CERIAS

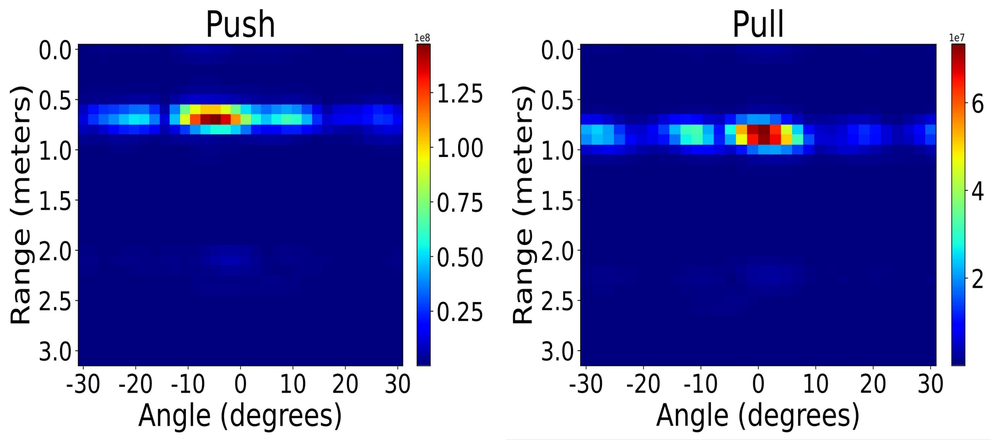
The Center for Education and Research in Information Assurance and Security

Securing Contrastive mmWave-based Human Activity Recognition against Adversarial Label Flipping

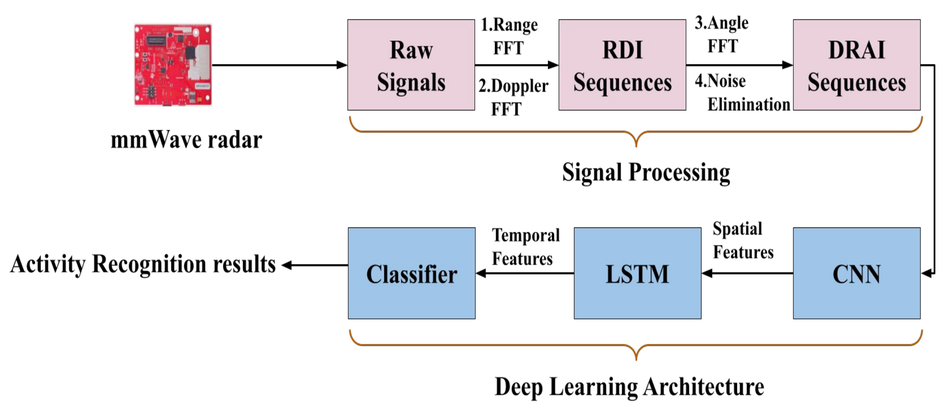
Amit Singha, Ziqian Bi, Tao Li (Purdue University) {singha3, bi32, li4270}@purdue.edu
Yimin Chen (University of Massachusetts Lowell) ian_chen@uml.edu
Yanchao Zhang (Arizona State University) yczhang@asu.edu

The objective of this research is to identify vulnerabilities within mmWave-based Human Activity Recognition (HAR) systems to adversarial label poisoning attacks under supervised contrastive learning (SCL) frameworks. We identify three types of label poisoning attacks on contrastive mmWave-based HAR systems and propose a corresponding defense termed selective supervised contrastive learning (Sel-CL).

