Peer-reviewed version available at Appl. Sci. 2020, 10, 3980; doi:10.3390/app10113980

Article

Identification of NLOS and Multi-path Conditions in UWB Localization using Machine Learning Methods

Cung Lian Sang*, Bastian Steinhagen, Jonas Dominik Homburg, Michael Adams, Marc Hesse and Ulrich Rückert

Cognitronics and Sensor Systems Group, CITEC, Bielefeld University, 33619 Bielefeld, Germany; csang@techfak.uni-bielefeld.de; bsteinhagen@techfak.uni-bielefeld.de; jhomburg@techfak.uni-bielefeld.de; madams@techfak.uni-bielefeld.de; mhesse@techfak.uni-bielefeld.de; rueckert@techfak.uni-bielefeld.de;

* Correspondence: csang@techfak.uni-bielefeld.de; Tel.: +49 521 106 - 67368

Abstract: In Ultra-wideband (UWB)-based wireless ranging or distance measurement, differentiation between line-of-sight (LOS), non-line-of-sight (NLOS), and multi-path (MP) conditions are important for precise indoor localization. This is because the accuracy of the reported measured distance in UWB ranging systems is directly affected by the measurement conditions (LOS, NLOS or MP). However, the major contributions in literature only address the binary classification between LOS and NLOS in UWB ranging systems. The MP condition is usually ignored. In fact, the MP condition also has a significant impact on the ranging errors of the UWB compared to the direct LOS measurement results. Though, the magnitudes of the error contained in MP conditions are generally lower than completely blocked NLOS scenarios. This paper addresses machine learning techniques for identification of the mentioned three classes (LOS, NLOS, and MP) in the UWB indoor localization system using an experimental data-set. The data-set was collected in different conditions at different scenarios in indoor environments. Using the collected real measurement data, we compare three machine learning (ML) classifiers, i.e., support vector machine (SVM), random forest (RF) based on an ensemble learning method, and multilayer perceptron (MLP) based on a deep artificial neural network, in terms of their performance. The results show that applying ML methods in UWB ranging systems are effective in identification of the above-mentioned three classes. In specific, the overall accuracy reaches up to 91.9% in the best-case scenario and 72.9% in the worst-case scenario. Regarding the F1-score, it is 0.92 in the best-case and 0.69 in the worst-case scenario. For reproducible results and further exploration, we (will) provide the publicly accessible experimental research data discussed in this paper at PUB - Publications at Bielefeld University. The evaluations of the three classifiers are conducted using the open-source python machine learning library scikit-learn.

Keywords: UWB; NLOS identification; multi-path detection; NLOS and MP discrimination; machine learning; SVM; random forest; multilayer perceptron; LOS; DWM1000; indoor localization

1. Introduction

Indoor localization systems enable several potential applications in diverse fields. A few examples where positioning is crucial include tracking valuable assets and personal devices in IoT, ambient assisted living systems in smart home and hospital, logistics, autonomous driving system, customer tracking system in shopping and public areas, positioning systems in industrial environments, and mission-critical systems such as an application for firefighters and soldiers [1–3]. Among several technologies available for indoor localization described in the literature, an ultra-wideband (UWB) technology [1,4,5] plays an increasingly important role in a precise indoor localization system due to its fine ranging resolution and obstacle-penetration capabilities [2,3,6].

Peer-reviewed version available at Appl. Sci. 2020, 10, 3980; doi:10.3390/app10113980

In wireless ranging systems including UWB technology, the distance between the transmitter and receiver is estimated by measuring the time-of-flight (TOF) between the two transceivers and multiplying it with the speed of light [7,8]. But the ranging algorithm assumes that the *TOF* signal is always in a direct line-of-sight (LOS) condition. Therefore, non-line-of-sight (NLOS) [9–11] and multi-path (MP) [3] conditions cause a positive bias in the estimated distances. Figure 1 expresses an abstract view of LOS, NLOS and MP conditions in typical wireless communications. The figure shows how a signal sent from a tag device (green pyramid shape in the middle) can be received in different scenarios at the anchor nodes (yellow pyramid shapes). Therefore, differentiation between the LOS, NLOS, and MP conditions in wireless ranging systems are important for precise localization systems.

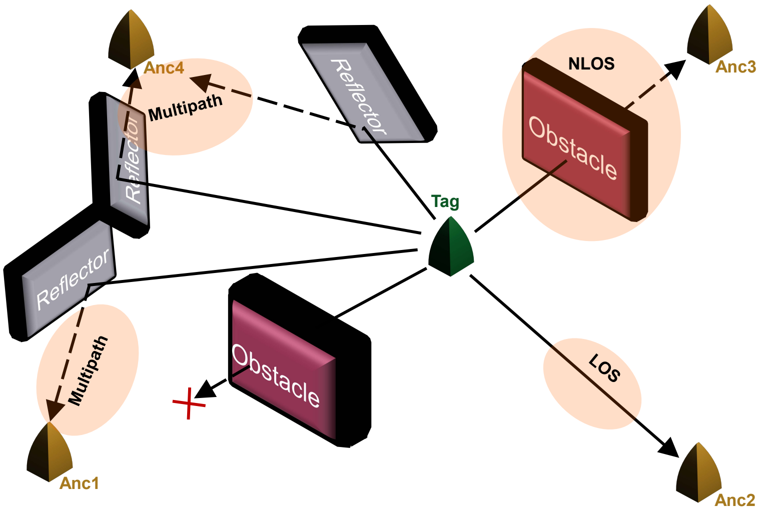

Figure 1. Illustration of LOS, NLOS and Multipath (MP) scenarios in UWB-based ranging system.

This paper discusses the vital role of classifying LOS, NLOS and MP scenarios in UWB ranging system using machine learning (ML) approaches. By understanding the defined three classes, a positioning algorithm [6,12] can mitigate the biases caused by NLOS and MP conditions, i.e. by giving different weights to each class. We have proposed such a mitigation technique in our previous work [2]. The common identification and mitigation techniques for NLOS condition in UWB can be found in [9,13,14] and the references provided in there.

In fact, the multi-class identification of UWB measurement data (LOS, NLOS, and MP) in real-world is challenging in indoor environments because a variety of physical effects can distort the direct path LOS signal in different ways [3,15] (Figure 1). This includes walls, furniture, humans, the orientation of the UWB antenna, etc. Therefore, machine learning methods are attractive solutions for solving such a problem.

Identification and mitigation techniques of NLOS condition in UWB, or wireless communications in general, using ML methods are not new. It has been received significant interests in recent years [9–11,15–21]. However, the major contributions in the literature address the binary classification between the LOS and NLOS in UWB ranging system.

In contrast, this paper addresses machine learning techniques for direct identification of the mentioned three classes (LOS, NLOS, and MP) in UWB indoor localization system using experimental data collected at seven different environments in two different test scenarios (section 4). Using the collected real measurement data, we compare three machine learning methods, i.e., support vector machine (SVM), random forest (RF), and multilayer perceptron (MLP) in terms of their prediction accuracy, training time, and testing time. The classifiers are chosen by bearing in mind that the evaluated ML models can be used in a low cost and power efficient real-time system such as microcontroller-based platforms [22].

For the sake of reproducible results and further exploration, we provide all the experimental research data and its corresponding source codes as a supplementary data of this manuscript in

PUB - Publications at Bielefeld University [PUB link will be provided upon the acceptance], which is publicly available. The evaluation of the algorithms is conducted using the open-source python machine learning library scikit-learn [23].

2. Problem Description

The primary goal of the identification process in wireless communications is to detect the existence of a NLOS and/or MP condition in a communication between a transmitter and a receiver. This process is crucial because the multi-path effects and the NLOS conditions strongly influenced the accuracy of the measured distances in wireless communications. As an example, Figure 2 (d) compares the error of measured distances at static scenarios in LOS, NLOS and MP conditions in UWB based on our experimental data. Figure 2 ((a) - (c)) also illustrates the comparison of the conventional identification techniques for the mentioned three classes (LOS, NLOS, and MP) based on the first-path (FP) power level and the channel impulse response (CIR) (more details description in subsection 3.1).

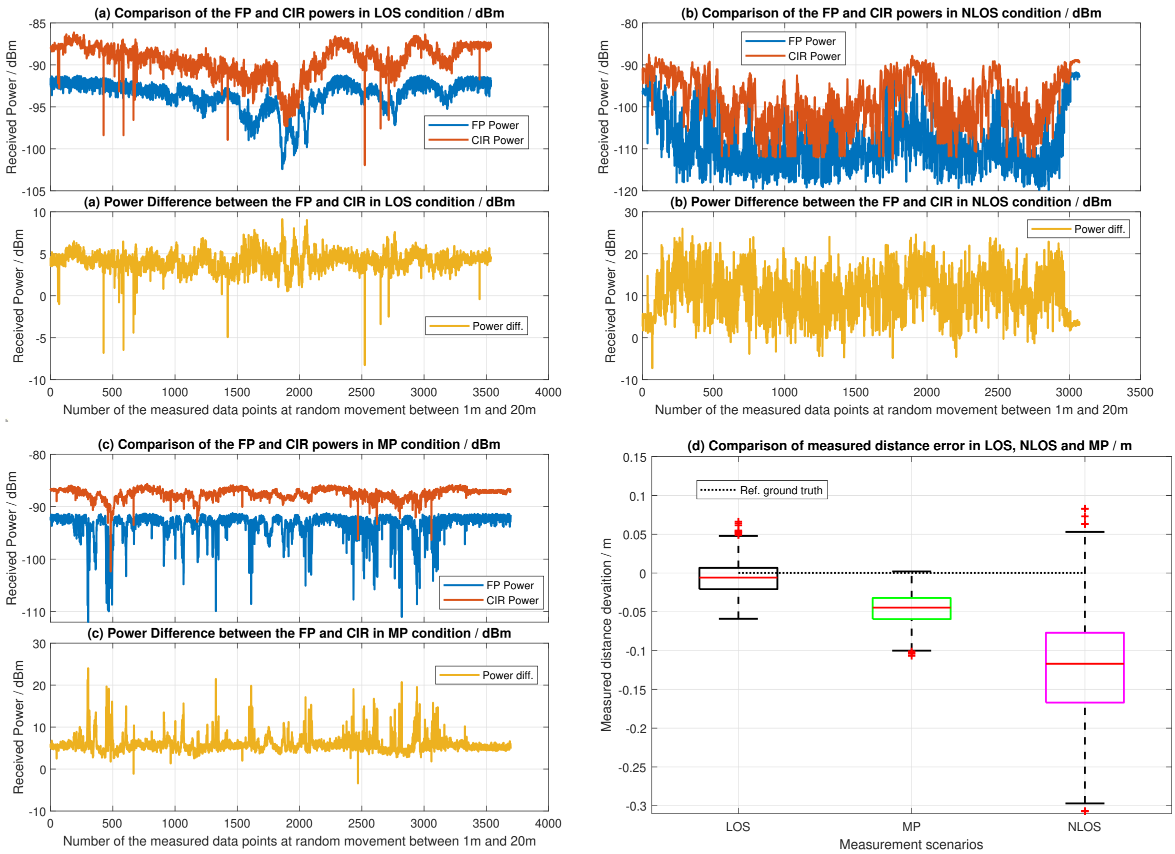

Figure 2. Illustration of LOS, NLOS and MP conditions in UWB ranging systems: (a) - (c) Comparison of the FP power, the CIR power, and the difference between the two power levels (FP and CIR power) at three scenarios: (a) LOS, (b) NLOS, and (c) MP conditions. The measurement was conducted for the three scenarios at random movement between 1 m to 24 m distances. (d) Comparison of measured distance errors at the mentioned three conditions (LOS, NLOS, and MP) in static scenarios.

