WiFi-Based Gesture Perception: Enabling Non-Contact Hand Gesture Recognition for IoTs

CIS 5603: Artificial Intelligence | Term Project

Weiyi Jin

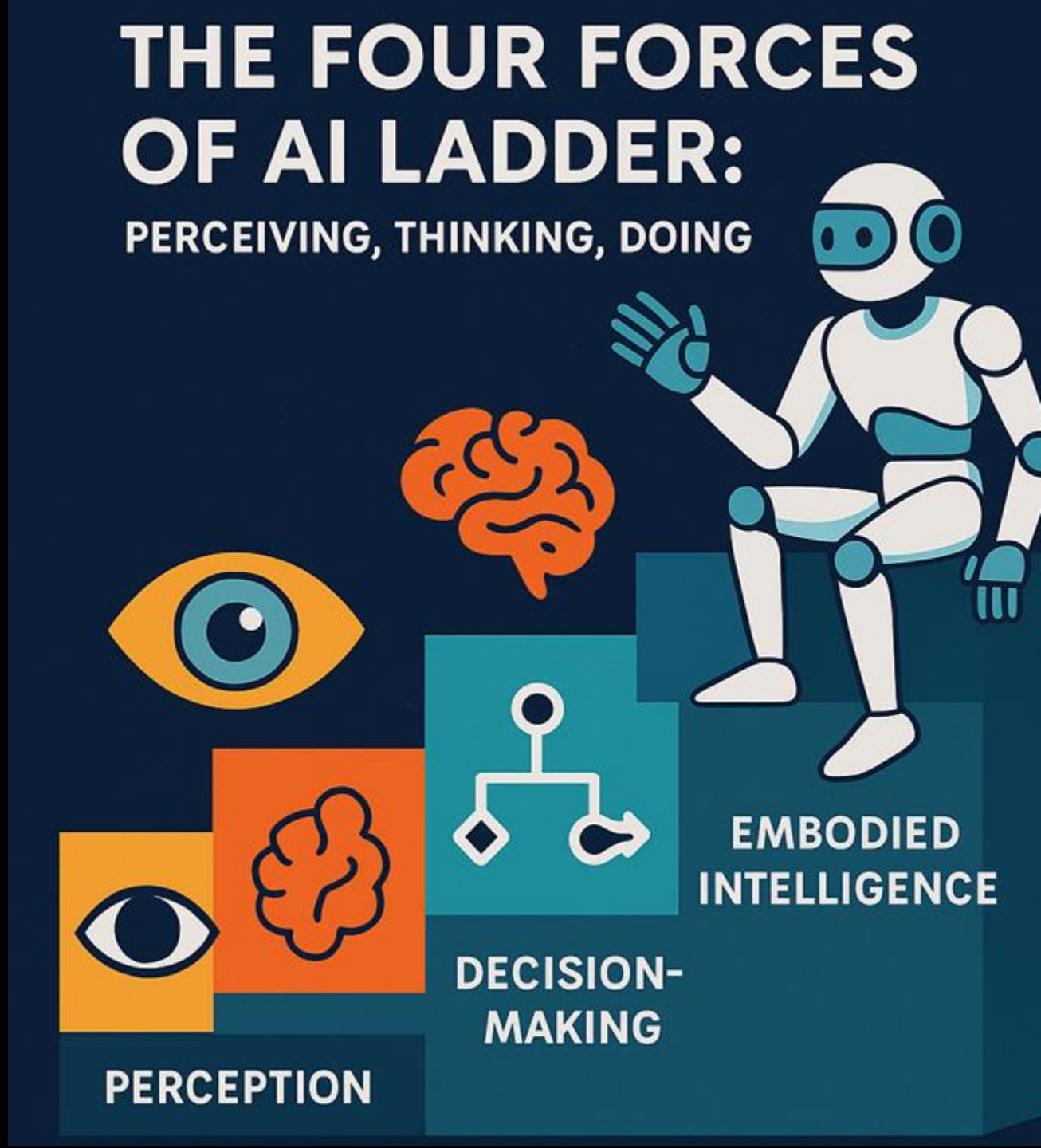
The "Perception" Problem

The Environment

We are operating in a **continuous, dynamic, and noisy** state space. Unlike a camera image (pixels), WiFi signals are invisible and abstract.

Multipath Effect

WiFi signals bounce off walls and objects. A moving hand introduces a **Doppler Shift** and alters these paths. The AI challenge is *Pattern Recognition*: distinguishing the specific signal distortion caused by a hand from background environmental noise.



