Seminar at NTNU RAMS lab

Remaining useful life prediction based on machine learning approaches

Yuchen Jiang & Shimeng Wu

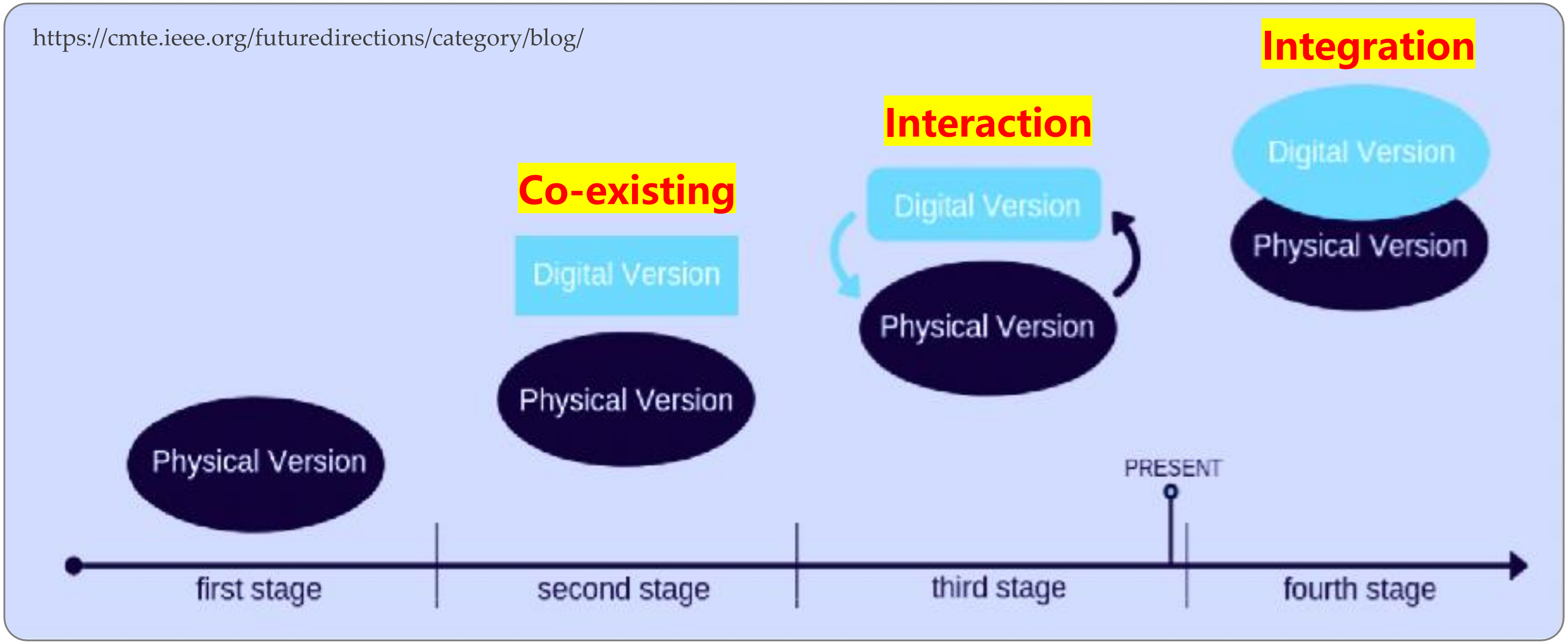
Harbin Institute of Technology

2021. 09. 30

Content

The impact of digital transformation

HARBIN INSTITUTE OF TECHNOLOGY



- The four stages have shown significant differences and led to entirely different working and life style.
- With the aid of deep integration of physical and digital world, many impossibilities become possible. At the upcoming stage, digital twins becomes a part of the feature sets of the physical twins.

The role of DT in PHM

In 2016, the GE digital team showed how they designed DTs for PHM purposes.

Physical assets

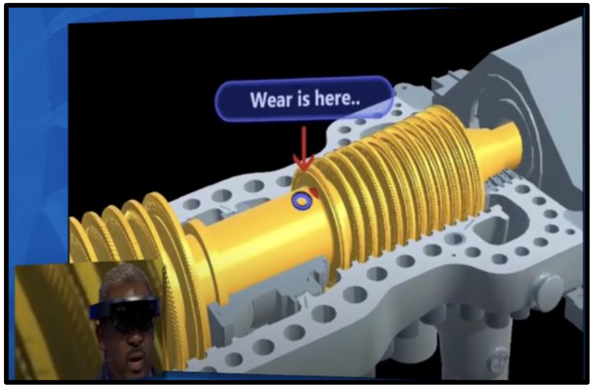
E.g., a total of 125 D11-type turbines practically deployed and operating in Southern California, USA

Digital twin

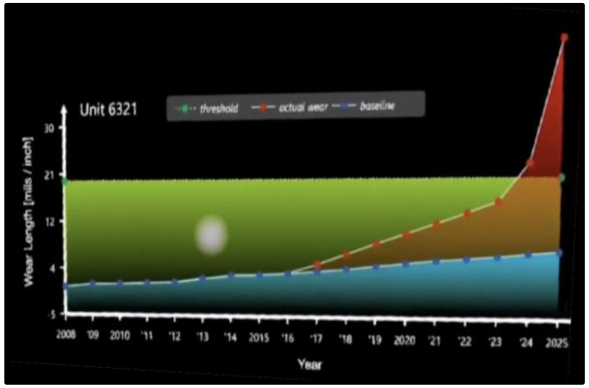
A virtual replica of the physical entity in the digital space.
Eg., A product, a component, a process, a dataset, a being, etc.

The role of DT in PHM

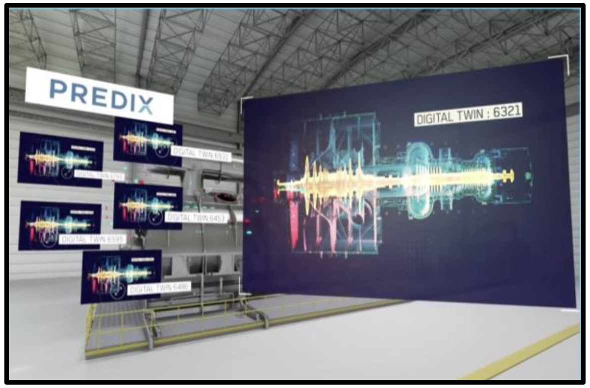
A virtual replica which characterizes the full life-cycle of the physical entity and is operational.



