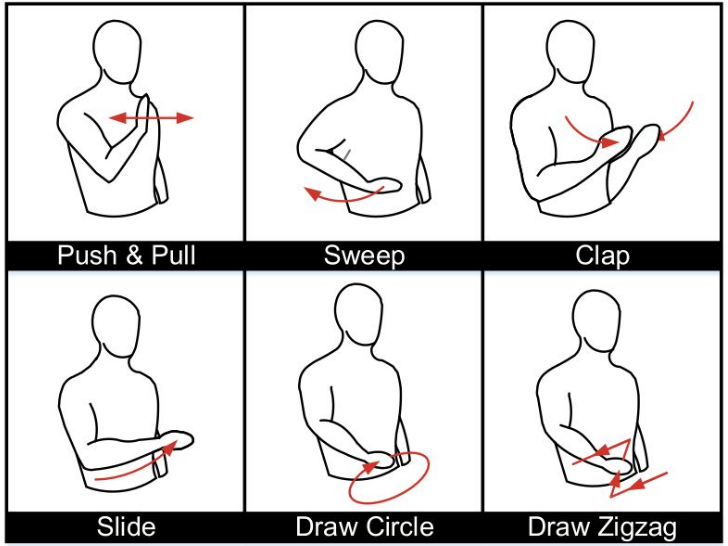
Data Exploration on WiFi Human Activity Recognition

Emmanuel Magana Mendez and Pooya Aghanoury

Table of Contents

Introduction

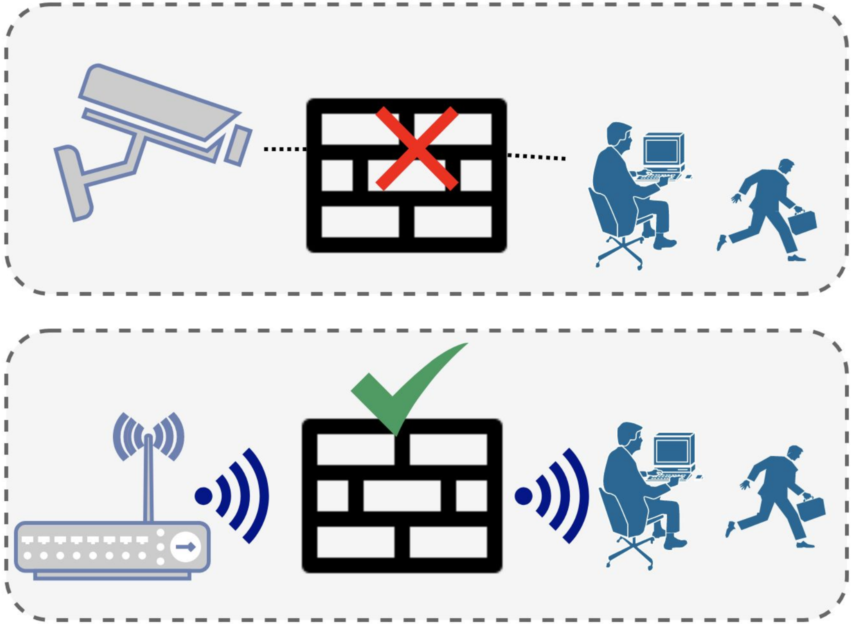
- What is Human Activity Recognition (HAR)
  - Identifying and interpreting the actions and activities performed by humans based on data from various sensors.
- Data Acquisition
  - Vision based
  - IMU
  - Microphones [12]
  - RF (*this work*)
- WiFi Based HAR
  - Utilize native MIMO Channel State Information (CSI) to detect changes in WiFi signal strength
- Goal of this work: Focus on WiFi and pre-existing datasets
- Contribution: we find that each dataset provided has ample room for sample reduction without sacrificing performance.



[8] Set of measured gestures

Background

- Why WiFi-based HAR?
  - Camera based systems: Require line-of-sight (LOS) and may have privacy concerns
- WiFi is ubiquitous in indoor settings, therefore can provide a method for passive detection

