



Dalle Molle Institute  
for Artificial Intelligence



SUPSI

# Predicting the Intention to Interact with a Service Robot: the Role of Gaze Cues

Simone Arreghini, Gabriele Abbate, Alessandro Giusti and Antonio Paolillo

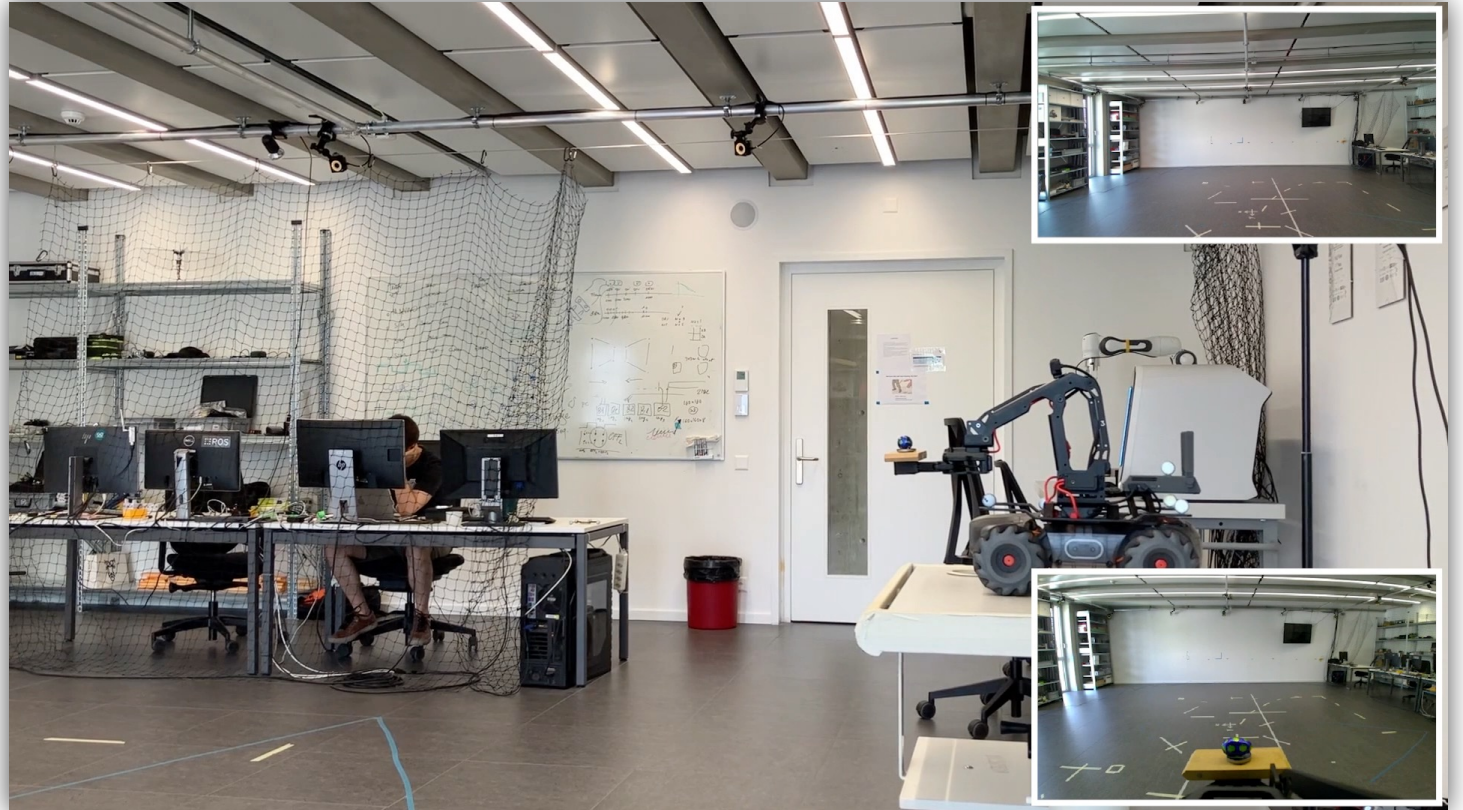
IDSIA USI-SUPSI, Lugano, Switzerland



**ICRA2024**  
YOKOHAMA | JAPAN

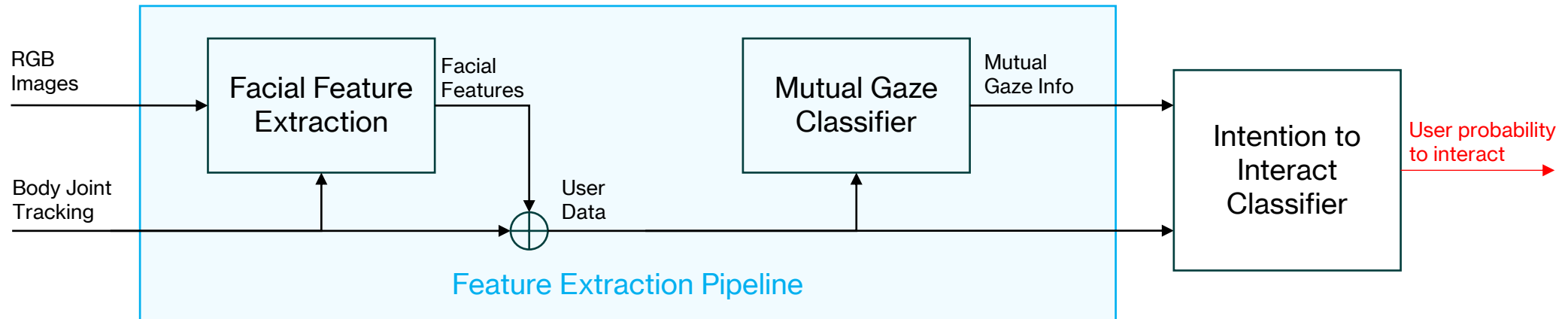
# Why is intention detection useful for HRI?

- More **natural** interactions
- React **only** to **interested** users
- Increase **social acceptance**



