Channel State Information for Human Activity Recognition with Low Sampling Rates

Jeroen Klein Brinke
Pervasive Systems
University of Twente
Enschede, the Netherlands
j.kleinbrinke@utwente.nl

Alessandro Chiumento
Pervasive Systems
University of Twente
Enschede, the Netherlands
a.chiumento@utwente.nl

Paul Havinga
Pervasive Systems
University of Twente
Enschede, the Netherlands
p.j.m.havinga@utwente.nl

***Abstract*—In this paper, it is shown that lower transmission/sampling rates can be used in human activity recognition using channel state information ($\mathbf{F_1}$-scores $> 85\%$) and that extremely high sampling rates are unnecessary once the system has been deployed. This is done by analysing the effects of interpolating different sampling rates on Wi-Fi dynamic channel state information for human activity recognition. While current research focuses on training and testing with homogeneous and very high sampling rates (> 100 Hz), this paper outlines some issues with higher sampling rates and explores the impact of training and testing with heterogeneous sampling rates in order to advance more towards joint communication and sensing, where one cannot be certain of the received data rate over time while not knowing the exact training set due to weight sharing in Federated Learning. This paper shows the effect of training and testing with heterogeneous sampling rates (including interpolated datasets) on convolutional neural networks in WiFi sensing.**

***Index Terms*—joint comunication and sensing, channel state information, human activity recognition, device-free sensing, 802.11n, data stability, scalability**

I. Introduction

The desire to monitor the environment and its inhabitants is increasing. This is can be seen in the research trends to develop smaller and less obtrusive sensors, while entire sensing solutions in pervasive computing are getting smarter. The increase in intelligence of these systems is due to the ever-decreasing size of physical hardware, which at the same time is getting more powerful. This allows for more complex and deeper neural networks, which are becoming the brain behind intelligent computing and systems. Moreover, sensing technologies are expanded from physical sensing into truly-unobtrusive, or device-free, sensing.

One of these promising device-free sensing techniques is radio frequency-based sensing through channel state information in the Wi-Fi frequency range. It has been a steadily growing field over the past decade and is applicable in many domains of human activity recognition [1]–[3], vital sign monitoring [4]–[6], and localization [7]–[9]. Wi-Fi channel state information leverages the multi-path propagation of Wi-Fi point-to-point networks, as the same signal can traverse different paths and arrive multiple times at the same endpoint, albeit with different phases and amplitudes. These can be analyzed in both time and frequency domain in order to create insight about changes in the enviroment.

Currently, most research involving RF-based sensing looks at the boundaries of what is physically possible with any given static sampling/transmitting frequency. This sampling frequency commonly ranges between 10^1 and 10^3 Hz (1 kHz). These higher frequencies are determined by the Nyquist rate which is considered anywhere between 300 [10] and 800 Hz [11], depending on the highest frequency of the human activity to be monitored. However, this depends on the data gathered and processing methods: phase is more sensitive to smaller and rapid movements than amplitude, and classification of the frequency domain may require higher sampling rates for detailed analysis than analysing the temporal domain. Therefore, the defined Nyquist frequency applies is dependent on these combinations. For classification purposes, interpolation can be used to predict missing data points when the rate is above the Nyquist rate, and research has shown lower sampling rates can be used for vital sign monitoring [12], [13] and human activity recognition [8], as most human activities and vital signs span the range of 2 - 5 Hz in human activity recognition.

In order to truly facilitate joint communication and sensing, current sensing techniques should be optimized to work in real-life scenarios, where one cannot assume to always receive packets following a steady stream - or in other words, always receive enough data to use the same sampling frequency to collect data. Furthermore, in the future, large-scale Wi-Fi sensing deployments may employ decentralized federated learning in order to reduce resource consumption (only sharing weights and not the training data) and increase privacy. In this case, even when the received data is collected at the Nyquist rate of the activity and can thus be interpolated, the knowledge of which sampling frequency was used for training the received weights may be unknown.

A. Problem statement

Most research in human activity recognition using channel state information is dependent on a static stream of data in the range of 10^1 and 10^3 Hz, where only minor inconsistencies are considered and interpolated. However, in order to achieve joint communication and sensing, a symbiotic relationship would need to be created, in which sensing happens when there is data to transmit; or rather, sensing happens when data is being transmitted. This means that the sampling rate used for training

(F_{train}) could differ considerably from the sampling rate of an actual received stream of packets (F_{test}).

Furthermore, while the received signal can successfully be interpolated above the Nyquist rate (examples given in this paper), in complex mutli-receiver setups covering larger areas (e.g. whole building or large space) employing federated learning, the knowledge of which sampling rate was used to train or optimize the original weights may be lost, as only the weights itself are shared. Therefore, gaining insights in the impact of heterogeneous sampling rates for training and testing on overall system performance could be key to enable autonomous Wi-Fi sensing nodes in decentralized federated learning, as it could enable nodes to detect weights trained at different sampling rates either choose not taint their current weights, or in the future employ smart techniques to combine these weights.