The experimental evaluation results in Figure 2 (d) suggests that the magnitude of the error in NLOS and MP conditions is considerably larger than the LOS condition compared to the ground truth reference. Moreover, the error introduced by MP condition is significantly lower than the completely blocked NLOS condition, where the signal needs to penetrate the obstacle to reach the receiver. Indeed, this depends on the materials and other factors of the obstacles [3]. However, the result in Figure 2 (d), which is blocked by a human in this experiment, indicates that the NLOS condition introduced the highest impact on the measured distance errors. This motivates us the classification of

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 29 April 2020 doi:10.20944/preprints202004.0503.v1

Peer-reviewed version available at Appl. Sci. 2020, 10, 3980; doi:10.3390/app10113980

the UWB ranging systems into three classes (LOS, NLOS, MP) to improve the location accuracy in UWB localization system. The classified ranging information is applicable in any positioning algorithm [6,12] to effectively mitigate the biases [2,16,18] caused by the NLOS and MP conditions.

3. Related Works

In literature, the problem solving strategy for ranging errors in UWB due to the effects of NLOS and/or MP can be coarsely classified into two steps [9,13]: (i) NLOS identification process [14,24,25] and (ii) NLOS mitigation process [9,14,26]. This paper is solely focused on the former case. In fact, there exists a method that by-pass the identification process and directly mitigate the ranging error using channel statistics and SVM as a ML-based classifier [16,27]. However, this method restricts the flexibility to choose different positioning algorithms in the later case since the mitigation technique is limited to a few compatible algorithms.

The common approach is to detect the non-direct path signal (i.e. NLOS and/or MP) and use the detected information to modify the location algorithm in order to mitigate the biases caused by the NLOS and/or MP conditions [9,10,15,16,18,19,27]. In this manuscript, we divided the related works into two subsections: (i) conventional approaches without using ML techniques (subsection 3.1), and (ii) ML-based approaches (subsection 3.2). Our proposed technique regarding the multi-classes identification process in UWB is based on the ML-based approach.

3.1. Conventional NLOS Identification Techniques in UWB

As already mentioned in section 1, identification of NLOS and LOS in UWB communications is not new. There has been several proposals in literature [9,14,24,25,28–30] to identify and mitigate the NLOS conditions in UWB. However, the identification process of the MP condition is usually ignored in literature, although the effects of the MP conditions in UWB ranging systems were acknowledged as important aspects in [3,8,19,31]. Conventionally, the NLOS detection in UWB is always regarded as a binary classification problem. The traditional NLOS detection methods can be coarsely categorized as:

- Identification of NLOS situation based on binary hypothesis test [24]
- NLOS detection based on the change of Signal to Noise ratio (SNR) [25]
- NLOS identification based on channel impulse response [9,28]
- NLOS detection techniques based on the multi-path channel statistics such as the kurtosis, the mean excess delay spread, and the root mean square delay spread [14,30]
- Detection of the NLOS condition using Received Signal Strength (RSS) [29,30]

In brief, the conventional NLOS identification approaches mainly rely on the statistically condition of the received signals in UWB communications. Figure 2 (a) - (c) demonstrates these scenarios by comparing the first-path (FP) signal power and the channel impulse response (CIR) power for three conditions (LOS, NLOS, and MP). Among the mentioned NLOS detection methods, the threshold approach presented by Decawave in [32] was widely used in different UWB applications and system implementations [21,32,33]. This is accomplished by taking the difference between the estimated total received (RX) power and the First-Path (FP) Power using the following equations [32] (Figure 2 (a) - (c)):

$$FP\ Power\ Level = 10 \cdot log_{10}(\frac{F_1^2 + F_2^2 + F_3^2}{N^2}) - A \tag{1}$$

where, F_1, F_2, and F_3 are the first, second, and third harmonics of the first path signal amplitudes for a signal propagation through wireless media as in a multi-path, NLOS, and/or LOS scenarios [32,33]. N is the value of preamble accumulation count reported in the DW1000 chip from Decawave. A is a predefined constant value, which has 133.77 for a pulse repetition frequency (PRF) of 16 MHz and 121.74 for a PRF of 64 MHz.

 doi:10.20944/preprints202004.0503.v1

Peer-reviewed version available at Appl. Sci. 2020, 10, 3980; doi:10.3390/app10113980

The estimated received power (RX) level can be defined as:

$$RX\ Power\ Level = 10 \cdot log_{10}(\frac{C \cdot 2^{17}}{N^2}) - A \tag{2}$$

where, C is the value of the channel impulse response power reported in DW1000 chip.

Therefore, the metric that specifies the conditions of LOS and NLOS in the threshold method can be achieved by computing the difference between the received and first path power [32] as:

$$Threshold\ Power = RX\ Power\ Level - FP\ Power\ Level \tag{3}$$

In conventional threshold approach, the measured distance is classified as a LOS when the threshold power using (3) is less than 6 dBm and defined as a NLOS when it is more than 10 dBm [32]. This is a sub-optimal acceptable solution as our particular experimental evaluation results shown in Figure 2. That is the mean value of the threshold power including its standard deviation in LOS condition using (3) is 4.12 ± 1.13 dBm (Figure Figure 2 (a)) and the NLOS condition is 10.75 ± 5.51 dBm (Figure 2 (b)). However, the solution is not optimal as a lot of fluctuation can occur as described in the experimental measurement data (Figure 2 (a) - (c)). The condition is more harsh to solve in MP condition, where the first-path signals are hard to clearly distinguish from the received signal (Figure 2 (c)).

Nevertheless, the classification of the mentioned three classes is not straightforward. The complexity of the classification problem increases especially in indoor environments because of several factors such as material characteristics [34], refractive index of different materials and so on [1,3]. Moreover, the phenomenon of the multi-path effects and NLOS depends on the properties of the medium through which the signal travels, the location (dimension of the places and rooms) where the signals are measured, the presence of other objects within the measured environment, the orientation of the UWB antenna, etc. Therefore, ML approaches has been regarded as an attractive strategy for solving this complex task in recent years (subsection 3.2).

3.2. Identification of NLOS and MP Conditions in Literature based on Machine Learning Techniques

One of the earlier ML-based NLOS identifications in UWB was conducted in [9] using SVM as a classifier. In that paper, the identification process was considered as a binary classification problem (LOS vs. NLOS) showing the ML approaches outperform the traditional parametric techniques and signal processing approaches from the literature.

Consequently, several investigations of NLOS identification process in UWB were examined in the literature using different ML techniques as a classifier such as SVM in [9,10,15,19], MLP in [21,35,36], Boosted decision tree (BDT) in [36], recursive decision tree in [33], and other ML techniques such as kernel principal component analysis in [17], etc. Moreover, the unsupervised machine learning technique called "expectation maximization for Gaussian mixture models" was recently applied in [37] to classify the LOS and NLOS conditions in UWB measurement. Likewise, deep learning approaches such as convolutional neural network (CNN) were also explored to distinguish the NLOS condition from LOS in UWB ranging [20,31]. In CNN-based deep learning approaches, the authors generally modified the existing CNN network such as GoogLeNet [20], VGG-architecture (i.e. VGG-16 or VGG-19) [20,31], AlexNet [20], etc. to be usable for the low cost UWB systems. The reported overall accuracy ranges starting from 60 % (using typical ML technique such as SVM) up to 99 % (using CNN approach). In all of the above-mentioned approaches, the focus is solely on detecting the NLOS condition in UWR ranging, i.e. the binary classification between LOS and NLOS.

Moreover, the performance comparison of different ML techniques for identification of NLOS in UWB ranging was conducted in [11,15,36]. The main purpose of these analyses is to compare the impact of model selection in ML-based system applications in UWB. In [36], the performance comparison of two ML methods namely MLP and BDT was carried out for the binary classification (LOS vs.

Peer-reviewed version available at Appl. Sci. 2020, 10, 3980; doi:10.3390/app10113980

NLOS). The resultant report concluded that the BDT outperforms the MLP. Likewise, the comparison of five classifiers using Matlab (i.e. SVM, k-nearest neighbor (KNN), binary decision tree, Gaussian process (GP), generalized linear model) was performed in [11]. The authors concluded that KNN and GP performed better than the other three models. Similarly, the authors in [15] evaluated three ML models (SVM, MLP, RF) to classify the LOS and NLOS in narrowband wireless communications (i.e. not specifically for UWB systems in this case). The author reported that RF and MLP performed better than SVM in all of their evaluations.

In contrast to the binary classification between LOS and NLOS in UWB, the binary classification of the MP from LOS conditions was investigated in [31]. The author reported that MP effects can cause an error in UWB ranging from a few centimeters up to 60 cm. This can also be verified in our experimental evaluation presented in Figure 2 (d).

Throughout the literature, the problem has been treated as a binary classification problem or hypothesis test (i.e. LOS vs. NLOS or LOS vs. MP). To the best of the authors' knowledge, only two papers addressed UWB-based ranging errors as a multi-class problem [11,19]. The first paper was based on a two-step identification process [19] using SVM as a classifier. In that paper, the LOS and NLOS was identified in the first step. Then, further classification (MP vs. NLOS) is categorized in the second step if NLOS was detected in the first step. The second paper categorized the NLOS conditions into two types (soft-NLOS vs. hard-NLOS) in addition to LOS while ignoring the MP effects [11]. The differentiation between the two NLOS types is primarily based on the material of the obstacles which the UWB signal is passing through by penetration. The authors used two types of walls in their evaluation to classify a soft-NLOS and a hard-NLOS.

On the contrary to the above-mentioned approaches, we perform a direct identification of the multi-class classification for UWB-ranging systems in this paper. The classified classes are LOS, NLOS, and MP conditions. Based on the measurement data, we performed three ML models (section 5) namely SVM, RF, and MLP to compare their performances (section 6 and 7). The experimental research data utilized in this paper are provided at a public archive for reproducible results and further exploration.

4. Measurement Scenarios and Data Preparation

In this section, we describe the experimental set-up of the evaluations (subsection 4.1), the data collection processes including labeling and data separation (subsection 4.2), and the feature extraction based on the collected data (subsection 4.3) for the evaluated three ML models.

4.1. Experimental Set-up

For the UWB data measurement process in experimental evaluations, we used a DWM1000 module [32] manufactured by Decawave as the UWB hardware and the STM32 development board (NUCLEO-L476RG) [38] manufactured by STMicroelectronics as the main Microcontroller (MCU). Table 1 provides the hardware configurations used in the experimental evaluations.

Table 1. Configurations of the primary hardware used in the experimental evaluation

Types of hardware	**Properties**	**Values**
UWB module	Module name	DWM1000
	Data rate	6.8 Mbps
	Center frequency	3993.6 MHz
	Bandwidth	499.2 MHz
	Channel	2
	Pulse-repetition frequency (PRF)	16 MHz
	Reported precision	10 cm
	manufacturer	Decawave
Microcontroller (MCU)	Module type	STM32L476RG
	Development board	NUCLEO-L476RG
	manufacturer	STMicroelectronics

Peer-reviewed version available at Appl. Sci. 2020, 10, 3980; doi:10.3390/app10113980

Our previous work [7,8] pointed out that the Alternative Double-sided Two-way ranging (AltDS-TWR) method outperformed other available TWR methods in literature at different tested scenarios. Therefore, we applied AltDS-TWR as a wireless ranging method in our evaluations. Furthermore, AltDS-TWR operated well without the need to use high precision external oscillators in the MCU [8]. Hence, the built-in high-speed internal (HSI) clock source (16 MHz) from the MCU was applied to all of the evaluation results presented in this article. According to the data sheet [39], the HSI has an accuracy of $\pm 1\%$ using the factory-trimmed RC oscillator.