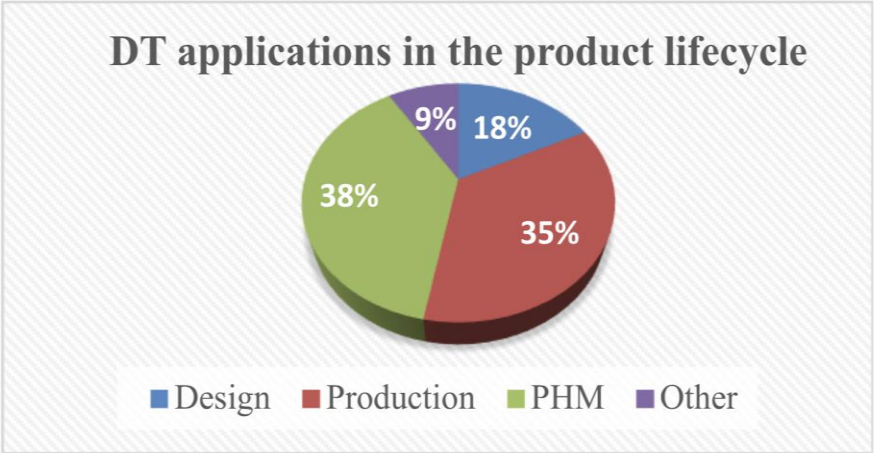
a



b



c



PHM applications powered by DT

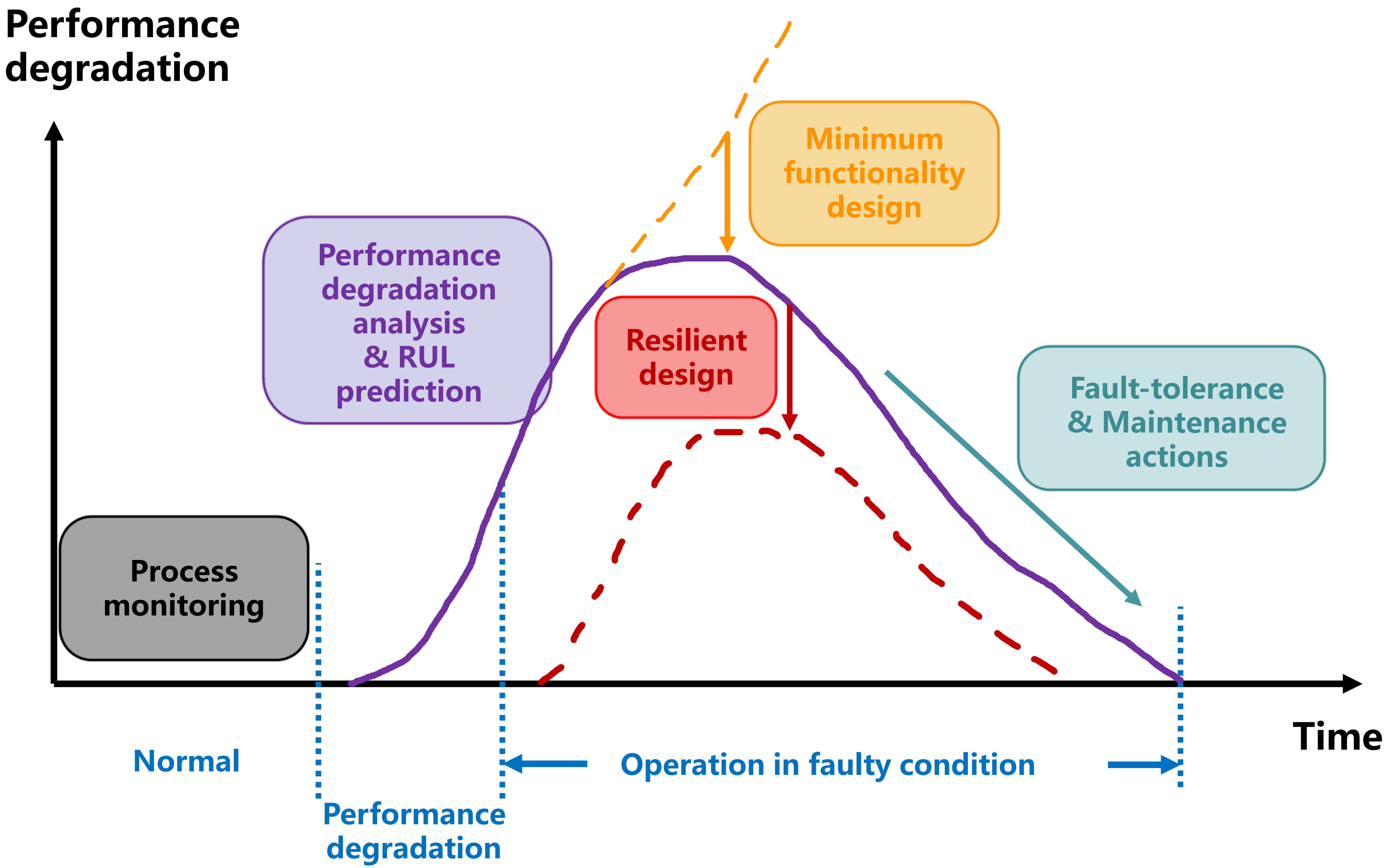
- **Online parallel simulation and optimization (scheduling, set-point, etc) for monitoring, diagnosis, and prognosis**
- **Colossal amount of process history data**
- **Interaction and learning from the fleet**

Digital Twin in Industry: State-of-the-Art, IEEE Trans. Ind. Informat. 15(4), 2019.

Industrial Applications of Digital Twins, Philosophical Transactions of the Royal Society A, 379(2270): 20200360, 2021.

Full life-cycle of system performance

HARBIN INSTITUTE OF TECHNOLOGY



Content

Outline:

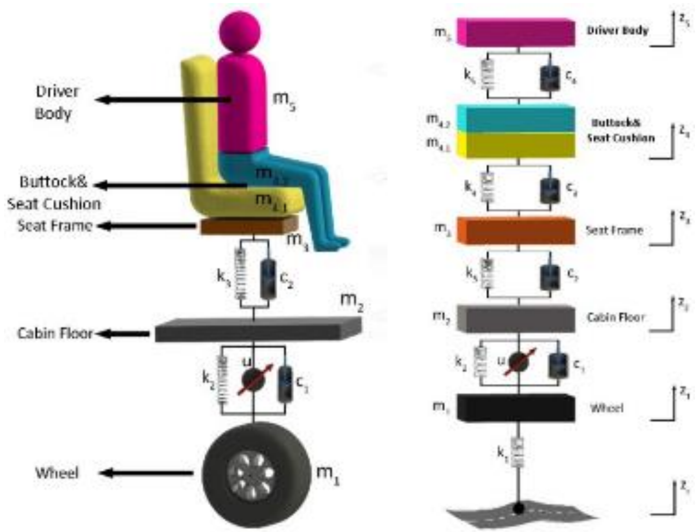
- Basics and subject of study
- Scientific problem 1
- Scientific problem 2