Additionally, the radio-frequency spectrum is an increasingly scarce resource and flooding it with unnecessarily high sampling rates is thus undesired. Moreover, to achieve joint communication and sensing, it is important that any device can participate in both communication and sensing. This also includes lightweight IoT devices with minimum computing resources and power, which in turn means fewer data samples (thus lower transmitting or sampling rates).

B. Contributions

Based on the aforementioned problems, this paper provides insight into lower sampling (and thus transmission) rates and the performance on human activity recognition to reduce bandwidth and neccesarily flooding the wireless environment. Additionally, it explores the combination of different sampling rates for training and testing and their impact to enable heterogeneous transmission and recognition sampling rates (e.g. different devices).

The main contributions of this paper are to gain insight into:

- The effect of cross-sampling frequencies on both training and testing sets to simulate more variable network traffic, including training with lower sampling rates.
- Verifying the findings found in this paper by applying the methodology to different types of datasets.

The rest of the paper is organised as followed: § II outlines current trends in regards to sampling rates, § III describes the process of data collection, § IV explains the methodology, § V and § VI highlight and discuss results found, § VII target limitations and directions for future work, and finally § VIII concludes this paper.

II. State of the Art

As discussed in § I, most research focuses on high sampling rates (> 100 Hz), with only a few focusing on lower sampling rates. In this paper, we will outline current trends regarding sampling rates used in channel state information for human activity recognition in recent literature.

A. Comparative analysis of sampling rates in literature

Widely available channel state information datasets are collected using sampling frequencies over 100 Hz and some offer a comparative analysis for different sampling rates (in such a way that $F_{train} = F_{test}$). The related works covered here are based on human activity recognition in a real-life scenario, thus focusing on regular activities (e.g. walking, sitting, eating), rather than specific or niche scenarios (e.g. human-computer interaction [14] or sign language [15]).

To the authors' best knowledge, the dataset employing the highest sampling frequency has been collected by Wang *et al.* [10], as it is collected at 2,500 Hz. A total of eight activities were collected, mostly resembling activities of daily living (including running/walking, sitting, opening a refrigerator, brushing teeth) for 25 volunteers (aged 19-22). The authors calculated the minimum sampling rate as $600Hz$, by employing the Nyquist sampling rate defined as $F_s \geq 2f_{ha}$ and by observing no human movement was found over $300Hz$ in the channel frequency response power while collecting the dataset. It was found that above the Nyquist sampling rate, the performance did not degrade significantly (94.8% and 96.5% for 800 and 2,500 Hz, respectively). Under the sampling rate, performance seems to degrade, with 87% for 400 Hz. For classification, spectrograms were used, meaning that a lower sampling rate does not result in a differently shaped input, but rather a lower resolution of the spectogram.

Yousefi *et al.* [16] collected a dataset at 1,000 Hz, monitoring activities related to lying down, falling, walking, and sitting, resulting in a total of six classes. It is noted by the authors that 1,000 Hz is "*a good trade-off between capturing fast movements and computational cost*". The authors mention "*severely degraded performance*" at a sampling rate of 50 Hz. The authors use an LSTM for classification and input a 90-dimensional vector (namely 3 antennas $\times$ 30 subcarriers) and an accuracy of 75% is reached. Per trial, only one activity was recorded, meaning each activity is surrounded by "empty" data in which nothing occurs.

A dataset containing information for line-of-sight and non-line-of-sight environments was collected by Alsaify *et al.* [1] and used a sampling rate of 320 Hz. The collected data contains three environments, 30 subjects, five experimental setups, 12 activities and 20 trials. While the paper itself does not include any accuracy metrics, future research done by the author using that dataset included these metrics [17], [18]. Using two environments, six activities, and 20 participants, an F_1-score of $\geq 91\%$ was reached using an SVM with the original 320 Hz. No information is given on the effect of downsampled signals on the performance.

Wiar, collected by [19], is sampled at 30 Hz, which is significantly lower than the others. The authors differentiate between upper-body activities (10 classes), lower-body activities (2 classes), and whole-body activities (4 classes), most of which are related to activities of daily living. In total, 10 participants participated and three locations (rooms) were considered. Like Yousefi *et al.* [16], the activity time is shorter

than the actual time frame, resulting in the activity being surrounded by "empty" data. Using the 30 Hz sampling rate, different machine learning methods were employed, including CNN and LSTM, for three participants. These achieved an accuracy of 0.91 and 0.92 on average, respectively.

B. Sampling rates in recent literature

To explore the recent advances in sampling rates in literature, different types of applications are included, such as gestures, vital sign monitoring, and localization.

In recent years, most research has shifted into higher ($\geq$ 100 Hz) sampling rates. Alizadeh *et al.* [11] used the dataset collected by Wang *et al.* [10] at a sampling rate of 1 kHz, as they calculated the Nyquist sampling rate for human activity monitoring at 800 Hz by calculating the Doppler frequency shift for running and walking targets. An accuracy of $94\% \pm 1.4\%$ was achieved by employing a time-series LSTM-RNN model.

