RSSI based Device-free Human Identification

Isuru Dharmadasa*, Malith Weerasekara†, Chamath I Keppitiyagama‡ Chandana Gamage§

*The University of Adelaide, Australia
†SCoRe Lab, Sri Lanka
‡University of Colombo School of Computing, Sri Lanka
§University of Moratuwa, Sri Lanka

Abstract—Researchers have explored many methods and techniques for human detection and identification in diverse contexts. One such approach is studying variations of Radio Frequency (RF) signals (e.g., Received Signal Strength Indicator (RSSI)). RSSI based techniques have been widely used in human detection, but not for recognizing the identity of a person from a group of people due to its noisy nature. This research focused on investigating the possibilities and limitations of device-free human identification using WiFi RSSI data with machine learning-based classification techniques. To inspect the characteristics of WiFi RSSI data, the authors have conducted multiple statistical analyses. A Kalman filter was applied to minimize the noise in WiFi RSSI data, followed by a feature extraction process. Furthermore, the authors have conducted several research experiments in different configurations of receivers and participant numbers. The experimental results show that the human identification accuracy level increases with the number of receivers used for the data collection. Moreover, the authors have identified that human identification accuracy can be further improved by leveraging proper noise reduction methods and feature extraction processes. For a Kalman filter applied and feature-extracted WiFi RSSI dataset of 20 people, the Support Vector Machine (SVM) - Radial Kernel model recorded the highest average identification accuracy of 99.58%.

Index Terms—Human Identification, WiFi RSSI, Device-Free, Machine Learning, Kalman Filter

I. INTRODUCTION

Human detection and identification in open spaces are vital in many research as well as for industry-level applications. Human detection refers to distinguishing a human's presence from other physical objects (e.g., animals). In contrast, human identification refers to the unique identification of a particular person from a finite set of individuals. In this paper, the size of that particular set is limited to 20 since a typical office or meeting room accommodates around 15- 20 people.

Some applications, such as trip-wire systems, need only to detect a human's presence, whereas sophisticated authentication systems need to uniquely identify the person. Hence, depending on the system requirements, system developers/researchers must deploy suitable human detection and identification models, accordingly.

Many research studies have investigated and implemented human detection and identification in the current literature. The use of properties and behaviors of Radio Frequency (RF) signals is one such approach. When RF signals travel through solid objects and open spaces, it attenuates and demonstrates variations in the RF signal strength. This behavior has been researched to apply in the localization and human identification applications. Moreover, since RF signals can penetrate through solid objects, these techniques can be used in dark environments as well [1]. Therefore, it has the potential to be used in many scenarios even when the traditional human identification methods such as video cameras and thermal cameras fail to provide sufficient accuracy.

RF signals comprise multiple attributes from both frequency and time domains. The Received Signal Strength Indicator (RSSI) is one such obtainable attribute from any wireless transmitter. Moreover, it is available in widely used technologies such as WiFi, Bluetooth, and ZigBee. The primary purpose of RSSI is to provide information on the signal strength to other wireless devices [2]. It has been identified that when RF signals travel through objects, RSSI variations occur, and these variations can be used for human detection in open spaces [1]. Moreover, Wilson and Patwari have named this RSSI reduction as 'shadow-loss' in their research [1].

According to the existing literature, the emerging device-free localization approach 'Radio Tomographic Imaging (RTI)' is based on 'shadow-loss' of RSSI values [1]. Furthermore, researchers have used these RTI-based techniques for human tracking and activity detection as well [3], [4], [5]. Since participants do not need to carry any device dedicated for the identification, these approaches are considered device-free methods. However, it is important to note that RTI or RSSI based approaches for unique human identification are limited in the current literature.

Researchers have highlighted that RSSI data suffers from high noise and lack sufficient information to identify humans, uniquely [6], [7]. Therefore, WiFi Channel State Information (CSI) has been used for human identification studies [8], [9], [10]. It is due to the reason that CSI data are considered information-rich, as those provide the information including, but not limited to signal amplitude, phase, Doppler-shift, delay-spread and channel frequency response. Moreover, these

studies utilize supervised machine learning models for the classification process. Similar to RSSI based approaches, these techniques also follow device-free principles.

Even though the current literature has limited attention over RSSI based human identification, RSSI data are widely available from many wireless networks and can be extracted even without connecting to the network. Most importantly, opposed to CSI, RSSI does not need any special device or equipment for the data extraction as any wireless receiver can capture it [11]. Hence, RSSI based approaches have a wide range of applicability in real-world scenarios. However, due to the noisy nature and lack of information, it is crucial to examine the feasibility of using RSSI data for unique human identification. Therefore, it is evident that there is a clear research gap on identifying and exploring possibilities of using RSSI based human identification methods and designing solutions while addressing the real-world setting limitations.

This research contributes to the human detection and identification research domains by empirically investigating the possibilities and limitations of using WiFi RSSI as a device-free human identification method. This study aims to observe, analyze, and critically evaluate the human identification accuracy levels based on different machine learning models, experimental setup arrangements, and data processing methods. As a preliminary study, a statistical analysis was conducted to examine the properties and behavior of WiFi RSSI data. Then, several research experiments were performed to explore the human identification accuracy differences according to the used machine learning model, experimental setup configurations and multiple data processing steps.

The rest of the paper is arranged as follows. Section 2 presents the most related existing literature on human detection and identification approaches based on radio waves. The research methodology, consisting of preliminary studies and research experiments, is explained under section 3. A critical analysis of the research experiment results is presented in section 4, and finally, section 5 presents the research conclusion and possible future directions.