During measurement, one of the transceivers among the two is connected to a computer for logging the data via serial USART port. Both transceivers were executed with a two-way ranging software provided by Decawave for production testing of their evaluation kit (EVK1000), which is available online (https://www.decawave.com/software/) on Decawave's website. We modified the library of this software to extracted all required features provided in subsection 4.3. Then, we logged and saved the extracted features into a file at each trial in our measurement campaign. To avoid the effect of Fresnel zones in our measurement results, the antenna height was always maintained at 1.06 m in one of our UWB devices, i.e. the static one that recorded the measurement data via PC.

4.2. Data Collection Process

The required data for experimental evaluations presented in section 7 were collected in three scenarios (two small rooms, a hall, and four corridors) at 7 different places in indoor environments (Figure 3). The two rooms were the (6 m x 6 m) laboratory environment and approximately the (8 m x 6 m) communication room in which different items of furniture were placed. The collected data in narrow corridors were intended for MP conditions, where the direct LOS cannot be distinguished because of multiple signal reflections from the narrow walls. Figure 2 (c) illustrates a concrete example of this MP condition in terms of the FP and RX powers. In all cases, the data were collected for both static and dynamic conditions. In the dynamic case, the device attached to the PC stays static while another device was held by a human at random walks. Moreover, the NLOS conditions by blocking the communication between two transceivers using a human as an obstacle were conducted in all cases depicted in Figure 3.

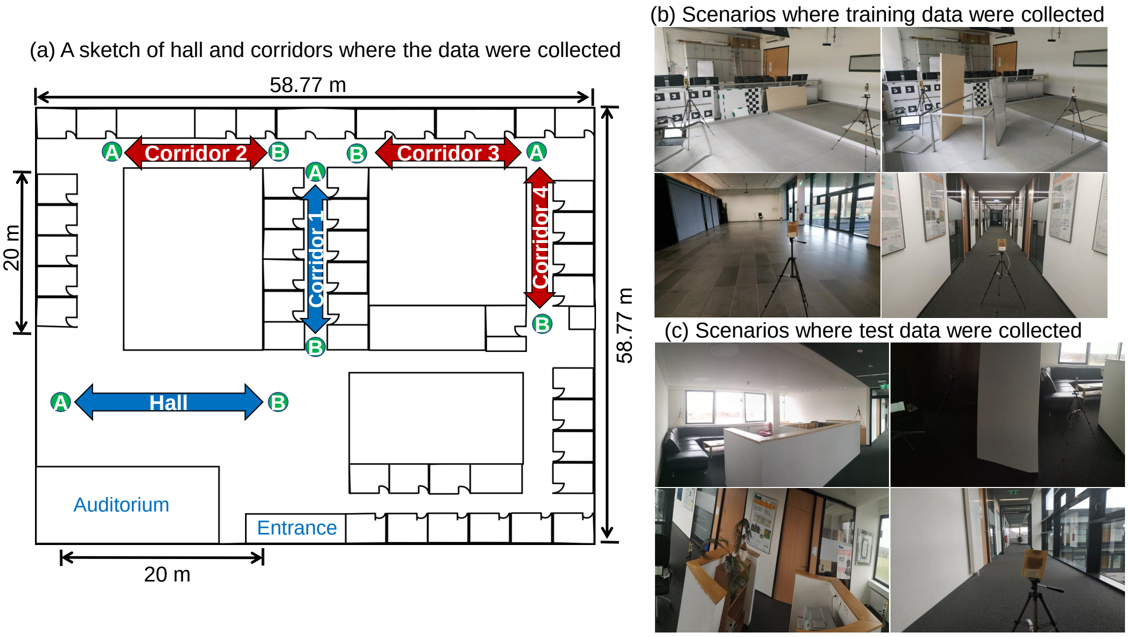

Figure 3. Illustration of the scenarios where training and test data were collected for evaluation: (a) a sketch of the building where the experimental data for training and test were collected. (b) training data were collected for both LOS, NLOS (including human blocking), and MP conditions in a laboratory, a large hall and a corridor (blue color in (a)). (c) Similar to (b), test data were collected in a different room (including different types of furniture and NLOS human blocking), and different corridor (red in (a)).

Peer-reviewed version available at Appl. Sci. 2020, 10, 3980; doi:10.3390/app10113980

4.2.1. Labeling the Measured Data and Dealing with the Class Imbalance Case

The class labels (LOS, NLOS, and MP) were manually annotated in the data preprocessing phase after the measurement campaign. During the data measurement process, a block of observations for each trial regarding the three categories (LOS, NLOS, MP) was carried out and saved individually into the PC. In this way, the whole block of that data can easily be labeled as either a LOS, NLOS, or MP.

The initial UWB data-set achieved from the measurement is imbalanced for the demanded three classes (LOS, NLOS, and MP), which is a typical phenomenon in data collection. Class imbalance refers to a scenario where the number of observations in each class is not the same in the measurement. In other words, the number of samples in one class or more classes is significantly lower or higher than those belonging to the other classes. There are several techniques to deal with the imbalanced data in classification problems including resampling techniques and algorithms [40].

We chose a random undersampling technique [40] in our evaluation to balance the mentioned three classes equally. This ensures that no artificial data points were created outside of the measured experimental data. The undersampling was performed by setting the class belonging to the smallest number of observations as a base class. Then, the classes belonging to higher samples of observation were reduced to balance with the total number of the base class by randomly selecting their elements.

4.2.2. Separation of Training, Validation, and Test data-set

Two independent test data-sets were used in our experimental evaluations. The first test data-set was separated from the measurement environments provided in Figure 3 (b), i.e. the environments where the training data were come from. This is the typical scenario for the presentation of classification results in most UWB-based literature [10,11,19]. In some cases in the literature, the test data-set was separately collected by intentionally switching the subject of the experiment, i.e. a person who carried the UWB device in [11]. However, the environment of the measurement stayed unchanged in the evaluation. As already mentioned in section 2, the propagation of the measured UWB signal can be affected by several environmental factors such as the refractive index of materials, placement of objects in the measured circumstances, etc. To examine this incident in our results, we collected a second test data-set that is independent and different from the training environments (Figure 3 (c)). This second data-set was solely split out for the purpose of testing in our evaluation. The results using both test data-sets are discussed in section 7.

For training the evaluated ML models including the validation process, more data were gathered in the laboratory room, the hall and the first corridor (Figure 3 (b)). In particular, 185 790 observations in total were gathered after balancing the three classes in this scenario. This means each class belongs to 61 930 data points. For testing purposes, the 30 % of the data points was left out by random shuffling in each trial conducted in section 7. The results using this test scenario in Figure 3 (b) are expressed as a scenario when test and training are in the same condition (section 7).

In contrast, the measurement campaign, particularly for testing purposes, is conducted in different scenarios from the training. These measurements were carried out in a different room with various items of furniture, and three different corridors (Figure 3 (c)). The results achieved from these second test scenarios are expressed in section 7 as a scenario where the test and training conditions are different. The total number of 36 015 data points, 12 005 for each class, was used for conducting this test scenario in our evaluation after balancing equally to the three classes.

4.3. Feature Extraction

In total, 12 features were extracted from the DWM1000 UWB modules manufactured by Decawave [32] using the configuration described in Table 1. The extracted features are based on the typical parameters that are necessary in the traditional NLOS identification methods as expressed in equations (1), (2) and (3). This means that no extra burdens are involved by using these extracted features in our ML application. For the sake of completeness, two more features namely standard

 doi:10.20944/preprints202004.0503.v1

Peer-reviewed version available at Appl. Sci. 2020, 10, 3980; doi:10.3390/app10113980

and maximum noises supported by the DW1000 module are included in the feature extraction of our evaluation. Therefore, the full features extracted and saved during the experimental evaluation are:

1. the reported measured distance
2. the compound amplitudes of multiple harmonics in the FP signal
3. the amplitude of the first harmonic in the FP signal
4. the amplitude of the second harmonic in the FP signal
5. the amplitude of the third harmonic in the FP signal
6. the amplitude of the channel impulse response (CIR)
7. the preamble accumulation count reported in DW1000 chip module
8. the estimated FP power level using (1)
9. the estimated RX power level using (2)
10. the difference between the FP and RX power level using (3)
11. the standard noise reported in DW1000 chip module
12. the maximum noise reported in DW1000 chip module

5. Machine learning Models for Identification of LOS, NLOS and MP conditions

Three machine learning models (SVM, RF, and MLP) are evaluated in this paper. SVM is regarded as a baseline model in the evaluation since it is the most commonly and frequently used model for the UWB-based identification of NLOS conditions in literature [9,10,15,19]. The configuration and setup for each classifier are discussed in the subsequent subsections.

The training and test times of each classifier reported in this section are based on a single concurrent CPU core without using any parallel computing devices such as GPU. The evaluation was done on the same machine for all classifiers. The reported results for all classifiers in this section (SVM in subsection 5.1, RF in subsection 5.2, and MLP in subsection 5.3) are based on 10 iterations of randomly splitting the measured training, validation and test data. The extracted features used for all classifiers are based on the discussion and selection observed in section 6. The reported training and test times per sample (mean value) for each classifier are estimated in two steps. First, we estimated the total amount of time it takes for the whole data-set in training and test phases using the corresponding training and test data-set. Then, the measured time is divided by the total number of samples to get the mean value per sample.

Generally, several parameters in the ML classifiers are tuned to achieve the optimized results. Moreover, each classifier has its specific hyper-parameters, that are not compatible with one another. Therefore, a direct comparison using exactly the same parameters for all classifiers is impossible. For the sake of simplicity and better representation of the results, the comparison was done by choosing the most important and influential parameters for each classifier in this section. This implies that the kernel type was chosen for SVM, and the number of decision trees in the forest was selected for RF. For MLP, the number of hidden layers including the total number of neurons in each layer was evaluated to choose the best option for the given problem.

For the reproducible results, the parameters for each classifier such as activation function, optimizer, earlier stopping criteria for the training, learning rate, etc. are based on the default setting of the scikit-learn [23] library if nothing is explicitly mentioned in the following. The applied stable version of the scikit-learn library was 0.22.1 as the time of writing this paper.

5.1. Support Vector Machine Classifier for UWB Localization System

The SVM is a supervised machine learning technique suitable for solving both classification and regression problems [41,42]. It is strongly based upon the framework of statistical learning theory [43]. SVM is also recognized as one of the most frequently used classification techniques in the machine learning community in the past due to its robustness and superior performance without the need to tune several parameters compared to deep neural networks [9]. In short, SVM takes the data as an input and determines a hyper-plane that separates the data into predefined classes. The hyper-plane

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 29 April 2020 doi:10.20944/preprints202004.0503.v1

Peer-reviewed version available at Appl. Sci. 2020, 10, 3980; doi:10.3390/app10113980

Table 2. Comparison of the SVM configurations based on the kernel functions

Kernel Types	Mean Accuracy with std (%)	Mean Training Time per sample (ms)	Mean Test Time per sample (ms)
Radial basis function (RBF)	**82.96** $\pm$ 0.14	2.06 $\pm$ 0.18	0.99 $\pm$ 0.01
Linear function	72.59 $\pm$ 0.25	**1.92** $\pm$ 0.08	**0.53** $\pm$ 0.01
3rd order polynomial function	70.82 $\pm$ 0.19	3.05 $\pm$ 0.09	0.80 $\pm$ 0.02
Sigmoid function	50.59 $\pm$ 3.05	3.01 $\pm$ 0.27	1.59 $\pm$ 0.09

was established in the SVM algorithm by maximizing the margin between the separatable classes as wide as possible. Table 2 presents the comparison of four kernel types in SVM using the UWB measurement data and extracted features examined in section 4.

