Inferring Severe Accident Scenarios in NPPs with Reinforcement Learning (RL) and Supervised Learning (SL) Approaches

- Part 3 -
Sensitivity of RL to SL

Semin Joo

Master's Student
Nuclear Power and Propulsion Laboratory (NPNP)
Dept. of Nuclear and Quantum Engineering, KAIST

CONTENTS

01
INTRODUCTION

Introduction

Background

1. Severe accidents are highly non-linear and chaotic in nature.

2. DSA & PSA-based methods require large computational resources.

3. Need to develop an alternative method that can incorporate uncertainties more easily with fewer computational resources

Q Can AI be a tool for the prediction and management of severe accidents?

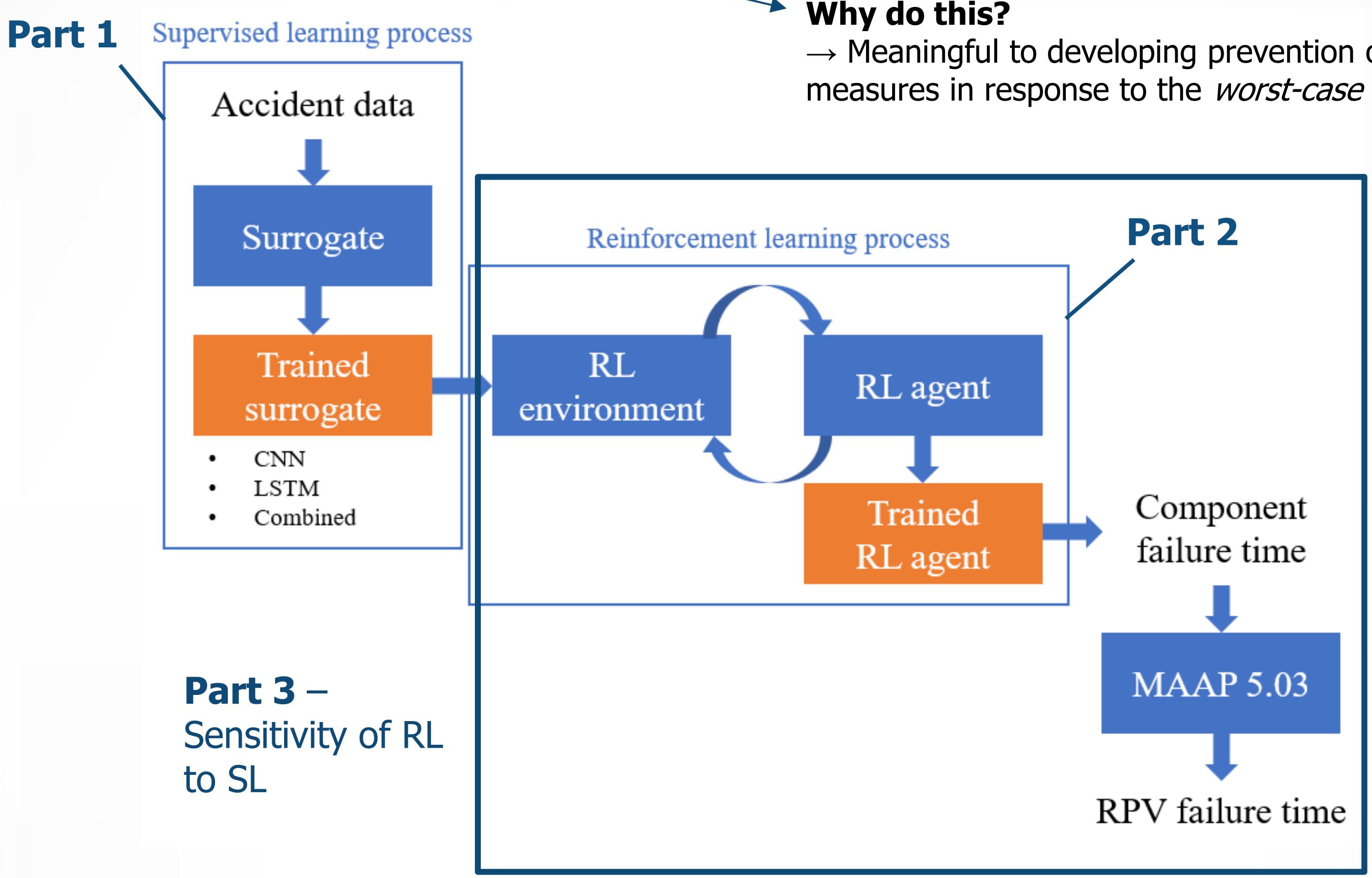
Introduction

Objectives

Main Goal:

Develop an artificial neural network (ANN)-based method that predicts the progression of a severe accident in an **accelerated** manner.