# Problem definition and proposed solution

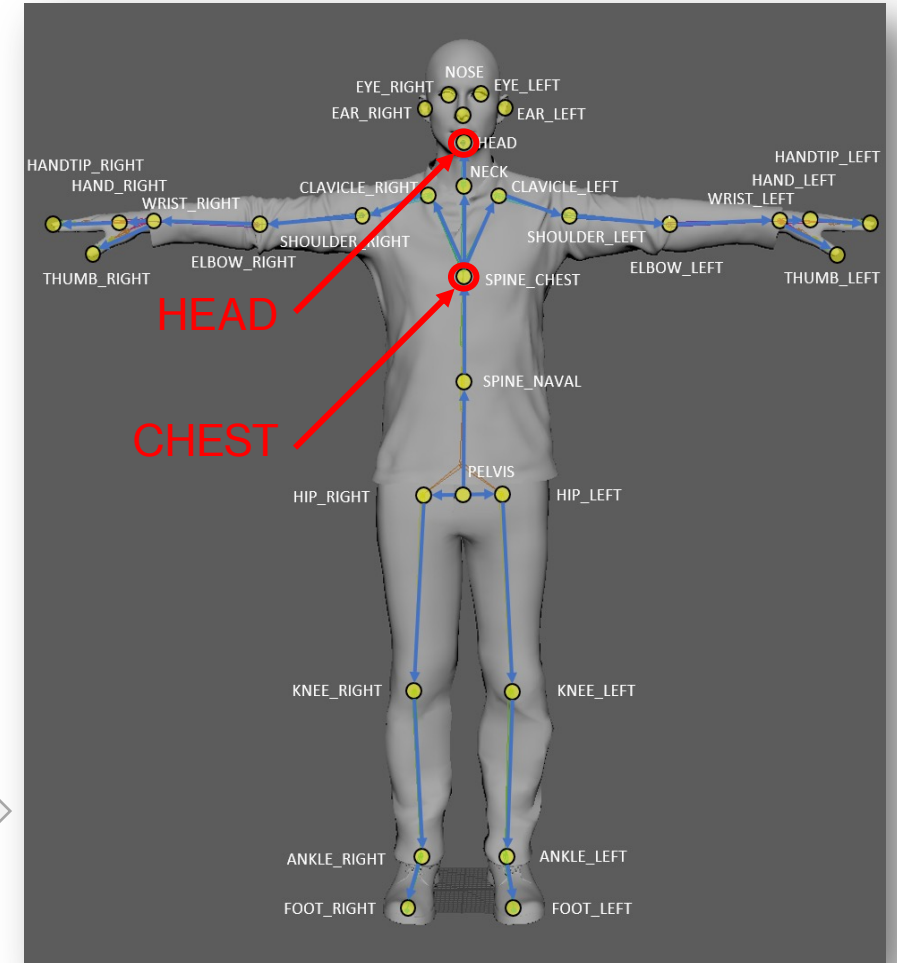
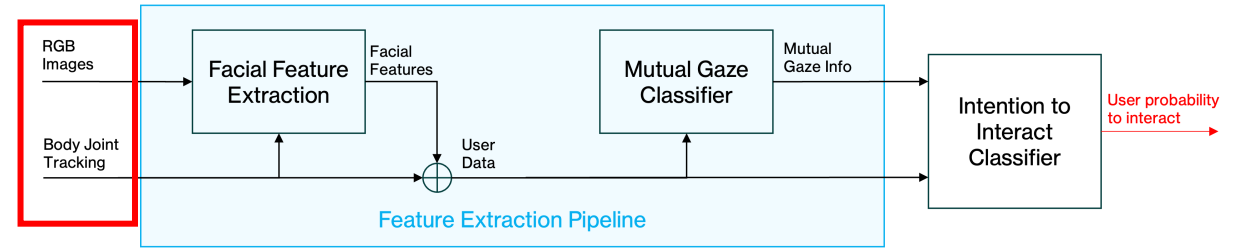
- Long-range prediction of the **intent to interact** using **only non-verbal** cues
- **Include gaze cues and facial landmarks** to enhance performances



# Sensor signals

- Sensor: **Azure Kinect**

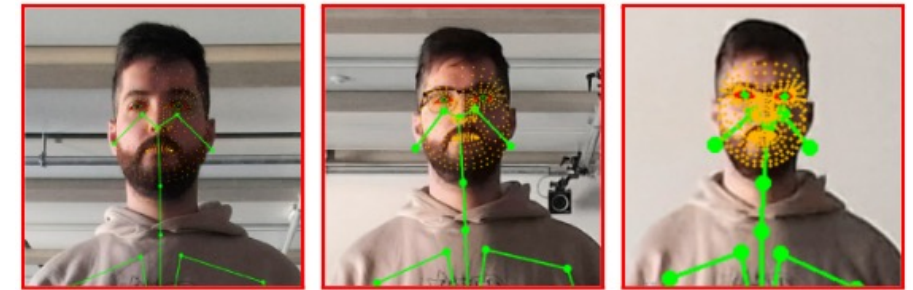
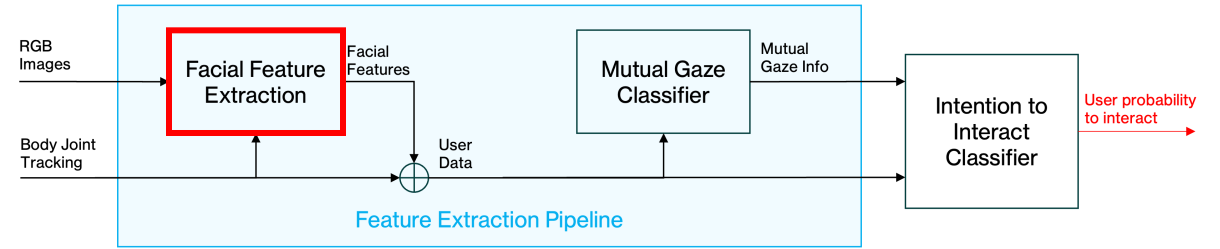
- **High resolution RGB images**
- **Body Tracking (BT)**
  - Tracks up to **32** body joints poses in 3D



# Facial features

- Facial feature: **MediaPipe**

- Track up to **478** facial landmarks
- Get unified **user data** vector containing:
  - Body joints poses
  - Facial landmarks



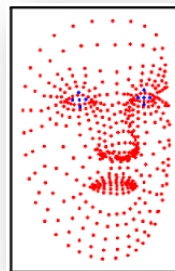
1 m

2.5 m

5 m

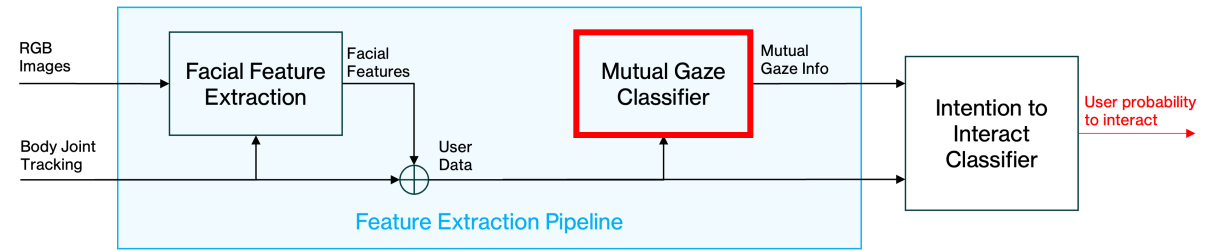


RGB+BT



User Data

# Mutual gaze



- Predict the **probability** that **someone** is **looking at the robot camera**

Only head and eyes movements

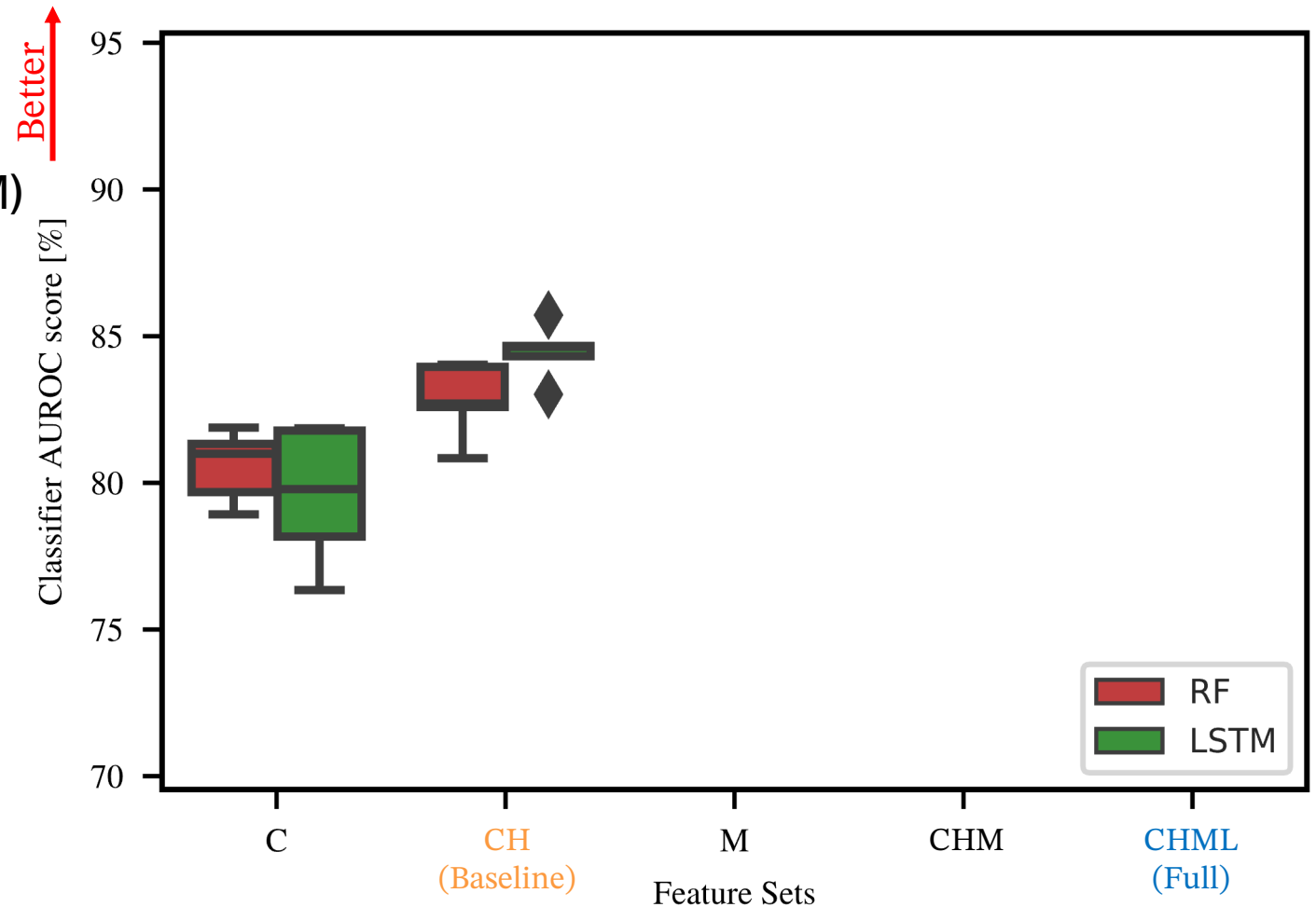
# Stateful model better captures problem dynamics

