# Improving the Correction of NLoS-Induced Ranging Errors in UWB Systems through Enhanced Labeling

Elisei Ember[1,2], Jesus Pestana[1], Michael Krisper[1,2], Michael Stocker[2], Kay Römer[2], Carlo Alberto Boano[2], and Pablo Corbalán Pelegrín[3]

[1]Pro2Future GmbH, Graz, Austria [2]Institute of Technical Informatics, Graz University of Technology, Graz, Austria
[3]NXP Semiconductors Austria GmbH & Co KG, Gratkorn, Austria

Email: {name.surname}@pro2future.at ; {michael.stocker, roemer, cboano}@tugraz.at ; pablo.corbalanpelegrin@nxp.com

***Abstract*—Ultra-wideband (UWB) ranging estimates taken in Line-of-Sight (LoS) conditions are typically centimeter accurate, but large errors may occur when devices are in Non-Line-of-Sight (NLoS). Several techniques, often based on machine learning, have been proposed to correct NLoS-induced ranging errors. These techniques are typically designed to effectively correct ranging estimates obtained in NLoS conditions, but significantly worsen the ranging estimates taken in LoS. We show that this problem stems from the method commonly used to label LoS/NLoS conditions and to train classification and regression models, which results, among others, in error correction techniques being applied to LoS samples that are erroneously classified as NLoS. To tackle this problem, we propose ranging error labeling (RELa); an approach that defines more suitable labels, based on the severity of the ranging error, to guide the decision of whether applying ranging error correction and the selection of which data should be used to train the error correction model. The effectiveness of RELa hinges on the careful selection of two thresholds: one identifying which ranging estimates should be corrected ($\alpha$), and one defining the training set used for the error correction model ($\beta$). After deriving suitable values for these two thresholds empirically and showing their applicability to different settings and UWB platforms, we evaluate and prove the effectiveness of RELa. Our experimental results show that RELa reduces the error introduced on ranging estimates taken in LoS by up to 99%, without significantly impacting the effectiveness of ranging error correction in NLoS conditions.**

***Index Terms*—Channel impulse response, UWB, NLoS conditions, Ranging correction, Machine learning, Embedded systems.**

## I. Introduction

The popularity of UWB systems is soaring, driven by their increasing integration into modern smartphones [1] and adoption in both industrial [2] and automotive [3] applications. UWB radios spread the signal energy across a large bandwidth ($\geq 500\,\mathrm{MHz}$) by transmitting short signal pulses ($\approx 2\,\mathrm{ns}$). This provides an excellent time resolution allowing UWB receivers to precisely derive the time-of-arrival (ToA) of a signal and estimate the distance between two devices, which is pivotal for accurate ranging and localization [4].

Whilst UWB ranging generally achieves cm-level accuracy in Line-of-Sight (LoS) conditions, its performance degrades in Non-Line-of-Sight (NLoS) conditions, which may introduce errors up to several meters [4], [5]. To improve the accuracy of UWB ranging in NLoS, state-of-the-art (SoA) approaches [5], [6] commonly follow a two-stage approach: first, they classify whether devices are in LoS or NLoS; then, they correct the ranging estimates in case of NLoS conditions, typically using regression-based machine learning (ML) models [7], [8].

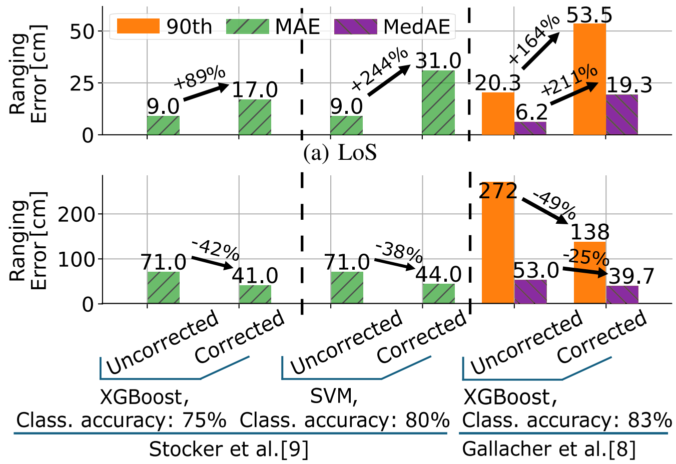



Fig. 1: Ranging error when applying (*corrected*) or not applying (*uncorrected*) SoA ranging correction methods [8], [9]. The correction is clearly effective in NLoS conditions (b), but has a negative impact on ranging estimates taken in LoS (a). *90$^{th}$*: 90$^{th}$ percentile of the absolute ranging error; *MAE:* mean absolute error; *MedAE:* median absolute error.

**Limitations of existing solutions.** SoA approaches are designed to effectively correct ranging estimates obtained in NLoS conditions [7]–[11], but tend to affect negatively ranging estimates taken in LoS, due to misclassification of LoS samples. Fig. 1 illustrates this problem using concrete examples from the literature based on either support vector machines (SVMs) with handcrafted features extracted from the channel impulse response (CIR) or extreme gradient-boosted trees (XGBoost) fed directly with raw CIR samples [8], [9]. Whilst these approaches can effectively reduce NLoS-induced ranging errors by 25–49% (Fig. 1b), they also significantly worsen the ranging error in LoS conditions (Fig. 1a). For example, Gallacher et al. [8] have shown that the median absolute error (MedAE) of ranging estimates taken in LoS increases from 6.2 to 19.3 cm (+211%), whereas the 90$^{th}$ percentile of the absolute ranging error (90$^{th}$) in LoS – which captures how far the error is spreading – increases from 20.3 to 53.5 cm (+164%). This is due to ranging correction being applied to LoS samples that are misclassified as NLoS. Note that the accuracy of the LoS/NLoS classification used to infer if the correction should be applied is between 75 and 83%.

**The gap to fill.** These results highlight that the impact of existing error correction approaches on ranging estimates taken in LoS conditions is *not negligible*. We hence aim to find a way to effectively correct ranging estimates taken in NLoS, but *without affecting* significantly those taken in LoS. After analyzing one of the most comprehensive public UWB datasets in the literature (§III), we found that every fourth LoS sample is commonly misclassified as NLoS (i.e., it undergoes error correction, which lowers the estimated distance by a certain amount). This results in the estimated distance of misclassified LoS samples (which commonly have a relatively small ranging error of $\pm 25$ cm) being lowered by up to 65 cm after applying SoA correction methods (i.e., there is an unnecessary over-correction of ranging estimates that have small errors).