Why do this?
→ Meaningful to developing prevention or mitigation measures in response to the *worst-case* scenario



02
METHODOLOGY

Methodology

Selection of Accident Scenario

Accident type: Loss of Component Cooling Water (LOCCW)

Components that can fail over 72 hrs

- RCP
- HX
- HPI pump
- LPI pump
- CSS pump
- MDAFW pump
- CHP

→ 10,679 accident scenarios → MAAP 5.03 Code →

TH variables

- Primary system pressure
- Cold leg temperature
- Hot leg temperature
- RV water level
- SG pressure
- SG water level
- Max. core exit temperature

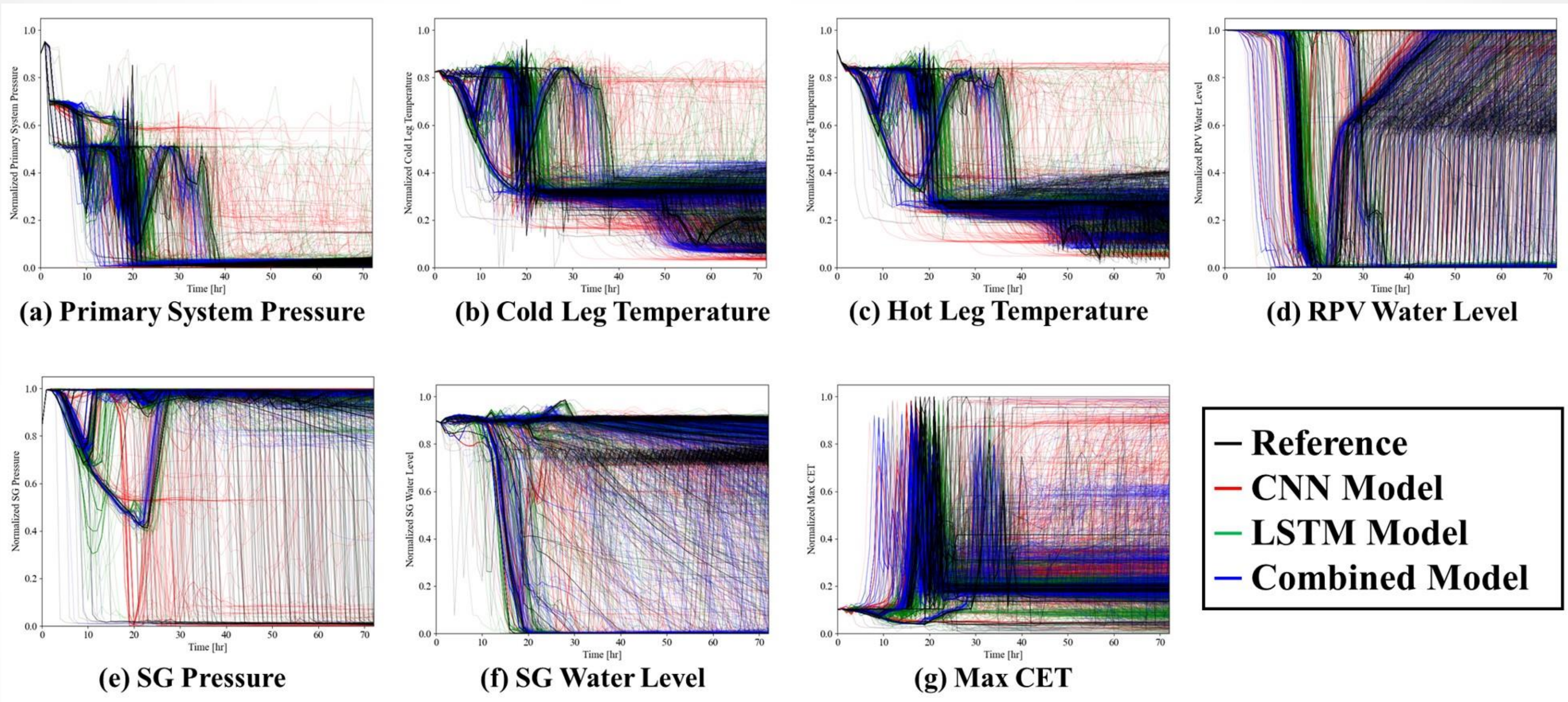
*RCP = Reactor Coolant Pump
*HPI = High-Pressure Injection
*LPI = Low-Pressure Injection
*CSS = Containment Spray System
*MDAFW = Motor-Driven Auxiliary Feedwater
*CHP = Charging Pump

