

WiSe4Car: A Dataset for WiFi Sensing in Automotive Vehicles

Daniel Markert, Estelle Wünsche, Hussein Kerdi, Erik Leo Hass, Ian Blake, Yannic Warias,
Ana Perez Grassi, Alejandro Masrur
Department of Computer Science, TU Chemnitz, Germany

Abstract—WiFi sensing, i.e., using WiFi signals to scan the environment, is a promising technology with a plurality of potential applications in the automotive domain. In addition to being cost-effective, since WiFi sensing does not rely on images, it addresses privacy concerns and/or regulations in vehicle interiors. However, there is a notorious lack of publicly available datasets necessary to train the underlying machine learning (ML) models and further develop the technology in this domain. In this paper, we propose one such dataset consisting of CSI (Channel State Information) recordings in more than 100 different scenarios involving multiple individuals and different vehicles. We have not only labeled CSI data by participants' height, weight, gender, and seat position among others, but also include anonymized video recordings for further labeling in the future (say, for example, activity, etc.). In addition, CSI data was collected every time from multiple spots within the vehicle allowing for later studies concerning sensor positioning in an application of interest. The proposed dataset is made publicly available and intended to help establishing WiFi sensing to be used within automotive vehicles, which we demonstrate by an exemplary case study.

Keywords: WiFi-Sensing, Channel State Information (CSI), Dataset

I. INTRODUCTION

WiFi has become pervasive nowadays and, hence, it can be encountered almost everywhere. Besides serving for communication between mobile devices (such as smartphones, tablets, etc.), WiFi signals can also be used to sense the environment, which is known as *WiFi sensing*.

In particular, WiFi signals are reflected and absorbed by the human body, which affects the CSI (Channel State Information) data in them. This can be utilized by machine learning (ML) approaches to draw conclusions from the environment such as estimating the number of persons in a room [1], recognizing human activity [2] [3], identifying specific individuals [4], etc.—see Fig. 1.

In general, since WiFi sensing is a non-imaging technology, it is unsensitive to lighting conditions and occlusion, further preserving people's right to privacy. These characteristics make WiFi sensing ideal for numerous application scenarios. In this work, we are particularly interested in its use within automotive vehicles. Our goal in the future is to personalize safety and comfort features such as, e.g., airbags, seat belts, headrests, etc. based on WiFi sensing.

There have already been some works in this area, e.g., to detect fatigue levels [5] and head position [6]. However, to the best of our knowledge, there is still no publicly available CSI dataset for the automotive domain yet, which hinders further research and development in said context.

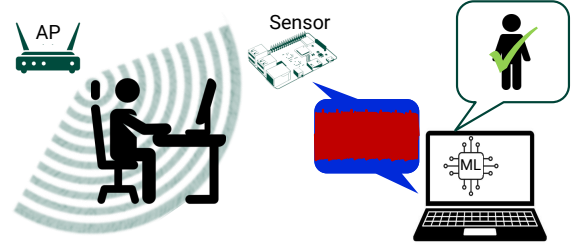


Fig. 1: WiFi Sensing: ML models are trained to infer information out of changes in the CSI data from the environment.

To overcome this limitation, we present one such CSI dataset. In particular, we recorded CSI data from automotive vehicles on different days over a period of multiple months. Each time, we annotated data by seat position, height, weight, gender, etc. of persons. Our contributions can be summarized as follows.

Contributions: The work in this paper revolves around collecting CSI data from vehicle cabins and in different scenarios, with a varying number of passengers sitting in different positions. Our overall contributions are stated next:

- We present, to the best of our knowledge, the first publicly available CSI dataset collected in standard vehicle cabins and accessible under <https://mobilithek.info/offers/872135089723445248>. As stated above, CSI data was collected on different days and vehicles, and labeled by seat position, height, weight, gender, etc. of currently 48 people participating in the experiments.
- A full description of our experimental setup is presented. This basically consists of an easy-to-mount, portable demonstrator composed of multiple *CSI sensors*, i.e., adapted WiFi receivers, a standard WiFi router/AP and a temperature/humidity sensor.
- We recorded anonymized videos during every measurement session, which we also added to our dataset. These videos contain useful information for further labeling, e.g., for activity or posture recognition.
- Finally, we discuss a case study, where we train an ML model based on the proposed CSI dataset to identify the driver's weight and height respectively. We believe our results demonstrate the usefulness and potential of the proposed dataset.