Gao *et al.* [20] compared the performance of WiGesture over different sampling rates. Data was collected independently for each sampling rate, meaning no downsampling was used. The signal was originally captured at 400 Hz, which resulted in an accuracy of 0.9. When the sampling rate is set to $50 > F_s > 400$ Hz, the accuracy drops minimally for all digits (to > 0.8). For most classes, when the sampling rate is set to 50 Hz or below, the accuracy drops to $0.55 - 0.6$.

For general human activity recognition, most papers consider sampling rates over 100 Hz. Chen *et al.* [3], Shalaby *et al.* [2], Li *et al.* [21], and Muaaz *et al.* [22] all employ 1,000 Hz sampling frequency for human activities successfully with varying techniques, number of participants, and number of activities. Cheng *et al.* [23] used 800 Hz for human activity recognition successfully. Yang *et al.* [24], Cui *et al.* [25], Li *et al.* [26], and Hwang *et al.* [27] set the sampling rate to 500 Hz.

Less research explores lower sampling rates, such as 100 Hz and below. Zeeshan *et al.* [8] used a 100 Hz sampling rate to achieve joint localization and human activity recognition. For an omnipotent model, the authors report an accuracy of 0.99 for both activity recognition and localization, while for a model that has not seen all classes, the accuracy drops to 0.93 for activity recognition. For vital sign monitoring, Wang *et al.* [28] used 20 Hz for phase analysis to achieve a good performance. Additionally, 10 Hz was also used to achieve resilient respiration rate monitoring by Wang *et al.* [12]. Li *et al.* successfully monitored the heart rate using 10 Hz [13]. For human activity recognition, Shruti *et al.* used 20 Hz to reduce the false negative rate of six different activities. Yang *et al.* [29] used a comparable rate (30 Hz) with a focus on reducing data by selecting the most sensitive antenna and enhancing active data streams, while limiting the effect of inactive streams.

As can be seen, there is a wide range of sampling rates being used in research, with most being over 100 Hz. However, for problems outlined in § I-A, this is undesirable and an effort should be made to find a way to use a minimal sampling rate.

III. Data acquisition

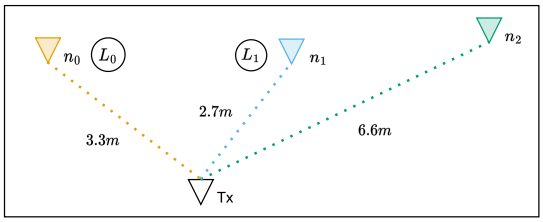

Fig. 1: Schematic of the experimental setup. Furniture and other obstructions removed for clarity. Note that this image is not to scale, but the distances mentioned are correct.

1) Hardware and software: Each transceiver node for this research was modified to fit an Intel Ultimate Wi-Fi Link 5300 NIC and three antennas. The Intel NIC was chosen in order to use the open CSI platform by D. Halperin, W. Hu, A. Sheth, and D. Wetherall [30]. The access point was put in the *injector mode*, while the receivers were placed in *monitor mode* (conforming to the 802.11n specifications). Channel 64 was chosen with a center frequency of 5.32 GHz. The transmitter injected random packets at a frequency of $F_{orig} = 100$ Hz using all three antennas. This results in a CSI matrix of 3x3x30.

2) Participants: The total data was collected from twenty-one healthy participants (11F, 10M), with an average height of 173.52 ± 8.89cm and an average weight of 67 ± 11.25kg.

3) Activities: The selected activities were associated with agitation monitoring in Alzheimer patients (based on the SOAPD scale). Six micro-activities were selected based on this scale, namely *flipping objects*, *kicking table leg*, *rubbing table top*, *sitting and standing*, *nervously tapping on leg or armrest*, and *wringing the hands*.

All nodes are visualised in Fig. 1, with n_i being the identifier used in the remainder of the paper. The two locations location L_0 and location L_1 are also highlighted. Note that any obstructions (e.g. furniture or kitchen appliances) have been removed from the schematic, but were stationary over the experiments.

IV. Methodology

Wireless signal propagation is heavily affected by environmental components causing reflections, diffraction, refraction, and scattering. Every channel between two antennas can take a different path, resulting in more or fewer obstacles and environmental influences. The states of these channels can be logged in a channel state information matrix $\mathbf{H}$.

This matrix has a two-dimensional shape of $N_T \times N_R$ elements (number of transmitting and receiving antennas, respectively), where each element H_{ij} is defined as an amplitude and phase of the arrived signal between any T and R This relation is expressed as $\mathbf{H}_{ij} = ||\mathbf{H}_{ij}||e^{j\angle\mathbf{H}_{ij}}$, where $||\mathbf{H}_{ij}||$ is the received amplitude and $\angle\mathbf{H}_{ij}$ the received phase, given as $a + bj$ by the used software

A. Spline interpolation

Spline interpolation is used for interpolating the down-sampled signal. Low-order polynomials are used between subsequent data points, rather than a high-order polynomial between all the points. Note that the assumed time frame for these interpolations is a window of 1 second. Four different interpolation techniques are averaged (if the minimum number of required data points was available), namely cubic ($F_s >= 4$ Hz), quadratic ($F_s >= 3$ Hz), linear ($F_s >= 2$ Hz), and previous point replication ($F_s >= 1$ Hz).