**Contributions.** We show that the problem lies in the method that is commonly used to label LoS/NLoS conditions (we refer to this labeling strategy as *traditional labeling (TraLa)*) and to train classification and regression ML models used to correct the ranging estimates. In this work, we propose instead *Ranging Error Labeling* (RELa), a method that employs labels *based on the severity of the ranging error* to train NLoS classification and error correction models (§IV). Specifically, RELa defines two ranging error thresholds: $\alpha$ for classification (to inform whether applying error correction), and $\beta$ for regression (to select the data used to train the correction model).

As the effectiveness of RELa depends on the selection of these two thresholds, we show how to empirically derive suitable values for $\alpha$ and $\beta$ through the use of a specifically-designed cost function (§V). We benchmark RELa on four datasets: one that is publicly available [7], [12], and three that we collected using UWB devices from different manufacturers (Qorvo and NXP). Our results indicate that the use of RELa allows us to achieve a ranging error correction performance comparable to TraLa, while *not affecting* (or affecting in a negligible way) ranging estimates taken in LoS. That is, RELa effectively corrects ranging estimates taken in NLoS, while reducing the negative impact of error correction on LoS ranging estimates by up to 99% compared to SoA approaches. Our results also show that we can derive generic values for $\alpha$ and $\beta$ that perform well across the four datasets considered in our study, regardless of the employed UWB platform (§VI).

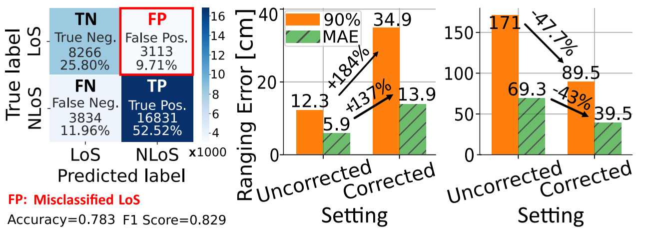
## II. State of the Art Review

UWB receivers can precisely estimate the time-of-arrival (ToA) of a received signal by identifying the direct path in the estimated CIR. This precise ToA estimation is essential for determining the time-of-flight (ToF) of the radio signal between two devices, which in turn allows for accurate distance (ranging) measurements. In LoS cases, i.e., when no obstacle is present between two devices, the direct path is correctly recognized as the first peak in the estimated CIR exceeding a given noise threshold. In NLoS cases, the direct path may be attenuated by the presence of some obstacle(s) and deemed as noise: as a result, the UWB receiver may erroneously identify later peaks in the CIR that correspond to reflections from walls and scattering from other objects as the direct path. This introduces an error in the ToA estimates and skews the ranging measurements, a well-known problem in indoor environments where NLoS conditions are common [13].

**NLoS detection and correction.** Many techniques have been developed to identify (i.e., classify) whether UWB ranging occurs in NLoS conditions [5], [6], [14], [15] and to correct ranging estimates taken in NLoS [7], [13][1]. Numerous works leverage ML solutions based on *manually-crafted* features for NLoS classification and ranging error correction [16], [17]. For example, Maranò et al. [13] and Stocker et al. [7] make use of six to nine features in conjunction with SVMs. How to improve feature selection has also been the goal of several studies, in order to optimize accuracy [7], [11], and to reduce computation times by processing less CIR samples [8] and by reducing the feature extraction times [18]. The community has also investigated ML solutions based on *self-learning* features, including deep neural networks (DNNs) [10], [19] and transformer-based classifiers [20]. In this context, automatic feature extraction by directly inputting the CIR to XGBoost decision trees can also be leveraged to create lightweight models [8]. The latter can run on constrained embedded devices and are hence becoming increasingly more attractive to reduce latencies and avoid off-loading tasks to edge devices [18], [21].

**Enhanced labeling.** Existing works commonly use manually-annotated LoS/NLoS labels. A few works have created a third label to identify ranging measurements taken in "weak" NLoS conditions (WLoS), i.e., in scenarios where the presence of obstacles of certain size or material (such as wooden chairs or PC monitors) introduce a relatively small ranging error [7], [22]. However, if and how this additional class can be leveraged to improve NLoS classification and error correction performance is not fully clear: for example, in [7] WLoS samples have not been considered for training, whereas in follow-up work [9] the WLoS and NLoS classes have been merged and evaluated together. Other works have proposed to use labels based on the ground truth ranging error for classification. Stocker et al. [9] have observed on-par classification performance when using TraLa and synthetic labels derived by applying a threshold on the ranging error. In [11] and [23], data was labeled as LoS when the true distance was below 25 or 30 cm. Kim et al. [24] have split the ranging samples in decile classes according to the ranging error (i.e., in ten classes with equal numbers of samples), providing an extended Kalman filter framework with the uncertainty range of the ranging estimate as given by the estimated decile class of the sample. This was shown to achieve an almost-perfect classification performance [25], but its use for NLoS-induced error correction [24] was not explicitly investigated. In contrast, in this work, we are the first to show that the use of labels embedding the severity of the ranging error can significantly benefit the performance of NLoS-induced error correction, especially in LoS conditions.

[1] Please note that several methods in the literature focus on leveraging consecutive ranging estimates for NLoS classification [5], [14], or mitigate the NLoS problem by leveraging device diversity (e.g., by performing a selection of the best available anchors at runtime [8]). In this study, we instead focus on the NLoS classification and correction of *individual* ranging estimates.




(a) Confusion matrix (b) LoS (c) NLoS

Fig. 2: Classification and error correction performance of XGBoost [8] evaluated on the STOCKER CPS dataset using TraLa.

## III. DEMYSTIFYING NLOS-INDUCED RANGING ERRORS

As teased in Fig. 1, SoA error correction approaches based on TraLa negatively affect ranging estimates taken in LoS, which are commonly centimeter accurate. In this section, we identify the root cause of the problem by re-implementing SoA models (specifically, the lightweight XGBoost models from [8]) and by evaluating their performance in detail on publicly-available datasets (specifically, the STOCKER CPS dataset [7], [12]).