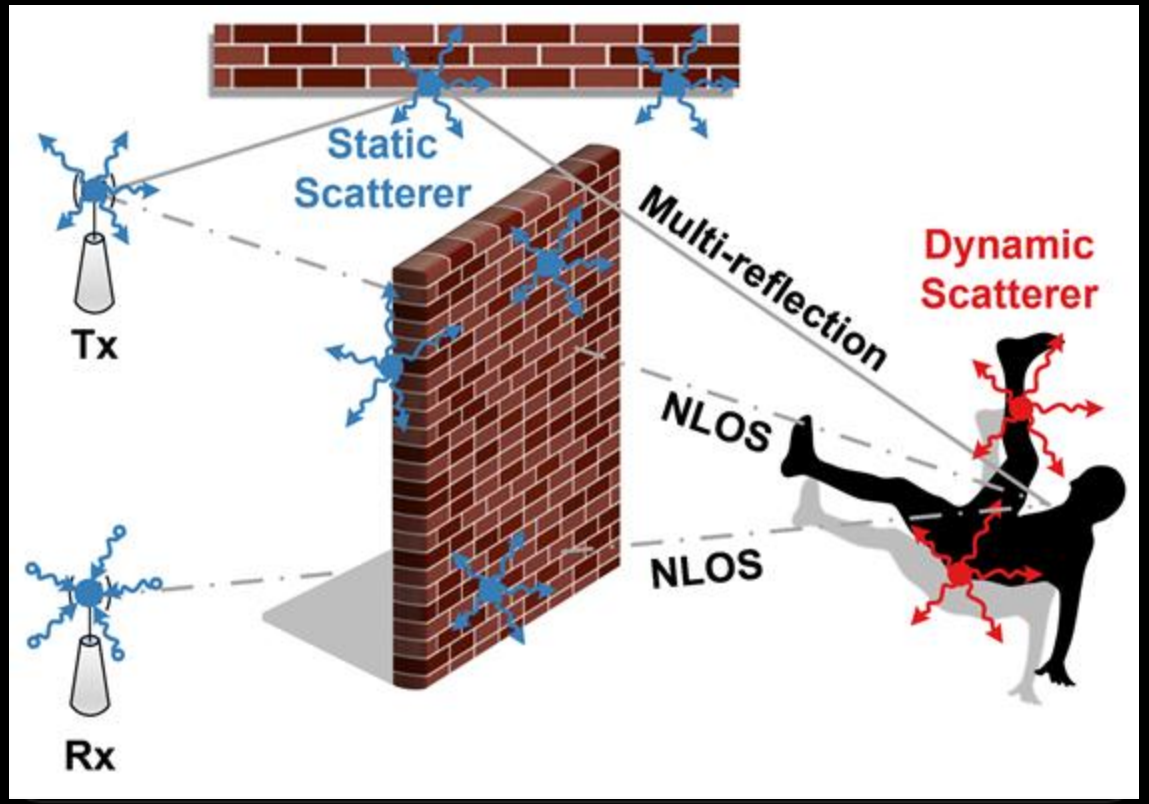
Input Representation: CSI

Channel State Information (CSI)

CSI reveals the channel properties of a communication link. It describes how a signal propagates from the transmitter to the receiver and represents the combined effect of, for example, scattering, fading, and power decay with distance.