II. BACKGROUND

This section presents the existing literature on RF signal-based device-free human detection and identification under three main subsections.
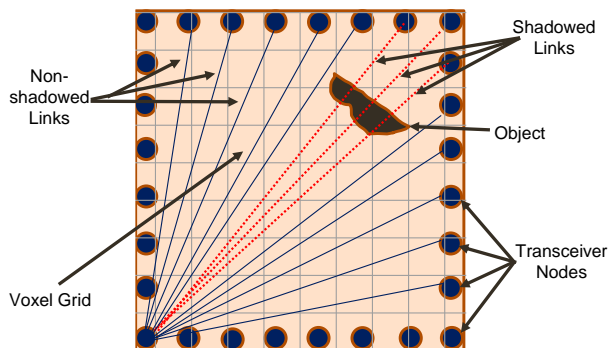
A. RSSI based Human Detection and Tracking

Wilson and Patwari have presented RSSI based Radio Tomographic Imaging (RTI) approach for localization and human presence detection [1]. They have stated that this method can be used in many real-world scenarios, such as emergency rescue missions and physical intrusion detection systems. Furthermore, RTI is considered a device-free approach since external transceivers collect RSSI data instead of participants carrying devices. The RTI method is based on *Computed Tomography* principles, where signal strength is considered as the main measurement. Figure 1 presents a typical setup used in RTI-based research studies. In one of their studies, Wilson and Patwari have collected RSSI data by using 28 transceiver nodes in a 21x21 foot square open space. For the data collection, the researchers have used TelosB wireless nodes that are based on IEEE 802.15.4 standard and function on the 2.4GHz frequency. They have concluded that the proposed RTI method can visualize human's presence based on RSSI attenuation values [1]. As an extension to this research, the same researchers have presented an improved human tracking method using RTI techniques on RSSI data [3]. In their approach, a Kalman filter has been used for noise reduction, and variance-based RTI, a variation of RTI, has been used to improve human presence detection and tracking accuracy [3]. Following a similar approach Piumwardane et al. [12] and Niroshan et al. [13] have empirically investigated the possibility of applying these RTI techniques in WiFi networks and increase the human detection accuracy through a human-interference model, respectively.

Scholz et al. [14] have described an RSSI-based human activity recognition approach, which compares the accuracy levels of both device-free and device-bound methods . The researchers have deployed transceivers based on IEEE 802.15.4 standard and have used machine learning models for activity classification. It was asserted that the both device-free and device-bound methods had yielded similar activity recognition accuracy levels [14].

Konings et al. [15] have proposed an RSSI based device-free localization method, in contrast to RTI methods, which does not require an offline calibration. However, similar to RTI methods, transceiver locations must be prior known in this approach. Their introduced spring relaxation approach has utilized RSSI data from commercial off-the-shelf (COTS) wireless devices, and the researchers have discussed the CSI data unavailability from COTS devices as a limitation. Booranawong et al. [16] have conducted a literature review on existing filtering methods for RSSI data and have presented a novel filtration method. In this method, the researchers have considered both the accuracy and computational complexity of the RSSI data filtration. Subsequently, the researchers have introduced an adaptive filtration method for human detection and tracking with reduced computational complexity [16].

Kaltiokallio et al. [17] have proposed a novel Bayesian filter for RSSI based human localization and tracking. The proposed Bayesian filter method leverages an image-based position

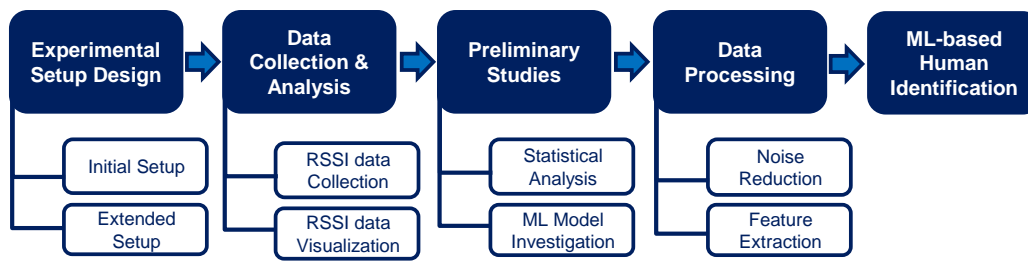Fig. 1: An illustration of a typical RTI setup

Fig. 2: Research methodology

estimation model and it resolves the divergence issue faced by Bayesian filter based methods. The researches have stated that their approach outperforms the state-of-the-art imaging based localization methods by 30% - 48%. Panwar et al. [18] have proposed a novel localization method leveraging both time-of-arrival and RSSI information. The authors have claimed that their approach works in non-line of situations without prior knowledge of the sensing environment. Furthermore, they have used a majorization-minimization algorithm to increase the computational efficiency and complexity.

B. CSI based Human Identification

Zhang et al. [8] were the first to use WiFi signals for unique human identification . In their research, WiFi CSI data has been collected for 6 people by asking them to walk between two transceivers. Their research methodology is comprise of silence removal, segmentation, feature extraction, and classification. For a group of 6 people, they have achieved a human identification accuracy level of 77% [8]. Hong et al. have also followed a similar approach for unique human identification [9]. Their research has explored the possibility of using WiFi CSI data for human identification in 3 scenarios. Those are standing, walking, and marching. They have proposed a novel CSI feature named as subcarrier-amplitude frequency (SAF) to use in the classification process with SVM - Linear kernel machine learning model [9].

'Wii' [10] is another attempt to identify humans uniquely using WiFi CSI data with SVM - Radial Kernel model. A Principle Component Analysis (PCA) has been conducted, which is followed by a low pass filter to minimize the noise in CSI data. In their research experiments, over 1500 gait instances have been recorded for eight human subjects. They have achieved over 90% accuracy level and have discussed the possibility of using 'Wii' in home security systems as well [10].

Nipu et al. [7] have evaluated two machine learning-based classification models to determine the human identification accuracy level for a group of five participants. According to the yield results, decision tree model and random forest model has recorded average human identification accuracy of 84% and 78%, respectively. Mo et al. [19] have developed a deep learning model, Convolutional Long and Short-Term Memory (CLSTM), to identify humans uniquely leveraging WiFi CSI data. They have proposed a data augmentation method to reduce the data collection cost. Their approach has achieved

92% accuracy for a group of 8 people. Zou et al. [20] have proposed a WiFi CSI data underpinned human identification system, *AutoID*. They have leveraged human gait information and considered CSI data as human *fingerprint*. Their approach has achieved 91% accuracy for a group of 20 people. Similarly, Ming et al. [21] also have leveraged human gait for uniquely identifying humans using CSI data. They have used a LSTM model with WiFi CSI data and achieved 96% accuracy for a group of 24 people.

In a different approach, Chen et al. [22] have proposed a *fusion* method to integrate the information yielded through WiFi CSI data and camera videos. Importantly, this approach has achieved 97.01% *real-time* detection accuracy for a group of 25 people. However, in this method, the participants had to carry a WiFi device and the device ID alongside with WiFi CSI data were being used to build user profiles.