- **Classifier architectures**

- Stateless Random Forest (RF)
- Stateful Long-Short Term Memory (LSTM)

- **Input features** are composition of:

- **Body joint information**
  - Chest (**C**)
  - Head (**H**)
- Mutual gaze probability (**M**)
- Facial landmarks (**L**)





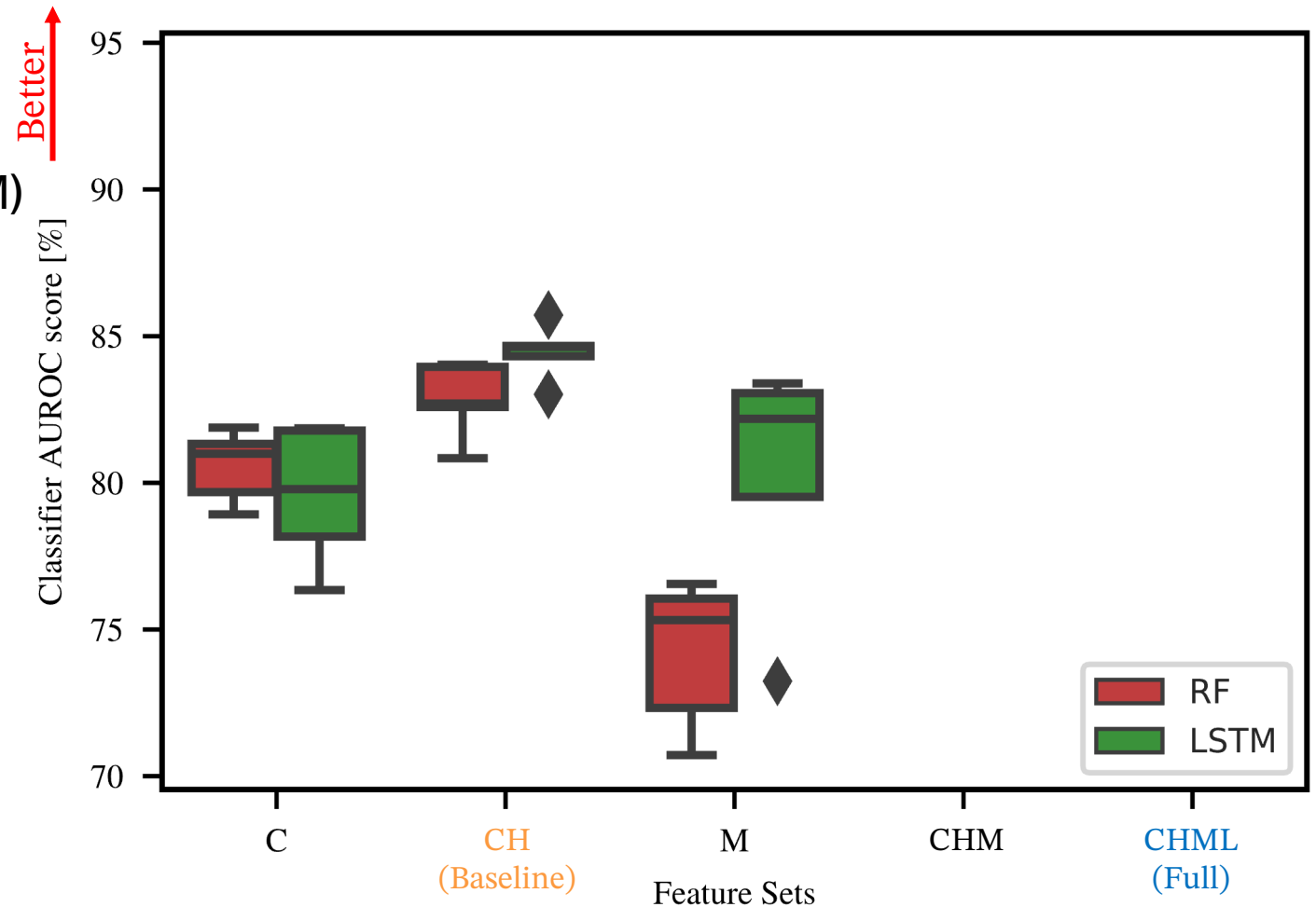
# Stateful model better captures problem dynamics

- **Classifier architectures**

- Stateless Random Forest (RF)
- Stateful Long-Short Term Memory (LSTM)

- **Input features** are composition of:

- Body joint information
  - Chest (**C**)
  - Head (**H**)
- Mutual gaze probability (**M**)
- Facial landmarks (**L**)





