# Cross Domain gesture recognition using wifi-signals

*Abstract*— **In my endeavor to address the complexities of human gesture recognition using Wi-Fi-based systems, I am incorporating the dataset developed by the Widar3.0 model. Widar3.0 introduces a novel methodology, extracting domain-independent features from lower-level signal data to create a versatile general model. By utilizing the dataset from Widar3.0, I aim to leverage the unique insights and findings embedded in their data to verify and enhance the adaptability and performance of my own machine learning model.**

**This approach involves implementing Widar3.0's dataset and adopting its methodology as a reference point for my own experimentation. As I explore the accuracy of cross-domain recognition without the need for model re-training, Widar3.0's dataset serves as a foundation for benchmarking against state-of-the-art solutions.**

## I. Introduction

Recent advancements in Wi-Fi-based gesture recognition have encountered challenges in achieving cross-domain generalization due to the domain-dependent nature of signal features. Features such as amplitude, phase, and Doppler Frequency Shift (DFS), along with their statistics, are influenced by various factors such as locations, orientations, and environments, making them unsuitable for constructing a general model applicable across different domains. The term "domain" here refers to the specific set of conditions, including physical locations, orientations, and environmental settings, that influence the Wi-Fi signals used for gesture recognition.

It is crucial to acknowledge the significance of domain-dependent data as it reflects the unique context in which the Wi-Fi signals are captured. However, existing solutions, including those incorporating transfer learning and adversarial learning methodologies, or utilizing translation functions like WiAG, still necessitate additional training efforts for each new target domain, making the prospect of a "one-fits-all" model impractical.

In addressing this challenge, a noteworthy contribution comes from the development of a domain-independent feature called body-coordinate velocity profile (BVP) by the authors of Widar3.0[1]. Unlike traditional domain-dependent features, BVP captures the power distribution over different velocities involved in gesture movements. The uniqueness of BVP lies in its theoretical independence from domain-specific information in raw Wi-Fi signals, making it a promising indicator for human gestures that can transcend various domains.

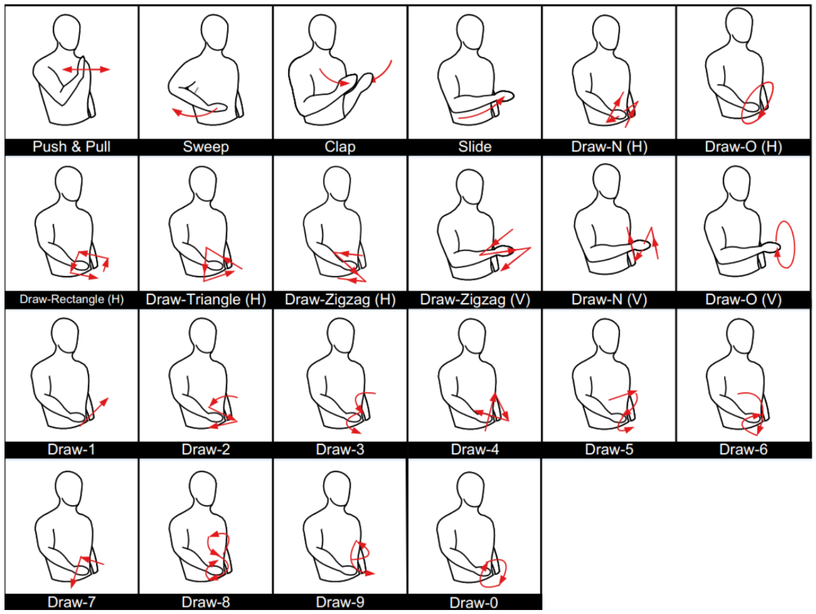
Recognizing the potential of BVP as a domain-independent feature, I am inclined to use it as a foundation for developing a "one-fits-all" model. By leveraging BVP and adopting a learning method that exploits its spatial-temporal characteristics, I believe it is possible to construct a model that achieves cross-domain gesture recognition without the need for additional data collection or model re-training. This approach represents a departure from the conventional reliance on domain-dependent features, offering the prospect of a more generalized and adaptable solution for Wi-Fi-based gesture recognition across diverse domains.