C. RSSI based Human Identification

'Radio Biometrics' by Xu et al. [6] is one key research that has examined and compared both WiFi CSI and RSSI data for human identification. Data collection has been carried out with the participation of 12 people. Furthermore, for data collection, a WiFi chipset comprise of 3×3 multi-in multi-out (MIMO) antenna has been used. According to their novel approach, *Radio Biometrics Refinement Algorithm*, CSI based technique has recorded a 98.78% accuracy level, whereas the RSSI based method has achieved only 31.93% accuracy level for human identification. According to research conclusions, they researchers have asserted that it is mainly due to the noise in WiFi RSSI data [6]. In a similar study, Zhanyong et al. [23] have proposed, *CrossSense*, a Wifi sensing data approach that consider both RSSI and CSI data for gait and gesture recognition. Their experimental results also suggests that WiFi CSI data based models have a higher accuracy compared to RSSI models.

Dharmadasa et al. [24] have proposed WiFi RSSSI based human identification approach using a supervised machine learning method. The authors have applied multinomial logistic regression model on WiFi RSSI data for both human detection and identification. But the study has limited to 5 participants and has not considered any noise reduction techniques. Moreover, it has used 7 transmitters and a receiver for data collection, which is challenging to implement in real-world scenarios.

D. Summary

In summary, it is evident that RSSI based methods and techniques are widely being used for human detection and localization. Approaches such as RTI have employed a large number of transceivers for data collection, making them impractical to use in real-world settings. Furthermore, to use RTI methods, prior knowledge of transceiver placement is also needed, which makes it further challenging to use in closed or private environments. Since RSSI suffer from noise, researchers are more interested in using WiFi CSI data for human identification. Hence, the literature on RSSI based human identification is limited. Many WiFi CSI based approaches have used machine-learning models for human classification. However, it was identified that participant number (sample size) is low in many research and has not investigated the possibility of using multiple transceivers for data collection. Hence, a research niche is visible in WiFi RSSI based human identification methods in terms of data collection, processing and classification phases.

III. RESEARCH METHODOLOGY

Figure 2 shows the main stages of followed in this research study.

A. Experimental Setup Design and Implementation

All research experiments were conducted in a covered empty room. Figure 3a illustrates the room dimensions and the placement of the partition board wall. The main reason for separating the WiFi receivers from the WiFi transmitter was to simulate the scenario where an outsider unauthorizedly monitoring the inside of a room. Furthermore, as shown in Figure 3b, the WiFi transmitter was placed 2.75m above the room floor, where it was centered on the wall. As per Figure 3b, WiFi receiver was placed 0.5m above the room floor. Both these placements were made to ensure that the experimental setup resembles a real-world setting. To reduce any interference that could cause due to movements, the researchers ensured that, except for individuals who participated in the data collection and the researcher who was conducting the experiments, no one else was present in the experimental setup environment while collecting the data. Since this research aimed at exploring the possibilities of using RSSI based for human identification in real-world settings, all other environmental factors were kept unchanged. For example, the researchers did not use shields to reduce the interference from other WiFi networks to ensure that the experimental environment is as similar as possible to real-world settings.

Prolink - H5004NK WiFi router was used as the WiFi transmitter, and Dell Inspiron 5110 laptops were used as WiFi receivers, which were equipped with Intel Centrino N1030 WiFi adapters. The laptops were installed with Ubuntu-18.10-Desktop (64bit) operating system. The researchers developed a shell script to scan and capture RSSI values from the WiFi transmitter and store them in laptop hard-drives. Furthermore, after several attempts, 0.1s was selected as the scanning time. Details on the enhanced experimental setup design are presented at the end of this section.

B. Data Collection and Analysis

1) RSSI data collection: The data collection was carried out with the participation of 20 people. A convenient sampling method was used for selecting these participants, and their composition is as follows.

- Age: 20 - 60 years
- Gender: Male - 11, Female - 09
- Height: 1.48m - 1.86m
- Weight: 44kg - 92kg

As shown in Figure 4, the room floor center was marked to place human subjects. The following steps were carried out in the data collection process.

- **Step 01:** Turn on the WiFi Transmitter (Tx)
- **Step 02:** Turn on and set WiFi Receiver (Rx) to listening/scanning mode (scanning time - 0.1 s)
- **Step 03:** Place the human subject in the **X** marked location
- **Step 04:** Record WiFi RSSI values from the transmitter through the receiver (time duration - 120s)
- **Step 03** and **Step 04** were repeated for all twenty participants.

2) RSSI data visualization: To examine patterns and anomalies of the collected WiFi RSSI data, those were visualized using line graphs. As shown in Figure 5, it was observed that in the first and last 10 seconds (approximately) of the data collection period, WiFi RSSI data shows exceptional variations. It was identified that it was due to the movements of the person conducting the research experiments. As the researcher walks near WiFi receivers to start and end the data collection process, the researcher's movements have impacted the WiFi RSSI values. Hence, to minimize the interference that happened due to these external factors, the first and last 10 seconds of each WiFi RSSI data record were filtered out.

C. Preliminary Studies

1) Statistical analysis on WiFi RSSI data: As the first step, a normality test was conducted on the collected WiFi RSSI data to investigate whether WiFi RSSI data follows a normal distribution in the presence of people. For each data record, *Shapiro-Wilk Normality Test* was performed with the following null hypothesis.

Hypothesis H0: WiFi RSSI data has a normal distribution. Since the returned *p-value* for each data record was lower than 0.05, null hypothesis was rejected and decided to use non-parametric tests for the next steps.

To determine whether the collected RSSI data is statistically different from person to person, the researchers conducted *Mann Whitney U Test* for all participant pairs with the following null hypothesis.