The choice of the kernel types in SVM has a strong influence on its accuracy regarding our particular measurement of UWB data. The results in Table 2 shows that the radial basis function (RBF) kernel reached the highest accuracy with 82.96 % while the sigmoid function provided the poorest with 50.59 %. Both linear and third-order polynomial functions had comparable results. In terms of training and test times, the linear function achieved the lowest time per sample while the sigmoid function showed the worst performance with the highest time per sample. In all circumstances, the training and test times in SVM are in the order of milliseconds. This means that the SVM has the poorest performance in terms of test time compared to RF (subsection 5.2) and MLP (subsection 5.3).

5.2. Random Forrest Classifier for UWB Localization System

According to the original paper in [44], Random forests (RF) are a combination of decision-tree predictors in the forest such that each tree depends on the values of a random vector, which is sampled with the independent and identical distribution for all the trees. In brief, RF is built upon multiple decision trees and merges them to get a more accurate and stable prediction as its final output. Two significant advantages of RF are (i) reduction in over-fitting by averaging several trees, and (ii) low risk in prediction error since RF typically makes a wrong prediction only when more than half of the base classifiers (decision trees) are wrong. The disadvantage, though, is that RF is typically more complex and computationally expensive than the simple decision tree algorithm. In general, the more trees in the forest, the better the prediction. However, this flexibility comes with the cost of the processing time (training and test times) as described in Table 3.

Table 3. Comparison of the RF configurations based on the numbers of decision trees in the forest

No. of Decision Trees in the Forest	Mean Accuracy with std (%)	Mean Training Time per sample (μs)	Mean Test Time per sample (μs)
5 decision trees	90.91 $\pm$ 0.18	**4.84** $\pm$ 0.25	**1.26** $\pm$ 0.02
10 decision trees	91.55 $\pm$ 0.09	9.38 $\pm$ 0.27	2.33 $\pm$ 0.08
20 decision trees	91.83 $\pm$ 0.10	18.68 $\pm$ 0.52	4.66 $\pm$ 0.08
30 decision trees	91.89 $\pm$ 0.11	27.30 $\pm$ 0.41	6.82 $\pm$ 0.13
50 decision trees	91.99 $\pm$ 0.11	45.44 $\pm$ 0.62	11.30 $\pm$ 0.27
100 decision trees	92.07 $\pm$ 0.09	90.42 $\pm$ 1.39	22.43 $\pm$ 0.34
200 decision trees	92.12 $\pm$ 0.13	179.85 $\pm$ 3.10	44.85 $\pm$ 0.47
500 decision trees	**92.13** $\pm$ 0.12	460.04 $\pm$ 10.27	113.39 $\pm$ 1.44

The prediction accuracy in RF increases steadily as the number of decision trees in the forest is increasing (Table 3). However, the improvement was getting slowed down when the number of trees in the forest is more than 50 in this particular UWB data. In contrast, the training and test time keeps increasing linearly by the increase of decision trees in the forest. This infers that the training and test times (the smaller the magnitude of the metric, the better the performance) are negatively affected by the growth of trees in the forest. Therefore, the trade-off between the accuracy by growing trees in the forest and the efficiency of test time should be thoroughly made. In terms of training time, RF performed the fastest among the three classifiers compared to SVM (subsection 5.1) and MLP (subsection 5.3), i.e. the training time per sample in RF is in the order of microseconds.

Peer-reviewed version available at Appl. Sci. 2020, 10, 3980; doi:10.3390/app10113980

5.3. Multi-layer Perceptron Classifier for UWB Localization System

MLP is a type of deep feedforward artificial neural networks, which contains at least three layers (an input layer, a hidden layer, and an output layer) in a single network [45]. Typically, the neurons in the hidden and output layers of the MLP use nonlinear activation functions such as sigmoid, ReLU, and Softmax. The term deep is usually applied when there is more than one hidden layer in the network. MLP utilizes the backpropagation algorithm [46] for training the network. In this paper, the MLP classifier is configured using the rectified linear unit (ReLU) as the activation function for the hidden layers, the Softmax function as the output layer, and the Adam (adaptive moment estimation) as an optimization algorithm. The maximum number of epochs was set to 500 allowing early stopping if the training loss has not improved for 10 consecutive epochs.

Table 4. Comparison of the RF configurations based on the numbers of decision trees in the forest

No. of Neurons in each Hidden Layers	**No. of Hidden Layers**	**Mean Accuracy with std** (%)	**Mean Training Time per sample** (ms)	**Mean Test Time per sample** (μs)
50	1	84.93 ± 0.26	$\mathbf{0.68} \pm 0.27$	$\mathbf{1.26} \pm 0.06$
	2	88.77 ± 0.53	1.45 ± 0.07	3.23 ± 0.55
	3	90.45 ± 0.51	2.57 ± 0.25	5.46 ± 0.20
	4	90.95 ± 0.35	2.84 ± 0.63	8.79 ± 0.41
	5	$\mathbf{91.04} \pm 0.66$	3.50 ± 1.07	12.20 ± 0.26
	6	90.61 ± 1.26	3.70 ± 0.74	15.25 ± 3.28
100	1	85.88 ± 0.12	$\mathbf{1.02} \pm 0.01$	$\mathbf{2.51} \pm 0.05$
	2	89.78 ± 0.59	3.75 ± 0.19	12.20 ± 1.53
	3	91.36 ± 0.54	5.68 ± 1.23	18.12 ± 1.71
	4	$\mathbf{91.38} \pm 0.44$	5.08 ± 1.67	23.32 ± 3.75
	5	90.85 ± 0.67	7.90 ± 2.60	29.24 ± 0.94
	6	91.33 ± 0.44	9.42 ± 3.90	32.47 ± 1.81

The evaluations of MLP were conducted for up to 6 fully connected hidden layers in two conditions using 50 and 100 neurons in each hidden layer (Table 4). The results show that there was a significant increase in the overall accuracy by adding a second and third hidden layer to the network. However, the improvement is thin when more than 3 hidden layers are used for a network. In terms of the number of neurons per layer in the network, the use of 100 neurons in each hidden layer constantly beats the use of 50 up to four consequent layers. However, the difference cannot be clearly distinguished when more than 4 layers are used in the network.

In terms of the processing time (training and test times), adding more hidden layers and more neurons in the network have negative impact on the performance (the last two columns in Table 4), i.e. the lower the processing time, the better the performance. Therefore, the trade-off between the accuracy and processing time is necessary to make for efficient performance. The results in Table 4 suggests that the use of 3 hidden layers in which each contains 100 neurons seems a good choice for solving the evaluated UWB-based multi-class classification problem.

5.4. Section Summary

In summary, RF is the fastest among the three methods for the given data-set in terms of the training time (Figure 4). This is because the training time per sample data in RF only takes in the order of microseconds. Meanwhile, SVM and MLP are in the order of several milliseconds depending on the configuration and setup. In terms of test time, SVM performs worst among the three with a test time in the order of milliseconds while RF and MLP are in the order of microseconds.

Taking into consideration the evaluated results presented in this section, we established the configurations of three classifiers for further processing in section 7. The summarized overview is represented in Figure 4. The selected configurations are the radial basis function kernel approach for SVM, 50 independent decision-tree estimators for RF, and three hidden fully connected layers with 100 neurons in each layer for MLP network.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 29 April 2020 doi:10.20944/preprints202004.0503.v1

Peer-reviewed version available at Appl. Sci. 2020, 10, 3980; doi:10.3390/app10113980

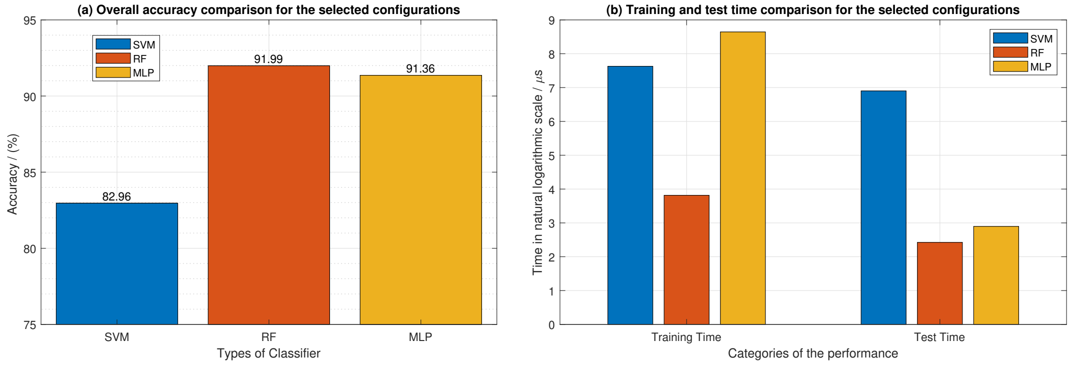

Figure 4. Summary of the chosen configurations for the three classifiers, which is RBF kernel for SVM, 50 estimators for RF, and 3 hidden layers with 100 neurons in each for MLP.

6. Data Preprocessing and Feature Selection

This section examines the impact of feature selections (subsection 6.1) and features scaling, i.e. standardization technique in this manuscript (subsection 6.2), for the evaluated three ML models. The experimental results presented in section 7 were performed based on the outcomes of this section.

6.1. The Impact of Feature Extraction in the evaluated Machine Learning Models

Based on the extracted features defined in subsection 4.3, the performance comparison of feature extractions for five categories are illustrated in Figure 5. The five categories are built upon: (i) 12 extracted features (i.e. the full features in our evaluation), (ii) 10 features excluding standard and maximum noises, (iii) 5 features, i.e. the reported distance, the CIR, and the first, second and third harmonics of the FP, (iv) 3 features, i.e. reported distance, the CIR, and the first harmonics of FP, and (v) 2 features, i.e. the CIR and first harmonic of FP.