*Observable from the MCR and SAMG supervisory variables

Methodology

Selection of Accident Scenario

Accident scenarios generated by MAAP 5.03



Methodology

Surrogate Models

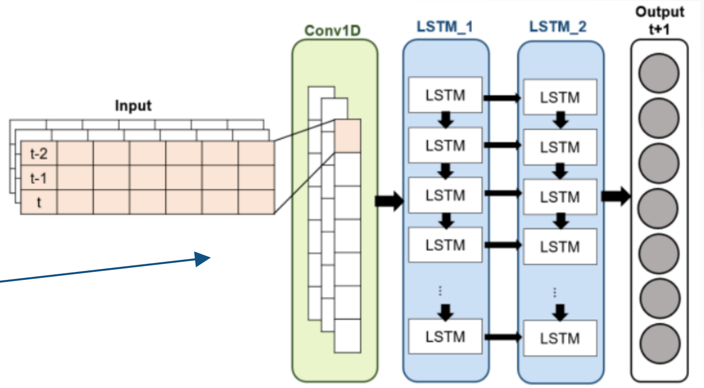
1 **A surrogate model is a supervised learning technique to predict an outcome using a data-driven approach.**

It quickly predicts the TH variables of an NPP in real time.

The surrogate model is coupled to the RL environment.

2 **Three Deep Neural Network (DNN) types were considered as surrogate models.**

1. Convolutional Neural Network (CNN)
2. Long-term Short Memory (LSTM)
3. CNN-LSTM combined network



3 **Each DNN is trained by the LOCCW data generated by MAAP 5.03.**

The performances of the surrogate models are evaluated by mean absolute error (MAE).

Methodology

Reinforcement Learning

Action

The RL agent chooses the following:

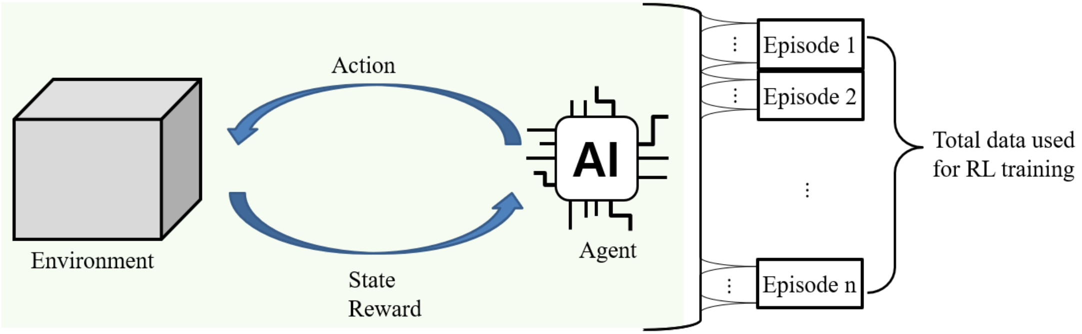
RCP	→ Fail at t = 1hr
HX	→ Fail or not?
HPI pump	→ Fail or not?
LPI pump	→ Fail or not?
CSS pump	→ Fail or not?
MDAFW pump	→ Fail or not?
CHP	→ Fail or not?