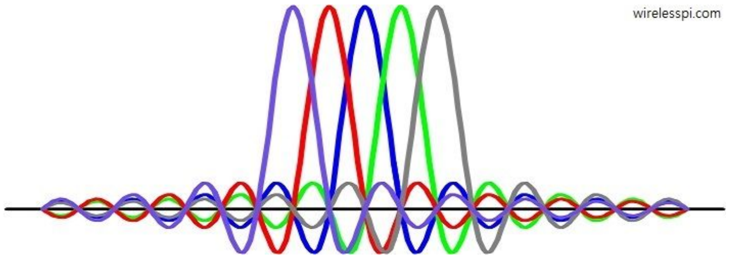


Background: CSI

CIR: $$h(\tau) = \sum_{l=1}^{L} \alpha_l e^{j\phi_l} \delta(\tau - \tau_l)$$

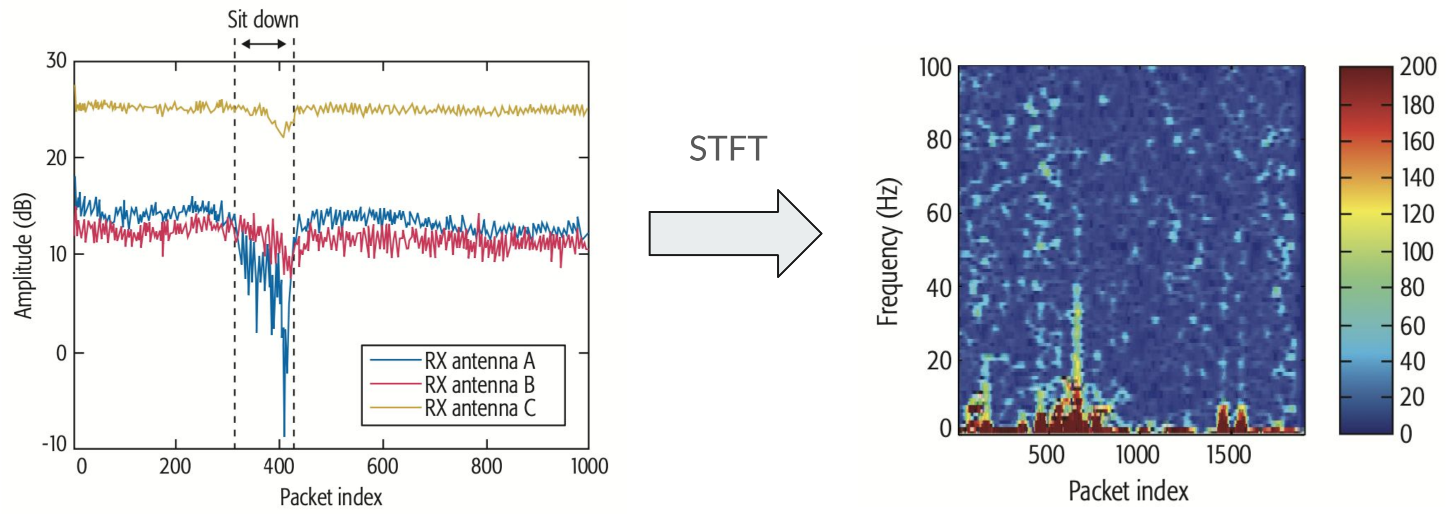
Estimate: $$H_i = ||H_i||e^{j\angle H_i}$$

- Channel State Information (CSI)
    - Describe how a signal propagates in an environment full of reflections, diffraction, and scattering.
    - MIMO and OFDM increase the number of channels
    - Modern APs have multiple antennas and many subcarriers per.
- How can it be used for HAR
    - RSS is an averaged over entire bandwidth
    - Tools like Intel NIC or Atheros CSI tool can be used to gather this information – where CSI is composed of the amplitude and phase of each channel *i*
    - CSI provides a *more diverse* set of information about the HAR – it is essentially a “WiFi Image”



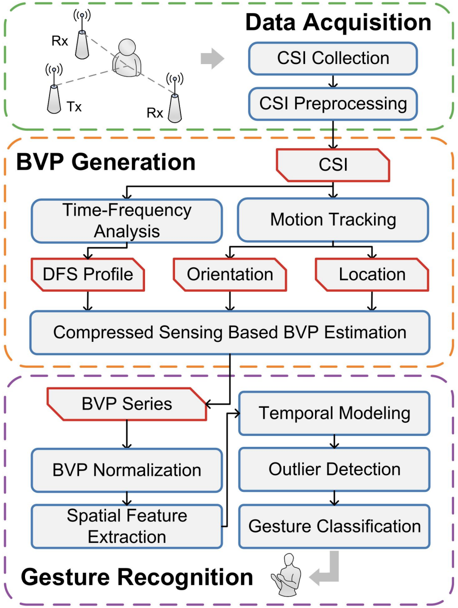
Example OFDM [10]

Background: CSI Feature Extraction - UT-HAR [7]

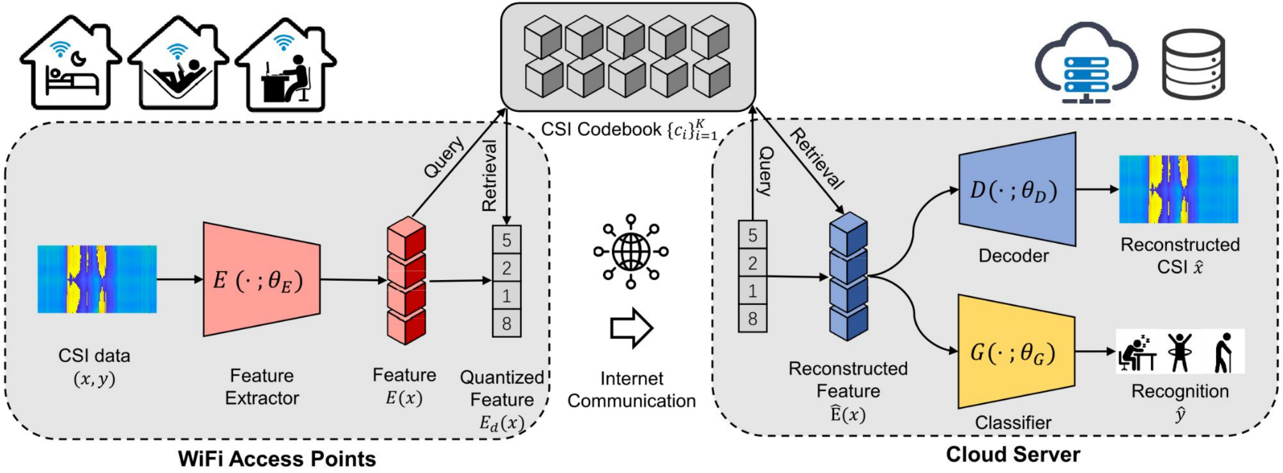


Background: CSI Feature Extraction - WIDAR [3]