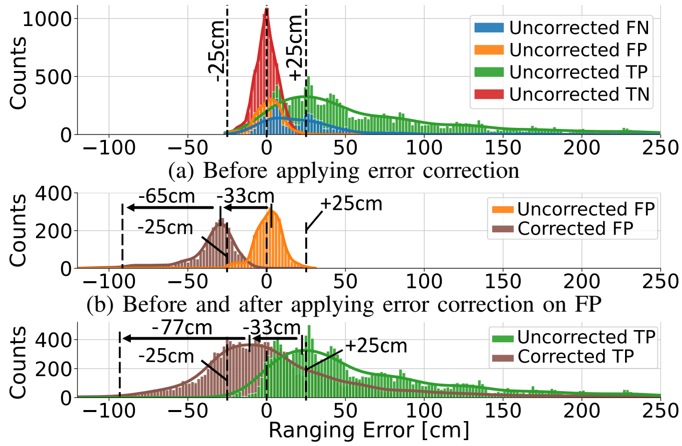
**Classification performance.** Fig. 2a shows the classification performance of the re-implemented XGBoost classifier based on the classifier's predictions and the ground truth labeling as a confusion matrix. The latter splits the data into four cases:

- true negatives (TN): correctly classified LoS samples,
- false positives (FP): LoS cases misclassified as NLoS,
- false negatives (FN): NLoS cases misclassified as LoS,
- true positives (TP): correctly classified NLoS samples.

The ranging correction is applied to samples predicted as NLoS, i.e., the FP and TP. It follows that: *(i)* the improvement of the NLoS ranging performance is achieved through ranging correction on the TP samples, whereas *(ii)* the negative effects on LoS ranging are caused by applying erroneous ranging corrections to the FP through a model trained on NLoS samples. We observe a FP rate of 9.71 % (see Fig. 2a), which corresponds to 27.36% of all LoS samples, i.e., roughly every fourth LoS sample is misclassified as NLoS.

Fig. 3a shows the distribution of the ranging error for each of the four cases identified in the confusion matrix (TN, FP, FN, TP) before applying error correction (uncorrected). We can notice that a significant amount of NLoS samples (50% of FN and 33% of TP) have ranging errors in the same range as LoS samples (TN and FP). This shows that the traditional LoS and NLoS labels only refer to whether the direct path was visually free or blocked by an obstacle, and do not accurately reflect whether the ranging accuracy is affected. Whilst the impact of NLoS cases misclassified as LoS (FN) on the final ranging estimates is not severe (as they do not get corrected and have a MedAE of only 18.9 cm), the impact of FP and TP is much more relevant, as discussed next.

**Error correction performance.** Fig. 2b and 2c show the mean absolute ranging error (MAE) and the $90^{th}$ percentile of the absolute ranging error ($90^{th}$) before (uncorrected) and after (corrected) applying error correction using XGBoost with TraLa on LoS and NLoS samples, respectively. Because error correction is applied to almost every fourth LoS sample, the MAE in LoS conditions increases from 5.9 to 13.9 cm (+137%), whereas the $90^{th}$ increases from 12.3 to 34.9 cm (+184%). Fig. 3b and 3c show the distribution of the ranging error for FP and TP, respectively, before and after applying error correction. When LoS samples are misclassified (FP case), they undergo error correction, which lowers their (originally rather accurate) ranging estimates by up to 65 cm. NLoS samples (TP) with a small ranging error are affected in a similar way, with unnecessary over-corrections of the ranging estimates resulting in errors up to -100 cm.




(c) Before and after applying error correction on TP

Fig. 3: Ranging error distribution as a function of LoS/NLoS (mis)classification before and after applying error correction.

**Takeaways.** These results highlight the limits of TraLa and confirm the need to improve the performance of ranging error correction schemes following the traditional two-stage approach. First, the classifier used to inform whether error correction should be applied needs to identify ranging samples with high error (so to avoid that correction is applied to ranging samples with low error). Second, to avoid over-correcting ranging samples with relatively low error, the error correction model (i.e., the regression model) needs to be trained using a high amount of ranging samples with low error, possibly by including samples taken in LoS conditions in the training set.

## IV. RANGING ERROR LABELING (RELA)

Building upon these results, we design RELa: a novel labeling method that improves the performance of classifier and regression models used to correct NLoS-induced ranging errors.

RELa uses two thresholds that capture the severity of the ranging error: $\alpha$ and $\beta$, as shown in Fig. 4. $\alpha$ splits the data employed to train the classifier (used to inform whether error correction needs to be applied) into the low ranging error (LoRE) and high ranging error (HiRE) classes. Samples classified as HiRE undergo correction, whereas samples classified as LoRE are not corrected. $\beta$ is used to select which data should be used to train the regression model: all samples with ranging error $\geq \beta$ are included in the training set.

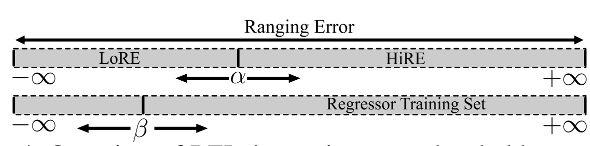



Fig. 4: Overview of RELa's ranging error thresholds $\alpha$ and $\beta$.

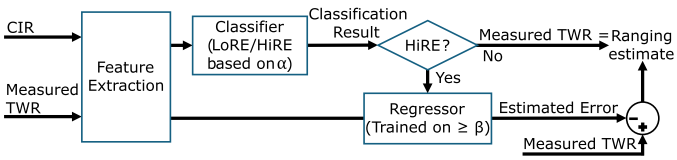



Fig. 5: Two-stage ranging error correction using RELa.

Fig. 5 summarizes how to apply RELa when performing ranging error correction. Ranging samples classified as HiRE (where $\alpha$ specifies what is considered high or low ranging error) are corrected by a regression model trained on samples with a ranging error $\geq \beta$. Both classifier and regression models use features extracted from the CIR (either manually or automatically), as commonly done in the literature (see §II).

## V. RELa: The Role of $\alpha$ and $\beta$

The effectiveness of RELa hinges on the careful selection of its two thresholds. We hence need to study the classification and regression performance of SoA models as a function of different $\alpha$ and $\beta$ values, and compare it with the use of TraLa. After detailing our empirical methodology (§V-A), we present the results characterizing RELa's performance as a function of $\alpha$ (§V-B) and $\beta$ (§V-C). We then perform an ablation study comparing the use of RELa and TraLa (and their combinations) to train classifier and regression models (§V-D).

### *A. Experimental Setup & Methodology*

**Models.** We perform our study using the XGBoost classification and regression models based on decision trees (DTs) proposed by Gallacher et al. [8], as these represent the SoA of lightweight ML models that can be implemented on constrained UWB devices. These models are fed with raw CIR samples, and autonomously learn the set of relevant features. We parametrize these models following the authors' guidelines (i.e., with a maximum of 30 trees and a maximum depth of 3 [8]), which results in compact models of 7 and 12 kB for classification and regression, respectively, after quantization.

