IAIA Project #1

Drying Process RUL Estimation & Classification using ML

School of Mechanical and Control Engineering

Jin Kwak(21900031)

Sunwoo Kim(22000090)

24.10.29

Industrial AI and Automation

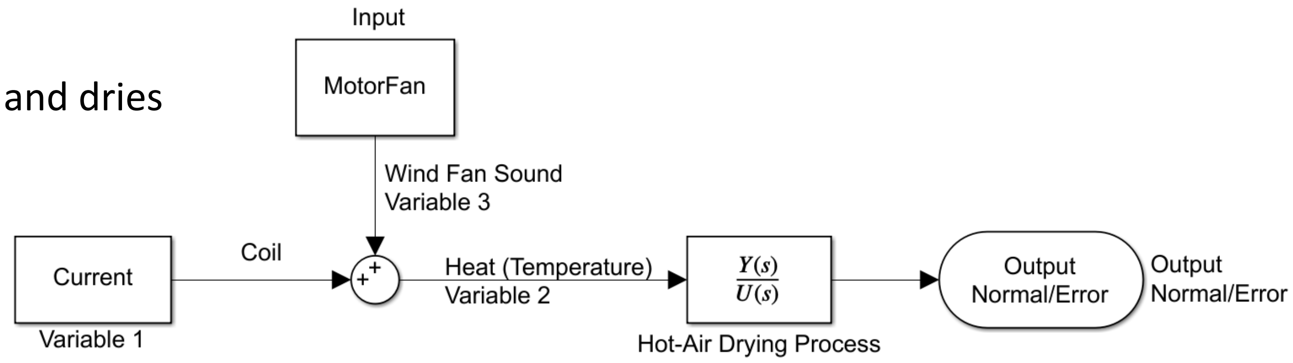
Prof. Young-Keun Kim

한동대학교
HANDONG GLOBAL UNIVERSITY

1. Introduction

1) Background

- Hot-Air Drying Process stabilizes metal surface after plating and dries moisture from the product’s surface.
- Generates hot air through coils and motor.



2) Problem

- Errors throughout the process

→ Classification of 12 states(normal and 11 errors)

→ Prediction of Remaining Useful Life(RUL) of equipment

3) Goal of this project

- Development of RUL model that analyzes RUL trend
- Classification model for 12 states

4) Specific Goals

	RUL Model	Classification
Performance	80% Accuracy within α bound after train-test breakpoint	Over 90% F1-Score

2. Dataset

1) [AI dataset for early detection of equipment abnormalities]
2) Data Measurement
   - PLC data (Normal: 1346 / Error: 87 / Total: 1419)
   - Sound data (Normal: 170 / Error: 13 / Total: 183)
   - PLC data ↔ Sound data (Asynchronous set)
3) Data Categories
   - PLC data: **Normal**(0) or **11 types of errors**(1-11)
   - Sound data: Normal / Error
4) Limitation on the PLC data
   - Sampling Frequency ↓($F_s = 0.2Hz$)
   - Various amount of data per process

     → Some processes **do not have** enough information
   - Defect frequently occurs → **Ambiguous to estimate RUL**

	Attribute	Explain
Variables	**Index**	Auto-generated value when collecting data
	Process	Tracking the process by assigning the same number to the same process
	Time	Recording time down to seconds in the format (H:MM:SS)
	Temp	Temperature within the hot-air drying system
	Current	Current within the hot-air drying system
Label	**0-11**	Status Number

