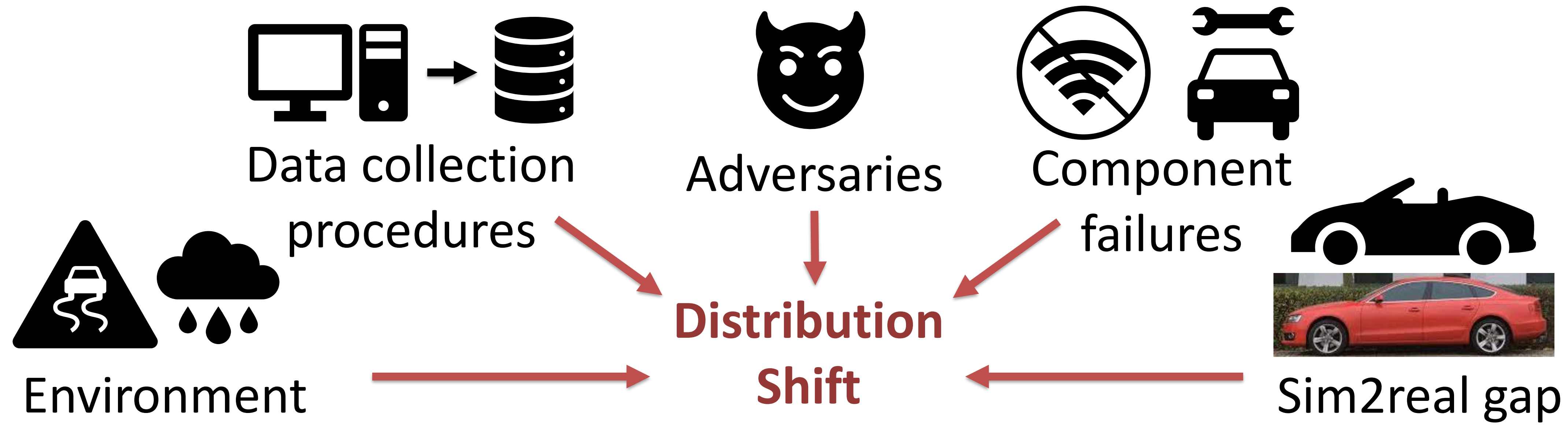
Distribution Shift in Learning-Enabled Cyber-Physical Systems: Safety Monitoring and Recovery

Insup Lee (PI), Vivian Lin (GRA)
University of Pennsylvania

Challenge: Maintaining reliability in learning-enabled cyber-physical systems (LE-CPS) under distribution shift.