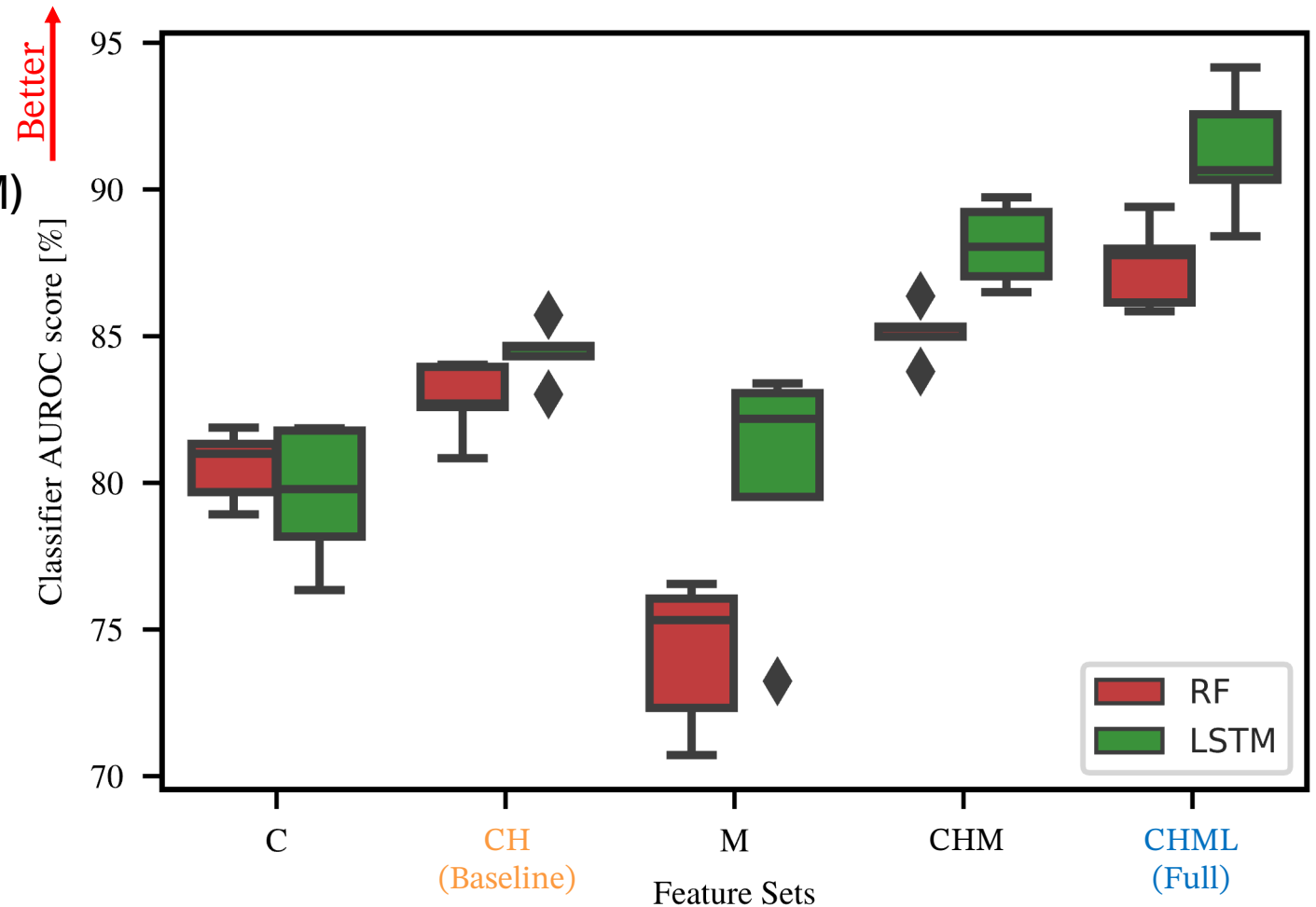
# Stateful model better captures problem dynamics

- **Classifier architectures**

- Stateless Random Forest (RF)
- Stateful Long-Short Term Memory (LSTM)

- **Input features** are composition of:

- Body joint information
  - Chest (**C**)
  - Head (**H**)
- Mutual gaze probability (**M**)
- Facial landmarks (**L**)



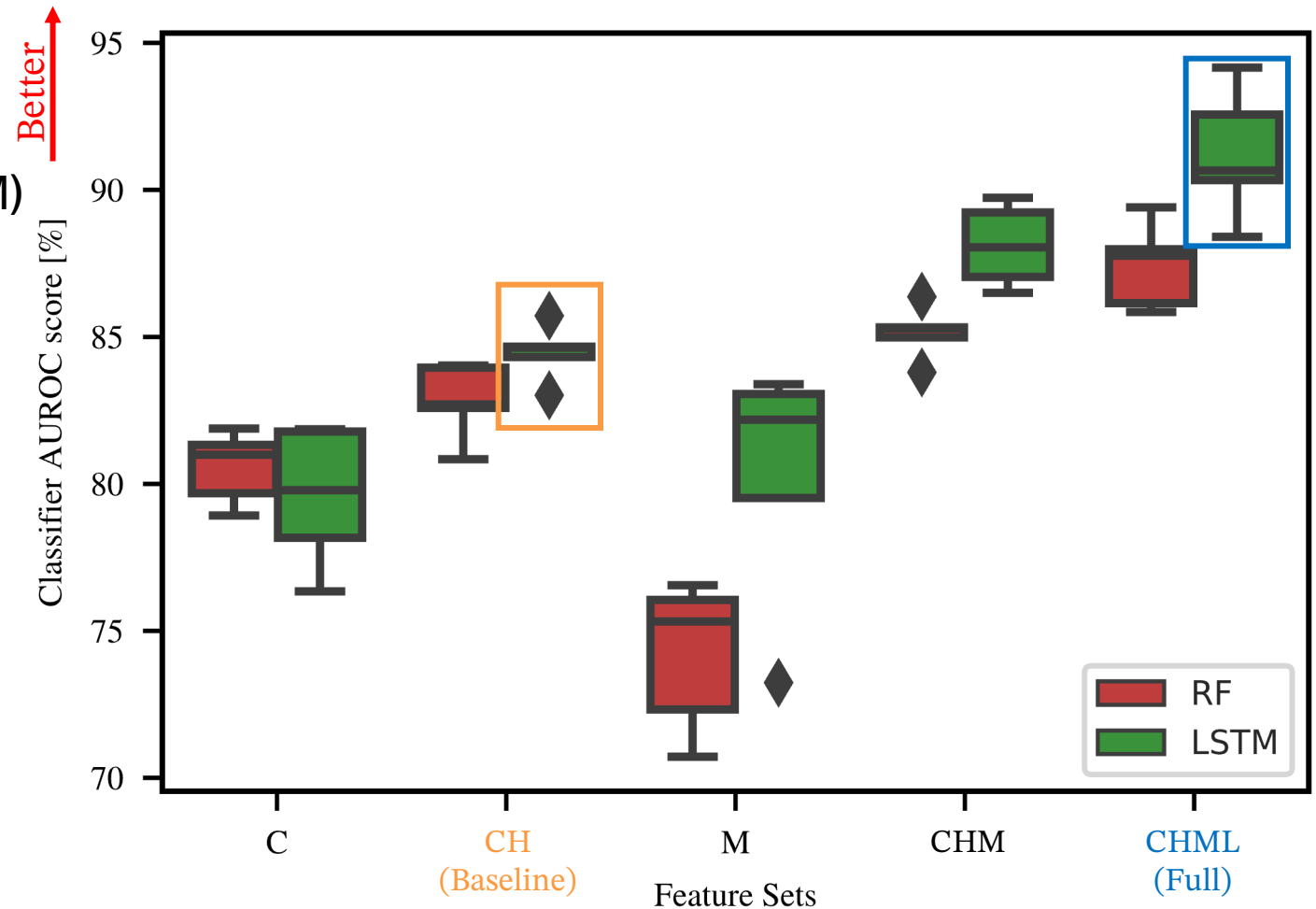
# Stateful model better captures problem dynamics