Regarding feature extractions of the UWB measurement data, Figure 5 indicates that a notable degradation in accuracy occurred to the three ML models when two features are applied in the evaluation. The rest of the categories (starting from 3 to 12 features) provides more or less comparable results. Moreover, we noticed during the evaluation that using the reported distance as a feature plays an important role in feature extractions for UWB data. Furthermore, this is also the metric that we are mostly interested in estimating the position in UWB. We also observed that the contribution of the amplitudes of the three harmonics (first, second, and third) in the FP signal implies comparable

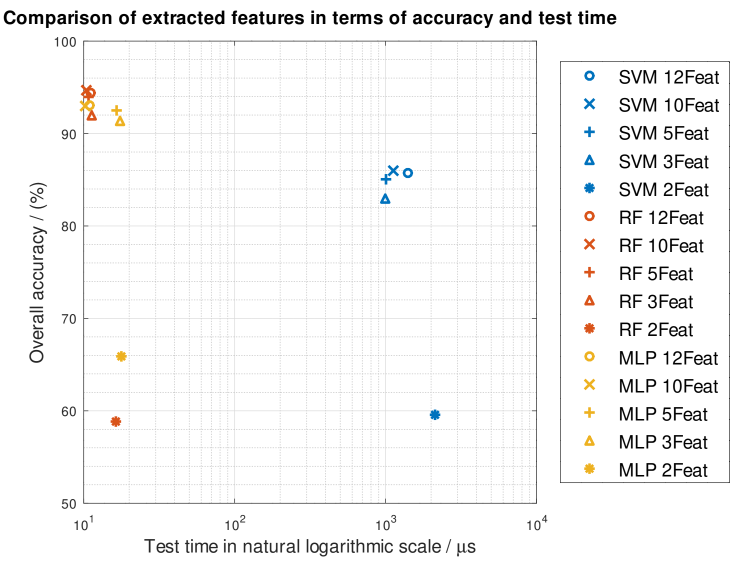

Figure 5. Performance comparison of three ML models (SVM, MLP, and RF) using different extracted features at training and test data collected at the same scenarios

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 29 April 2020 doi:10.20944/preprints202004.0503.v1

Peer-reviewed version available at Appl. Sci. 2020, 10, 3980; doi:10.3390/app10113980

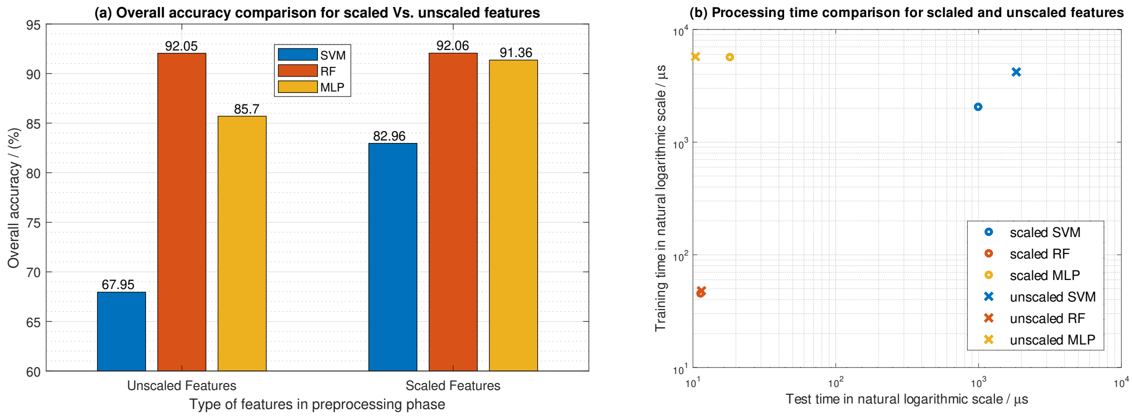

Figure 6. Comparison of the overall accuracy, training and test times for the scaled vs. unscaled features in preprocessing phase.

impacts. This infers that picking any one of them as a feature provides an equivalent performance in case of feature reductions. The amplitude of the CIR is undoubtedly important features in UWB, which represents a vital role in the identification of NLOS in the conventional technique using (2).

The empirical results in Figure. 5 suggests that the most suitable choice for the evaluation in terms of minimum features and optimal performance is to use three features. Therefore, the experimental evaluation results presented in section 7 are based on three features specifically the reported distance, the amplitude of CIR signal, and the amplitude of the first harmonics in the FP signal.

6.2. The Impact of Feature Scaling in the evaluated Machine Learning Models

Feature scaling is a model-dependent parameter in ML. It is a technique to normalize or standardize the range of independent variables or features of input measured data in a preprocessing step. This typically allows a faster training time and a better performance in many ML models. This section briefly reveals the effects of feature scaling in the three evaluated ML models. Besides, there exist ML models, where their performance is not affected by the feature scaling in the preprocessing of the input data. A good representative of such model in our evaluation is the RF classifier (Figure 6). In this paper, the feature scaling was performed using standardization technique. This typically means rescaling the data in preprocessing to have a mean of 0 and a standard deviation of 1 (unit variance).

Figure 6 (a) depicts the impact of feature scaling for the three ML models in terms of the overall accuracy. Scaling the input data in the preprocessing phase has a notable impact on the SVM and MLP classifiers in terms of accuracy. In SVM, the overall accuracy was improved from 67.95 to 82.95 % by scaling the features of the input data. Similarly, the accuracy of the MLP was increased from 85.70 to 91.36 %. However, the RF gave equivalent outcomes in both scaled and unscaled features.

In terms of training time using three features in each model, SVM learns significantly faster when feature scaling is used (Figure 6 (b)). To be precise, the training time of SVM reduced from 4195.14 ± 210.38 µs to 1977.12 ± 185.45 µs by scaling the features. However, both RF and MLP do not show obvious improvement for training time in our small-scaled three features evaluation. Specifically, the training time of RF for the scaled and unscaled features are 44.97 ± 2.38 µs and 47.98 ± 4.74 µs respectively. Likewise, the training time of MLP for the scaled and unscaled features are 5562.66 ± 1313.85 µs and 5737.97 ± 1949.05 µs respectively.

In terms of test time, feature scaling hurts the performance of MLP (Figure 6 (b)). In specific, the value of the test time (the smaller, the better) in MLP degraded from 10.41 ± 0.54 µs to 17.58 ± 1.27 µs by feature scaling. On the contrary, the performance of test time in SVM improves when feature scaling is applied, i.e. 1839.53 ± 53.70 µs for unscaled features and 956.28 ± 13.94 µs for scaled features. Again, RF does not show any significant improvements except a small variation in its standard deviation, which implies 11.48 ± 0.99 µs for unscaled features and 11.48 ± 1.42 µs for scaled features respectively.

 doi:10.20944/preprints202004.0503.v1

Peer-reviewed version available at Appl. Sci. 2020, 10, 3980; doi:10.3390/app10113980

7. Evaluation Results

This section examined the experimental evaluation results of three ML classifiers based on two quantitative metrics: (i) an F1-score which is used to compare the performance of the three evaluated classifiers in this paper (subsection 7.1), and (ii) a confusion matrix that gives the insightful representation of the reported results for each individual classifiers (subsection 7.2).

7.1. Performance comparison of the Three Classifiers using Macro-averaging F1-score as a metric

To give an overview of the actual state in each trial conducted for 10 times, we use the macro-averaging F1-score to compare the performance of the three classifiers in this section. F1-score, i.e. in contrast to the overall accuracy score in confusion matrix (subsection 7.2), is extensively used to quantify the classifier's performance in ML because it takes into account both the precision and recall to compute the decisive score [11,47]. It is the harmonic mean of the precision and recall that can be expressed for a binary classification as:

$$F1 = 2 \cdot \frac{Precision \cdot Recall}{Precision + Recall} \tag{4}$$

For multi-class classification, there are two typical ways (macro-averaging and micro-averaging) to compute the overall F1-score for a classifier [47]. We applied the macro-averaging technique in our evaluation that treats all the classes equally. Based on the mentioned macro-averaging F1-score, Figure 7 compares the experimental evaluation results of the three classifiers at two test environments (the scenario which is the same as vs. different from the training state) defined in subsubsection 4.2.2. The solid lines denote the results of the test data-set when the data in the test state is collected at a different environment from the training state. The dotted lines tell the results of the test data-set achieved from the same scenario as the training state.

In general, a significant gap between the two test scenarios was discovered in the experimental evaluation results of the three classifiers (Figure 7). The figure reveals that impressive outcomes were achieved in RF and MLP classifiers when the test and training states were conducted in the same environments. However, the performance of SVM is relatively low in this scenario compared to RF and MLP. In contrast, the performance of all classifiers was notably degraded when the test environment is different from the training state. Specifically, the resultant mean of the SVM classifier based on the macro-averaging F1-score is reduced from 0.83 (when training and test scenarios are the same) to 0.75 (when training and test scenarios are different). Similarly, the performance of the RF decreased

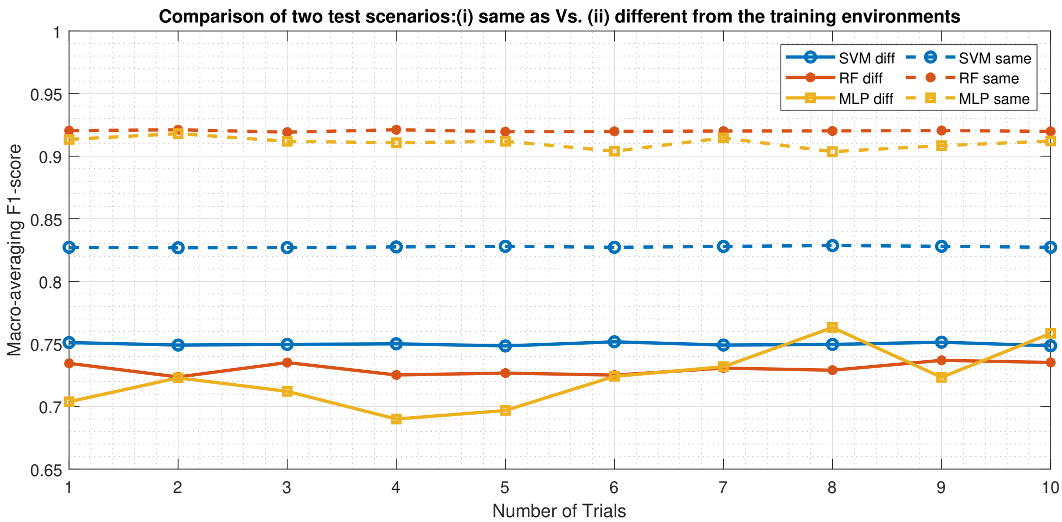

Figure 7. Performance comparison of the evaluated three classifiers based on the macro-averaging F1-score at two different scenarios: (i) dotted lines denoted the training and test data came from the same environment, (ii) solid lines denoted the training and test data came from different environments.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 29 April 2020 doi:10.20944/preprints202004.0503.v1

Peer-reviewed version available at Appl. Sci. 2020, 10, 3980; doi:10.3390/app10113980

from 0.92 to 0.73. MLP classifier shows a degradation from 0.91 to 0.72. The results show that an immediate conclusion and judgment upon the choice of classifiers based on a single test scenario or environment could be misleading. The core reason is that the quality of the measured wireless signal, i.e. the UWB ranging data in our evaluation, is affected by a variety of physical impacts in indoor environments, as previously mentioned several times in this manuscript.

It is interesting to see that SVM stands out to be the best classifier in our evaluation when the test scenario is different from the training state (Figure 7). Though it is the poorest performance among the three classifiers when the same environments of training and test were applied. Moreover, the outcomes of SVM were consistent in all of the evaluated trials at both of the two scenarios. Indeed, the outcomes of RF were also quite stable across the whole trials compared to the MLP classifier. However, a lot of fluctuation were evident in the predicted outputs of the MLP, especially in the condition where the training and test environments were different.

In all experiments, the lowest F1-score was 0.69 in trial no. 4 when MLP is used as a classifier and the highest score reached 0.92 using RF. This outcome shows that the ML-based classifications, regardless of the type of the classifier, are more effective in the multi-class identification of UWB data than traditional approaches described in section 2.