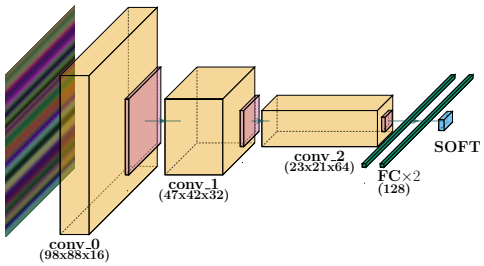

Fig. 2: Visualization of the neural network used in this research. Note that every convolutional layer (yellow) is followed by a leaky ReLU and dropout layer prior to the pooling layer (red).

B. Convolutional neural network

A single network was used for training, but it was split in three different networks were trained and tested to validate the test results, with an increasing depth to simulate the activation layers. The training and testing set were randomly generated three times per classification trial, while each trial used 10-fold cross-validation. It should be noted that the main purpose of this paper is not to find the best convolutional neural network, but rather explore the effect of downsampling training and testing sets.

For each training set, three different training sets were generated: one where all data was considered, one where every odd or even frame was considered, and finally one where only frames modulo 5 were considered. This was done to reduce the temporal dependency of the training set and thus to reduce chances for overfitting. These are not individually discussed. Rather, these were compared to ensure they had similar performance. As this turned out to be the case, these were all aggregated in the remainder of the paper.

For each trained network, the weights were randomly initialized and optimized using the Adam algorithm for a total of 250 epochs. The learning rate was set to $1 * 10^{-3}$ with a learning rate decay of 0.

V. Results

Fig. 3 shows the effect of different sampling frequencies for either training or testing, with the opposite one fixed to F_{orig}: for *a* and *b*, the testing rate was downsampled, while for *c* and d the training rate was downsampled. Additionally, observations were made for two different locations of activities, namely location L_0 (*a*, *c*) and location L_1 (*b*, *d*).

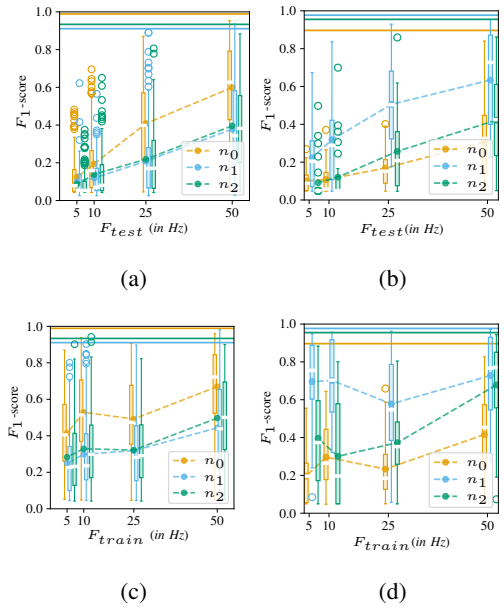

Fig. 3: Visualization of the effect of downsampling testing set (top row) and training set (bottom row) of the self-collected dataset when training with $F_s = 100$ Hz for both location L_0 (left column) and location L_1 (right column) averaged over $N = 21$ participants over all randomization trials and 10-fold cross validation (30 classifications per person). Horizontal lines show the average accuracy for $F_{train} = F_{test} = 100$ Hz.

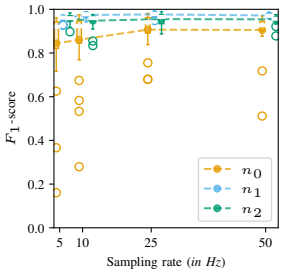

Fig. 4: Figure showing the results of interpolation using four techniques on aggregated down-sampled data over both location L_0 and location L_1. These results are aggregated over different participants using different data segmentation (regular, skip-1, and skip-4) for $N = 21$ and all different interpolation techniques due to similar performance.

It can be seen that for location L_0 node n_0 remains more stable than node n_1 and node n_2 ($0.5-0.8$ and $0.2-0.55$, respectively) when looking at downsampled testing data between 25 and 50 Hz. Below 25 Hz, the F_1-scores are < 0.6, below 10 Hz < 0.3, and below 5 Hz it becomes random guessing (0.2) for $\forall n \in N$. Likewise, for downsampled training data, node n_0 has a slightly higher performance than node n_1 and