The rest of the paper is organized as follows. Section II briefly introduces WiFi sensing and associated concepts, whereas Section III gives an overview of the related work.

Section IV and Section V discuss data collection and the resulting dataset respectively. Finally, we present our case study in Section VI and conclude the paper with Section VII.

II. FUNDAMENTALS

In general, WiFi sensing consists in extracting the CSI data from WiFi signals. Even though any WiFi packet/frame contains CSI data, typically, beacons are used, since these are sent periodically by the router or access point (AP) without request. As mentioned before, variations in the CSI data can be used by an ML approach to infer information from the environment. Current WiFi standards such as the IEEE 802.11ac protocol divide the communication channel into individual *subcarriers* via orthogonal frequency division multiplexing (OFDM). As subcarriers are equidistantly spaced (every 312.5 kHz), their total number depends on the bandwidth, ranging from 64 subcarriers for a single channel of 20 MHz to 512 for 8 such channels (amounting to 160 MHz). Some of these subcarriers are guard and pilot subcarriers though and contain no CSI data [7].

While our CSI sensors (i.e., modified WiFi receivers) measure both primary and secondary channels—see Fig. 2, both of 20 MHz, the router/AP sends its beacons on the primary channel only. Secondary channels are only used as backup, if communication fails on the primary channel. This leads to 64 subcarriers that are recorded, 16 of which are reserved and hence carry no relevant CSI data.¹

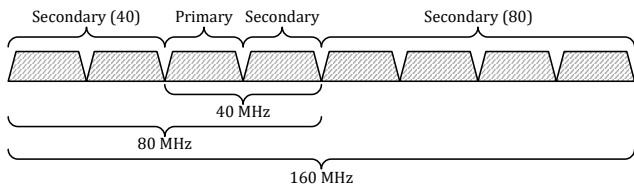


Fig. 2: WiFi channels according to the 802.11ac standard [8].

Changes in the CSI data can then be used to train ML models for different applications so as to draw conclusions from the environment and objects/persons in it—see again Fig. 1.

In simple terms, for a single beacon received from the WiFi router/AP, CSI data is obtained in the form of a vector (of complex values) denoted by $\mathbf{h}$ as follows:

$$\mathbf{y} = \mathbf{h} \circ \mathbf{x} + \mathbf{n},$$

where $\mathbf{x}$ and $\mathbf{y}$ are also vectors representing the sent and received signal respectively, with an element for each of the subcarriers considered. Further, $\mathbf{n}$ is a vector representing noise on the communication channel, and we denote the pairwise multiplication of two vectors (element by corresponding element) by $\circ$.

The WiFi receiver/CSI sensor basically estimates $\mathbf{h}$ from $\mathbf{x}$ and $\mathbf{y}$ considering an expected $\mathbf{n}$. From the complex

values/elements in $\mathbf{h}$ a CSI amplitude and phase can be obtained and collected.

Over time, when multiple beacons are received, $\mathbf{h}$ becomes a matrix $\mathbf{H}$, where each of the columns results from processing each of the different beacons.

III. RELATED WORK

Since ML models need to be trained, CSI datasets have become of paramount importance for the WiFi sensing community. These datasets are, however, very much dependent on the application of interest and, hence, they cannot be easily reused in other contexts.

So far, there exist already a number of publicly available CSI datasets in different domains and settings, however, to the best of our knowledge, none covering applications within vehicle cabins.

A dataset called eHealth CSI was presented in [9]. It includes more than 100 participants performing various actions inside a controlled environment, placing clear focus on healthcare contexts and including heart rate and pre-existing conditions (such as cardiovascular or respiratory diseases) of participants. Similar to this paper, eHealth CSI includes phenotypic information (i.e., gender, height, and weight) of participants. However, CSI data in eHealth CSI is collected in a room using a single Raspberry-Pi board with only one person present at a time. While this dataset is interesting for training models to estimate an individual's medical conditions, its application in multi-person settings, especially in an automotive context, is very limited.

Another dataset called Wi-MIR was presented in [10] with focus on recognizing interactions between two individuals. Wi-MIR presents recorded CSI data from a group of six persons performing around 20 different interactions, all of which taking place in a limited space and while standing. Even though Wi-MIR includes phenotypic data such as weight, height, and gender of participants, its small size and its reduced variability limit applicability in other contexts.