7.2. Results Representation of the Evaluated Three Classifiers using Confusion Matrix

To examine a more extensive study, the comparative analysis of the two test scenarios for each classifier was conducted using the confusion matrix in this section. In the confusion matrix (Figure 8, 9, and 10), the output class in Y-axis refers to the prediction of the classifier and the target class in X-axis refers to the true reference class. The overall accuracy of the classifier is given in the bottom right corner of each confusion matrix. The last column in each category of the confusion matrix indicates the precision (positive predictive value) and its counterpart false discovery rate (FDR) of the classifier. Likewise, the last row in each category gives the recall (sensitivity or true positive rate) and its complement false-negative rate (FNR). The correct predictions for each category are expressed in the diagonal of the confusion matrix. The values in off-diagonal correspond to the Type-I and Type-II errors. For a scenario where training and test data-sets were collected in different environments, the confusion matrices presented in this section are based on the mode of each classifier (i.e. we chose the most frequently predicted class in each trial as our estimator output). For a scenario where the training and test environments are the same, the confusion matrix is based on the trial no. 5 out of 10 trials

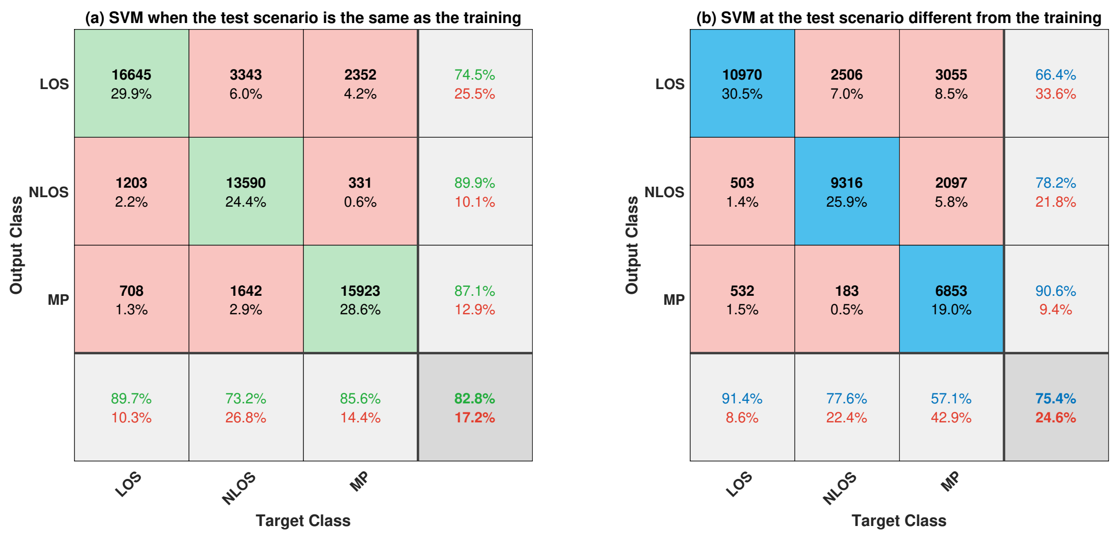

Figure 8. Comparison of the multi-labels classification results for SVM using the confusion matrix in two different scenarios: (a) the test data-set was obtained from the same environments as in the training state, (b) the test data-set was collected in the environments different from the training state. In the evaluation, the radial basis function was used as the kernel for the SVM classifier.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 29 April 2020 doi:10.20944/preprints202004.0503.v1
Peer-reviewed version available at Appl. Sci. 2020, 10, 3980; doi:10.3390/app10113980

described in subsection 7.1. The reason is that the random splitting of the test data-set for the true class in this scenario was different for each trial. Moreover, all trials in this scenario gave comparable results as reported in Figure 7.

7.2.1. Comparative Analysis of the Two Test Scenarios for SVM Classifier

The insightful comparison of two test scenarios based on the confusion matrix for SVM is presented in Figure 8. The result shows that the overall accuracy of the SVM significantly dropped, i.e., from 82.8 % to 75.4 %, when the tested data set was different from the training state. We observed that this is the cause of a significant decrease in the identification process of the MP condition. By comparing the two test scenarios in Figure 8 (a) and 8 (b), the predicted accuracy of the MP conditions in SVM was declined from 28.6 % to 19.0 %. Meanwhile, there existed no sharp deviations in the predicted accuracy of the LOS and NLOS conditions at both test scenarios. This increases the misclassification rate of LOS and NLOS conditions as an MP in the SVM classifier.

To be precise, a significant misclassification rate of MP condition as an NLOS was detected in the evaluated results, i.e. the value rises from 0.6 % (when training and test are in the same condition) to 5.8 % (when training and test are in different conditions) as provided in Figure 8. The misclassification rate of LOS as MP was also quite high, i.e. it increases from 4.2 % to 8.5 %. The main reason could be the data collected for MP conditions when the two transceivers are too close to each other. In that case, it is acceptable to interpret the received signal in MP condition as a LOS.

Regarding the NLOS condition in both test scenarios, we observed that a quite good outcome in the predicted accuracy, precision, and recall of the SVM classifier (Figure 8). This result is crucial because the main impact on the performance of the UWB localization algorithm is the NLOS condition (section 2). The misclassification of the LOS as an NLOS does not produce a severe consequence on the overall performance of the UWB system. The reason is that the location algorithm assigns different weights upon the classification results and giving LOS as a smaller weight doesn't hurt the system performance.

7.2.2. Comparative Analysis of the Two Test Scenarios for RF Classifier

Similar to the SVM classifier, the overall accuracy of the RF classifier degraded strikingly from 91.9 to 73.5 when the test and training environments were different (Figure 9). Again, the cause of this significant decrease in RF was evident in the predicted accuracy of the MP conditions, i.e. it reduces

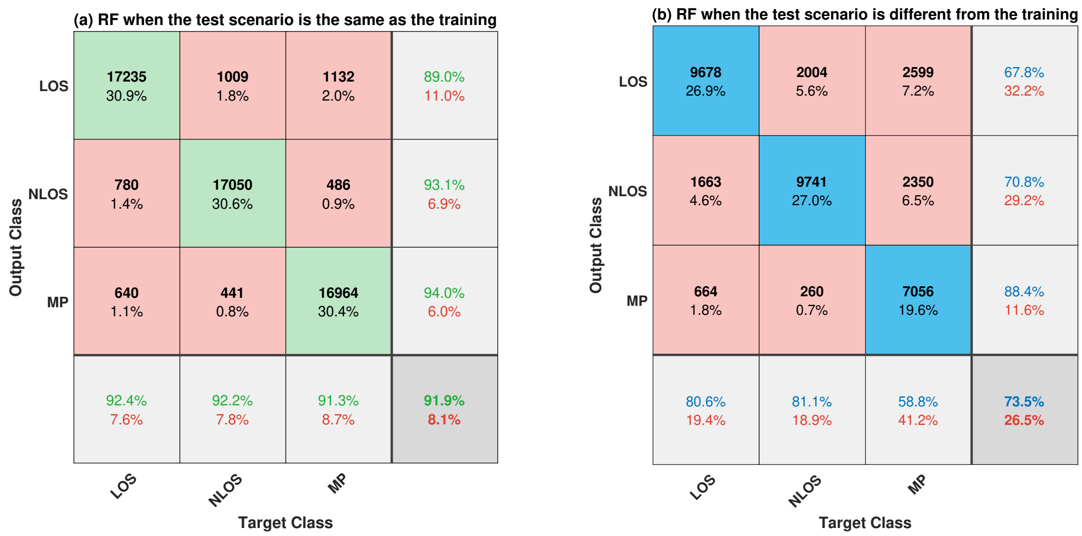

Figure 9. Comparison of the multi-class classification results for RF classifier using the confusion matrices at two scenarios: (a) the test data-set obtained from the same conditions as the training data, (b) the test data-set collected at a different condition from the training. In the evaluation, 50 decision trees were used in the forest of the applied RF classifier.

 doi:10.20944/preprints202004.0503.v1

Peer-reviewed version available at Appl. Sci. 2020, 10, 3980; doi:10.3390/app10113980

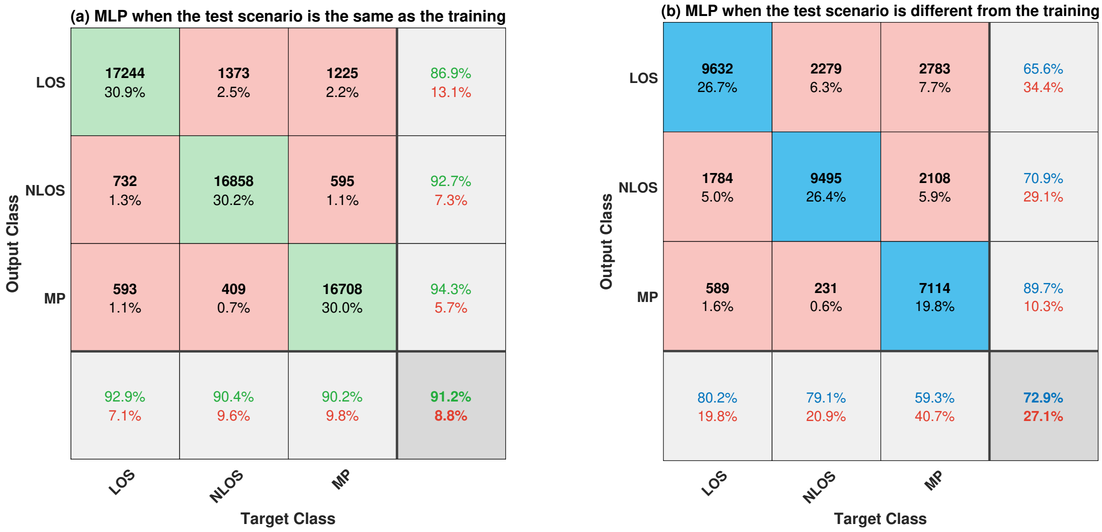

Figure 10. Comparison of the multi-class classification results for the MLP classifier using confusion matrices at two scenarios: (a) the test data-set obtained from the same conditions as the training phase, (b) the test data-set collected at a different condition from the training. In the evaluation, three fully connected hidden layers with 100 neurons in each layers were used for the MLP classifier.

from 30.4 % to 19.6 %. Unlike the SVM classifier, the predicted accuracy of the RF classifier in both LOS and NLOS conditions are noticeably declined as well. This implies the predicted accuracy of LOS degraded from 30.9 % to 26.9 % while NLOS decreased from 30.6 % to 27.0 %. This pushes the overall accuracy of the RF classifier worse than the SVM in the scenario where the training and test environments are different.

Similar to the aforementioned reason and condition in the SVM classifier, the misclassification rate of the MP condition as either the LOS or NLOS is quite high in the RF classifier as well (Figure 9). To give with the exact values, the misclassification rate of the MP condition as NLOS was increased from 0.9 % to 6.5 %. Similarly, the misclassification of the MP as LOS was grown from 2.0 % to 7.2 %. In the RF classifier, the misclassification rate of the NLOS as a LOS was also noticeably high, i.e. it increases from 1.8 % to 5.6 %. Again, the predicted accuracy of the NLOS condition in both of the two test scenarios was quite satisfying. That is 30.6 % out of 100 % for the three classes in the scenario when the training and test data were in the same environments. Similarly, it is 27.0 % when the data for the training and test environments were different.