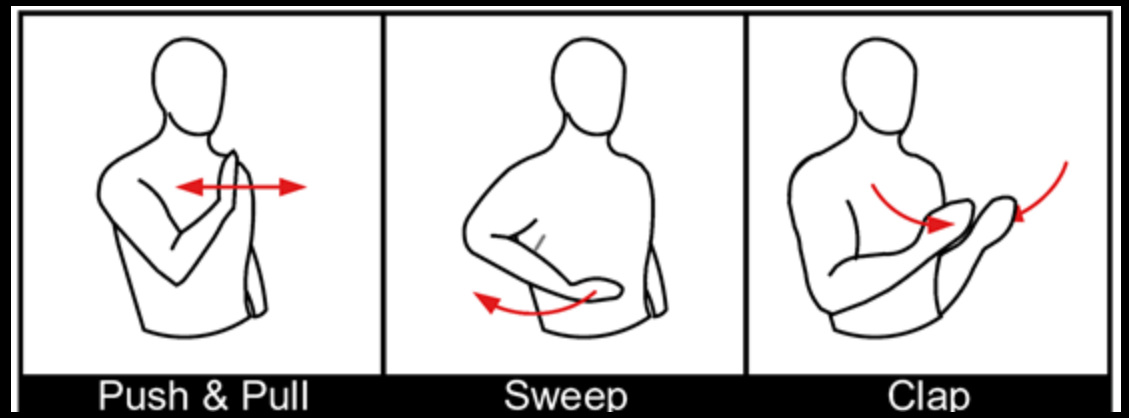
$$H(f,t) = |H(f,t)|e^{j\theta(f,t)}$$

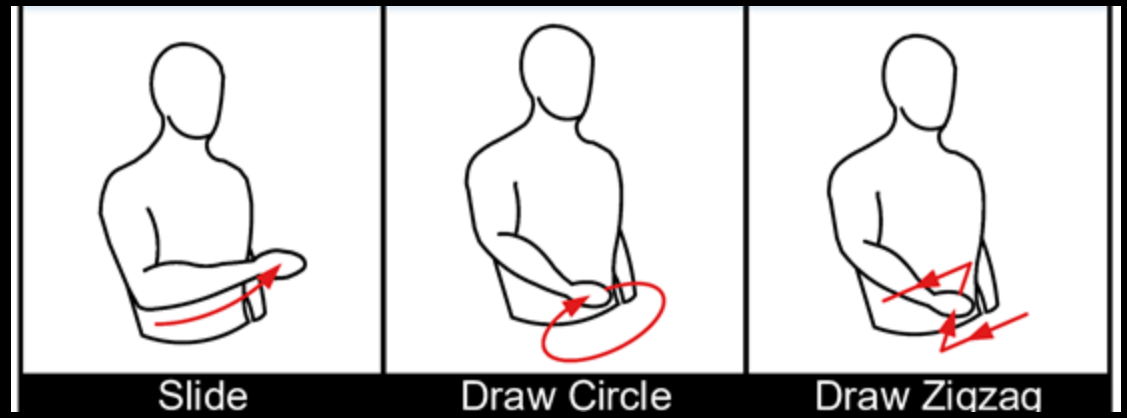
Amplitude and Phase representation



WIDAR 3.0 Dataset

The system learns from the WIDAR 3.0 dataset, focusing on 6 distinct gesture classes (4500 samples).





Learning as Approximation

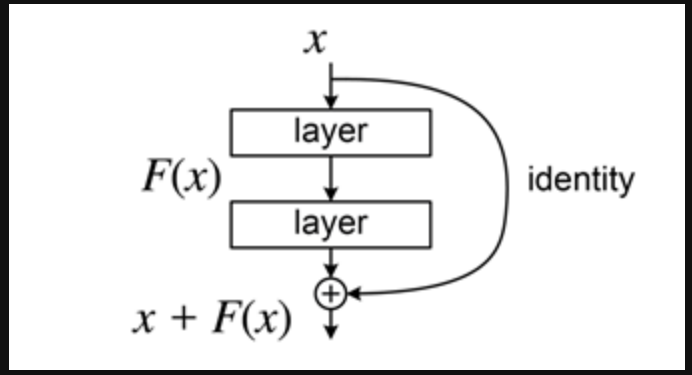
Signal Preprocessing

Raw sensor data is too noisy for direct inference. We apply a **Butterworth Filter** to smooth the time-series data, effectively reducing the stochastic uncertainty before the learning phase.