In [11], a CSI dataset for identification tasks and estimating people counts is presented, covering seven distinct activities and including 10 people in seven different (indoor) environments. Three different configurations of hardware transmitter and receiver pairs are provided. The dataset proposes the assessment of obstructions in the direct path by collecting data using a wooden bookcase and a concrete block wall as a barrier, and also provides data in a semi-anechoic chamber. However, since no phenotypic data is included and due to the singular sensor position, the use of this data in vehicular contexts is also limited.

Although no public datasets exists for automotive contexts, there are several applications of WiFi sensing in vehicle interiors, such as occupant detection [12], fatigue level detection [5] and head position detection [6]. However, these publications do not allow access to the data used for training said models and techniques.

In this paper, we address this problem and present the first publicly available CSI dataset for applications of WiFi sensing in automotive vehicles, as described next in detail.

¹More specifically, these subcarriers are 11 guard subcarriers, the DC subcarrier in the center, and 4 pilot subcarriers.

IV. DATA COLLECTION

All data contained in the presented dataset was collected from parked vehicles featuring different settings with the driver alone and a varying number of passengers. We conducted our experiments on different days along a year to account for different environmental conditions. The dataset also includes temperature and humidity data for each measurement, which we recorded at 15 s intervals. Different vehicle models were considered to introduce background variability, as mentioned later in Section V, but we have also consistently used the same reference vehicle throughout our measurements for comparison purposes.

A. Demonstrator

We designed a portable, easy-to-mount demonstrator for our experiments following the general concept presented in [13]. Our demonstrator consists of a standard WiFi router/AP, i.e., an AVM FB W7590AX, that we place on the vehicle's dashboard and seven Raspberry-Pi boards (of type 4B) that are distributed within the cabin and act as CSI sensors. In addition, we use a LogiLink UA0377 camera for videos and an Aosong DHT22 digital temperature/humidity sensor for environmental data.

All components are then connected over an Ethernet switch (of type Netgear GS108PP) to a laptop computer from which we start/stop and monitor measurements. An overview of the demonstrator, including the distribution of components inside the vehicle, can be seen in Fig. 3. Further, Fig. 4 shows a picture of the individual, unconnected components, whereas Fig. 5 shows how CSI sensors are attached to doors.

The router/AP is configured for the 5 GHz spectrum using the 802.11ac standard, transmitting approximately 10 beacons per second on the primary channel with a bandwidth of 20 MHz. The CSI sensors then extract CSI data from the 64 subcarriers of each beacon transmission, i.e., approximately every 100ms.

Now, for collecting and recording CSI data, we have used the freely available WirelessEye framework [7]. WirelessEye

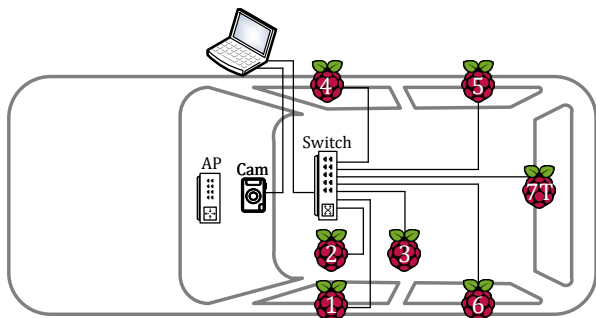


Fig. 3: Demonstrator overview including the exact positioning of CSI sensors (i.e., modified Raspberry-Pi boards). In particular, sensor 2 is below the driver's seat, whereas sensor 3 is attached to the driver's headrest. CSI sensor 7 is located centrally in the back of the vehicle and attached to the temperature/humidity sensor.



Fig. 4: Demonstrator components without cabling.



Fig. 5: CSI sensors 1 and 6 temporarily mounted on the driver's side doors by hook-and-loop fasteners.

deploys a CSI server onto the Raspberry-Pi boards, which have been modified with a Nexmon-based firmware [14] to retrieve CSI data. This data is then collected and stored on the laptop computer running the client software, called *WirelessEye Studio*, see Fig. 6.

