Exploring the Impact of Locations and Activities in Person-wise Data Mismatch in CSI-based HAR

1st Nikita Sharma
University of Twente, The Netherlands
University of Oulu, Finland
0000-0002-2803-9834

2nd Duc V. Le
University of Twente, The Netherlands
0000-0001-7851-0869

3rd Paul J.M. Havinga
University of Twente, The Netherlands
TNO, The Netherlands
0000-0002-3399-1790

***Abstract*—Over the past decade, research has demonstrated the potential of Wi-Fi Channel State Information (CSI) in Human Activity Recognition (HAR). However, its real-world implementation is lacking due to the inability of CSI-based HAR to generalize across different domains (persons, locations, etc.). This inability of CSI can be attributed to the dynamic nature of CSI, leading to the issue of data mismatch. Therefore, in order to efficiently employ CSI-based HAR in real-world applications, a comprehensive understanding of the interplay between various data mismatch domains is essential. In that direction, the presented work aims to gain analytical insights into the impact of varying locations and activities in the person-wise data mismatch in realistic scenarios. To understand the person-wise data mismatch, three different analysis types namely subject-specific, mixed-subject, and generic-subject were defined. To assess the impact of locations in person-wise data mismatch, two activity locations and four receiver locations were considered. Whereas to assess the impact of the type of activities, four different activity sets, including full-body activities, fine-grained hand and leg activities, only fine-grained hand activities, and a mix of all activities, were evaluated. F1 score degradation by 43% for full-body activities, 72% for only fine-grained hand activities, and 76% for a mix of all activities in the person-wise domain indicate that person-wise data mismatch has a significant impact on the performance of CSI-based HAR. Furthermore, the impact of receiver location and activity location varied based on the activity set but was found comparatively insignificant when observed individually.**

***Index Terms*—Realistic scenarios, Channel state information, Human activity recognition**

I. Introduction

In the past decade, Human Activity Recognition (HAR) technology for healthcare and home automation has gained significant attention, especially with device-free sensing [1] [2]. It minimizes the disadvantages of other sensing solutions such as privacy risks, continuous wearing, and line-of-sight monitoring [3]. Among other existing device-free technologies for HAR, Wi-Fi Channel state information (CSI) is intriguing because it utilizes readily available Wi-Fi routers in households. CSI contains the propagation characteristics of transmitting signals including scattering, fading, and power decay along the way [3] which change dramatically due to the human body. Thus, human activities can be recognized by detecting the change in amplitude attenuation and phase shift of the radio signal.

The available research demonstrates the use of CSI for recognizing *physical activity* such as falling, walking, sitting, and hand gestures; *physiological activity* such as heart and breathing rate; and *behaviors* such as sleep patterns, eating-drinking habits, and physical fitness [3] [4] [5]. Though the achieved success of CSI in HAR is fascinating, its inability to generalize across domains due to data mismatch hinders its real-world implementation [6] [7]. Data mismatch occurs when there is a significant difference between the distributions of trained and testing data.

Recent works in CSI have tried to identify and propose solutions to overcome the data mismatch in various domains. For example, the work by Zinys et al. [6] proposes the use of GAN-based architecture to improve performance when different environments or persons data is used for training and testing. In another study, a novel approach using a wavelet template and stretch-limited dynamic time wrapping was proposed to overcome the impact of distance and orientation on the performance of CSI-based HAR [7]. Since the majority of available studies were conducted in labs or in controlled setups, a clear understanding of the domains causing data mismatch (individually or when combined) as expected in CSI-based HAR in real-world scenarios cannot be drawn.

In that direction, this work presents a step-by-step analytical understanding of the data mismatch due to persons when varying activity performance locations, receiver locations, and activities sets in realistic (close to real-world) scenarios on the performance of CSI-based HAR. To understand the impact of persons, three different analysis methods subject-specific (when train-validation-testing occurs subject-wise), mixed-subject (when train-validation-testing occurs on the same group of subjects), and generic-subject (when train-validation-testing occurs on different groups of the subject) were proposed. These analyses were utilized to evaluate the impact of two different activity performance locations and four receiver locations. Four different activity sets including, full-body activities, fine-grained hand and leg activities, fine-grained hand activities, and all activities together were defined to assess the impact of types of activities on these domains. Along with gaining an understanding of the interplay between

This research was funded by the European Union's Horizon 2020 research and innovation program under the Marie Skłodowska-Curie grant agreement number 814072 for the ITN ENTWINE informal care. The authors would like to thank Prof.Dr. J.E.W.C. van Gemert - Pijnen, Dr. L.M.A. Braakman - Jansen, and Dr. J.H. Croockewit for their valuable feedback.

multiple factors in person-wise data mismatch, these analyses can also help in underlining the potential and limitations of CSI technology for HAR in realistic settings.

Briefly, the contributions of these exploratory analyses are:

- **Explored person-wise data mismatch** by defining and evaluating three types of analysis namely, subject-specific (when train-validation-testing occurs subject-wise), mixed-subject group (when train-validation-testing occurs on the same group of subjects), and generic-subject group (when train-validation-testing occurs on different groups of the subjects) in a realistic scenario.
- **Highlighted the impact of different activity sets in person-wise data mismatch** with the help of four different activity sets namely full-body activities, fine-grained hand and leg activities, fine-grained hand activities, and all activities together. Analyses based on activity sets can help in understanding the impact of the size or combination of different activities on the classification performance of CSI-based HAR.
- **Highlighted the impact of different locations in person-wise data mismatch** by using two different activity locations and four different receiver locations (also beyond the wall). A rough estimation of the placement of receivers in houses can be made by knowing the impact of activity performance and the receiver's locations.