- Generate Body-coordinate velocity profile (BVP) from CSI
- Pipeline first converts CSI to BVP
- Features then extracted from BVP



Background: CSI Feature Extraction - EfficientFi [3]

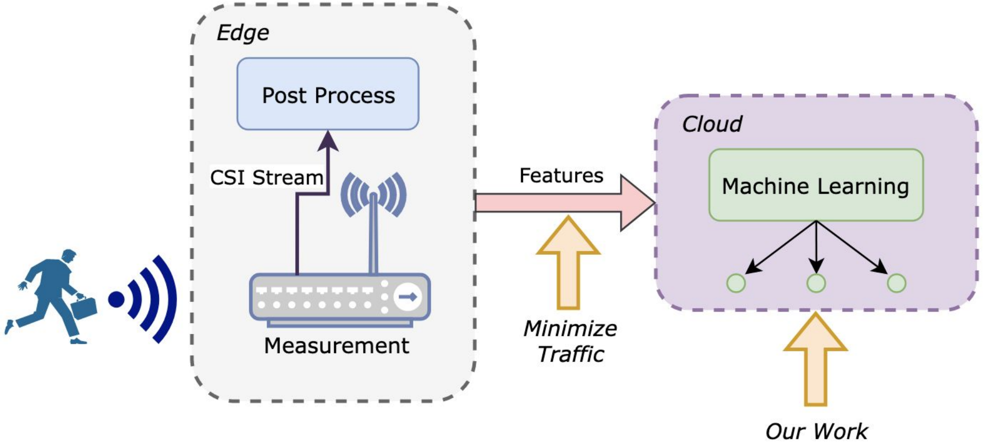


Background: Deep Learning Models

- MLP
  - Pro: Simple and robust architecture
  - Con: Slow convergence and significant computational costs
- CNN
  - Pro: Excel in capturing spatial and temporal features in data
  - Con: CNNs may have an insufficient receptive field due to limited kernel size, and they traditionally stack all feature maps equally
- RNN
  - Pros: Recurrent Neural Networks (RNNs) are capable of memorizing arbitrary-length sequences of input patterns, making them highly effective for handling time sequence data, such as video and CSI, due to their ability to process multiple inputs and outputs.
  - Con: RNNs face challenges with capturing long-term dependencies in data; suffer from the vanishing gradient problem during backpropagation.
- LSTM
  - Pro: Long Short-Term Memory (LSTM) addresses the vanishing gradient problem of traditional RNNs, allowing for better handling of long-term dependencies in sequential data.
  - Con: The LSTM architecture introduces increased complexity compared to standard RNNs

Motivation

- CSI is gathered at the edge
- Post processing (may include denoising) and then feature extraction
    - Short-Time Fourier Transform (STFT)
    - Velocity Profile
- Feature data is offloaded to servers for ML Classification
- **Goal**: Analyze the relationship between sampling rate and classification performance
- **Benefit**: *minimized* traffic between Edge and Cloud



Related Works

- UT-HAR [7]
  - Provides a data set containing measurements of 7 different activities
- NTU-FI [6]
  - HAR
    - Dataset containing 6 different activities
  - HumanID
    - Dataset containing the gait of 15 people
- Widar [3]
  - Dataset containing 22 different activities
- SignFi [5]
  - Dataset containing 256 different signed symbols

Technical Approach

Datasets

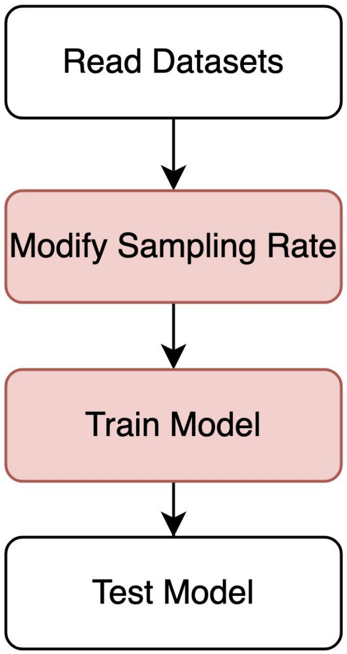
- NTU-Fi-HAR [6]
- NTU-Fi-HumanID [6]
- UT-HAR [7]
- Widar [3]
- SignFi [5]

Platforms

- Python
  - Pytorch
  - Sklearn
  - SenseFi
- Matlab

Models:

- Multi-layer Perceptron (MLP)
- Recurrent Neural Network (RNN)
- Gated Recurrent Unit (GRU)
- Gated Recurrent Unit + Convolutional Neural Network (GRU+CNN)
- LeNet (Special Type of CNN) [11]
- CNN – Used in SignFi