B. Data Measurement

As mentioned before, we collected CSI data on different days throughout the year. For each measurement appointment, we consider—with few exceptions—two vehicles, i.e., reference vehicle and an additional random vehicle. Further,

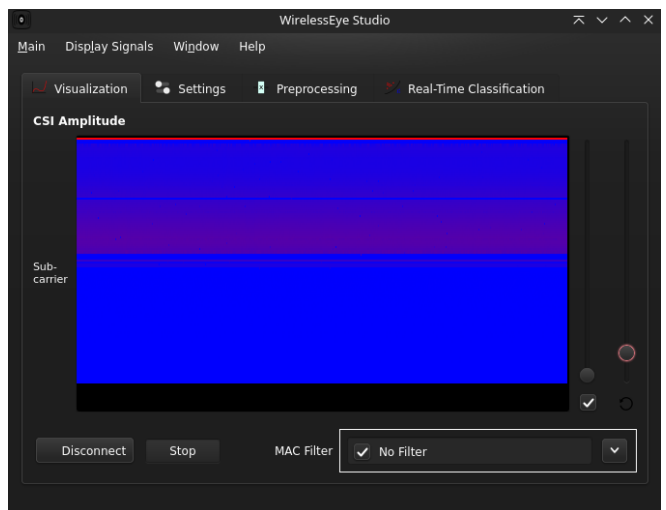


Fig. 6: WirelessEye Studio [7]

up to eight test persons/participants were invited for each measurement session—four for each vehicle, who were recruited via local advertisements and mailing lists available for this purpose.

Each measurement took approximately one hour, including briefing participants and collecting phenotypic data (i.e., height and weight with a standard stadiometer and a scale). Participants were divided into two groups of four people and assigned to a particular vehicle. Measurements of 120 s length were then conducted inside the stationary vehicles, alternating between scenarios with the driver only and the driver with passengers, every time rotating the person at the driver’s seat—clearly, with exception of the participating minors. Finally, the two groups switch vehicles and repeat the above measurements.

C. Anonymized Videos

As previously mentioned, video data was recorded during each measurement by the camera placed centrally on the rearview mirror—see Fig. 7.



Fig. 7: Camera mounted on the rearview mirror pointing at vehicle’s interior.

After every measurement, all video data recorded was anonymized in an irreversible manner by detecting all faces using Ultralytics [15] and yolo-face [16] and masking them out entirely. Videos were then manually checked to make sure that no partial faces are visible, for example, behind headrests, that are overlooked by the face-detection models used. Examples for the anonymized videos can be seen in Figs. 8 and 9.

Since participants were not asked to restrict or restrain themselves in any way during our measurements as long

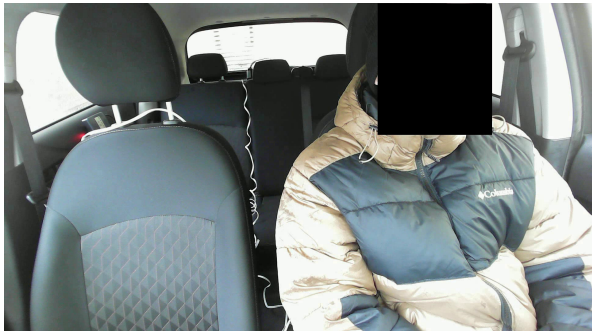


Fig. 8: Anonymized video frame with the driver alone.



Fig. 9: Anonymized video frame with the driver and two passengers.

as they stay in their assigned seats, activities caught in video can be used for further labeling of the CSI data. Examples for observed activities are talking (with other participants), reading from smartphone or inspecting the dashboard/attached sensors, etc.

V. OUR DATASET

As of now, the proposed dataset contains 110 individual measurements with a total of 48 participants, consisting of 19 (39.6%) female and 29 (60.4%) male participants. Further, 3 participants (6.25%) were minors. However, we plan to perform some more measurements in the coming months to extend this dataset further.

Participants’ height is in the range between 144 *cm* and 204 *cm*—see Fig. 10. In addition, participants’ weight ranges from 38 *kg* to 135 *kg* as depicted in Fig. 11.

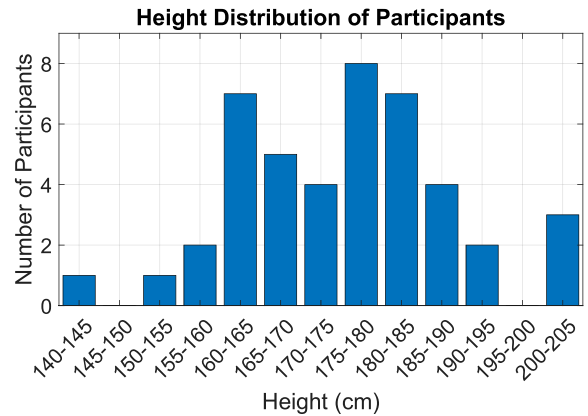


Fig. 10: Height distribution of participants in our dataset.