Reward

Logic: over-pressurization of the primary system may cause RPV failure.

The earlier, the better.

$$reward = \begin{cases} \Delta P_1 \cdot \boxed{(72 - t),} & \Delta P_1 \geq 0 \\ 0, & \Delta P_1 < 0 \end{cases}$$

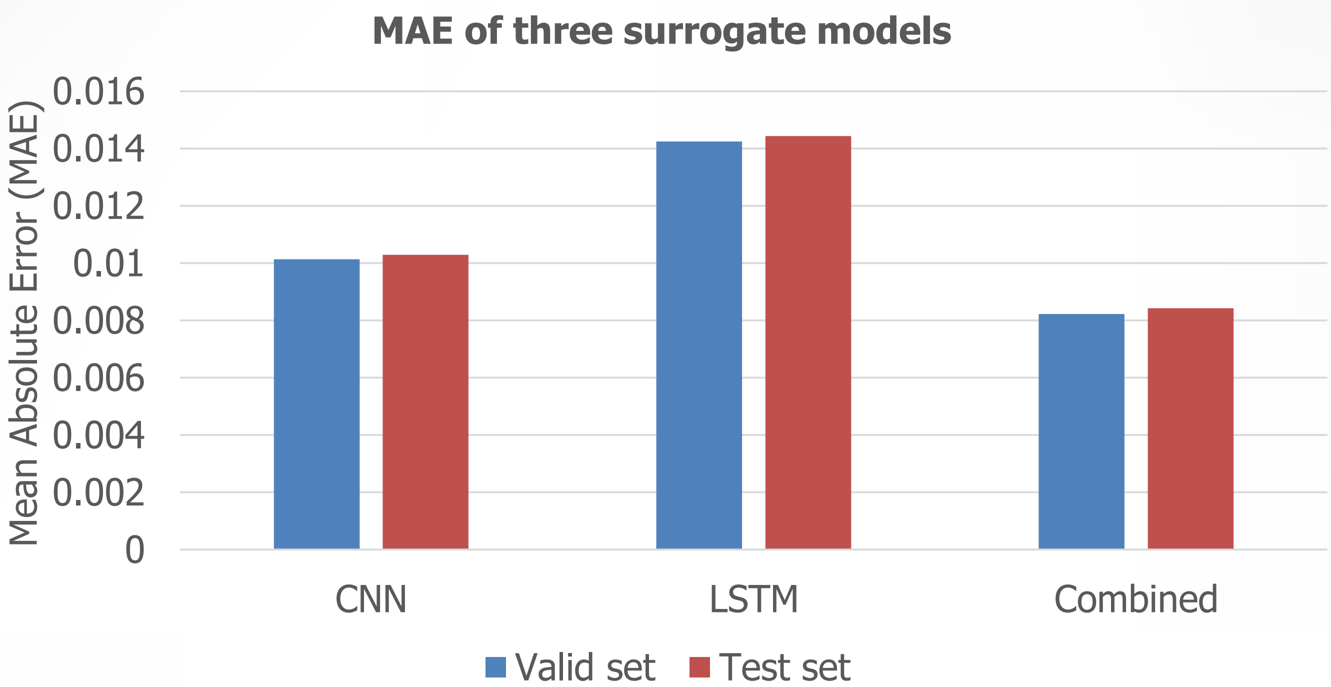


▲ Interaction between the RL agent and environment

03
RESULTS AND DISCUSSION

Results and Discussion

Performance of surrogate models



Mean Absolute Error (MAE) comparison: CNN-LSTM < CNN < LSTM

Reason 1) CNN layer extracts important features from the time series data.