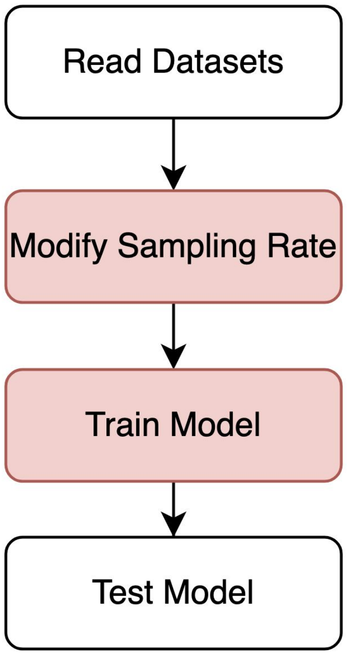


Summary of Datasets

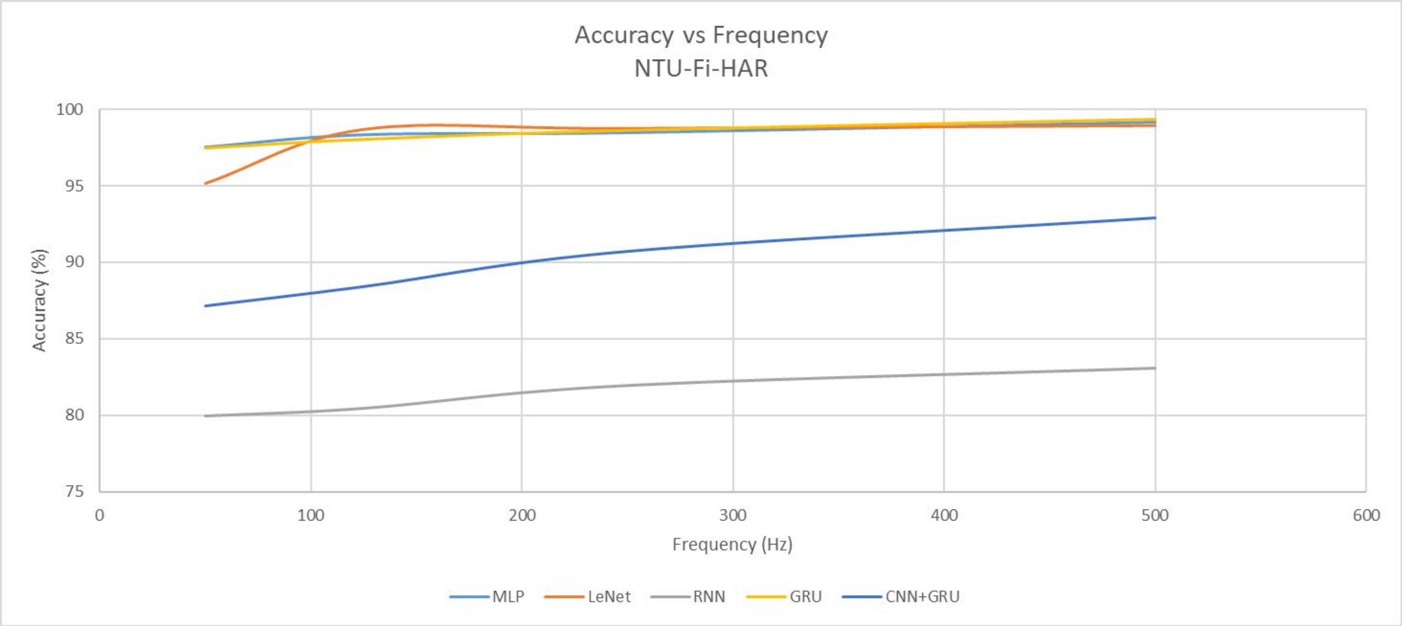
Datasets	Activities	Environment	Number of subjects/samples	BW
NTU-Fi-HAR	6: running, walking, falling, boxing, circling arms, cleaning floor	Lab	20 people, each activity 20 times	40 MHz
NTU-Fi-HumanID	15 people's gait	Lab, 3 scenario	15 people	40 MHz
UT-HAR	6: lie down, fall, walk, run, sit down, stand up	Office	6 people, 20 trials per activity Data collected continuously	20 MHz
Widar	22: Push, Pull,Sweep,Clap,Slide, 18 types of Draws	3 environments: classroom, hall, office	16 volunteers	20 MHz
SignFi	276 signed gestures	Lab, home	5 people	20 MHz

Technical Approach (cont.)