Our dataset shows an evenly spread distribution of weight classes to form cross validation groups as shown in Table I below.

Group	Count
Below 66 <i>kg</i>	14
66 <i>kg</i> to 80 <i>kg</i>	12
80 <i>kg</i> to 92 <i>kg</i>	12
Above 92 <i>kg</i>	10

TABLE I: Exemplary weight classes for cross validation groups.

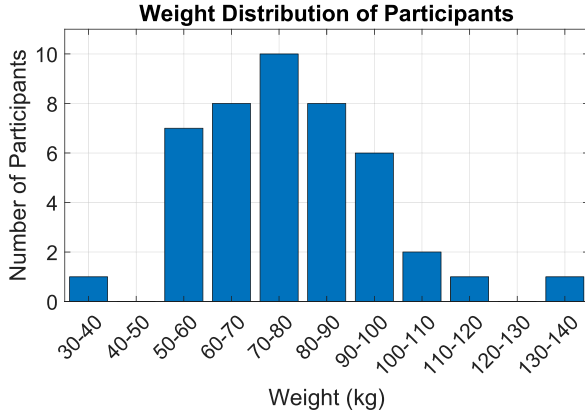


Fig. 11: Weight distribution of participants in our dataset.

Fig. 12 further shows the combined height/weight distribution of all participants.

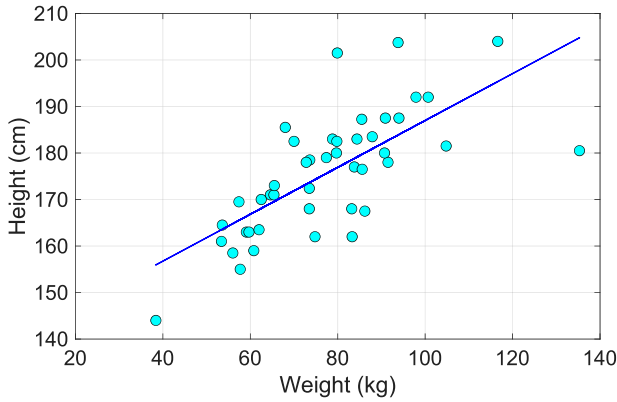


Fig. 12: Combined height/weight distribution with diagonal line depicting the average value.

As mentioned above, for these measurements, a total of six different vehicles were used. These are the following:

- 2017 Volkswagen Golf 7 Combi,
- 2021 Kia Ceed SW,
- 2024 Seat Ateca,
- 2024 Volkswagen Touran,
- 2024 Suzuki Swift, and
- 2024 Mitsubishi SpaceStar.

Note that the 2017 Volkswagen Golf 7 Combi was used for the majority of the measurements to serve as a reference vehicle.

Each of the 110 measurement rounds can be classified as having 1 to 4 participants inside the vehicle. More specifically, measurements can be divided as shown next in Table II.

The majority of our measurements fall into the category of one-person recordings, as this case is treated with priority. Next priority was recording full-vehicle settings, i.e., four-person recordings. Other settings with were recorded depending on the availability of participants, i.e., when less than eight people were available for a measurement session.

Scenario	Count
One-person recordings (driver alone)	67
Two- and three-person recordings	19
Four-person recordings	24

TABLE II: Recorded scenarios.

VI. CASE STUDY

In this section, we demonstrate that the proposed CSI dataset can be used to train ML models to draw conclusions from a vehicle cabin. To this end, we present an illustrative case study where a modified Resnet18 [17] is used to identify individual test persons/participants when sitting alone in one of the vehicles on the driver's seat.

We proceeded slicing data into segments of 45 consecutive beacons (i.e., $\approx 4.6s$). In addition, for each measurement performed on a particular day, 75% of the collected data was used for training, whereas the remaining 25% was used for test and evaluation.

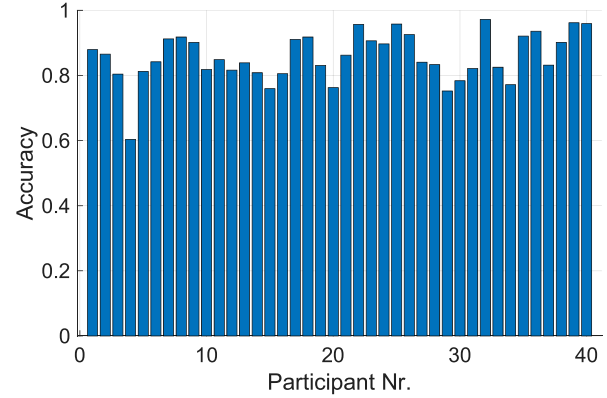


Fig. 13: Identification accuracy of individual participants using CSI data.