## II. Materials and Methods

The Widar3.0 project has curated an expansive and open-source dataset specifically designed for WiFi-based hand gesture recognition, playing a pivotal role in shaping the trajectory of my current research. This dataset is characterized by the collection of RF data from commodity WiFi NICs, presenting information in the form of Received Signal Strength Indicator (RSSI) and Channel State Information (CSI). Comprising a rich diversity of 258,000 instances of hand gestures, the dataset spans a total duration of 8,620 minutes and encompasses data from 75 distinct domains. Notably, the dataset incorporates two sophisticated features derived from the raw RF signal: Doppler Frequency Shift (DFS) and a novel feature known as Body-coordinate Velocity Profile (BVP). The open-source nature of this comprehensive dataset not only facilitates transparency and reproducibility in research but also fosters collaborative efforts within the scientific community. This resource provides a robust foundation for my investigation into WiFi-based hand gesture recognition, leveraging its diverse and well-annotated content to enhance the accuracy and generalization of my machine learning models. Link to Dataset

### A. Dataset

Gesture data is gathered from five distinct locations and five orientations within each sensing area. Two datasets are compiled, each serving specific purposes. Dataset 1 encompasses common hand gestures utilized in human-computer interaction, such as pushing and pulling, sweeping, clapping, sliding, drawing a circle, and drawing zigzag. Dataset 1 comprises gesture samples from 16 users across five positions, five orientations, six gestures, and five instances each. Notably, eight users contribute to 6,000 data samples collected in the classroom, five users contribute to 3,750 data samples collected in the hall, and four users contribute to 3,000 data samples collected in the office. One user has data in both the hall and office, while the remaining users have data in only one room. Moving on to Dataset 2, it is specifically curated for a case study involving more intricate and semantically rich gestures. In this dataset, two volunteers (one male and one female) draw numbers 0 to 9 in the




(a)

Fig. 1: (a) orientations

horizontal plane, resulting in a total of 5,000 samples (2 users × 5 positions × 5 orientations × 10 gestures × 10 instances).

### B. Data Recognisation Mechanism

The selection of the Deep Neural Network (DNN) architecture in our study is grounded in a comprehensive literature review across diverse sources. Our investigation revealed that the combination of Convolutional Neural Network (CNN) for spatial feature extraction and Recurrent Neural Network (RNN) for temporal modeling has demonstrated promising performance in related studies. However, the primary motivation behind opting for this specific model is twofold. Firstly, it aligns with existing findings suggesting its potential superiority in capturing complex spatial-temporal characteristics for hand gesture recognition. Secondly, our aim extends beyond merely validating its effectiveness; we seek to scrutinize if there are opportunities for improvement or if alternative models can yield superior results. Thus, the chosen DNN serves as a starting point, with the overarching goal of not only assessing its performance but also exploring avenues for innovation and enhancement in the realm of hand gesture recognition methodologies.

## III. Cross-Domain Evaluation

In the evaluation of Widar3.0, the overall performance was assessed across various domain factors, including environmental conditions, person diversity, and the location and orientation of individuals. To conduct a comprehensive evaluation, leave-one-out cross-validation was employed on the datasets, holding the other domain factors constant. Additionally, scenarios were explored where multiple domain factors changed simultaneously, simulating real-world deployment conditions for Widar3.0. Furthermore, the system's performance was examined when users wore different outfits on different dates. One critical aspect of the evaluation focused on orientation sensitivity. This experiment involved selecting each orientation as the target domain while considering the other four orientations as the source domain. In the assessment, orientation 2 exhibited the highest performance, while orientation 5 demonstrated a decline in accuracy. This phenomenon can be attributed to gestures being potentially shadowed by the human body in these orientations, leading to a decrease in the number of effective wireless links for Body-coordinate Velocity Profile (BVP) generation.

## IV. Reuslts

The evaluation results encompass a detailed examination of orientation sensitivity through the utilization of a confusion matrix. This matrix was derived from a meticulous leave-one-out cross-validation procedure, where each orientation—namely, 1, 3, 4, and 5—was systematically treated as the target domain, while the remaining orientations served as the source domain. The assessments revealed distinctive performance characteristics for each orientation during the Widar3.0 dataset evaluation.