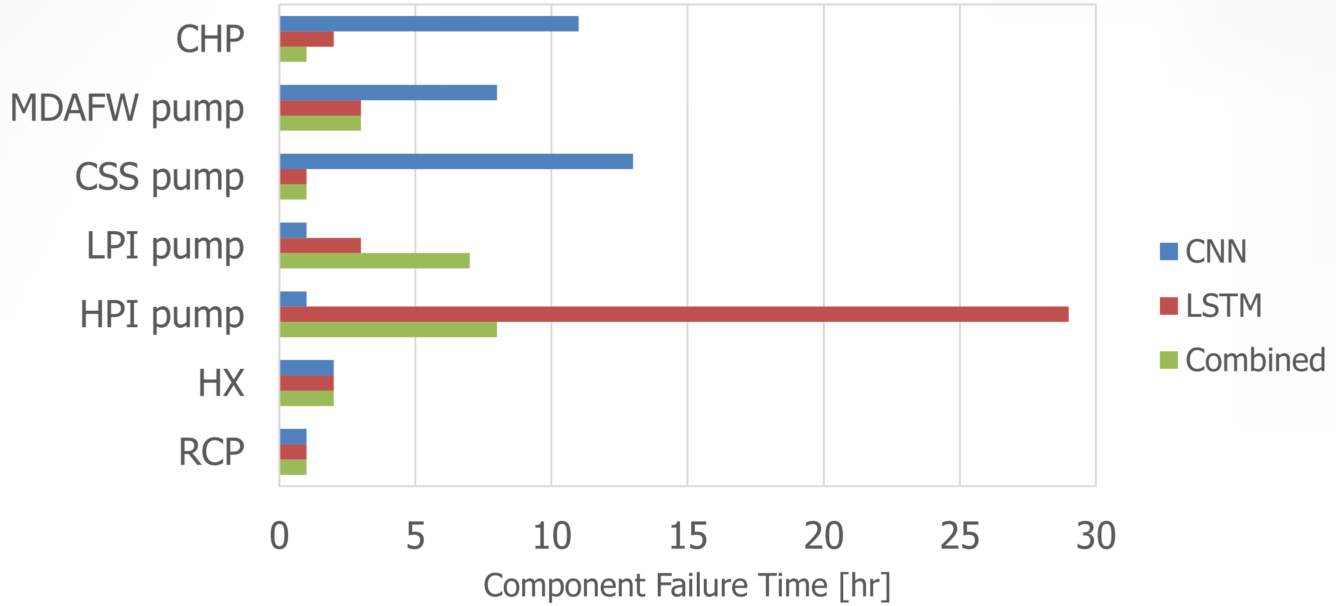
Reason 2) CNN layer reduces the # of parameters that needs to be optimized at the LSTM layer.

Results and Discussion

Performance of RL agents

The RL agents were given 1,000 chances to choose the component failure times.