- **Classifier architectures**

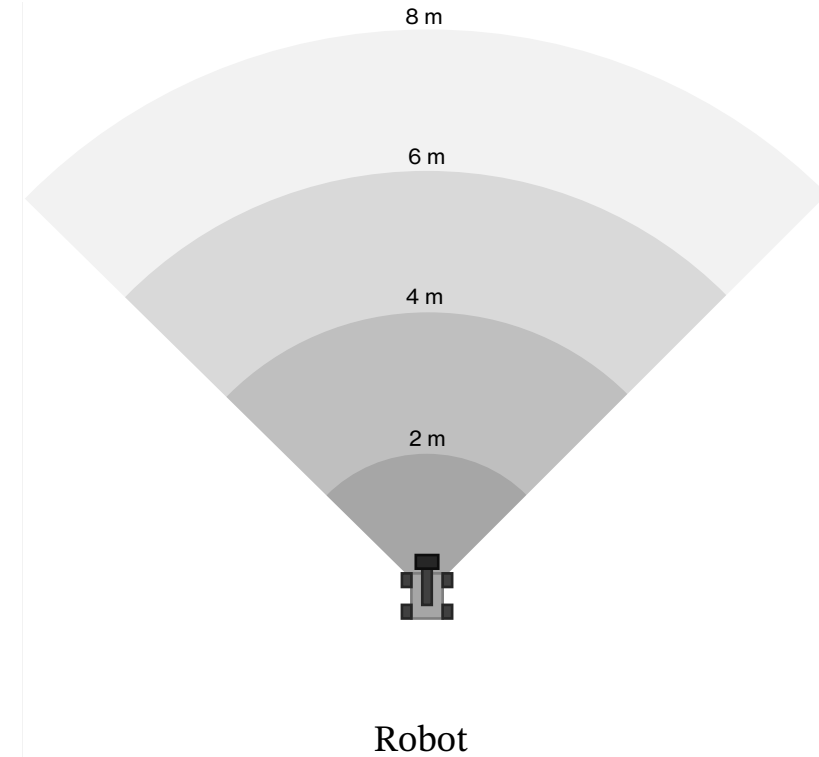
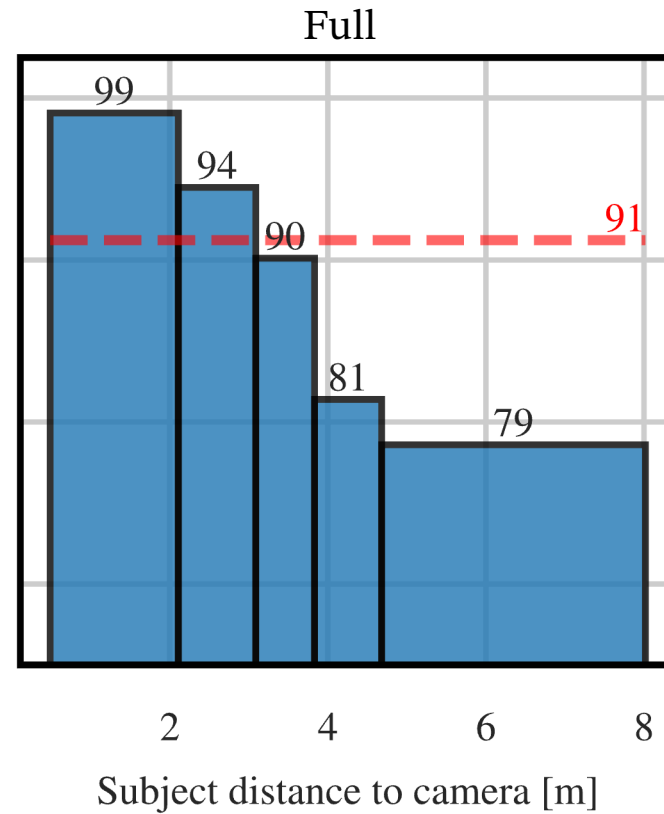
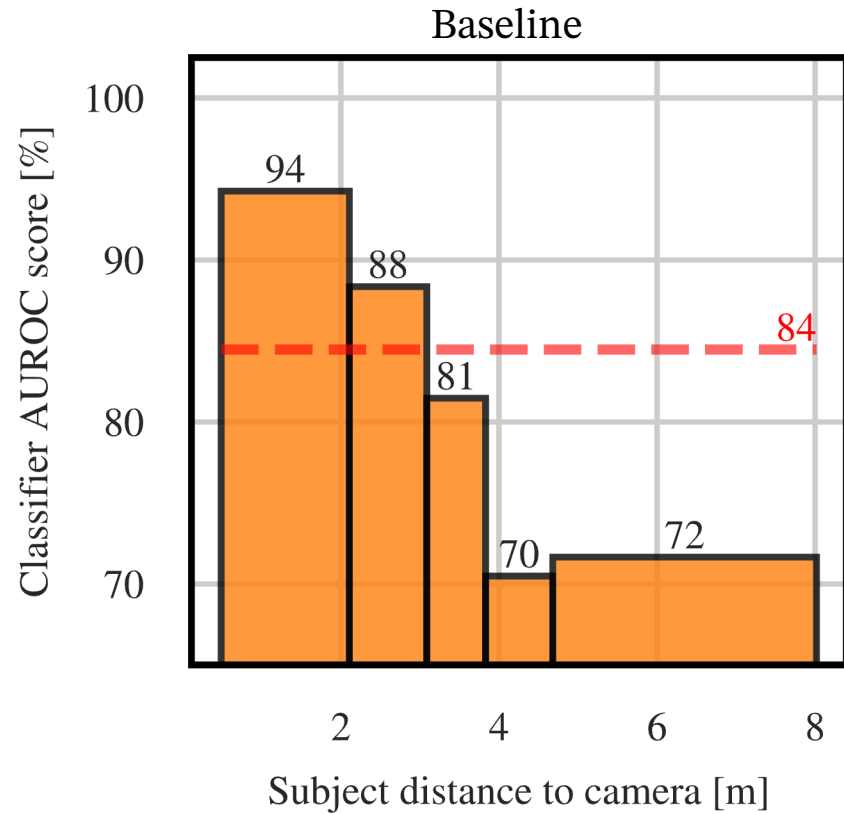
- Stateless Random Forest (RF)
- Stateful Long-Short Term Memory (LSTM)

- **Input features** are composition of:

- Body joint information
  - Chest (**C**)
  - Head (**H**)
- Mutual gaze probability (**M**)
- Facial landmarks (**L**)

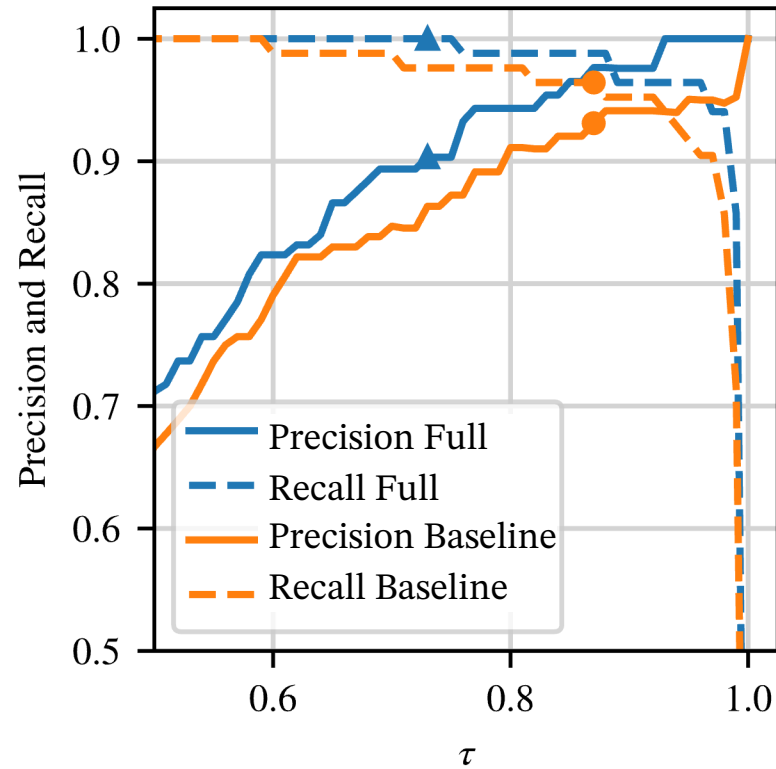
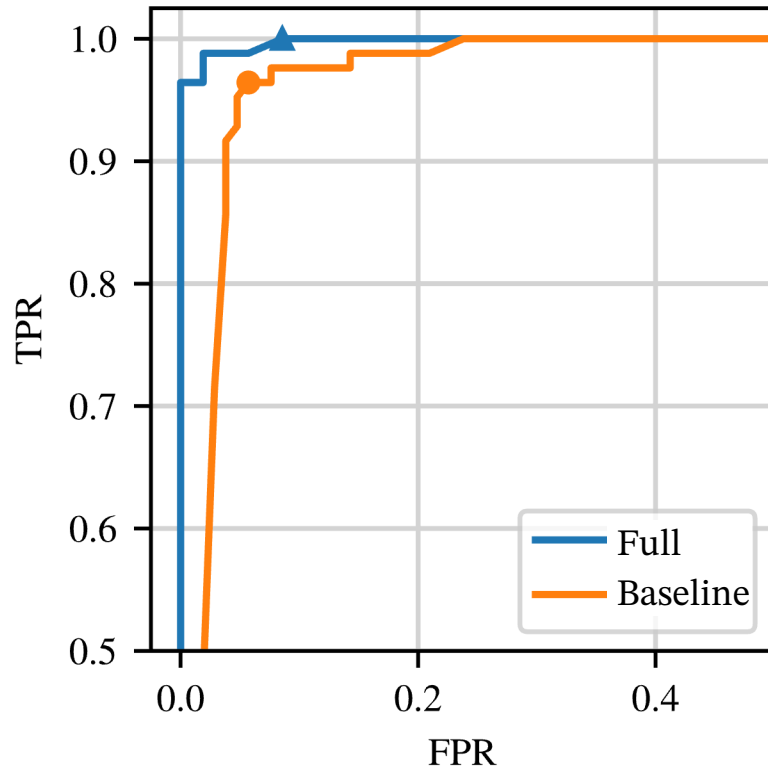