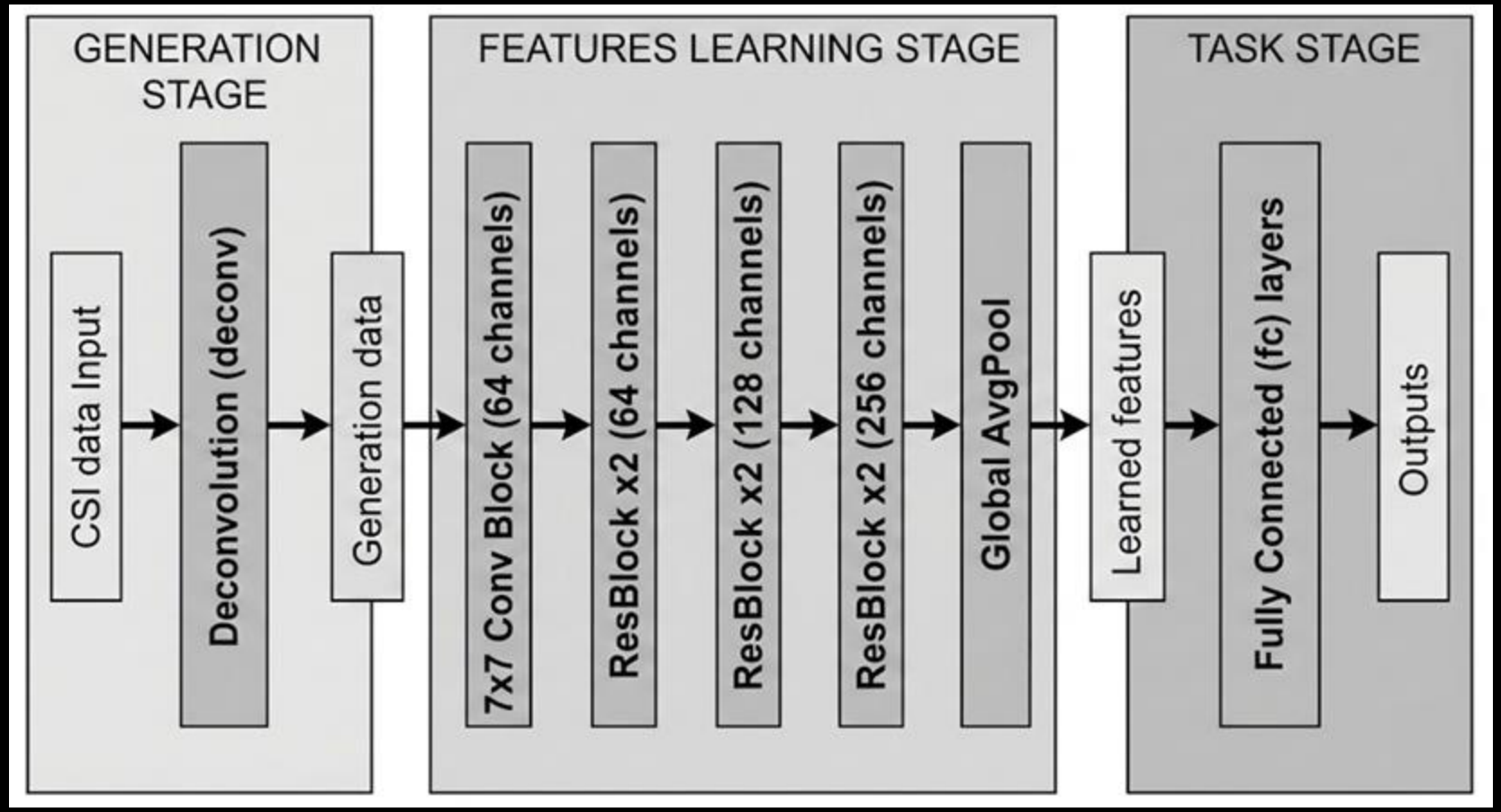
Deep Residual Learning

We use a **ResNet** to approximate the target function. Deep networks often suffer from degradation, but residual skip connections allow the agent to learn identity mappings, enabling deeper feature extraction without vanishing gradients.

[Image of ResNet block diagram]



Model Structure



Training Setup

- **Supervised Learning:** The agent learns from labeled examples (Inductive Learning).
- **Data Split:** 60% Training, 20% Validation, 20% Testing (Evaluating generalization).
- **Input Tensor:** (1500, 30, 3, 1) — 1500 time steps, 30 subcarriers, 3 antennas.
- **Optimization:** Adam Optimizer (Learning Rate: 0.001, Epochs: 50).
- **Loss Function:** Cross-entropy (Measuring divergence between predicted probability and actual class).