**Datasets.** We evaluate in detail the role of $\alpha$ and $\beta$ using the Stocker CPS dataset [7], [12], which is one of the most comprehensive publicly-available datasets of UWB ranging measurements in LoS and NLoS conditions[2], and the same one used for our initial analysis in §III. Note that, although we only present results for Stocker CPS and XGBoost in this section due to space limitations, similar trends are observed for different datasets and models, as discussed in §VI.

**Data preparation.** UWB ranging estimates often contain a small constant bias that is device-dependent. We estimate and correct this bias by considering all LoS samples taken at a certain distance (3 m), and by computing the corresponding deviation in the ranging estimates. For the Stocker CPS dataset, we have observed (and subtracted) a bias of 3.41 cm. We filter ranging outliers as in [8], i.e., by calculating the median of measurements taken in the same position, and by removing samples that exceed `2.5x` the standard deviation.

[2] In this dataset, data is labeled using three classes: LoS, WLoS, and NLoS (see §II). Hence, we merge WLoS and NLoS samples together as done in [9].

TABLE I: $\omega_{c,m}$ factors for the Stocker CPS dataset.

| Metric | Line-of-Sight (LoS) exp. best [cm] | required [cm] | $\omega_{c,m}$ [cm] | Non-Line-of-Sight (NLoS) exp. best [cm] | required [cm] | $\omega_{c,m}$ [cm] |
|---|---|---|---|---|---|---|
| $90^{th}$ | 12.29 | 15.36 | 3.07 | 89.49 | 89.44 | 8.95 |
| MAE | 5.88 | 7.35 | 1.47 | 39.5 | 43.45 | 3.95 |

To tackle dataset imbalance, we perform a *stratified 5-fold split* to divide the dataset into training and testing sets that are different from each other, but that share the same distribution of LoS/NLoS labels. We do this using `StratifiedGroupKFold` (a Python `scikit-learn` library), and by performing the group split based on characteristics of the dataset (e.g., room and setup ID) that correspond to different real-world environments.

For each ranging sample, we scale the CIR and align the position of the first peak at 20 ns: we then feed the classifier and regression models with the first 60 ns of the CIR.

**Performance metrics.** We evaluate performance using the same metrics used in §III. That is, we analyze the distribution of TP, TN, FP, and FN after performing classification, and we compute the mean absolute ranging error (MAE) as well as the $90^{th}$ percentile of the absolute ranging error ($90^{th}$) before and after applying error correction. We calculate performance separately for measurements taken in LoS and NLoS to allow a direct comparison with SoA approaches using TraLa.

To quantitatively rank the analyzed models based on their error correction performance, we define a *cost function* (Eq. 1) that accounts for the sustained MAE and $90^{th}$ in both LoS and NLoS conditions, and privileges models having a good performance across both conditions. The costs are defined with respect to a given list of models to be compared (e.g., XGBoost trained with different values of $\alpha$ and $\beta$) per dataset. From this list, the lowest value for each of the four performance metrics (MAE and $90^{th}$ in LoS and NLoS) is used as a reference to compute a positive cost (Eq. 2). Specifically:

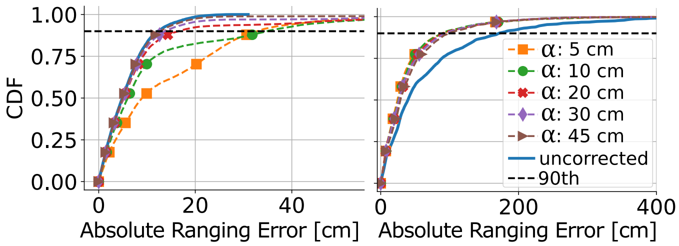
$$cost(model, list) = \sum_{m} \sum_{c} \frac{\Delta M_{c,m}}{\omega_{c,m}} \tag{1}$$

$$\Delta M_{c,m} = |M_{c,m}(model) - M_{c,m}^{best}(list)| \tag{2}$$

$$\omega_{c,m} = |M_{c,m}(required) - M_{c,m}(exp.\,best)| \tag{3}$$

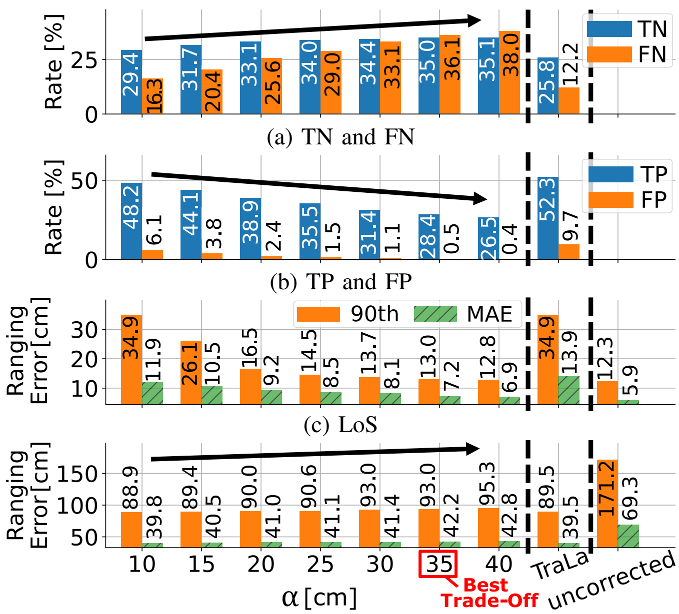
- $M_{c,m}(model)$ is the value of the performance metric $m$ achieved by $model$ under the ranging condition $c$, with $m$ being MAE or $90^{th}$, and $c$ being LoS or NLoS.
- $M_{c,m}^{best}(list)$ is the best value for the metric $m$ under the ranging condition $c$ achieved by the models in the list.
- $\omega_{c,m}$ is a constant scaling factor that represents the relative weightings between cost terms.

Suitable values of $\omega_{c,m}$ are derived using Eq. 3. An example of their calculation for the Stocker CPS dataset is shown in Table I. For LoS conditions, the $expected\,best$ value is extracted from the performance metrics computed over the uncorrected samples. For NLoS conditions, the $expected\,best$ value is extracted from the performance metrics computed with models using TraLa. The $required$ value is set to be 25%




(a) LoS (b) NLoS

Fig. 6: CDF of the ranging error as a function of $\alpha$.