The resulting model is able to identify individual participants from the dataset with an average accuracy of 0.86 — see Fig. 13. Moreover, we utilized these identification rates to obtain average weight and height estimates. More specifically, an average weight estimate is obtained as follows:

$$W_{avg} = \frac{1}{c} \sum_{n=1}^c (W_n \cdot p_n),$$

where c is the number of participants (i.e., 48 at the moment), W_n is the annotated weight of the n -th participant and p_n is the prediction rate of the n -th participant, i.e., the estimated likelihood that the corresponding CSI data represents the n -th participant. Similarly, an average height is estimated as follows with H_n being the annotated height of the n -th participant:

$$H_{avg} = \frac{1}{c} \sum_{n=1}^c (H_n \cdot p_n).$$

A comparison between the average estimated and actual values of weight and height is displayed in Fig. 14 and Fig. 15 respectively.

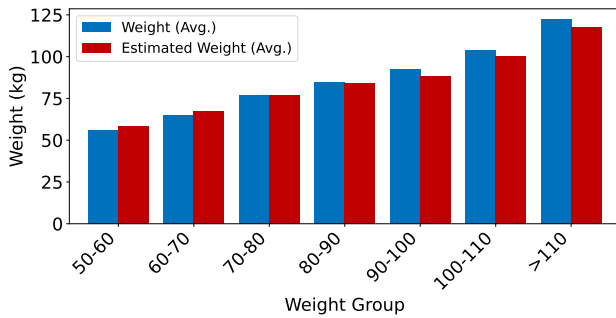


Fig. 14: Accuracy of weight estimation using CSI data.

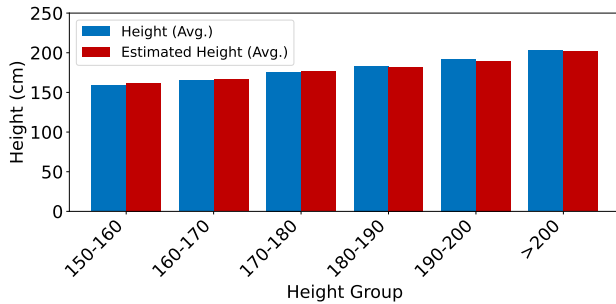


Fig. 15: Accuracy of height estimation using CSI data.

VII. CONCLUSION

In this paper, we present a public CSI dataset focused on vehicle interiors. The proposed dataset includes 110 separate recordings in six different vehicles, each time using seven CSI sensors at different positions. At the moment, our dataset presents data recorded from 48 test persons, however, we plan for more experiments in the coming months.

CSI data is basically annotated with the seat position and phenotypic data of participants, i.e., weight, height, and gender. However, we also include anonymized video material of each recording, which can be useful to further annotate CSI data, for example, by specific activities or movements.

This provided data can be used for a wide range of Wi-Fi-sensing applications inside vehicles, which we demonstrated by an exemplary case study consisting of estimating weight and height of a driver by identification. In future, we intend to refine this and work on generalizable, regression-based height and weight estimations using this CSI dataset.

Overall, the idea is to better configure safety and comfort features of a vehicle, e.g., airbag inflation, headrest position, seat belts, etc., based on actual phenotypic information obtain by Wi-Fi sensing from within the cabin.

ETHICS STATEMENT

Prior to conducting our experiments to build the CSI dataset from this paper, our data-protection concept and procedures were carefully evaluated and subsequently approved by the ethics committee within our institution.

At every measurement session, all participants were informed about the nature the collected data and of our experiments in general. All participants and/or their legal guardians then signed corresponding consent forms. However, no personal information is contained in the presented dataset.

ACKNOWLEDGEMENT

This research is supported by the German Federal Ministry for Digital and Transport (BMDV) innovation initiative mFUND (WiSe4Car — 01F1177A).

The authors would like to thank the reviewers, whose useful and constructive criticism significantly improved the paper.