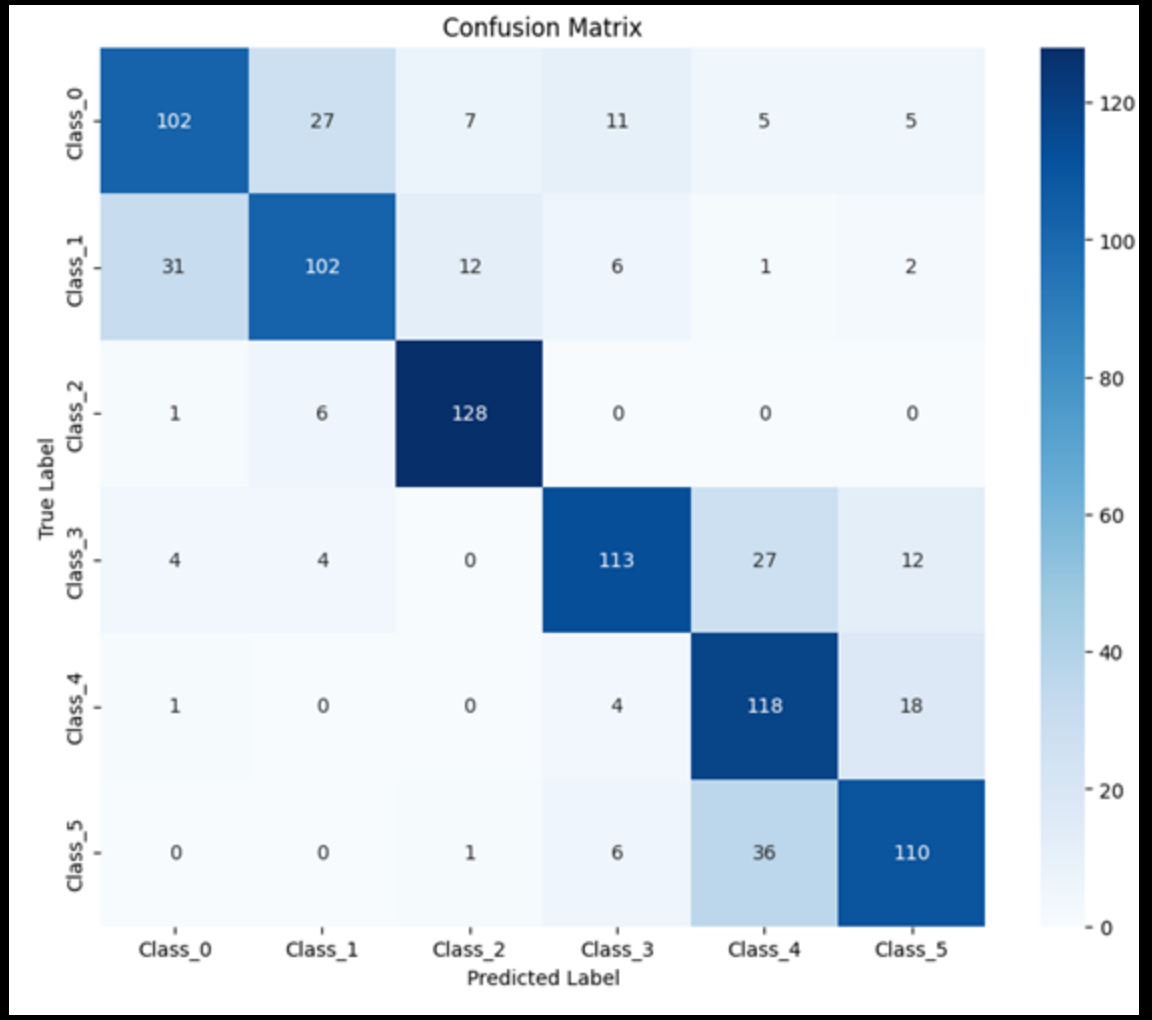
Confusion Matrix Analysis

Ambiguity in Perception

The confusion matrix reveals where the agent's "sensors" failed to distinguish states:

1. Push & Pull vs. Sweep:
Both create similar Doppler shifts (horizontal movement). The feature separation here is non-trivial.