(a) TN and FN

(b) TP and FP

(c) LoS

(d) NLoS

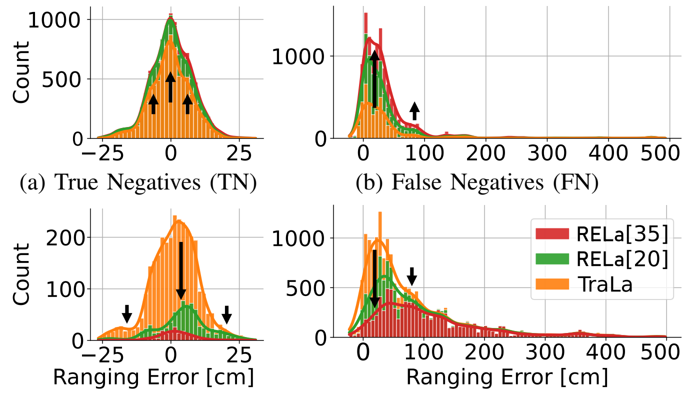
Fig. 7: Classification and error correction performance as a function of $\alpha$. The regression model is trained on the classifier's prediction (TrPr).

and 10% higher than the *expected best* for LoS and NLoS conditions, respectively: we choose these values empirically, as they allow us to balance the importance of having a good performance in LoS (a low error compared to the uncorrected value) and in NLoS (a low error compared to TraLa).

### B. Parametrization of $\alpha$

We perform a sweep over different values of $\alpha$ (between 0 and 50 cm), and train the XGBoost classifier accordingly. To account only for the role of $\alpha$, we train the regression model based on the classifier's prediction (TrPr), and not based on $\beta$. That is, we train the classifier using LoRE/HiRE labels, and the regression model using the samples classified as HiRE.

Fig. 6 shows the cumulative distribution function (CDF) of the ranging error in LoS and NLoS conditions as a function of $\alpha$. In LoS, the higher the value of $\alpha$, the closer is the ranging error to the original value before applying correction (uncorrected). In NLoS, the differences in the ranging error for different $\alpha$ values are negligible. Fig. 7c and 7d show the MAE and 90$^{th}$ as a function of $\alpha$, and directly compare them with those obtained using TraLa (i.e., using XGBoost trained on classical LoS/NLoS labels) and with those obtained before applying correction (uncorrected). For $\alpha \geq 20$, the 90$^{th}$ of the ranging error in LoS conditions is less than 4.2 cm higher than before applying correction; in contrast, when using TraLa, it grows by up to 22.6 cm. The ranging error in NLoS conditions is relatively stable (and comparable to that obtained using TraLa) independently of the chosen $\alpha$, although lower values of $\alpha$ exhibit a slightly better performance.




(a) True Negatives (TN) (b) False Negatives (FN)

(c) False Positives (FP) (d) True Positives (TP)

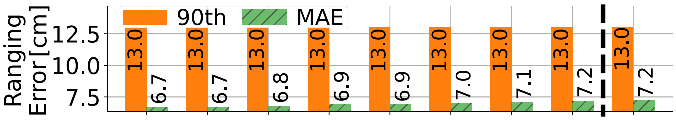
Fig. 8: Distribution of the ranging error before applying correction when using RELa (with different $\alpha$) and TraLa.

Fig. 8 shows the distribution of the ranging error before applying error correction for each of the possible outcomes of the classification (TN, FN, FP, TP) when using RELa with $\alpha = 20$, RELa with $\alpha = 35$, and TraLa. As $\alpha$ increases, the number of FP and TP samples decreases (see Fig. 8c and 8d), while the number of TN and FN samples increases (Fig. 8a and 8b). This is also visible in Fig. 7b and 7a, respectively. For models based on RELa, the majority of the FN cases are in the range [-20, +50] cm, with only a few samples in the range [+50, +100] cm, and a negligible amount above 100 cm (see Fig. 8b). As discussed in §III, this is desirable, as samples with a low ranging error should rather be left uncorrected.

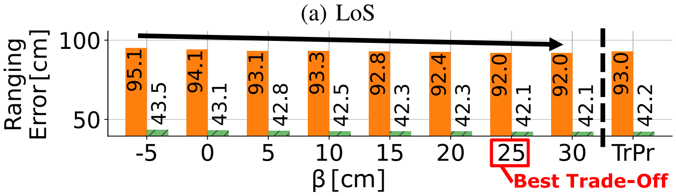
According to the cost function defined in Eq. 1, among the studied models, those with $25 \leq \alpha \leq 35$ achieve the best performance. First, they do not worsen ranging estimates in LoS significantly (as they minimize the number of FP to 0.5–1.5% only). Second, they exhibit only a relative drop in NLoS correction performance (≈5 %) compared to the use of TraLa. This competitive correction performance is achieved at an apparently low TP rate of 28–35%, i.e., ≈50% of the NLoS samples (in contrast, when using TraLa, the TP rate is 52.3 %). This indicates that, to maximize the ranging correction performance, it is more important to minimize the number of FP (i.e., avoiding the misclassification of LoS samples) than maximizing the number of TN+TP (i.e., ensuring that all NLoS samples are sent for correction).

### C. Parametrization of $\beta$

We fix $\alpha$ to the best value obtained in §V-B (i.e., $\alpha = 35$), and perform a sweep over different values of $\beta$. We denote the use of RELa with specific values for $\alpha$ and $\beta$ as $RELa[\alpha, \beta]$. Fig. 9 shows the ranging correction performance as a function of $\beta$ and also compares it with that obtained when training the

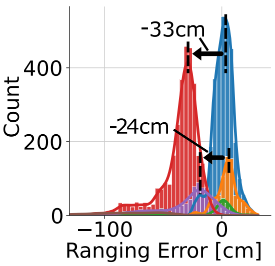



(a) LoS

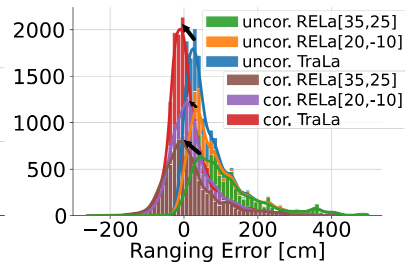



(b) NLoS

Fig. 9: Ranging correction performance metrics for RELa[$\alpha$=35, $\beta$] models, RELa regression $\beta$-sweep.