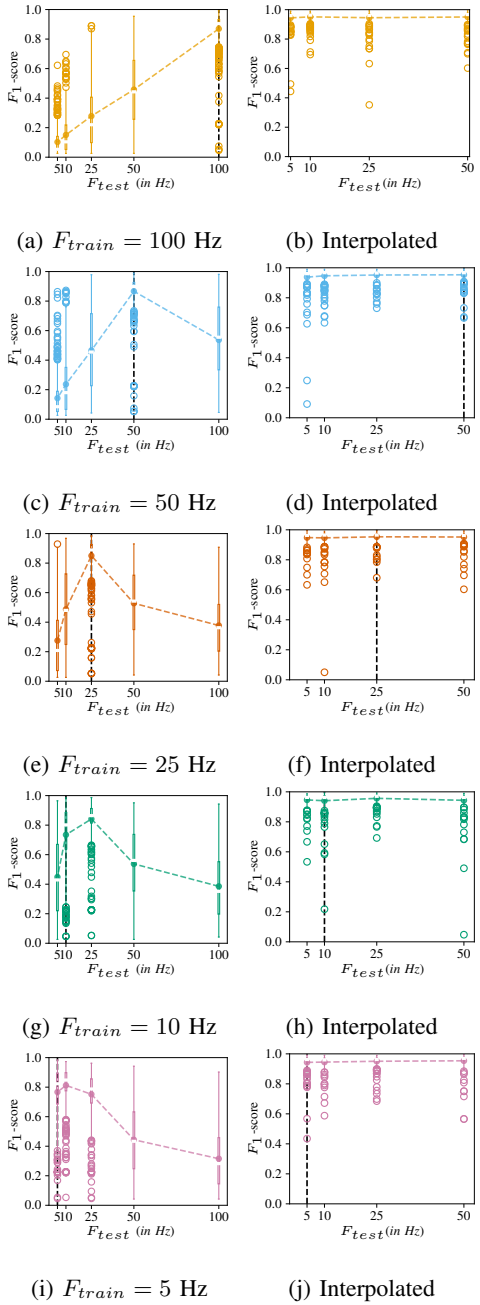

Fig. 5: Performance of different sampling rates for training and testing sets without interpolation (left) and showing the effect of interpolation on the performance (right).

node n_2 ($+0.05$). Unlike the downsampled training set, the accuracy does not seem to fall below 0.3 for $F_s < 25$. It can be observed that all nodes seem to stabalize more for $F_s \geq 25$ Hz, with only node n_1 dropping significantly between $F_s = 10$ and $F_s = 10$, from an F_1-score of 0.5 to 0.39.

For location L_1, node n_1 achieves the highest performance when considering the downsampled testing sets, with F_1-scores of 0.6, 0.5, and < 0.35 for $F_s = 50, 25$, and< 10 Hz, respectively. Node n_0 is overall the worst performer (-0.4 compared to node n_1 on average). Comparable to location L_0, the average performance is slightly higher for a downsampled training set $+0.05$ on average.

For both locations, $\forall n \in N$ achieved an F_1-score > 0.9 for $F_{train} = F_{test} = F_{orig}$.

The effects of previous, linear, quadratic, and cubic interpolation are averaged in Fig. 4, due to being close in performance. Here, it can be observed that the original performance can be restored by using any spline interpolation technique outlined in § IV-A.

For $F_s \leq 10$, the performance is in some cases restored, but the average is usually lower than F_{orig}. This is likely due to losing some information from the activities. While the performance overall is still significant (F_1-score > 0.8) for lower sampling rates $F_s < 5$ Hz, certain participants may finger tap more rapidly than once every 0.2 seconds when nervous. It should be noted that quadratic and cubic (c and d, respectively) require $F_s > 2$ as they require more than 2 sampling points, and thus there is no value given for $F_s = 2$. Therefore, for comparison reasons, no interpolation is done for $F_s < 5$ Hz.

The effect of downsampling with different sampling rates for the training and testing set can be seen in Fig. 5. Results per sampling rate for F_{test} are visualized per F_{train} (a,c,e,g,i *for* $F_{train} = 100, 50, 25, 10, 5$, respectively). It can be seen that for $F_{train} = F_{test}$ performance remain high (> 0.8). For $F_{train} = 5, 10$, and 25 Hz, the average performance is higher than for $F_{train} = 50$ or 100 Hz, namely 0.61 and 0.47, respectively.

The vertical black line in Fig. 5. represents the case where $F_{train} = F_{test}$. It can be seen that for $F_{train} > F_{test}$, the performance degradation is less severe: on the right side of the line, the drop off is less steep (less performance lost per Hz).

As with Fig. 4, it can be seen interpolation approximates the original performance Fig. 5 (b,d,f,h, and j). Between the comparison in Fig. 3, 4, and 5, it appears the latter has more outliers. This is due to the averaging of participants, nodes, and locations, which are individually plotted in Fig. 3 and 4.

VI. Discussion

As can be seen in § V and Fig. 3, the performance for $\forall n \in N$ for $l \in L$ is quite similar for $F_{test} = F_{train} = 100$ Hz. While the magnitude of accuracy drops is different depending on the location when downsampling, the overall degradation follows a similar curve. This implies each node learns similar features from the performed activities (albeit in different gradients) and are thus similarly affected by downsampling. Better performance can thus be seen for node n_0 in location L_0 and node n_1 in location L_1, as these are being less affected by the extra distance between node and location. Due to the comparable behaviour for $\forall n \in N$ for $l \in L$ and the focus of the paper being looking at the broader effect of downsampling over any node and location, results between nodes and locations are aggregated from here on.