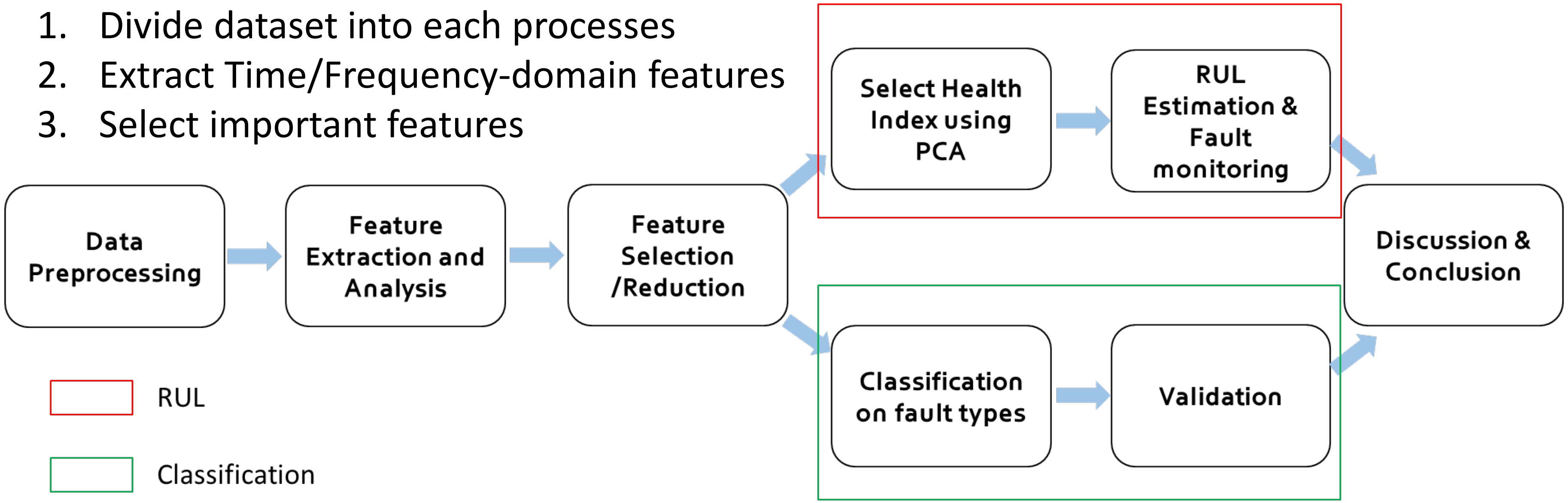
- FAN_sound_error
  - FAN_sound_error_01.wav
  - FAN_sound_error_02.wav
  - FAN_sound_error_03.wav
  - FAN_sound_error_04.wav
  - FAN_sound_error_05.wav
  - FAN_sound_error_06.wav
  - FAN_sound_error_07.wav
  - FAN_sound_error_08.wav
  - FAN_sound_error_09.wav
  - FAN_sound_error_10.wav
  - FAN_sound_error_11.wav
  - FAN_sound_error_12.wav
  - FAN_sound_error_13.wav
- FAN_sound_OK
  - FAN_sound_01.wav
  - FAN_sound_02.wav
  - FAN_sound_03.wav
  - FAN_sound_04.wav
  - FAN_sound_05.wav
  - FAN_sound_06.wav
  - FAN_sound_07.wav
  - FAN_sound_08.wav
  - FAN_sound_09.wav
  - FAN_sound_10.wav
  - FAN_sound_11.wav
  - FAN_sound_12.wav
  - FAN_sound_13.wav

3. Methodology (1/2)

Strategy of the project

4.1.
- Select Health Index
- Set RUL model and validate the model using $\alpha - \lambda$ plot

1. Divide dataset into each processes
2. Extract Time/Frequency-domain features
3. Select important features



5.
- Compare with other methods
- Performance analysis

4.2.
- Set Classification models (SVM, KNN, Decision Tree)
- Draw Confusion matrix and compare performance

3. Methodology (2/2)

1) Time/ Frequency-domain based Feature extraction

I. Extract system Parameters & Statistical property of model parameters

2) Feature Selection/Reduction

I. Important technique to reduce dimensionality

II. Improve performance of parameters of the system

$$X_{\text{rms}} = \left(\frac{1}{N}\sum_{i=1}^{N} x_i^2\right)^{1/2} \qquad X_{\text{sra}} = \left(\frac{1}{N}\sum_{i=1}^{N} \sqrt{|x_i|}\right)^{2}$$