II. Related work

Mathematically, the CSI can be represented by a 3D matrix of complex values, consisting of amplitude attenuation and phase shift of channels at any given time. CSI signals, when captured over a time period, can predict the behavior of the signal in the time, frequency, and spatial domains with respect to the surrounding objects and persons. This predictive property makes CSI useful for wireless HAR [3]. However, due to multipath propagation, CSI signals are susceptible to dynamic indoor environmental cues. This means that along with human movements, a change in signal properties can be expected because of any other (slight) changes in the surrounding environment (such as moving chairs or having a visitor) [3] [6]. Therefore, if HAR signals are captured from different domains (different persons, locations, etc.), they will suffer from the data mismatch issue, which means that there will be a significant difference between the distribution of the training data and test data.

Recent research has highlighted and proposed solutions to overcome data mismatch for various domains influencing the performance of CSI-based HAR systems. These domains are quite diverse, ranging from the time of day to hardware, data quality, environmental cues, a person's physical characteristics, etc. In the study by Gao et al. [8] it was shown that the location of activity performance impacts the quality of data and to overcome that a novel error of dynamic phase metric to accurately quantify the sensing quality was proposed. Another study by Nguyen et al. [9] focused on hardware impairments and demonstrated that they can reduce not only the system's performance but also the self-interference cancellation capability of full-duplex devices.

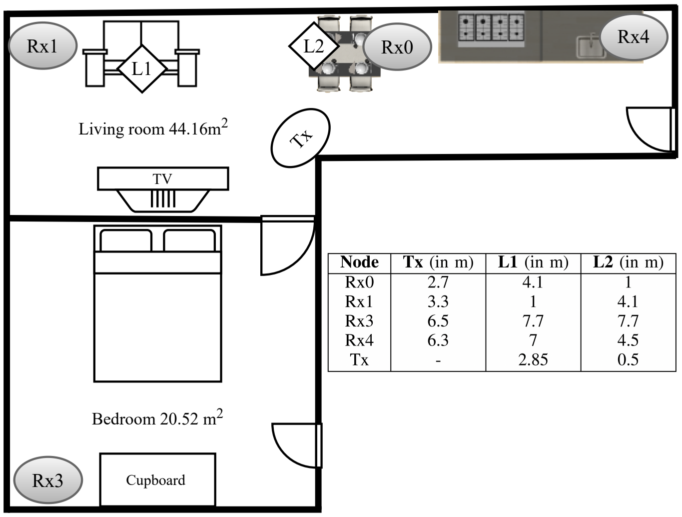
In a study by Brinke et al. [10], the impact of participants and even the time of the day on HAR performance were identified and evaluated by using the cross-participant validation method followed by transfer learning to solve the degraded performances. Zheng et al. [11] highlighted the performance variability in a gesture recognition system due to changes in orientation and location with respect to the sensing device. Zhang et al. [12] studied the impact of distance and orientation on HAR performance and proposed a novel approach that utilized a wavelet template and stretch-limited dynamic time wrapping to overcome these issues. In another similar study to identify sign language gestures, the impact of different physical properties of participants or a slight difference in movement patterns on the performance has been underlined [13]. In a recent study, the problem of domain change due to environment and person was identified, and to mitigate that a domain-independent generative adversarial network was proposed and evaluated for gesture recognition [7].

III. Gap in research

While the use of advanced algorithms or analysis to address data mismatch in CSI is motivating, their effectiveness in real-world scenarios remains uncertain. First, the impact of the above-discussed domains was identified in controlled setups either in lab environments or by tightly controlling the experiment protocol (such as defined participant's position or orientation, etc.). Thus, there is a need to cross-examine the identified domains in realistic scenarios. Second, in the real-world, multiple factors can impact the data mismatch in CSI at once, requiring a thorough analysis of the interplay between multiple factors impacting CSI. Third, the performance of CSI-based HAR across studies varies due to differences in study setups, environmental conditions, activity types, and analysis approaches in the literature. These analyses can help in underlining the reasons behind varied performances. Lastly, understanding the behavior of raw CSI data is crucial before proceeding with advanced signal or data analytical approaches as it might provide new insights into the behavior of CSI or might open up new challenges. Considering these gaps, the work aims to conduct step-by-step exploratory analyses to gain an analytical understanding of the person-wise data mismatch while considering different locations (activity performance and receiver) and activity sets (fine-grained and full-body) in a realistic setup.

IV. Dataset used and Analysis protocol

For this work, a realistic dataset (named Wi-Gitation) collected in a fully-furnished one-bedroom apartment (Figure 1) by using a semi-protocolized experiment paradigm was used. Here, participants were given certain degrees of freedom, such as participants were not strictly instructed on the exact place of sitting (they were asked to sit on the sofa but not at the same spot); they were allowed to choose the orientation of sitting (which side they want to face); they were asked to choose their



Node	**Tx** (in m)	**L1** (in m)	**L2** (in m)
Rx0	2.7	4.1	1
Rx1	3.3	1	4.1
Rx3	6.5	7.7	7.7
Rx4	6.3	7	4.5
Tx	-	2.85	0.5

Fig. 1. Layout of eHealth house [14].

own the path of walking across kitchen and living room; they were free to improvise the way of performing the fine-grained activity (speed or use of any or both hands or legs). The other important characteristics of the dataset are explained below in accordance with its organization for exploring the impacts of different data mismatch domains:

To explore person-wise data mismatch: In general, the impact of data mismatch due to persons on classification performance is very evident in CSI-based HAR. The used dataset is collected from twenty-three healthy participants (age: 25.26 $\pm$ 9.49) in a simulated one-bedroom apartment. Thus, it helps in cross-examining the impact of data mismatched caused by different persons in real-world scenarios. Therefore, here we define three cases showing different splits in training-validation-testing data.