Hypothesis H0: X and Y data samples are taken from the same population. Here, *X* and *Y* refer to WiFi RSSI data records of two people that participated in the data collection process. The Mann Whitney U Test was conducted with a 95% confidence level and since the returned *p-value* was lower than 0.05, null hypothesis was rejected. Hence, it was asserted that the WiFi
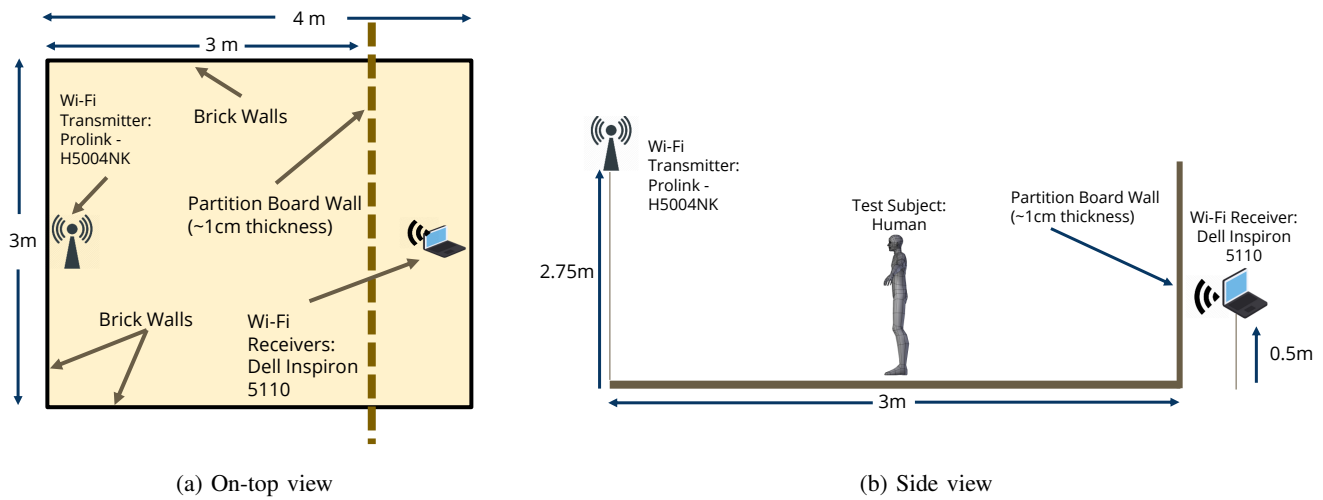
(a) On-top view

(b) Side view

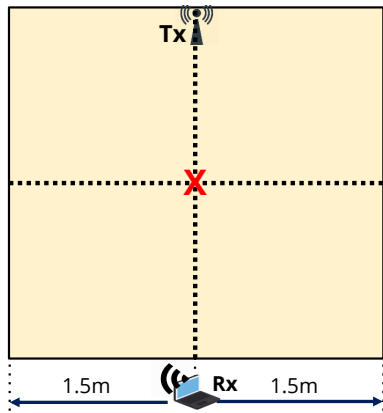Fig. 3: Experimental setup design and arrangement.
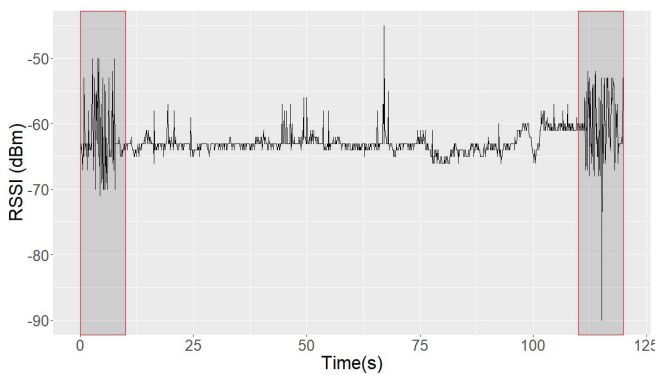

Fig. 4: Human subject placement

Fig. 5: WiFi RSSI visualization

RSSI data statistically different from person to person. In other words, it was decided to consider WiFi RSSI data as the *signature* of a particular person.

2) Investigation on machine learning models: The existing literature was examined to determine the machine learning models that are applied in RSSI/CSI based human identification studies. It was identified that researchers have leveraged Support Vector Machine (SVM) based machine learning models and Multinomial Logistic Regression model for human identification in both CSI and RSSI based research [9], [10], [24]. Hence, SVM-Linear Kernel (SVM-L), SVM-Radial Kernel (SVM-R), SVM- Polynomial Kernel (SVM-P), SVM-Sigmoid Kernel (SVM-S) and Multinomial Logistic Regression (MLR) model were examined in the preliminary experiments of this study. Here, Human Identification Accuracy (HIA) refers to correctly identifying a particular person from a given set of participants. The following formula was used to calculate the HIA for a particular person. **CP** and **TP** denote the number of correctly identified RSSI data points and total number of data points, respectively. The average human identification accuracy refers to the average value of all HIAs for a particular set of participants.

$$HIA = (CP/TP) * 100 \qquad (1)$$

For the preliminary study, all 20 participants were considered but limiting to WiFi RSSI data from a single receiver (as shown in Figure 4). Table I presents the average human identification accuracy along side with the used machine learning model. Since, Multinomial Logistic Regression, SVM - Linear kernel, and SVM-Radial Kernel models outperformed other models, those were selected for the next phase of the research experiments. Furthermore, *k-Fold Cross-Validation* method was used for the evaluation, where *k* value was set to 10, after considering the size of the WiFi RSSI dataset.

D. Data Processing

1) WiFi RSSI data noise reduction: WiFi RSSI data considered to be noisy and highly susceptible to interference.

TABLE I: Preliminary Results - Average human identification accuracy (for a set of 20 people) with the applied machine learning model

Model	Accuracy (%)
SVM-Radial Kernel	32.63
SVM-Linear Kernel	31.05
SVM-Polynomial Kernel	10.08
SVM-Sigmoid Kernel	9.82
Multinomial Logistic Regression	30.51

Physical phenomenon such as reflection and scattering can cause multi-path fading, which leads to unreliable RSSI. Furthermore, WiFi strength can be drastically reduced when traveling through physical objects. Hence, it is essential to use suitable noise filtering/reduction methods when applying RSSI data driven method on real-world applications [2], [25].

In RSSI based indoor localization approaches, Bulten et al., Li et al., and Xu et al. have used Kalman filter to smoothen and reduce noise in RSSI data [2], [26], [27]. Furthermore, Joey Wilson and Neal Patwari also have used a Kalman filter for RSSI based human detection and tracking [3]. Kalman filter, which is also known as linear quadratic estimation (LQE) is an iterative process which provides state estimations on non-observable variable based on observed noisy data. There are two conditions that must be fulfilled to apply a regular Kalman filter [28]. Those are, model or the system should be linear and observed error must follow a Gaussian distribution. The collected RSSI data of this study fulfills both those conditions. RSSI data were collected on stationary people, hence the estimation model can be approximated as a linear model. And as per the existing literature, RSSI data noise (e.g., multi-path fade, device sensing errors) can be approximated as a Gaussian distribution [3], [2].

The researchers adopted the approach followed by Wouter et al. [2], where they apply a Kalman filter on noisy RSSI data in a stationary environment setting. The following steps explains the application of a Kalman filter on RSSI data.