# Performance gains across all distances



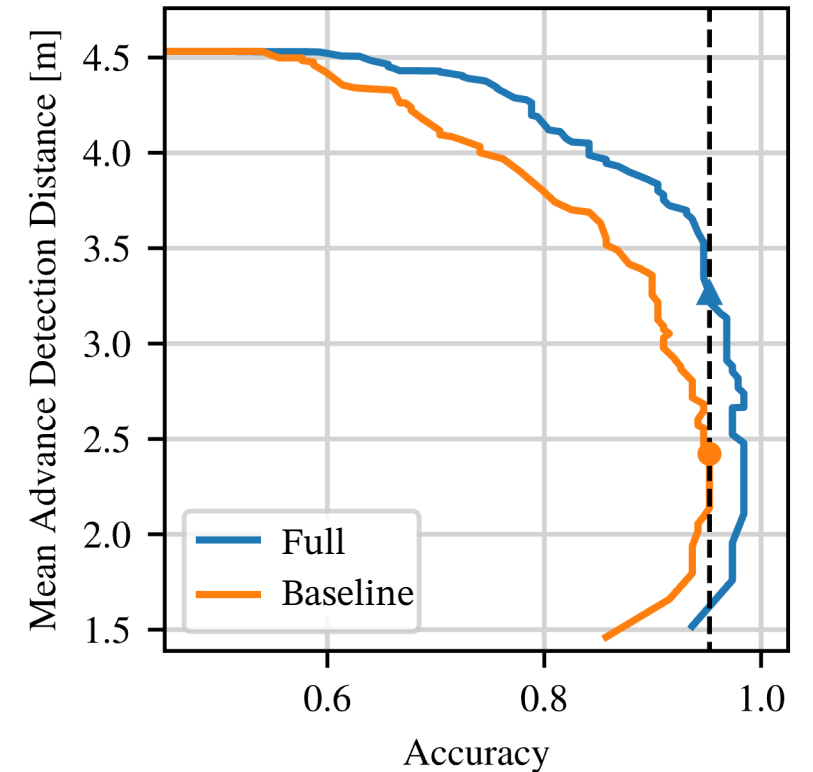
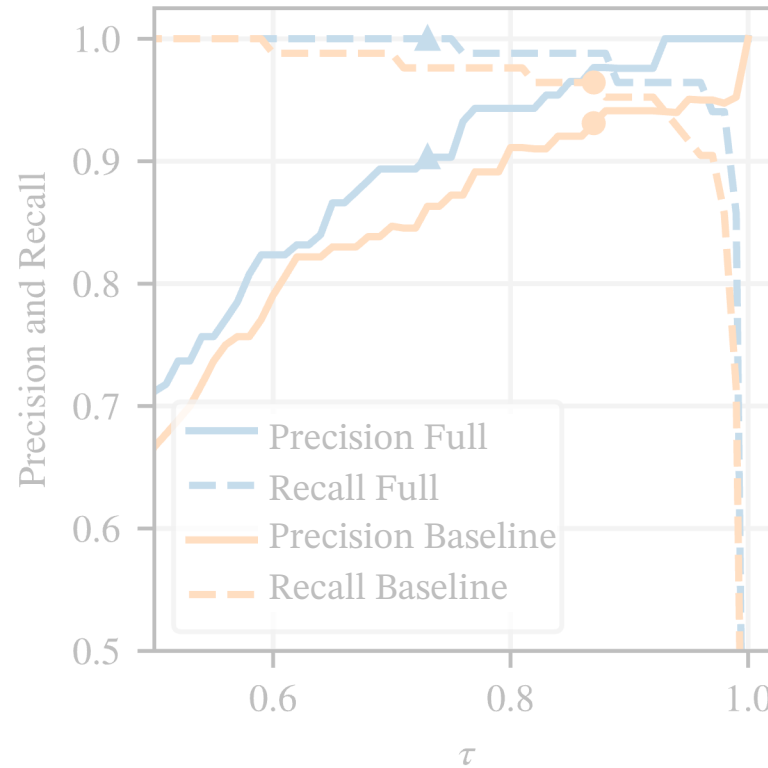
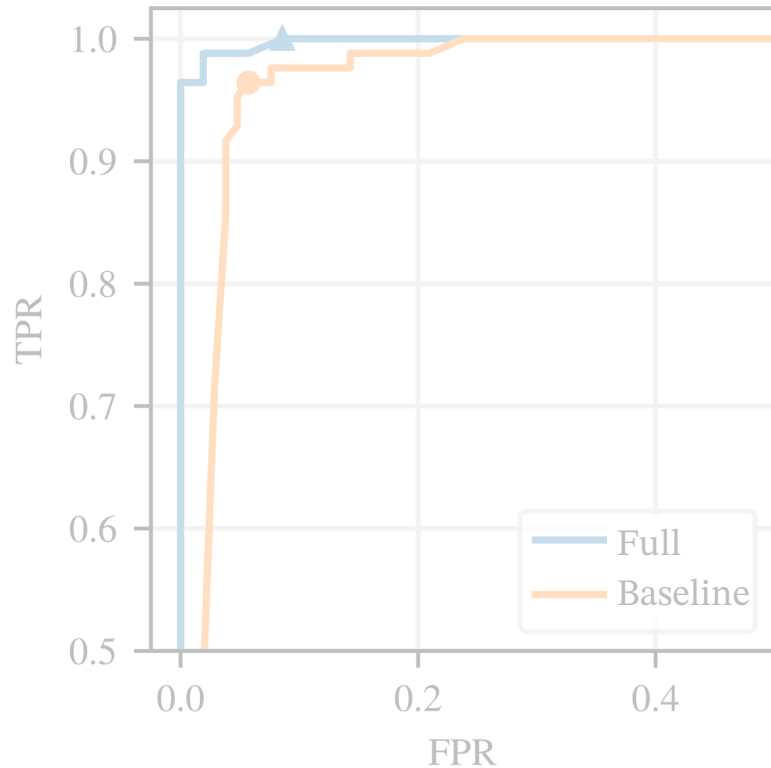
- **Full model improves AUROC consistently across all distances (+5 – +11%)**