- **Subject-specific data analysis** (train-validate-test on the same or another person): It means that individual participants are taken for training, validation, and testing. Within the subject-specific analysis, two possibilities can be observed. First, when the algorithm was trained and tested on the same subject to observe the behavior of the system when it has an opportunity to train on the test subject (Case 1.1). Contextually, it represents the case when monitoring needs to be done on only one person living alone. Second, when the algorithm was trained on one individual and tested on other individuals (Case 1.2). This type of analysis is used when training data is limited but testing will be done on wider audiences.
- **Mixed-subject data analysis** (train-validate-test on mixed data): It means data from all participants are mixed and then sliced into training (68%), validation (17%), and testing (15%) (Case 2). Here, all three splits can have data from the same person but different samples. It can be useful for scenarios when the system is expected to predict only the seen participants.
- **Generic-subject data analysis** (train-validate-test on different groups of participant data): It means data from different groups of persons are used for training (12 participants), validation (4 participants), and testing (4 participants) (Case 3). This is the most common scenario as more data is used for training and testing is done on unseen individuals.

Table I presents the overview of these cases with the possible scenarios where these cases can be used in a future real-world implementation. These cases were applied for all types of activities, both locations, and all receivers so that a fair idea of the interplay between different factors can be drawn.

To highlight the impact of different activity sets in person-wise data mismatch: From the used dataset, two full-body (normal walking and disturbed walking), one fine-grained leg activity (kicking on the leg of a table), and four fine-grained hand activities (rubbing hands on a table, hand wringing, tapping on a table, flipping objects) were utilized to define four different activity sets by considering the kind of activities. Here each activity was conducted for two minutes. The rationale was to see the difference in classification performances for similar (for example, only hand activities) or mixed (full-body and fine-grained hand/leg) activity types when performed by different participants, at different locations, and measured from different receivers i.e. different data mismatch domains. These activity sets are illustrated in Table II.

- **Full-body activities (A1 set)**: It includes two full-body activities: normal and disturbed walking. In disturbed walking activity (usually identified in older adults) participants used a cane, showed different paces of walking (slow/fast), and were also limping whereas participants were asked to walk normally as they do for normal walking activity. This could also be helpful when disturbed walking needs to be recognized from normal walking in older adult care. Note these activities are not location-dependent as they were performed across the kitchen and living room areas.
- **Fine-grained hand and leg activities (A2 set)**: A combination of fine-grained hand activity (Tapping, wringing hands, rubbing tables, flipping objects) and leg activity (Kicking) were used in this set. This activity set can provide insights when the aim is to classify fine-grained but different activities into broader categories without going into more specific classifications.
- **Fine-grained hand activities (A3 set)**: Fine-grained hand activities (Tapping, wringing hands, rubbing tables, flipping objects) were used to see the individual classification performance within a group of similar fine-grained activities, adding evidence for exploring the limits of CSI.
- **All activities together (A4 set)**: This set utilizes a mix of full-body, fine-grained hand, and leg activities. Usually, for a context-based HAR system, it is common to have different types of activities and this set can give insight into the behavior of CSI in such cases.

To highlight the impact of different locations in person-wise data mismatch: A system can be ubiquitous if it can

TABLE I

TYPE OF ANALYSIS, CORRESPONDING CASES, AND POSSIBLE SCENARIOS ENCOUNTERED IN REAL-WORLD IMPLEMENTATION

Analysis type	Cases	Possible scenarios
Subject-specific	**Case 1.1:**Training and testing on the same participant	The model has an opportunity to train on the person to be tested, for example when an individual is living alone
Subject-specific	**Case 1.2:** Training on one and testing on the different participants	The model is trained on a different person than the test person. This is usually the case when less data is available for training.
Mixed data	**Case 2:** Mixed data split into 68% Train, 17% validation, 15% test	The model has the opportunity to learn from a group of persons that will be used in the future for testing.
Generic data	**Case 3:** Using different 12P for Train, 4P for validation, 4P for test	The model is trained on a group of persons with testing on an unseen group of users. For example, a product is developed with the aim to deploy in different households with unknown users.

TABLE II

FOUR DEFINED ACTIVITY SETS: A1, A2, A3, AND A4

Activity sets	Activities includes
A1	Full body activities: Normal walking (NW), Disturbed walking (DW)
A2	Fine-grained hand and leg activities: Hand activity (Tapping (TA), Hand wringing (HW), Rubbing table (RT), Flipping objects (FO)) and leg activity (Kicking on the table (KA))
A3	Fine-grained hand activities: All hand activities (Tapping (TA), Hand wringing (HW), Rubbing table (RT), Flipping objects (FO))
A4	All activities together: Mix of full-body, fine-grained hand, and leg activities (Disturbed walking (DW), Sitting down standing up (SS), Kicking (KA), Tapping (TA), Hand wringing (HA), Rubbing tables (RT), Flipping objects (FO))

monitor activities at all possible locations in a given area or house. For example, if a system can recognize kicking activity at location L1 then ideally it should also recognize when the same activity is happening at location L2. But due to the vulnerable nature of CSI with changes in activity location, a variation in classification performances can be observed [12]. But its impact in realistic scenarios within the same house is unknown. Thus, to demonstrate this, the work utilizes the CSI data of different activities at Locations L1 (on a sofa) and L2 (on a dining chair) from all four receivers (see Figure 1).