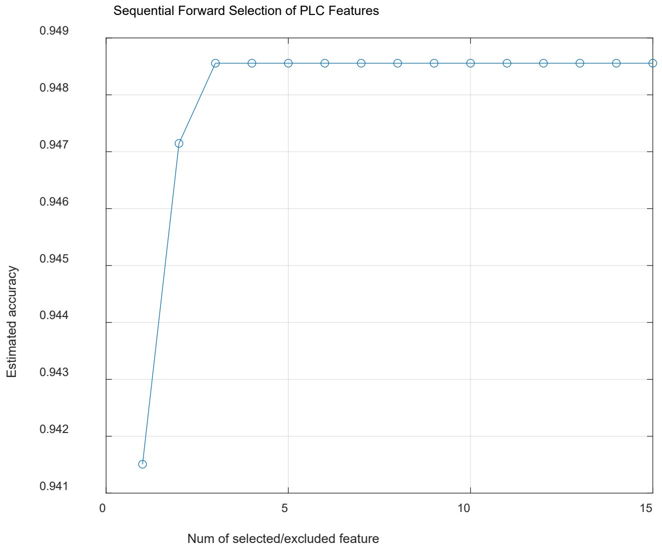
$$X_{\text{kv}} = \frac{1}{N}\sum_{i=1}^{N}\left(\frac{x_i - \overline{x}}{\sigma}\right)^4 \qquad X_{\text{sv}} = \frac{1}{N}\sum_{i=1}^{N}\left(\frac{x_i - \overline{x}}{\sigma}\right)^3$$

$$X_{\text{ppv}} = \max(x_i) - \min(x_i) \qquad X_{\text{cf}} = \frac{\max(|x_i|)}{\left(\frac{1}{N}\sum_{i=1}^{N} x_i^2\right)^{1/2}}$$

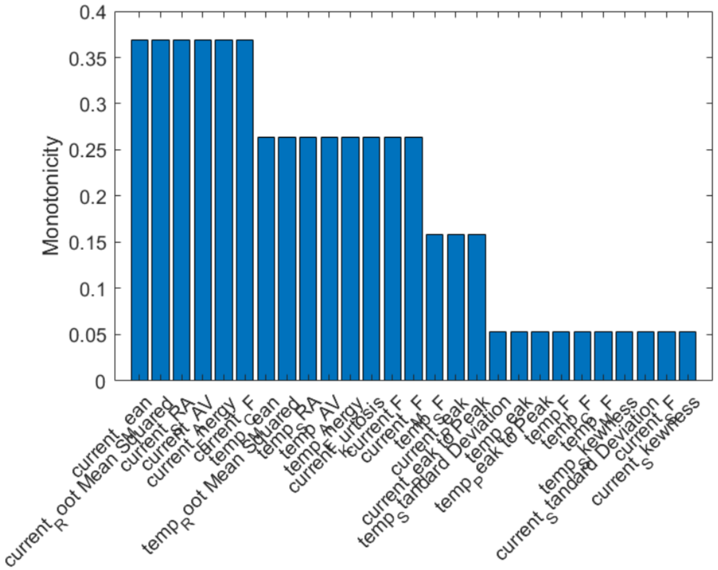
$$X_{\text{if}} = \frac{\max(|x_i|)}{\frac{1}{N}\sum_{i=1}^{N} |x_i|} \qquad X_{\text{mf}} = \frac{\max(|x_i|)}{\left(\frac{1}{N}\sum_{i=1}^{N} \sqrt{|x_i|}\right)^{2}}$$

$$X_{\text{sf}} = \frac{\max(|x_i|)}{\left(\frac{1}{N}\sum_{i=1}^{N} x_i^2\right)^{1/2}} \qquad X_{\text{kf}} = \frac{\frac{1}{N}\sum_{i=1}^{N}\left(\frac{x_i - \overline{x}}{\sigma}\right)^4}{\left(\frac{1}{N}\sum_{i=1}^{N} x_i^2\right)^{2}}$$