Three cases can be identified worth discussing when considering training and testing with different sampling rates, namely $F_{train} = F_{test}$ (case 1), $F_{train} > F_{test}$ (case 2), and $F_{train} < F_{test}$ (case 3).

- **Case 1.** When $F_{train} = F_{test}$, the overall performance is comparable to F_{orig} (F_1-score ≈ 0.85). This is inline with literature, where a successful performance can be received with lower sampling rates as long as the sampling rates are equal to one another. This corresponds to Fig. 5 *a*, *c*, *e*, *g*, and *i*, where $F_{train} = F_{test}$.
- **Case 2.** When $F_{train} > F_{test}$, the overall performance ultimately drops to random guessing (F_1-score of 0.167). This effect is greatest when the difference between the two rates is highest ($F_{train} = 100$ Hz and $F_{test} = 5$ Hz, 5 *a*). Likewise, the drop in performance is lesser when the sampling rates are closer, with a minimum effect seen when the two are closest, namely $F_{train} = 10Hz$, $F_{test} = 5$ Hz (Fig. 5 *g*). It is likely that when the sampling rate of the training set is higher, the features the neural networks extracts are better preserved with similar sampling rates. This explains why for drastically lower sampling rates for the test set with a higher sampling rate for the train set (e.g. 5 and 50 Hz, respectively) the accuracy drops off after 25 Hz for both $F_{train} = 50, 100$, as these are far away from the remaining sampling rates $(25, 10, 5)$. These effects are visualized in Fig. 5.
- **Case 3.** When $F_{train} < F_{test}$, the overall performance impact seems lesser than for Case 2. For the largest difference in rates ($F_{train} = 5$ Hz and $F_{test} = 100$ Hz, 5 *i*) an F_1-score of 0.39 is measured. While not usable as is, this is still twice as high as random guessing. In certain cases, either with fewer classification classes or more diverse activities (e.g. anomaly detection), this implies devices with different transmission rates can use the same neural network, reducing cost resources. Additionally, it can be observed that when the sampling rates are closer ($25-5$ Hz) in favour of the test set, the performance does not degrade as much and in some cases is even higher (*g* and *i*). This implies that it is easier to train on features from a lower resolution training set and finding these features in a higher resolution testing set ($F_{train} < F_{test}$) than vice versa (Case 2). This may be attributed to it being easier to extract already trained features from a high resolution image, rather than finding high resolution features in a low resolution image.

Outliers in the results can be found for $F_{train} = 5, 10$ Hz and $F_{test} = 5, 10$ Hz, where the accuracy for both is higher when testing with a sampling rate slightly higher than the training rate ($+0.05$ to $+0.1$). This could be due to the training set providing more possible data points for the training set to be mapped to while testing.

Overall, without applying interpolation, these results imply training a convolutional neural network network slightly under the expected received signal rate seems most optimal. This is due to a combination of having similar rates for training and testing is beneficial in all three cases, where $F_{train} = F_{test}$ achieve the highest performance. However, in joint communication and sensing one cannot assume that the data rate is always the same. Comparing case 2 and case 3, overall higher performance is achieved when F_{train} is slightly lower than F_{test}, within the range of 10 Hz. Moreover, the slope of the loss in performance for case 2 and 3 is different and this feature could potentially be used to identify the original training rate of the received weights in federated learning.

The same and sometimes slightly higher performance is gained when interpolating the downsampled signal as shown in Fig. 4 and Fig. 5 (*b*, *d*, *f*, *h*, and *j*). This is inline with current knowledge and expectations that signals can be well replicated interpolated as long as the sampling rate is gathered at the Nyquist rate, which is around 5 Hz for these activities. It is likely that spline interpolation can be seen as a filter: low-level polynomials between a small set of points reduce the number of outliers caused by noise in the noisy wireless environment and thus reducing the overall variance. The reason this holds for the lower sampling rates $F_s < 10$ Hz is that the activities are significantly slow to still be distinctive.

With spline interpolation, a polynomial is applied over a low number of data points, which corresponds to $\frac{1}{F_s'}$ seconds (as discussed in § IV-A). So, in the case of $F_s' = 50$ Hz, this means over every 0.02 seconds. This window is relatively small compared to the duration of an activity performed during the experiments, which resulted in all interpolation methods worked equally well.

VII. Future work

This paper considered classification windows of 1 second, which would result in blank frames (frames containing no data) when the sampling frequency is below 5, or even 1 Hz. Therefore, future research should consider longer classification frames so that sampling frequencies below 1 Hz can be explored, as these results imply interpolation could restore the performance to a certain degree, likely depending on the activity.

Another aspect of joint communication and sensing worth exploring in human activity recognition is simulating proper network behavior, especially the aspect of burst data and more correct arrival time distributions (e.g. Poisson arrival process). At times, receivers may receive a burst of a few seconds, followed by a period of little to no packets. It is worth considering how unsupervised networks can be adopted to learn from the higher rates in order to classify (interpolated) lower sampling rates.