Specifically, orientation 2 demonstrated superior performance, showcasing the highest accuracy among the orientations considered. Conversely, orientation 5 exhibited a decrease in accuracy compared to the overall average for the orientation domain, hinting at potential challenges or limitations associated with this particular orientation. The overall accuracy for the orientation domain consistently remained at 80% , providing a comprehensive understanding of Widar3.0's dataset performance across different orientations in the context of leave-one-out cross-validation. The confusion matrix offers granular insights into the system's ability to accurately recognize and distinguish gestures based

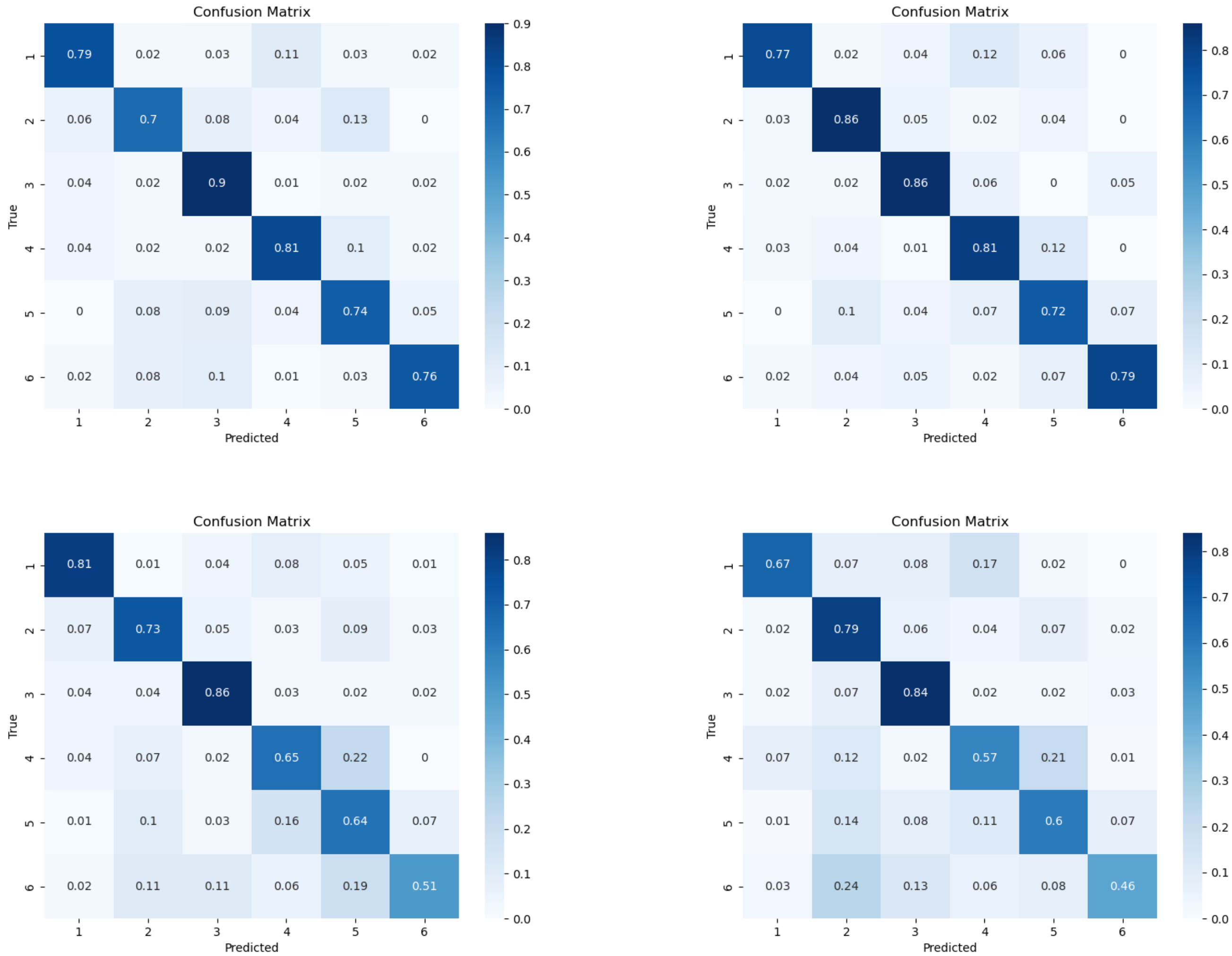



Fig. 2: (a) oriantaition 1; (b) orinatation 3; (c) orinatation 4; (d) orinatation 5 - anticlockwise respectively

on varying user orientations.

## References

[1] 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 Xplore*, 2021.