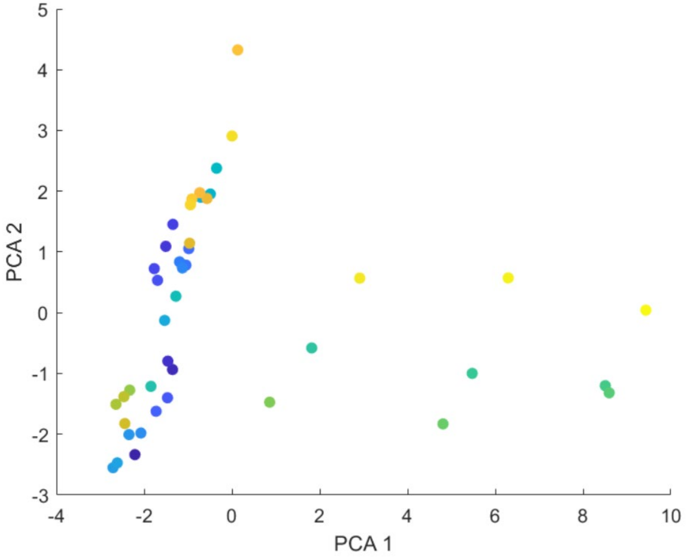
① Sequential Forward Selection



② Monotonicity



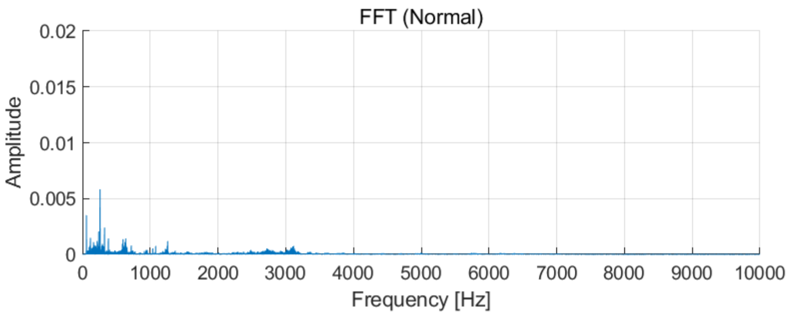
③ PCA

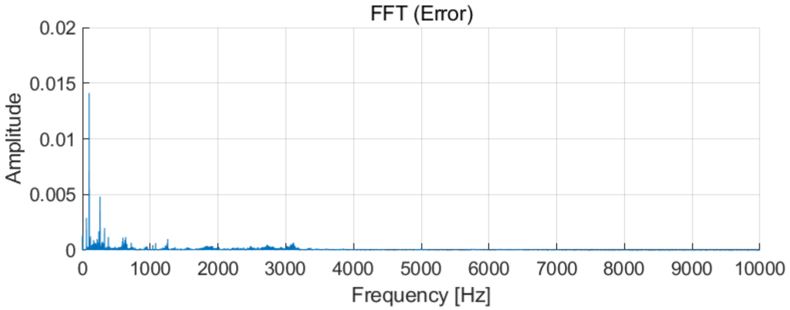


3.1. Classification (Sound Data)

3.1.1 Binary classification of two states (Normal/Error) using sound data

1) Binary Classification by sound data





2) Dataset

- Sampling Frequency: $44100\ Hz$

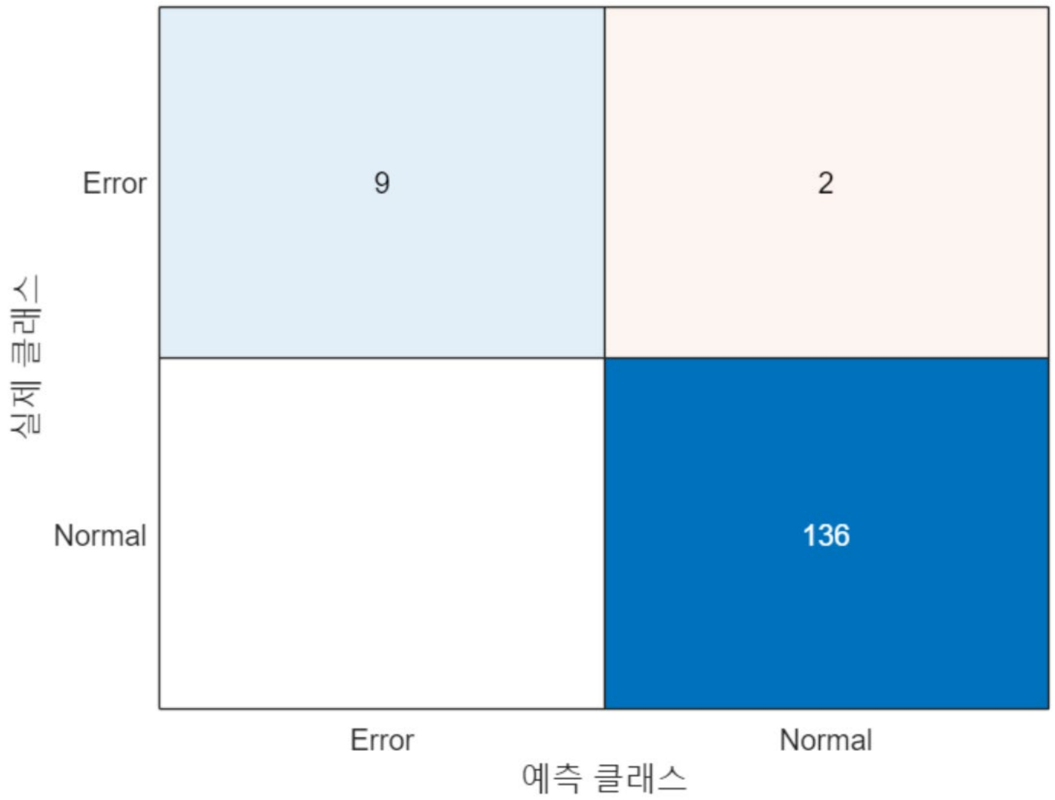