7.2.3. Comparative Analysis of the Two Test Scenarios for MLP Classifier

Figure 10 compares the multi-class classification results of two test scenarios using MLP as a classifier. Similar to previously mentioned two classifiers, the overall accuracy of the MLP considerably declined from 91.2 % to 72.9 % when the test environments of the UWB was different from the training state. Repeatedly, this is caused by the false discovery rate of the MP conditions in the measured UWB data. In specific, the predicted accuracy of MP condition in MLP classifier decreases from 30 % to 19.8 % out of the 100 % for three classes. This outcome shows that the MP condition is the challenging class to identify throughout our evaluation in all classifiers (Figure 8, 9, and 10).

The decrease of performance in a particular class makes an increase in the false discovery rate of other classes. Specifically for our evaluation in MLP, the misclassification rate of MP as LOS increases from 2.2 % to 7.7 % while MP as NLOS increases from 1.1 % to 5.9 % (Figure 10). In fact, the predicted accuracy of both the LOS and NLOS were also declined in MLP classifier, i.e. from 30.9 % to 26.7 % in LOS condition and from 30.2 % to 26.4 %.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 29 April 2020 doi:10.20944/preprints202004.0503.v1

Peer-reviewed version available at Appl. Sci. 2020, 10, 3980; doi:10.3390/app10113980

Table 5. Summary of the results based on the F1-scores and overall accuracy

Scenarios	Classifiers	Individual F1-scores			Macro-averaging F1-scores	Overall Accuracy (%)
		LOS	NLOS	MP		
Training and Test environments are different	SVM	0.77	**0.78**	0.70	**0.75**	**75.35**
	RF	0.74	**0.76**	0.71	0.73	73.52
	MLP	0.72	**0.75**	0.71	0.73	72.86
Training and Test environments are the same	SVM	0.81	0.81	**0.86**	0.83	82.80
	RF	0.91	**0.93**	**0.93**	**0.92**	**91.90**
	MLP	0.90	**0.92**	**0.92**	0.91	91.20

7.3. Summary of the Experimental Evaluation Results

In this paper, the multi-label classification results of the UWB data were quantified based on two metrics, i.e. F1-score and confusion matrix. F1-score is typically used in literature to quantify the performance of ML-based classifiers because it provides a convenient single value score [11,40,47]. However, it can sometimes overlook the insightful information of some classes. Table 5 gives the typical summary of the two evaluated scenarios using both the F1-score and overall accuracy. The individual F1-score for each class (LOS, NLOS and MP) is also given in the table.

Based on the data given in Table 5, both the macro-averaging F1-score and overall accuracy show that the SVM classifier gave the best performance when the training and test environments are different. The consistent outcome in SVM also reveals the reason why it is one of the most frequently used classifiers in literature [9,10,15,19]. In contrast, RF performed the best among the three evaluated classifiers regarding the same training and test environments.

However, it is hard to clarify using Table 5 that the predicted accuracy of MP condition is significantly low compared to the other two classes in all evaluated classifiers at a scenario when training and test are in different environments. This phenomenon can be detected using the confusion matrix as previously described in subsection 7.2.

8. Discussions

The identification process of the LOS, NLOS, and MP condition in a wireless ranging system, especially in UWB, is crucial because they strongly influence the quality and accuracy of the actual measurement. For that reason, several contributions have been proposed in the literature to identify these conditions as already mentioned in section 1. However, most of the contributions treated the issue as a binary classification problem (section 2). To the best of our knowledge, only two papers [11,19] addressed the UWB-based classification process as a multi-class problem. In this paper, we defined three classes in UWB measurement data (LOS, NLOS, and MP conditions as presented in section 2) and evaluate three ML classifiers (SVM, RF, and MLP as presented in section 5) to identify these defined three classes.

Two metrics, F1-score and confusion matrix, were used for evaluating the performance of each classifier. Apart from these scores, the training and test times for each classifier were also given in our evaluation. As a matter of fact, this type of metrics (training and test times) is typically ignored in literature. However, it is undeniable that the magnitude of a test time in a certain classifier is usually vital in the overall performance of the system. Our results presented in section 5 reveals that SVM has the poorest performance among the three classifiers in terms of the test time. In contrast, SVM gave the best performance in terms of F1-score when different environments of training and test states were applied (section 7). The mentioned two results told us the interesting contradiction of two metrics in SVM classifier, which could be an important factor in system implementation of some ML applications.

The evaluation results based on the F1-score and overall accuracy pointed out that the measured environments in UWB have a strong effect on the performance of the three classifiers (section 7). This refers to the striking degradation of the performance in all classifiers when different environments of training and test were applied in the evaluation. One may argue that it is caused by the overfitting of the model in the training state. However, this outcome occurred in all of the three evaluated classifiers regardless of the hyper-parameters used in each classifier. Indeed, it is a part of the generalization problem caused by the inadequate representation of the conditions in the data. However, the data collection process in UWB, especially for the mentioned three classes, is quite time-consuming, elaborate, and costly because of the nature of the wireless signal. This means the evaluated outcomes will be affected differently in several different ways by the conditions of materials, types of walls, types of furniture, etc. in the measured environment. Our attempt is to provide the feasibility of the ML approach in the multi-class classification of the UWB measurement data in contrast to the conventional technique given in section 2. The results based on F1-score in our experiments show that 0.69 in the worst-case scenario and 0.92 in the best-case scenario. The outcomes are quite satisfying and promising.

In general, the classification results of UWB in literature were usually reported using single value metrics, i.e. the overall accuracy score [19,21], and F1-score [11]. In some cases, the recall and true negative rate were applied in addition to the accuracy score as in [36]. Typically in a binary classification problem, receiver operating characteristic (ROC) curve [48] and cumulative distribution function [9,16] were widely used. Those scores and metrics are particularly helpful for the comparative analysis of different classifiers. However, the insightful details of the actual conditions are usually overlooked or missed in some cases. Therefore, we used the complete confusion matrix in this paper to examine the individual outcomes of each class in all evaluated classifiers (subsection 7.2). Using the confusion matrix, we observed that the predicted accuracy of the MP condition is significantly dropped in all classifiers when different environments of training and test were used in the data. This condition cannot be clearly evident using other above-mentioned metrics. This incident also proves that the identification of the MP condition is more challenging than the other two classes (NLOS and LOS). Moreover, we observed that the predicted accuracy of the NLOS condition is generally quite good in all of the evaluations regardless of the classifiers. At the same time, the misclassification rate of NLOS as either a LOS or MP condition is also moderately low. This criterion is crucial because the highest error rate in UWB measurements was usually caused by the NLOS condition (section 2). In contrast, the misidentification of LOS as an NLOS condition in a certain model does not hurt the overall system performance of the UWB application.

9. Conclusions

In this paper, the multi-label (LOS, NLOS, and MP) identification for UWB ranging was conducted using three machine learning techniques (SVM, RF, MLP) as a classifier. This is in contrast to the typical binary classification approaches in the literature. The experimental evaluation results based on F1-score proved that ML-based classifiers can identify the defined three classes with a high score, i.e. 0.69 in the worst-case scenario and 0.92 in the best-case scenario. However, it is unreasonable to single out the best classifier out of the three because their performances depend extensively on the environmental changes and the metrics used to quantify them. Moreover, the insightful results based on the confusion matrix revealed that the MP condition is the most challenging to identify among the three classes. It is also evident in the confusion matrices that the predicted accuracy of the NLOS condition is quite high throughout all the evaluated experiments.

As future work, the NLOS condition can be further classified into two classes based on the study conducted in [11]. This will lead the defined classes of UWB data into four, i.e. LOS, MP, soft-NLOS, and hard-NLOS. Indeed, this also means much more data will be needed to collect to identify these four conditions in several environments using several different materials and objects. Besides, the experimental evaluation will be conducted in the actual hardware of the microcontroller-based platform instead of PC.

Author Contributions: "conceptualization, C.L.S.; methodology, C.L.S. and J.D.H.; validation, B.S., J.D.H., M.A., M.H. and U.R.; software, C.L.S., B.S. and M.A.; writing-original draft preparation, C.L.S.; writing-review and

editing, B.S., J.D.H., M.A., M.H. and U.R.; visualization, C.L.S., J.D.H., M.A. and M.H.; project administration, M.H. and U.R.; supervision, U.R.; funding acquisition, U.R."

Funding: We acknowledge support for the Article Processing Charge by the Deutsche Forschungsgemeinschaft and the Open Access Publication Fund of Bielefeld University.

Acknowledgments: This work was supported by the Cluster of Excellence Cognitive Interaction Technology 'CITEC' (EXC 277) at Bielefeld University, funded by the German Research Foundation (DFG). Author Cung Lian Sang was partly supported in this work by the German Academic Exchange Service (DAAD). The authors are responsible for the contents of this publication.

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

Abbreviations

The following abbreviations are used in this manuscript:

AltDS-TWR	Alternative Double-sided Two-way Ranging
BDT	Boosted Decision Tree
CIR	Channel Impulse Response
CNN	Convolutional Neural Network
FP	First Path
GP	Gaussian Process
HSI	High Speed Internal (clock)
KNN	K-nearest Neighbor
LOS	Line of Sight
MCU	Microcontroller Unit
ML	Machine Learning
MLP	Multi-layer Perceptron
MP	Multi-path
NLOS	Non-line of Sight
RBF	Radial Basis Function
RF	Random Forest
RSS	Received Signal Strength
RX	Received or Receiver
SNR	Signal to Noise Ratio
SVM	Support Vector Machine
TOF	Time-of-Flight
UWB	Ultra-Wideband

References

1. Win, M.Z.; Dardari, D.; Molisch, A.F.; Wiesbeck, W.; Zhang, J. History and Applications of UWB [Scanning the Issue]. *Proceedings of the IEEE* **2009**, *97*, 198–204. doi:10.1109/JPROC.2008.2008762.
2. Sang, C.L.; Adams, M.; Korthals, T.; Hörmann, T.; Hesse, M.; Rückert, U. A Bidirectional Object Tracking and Navigation System using a True-Range Multilateration Method. 2019 International Conference on Indoor Positioning and Indoor Navigation (IPIN), 2019, pp. 1–8. doi:10.1109/IPIN.2019.8911811.
3. Dardari, D.; Conti, A.; Ferner, U.; Giorgetti, A.; Win, M.Z. Ranging With Ultrawide Bandwidth Signals in Multipath Environments. *Proceedings of the IEEE* **2009**, *97*, 404–426. doi:10.1109/JPROC.2008.2008846.
4. Win, M.Z.; Scholtz, R.A. Impulse radio: how it works. *IEEE Communications Letters* **1998**, *2*, 36–38. doi:10.1109/4234.660796.
5. Win, M.Z.; Scholtz, R.A. Ultra-wide bandwidth time-hopping spread-spectrum impulse radio for wireless multiple-access communications. *IEEE Transactions on Communications* **2000**, *48*, 679–689. doi:10.1109/26.843135.
6. Sang, C.L.; Adams, M.; Hesse, M.; Hörmann, T.; Korthals, T.; Rückert, U. A Comparative Study of UWB-based True-Range Positioning Algorithms using Experimental Data. 2019

16th Workshop on Positioning, Navigation and Communications (WPNC), 2019, pp. 1–6. doi:10.1109/WPNC47567.2019.8970249.