It is also important to understand the impact Rx location can have on HAR when considering the wider implementation. Particularly, due to the varied distance between the Tx-Rx and when the receiver is placed beyond the wall. Upon gaining this understanding, an idea for strategies concerning the placement of Tx-Rx as per the size/shape of a house can be determined. Therefore, four different Rx placed at different distances from a Tx 2.7m (Rx0), 3.3m (Rx0), 6.7m (Rx3, beyond the wall), and 6.3m (Rx4) respectively in a simulated real-world one-bedroom apartment, when the persons were performing the activity at L2 were utilized (see Figure 1). One location is chosen to zoom in on the impact of Rx locations in person-wise data mismatch.

V. DATA ANALYSIS

After exploring multiple networks, ResNet-50 was chosen to conduct these analyses. ResNet-50 is a 50-layer deep convolutional neural network widely used for image classification. The untrained network architecture of ResNet-50 was obtained from Matlab and adapted as per the required input and output size. In the Wi-Gitation dataset, a three-dimensional channel state matrix, $\mathbf{H} = 3(Ntx) \times 3(Nrx) \times 30(Nsc)$, where Ntx is the number of Tx antennas, Nrx is the number of Rx antennas, and Nsc the number of sub-carriers was obtained. This means 270 channels for each sample were present in the raw data. Moreover, each activity was conducted for two minutes thus a total of 12000 samples (120 sec $\times$ 100 Hz) were expected (without packet loss). Thus, a matrix of dimension 270 channels $\times$ 12000 samples (only CSI amplitude) were obtained for each activity and each participant. Since the aim of this paper is to showcase exploratory analysis, minimum pre-processing steps including nearest neighbor interpolation to make samples consistent [15] and sliding window with 50% overlap to overcome the information loss, if any, during segmentation in further processing [16] were applied.

VI. RESULTS

In this section, the obtained results are presented to show the person-wise data mismatch when activities, activity performance, and receiver locations are varied. Here, $F1-scores$ (harmonic mean of precision and recall) being a better evaluation matrix (than accuracy) for trained models specifically when the dataset is imbalanced is used to report these results.

A. Person-wise Data Mismatch

Figures 2 and 3 show that subject-specific data analysis when trained and tested on the same person (case 1.1) and mixed-subject data analysis (case 2) gave higher and comparable performances for all the activity sets. It can be because in case 1.1 algorithm is trained and tested on the same person and thus might be highly overfitted to the person's physical traits. Similarly, in case 2, the model had an opportunity to see the samples from the same person thus it again gets trained well on the given persons. Note that, there was no data mismatch in these two cases, thus yielding better performances. Furthermore, the degraded performance due to data mismatch can be observed in generic-subject (case 3) and subject-specific data analysis when trained and tested on a different person (case 1.2). Case 3 performed slightly better than case 1.2 as it had more data for training thus maybe it is able to generalize better in comparison to case 1.2 where only one person's data is given for training which might lead to overfitting on the person it was trained on. The $F1-scores$ presented in figure 2 and 3 are averaged over both locations and all receivers.

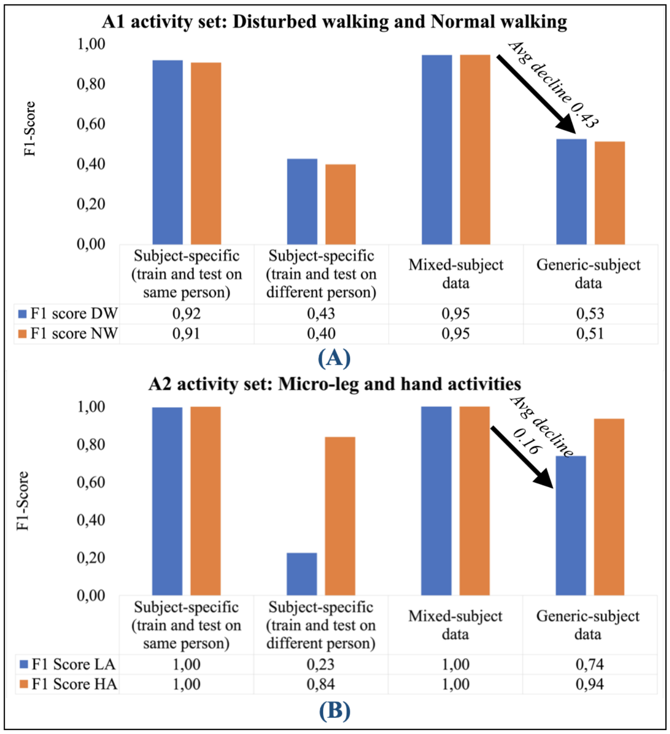


Fig. 2. $F1-scores$ for person-wise data mismatch in (A) A1 Activity set: Disturbed walking (DW) & Normal walking (NW); (B) A2 Activity set: Leg activity (LA) & Hand activities (HA) corresponding to subject-specific, mixed-subject, and generic-subject analysis.