- # of Error Sound file: 13
- # of Normal Sound files: 170

3) Selection of Model

- Logistic Regression is suitable for binary classification
- Probability calculation using a logistic function

3.2. Classification Analysis

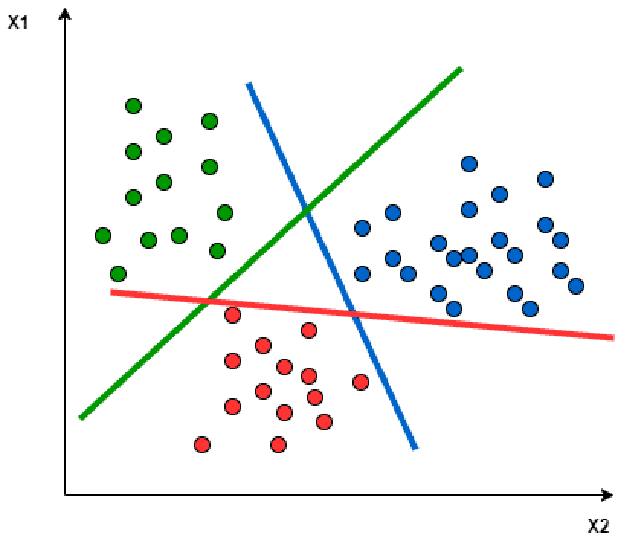
1) Sound Sensor Classification shows 95% of F1-Score
2) Through the binary classification, it good for classify the error & normal state obviously before multi-class classification.