Basics and subject of study

哈爾濱工業大學
HARBIN INSTITUTE OF TECHNOLOGY

Categories of RUL prediction approaches

Model-based RUL prediction approaches

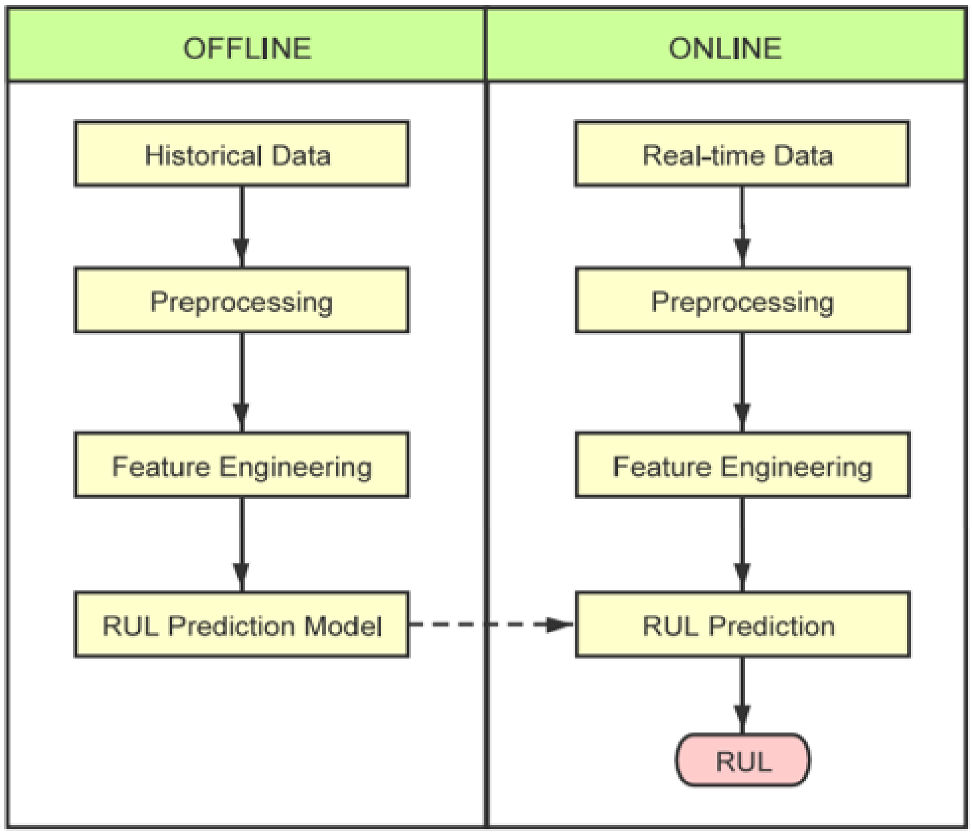


Data-based RUL prediction approaches

Machinery health prognostics: A systematic review from data acquisition to RUL prediction. Mechanical systems and signal processing, 2018.

HARBIN INSTITUTE OF TECHNOLOGY

RUL prediction procedure



Offline modeling

- Establishment of the prognostic models
- The historical monitoring data or the physical models of subsystems or components.

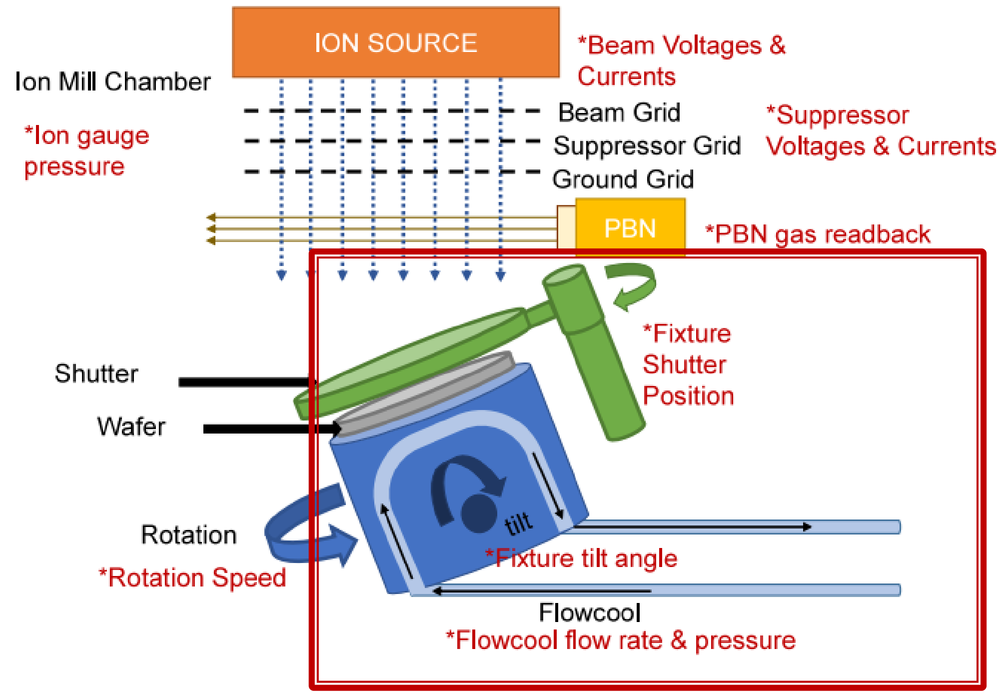
Online testing

- Real-time prediction of RUL
- Based on the models obtained at the offline stage.

HARBIN INSTITUTE OF TECHNOLOGY

Case study: Flowcool system

- The **ion etching technology** can process the surface structure of **wafers** at the micron level.
- **Ion mill etching (IME) machines** have become necessary in wafer processing.



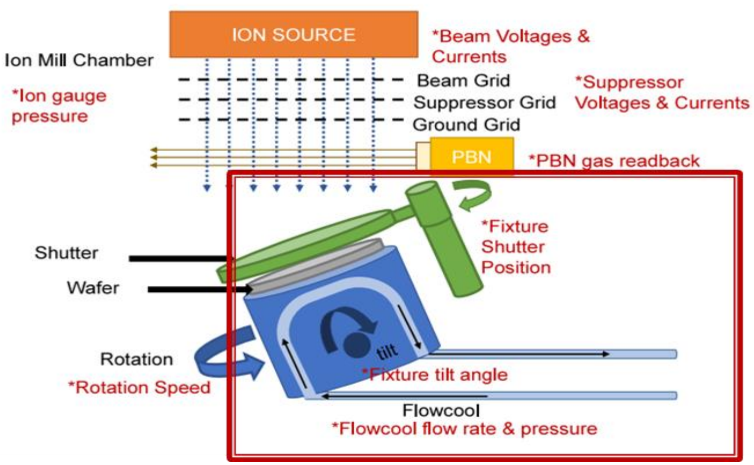
- The **flowcool system** plays an important role in the etching process
- Prevent damage to the wafer surface caused by overheat
- Many different failure scenarios take place in the **flowcool system**

Remaining useful life prediction for ion etching machine cooling system using deep recurrent neural network-based approaches. Control Engineering Practice, Volume 109, 2021.

Basics and subject of study

Dataset is provided by 2018 PhM Data Challenge competition

- ✓ Sensor measurements and working condition configuration parameters of 20 IME machines
- ✓ 24 feature variables are recorded every 4 seconds
- ✓ The occurrence time of three kinds of failures



Fault types:

- F1: Flowcool leak
- F2: Pressure Too High Check Flowcool Pump
- F3: Pressure Dropped Below Limit

Outline:

- Basics and subject of study
- Scientific problem 1
- Scientific problem 2

Scientific problem 1

Find the early degradation point

- Extremely large possible RUL
- Average interval of given RUL interval: 60.82×10^4