Equation 2 defines a general transition model.

$$x_t = A_t X_{t-1} + B_t u_t + \epsilon_t \quad (2)$$

x_t is the current state vector where A is a state transition matrix of the previous state vector x_{t-1}. B is an input matrix of the u control input vector. ϵ represents system noise.

Since in this study transmitter and receivers are static and test subject (human participant) remains stationary throughout the data collection time it can be assumed that RSSI values should be consistent, and all the variations are due to the noise. Based on that assumption equation 2 can be simplified as below.

$$x_t \approx x_{t-1} + \epsilon_t \quad (3)$$

Equation 4 defines the observational model.

$$z_t = C_t x_t + \delta_t \quad (4)$$

z_t is the state which is produced due to the x_t measurement and δ_t noise. C_t is a transformation matrix. Since in this

research both state and measurement are equal, equation 3 and 4 can be combined as shown in the equation 5.

$$z_t = C_t x_t + \delta_t \approx x_t + \delta_t \quad (5)$$

To update the Kalman filter, the following steps were performed.

$$\overline{\mu}_t = \mu_{t-1} \quad (6)$$

$$\overline{\Sigma}_t = \Sigma_{t-1} + R_t \quad (7)$$

$\overline{\mu}_t$ defines the expected state, which is based on the previous state and $\overline{\Sigma}_t$ denotes certainty of the prediction which is again based on the previous certainty. R_t is the system noise.

Equation 8 defines a simplified Kalman gain.

$$K_t = \overline{\Sigma}_t (\overline{\Sigma}_t Q_t)^{-1} \quad (8)$$

Q denotes the measurement noise, which in this research related to the variance of the RSSI data.

$$\mu_t = \overline{\mu}_t + K_t (z_t - \overline{\mu}_t) \quad (9)$$

$$\Sigma_t = \overline{\Sigma}_t - (K_t \overline{\Sigma}_t) \quad (10)$$

Equation 9 and 10 defines the Kalman filter update steps. μ_t is the final predicted value from the system.

Accordingly, in this research, a Kalman filter was applied to reduce the noise in WiFi RSSI data. KalmanJS JavaScript library[1] was used to apply Kalman filter on the collected raw RSSI data. As shown in Figure 6, sharp fluctuations, and variations of the Raw-RSSI have been significantly reduced after applying this filter.

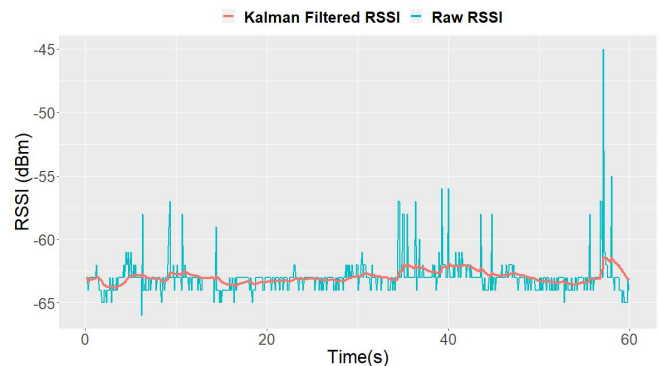[1]KalmanJS Library - https://github.com/wouterbulten/kalmanjs (Access Date: 01-06-2022)

Fig. 6: Raw RSSI and Kalaman filter applied RSSI

2) RSSI data feature extraction: Defining the feature vector is one of the critical steps that must be followed in a machine learning-based approach. Hence, to identify the most related features of WiFi data, existing literature was examined. After carefully reviewing machine learning-based human identification research, a set of RF features were listed. Subsequently, by leveraging *ReliefF*, a feature selection and ranking algorithm, the following features were selected [8], [9], [10].

- Mean
- Mode
- Max
- Min
- Skewness
- Mean-crossing-rate

E. Experimental setup enhancement

As observed from in the preliminary experiments, the human identification accuracy was too low when using only one WiFi receiver (less than 50%) and unsuitable for real-world settings. Therefore, the experimental setup was enhanced using multiple receivers to increase the average human identification accuracy. Figure 7 illustrates the receiver placement and arrangement for 2, 3, 4 and 5 WiFi receivers. The data collection process described in section III-B was followed to collect WiFi RSSI data for each of the above illustrated experimental setup configurations. A total of 300,000 WiFi RSSI data points were collected. Then, these data points were labeled with the participant's identification number (i.e., Person 1 - Person 20). For example, Table II shows a sample of a 5-receiver dataset that was later used with the machine learning models.

TABLE II: Dataset Structure

#	Label (Person No)	Rx1	Rx2	Rx3	Rx4	Rx5
1	Person 1	62	65	54	50	62
2	Person 1	62	64	53	53	60
-	Person -	-	-	-	-	-
-	Person -	-	-	-	-	-
19999	Person 20	42	55	51	50	42
20000	Person 20	45	55	61	64	62

IV. RESULTS AND ANALYSIS

This section discusses the experimental results with their interpretations and explanations.

A. Human Identification Accuracy with the Number of Participants

Figure 8 presents the box-plot diagrams prepared for average Human Identification Accuracy (HIA) with the number of participants (group size). Here, the number of participants refers to the sample size of the dataset. For example, if the number of participants is two, the dataset comprises of RSSI data records from two participants. When it is increased from two to three participants, another data record was added to the dataset. These data records were selected randomly, and each data record was considered once only for each conducted experiment. As Figure 8 clearly depicts, the average HIA has been drastically decreased with the number of participants and

a similar pattern is visible across all three machine learning models. Importantly, it was observed that the Interquartile Range (IQR) also has been increased with the number of participants resulting a low confidence on the HIA measures. In brief, the highest loss of accuracy is recorded in MLR model with a loss of 66.41% (from 90.8% to 30.5%) while SVM-R model recorded the lowest loss of 64.71% (from 93.8% to 33.6%). Furthermore, SVM-R model recorded the highest average HIA for every group size, while MLR model resulted the lowest. Thus, SVM-R model was selected for the next steps of this study.