4.1. Classification (PLC Data) (1/3)

3.1.2 Multi-class classification of PLC data

1) Support vector machines (SVM)

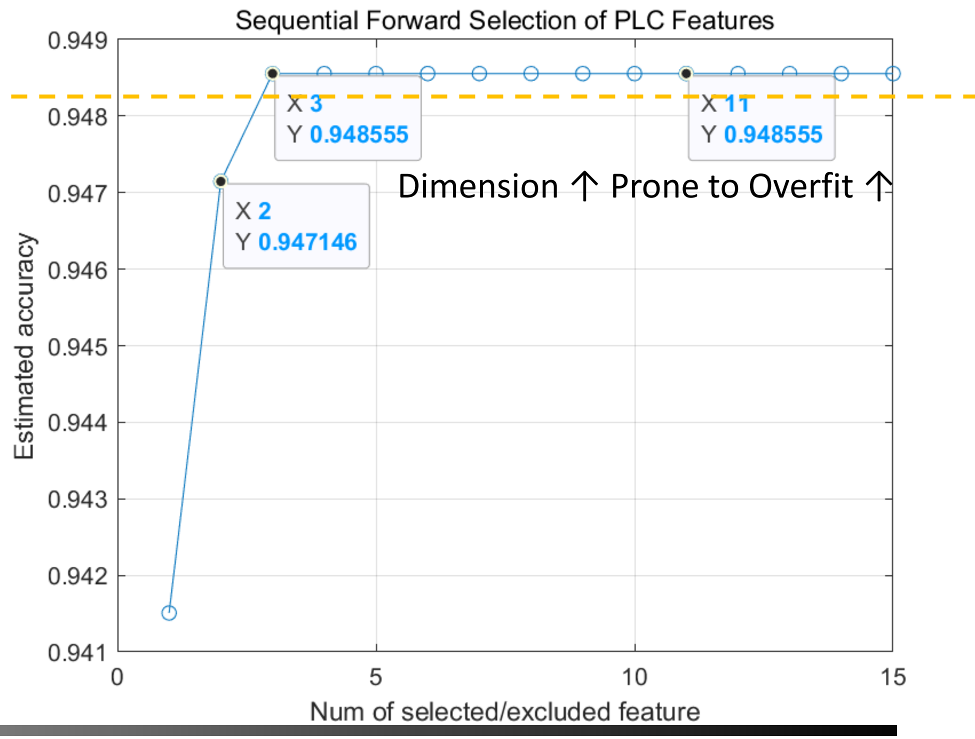


· Model: Multi-class SVM model

· Dimension: 3

· Reason for the hyperparameter :

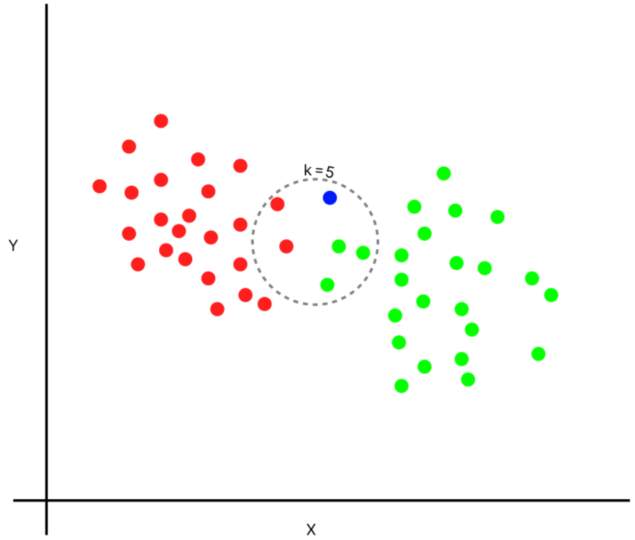
Among the dimensions with the **highest accuracy**, select the **lowest dimension(3)**.