Train the prediction model using all given instances

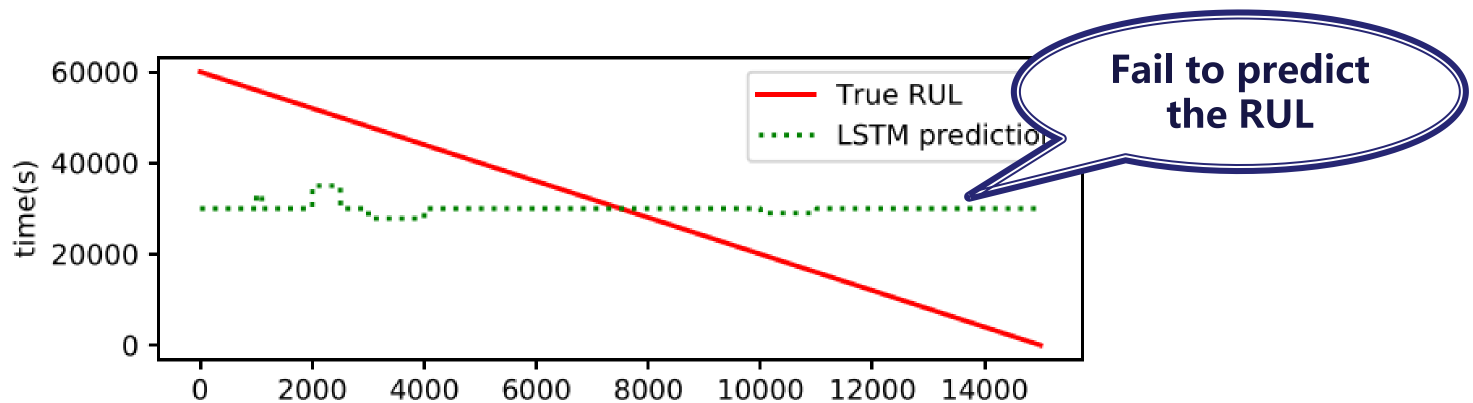
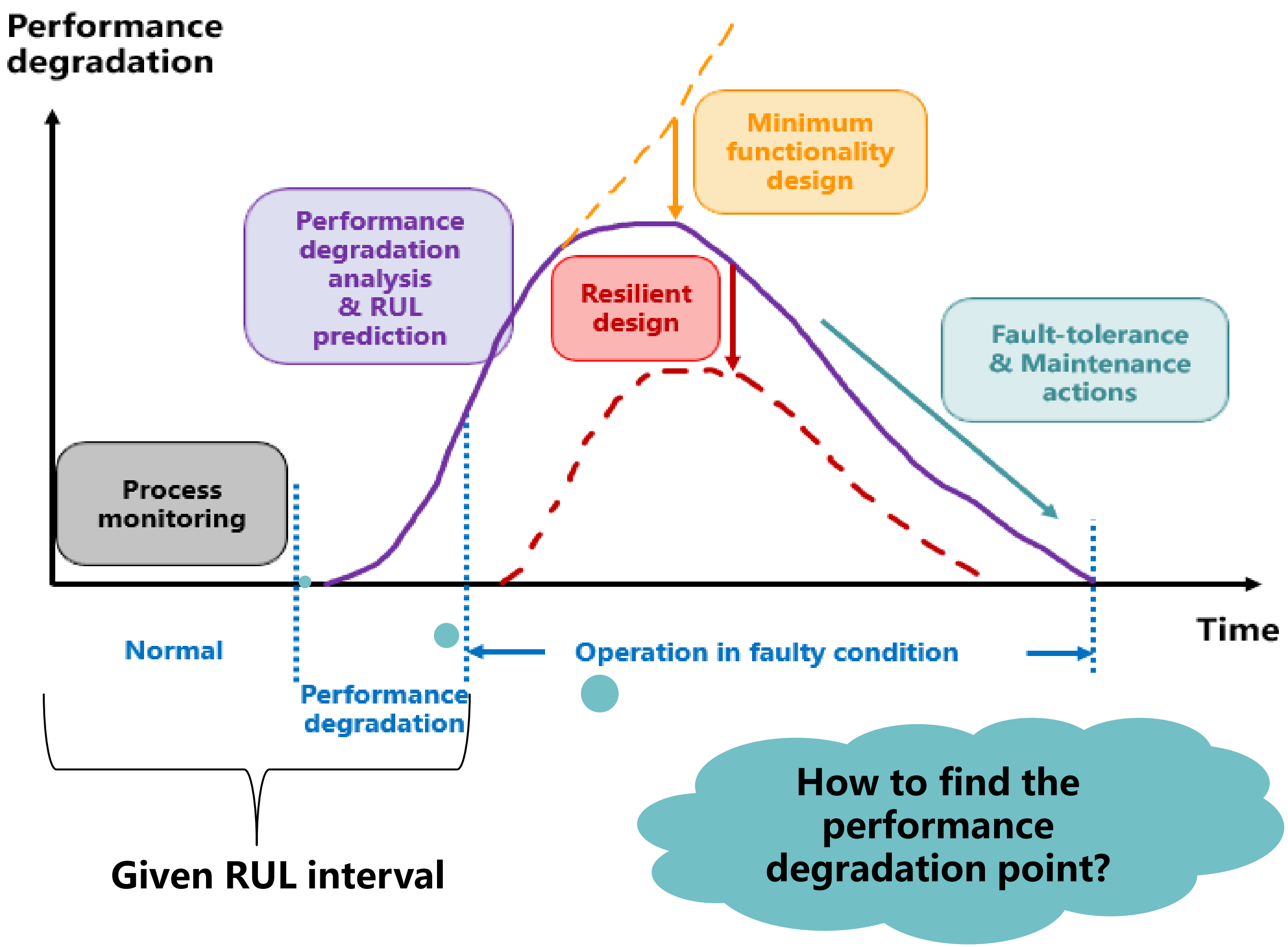


Fig. 4. RUL prediction results of a random run-to-failure sequence without degradation warning.

Find the early degradation point



Early Degradation Warning

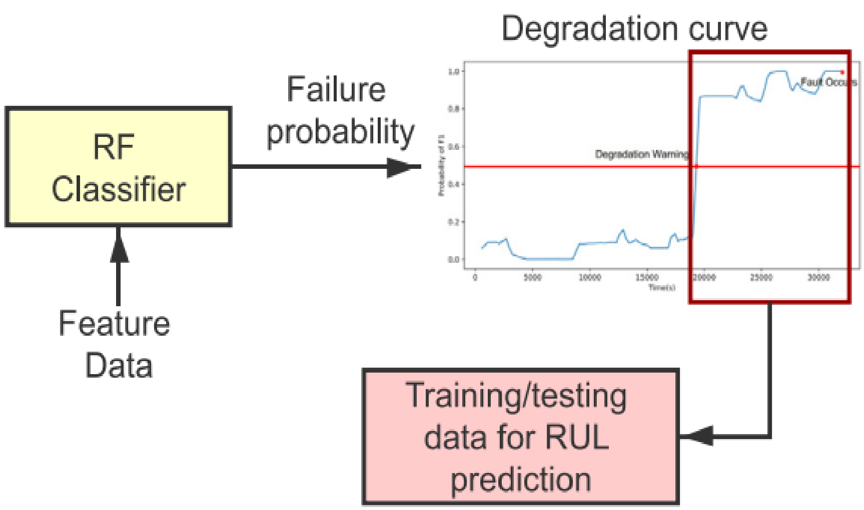


Fig. 3. Degradation warning procedure.

Random Forest (RF) classifier

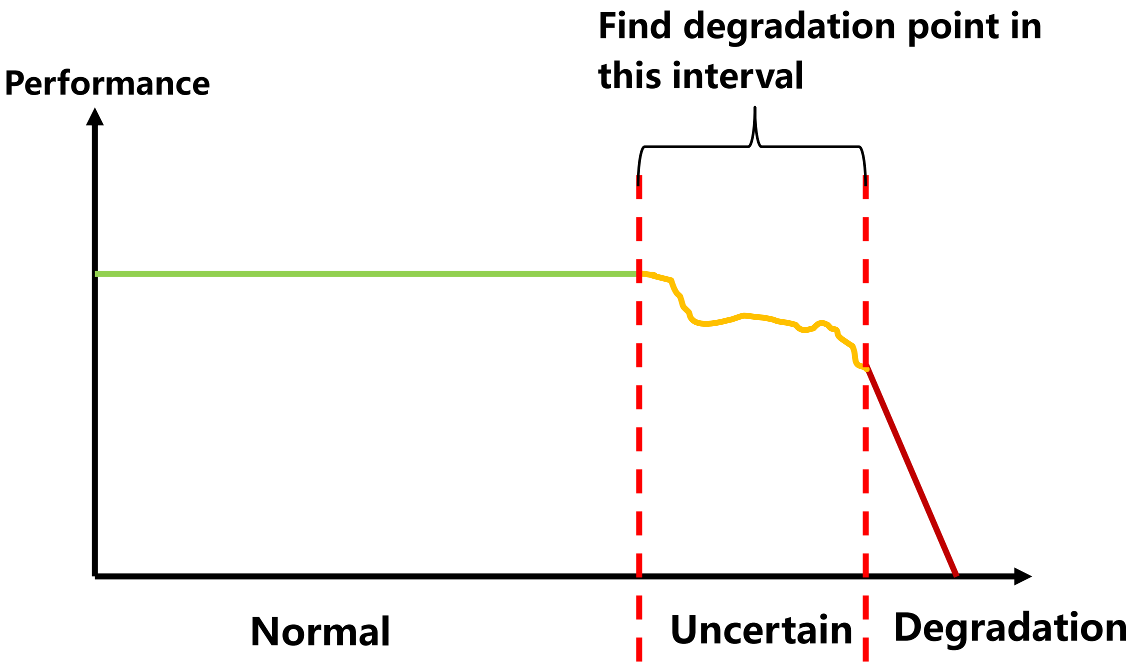
- **Input : the features of each run-to-failure sequence**
- **Output : is the probability of failure at each time point**
- **Degradation point: when the failure probability is larger than 0.5**