# Introducing facial and gaze cues helps



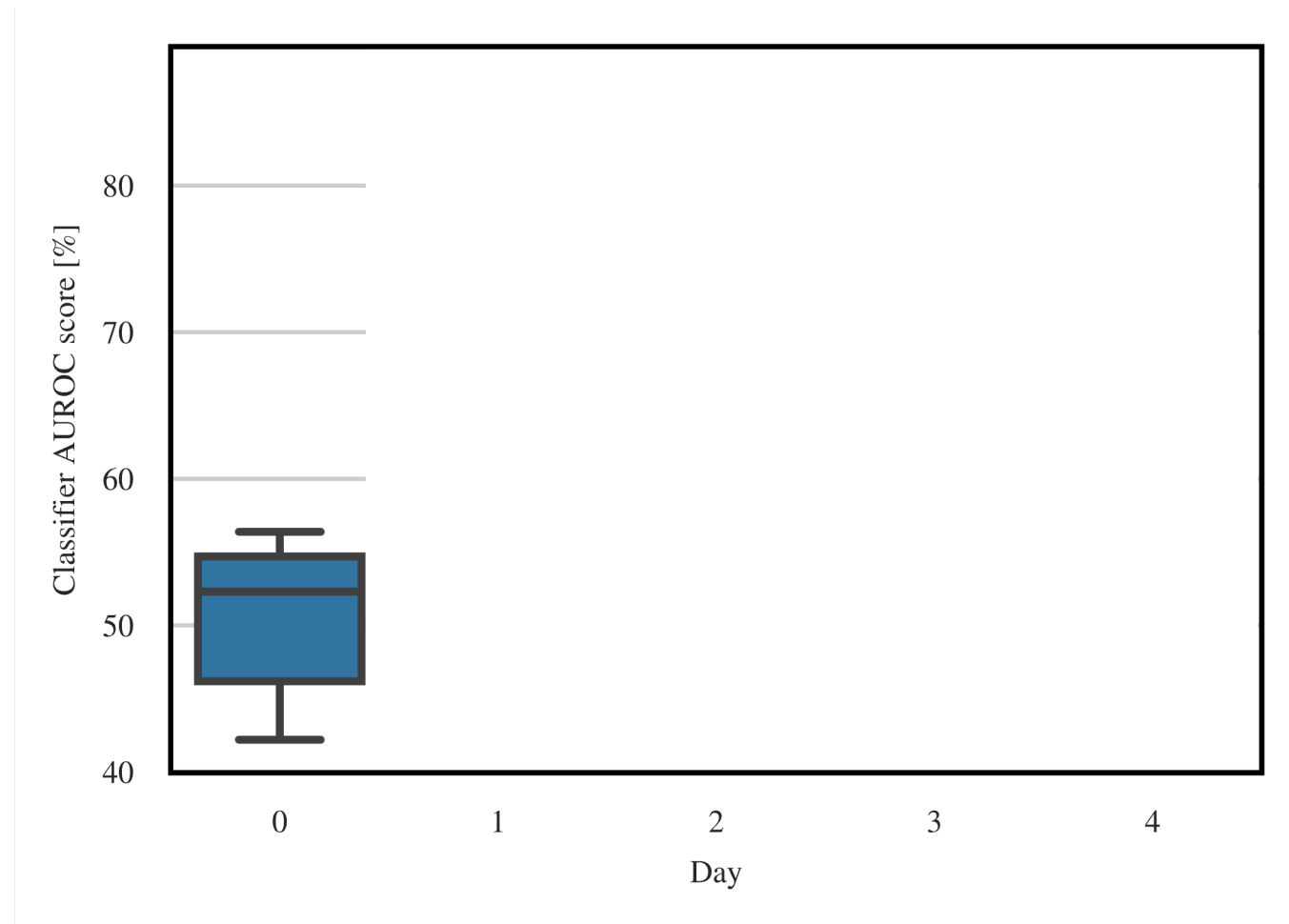
- **Full** model consistently **outperforms Baseline** model across considered metrics

# Introducing facial and gaze cues helps



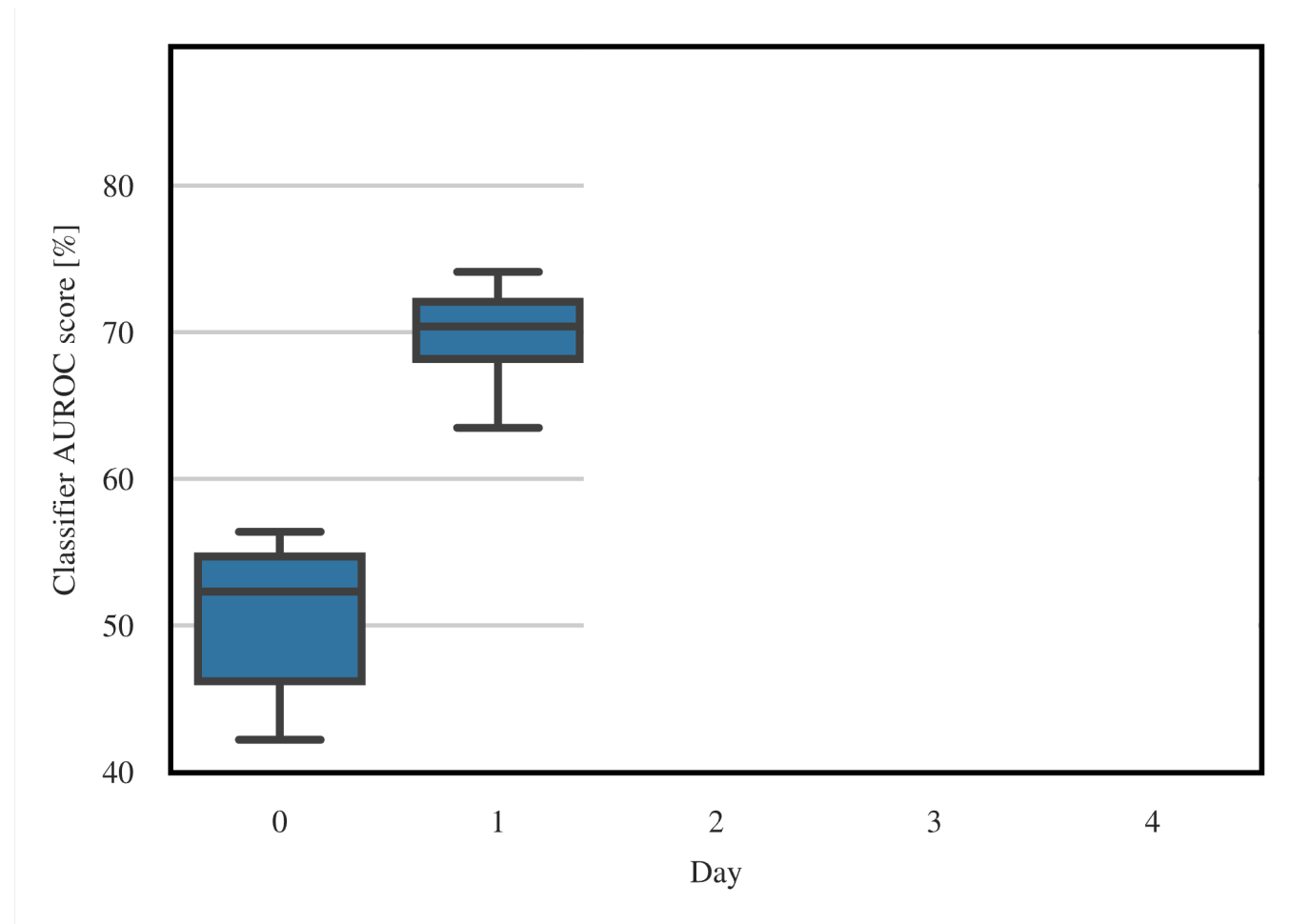
- **Full** model consistently **outperforms Baseline** model across considered metrics
- Achieving **same accuracy** Full model **gain 0.85 m** in mean advance detection distance

# Self-supervised adaptation



- System can **adapt to new environments** gathering and self-labelling data **autonomously**