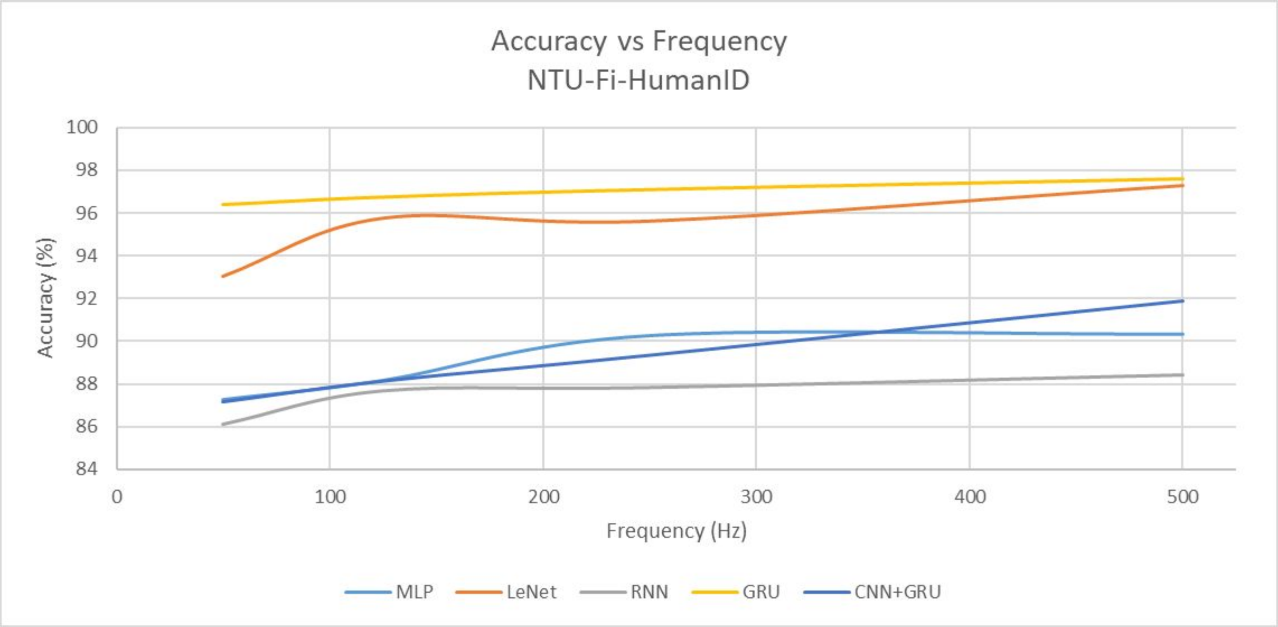
- Different methods for downsampling
  - Reducing the matrix
    - I.e. taking every second column for half the sampling frequency, every 4th for ¼ the sampling etc.
    - x=x[:,::2];
  - Using decimate function in Python
    - x = decimate(x, q=8, zero_phase=True)
    - x=x.copy() # needed to avoid any negative stride value
  - Using my own functions
- Different areas for downsampling
  - Pre vs Post converting the data to tensor format
- Added code to develop a confusion matrix to further analyze the data
- Began changing the num of subcarriers
  - Essentially changing the BW