B. Impact of Activities in person-wise Data Mismatch

To demonstrate the impact of activities in person-wise data mismatch, we zoom in on the generic-subject analysis (case 3) and compare it with mixed-subject data analysis (case 2) as it is a more prevalent type of analysis. In generic-subject analysis, higher $F1-scores$ were observed for fine-grained hand and leg activities when classified broadly (Figure 2B), followed by full-body activities (Figure 2A), only fine-grained hand activities (Figure 3A), and all activities together (Figure 3B). This could be because fine-grained hand and leg activities might have distinct variations in CSI signal as they were performed by using different limbs (either hands or legs). Here, in comparison to mixed-subject data analysis,$F1-scores$ dropped by 0.16 (on average) (Figure 2B). It is also worth noticing that the number of samples for hand activities was higher in training as it comprised of four activities namely tapping, hand wringing, rubbing hands, and flipping objects whereas, in the leg activity category, only one activity namely kicking was there. This might create a classification bias. Whereas used full-body activities normal and disturbed walking are quite close to each other as they require similar limb movement, thus might have more similarity in CSI signal. However, when compared to mixed-subject data analysis, $F1-scores$ dropped by 0.43 (on average) (Figure 2A). Despite that, the obtained $F1-scores$ shows the possibility that these two activities can be classified.

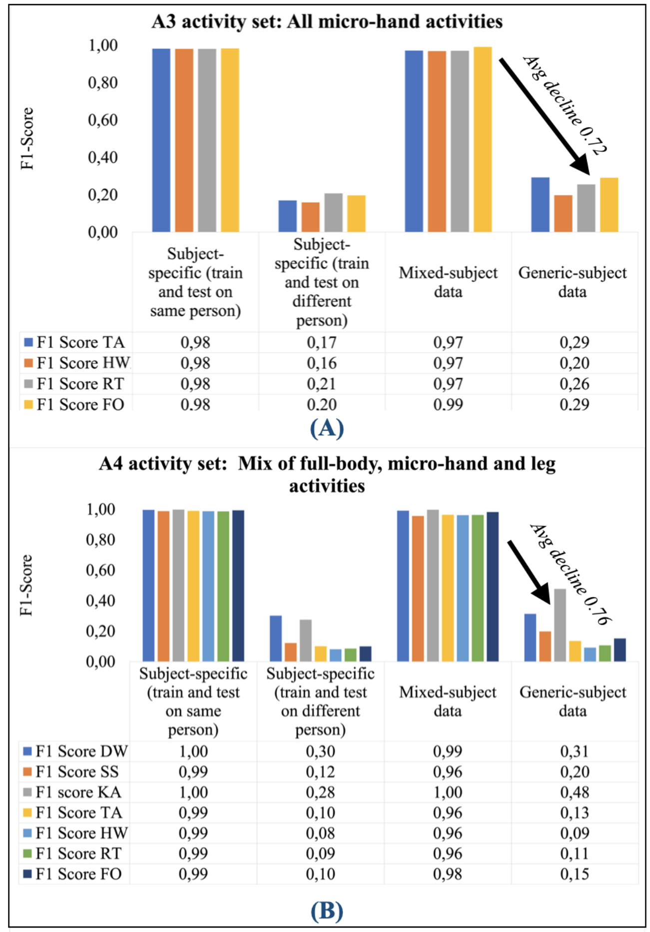


Fig. 3. $F1-scores$ for person-wise data mismatch in (A) A3 Activity set; (B) A4 Activity set for subject-specific, mixed-subject, and generic-subject analysis. (Abb: TA (Tapping), Hand Wringing (HW), RT (Rubbing Table), FO (Flipping Objects), Kicking (KA), SS (Sitting Standing), DW (Disturbed Walking))

Moreover, in generic-subject analysis for all the fine-grained hand activities (Tapping, Hand wringing, Rubbing hands, Flipping objects) $F1-scores$ were degraded by 0.72 (when compared to mixed-data analysis) but they were comparable between the activities of this set (Figure 3 A). This could be because when the number of classes using similar limb movements increases, the model becomes incapable to classify closely related classes. Lastly, when all the activities i.e. a mix of full body, fine-grained hand, and leg activities are classified together, a significant drop in $F1-scores$ (by 0.76) in comparison to mixed-subject data analysis was observed (Figure 3 B). Though a comparable $F1-scores$ can be observed for full-body activities and fine-grained leg activity. Overall, it can be understood that the chances of correctly classifying full-body activities are more compared to fine-grained hand activities in person-wise data mismatch.