Most frequently chosen component failure times

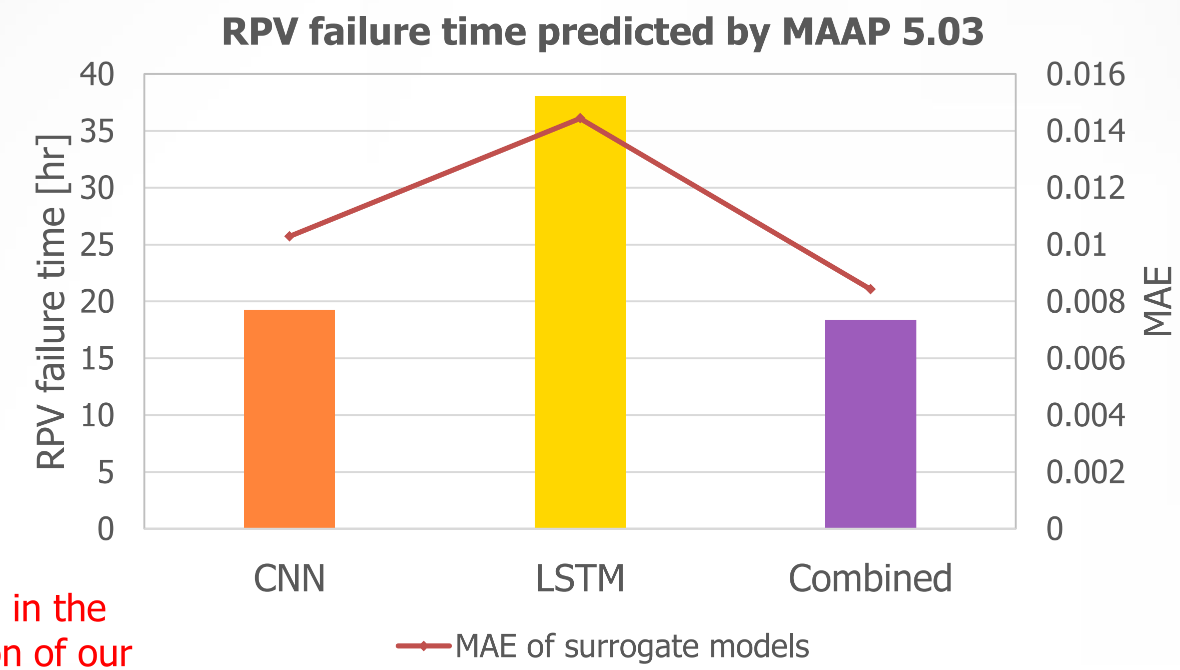


Comparison of the most frequently chosen component failure times selected by three RL agents:

LSTM-based RL agent tends to select a significantly delayed HPI failure time.

Results and Discussion

Performance of RL agents



Trained in the direction of our original intention

Comparison of RPV failure times generated by three RL agents

- RPV failure times: CNN-LSTM < CNN < LSTM
- MAE: CNN-LSTM < CNN < LSTM

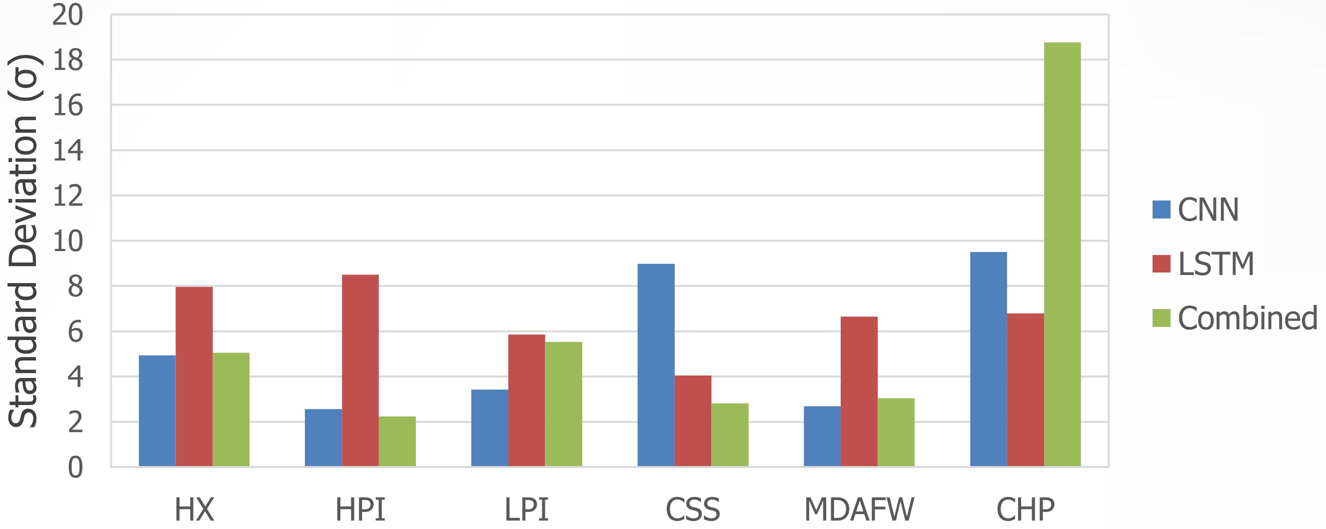
→ The performance of RL is improved by combining the CNN layer with the LSTM layer.

Results and Discussion

Uncertainty of RL models

Standard deviations of the component failure time distribution

→ Indirect measure of the uncertainty of RL



What does it mean to have a small σ?

- Means that those components play a big role in accelerating the RPV failure time.
- Intuitively, those components should be HPI, LPI, and MDAFW pumps.

Comparison among surrogate models

- LSTM-based RL models have larger σ on average.
- CNN, CNN-LSTM-based RL models have smaller σ for HPI, LPI, and MDAFW pump failure times.

    → They not only perform better but also have small uncertainties.