RUL less than 500 (about half an hour) are fault data, while the data with RUL larger than 6250 (about 6 h) are normal.

Scientific problem 1

HARBIN INSTITUTE OF TECHNOLOGY

Unbalanced normal and fault data



- A significant difference between the number of original fault samples and normal samples
- Make the model focus on the category with more samples

Up-sampling

	F1	F2	F3
up-sampling factors	500	1000	200

Early Degradation Warning

Table 3
Classification accuracies of RF models.

Fault mode	F1	F2	F3
Train	100%	100%	100%
OOB_score	99.88%	99.53%	99.91%

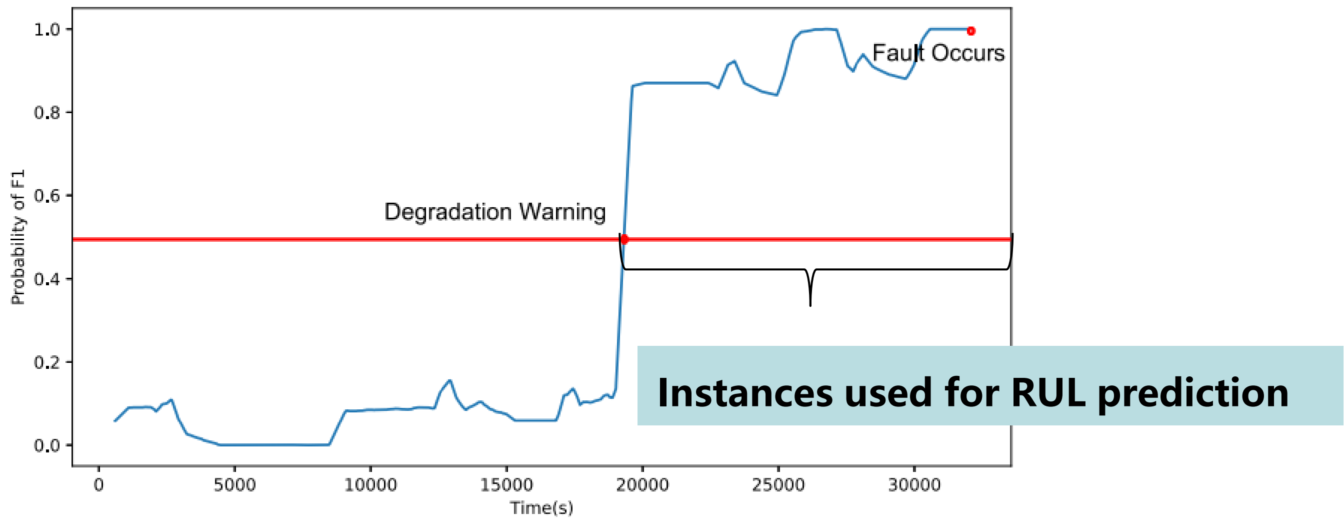


Fig. 7. Degradation warning results of a random run-to-failure sequence in F1 mode.

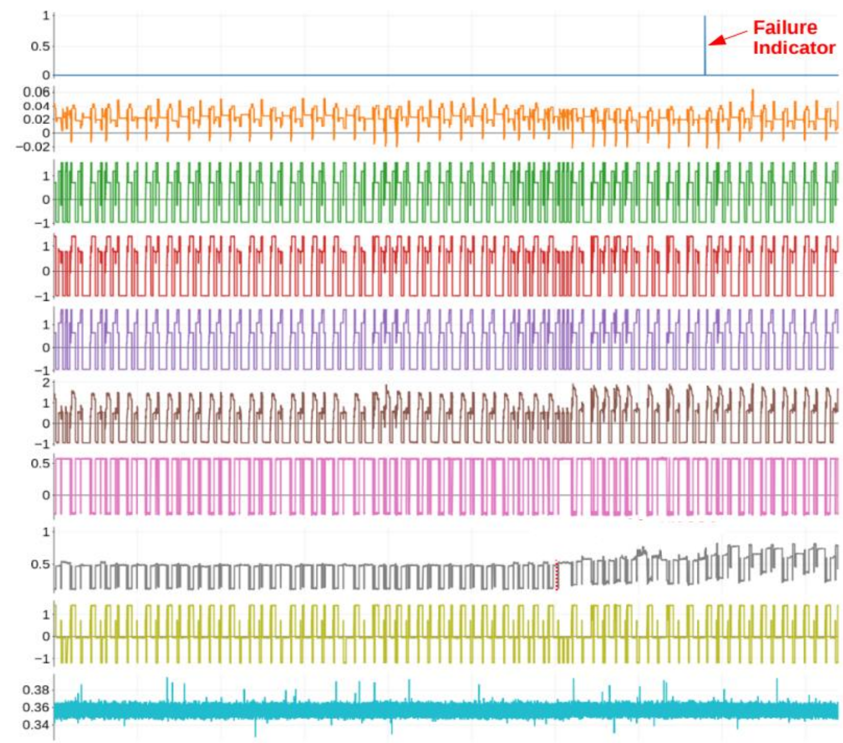
Outline:

- Basics and subject of study
- Scientific problem 1
- Scientific problem 2

Scientific problem 2

The nature of evolution of faults over time is not known & Multiple operating conditions

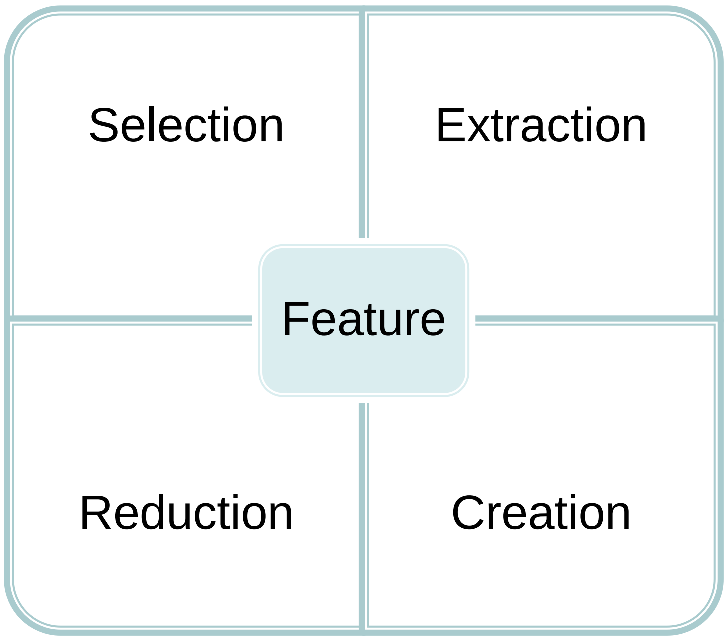
- There is no significant sign of degradation before failure
- 5 categorical parameters and 19 sensor parameters
- Has more than 500 different classification values



