y. Chen, D. Liu and Y. Bo, "UWB NLOS/LOS Classification Using Deep Learning Method," IEEE Communications Letters, vol. 24, no. 10, pp. 2226-2230, Oct. 2020.

[9] M. Stahlke, S. Kram, F. Ott, T. Feigl and C. Mutschler, "Estimating TOA Reliability With Variational Autoencoders," IEEE Sensors Journal, vol. 22, no. 6, pp. 5133-5140, 2022.

[10] C.L. Sang et al., "Identification of NLOS and Multi-Path Conditions in UWB Localization Using Machine Learning Methods," Applied Sciences, vol. 10, no. 11, pp. 3980-4005, June 2020.

[11] J. Fan and A. S. Awan, "Non-Line-of-Sight Identification Based on Unsupervised Machine Learning in Ultra Wideband Systems," IEEE Access, vol. 7, pp. 32464-32471, 2019.

[12] V.Barralet al., "NLOS identification and mitigation using low-cost UWB devices," Sensors, 19(16), 3464, 2019.

[13] V. B. Vales, T. Domínguez-Bolaño, C. J. Escudero and J. A. García-Naya, "Using the Power Delay Profile to Accelerate the Training of Neural Network-Based Classifiers for the Identification of LOS and NLOS UWB Propagation Conditions," IEEE Access, vol. 8, pp. 220205-220214, 2020.

[14] 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", in Sensors 2019, 19, 5438.

[15] S. Ruan, H. Li, C. Li and K. Song, "Class-Specific Deep Feature Weighting for Naïve Bayes Text Classifiers," IEEE Access, vol. 8, pp. 20151-20159, 2020.

[16] DW1000 User Manual. Available online: https://www.decawave.com/sites/default/files/resources

[17] F. Che et al., "Weighted Naive Bayes Approach for Imbalanced Indoor Positioning System Using UWB," IEEE International Black Sea Conference on Communications and Networking (BlackSeaCom), 2022, pp. 72-76.

[18] A. Samat, E. Li, P. Du, S. Liu and J. Xia, "GPU-Accelerated CatBoost-Forest for Hyperspectral Image Classification Via Parallelized mRMR Ensemble Subspace Feature Selection," IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, vol. 14, pp. 3200-3214, 2021.