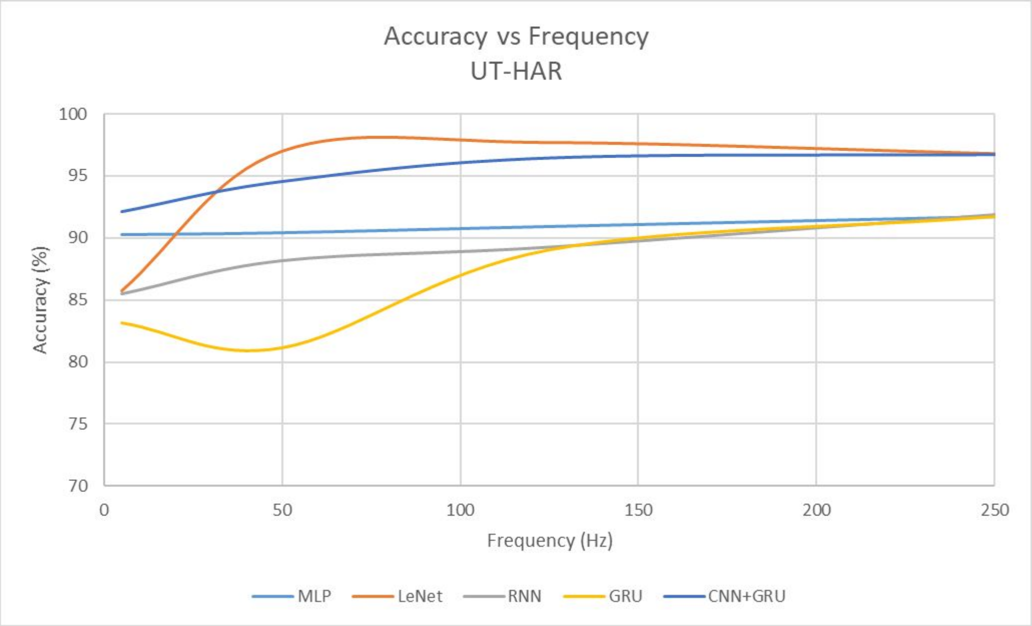
Results - NTU-Fi-HAR



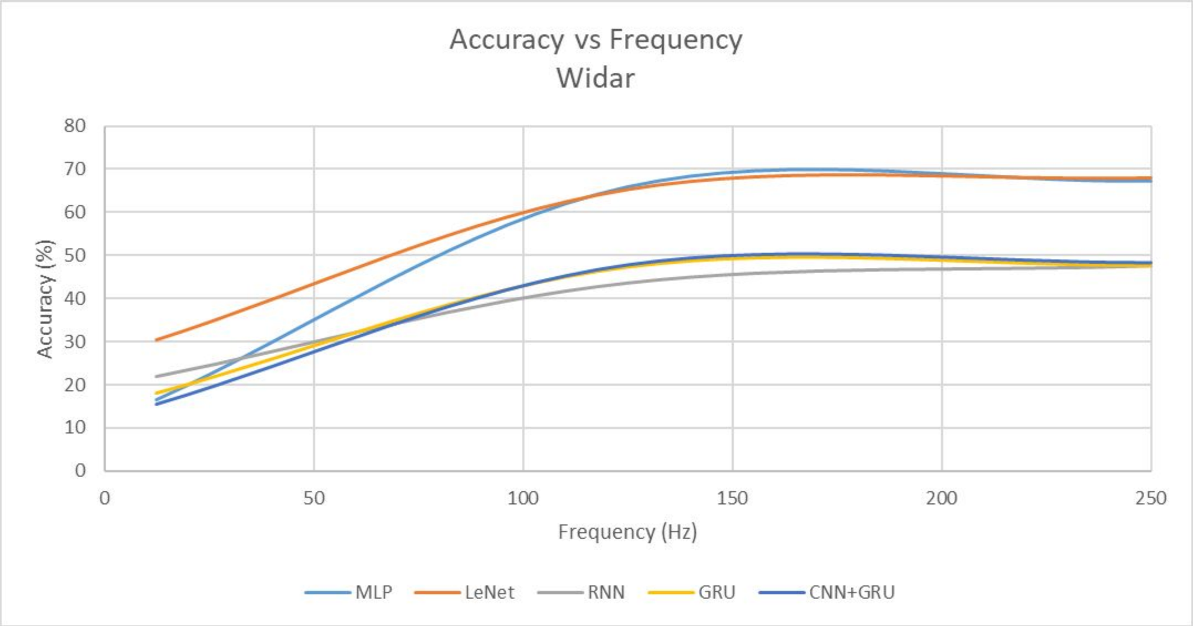
Results - NTU-Fi_HumanID (cont.)



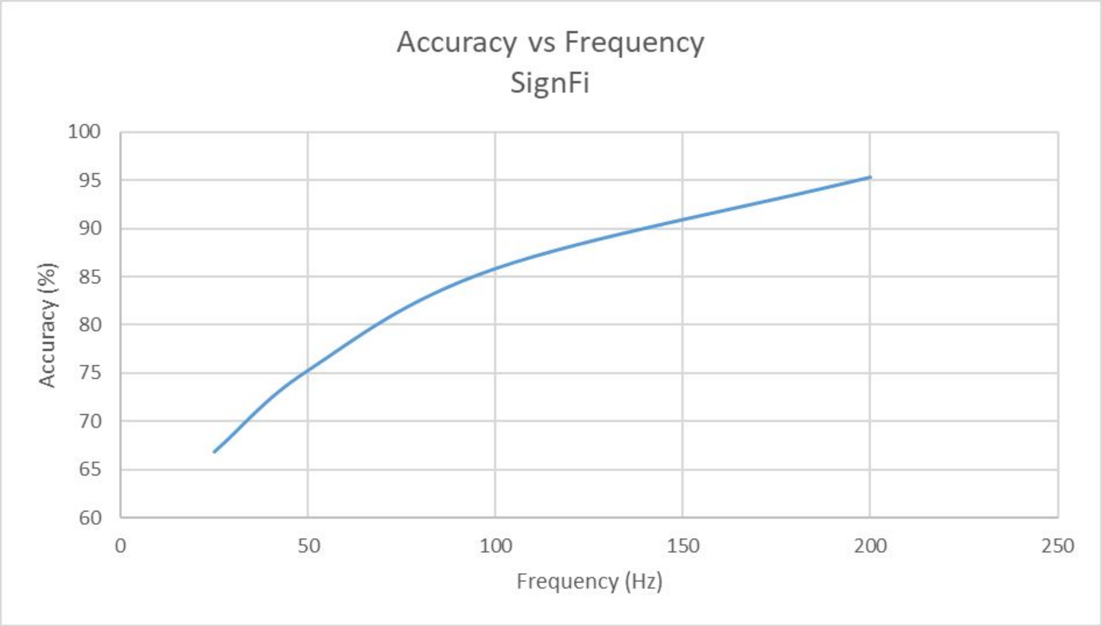
Results -UT-HAR (cont.)



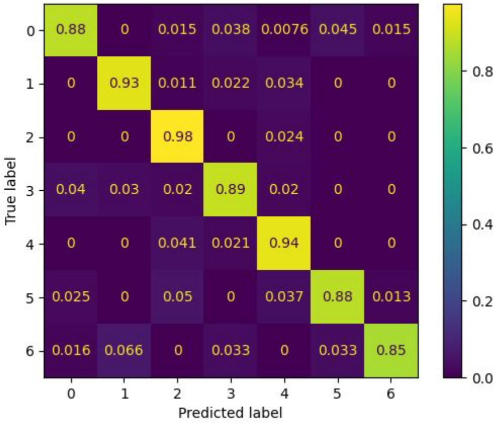
Results -Widar (cont.)