7. Sang, C.L.; Adams, M.; Hörmann, T.; Hesse, M.; Porrmann, M.; Rückert, U. An Analytical Study of Time of Flight Error Estimation in Two-Way Ranging Methods. 2018 International Conference on Indoor Positioning and Indoor Navigation (IPIN), 2018, pp. 1–8. doi:10.1109/IPIN.2018.8533697.
8. Lian Sang, C.; Adams, M.; Hörmann, T.; Hesse, M.; Porrmann, M.; Rückert, U. Numerical and Experimental Evaluation of Error Estimation for Two-Way Ranging Methods. *Sensors* **2019**, *19*, 616. doi:10.3390/s19030616.
9. Maranò, S.; Gifford, W.M.; Wymeersch, H.; Win, M.Z. NLOS identification and mitigation for localization based on UWB experimental data. *IEEE Journal on Selected Areas in Communications* **2010**, *28*, 1026–1035. doi:10.1109/JSAC.2010.100907.
10. Weijie Li.; Tingting Zhang.; Qinyu Zhang. Experimental researches on an UWB NLOS identification method based on machine learning. 2013 15th IEEE International Conference on Communication Technology, 2013, pp. 473–477. doi:10.1109/ICCT.2013.6820422.
11. Barral, V.; Escudero, C.J.; García-Naya, J.A.; Maneiro-Catoira, R. NLOS Identification and Mitigation Using Low-Cost UWB Devices. *Sensors* **2019**, *19*, 3464. doi:10.3390/s19163464.
12. Guowei Shen.; Zetik, R.; Thoma, R.S. Performance comparison of TOA and TDOA based location estimation algorithms in LOS environment. 2008 5th Workshop on Positioning, Navigation and Communication, 2008, pp. 71–78. doi:10.1109/WPNC.2008.4510359.
13. Khodjaev, J.; Park, Y.; Saeed Malik, A. Survey of NLOS identification and error mitigation problems in UWB-based positioning algorithms for dense environments. *annals of telecommunications - annales des télécommunications* **2010**, *65*, 301–311. doi:10.1007/s12243-009-0124-z.
14. Guvenc, I.; Chong, C.; Watanabe, F. NLOS Identification and Mitigation for UWB Localization Systems. 2007 IEEE Wireless Communications and Networking Conference, 2007, pp. 1571–1576. doi:10.1109/WCNC.2007.296.
15. Huang, C.; Molisch, A.F.; He, R.; Wang, R.; Tang, P.; Ai, B.; Zhong, Z. Machine Learning-Enabled LOS/NLOS Identification for MIMO System in Dynamic Environment. *IEEE Transactions on Wireless Communications* **2020**, pp. 1–1. doi:10.1109/TWC.2020.2967726.
16. Wymeersch, H.; Marano, S.; Gifford, W.M.; Win, M.Z. A Machine Learning Approach to Ranging Error Mitigation for UWB Localization. *IEEE Transactions on Communications* **2012**, *60*, 1719–1728. doi:10.1109/TCOMM.2012.042712.110035.
17. Savic, V.; Larsson, E.G.; Ferrer-Coll, J.; Stenumgaard, P. Kernel Methods for Accurate UWB-Based Ranging With Reduced Complexity. *IEEE Transactions on Wireless Communications* **2016**, *15*, 1783–1793. doi:10.1109/TWC.2015.2496584.
18. Zandian, R.; Witkowski, U. NLOS Detection and Mitigation in Differential Localization Topologies Based on UWB Devices. 2018 International Conference on Indoor Positioning and Indoor Navigation (IPIN), 2018, pp. 1–8. doi:10.1109/IPIN.2018.8533781.
19. Kolakowski, M.; Modelski, J. Detection of direct path component absence in NLOS UWB channel. 2018 22nd International Microwave and Radar Conference (MIKON), 2018, pp. 247–250. doi:10.23919/MIKON.2018.8405190.
20. Niitsoo, A.; Edelhäusser, T.; Mutschler, C. Convolutional Neural Networks for Position Estimation in TDoA-Based Locating Systems. 2018 International Conference on Indoor Positioning and Indoor Navigation (IPIN), 2018, pp. 1–8. doi:10.1109/IPIN.2018.8533766.
21. Cwalina, K.K.; Rajchowski, P.; Blaszkiewicz, O.; Olejniczak, A.; Sadowski, J. Deep Learning-Based LOS and NLOS Identification in Wireless Body Area Networks. *Sensors* **2019**, *19*, 4229. doi:10.3390/s19194229.
22. Leech, C.; Raykov, Y.P.; Ozer, E.; Merrett, G.V. Real-time room occupancy estimation with Bayesian machine learning using a single PIR sensor and microcontroller. 2017 IEEE Sensors Applications Symposium (SAS), 2017, pp. 1–6. doi:10.1109/SAS.2017.7894091.
23. Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Dubourg, V.; Vanderplas, J.; Passos, A.; Cournapeau, D.; Brucher, M.; Perrot, M.; Duchesnay, E. Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research* **2011**, *12*, 2825–2830.

24. Borras, J.; Hatrack, P.; Mandayam, N.B. Decision theoretic framework for NLOS identification. VTC '98. 48th IEEE Vehicular Technology Conference. Pathway to Global Wireless Revolution (Cat. No.98CH36151), 1998, Vol. 2, pp. 1583–1587 vol.2. doi:10.1109/VETEC.1998.686556.
25. Schroeder, J.; Galler, S.; Kyamakya, K.; Jobmann, K. NLOS detection algorithms for Ultra-Wideband localization. 2007 4th Workshop on Positioning, Navigation and Communication, 2007, pp. 159–166. doi:10.1109/WPNC.2007.353628.
26. Denis, B.; Daniele, N. NLOS ranging error mitigation in a distributed positioning algorithm for indoor UWB ad-hoc networks. International Workshop on Wireless Ad-Hoc Networks, 2004., 2004, pp. 356–360. doi:10.1109/IWWAN.2004.1525602.
27. Wu, S.; Zhang, S.; Xu, K.; Huang, D. Neural Network Localization With TOA Measurements Based on Error Learning and Matching. *IEEE Access* **2019**, *7*, 19089–19099. doi:10.1109/ACCESS.2019.2897153.
28. Maali, A.; Mimoun, H.; Baudoin, G.; Ouldali, A. A new low complexity NLOS identification approach based on UWB energy detection. 2009 IEEE Radio and Wireless Symposium, 2009, pp. 675–678. doi:10.1109/RWS.2009.4957442.
29. Xiao, Z.; Wen, H.; Markham, A.; Trigoni, N.; Blunsom, P.; Frolik, J. Non-Line-of-Sight Identification and Mitigation Using Received Signal Strength. *IEEE Transactions on Wireless Communications* **2015**, *14*, 1689–1702. doi:10.1109/TWC.2014.2372341.
30. Gururaj, K.; Rajendra, A.K.; Song, Y.; Law, C.L.; Cai, G. Real-time identification of NLOS range measurements for enhanced UWB localization. 2017 International Conference on Indoor Positioning and Indoor Navigation (IPIN), 2017, pp. 1–7. doi:10.1109/IPIN.2017.8115877.
31. Schmid, L.; Salido-Monzú, D.; Wieser, A. Accuracy Assessment and Learned Error Mitigation of UWB ToF Ranging. 2019 International Conference on Indoor Positioning and Indoor Navigation (IPIN), 2019, pp. 1–8. doi:10.1109/IPIN.2019.8911769.
32. Decawave. DW1000 User Manual: how to use, configure and program the DW1000 UWB transceiver, 2017.
33. Musa, A.; Nugraha, G.D.; Han, H.; Choi, D.; Seo, S.; Kim, J. A decision tree-based NLOS detection method for the UWB indoor location tracking accuracy improvement. *International Journal of Communication Systems* **2019**, *32*, e3997, [https://onlinelibrary.wiley.com/doi/pdf/10.1002/dac.3997]. e3997 dac.3997, doi:10.1002/dac.3997.
34. Smith, D.R.; Pendry, J.B.; Wiltshire, M.C.K. Metamaterials and Negative Refractive Index. *Science* **2004**, *305*, 788–792, [https://science.sciencemag.org/content/305/5685/788.full.pdf]. doi:10.1126/science.1096796.
35. Bregar, K.; Hrovat, A.; Mohorcic, M. NLOS Channel Detection with Multilayer Perceptron in Low-Rate Personal Area Networks for Indoor Localization Accuracy Improvement. Proceedings of the 8th Jožef Stefan International Postgraduate School Students' Conference, Ljubljana, Slovenia, 2016, Vol. 31.
36. Krishnan, S.; Xenia Mendoza Santos, R.; Ranier Yap, E.; Thu Zin, M. Improving UWB Based Indoor Positioning in Industrial Environments Through Machine Learning. 2018 15th International Conference on Control, Automation, Robotics and Vision (ICARCV), 2018, pp. 1484–1488. doi:10.1109/ICARCV.2018.8581305.
37. Fan, J.; Awan, A.S. Non-Line-of-Sight Identification Based on Unsupervised Machine Learning in Ultra Wideband Systems. *IEEE Access* **2019**, *7*, 32464–32471. doi:10.1109/ACCESS.2019.2903236.
38. STMicroelectronics. UM1724 User manual: STM32 Nucleo-64 boards (MB1136), 2019.
39. STMicroelectronics. STM32L476xx Datasheet - Ultra-low-power Arm Cortex-M4 32-bit MCU+FPU, 100DMIPS, up to 1MB Flash, 128 KB SRAM, USB OTG FS, LCD, ext. SMPS, 2019.
40. Charte, F.; Rivera, A.J.; del Jesus, M.J.; Herrera, F. Addressing imbalance in multilabel classification: Measures and random resampling algorithms. *Neurocomputing* **2015**, *163*, 3 – 16. Recent Advancements in Hybrid Artificial Intelligence Systems and its Application to Real-World Problems Progress in Intelligent Systems Mining Humanistic Data, doi:https://doi.org/10.1016/j.neucom.2014.08.091.
41. Cortes, C.; Vapnik, V. Support-vector networks. *Machine Learning* **1995**, *20*, 273–297. doi:10.1007/BF00994018.
42. Suykens, J.; Vandewalle, J. Least Squares Support Vector Machine Classifiers. *Neural Processing Letters* **1999**, *9*, 293–300. doi:10.1023/A:1018628609742.
43. Smola, A.J.; Schölkopf, B. A tutorial on support vector regression. *Statistics and Computing* **2004**, *14*, 199–222. doi:10.1023/B:STCO.0000035301.49549.88.

44. Breiman, L. Random Forests. *Machine Learning* **2001**, *45*, 5–32. doi:10.1023/A:1010933404324.
45. Hastie, T.; Tibshirani, R.; Friedman, J. *The Elements of Statistical Learning*; Springer Series in Statistics, Springer New York Inc.: New York, NY, USA, 2001.
46. Rumelhart, D.E.; Hinton, G.E.; Williams, R.J. Learning representations by back-propagating errors. *Nature* **1986**, *323*, 533–536. doi:10.1038/323533a0.
47. Sokolova, M.; Lapalme, G. A systematic analysis of performance measures for classification tasks. *Information Processing & Management* **2009**, *45*, 427 – 437. doi:https://doi.org/10.1016/j.ipm.2009.03.002.
48. Choi, J.; Lee, W.; Lee, J.; Lee, J.; Kim, S. Deep Learning Based NLOS Identification With Commodity WLAN Devices. *IEEE Transactions on Vehicular Technology* **2018**, *67*, 3295–3303. doi:10.1109/TVT.2017.2780121.

Sample Availability: The experimental research data and the corresponding source code used in this paper are publicly available in PUB - Publication at Bielefeld University [Pub link].