(a) False Positives (FP)

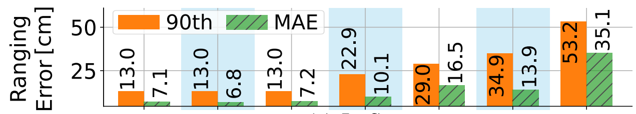



(b) True Positives (TP)

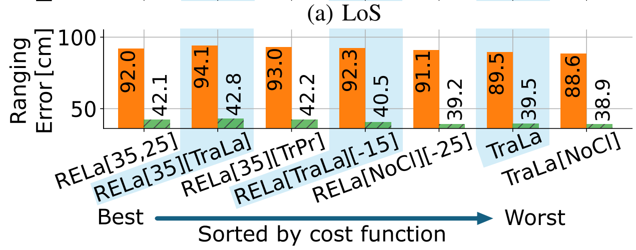
Fig. 10: Distribution of the ranging error before (uncor.) and after (cor.) applying error correction when using RELa (with different combinations of $\alpha$ and $\beta$) and TraLa.

correction model based on the classifier's prediction (TrPr). As $\beta$ increases, the MAE and 90$^{th}$ vary only slightly, with a minor increase of the ranging error in LoS and a minor decrease in NLoS. Essentially, given a fixed $\alpha$ (35 in this case), the choice of $\beta$ allows to fine-tune the trade-off between effectively correcting NLoS samples and negatively affecting LoS samples. According to the cost function defined in Eq. 1, among the studied models, the best performance is achieved when using $\beta = 25$, i.e., by $RELa[35, 25]$.

Fig. 10 shows the distribution of the ranging error before and after applying correction when using RELa with different combinations of $\alpha$ and $\beta$ as well as TraLa. For correctly-classified NLoS samples (i.e., the TP), the distribution is similar when using RELa and TraLa, with a peak around +25 cm. However, less samples are sent for correction when using RELa (i.e., the number of TP is lower), especially for higher values of $\alpha$ (Fig. 10b). A striking difference can be seen when observing the FP (Fig. 10a): when using TraLa, there is a long tail of LoS samples with large negative errors after applying correction. When using RELa[20,-10] and RELa[35,25] (whose curves overlap), this tail is smaller and the number of FP is also smaller. Moreover, the peak of the distribution is shifted by only -24 cm when using RELa[20,-10], compared to -33 cm when using TraLa. When using RELa[20,+10] (i.e., when increasing $\beta$), the distribution of the ranging error for the FP would have a peak with similar height to that of RELa[20,-10], but shifted by -41 cm (not shown in Fig. 10a to avoid cluttering the picture). This confirms that a higher $\beta$ has a negative impact on the correction of LoS samples.




(a) LoS




(b) NLoS

Fig. 11: Ablation study comparing the use of RELa and TraLa.

Note that when using a higher $\alpha$ (which results in less FP), one can have a more aggressive regression (i.e., use a higher $\beta$ to train for strong NLoS cases). Conversely, when selecting a lower $\alpha$, the regression model should be trained also with smaller errors (i.e., with a lower $\beta$). Fig. 10 exemplifies these two cases using $RELa[35, 25]$ and $RELa[20, -10]$.

### D. Ablation study

Fig. 11 shows the performance of different approaches (using various combinations of RELa and TraLa to train the classification and regression models) sorted by the cost function defined in Eq. 1. Approaches based on TraLa (which uses classical LoS/NLoS labels to inform the correction and NLoS labels to train the regression) perform poorly. Performance is worst when applying a blind correction without classification (NoCl) first, which is denoted as TraLa[NoCl].

Comparing TraLa[NoCl] and TraLa with RELa[NoCl][-25] and RELa[TraLa][-15] shows the importance of adding samples with a low ranging error to the training set for the regression model, as it reduces the negative impact on LoS samples by up to 59% and 53%, respectively, while keeping an on-par NLoS correction performance.

The most important reduction on the negative impact of error correction on LoS samples is achieved by training the classifier with a new label based on $\alpha$, which is clearly superior to TraLa. Note that, when training the classifier using $\alpha = 35$, there is little impact of the set used to train the regression model (i.e., RELa[35][25], RELa[35][TraLa], and RELa[35][TrPr] perform relatively on-par). Nevertheless, RELa[35][25] (i.e., fine-tuning $\beta$) allows to find a good balance between effectively correcting NLoS-induced errors and avoiding to negatively affect samples collected in LoS.

## VI. Evaluation

We study next the performance of RELa and compare it with that of TraLa using several datasets (§VI-A). Among others, we show that we can derive a *generic* configuration (i.e., RELa[20,-10]) that consistently outperforms TraLa and that offers good performance across all scenarios, regardless of the employed UWB platform (§VI-B). We also show the superiority of RELa compared to TraLa using a different model than XGBoost, and further show that the superiority holds true regardless of the employed feature set (§VI-C).

## A. Experimental Setup & Methodology

**Datasets.** We use four datasets in this evaluation: the publicly-available STOCKER CPS (as in the previous sections), which contains measurements collected using Qorvo EVK1000 devices (which embed the DW1000 radio), and three datasets that we acquired at our premises. These are collected using UWB platforms embedding either Qorvo or NXP chipsets. Specifically, the P2F DW3000 dataset is collected using an nRF52 device connected to a Qorvo DW3000 with a single chip antenna. The P2F CHIP and P2F PATCH datasets use the NXP Trimension™ SR150 (which embeds the QN9090 radio) with either two chip or two patch antennas, respectively. Across these three datasets, before applying correction, P2F PATCH exhibits the best performance in both LoS and NLoS conditions. P2F DW3000 and P2F CHIP, instead, exhibit the lowest ranging accuracy in LoS and NLoS, respectively.

All devices operate on channel 9, and are configured to use a pulse repetition frequency of 64 MHz, a preamble length and STS length of 64 symbols, and a data rate of 6.8 MBit/s. Data was acquired in a 60x40 m office by mounting the devices on two tripods at 1.5 m height, and by performing 100 ranging measurements (for which we logged the full CIR) between devices of the same type in 125 different positions.

**Models.** In §VI-B, we reuse the same XGBoost classification and regression models proposed by Gallacher et al. [8] described in §V-A. To show the applicability of RELa also to other models, we use in §VI-C Linear Support Vector Machines (LinSVM) with different feature sets. Specifically, we consider the following feature sets used in the literature:

- 30FS, which includes the 29 features from [18] plus the average signal power;
- ST, which includes the 9 features defined in [7];
- EM+, which includes the 3 features defined in [18], plus the estimated distance.