B. Impact of Data Processing Steps on the Human Identification Accuracy

Figure 9 illustrates the box-plot diagrams on average HIA with regards to the applied data processing steps. Figure 9(a), (b), and (c) denote the data processing steps, raw RSSI, Kalman filtered RSSI, and Kalman filtered, and feature extracted RSSI data, respectively. Overall, in all three datasets, the average HIA has decreased when the number of participants incremented. In brief, the average HIA has decreased from 93.8% to 31.6% (accuracy loss: 66.31%) in raw RSSI, 94.8% to 35.4% (accuracy loss: 62.66%) in Kalman Filtered RSSI, and 98.8% to 37.4% (accuracy loss: 62.15%) in feature-extracted RSSI datasets. The highest HIA result was recorded in the SVM-R classification model with the Kalman filter applied and feature extracted RSSI dataset, which also had the lowest average human identification loss with the incremented number of participants. Thus, these results assert that the use of proper noise reduction methods and feature-extraction approaches can increase the RSSI based HIA. However, the accuracy gains from these methods become insignificant when there is a large number of participants. Therefore, to further increase the HIA, the researchers decided to increase the number of receivers for data collection.

C. Human Identification Accuracy with the Number of Receivers

Figure 10 illustrates the average HIA for 20 participants with the number of receivers used for RSSI data collection. Note that, SVM-R model with Kalman filtered and feature-extracted dataset has been used in all these instances. It is evident that the average HIA has rapidly increased with the number of receivers employed for RSSI data collection. The main reason behind this improvement is the increased level of information captured by the WiFi receivers. For example, as illustrated in Figure 11, when only one receiver is used for the data collection, it captures only a limited number of WiFi links. But when two receivers are used for data collection, receivers can intercept more RSSI links. Hence, these incorporate more information regarding the participant's physical features. Therefore, machine learning model receives more information to be trained. Thus, leads to increased human identification accuracy. This claim is already supported by the existing research, where it has stated, multiple access points provide additional information about the participants
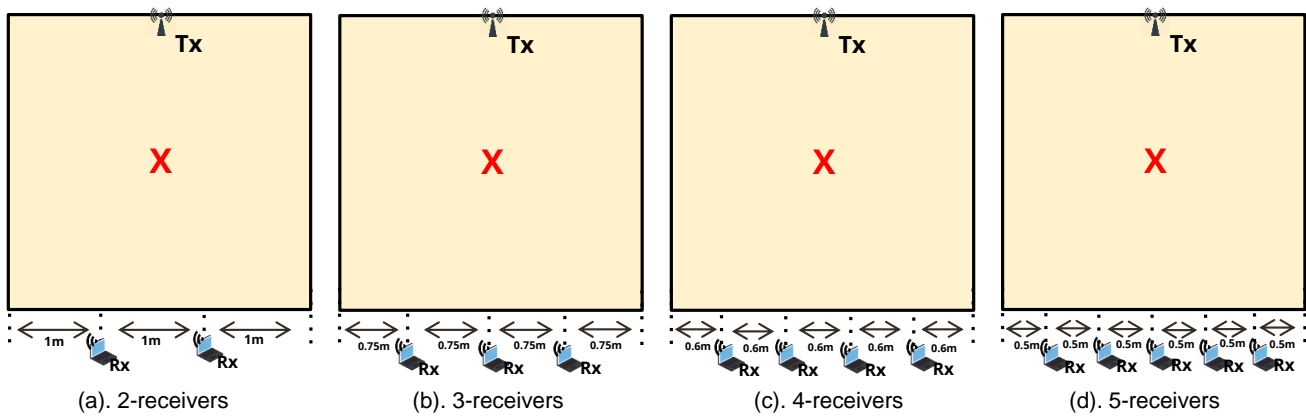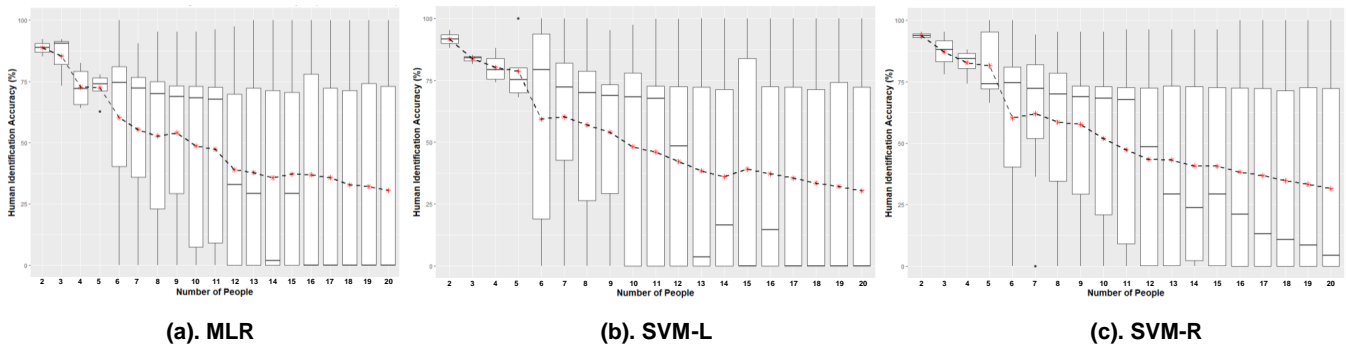
(a). 2-receivers   (b). 3-receivers   (c). 4-receivers   (d). 5-receivers

Fig. 7: Extended experimental setup

(a). MLR   (b). SVM-L   (c). SVM-R

Fig. 8: Average human identification accuracy with the number of participants.

(a). Raw   (b). Kalman filtered   (c). Feature extracted
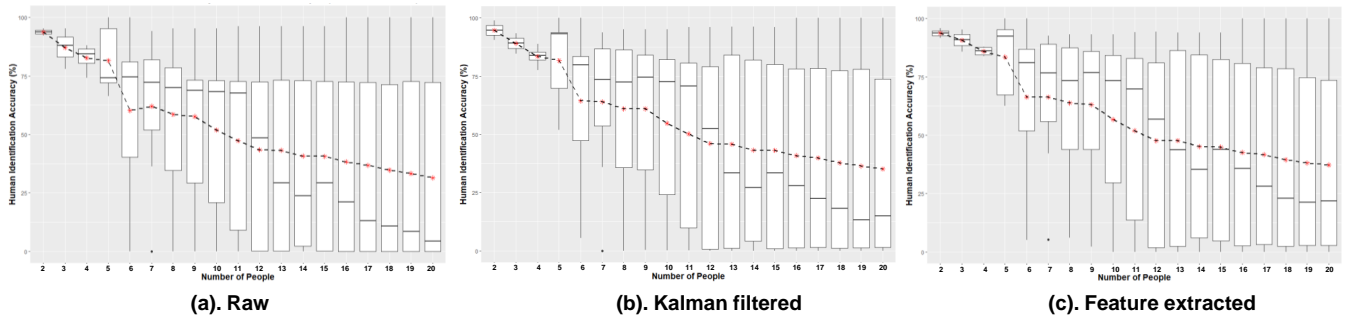
Fig. 9: Average human identification accuracy with SVM-R model and multiple data processing steps. (a). Raw RSSI, (b). RSSI data after applying a Kalman filter and (c). Feature extracted RSSI data after applying a Kalman filter

location [29]. Similarly, in 3-receivers, 4-receivers, and 5-receivers configurations also additional information has been incorporated into the machine learning models by adding extra receivers. However, as per Figure 10, it can be recognized that amount of accumulating new information gain has been decreased with the number of receivers since accuracy gain rate has decreased (e.g., 1-receiver to 2-receivers accuracy gain is 40.59% and 2-receivers to 3-receivers accuracy gain is 11.87%) with the number of receivers.