REFERENCES

- [1] A. I. G. Mowla, R. Hossain *et al.*, “CSI-Based People Counting in WiFi Networks: Leveraging Occupancy Detection,” in *2024 IEEE International Conference on Computing, Applications and Systems (COMPAS)*. IEEE, 2024, pp. 1–6.
- [2] Y. Bai, Z. Wang, K. Zheng, X. Wang, and J. Wang, “WiDrive: Adaptive Wi-Fi-based recognition of driver activity for real-time and safe takeover,” in *2019 IEEE 39th International Conference on Distributed Computing Systems (ICDCS)*. IEEE, 2019, pp. 901–911.
- [3] J. Schäfer, B. R. Barrsiwal, M. Kokkharova, H. Adil, and J. Liebenhenschel, “Human activity recognition using CSI information with nexmon,” *Applied Sciences*, vol. 11, no. 19, p. 8860, 2021.
- [4] C. Turetta, F. Demrozi, P. H. Kindt, A. Masrur, and G. Pravadeelli, “Practical identity recognition using WiFi’s channel state information,” in *2022 Design, Automation & Test in Europe Conference & Exhibition (DATE)*. IEEE, 2022, pp. 76–79.
- [5] W. Jia, H. Peng, N. Ruan, Z. Tang, and W. Zhao, “WiFind: Driver fatigue detection with fine-grained Wi-Fi signal features,” *IEEE Transactions on Big Data*, vol. 6, no. 2, pp. 269–282, 2018.
- [6] X. Xie, K. G. Shin, H. Yousefi, and S. He, “Wireless CSI-based head tracking in the driver seat,” in *Proceedings of the 14th International Conference on emerging Networking EXperiments and Technologies*, 2018, pp. 112–125.
- [7] P. H. Kindt, C. Turetta, F. Demrozi, A. Masrur, G. Pravadeelli, and S. Chakraborty, “WirelessEye-Seeing over WiFi Made Accessible,” in *2024 IEEE International Conference on Pervasive Computing and Communications Workshops and other Affiliated Events (PerCom Workshops)*. IEEE, 2024, pp. 562–567.
- [8] “IEEE Standard for Information technology– Local and metropolitan area networks– Specific requirements– Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications Amendment 5: Enhancements for Higher Throughput,” *IEEE Std 802.11n-2009 (Amendment to IEEE Std 802.11-2007 as amended by IEEE Std 802.11k-2008, IEEE Std 802.11r-2008, IEEE Std 802.11y-2008, and IEEE Std 802.11w-2009)*, pp. 1–565, 2009.
- [9] I. Galdino, J. C. Soto, E. Caballero, V. Ferreira, T. C. Ramos, C. Albuquerque, and D. C. Muchaluat-Saade, “eHealth CSI: A Wi-Fi CSI dataset of human activities,” *IEEE Access*, vol. 11, pp. 71 003–71 012, 2023.
- [10] M. S. Islam, M. H. Kabir, M. A. Hasan, and W. Shin, “Wi-MIR: A CSI dataset for Wi-Fi based multi-person interaction recognition,” *IEEE Access*, 2024.
- [11] F. Meneghello, N. Dal Fabbro, D. Garlisi, I. Tinnirello, and M. Rossi, “A CSI dataset for wireless human sensing on 80 MHz Wi-Fi channels,” *IEEE Communications Magazine*, vol. 61, no. 9, pp. 146–152, 2023.
- [12] Z. Hao, G. Wang, and X. Dang, “Car-sense: vehicle occupant legacy hazard detection method based on DFWS,” *Applied Sciences*, vol. 12, no. 22, p. 11809, 2022.
- [13] M. Ibrahim and K. N. Brown, “Vehicle in-cabin contactless Wi-Fi human sensing,” 2021.
- [14] F. Gringoli, M. Schulz, J. Link, and M. Hollick, “Free your CSI: A channel state information extraction platform for modern Wi-Fi chipsets,” in *Proceedings of the 13th International Workshop on Wireless Network Testbeds, Experimental Evaluation & Characterization*, 2019, pp. 21–28.
- [15] G. Jocher, A. Chaurasia, and J. Qiu, “YOLO by Ultralytics,” Jan. 2023. [Online]. Available: <https://github.com/ultralytics/ultralytics>
- [16] W. Chen, H. Huang, S. Peng, C. Zhou, and C. Zhang, “YOLO-face: a real-time face detector,” *The Visual Computer*, vol. 37, pp. 805–813, 2021.
- [17] K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition,” in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770–778.