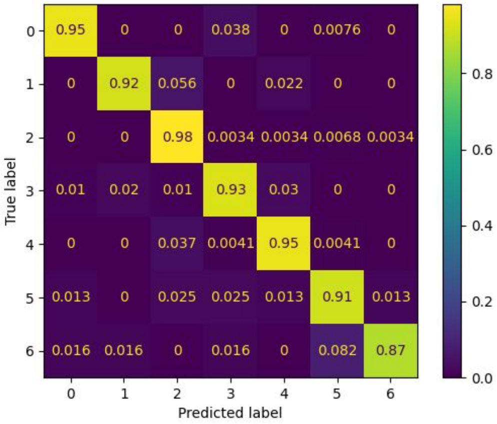
Results - SignFi (cont.)



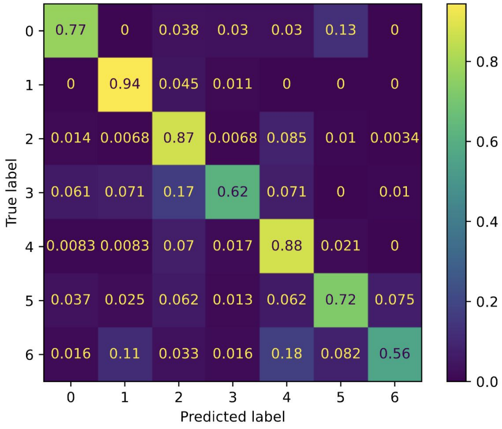
Results: Confusion Matrix UT-HAR GRU



250 Hz



125 Hz

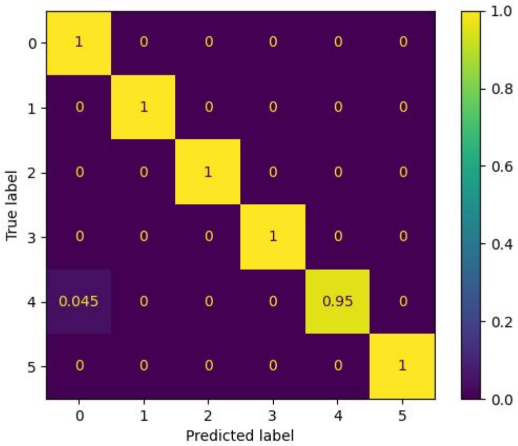


50 Hz

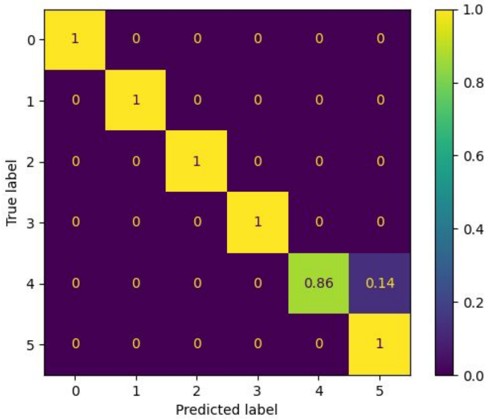
Following is the mapping:

0: Lie down
1: Fall
2: Walk
3: Pick up
4: Run
5: Sit down
6: Stand up

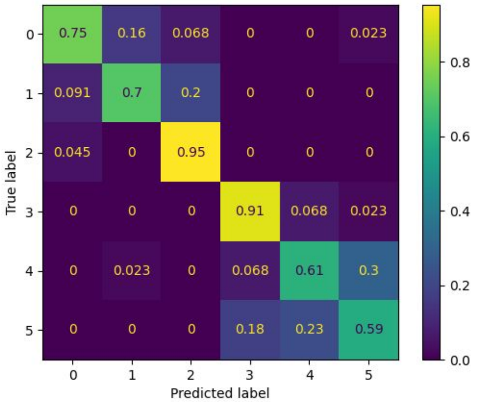
Results: Confusion Matrix NTU-Fi-HAR GRU



500 Hz



250 Hz

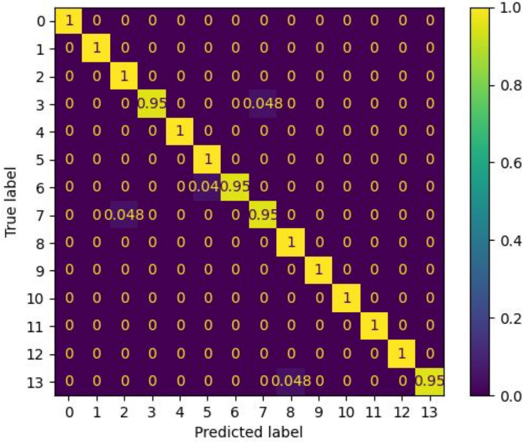


20 Hz

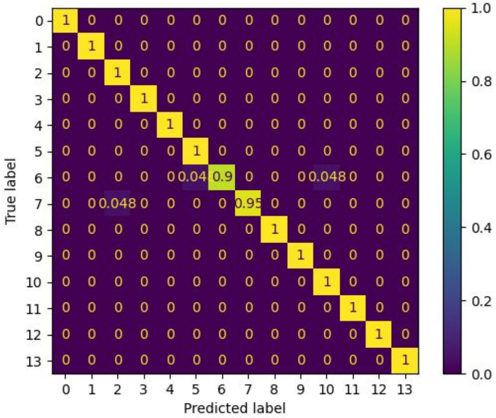
Following is the mapping:

0: Box	3: Fall
1: Circle	4: Walk
2: Clean	5: Run

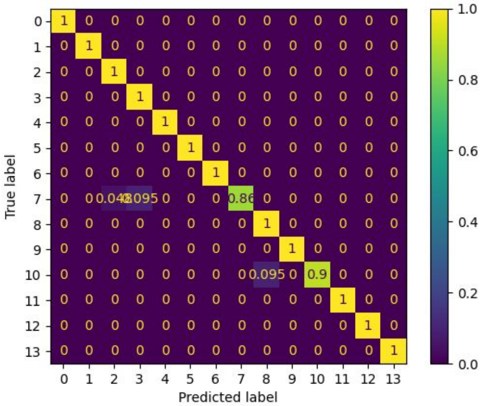
Results: Confusion Matrix NTU-Fi_HumanID GRU



500 Hz



250 Hz



50 Hz

Each index represents the gait of each individual person

Conclusion

From our results we can see the following when varying the sampling frequency:

- NTU-Fi-HAR saw very little change in accuracy because the limited amount of activities and all but 2 being distinct
- NTU-Fi-HumanID saw very little change as well because it measures the gait of 15 distinct people, which can be unique enough to provide accurate classification
- UT-HAR saw more of a change in accuracy because the activities are more similar
- Widar and SignFi saw the biggest difference because there are 22 and 256 activities being categorized respectively, and they are much more similar

Conclusions (cont.)

- We also noticed the following that can be explored further:
  - The accuracy of the datasets that were pre-processed and included only the amplitude component of the raw CSI data did not change too much when downsampling the frequency
    - This includes NTU-Fi-HAR, NTU-Fi-HumanID, UT_HAR
  - Datasets that contained more than just the amplitude component were affected by downsampling
    - Includes SignFi and Widar
    - SignFi has both amplitude and phase
    - Widar is a Body-Coordinated Velocity Profile which has the CSI amplitude and phase components
  - Changing which subcarriers were selected affected accuracy
    - Downsampling here did not change accuracy while keeping carrier count the same

Challenges

- Also both have lack of experience with different models of machine learning so had to understand those better
- Comparing multiple datasets in this context is a challenge due to the sheer number of variables and overall differences between the compositions of the data and the activities they're measuring.
- Finding computational resources took some initial time
  - Ran into memory issues because others were also using it to train their own models for other tasks
- Understanding the various ways datasets were normalized.

Future Work

- Explore downsampling the raw data
  - The datasets used had pre-processed data.
  - What would happen if the downsampling occurred on the raw data?
  - Would the downsampled raw data match the pre-processed downsampled data after being processed?
  - How would the accuracy change?
- Continue exploring varying other variables like subcarrier count
- Explore various models and architectures
- Multi-modal HAR
  - Combining multiple data sets/sensors
- Explore other methods of obtaining CSI
- Exploring which activities require CSI amplitude for proper classification vs which require amplitude and phase

Individual Contributions

- Emmanuel
  - Background research
  - Ran the simulations across different models and datasets
  - Research various methods of downsampling
  - Explored effect of changing BW
  - Slides
- Pooya
  - Set up infrastructure
  - Repository, docs webpage
  - Confusion matrix
    - Code and data collection
  - Slides

References

[1] Wang, Wei, et al. "Understanding and modeling of wifi signal based human activity recognition." MobiCom (2015).

[2] Yang, Jianfei, et al. "EfficientFi: Toward large-scale lightweight WiFi sensing via CSI compression." IEEE Internet of Things Journal (2022).

[3] Guo, Linlin, et al. "Wiar: A public dataset for wifi-based activity recognition." IEEE Access (2019).

[4] Hou, Huawei, et al. "DASECount: Domain-Agnostic Sample-Efficient Wireless Indoor Crowd Counting via Few-Shot Learning." IEEE Internet of Things Journal (2022).

[5] Ma, Yongsen, et al. "SignFi: Sign language recognition using WiFi." ACM IMWUT (2018).

[6] Yang, Chen, et al. "SenseFi: A Library and Benchmark on Deep-Learning-Empowered WiFi Human Sensing", Patterns (2023).

[7] Yousefi, Narui, et al. "A Survey on Behavior Recognition Using WiFi Channel State Information"IEEE Comunication Magazine (2017)

[8] Zheng Yang, Yi Zhang, Guidong Zhang, Yue Zheng, December 26, 2020, "Widar 3.0: WiFi-based Activity Recognition Dataset", IEEE Dataport, doi: https://dx.doi.org/10.21227/7znf-qp86.

[9] M. Cominelli, F. Gringoli and F. Restuccia, "Exposing the CSI: A Systematic Investigation of CSI-based Wi-Fi Sensing Capabilities and Limitations," 2023 IEEE International Conference on Pervasive Computing and Communications (PerCom), Atlanta, GA, USA, 2023, pp. 81-90, doi: 10.1109/PERCOM56429.2023.10099368.

[10] https://wirelesspi.com/advantages-and-disadvantages-of-ofdm-a-summary/

References

[11] From Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, “Gradient-based learning applied to document recognition,”Proceedings of the IEEE, vol.86,no.11,pp.2278–2324,1998.

[12] Sim JM, Lee Y, Kwon O. Acoustic Sensor Based Recognition of Human Activity in Everyday Life for Smart Home Services. International Journal of Distributed Sensor Networks. 2015;11(9). doi:10.1155/2015/679123