V. CONCLUSION AND FUTURE WORKS

This research investigates the possibilities and limitations of utilizing WiFi RSSI data with machine learning models as a device-free approach to identify humans, uniquely.

It was determined that the existing RTI based research has employed many transceivers for RSSI/CSI data collection and knowing the locations of these transceivers is a mandatory requirement. Thus, the applicability of these approaches in real-world settings is limited. Therefore, in this research, the researchers investigated the possible human identification accuracies that can be gained from minimal resources (i.e., one transmitter and one receiver) and the potential of increasing the
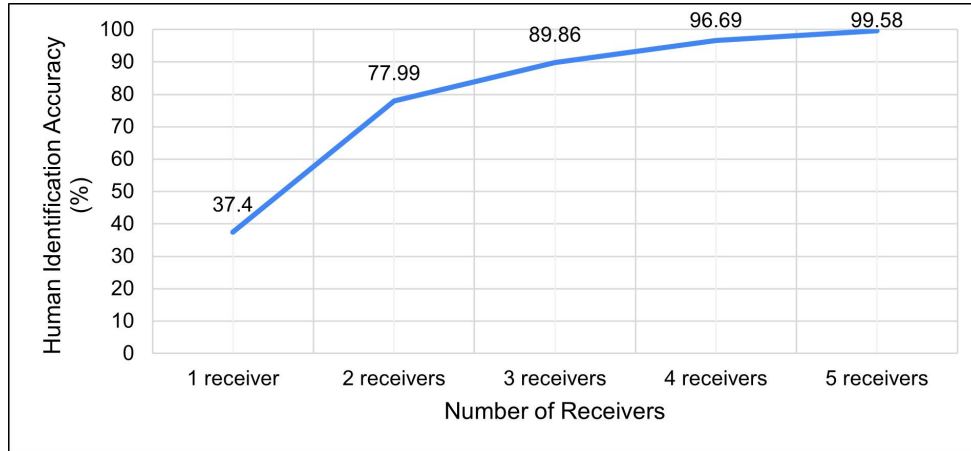
Fig. 10: Average human identification accuracy (by using SVM-R model on Kalman filter applied and feature extracted RSSI data) with the number of receivers used for RSSI data collection
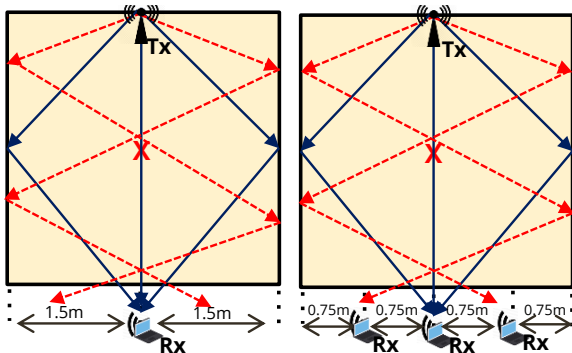

Fig. 11: An example of WiFi signal propagation

accuracy level by employing more resources (i.e., by increasing the number of receivers used for RSSI data collection).

As per the statistical analysis conducted under the preliminary study of this research, it was determined that the WiFi RSSI data can be used to distinguish humans uniquely since statistically different WiFi RSSI signatures were identified for different people.

By considering the results of all performed experiments, it was asserted that the human identification accuracy decreases with the number of participants (sample size). While data processing steps such as noise reduction and feature extraction help to increase the HIA, when there is a large number of participants, their accuracy gains become insignificant. In contrast, by increasing the number of receivers used for the RSSI data collection, a significant improvement of human identification accuracy can be achieved. However, deciding the number of receivers to be used for data collection is critical as the accuracy gaining rate decreases with the incremented number of receivers. Thus, researchers must carefully examine and balance their expected accuracy levels and the resources

at dispense.

In respect to the applied machine learning models, Support Vector Machine - Radial Kernel model showed the highest accuracy, whereas the Multinomial Logistic Regression model recorded the lowest. However, all three machine learning models followed a similar accuracy increase/decrease pattern in all experiments.

While this research presents promising results of human identification using WiFi RSSI data, it was observed that WiFi RSSI suffer from noise. Hence, as a future improvement, a comprehensive study on noise reduction mechanisms can be carried out by considering methods such as Extended Kalman Filter and Particle Filter. Furthermore, it is vital to investigate the human identification accuracy with the data collection time duration, as in this research, it was set to a fixed value. Furthermore, this study does not consider participants' physical properties (weight, height, body mass) or environmental factors (interference from other WiFi networks, different weather conditions). Therefore, future research will require to validate the proposed human identification methods accuracy and usability in different conditions and situations.

REFERENCES

[1] J. Wilson and N. Patwari, "Radio tomographic imaging with wireless networks," *IEEE Transactions on Mobile Computing*, vol. 9, no. 5, pp. 621–632, 2010.
[2] W. Bulten, A. C. Van Rossum, and W. F. Haselager, "Human slam, indoor localisation of devices and users," in *2016 IEEE First International Conference on Internet-of-Things Design and Implementation (IoTDI)*. IEEE, 2016, pp. 211–222.
[3] J. Wilson and N. Patwari, "See-through walls: Motion tracking using variance-based radio tomography networks," *IEEE Transactions on Mobile Computing*, vol. 10, no. 5, pp. 612–621, 2010.
[4] M. Bocca, O. Kaltiokallio, and N. Patwari, "Radio tomographic imaging for ambient assisted living," in *International Competition on Evaluating AAL Systems through Competitive Benchmarking*. Springer, 2012, pp. 108–130.