2. Draw-Z vs. Draw-O:
Complex shapes share temporal sub-sequences, confusing the sequential model.



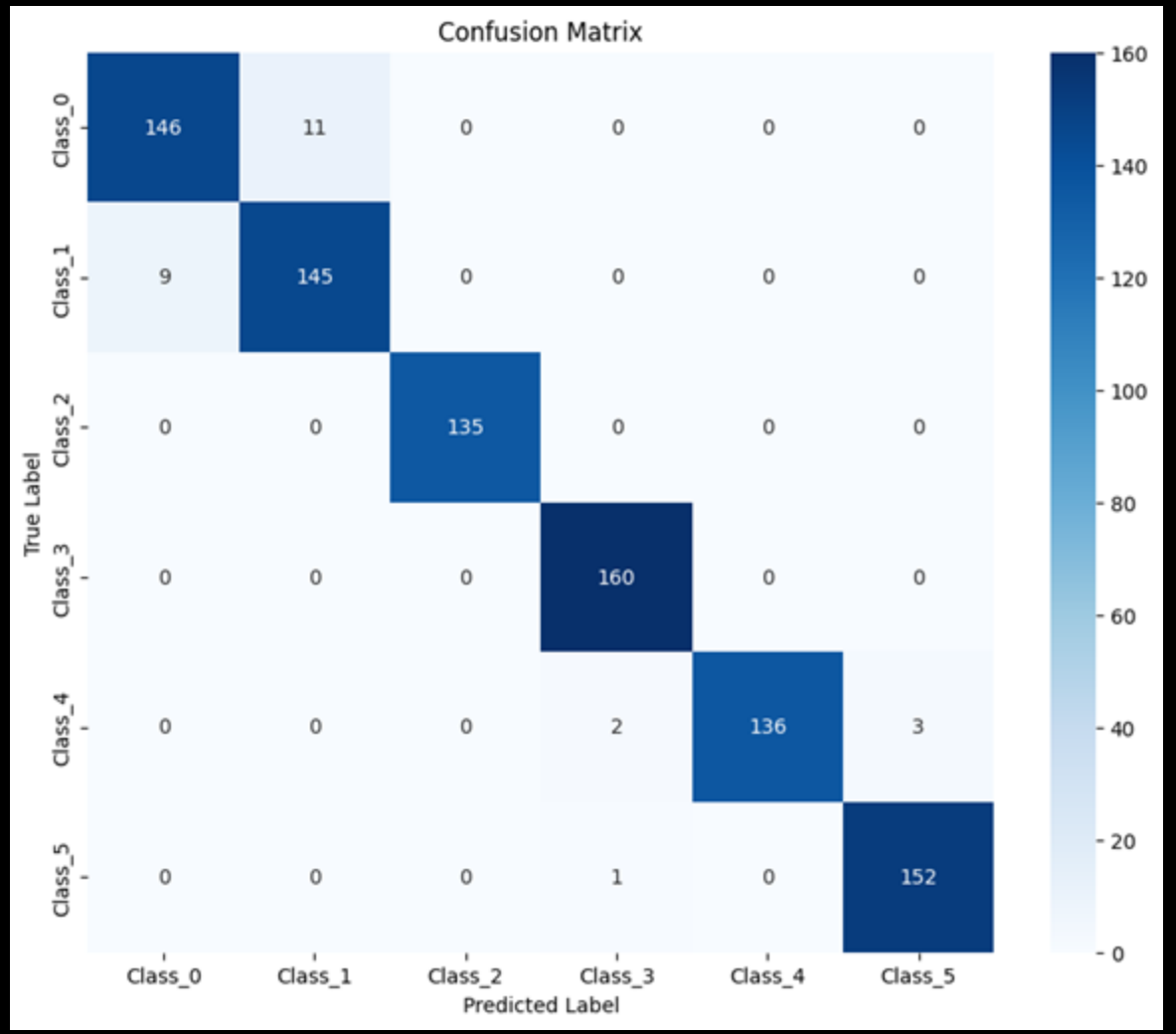
Confusion Matrix Analysis

Ambiguity in Perception

The confusion matrix reveals where the agent's "sensors" failed to distinguish states:

1. Push & Pull vs. Sweep:
Both create similar Doppler shifts (horizontal movement). The feature separation here is non-trivial.

2. Draw-Z vs. Draw-O:
Complex shapes share temporal sub-sequences, confusing the sequential model.



Effect of Preprocessing

Preprocessing (Noise Reduction) is critical for "Insufficient Resources" learning.

Performance Metrics

Gesture Class	Precision	Recall	F1-Score
Push & Pull	0.94	0.93	0.94
Sweep	0.93	0.94	0.94
Clap	1.00	1.00	1.00
Slide	0.98	1.00	0.99
Overall Average	**0.97**	**0.97**	**0.97**

High F1-Score indicates the agent has successfully learned the signal-to-gesture mapping.

Discussion & Perspectives

Limitations (Insufficient Knowledge)

The current agent is trained on a single user. In a true AGI context, this lacks **Generalization**. It does not yet possess the "adaptive" quality defined in class to handle unseen users or new rooms without retraining.

Future Directions

1. **Domain Adaptation:** Transfer learning to adapt to new environments (rooms) with minimal data.

2. **Resource Constraints:** Optimizing the model to run on standard WiFi routers (Edge AI), adhering to the constraint of "insufficient resources."

Thank you !