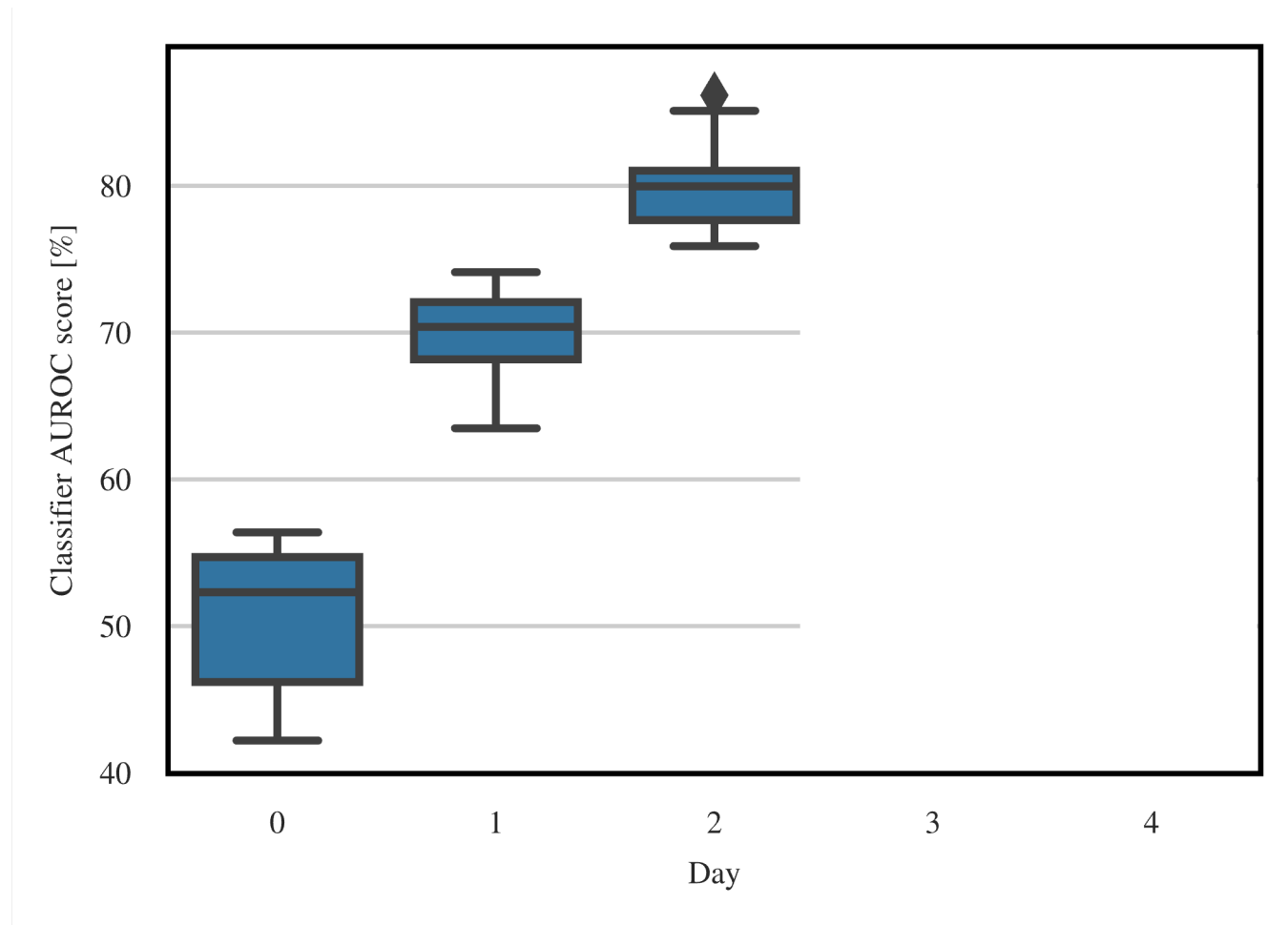
# Self-supervised adaptation



- System can **adapt to new environments** gathering and self-labelling data **autonomously**

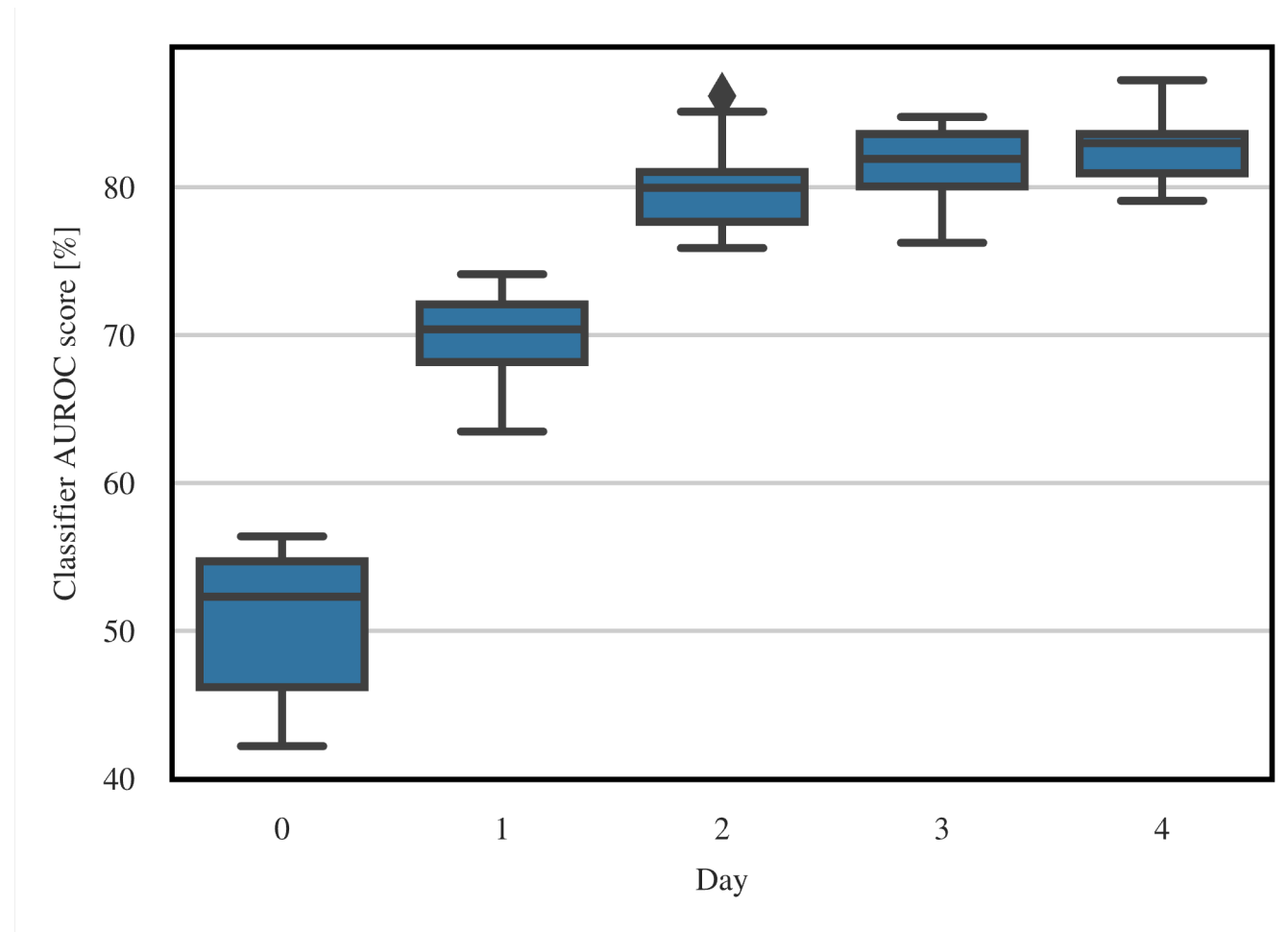


# Self-supervised adaptation



- System can **adapt to new environments** gathering and self-labelling data **autonomously**

# Self-supervised adaptation



- System can **adapt to new environments** gathering and self-labelling data **autonomously**

**Single user**





# Conclusions and Future Work

- Facial features and gaze cues significantly help in the intention detection process
- Explore the **impact** of robot reactions **on users** perceptions
- Deployment and validation **in the wild**



Project website:  
[bit.ly/idsia-iid-gaze](https://bit.ly/idsia-iid-gaze)