[5] B. Wei, A. Varshney, N. Patwari, W. Hu, T. Voigt, and C. T. Chou, "drti: Directional radio tomographic imaging," in *Proceedings of the 14th International Conference on Information Processing in Sensor Networks*, 2015, pp. 166–177.

[6] Q. Xu, Y. Chen, B. Wang, and K. R. Liu, "Radio biometrics: Human recognition through a wall," *IEEE Transactions on Information Forensics and Security*, vol. 12, no. 5, pp. 1141–1155, 2017.

[7] M. N. A. Nipu, S. Talukder, M. S. Islam, and A. Chakrabarty, "Human identification using wifi signal," in *2018 Joint 7th International Conference on Informatics, Electronics & Vision (ICIEV) and 2018 2nd International Conference on Imaging, Vision & Pattern Recognition (icIVPR)*. IEEE, 2018, pp. 300–304.

[8] J. Zhang, B. Wei, W. Hu, and S. S. Kanhere, "Wifi-id: Human identification using wifi signal," in *2016 International Conference on Distributed Computing in Sensor Systems (DCOSS)*. IEEE, 2016, pp. 75–82.

[9] F. Hong, X. Wang, Y. Yang, Y. Zong, Y. Zhang, and Z. Guo, "Wfid: Passive device-free human identification using wifi signal," in *Proceedings of the 13th International Conference on Mobile and Ubiquitous Systems: Computing, Networking and Services*, 2016, pp. 47–56.

[10] J. Lv, W. Yang, and D. Man, "Device-free passive identity identification via wifi signals," *Sensors*, vol. 17, no. 11, p. 2520, 2017.

[11] Z. Yang, Z. Zhou, and Y. Liu, "From rssi to csi: Indoor localization via channel response," *ACM Computing Surveys (CSUR)*, vol. 46, no. 2, pp. 1–32, 2013.

[12] D. Piumwardane, C. Suduwella, I. Dharmadasa, A. Sayakkara, P. Kumarasinghe, K. De Zoysa, and C. Keppitiyagama, "An empirical study of wifi-based radio tomographic imaging," in *Proceedings of the 15th ACM Conference on Embedded Network Sensor Systems*, 2017, pp. 1–2.

[13] L. Niroshan, I. Dharmadasa, S. Edirisinghe, C. Suduwella, A. Sayakkara, C. Keppitiyagama, and K. De Zoysa, "Enhancing the accuracy of wi-fi tomographic imaging using a human-interference model," in *2018 18th International Conference on Advances in ICT for Emerging Regions (ICTer)*, 2018, pp. 1–6.

[14] M. Scholz, T. Riedel, M. Hock, and M. Beigl, "Device-free and device-bound activity recognition using radio signal strength," in *Proceedings of the 4th augmented human international conference*, 2013, pp. 100–107.

[15] D. Konings, F. Alam, F. Noble, and E. M. Lai, "Springloc: A device-free localization technique for indoor positioning and tracking using adaptive rssi spring relaxation," *Ieee Access*, vol. 7, pp. 56 960–56 973, 2019.

[16] A. Booranawong, N. Jindapetch, and H. Saito, "Adaptive filtering methods for rssi signals in a device-free human detection and tracking system," *IEEE Systems Journal*, vol. 13, no. 3, pp. 2998–3009, 2019.

[17] O. Kaltiokallio, R. Hostettler, and N. Patwari, "A novel bayesian filter for rss-based device-free localization and tracking," *IEEE Transactions on Mobile Computing*, vol. 20, no. 3, pp. 780–795, 2019.

[18] K. Panwar, M. Katwe, P. Babu, P. Ghare, and K. Singh, "A majorization-minimization algorithm for hybrid toa-rss based localization in nlos environment," *IEEE Communications Letters*, vol. 26, no. 5, pp. 1017–1021, 2022.

[19] H. Mo and S. Kim, "A deep learning-based human identification system with wi-fi csi data augmentation," *IEEE Access*, vol. 9, pp. 91 913–91 920, 2021.

[20] H. Zou, Y. Zhou, J. Yang, W. Gu, L. Xie, and C. Spanos, "Wifi-based human identification via convex tensor shapelet learning," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 32, no. 1, 2018.

[21] X. Ming, H. Feng, Q. Bu, J. Zhang, G. Yang, and T. Zhang, "Humanfi: Wifi-based human identification using recurrent neural network," in *2019 IEEE SmartWorld, Ubiquitous Intelligence & Computing, Advanced & Trusted Computing, Scalable Computing & Communications, Cloud & Big Data Computing, Internet of People and Smart City Innovation (SmartWorld/SCALCOM/UIC/ATC/CBDCom/IOP/SCI)*. IEEE, 2019, pp. 640–647.

[22] H. Chen, S. Munir, and S. Lin, "Rfcam: Uncertainty-aware fusion of camera and wi-fi for real-time human identification with mobile devices," *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, vol. 6, no. 2, pp. 1–29, 2022.

[23] J. Zhang, Z. Tang, M. Li, D. Fang, P. Nurmi, and Z. Wang, "Crosssense: Towards cross-site and large-scale wifi sensing," in *Proceedings of the 24th annual international conference on mobile computing and networking*, 2018, pp. 305–320.

[24] M. Dharmadasa, C. Gamage, and C. Keppitiyagama, "Radio tomographic imaging (rti) and privacy implications," in *2018 18th International Conference on Advances in ICT for Emerging Regions (ICTer)*. IEEE, 2018, pp. 413–419.

[25] Q. Dong and W. Dargie, "Evaluation of the reliability of rssi for indoor localization," in *2012 International Conference on Wireless Communications in Underground and Confined Areas*. IEEE, 2012, pp. 1–6.

[26] G. Li, E. Geng, Z. Ye, Y. Xu, J. Lin, and Y. Pang, "Indoor positioning algorithm based on the improved rssi distance model," *Sensors*, vol. 18, no. 9, p. 2820, 2018.

[27] H. Xu, M. Wu, P. Li, F. Zhu, and R. Wang, "An rfid indoor positioning algorithm based on support vector regression," *Sensors*, vol. 18, no. 5, p. 1504, 2018.

[28] S. J. Julier and J. K. Uhlmann, "New extension of the kalman filter to nonlinear systems," in *Signal processing, sensor fusion, and target recognition VI*, vol. 3068. Spie, 1997, pp. 182–193.

[29] F. Della Rosa, M. Pelosi, and J. Nurmi, "Human-induced effects on rss ranging measurements for cooperative positioning," *International Journal of Navigation and Observation*, vol. 2012, 2012.