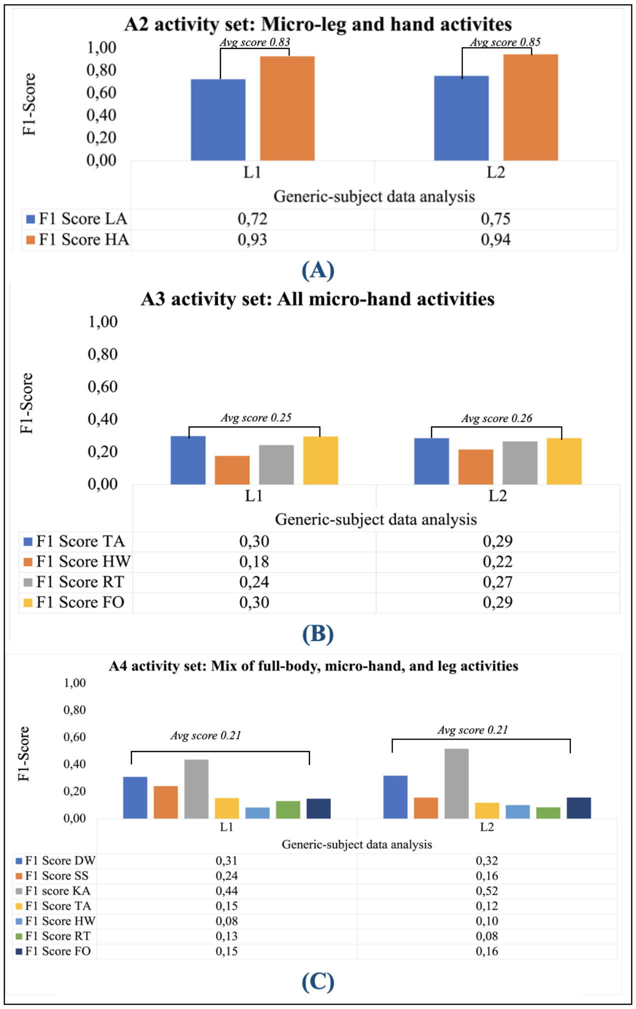


Fig. 4. $F1-scores$ for (A) A2 activity set; (B) A3 Activity set; (C) A4 Activity set at L1 & L2 location for genric-subject analysis. (Abb: LA (Leg Activity), HA (Hand Activities) TA (Tapping), Hand Wringing (HW), RT (Rubbing Table), FO (Flipping Objects), Kicking (KA), SS (Sitting Standing), DW (Disturbed Walking))

C. *Impact of Locations in person-wise Data Mismatch*

All four analysis cases and activity sets (except A1 set in the activity location as it is location independent) were used. In subject-specific data analysis when trained and tested on the same person (case 1.1) and mixed-subject data analysis (case 2) $F1-scores$ were significantly high (between 0.97 to 1), hence it does not deliver insights on the impact of locations or placement of receivers. Among subject-specific data analysis when trained and tested on a different person (case 1.2) and generic-subject data analysis (case 3), the latter is more commonly used. Thus, for both activity performance and receiver locations the results from only generic-subject analysis are presented and compared.

Within the generic-subject analysis, both locations gave comparable performances for individual activity sets (Figure 4). This indicates that when training and testing are done by using the same location's data similar performances can be expected despite varied distances from transmitter-receiver pairs. For the A2 activity set, Figure 4 A, the average $F1-scores$ for L1 and L2 were 0.83 and 0.85 respectively. In A3 activity set Figure 4 B, the average $F1-scores$ dropped to 0.25 and 0.26 for L1 and L2 respectively. Lastly, for A4 activity set Figure 4 C, the average $F1-scores$ further dropped to 0.21 for both locations. Note that, L2 was closer to the transmitter but not much differences in individual performances were observed. The $F1-scores$ presented in Figure 4 are averaged along the receivers to specifically demonstrate the impact of activity performance location.

Furthermore, in receiver location-wise data mismatch, different placements of receivers do impact the activity classification but not very significantly ($F1-score$ were 0.45, 0.48, 0.44, and 0.46 for Rx0, Rx1, Rx3, and Rx4 respectively, averaged over activity sets). Though significant differences with respect to different activity sets were observed. In the full-body activity set, participants were walking all around the house thus all the receivers gave varied performances (Figure 5A). Disturbed walking was best monitored by Rx4 ($F1-score$: 0.63) whereas normal walking was by Rx0 ($F1-score$: 0.59). This might be because for normal walking activity participants were walking more in the closer proximity of Rx0 and Rx4 for disturbed walking activity. Interestingly, in the fine-grained hand and leg activities set, the closest receiver (Rx0) gave the comparatively worst performance with $F1-score$ of 0.80 (Rx0 is the closest receiver as data from L2 is used here) whereas other receivers (including one placed beyond the wall) gave comparatively better results with $F1-scores$ 0.87, 0.87, and 0.85 for Rx1, Rx3, and Rx4 respectively (Figure 5B). It could be because this placement location of the receiver was not able to obtain good reflected signals from leg activities. Nevertheless, the classification performance for only fine-grained hand activities degraded with an average $F1-scores$ between 0.25 and 0.28 with respect to different receivers. This could be because these movements were of very small scale (Figure 5C). Lastly, similar degraded performances were observed when classifying all the activities together (Figure 5C). Interestingly, the performance of Rx3 (beyond the wall) was also at par with other receivers.

VII. Discussion & Challenges

A perspective on the interplay between different factors impacting CSI in person-wise data mismatch in realistic scenarios can be drawn on the basis of conducted exploratory analyses.

In a nutshell, the person-wise data mismatch (due to varied physical traits) is evident and dominant in realistic scenarios. For example, no significant differences were obtained between subject-specific analysis when trained and tested on the same person and mixed-subject analysis with respect to all activity sets, both locations and all receivers (Figure 2 and Figure 3). In these two cases model was trained on the samples of the person(s) to be tested i.e. no or less data mismatch, giving the model an opportunity to learn the (specific) way a person is performing an activity in combination with his/her physical traits. Whereas the performance degrades drastically in cases when a model is tested on the unseen person(s)