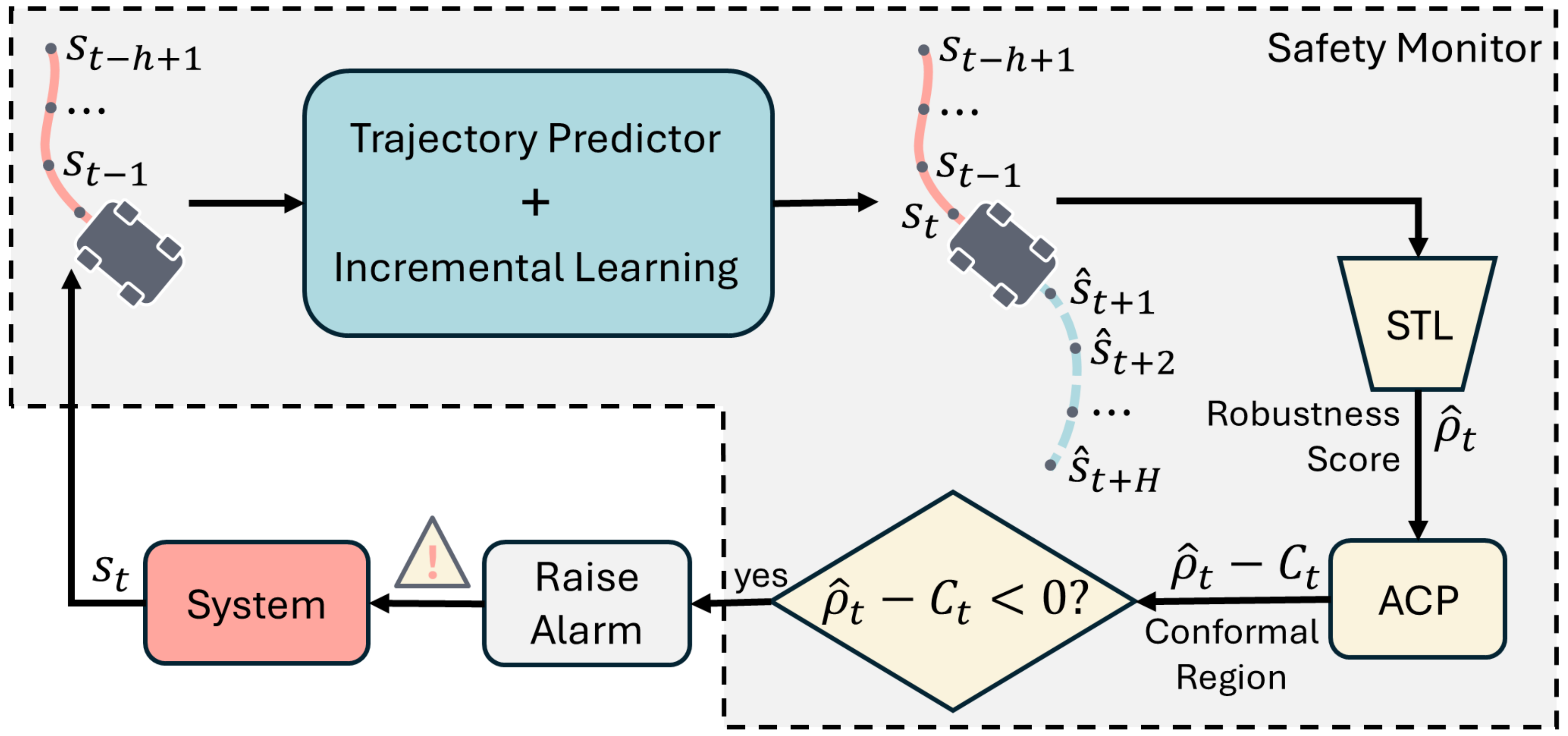
- Neural networks are fragile to differences between the train and test distributions, leading to downstream failures.
- Detecting distribution shift and abstaining from a decision is conservative and leads to inaction.

Scientific Impact: Distribution shift can affect any LE-CPS.



Solution: Monitoring STL safety properties directly is less conservative than detecting distribution shift.