Unless differently specified, we reuse the same data preparation strategy and performance metrics described in §V-A.

## B. RELa Performance for Different Datasets

For each dataset, we investigate the following models: (i) the best performing RELa[$\alpha$,$\beta$] model of each dataset and the corresponding version in which the regression model is trained using TraLa (i.e., RELa[$\alpha$][TraLa]); (ii) RELa[20,-10], which we show to offer good performance across all datasets, together with a version in which the regression model is trained using TraLa (i.e., RELa[20][TraLa]); (iii) RELa[35,25], which was the best performing model in STOCKER CPS based on our analysis in §V; as well as (iv) RELa[10,0] and RELa[15,-10] to show how additional combinations of $\alpha$ and $\beta$ perform.

Fig. 12 shows the error correction performance of RELa for each dataset, sorted according to the cost function defined in Eq. 1. On the right, the ranging error when applying TraLa and before applying correction are shown as baselines.

Across all datasets, as expected, TraLa worsens the ranging errors of LoS samples (with increases in $90^{th}$ from 6 to 22.6 cm). RELa, instead, effectively corrects ranging estimates in NLoS conditions with a performance that is on par with

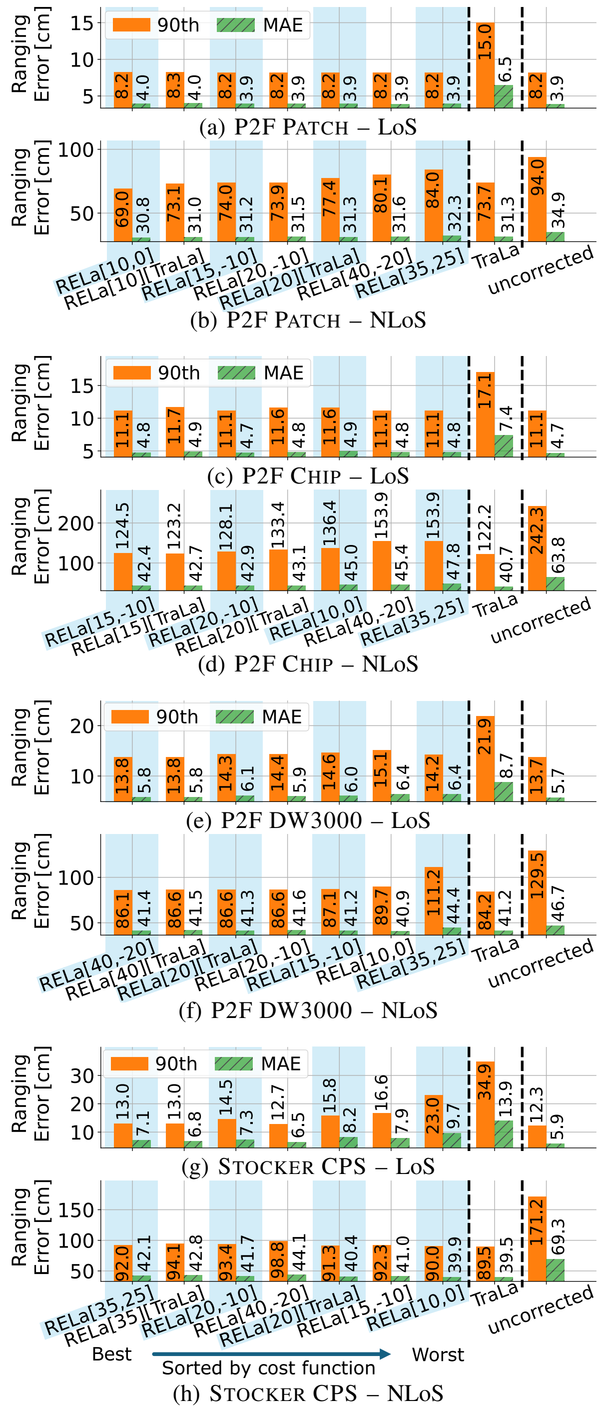



Fig. 12: Performance of RELa for different datasets.

TraLa (±5 cm), but with a negligible impact on LoS samples (where RELa reduces the introduced error by up to 99%[3]).

Note that the RELa[$\alpha$][TraLa] models also offer a competitive performance, being always the second best for each dataset according to our cost function. This confirms our observation in §V: with an optimized RELa classifier, the regression model alone has a minor impact on performance. The results in

[3]For example, in Fig. 12a, the ranging error ($90^{th}$) obtained with TraLa minus the uncorrected error is 15.00 - 8.18 = 6.82 cm, the ranging error obtained with RELa[20,-10] minus the uncorrected error is 8.19 - 8.18 = 0.01 cm. This corresponds to a reduction in the introduced error in LoS of 99.85%.

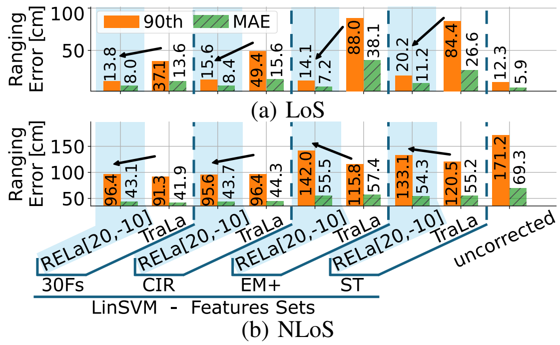



Fig. 13: Performance of RELa and TraLa when using LinSVM models with different feature sets (STOCKER CPS dataset).

Fig. 12 also allow us to observe that lower $\alpha$ values are optimal for devices that are more accurate in terms of their uncorrected LoS ranging, i.e., $\alpha$ of 10–15 cm for P2F PATCH and P2F CHIP, and 35–40 cm for P2F DW3000 and STOCKER CPS. We can hence conclude that RELa outperforms TraLa in all four datasets, showing that its advantages are not specific to certain environments. Moreover, we can consider RELa[20,-10] a *general setting* offering a balanced performance across all datasets. In fact, it worsens samples taken in LoS by only $\leq 1.5$ cm, and samples taken in NLoS by $\leq 5$ cm ($90^{th}$) compared to the best RELa setting for a given dataset.