4.1. Classification (PLC Data) (2/3)

3.1.2 Multi-class classification of PLC data

2) K-Nearest Neighbor (KNN)



- Neighbor: 3
- Reason for the hyperparameter :

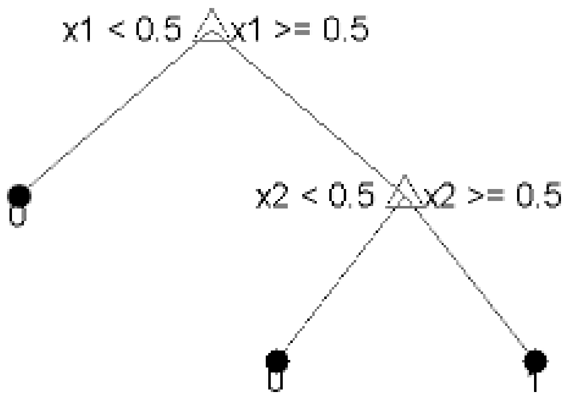
Using quantitative approach, the number of neighbor(K) that has highest accuracy is 3

Neighbor (K)	Accuracy
3	0.960
5	0.953
7	0.954

4.1. Classification (PLC Data) (3/3)

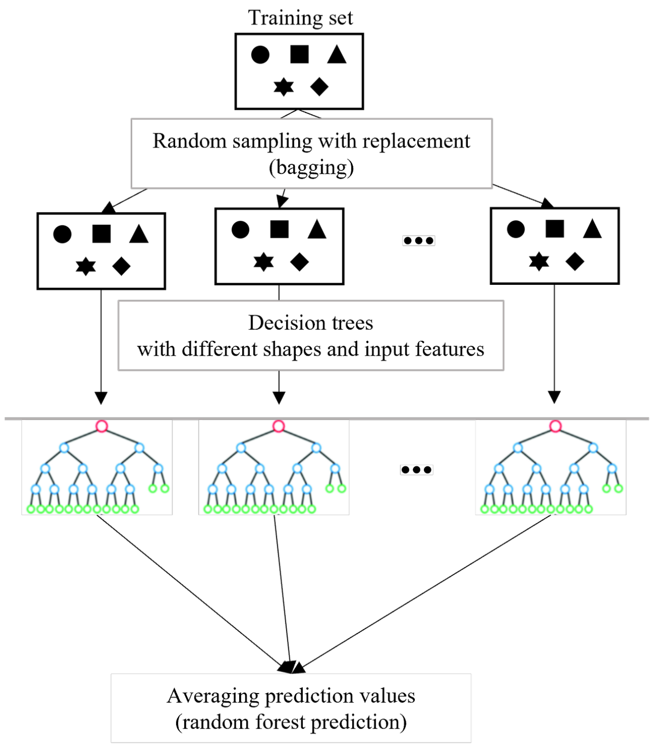
3.1.2 Multi-class classification of PLC data

3) Decision Tree



- Model: Decision Tree
- Branch: 24

4) Random Forest

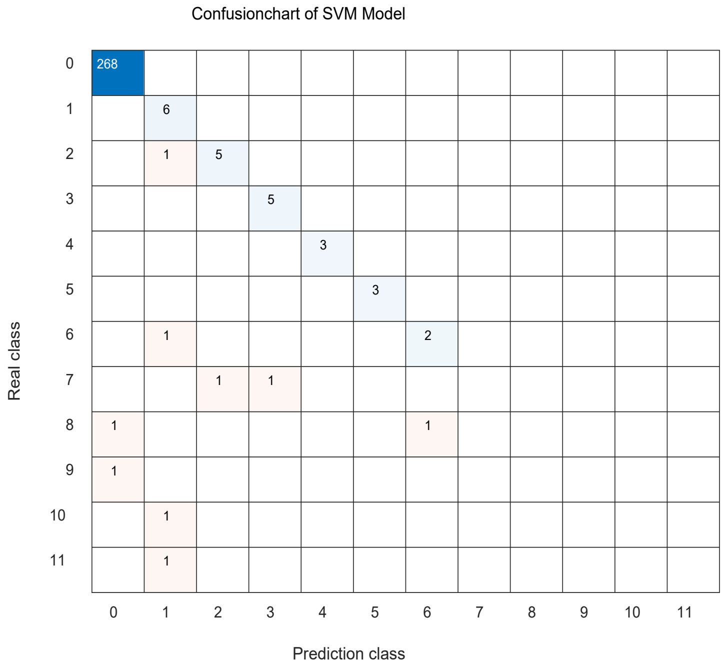


- Model: Random Forest
- Bagging decision tree: 50
- Reason for the hyperparameter : The well performed number of decision tree by experiment.