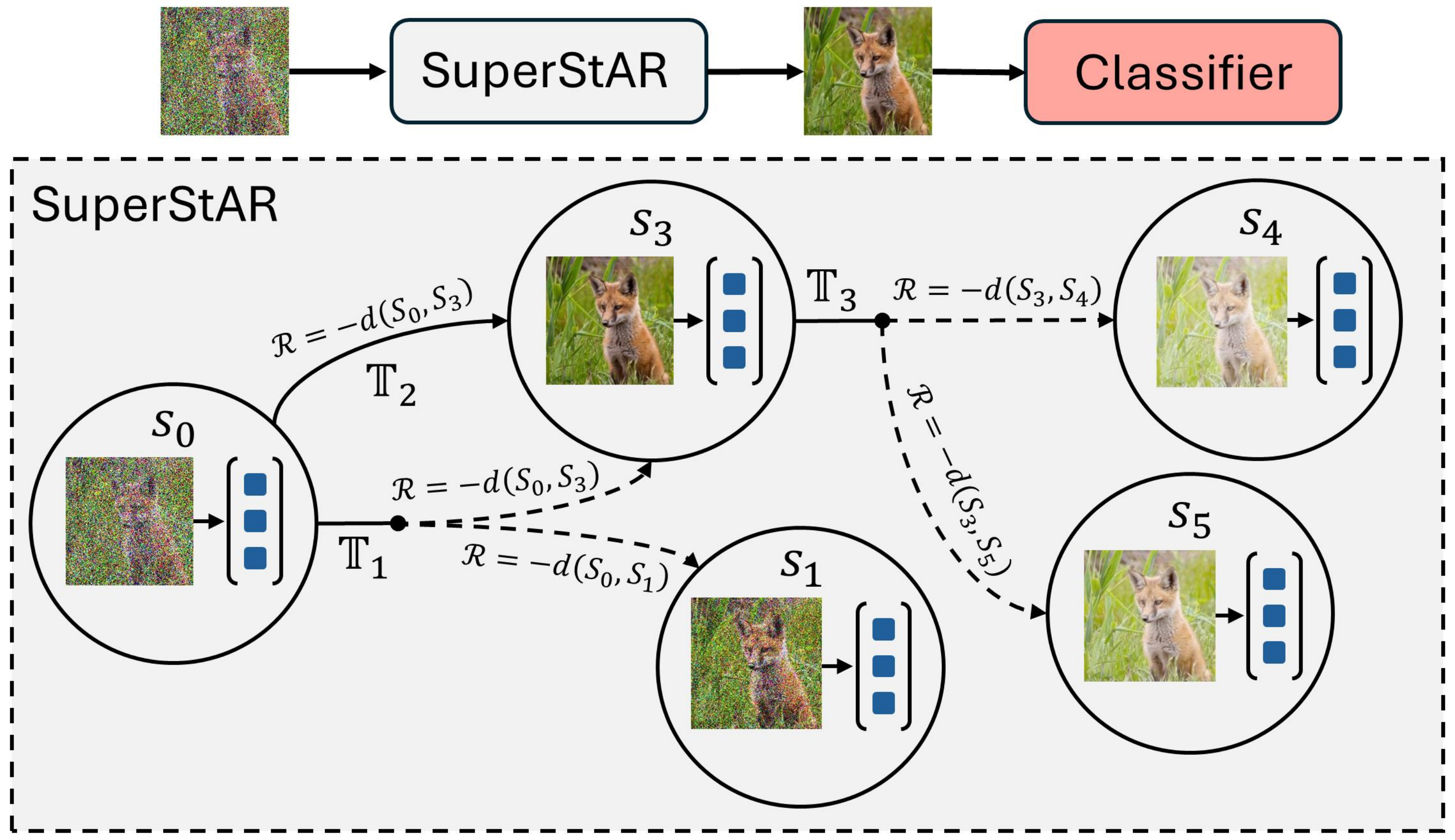
- Incremental learning plus adaptive conformal prediction leads to timely alarms with competitive recall.



V. Lin, R. Kaur, Y. Yang, S. Dutta, Y. Kantaros, A. Roy, S. Jha, O. Sokolsky, and I. Lee, "Safety Monitoring for Learning-Enabled Cyber-Physical Systems in Out-of-Distribution Scenarios," in ICCPS 2025.

Solution: Correcting data at inference time allows neural networks to make decisions and systems to act.

- By casting distribution shift recovery as a Markov decision process, up to 14.21% of average accuracy is recovered.


SuperStAR
Classifier
SuperStAR
S_0
S_1
S_3
S_4
S_5
$\mathbb{T}_1$
$\mathbb{T}_2$
$\mathbb{T}_3$
$\mathcal{R} = -d(S_0, S_3)$
$\mathcal{R} = -d(S_0, S_3)$
$\mathcal{R} = -d(S_0, S_1)$
$\mathcal{R} = -d(S_3, S_4)$
$\mathcal{R} = -d(S_3, S_5)$

V. Lin, K. J. Jang, S. Dutta, M. Caprio, O. Sokolsky, and I. Lee, "DC4L: Distribution Shift Recovery via Data-Driven Control for Deep Learning Models," in L4DC 2024.

Broader Impact – Societal:

Distribution shift limits the efficacy, safety, and profitability of LE-CPS, impacting practitioners and consumers.

Broader Impact – Education:

Provides research experience and exposure for graduate students.

Broader Impact – Quantified:

Probabilistic guarantees are obtained on the safety monitor. Recovery algorithm provably improves performance in optimal case.

Penn Engineering | PRECISE
PENN RESEARCH IN EMBEDDED COMPUTING AND INTEGRATED SYSTEMS ENGINEERING

Award ID#:
NSF CNS-2143274 and NSF CNS-2333980