(b) F_2 : Flowcool Pressure Too High Check Flowcool Pump (FCP High)

Scientific problem 2

Feature engineering



Feature selection:

- **To measure the dependency between each feature and the objective RUL , the mutual information (MI) was estimated.**
- **Higher values of MI mean higher dependency.**

Scientific problem 2

HARBIN INSTITUTE OF TECHNOLOGY

New feature construction based on MLP :

- There is a relationship between the flow rate and the pressure in the flowcool
- This relationship changes when there is a failure.

$$\hat{p} = f(q)$$

$$r = p - \hat{p}$$

Actual flow rate

Estimated pressure

- The MSE of the model reaches the order of 10^{-5}
- New feature improves the prediction accuracy

Table 1
Training and testing results of the pressure model.

Dataset	Training	Validation	Testing
MSE(10^{-5})	8.9310	8.8721	8.8931

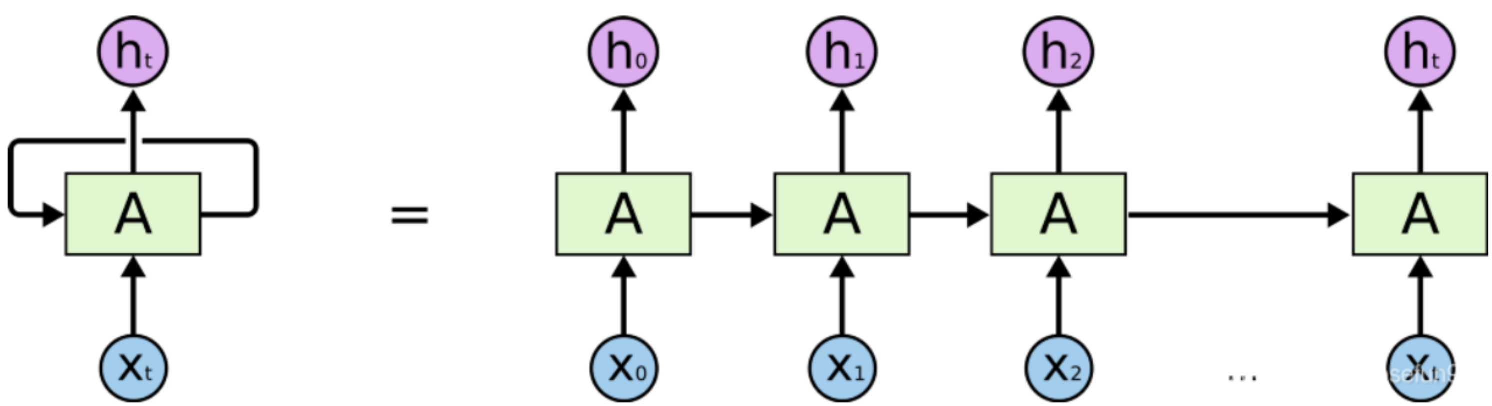
Table 2
Improvement of adding the new feature.

	RMSE (s)			SMAPE		
	F1	F2	F3	F1	F2	F3
Original feature	1074.92	748.44	1778.02	19.51%	18.06%	39.97%
New feature	954.78	489.70	1014.84	17.84%	15.04%	26.04%
Improvement	46%	35%	6%	9%	17%	35%

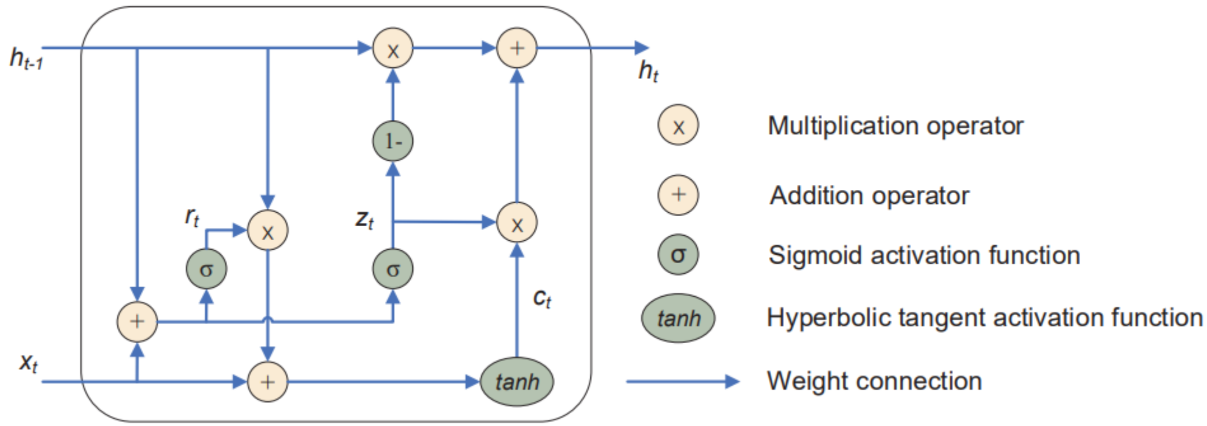
Scientific problem 2

Recurrent neural network (RNN)-based prediction scheme

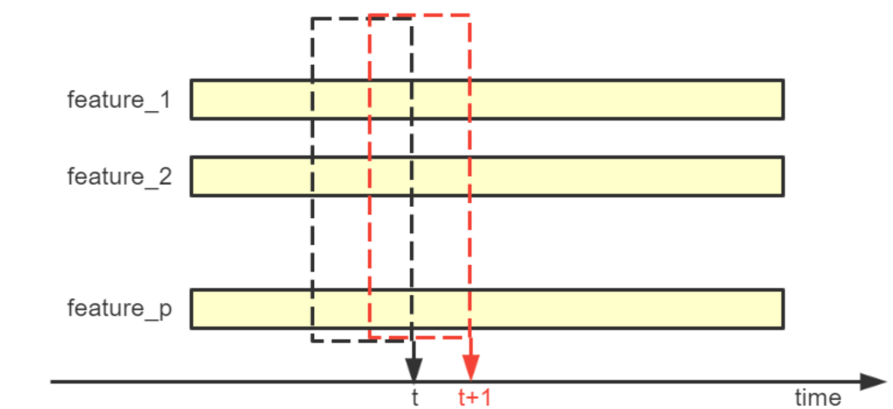
- **MLP cannot model the changes in sequences**
- **RNN can learn the relationship between the historical inputs and the current output**
- **Too many neural nodes and layers will lead to the vanishing of gradients in the process of propagation**



Gated recurrent unit (GRU) based prediction scheme



The hidden layer state is jointly ascertained by **update gate** and **reset gate**



Sliding time-window processing

- Change the discrete samples into the **time series**
- The objective output is the true RUL of the **last time step** in the time window sample.