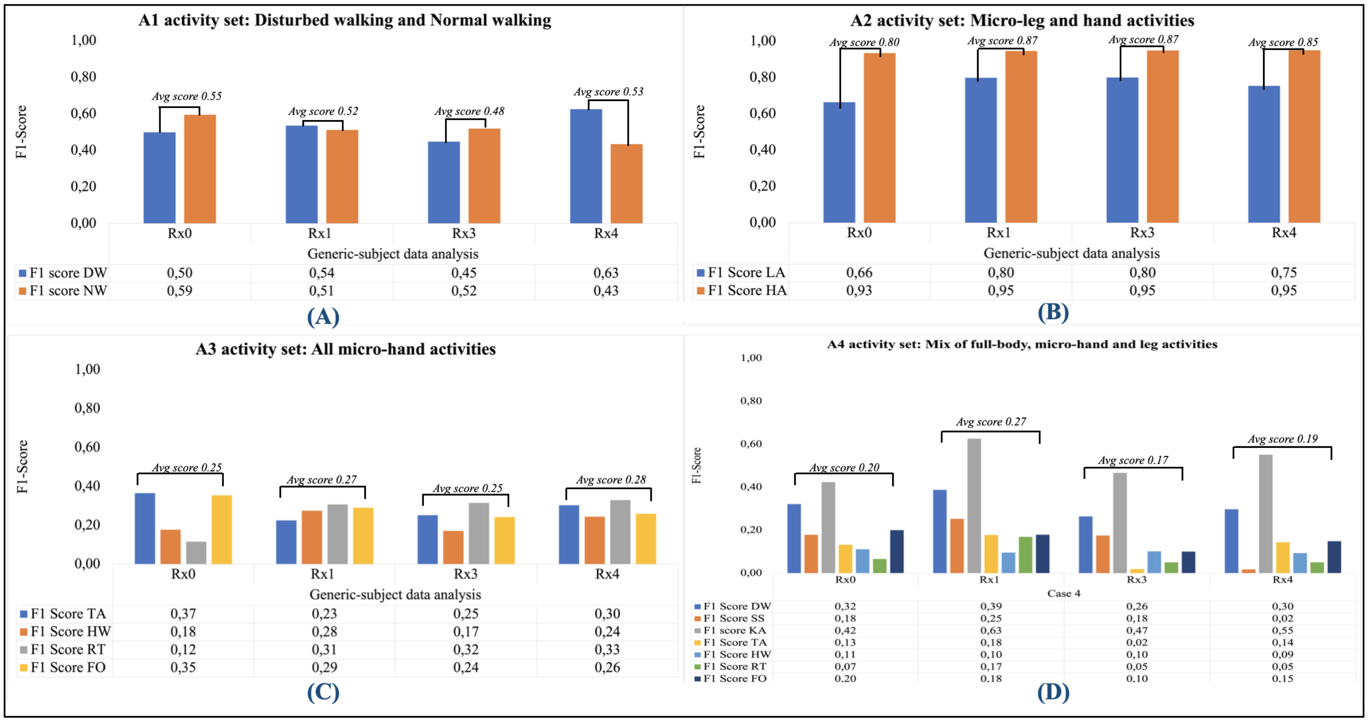


Fig. 5. $F1-scores$ for (A) A1 activity set; (B) A2 activity set; (C) A3 activity set and; (D) A4 activity set for different receiver locations corresponding to generic-subject analysis. (Abb: NW (Normal Walking), DW (Disturbed Walking), LA (Leg Activity), HA (Hand Activities), TA (Tapping), Hand Wringing (HW), RT (Rubbing Table), FO (Flipping Objects), Kicking (KA), SS (Sitting Standing)).

(subject-specific when trained and tested on different persons and generic-subject analysis). Every person impacts the CSI signals differently which creates a variation in training and testing data. The ML algorithm tends to learn the patterns in training data, expecting similar patterns in testing data. Thus for unseen persons in testing data, it is highly likely that testing performance will decline. This is a classic case of data mismatch in CSI-based HAR.

From this, it could be interpreted that the raw CSI signals pertaining to activities show less variation as compared to the combined variation due to physical traits such as height, weight; the way of performing the activity (some participants used their right hand some used their left hand for tapping on the surface); the pace of performing the activity; or way of sitting, etc. This makes it very difficult for the model to identify these small movements in raw data and the model gets trained on the overpowering variation due to the other person-dependent factors. Thus, it can be implied that the trained models are over-fitted to the physical traits of the person(s) in the training data hence yielding very poor testing performance with an unseen person(s) like in the generic-subject analysis. Nevertheless, the same person-specific factors will be present when CSI-based HAR will be deployed in real-world scenarios.

Another, interesting aspect outlined from these exploratory analyses is the varied impact of the type of activities used in data mismatch domains. Full-body activities like disturbed walking and normal walking showed almost similar $F1-scores$ when trained and tested on the different person(s). In the leg and hand activity set, $F1-scores$ in generic-subject group analysis were better and interesting to note. It supports the argument that varied data leads to generalization across unseen persons thus yielding better and more stable classification results. Overall, the set of hand activities always performed better than the kicking activity. While this could be because of classification bias due to higher training samples in hand activities, it does indicate that CSI has the potential to classify wider activity categories.

A level deeper when aiming to classify individual activities within the group of hand activities, the classification performances dropped drastically. It is because these activities were very similar and small in nature thus making it more difficult for a model to identify them among the similar set of activities in the raw CSI signals which are already varied due to the physical traits of the person(s). In addition, the increased number of classification categories in the only hand and all activities together set might have impacted the performance negatively. To prevent the degradation due to the type of activities, a step-by-step activity classification method can be used, where at first different activities are classified into broader categories followed by classifying individual activities.

Moreover, the activity locations do play a role in CSI-based HAR but not very significant differences were obtained at individual locations (i.e. training-testing on one location's data), for example, in generic-subject analysis, the performance of both locations was found comparable within each activity set (Figure 4). CSI data obtained from one particular location is a combination of reflections of multiple objects in the way along with the person performing activities. Upon changing the location multiple objects might impact in a different order (in terms of reflection), but since the same objects were in the house the obtained final signal might contain similar components but in different order hence giving comparable performances. The study by Wang et al. also indicates the limited impact of the activity performance locations [12].

Interestingly, the impact of the different receivers was comparatively more evident among different activity sets (Figure 5), with the receiver beyond the wall giving comparable performances. It might be because some receivers might not be able to get activity information in the CSI signal because of the angle or way some activities are performed. Overall, among individual activity sets the performances of different receivers do not degrade significantly despite a reasonable distance between Tx and Rx. This could possibly indicate the ubiquitous nature of CSI. While these are speculations, more rigorous testing is needed to conclude.