### C. RELa Performance of Different Models

Fig. 13 compares the performance of RELa[20,-10] with TraLa when using LinSVM with the different features sets described in §VI-A. In line with our previous results, RELa[20,-10] outperforms TraLa in avoiding to worsen the LoS samples (the $90^{th}$ is worsened by up to 8 cm when using RELa[20,-10] compared to the uncorrected case, whereas it grows up to 80 cm or more when using the EM+ and ST feature sets). The error correction performance for NLoS samples is mostly on par between TraLa and RELa, except for EM+, where the $90^{th}$ is worsened by 26.2 cm when using RELa (but note that, conversely, the $90^{th}$ for LoS samples is improved by 74 cm). In general, we can conclude that RELa outperforms TraLa regardless of the employed feature set. The feature sets 30FS and CIR perform best: among the two, the latter is preferred, as it requires a lower computation effort for feature extraction [18]. Finally, it is remarkable that the use of LinSVM with the CIR feature set performs similar (a difference of only $\pm 2.2$ cm in all metrics) to the XGBoost model (see Fig. 12): this is an interesting observation, as the LinSVM has less memory requirements than XGBoost (3 kB vs. 18 kB) and is hence more suitable for very constrained UWB devices.

## VII. CONCLUSION & FUTURE WORK

Error correction approaches for UWB ranging estimates using traditional LoS/NLoS labeling (TraLa) commonly reduce the error by roughly 50% in NLoS conditions. However, they apply correction on a large number of LoS samples, which worsens the ranging estimates by tens of centimeters. In this work, we have proposed Ranging Error Labeling (RELa), which redefines the training labels and sets of classification and error correction models used in the literature. We have shown experimentally that RELa can reduce the error introduced on ranging estimates taken in LoS by up to 99%, while maintaining the ability to effectively correct ranging estimates taken in NLoS with a performance that is on-par with TraLa. RELa currently uses binary labels for classification (LoRE/HiRE). In the future, we will explore if defining multiple ranging error classes can further boost performance.

## ACKNOWLEDGMENTS

This work has been supported by the BMK, BMAW, and FFG, Contract No. 881844: "Pro²Future" (Products and Production Systems of the Future), within the ENHANCE-UWB project.

## REFERENCES


[1] T. Brovko *et al.*, "Complex Kalman Filter Algorithm for Smartphone-based Indoor UWB/INS Navigation Systems," in *Proc. of the IEEE USBEREIT Symposium*, 2021.

[2] A. Ledergerber *et al.*, "A Robot Self-Localization System using One-Way UWB Communication," in *Proc. of the IROS Conference*, 2015.

[3] EETimes, "VW and NXP Show First Car Using UWB To Combat Relay Theft," 2019. [Last access: 2023-12-17] https://www.eetimes.com/volkswagen-and-nxp-show-first-car-using-uwb-to-combat-relay-theft/.

[4] I. Guvenc *et al.*, "NLOS Identification and Mitigation for UWB Localization Systems," in *Proc. of the IEEE WCNC Conference*, 2007.

[5] J. Borras *et al.*, "Decision Theoretic Framework for NLOS Identification," in *Proc. of the 48th IEEE VTC Conference*, 1998.

[6] V. Minh Le *et al.*, "Human Occlusion in UWB Ranging: What Can the Radio Do for You?," in *Proc. of the 18th MSN Conference*, 2022.

[7] M. Stocker *et al.*, "Performance of Support Vector Regression in Correcting UWB Ranging Measurements under LOS/NLOS Conditions," in *Proc. of the 4th CPS-IoTBench Workshop*, 2021.

[8] M. Gallacher *et al.*, "InSight: Enabling NLOS Classification and Error Correction on Resource-Constrained Ultra-Wideband Devices," in *Proc. of the 20th EWSN Conference*, 2023.

[9] M. Stocker *et al.*, "Applying NLOS Classification and Error Correction Techniques to UWB Systems: Lessons Learned and Recommendations," in *Proc. of the 6th CPS-IoTBench Workshop*, 2023.

[10] K. Bregar *et al.*, "Improving indoor localization using convolutional neural networks on computationally restricted devices," *IEEE Access*, 2018.

[11] Z. Zeng *et al.*, "UWB NLOS Identification with Feature Combination Selection based on Genetic Algorithm," in *IEEE ICCE Conf.*, 2019.

[12] M. Stocker *et al.*, "STOCKER CPS UWB NLOS Dataset," 2021. [Online] http://iti.tugraz.at/uwb-nlos – Last access: 2023-08-01.

[13] S. Maranò *et al.*, "NLOS Identification and Mitigation for Localization based on UWB Experimental Data," *IEEE JSAC*, vol. 28, no. 7, 2010.

[14] J. Xin *et al.*, "A Bayesian Filtering Approach for Error Mitigation in Ultra-Wideband Ranging," *Sensors*, vol. 19, no. 3, 2019.

[15] M. Ramadan *et al.*, "NLOS Identification for Indoor Localization using Random Forest Algorithm," in *Proc. of the 22nd International Workshop on Smart Antennas*, 2018.

[16] T. Van Nguyen *et al.*, "Machine Learning for Wideband Localization," *IEEE Journal on Selected Areas in Communications*, 2015.

[17] A. Ferreira *et al.*, "Feature Selection for Real-Time NLOS Identification and Mitigation for Body-mounted UWB Transceivers," *IEEE TIM*, 2021.

[18] E. Ember *et al.*, "Impact of Feature Selection and CIR Window Length on NLoS Classification for UWB Systems," in *MSN 2023*, 2023.

[19] K. Bregar *et al.*, "NLOS Channel Detection with Multilayer Perceptron in Low-Rate Personal Area Networks for Indoor Localization Accuracy Improvement," in *Proc. of the 8th IPSSC Conference*, 2016.

[20] S. Tomović and other, "Transformer-based nlos detection in uwb localization systems," in *Proc. of the 30th TELFOR*, IEEE, 2022.

[21] B. Jones *et al.*, "Tiny but Mighty: Embedded Machine Learning for Indoor Wireless Localization," in *Proc. of the 20th CCNC Conf.*, 2023.

[22] S. Venkatesh *et al.*, "Non-line-of-sight Identification in Ultra-wideband Systems based on Received Signal Statistics," *IET MAP*, 2007.

[23] Y. Chen *et al.*, "Exploiting Anchor Links for NLOS Combating in UWB Localization," *ACM Trans. Sen. Netw.*, vol. 20, no. 3, 2024.

[24] D.-H. Kim *et al.*, "UWB Positioning System Based on LSTM Classification With Mitigated NLOS Effects," *IEEE Internet of Things*, 2023.

[25] S. Hochreiter and J. Schmidhuber, "Long Short-Term Memory," *Neural computation*, vol. 9, no. 8, pp. 1735–1780, 1997.