哈爾濱工業大學
HARBIN INSTITUTE OF TECHNOLOGY

Gated recurrent unit (GRU) based prediction scheme

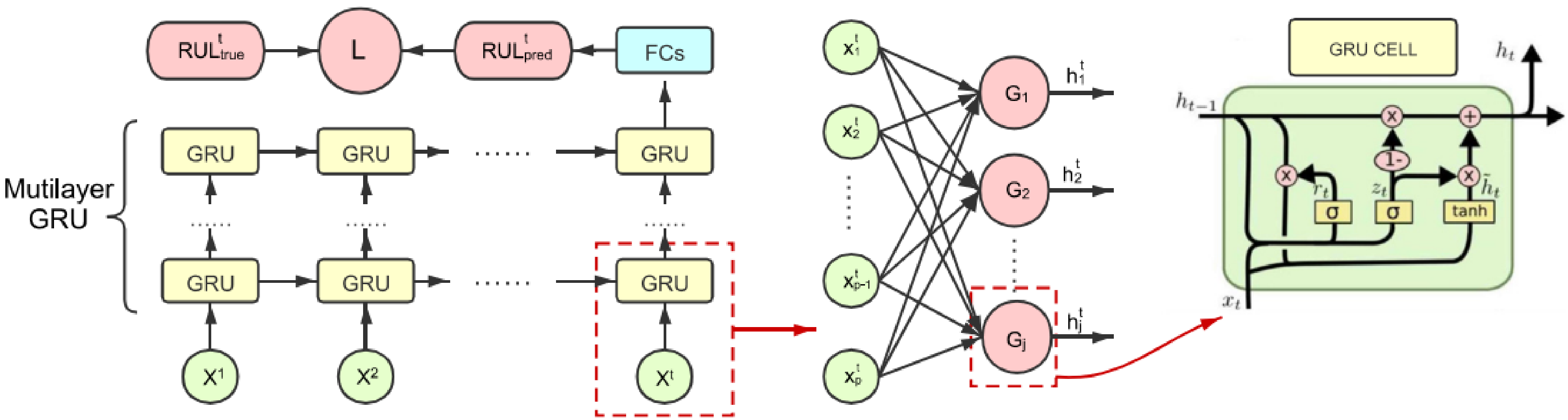


Fig. 5. GRU structure for RUL prediction.

- **GRU-FCs model is composed of multilayer GRUs followed by multilayer fully connected networks**
- **The training goal is to minimize the MSE between predicted RUL and groundtruth**
- **The FCs are good at integrating GRU features to regression outcomes.**

Gated recurrent unit (GRU) based prediction scheme

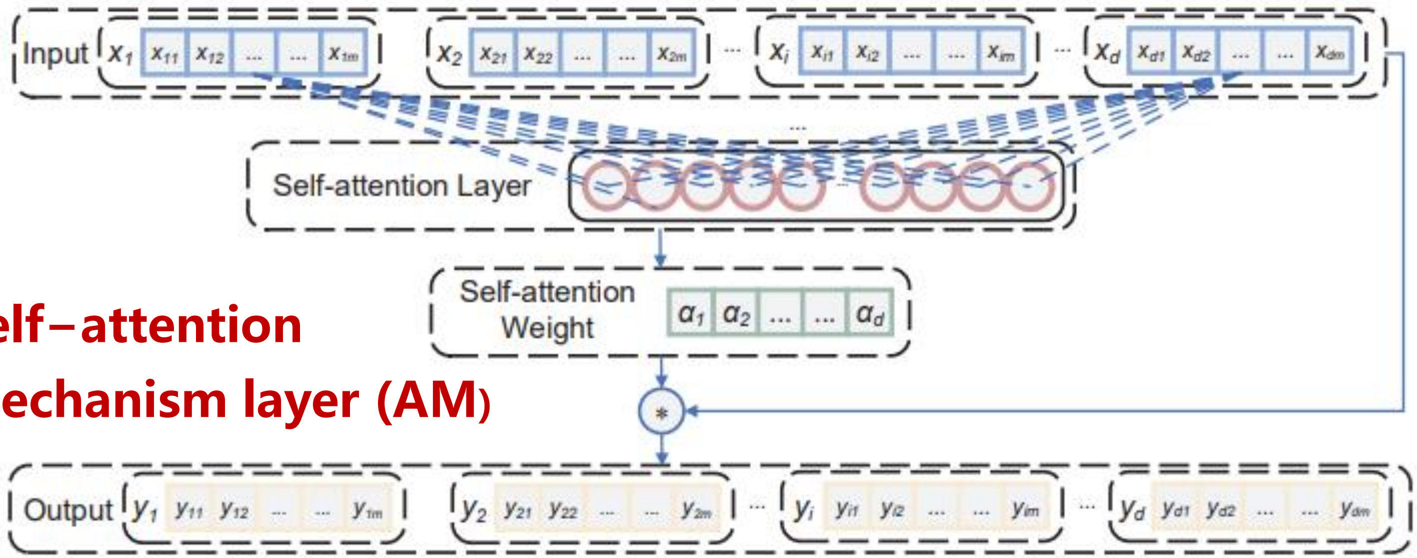
Table 8
Comparison with the existing works using the flowcool dataset.

Method	RMSE (s)			SMAPE		
	F1	F2	F3	F1	F2	F3
RFR (Gupta et al., 2018)	5476	5567	5294	59.84%	60.32%	58.54%
DW-SVR	5001	3185	5049	57.79%	50.34%	69.07%
MLP (Gupta et al., 2018)	5196	4113	5004	58.22%	47.56%	56.20%
DW-CNN	2184	546	1045	29.96%	16.92%	27.05%
LSTM (Gupta et al., 2018)	1469	2557	1877	23.48%	33.88%	27.80%
DW-LSTM	1849	618	1203	23.56%	19.63%	29.62%
DW-GRU	**686**	489	1014	**14.13%**	15.04%	26.04%
DW-GRU-FCs	703	**409**	**988**	14.53%	**14.44%**	**25.63%**
Improvement[a]	53.30%	84.00%	47.36%	39.82%	57.38%	7.81%

[a]Improvement of the proposed models in this paper over existing optimal predictive models for flowcool system, $Improvement = 1 - (Index_{\text{our model}}/Index_{\text{LSTM}})$.

HARBIN INSTITUTE OF TECHNOLOGY

Further improvement in feature engineering



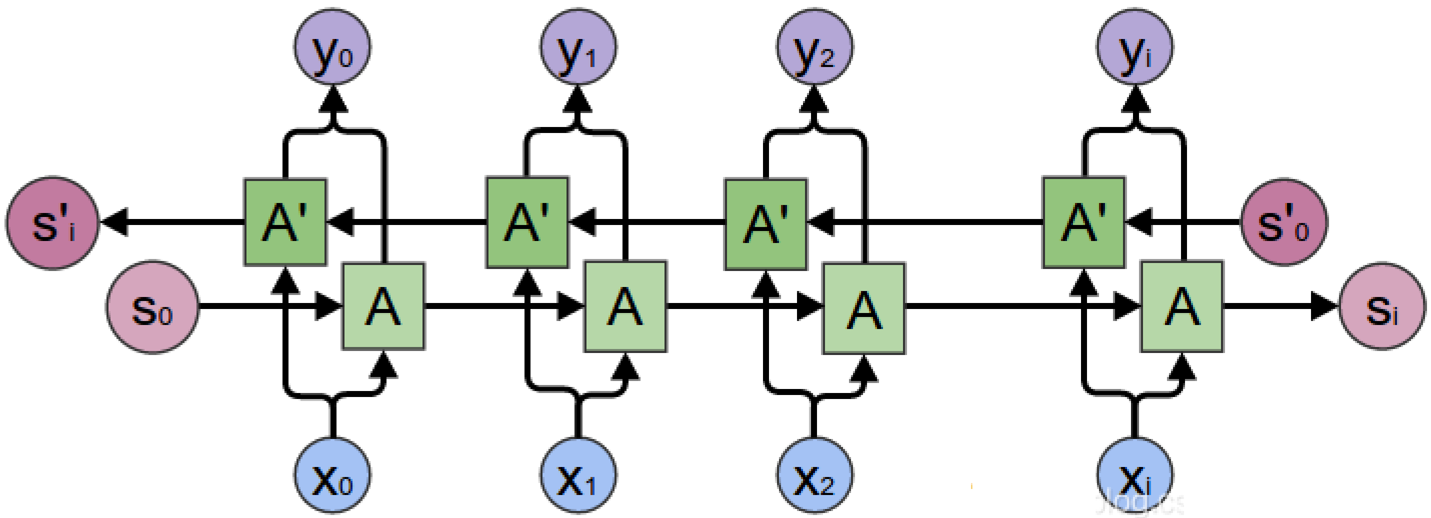
Self−attention mechanism layer (AM)