Altogether, these analyses demonstrate the increased person-wise data mismatch challenges when realistic setups are used as there are multiple factors impacting the CSI at once. For example, if an algorithm is designed with an aim to tackle person-wise data mismatch, varied performances with variations in activity sets or receiver locations can be expected. Therefore, to ensure the robustness of a novel algorithm developed for person-wise data mismatch, it should be tested on realistic datasets where multiple factors are impacting CSI at once. This also explains the varied performances of different datasets even though the same algorithms are used, each dataset has its own degree of variations in participants, activities, locations, the hardware used, etc. However, CSI also presents opportunities for HAR, such that if activities are performed at different locations or observed by different Tx-Rx pairs in a house, there are chances that almost similar performances might be obtained. In the current state, CSI-based sensing systems can be found comparatively reliable in cases when training and testing are done on the same person(s) and broader activity sets.

VIII. Conclusion & Future works

This analysis highlights the challenges associated with implementing CSI-based HAR in real-world scenarios. It emphasizes the much greater negative impact of person-wise data mismatch on its performance in comparison to testing in controlled environments. The type of activities used also significantly affects the classification performance of CSI-based HAR, with full-body or broader sets of activities performing better than fine-grained activities. However, developing different algorithms for different activity sets or overcoming the impact of different domains is impractical, necessitating the need for more robust and inclusive algorithms. Furthermore, this study reveals that the limited impact of locations within the same house when observed individually presents opportunities for further development of CSI-based HAR.

To advance the field, future work should focus on identifying domain-independent features by removing the physical traits of individuals from the CSI data and testing them in environments that closely resemble the real-world. Additionally, exploring the impact of persons in data mismatch due to locations and the impact of dynamically changing environments would be valuable for realizing the real-world implementation of CSI-based HAR.

References

[1] B. Tan *et al.*, "Exploiting WiFi Channel State Information for Residential Healthcare Informatics," *IEEE Communications Magazine*, vol. 56, no. 5, pp. 130–137, 2018.

[2] N. Sharma *et al.*, "Implementation of Unobtrusive Sensing Systems for Older Adult Care: Scoping Review," *JMIR Aging*, vol. 4, no. 4, p. e27862, Oct 2021.

[3] Y. Ma *et al.*, "Wifi Sensing with Channel State Information: A Survey," *ACM Computing Survey*, vol. 52, no. 3, jun 2019. [Online]. Available: https://doi.org/10.1145/3310194

[4] Z. Wang *et al.*, "A Survey on Human Behavior Recognition Using Channel State Information," *IEEE Access*, vol. 7, pp. 155 986–156 024, 2019.

[5] Z. Wang *et al.*, "A Survey on CSI-Based Human Behavior Recognition in Through-the-Wall Scenario," *IEEE Access*, vol. 7, pp. 78 772–78 793, 2019.

[6] R. Gao *et al.*, "Towards Position-Independent Sensing for Gesture Recognition with WiFi," *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.*, vol. 5, no. 2, jun 2021. [Online]. Available: https://doi.org/10.1145/3463504

[7] A. Zinys *et al.*, "A Domain-Independent Generative Adversarial Network for Activity Recognition Using WiFi CSI Data," *Sensors*, vol. 21, no. 23, 2021. [Online]. Available: https://www.mdpi.com/1424-8220/21/23/7852

[8] R. Gao *et al.*, "Towards Robust Gesture Recognition by Characterizing the Sensing Quality of WiFi Signals," *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.*, vol. 6, no. 1, mar 2022. [Online]. Available: https://doi.org/10.1145/3517241

[9] B. C. Nguyen *et al.*, "Impact of Hardware Impairments on the Outage Probability and Ergodic Capacity of One-Way and Two-Way Full-Duplex Relaying Systems," *IEEE Transactions on Vehicular Technology*, vol. 69, no. 8, pp. 8555–8567, 2020.

[10] J. K. Brinke *et al.*, "Scaling Activity Recognition Using Channel State Information Through Convolutional Neural Networks and Transfer Learning," ser. AIChallengeIoT'19. New York, NY, USA: Association for Computing Machinery, 2019, p. 56–62. [Online]. Available: https://doi.org/10.1145/3363347.3363362

[11] Y. Zhang *et al.*, "Widar3.0: Zero-Effort Cross-Domain Gesture Recognition with Wi-Fi," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 44, no. 11, pp. 8671–8688, 2022.

[12] W. Wang *et al.*, "Understanding and Modeling of WiFi Signal Based Human Activity Recognition," in *Proceedings of the 21st Annual International Conference on Mobile Computing and Networking*, ser. MobiCom '15, 2015, p. 65–76. [Online]. Available: https://doi.org/10.1145/2789168.2790093

[13] Y. Ma *et al.*, "SignFi: Sign Language Recognition Using WiFi," *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.*, vol. 2, no. 1, mar 2018. [Online]. Available: https://doi.org/10.1145/3191755

[14] U. o. T. Tech Med Centre. Living lab ehealth house. [Online]. Available: https://www.utwente.nl/en/techmed/facilities/htwb-labs/ehealth-house/l

[15] B. V. Dasarathy, "Nearest neighbor (nn) norms: Nn pattern classification techniques," 1991.

[16] J. Li *et al.*, "SlideAugment: A Simple Data Processing Method to Enhance Human Activity Recognition Accuracy Based on WiFi," *Sensors*, vol. 21, no. 6, 2021. [Online]. Available: https://www.mdpi.com/1424-8220/21/6/2181