Heatmaps of two activities



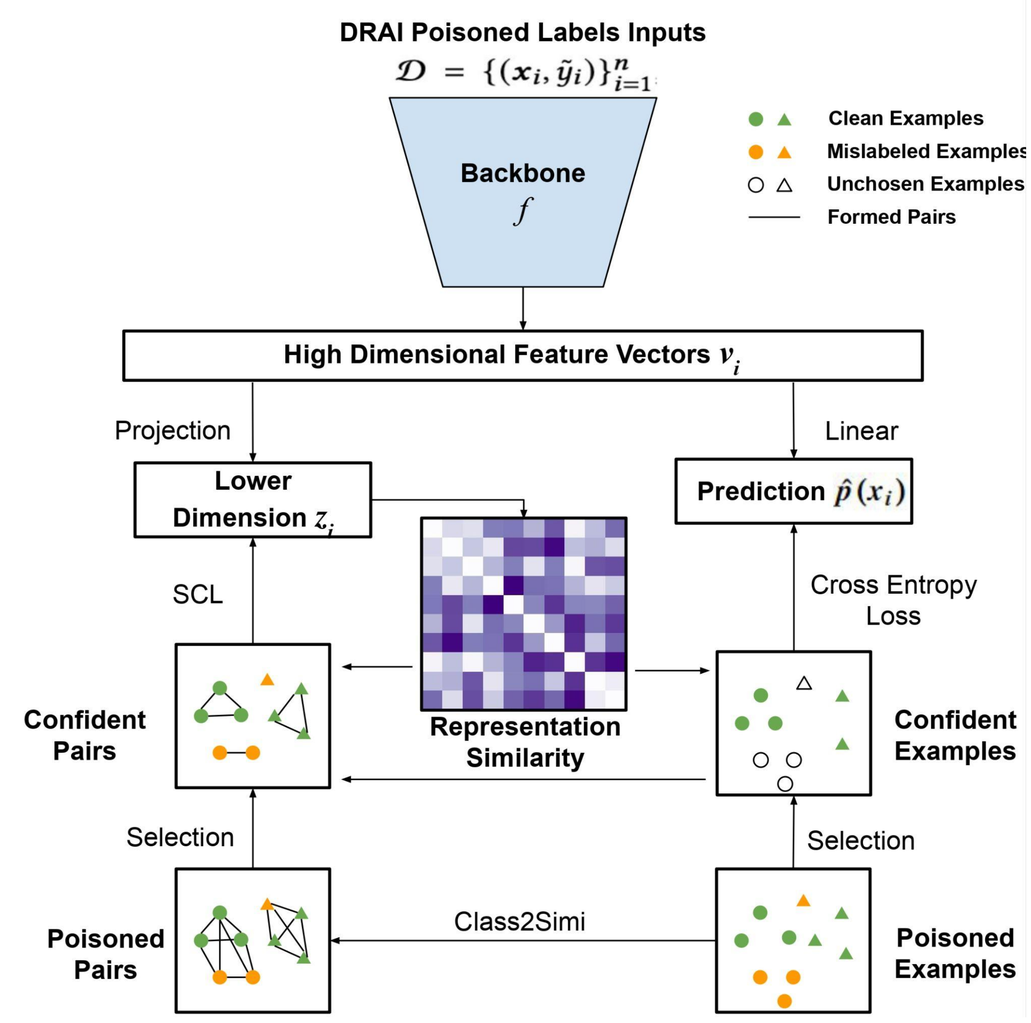
A basic mmWave-based HAR system



Attacks

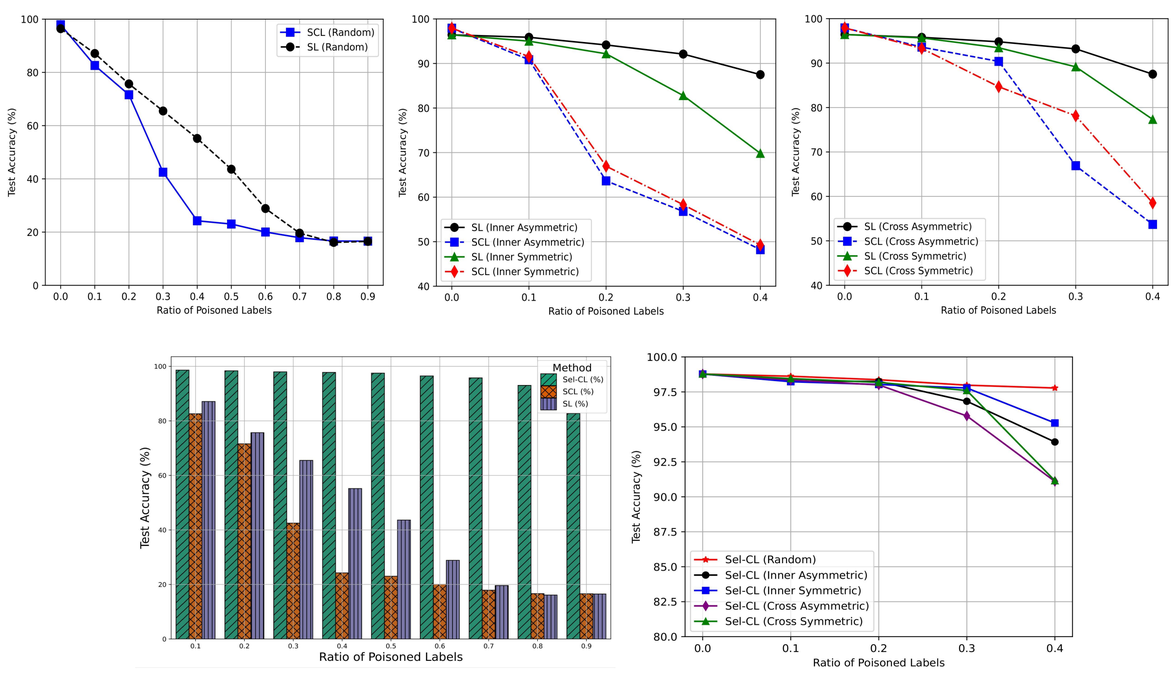
- **Adversary Model:** Attackers aim to degrade HAR system performance by manipulating activity labels in the training dataset. This can be done through intentional data mislabeling, acquiring mislabeled data, or outsourcing training to a malicious third party.
- **Random Attacks:** Randomly alter labels in the training dataset to other arbitrary labels.
- **Across Trajectory Attacks:** Modify labels of activities to those of other activities with different trajectories.
- **Inner Trajectory Attacks:** Involve altering the labels of activities to those of other activities with similar trajectories, aiming to hide malicious tampering.

Defenses



Results

Sel-CL is highly effective against label flipping attacks, outperforming traditional SL and SCL.



PURDUE UNIVERSITY

CERIAS