- **Reducing attention to useless information**
- **Different time instances to RUL is determined so that corresponding weights are assigned**

Prediction of remaining useful life via self-attention mechanism-based convolutional long short-term memory network. 31st European Safety and Reliability Conference, Angers, France 19-23 September 2021

Scientific problem 2

Further improvement in prediction model

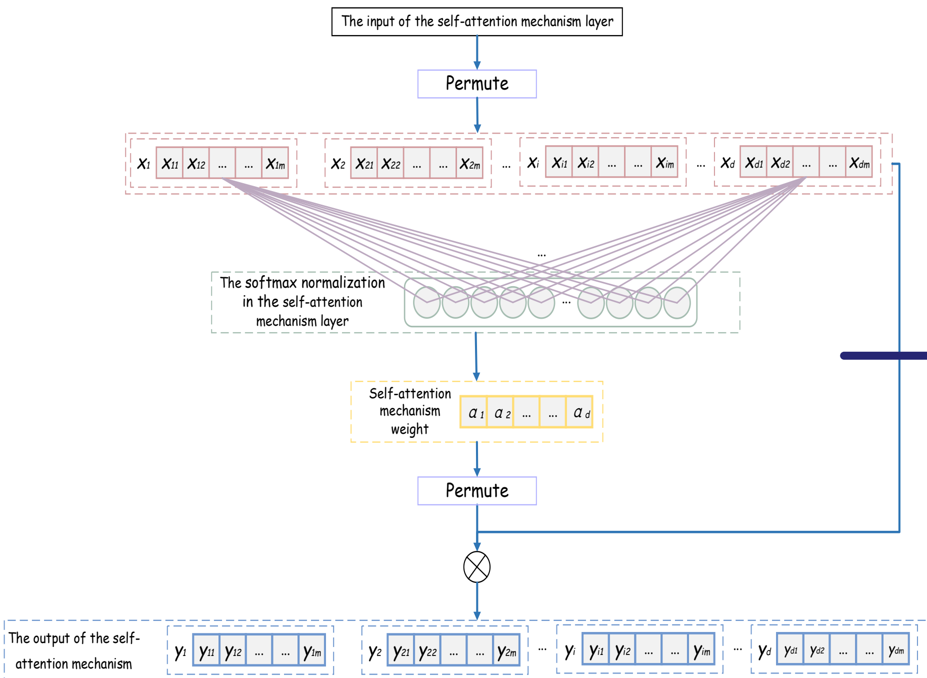


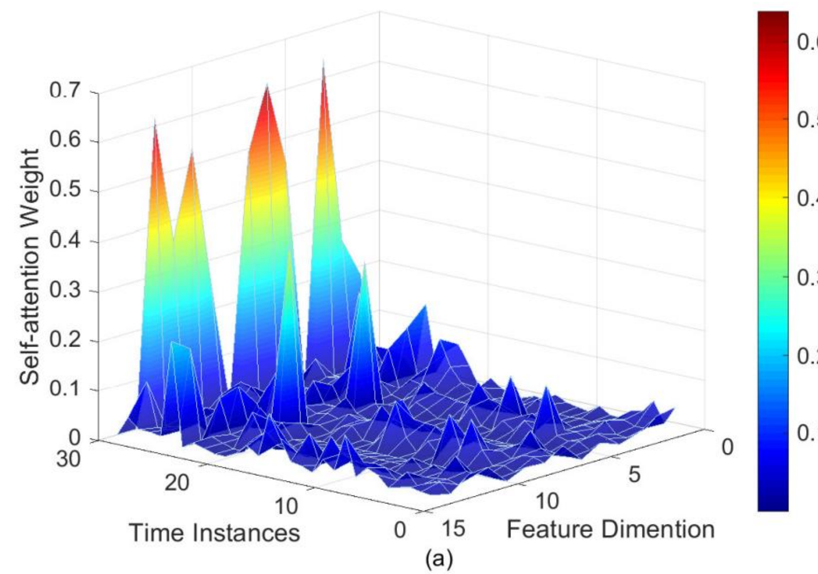
Bi-directional GRU (Bi-GRU)

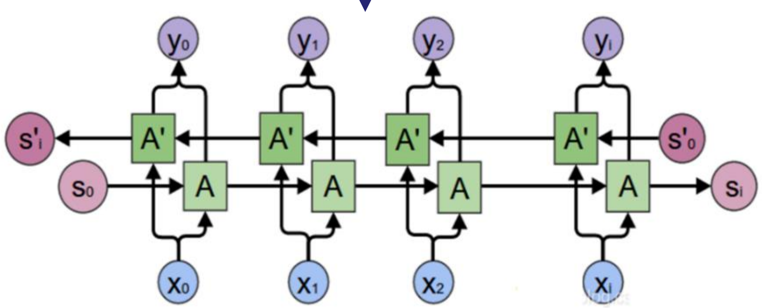
- Develop two different recurrent layers to montage the hidden layer state from the forward and backward orientations
- Study the reverse time series

Prediction of remaining useful life based on bidirectional gated recurrent unit with temporal self-attention mechanism. (under reviewer)

Scientific problem 2







- Learn the significance of the different **features** and **time instances**
- The larger self-attention weights are assigned to the latter time instances

Bi-GRU: Extract the temporal feature information

Content

Discussion

Two sides of a double-edged sword

Stochastic process-based approaches

Favorable points

① Precise mathematical model to describe degradation precisely
② Analytical solutions with reliability evaluation (confidence degree)
③ Component-level RUL prediction
④ Small online computing burden

Unfavorable points

1. Much dependency on domain experts
2. Hard to deal with complex systems
3. Wastes information in historical data
4. Application-specific design
5. Many assumptions to be met
6. Hard to interact/make use of other instances

ML-based approaches

Unfavorable points

1. Lack of interpretability
2. High demand of information infrastructure (computing, storage, data acquisition)
3. Need high-quality data for training
4. Need data processing for different cases
5. Wastes information in domain knowledge

Favorable points

① Little need of principle knowledge or mechanism knowledge
② Achieve system-level RUL prediction
③ Make use of historical process data
④ Easy generalization to other systems
⑤ Automatic extraction of features
⑥ Learning from the fleet (transfer learning)

Two sides of a double-edged sword

Stochastic process-based approaches

ML-based approaches

There is no universally favorable approach for all practical problems.

Future work

1) Evaluate the performance of different types of approaches with the same benchmarks from various perspectives

Uncertainty factors

- Various system types
- Working conditions
- Availability of sensor data
- Different scales/levels of the problem

New evaluation indexes for fair comparison:

- Computational complexity
- Interpretability, tractability
- Tunable parameters for optimization
- …

2) Integration of model, data, and knowledge

Thank you for listening!