04

CONCLUSIONS

Conclusions

Summary

Part 1
- SL development -

- LOCCW accident scenarios were generated using MAAP 5.03 code.
- These datasets were used to train three different surrogate models:
  1) CNN
  2) LSTM
  3) CNN-LSTM
- CNN-LSTM model showed the least MAE.

Part 2
- RL development -

- Develop an RL agent that predicts the progression of a severe accident in an accelerated manner.
- Two different reward systems were tested:
  1) Pressure reward
  2) CET reward
- Compare their accident consequences.

Part 3
- Sensitivity of RL to SL -

- Investigate the effect of the surrogate model on the RL agent's performance.
- Three different surrogate models were coupled to the RL environment:
  1) CNN
  2) LSTM
  3) CNN-LSTM
- The higher the performance of the surrogate model, the earlier the RPV failure time becomes.

Conclusions

Limitations and Further Works

1 **Surrogate model improvement**

∵ The performance of the surrogate model affects the RL agent's actions.

Ex) Hyperparameter adjustment, attempting different types of DNN layout

2 **Search for a better RL reward system**

- Since RPV failure is a complex and non-linear phenomenon, there is a need for a more sophisticated reward system.

∵ The action of the RL agent is directly affected by the reward system.

3 **Uncertainty quantification**

- Uncertainties associated with MAAP 5.03 code
- Uncertainties associated with the surrogate model → dynamic time-warping distance
- Uncertainties associated with the RL model → variance of the learned distribution

Q&A

Appendix

Selection of accident scenario

- Reference reactor type: OPR1000

- Level 2 PSA
  - Covers from core damage to CTMT failure

- Total Loss of Component Cooling Water (TLOCCW)
  - Event that possibly leads to RPV failure
  - One of the most frequent accidents at Level 2 PSA
  - PSA mission time = 72 hr → scenario length = 72 hr
  - Triggered by single/multiple failures of 7 safety components (HPI pump, LPI pump, HX, RCP seal, MDAFW pump, CSS pump, charging pump)

- Consequence of accident
  - <u>RPV failure</u>, rather than CTMT failure, was assumed to be the consequence of the TLOCCW accident. This is because preventing RPV failure is the primary objective of the mitigation strategies.

Appendix

Supervised Learning (SL) - Surrogate Model

- Time step
  - Time step size = 1 hour
  - Smaller step size → too much data & difficult to interact with RL agent
  - 3 step model performs better than 1 step model.

	1 step	3 step
Valid	0.0152	0.0085
Test	0.0155	0.0087

- Deep Neural Networks (DNN)
  - CNN, LSTM: specialized at predicting time series data
  - Performance metrics: regression performance (t → t+1) is calculated by mean absolute error (MAE)

$$MAE = \frac{1}{n}\sum_{i=1}^{n}\left|y_i - \hat{y}_i\right|, \text{ where } y_i \text{ : reference data, } \hat{y}_i \text{ : predicted value}$$

  - Deep neural network that is composed of CNN-LSTM layer often shows enhanced performance in predicting and classifying data. (D. W. Shin et al. (2016), T. Y. Kim, S. B. Cho (2019), A. Tasdelen, Baha Sen (2021), B. S. Seo et al. (2021))

Appendix

Supervised Learning (SL) - Surrogate Model

- DNN structures

<table>
<tr><th></th><th>CNN</th><th>LSTM</th><th>CNN-LSTM</th></tr>
<tr><td rowspan="4">Structure of layers</td><td>Conv1D(filters=100, kernel_size=(3,), activation='relu')</td><td>LSTM(100, return_sequences=True)</td><td>Conv1D(filters=100, kernel_size=(3,), activation='relu')</td></tr>
<tr><td>Dense(units=100, activation='relu')</td><td>LSTM(100, return_sequences=True)</td><td>LSTM(100, return_sequences=True)</td></tr>
<tr><td rowspan="2">Dense(units=7, activation='sigmoid')</td><td rowspan="2">Dense(units=7, activation=‘sigmoid')</td><td>LSTM(100, return_sequences=True)</td></tr>
<tr><td>Dense(units=7, activation='sigmoid')</td></tr>
<tr><td>Loss</td><td colspan="3">Mean squared error (MSE)</td></tr>
<tr><td>Optimizer</td><td colspan="3">Adam</td></tr>
<tr><td>Learning rate</td><td colspan="3">10^{-3}</td></tr>
<tr><td>Epochs</td><td colspan="3">500 with early stopping</td></tr>
</table>