VIII. Conclusion

This research has shown the impact of training and testing with different sampling rates in device-free WiFi sensing. This is a potential important aspect in joint communication and sensing combined with federated learning, as many devices will operate at different maximum transmission rates, not know the original training parameters for the received weights, or not have the resources to (re)train a network. Thus, being

able to use different sampling rates or identify the sampling rate a neural network is trained at by comparing the performances could be advantageous.

The results imply that lower sampling rates, or even different rates across the training and testing set, can be used for monitoring after interpolation and sensing does not need to be dependent on high sampling rates using the scarce wireless resources for human activity recognition ($2-5$ Hz), where the Nyquist frequency is more realistically around $4-10$ Hz.

Finally, it appears lowering the sampling rate of the training set has less of an impact on the overall performance. Especially when the sampling rates are between $5-25$ Hz and the training set has a comparable rate than the testing set, the performance could be likely be used in binary problems (e.g. presence or anomaly detection).

REFERENCES

[1] B. A. Alsaify, M. M. Almazari, R. Alazrai, and M. I. Daoud, "A dataset for wi-fi-based human activity recognition in line-of-sight and non-line-of-sight indoor environments," *Data in Brief*, vol. 33, p. 106534, 12 2020.

[2] E. Shalaby, N. ElShennawy, and A. Sarhan, "Utilizing deep learning models in csi-based human activity recognition," *Neural Computing and Applications*, vol. 34, pp. 5993–6010, 4 2022. [Online]. Available: https://link.springer.com/10.1007/s00521-021-06787-w

[3] Y.-S. Chen, C.-Y. Li, and T.-Y. Juang, "Dynamic associate domain adaptation for human activity recognition using wifi signals," vol. 2022-April. IEEE, 4 2022, pp. 1809–1814. [Online]. Available: https://ieeexplore.ieee.org/document/9771677/

[4] I. Shirakami and T. Sato, "Heart rate variability extraction using commodity wi-fi devices via time domain signal processing." IEEE, 7 2021, pp. 1–4. [Online]. Available: https://ieeexplore.ieee.org/document/9508523/

[5] M. I. Khan, M. A. Jan, Y. Muhammad, D.-T. Do, A. ur Rehman, C. X. Mavromoustakis, and E. Pallis, "Tracking vital signs of a patient using channel state information and machine learning for a smart healthcare system," *Neural Computing and Applications*, 1 2021. [Online]. Available: http://link.springer.com/10.1007/s00521-020-05631-x

[6] K. Ali, M. Alloulah, F. Kawsar, and A. X. Liu, "On goodness of wifi based monitoring of sleep vital signs in the wild," *IEEE Transactions on Mobile Computing*, pp. 1–1, 2021. [Online]. Available: https://ieeexplore.ieee.org/document/9423591/

[7] M. Z. Khan, A. Taha, W. Taylor, M. A. Imran, and Q. H. Abbasi, "Non-invasive localization using software-defined radios," *IEEE Sensors Journal*, vol. 22, pp. 9018–9026, 5 2022. [Online]. Available: https://ieeexplore.ieee.org/document/9738615/

[8] M. Zeeshan, A. Pandey, and S. Kumar, "Csi-based device-free joint activity recognition and localization using siamese networks." IEEE, 1 2022, pp. 260–264. [Online]. Available: https://ieeexplore.ieee.org/document/9668391/

[9] H. Choi, M. Fujimoto, T. Matsui, S. Misaki, and K. Yasumoto, "Wi-cal: Wifi sensing and machine learning based device-free crowd counting and localization," *IEEE Access*, vol. 10, pp. 24 395–24 410, 2022, overlap of 3s. [Online]. Available: https://ieeexplore.ieee.org/document/9724220/

[10] W. Wang, A. X. Liu, M. Shahzad, K. Ling, and S. Lu, "Understanding and modeling of wifi signal based human activity recognition." ACM, 9 2015, pp. 65–76.

[11] R. Alizadeh, Y. Savaria, and C. Nerguizian, "Human activity recognition and people count for a smart public transportation system." IEEE, 10 2021, pp. 182–187. [Online]. Available: https://ieeexplore.ieee.org/document/9604986/

[12] X. Wang, C. Yang, and S. Mao, "Resilient respiration rate monitoring with realtime bimodal csi data," *IEEE Sensors Journal*, vol. 20, pp. 10 187–10 198, 9 2020. [Online]. Available: https://ieeexplore.ieee.org/document/9076681/

[13] S. Li, A. T. Kristensen, A. Burg, and A. Balatsoukas-Stimming, "Complexbeat: Breathing rate estimation from complex csi," vol. 2021-Octob. IEEE, 10 2021, pp. 217–222. [Online]. Available: https://ieeexplore.ieee.org/document/9604942/

[14] Y. Zhang, Y. Zheng, K. Qian, G. Zhang, Y. Liu, C. Wu, and Z. Yang, "Widar3.0: Zero-effort cross-domain gesture recognition with wi-fi," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pp. 1–1, 2021.

[15] Y. Ma, G. Zhou, S. Wang, H. Zhao, and W. Jung, "Signfi," *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, vol. 2, pp. 1–21, 3 2018. [Online]. Available: https://dl.acm.org/doi/10.1145/3191755

[16] S. Yousefi, H. Narui, S. Dayal, S. Ermon, and S. Valaee, "A survey on behavior recognition using wifi channel state information," *IEEE Communications Magazine*, vol. 55, pp. 98–104, 10 2017.

[17] B. A. Alsaify, M. M. Almazari, R. Alazrai, and M. I. Daoud, "Exploiting wi-fi signals for human activity recognition." IEEE, 5 2021, pp. 245–250. [Online]. Available: https://ieeexplore.ieee.org/document/9464613/

[18] B. A. Alsaify, M. M. Almazari, R. Alazrai, S. Alouneh, and M. I. Daoud, "A csi-based multi-environment human activity recognition framework," *Applied Sciences*, vol. 12, p. 930, 1 2022. [Online]. Available: https://www.mdpi.com/2076-3417/12/2/930

[19] L. Guo, L. Wang, C. Lin, J. Liu, B. Lu, J. Fang, Z. Liu, Z. Shan, J. Yang, and S. Guo, "Wiar: A public dataset for wifi-based activity recognition," *IEEE Access*, vol. 7, pp. 154 935–154 945, 2019.

[20] R. Gao, M. Zhang, J. Zhang, Y. Li, E. Yi, D. Wu, L. Wang, and D. Zhang, "Towards position-independent sensing for gesture recognition with wi-fi," *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, vol. 5, pp. 1–28, 6 2021, should compare with this paper to see if the findings are the same. [Online]. Available: https://dl.acm.org/doi/10.1145/3463504

[21] W. Li, M. J. Bocus, C. Tang, R. J. Piechocki, K. Woodbridge, and K. Chetty, "On csi and passive wi-fi radar for opportunistic physical activity recognition," *IEEE Transactions on Wireless Communications*, vol. 21, pp. 607–620, 1 2022. [Online]. Available: https://ieeexplore.ieee.org/document/9497736/

[22] M. Muaaz, A. Chelli, M. W. Gerdes, and M. Pätzold, "Wi-sense: a passive human activity recognition system using wi-fi and convolutional neural network and its integration in health information systems," *Annals of Telecommunications*, vol. 77, pp. 163–175, 4 2022. [Online]. Available: https://link.springer.com/10.1007/s12243-021-00865-9

[23] X. Cheng, B. Huang, and J. Zong, "Device-free human activity recognition based on gmm-hmm using channel state information," *IEEE Access*, vol. 9, pp. 76 592–76 601, 2021. [Online]. Available: https://ieeexplore.ieee.org/document/9438707/

[24] J. Yang, X. Chen, H. Zou, D. Wang, Q. Xu, and L. Xie, "Efficientfi: Towards large-scale lightweight wifi sensing via csi compression," *IEEE Internet of Things Journal*, pp. 1–1, 2022, very interesting, contender for the Federated Learning paper. [Online]. Available: https://ieeexplore.ieee.org/document/9667414/

[25] W. Cui, B. Li, L. Zhang, and Z. Chen, "Device-free single-user activity recognition using diversified deep ensemble learning," *Applied Soft Computing*, vol. 102, p. 107066, 4 2021. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S1568494620310048

[26] L. Li, L. Wang, B. Han, X. Lu, Z. Zhou, and B. Lu, "Subdomain adaptive learning network for cross-domain human activities recognition using wifi with csi," vol. 2021-Decem. IEEE, 12 2021, pp. 1–7. [Online]. Available: https://ieeexplore.ieee.org/document/9763693/

[27] K.-M. Hwang and S.-C. Kim, "A study of cnn-based human behavior recognition with channel state information," vol. 2021-Janua. IEEE, 1 2021, pp. 749–751, half of the data removed, thus only 500 frames are considered. However, was collected (and thus intended) at 1000 hz. [Online]. Available: https://ieeexplore.ieee.org/document/9333879/

[28] X. Wang, C. Yang, and S. Mao, "On csi-based vital sign monitoring using commodity wifi," *ACM Transactions on Computing for Healthcare*, vol. 1, pp. 1–27, 7 2020. [Online]. Available: https://dl.acm.org/doi/10.1145/3377165

[29] J. Yang, Y. Liu, Z. Liu, Y. Wu, T. Li, and Y. Yang, "A framework for human activity recognition based on wifi csi signal enhancement," *International Journal of Antennas and Propagation*, vol. 2021, pp. 1–18, 2 2021. [Online]. Available: https://www.hindawi.com/journals/ijap/2021/6654752/

[30] D. Halperin, W. Hu, A. Sheth, and D. Wetherall, "Tool release: Gathering 802.11n traces with channel state information," *ACM SIGCOMM CCR*, vol. 41, no. 1, p. 53, Jan. 2011.