Appendix

Reinforcement Learning (RL)

- Proximal Policy Optimization (PPO) algorithm
  - Policy = an action that an agent can take with a probability
  - Limits the range of policy change → fast convergence

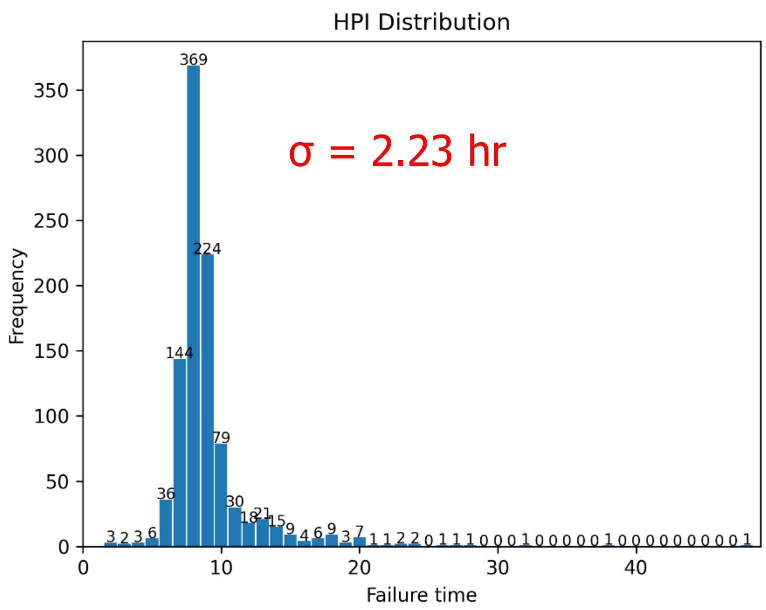
- Error propagation
  - The minimum/maximum RPV failure time could be identified (10 hr / 72 hr)
  - By not implementing any of the mitigation strategies, the RPV failure time could be further accelerated.

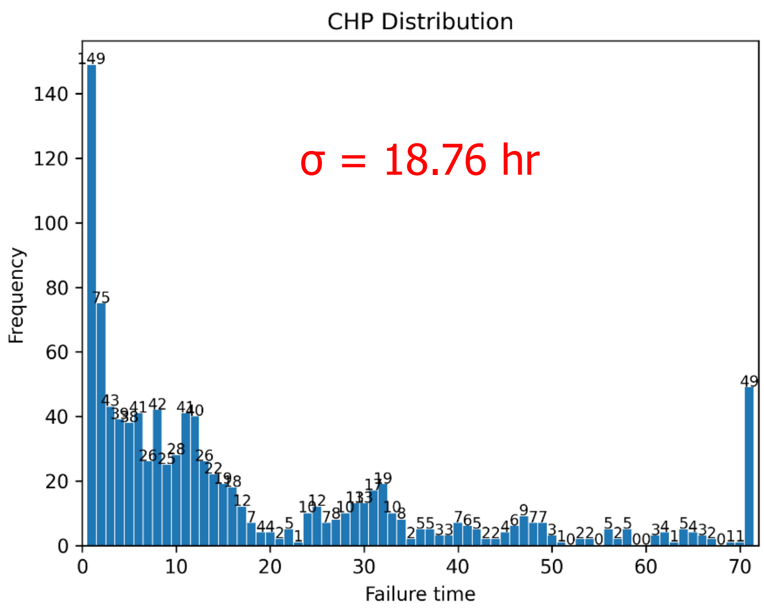
- Insights
  - The reward system significantly affects the RL agent's action and thus the RPV failure time.
  - The reward system should facilitate the learning process and accelerate the RPV failure time at the same time.

Appendix

Reinforcement Learning (RL)

- Component Failure Time Distribution
  - For example, the **CNN-LSTM**-based RL model selected the **HPI pump** and **CHP** failure times in the following manner:





→ HPI pump failure time tends to cluster at t = 8 hr.