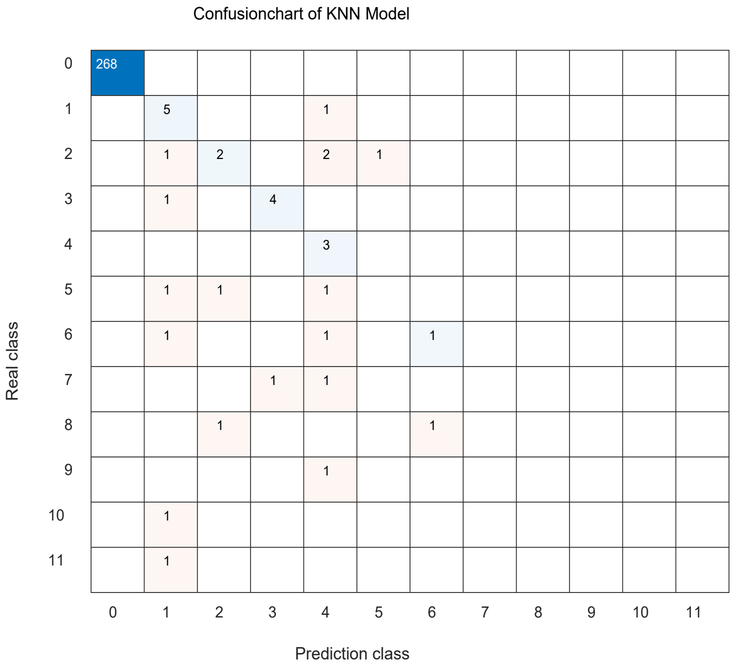
Tree	F1 Score
1	0.67
3	0.72
5	0.94
10	0.96
20	1

4.2. Classification Result (1/2)

1) Support vector machines (SVM)



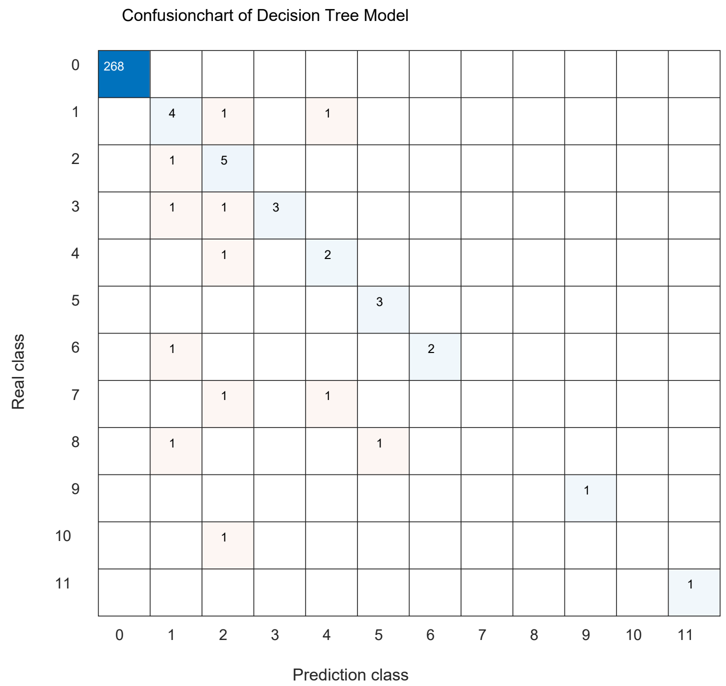
2) K-Nearest Neighbor (KNN)



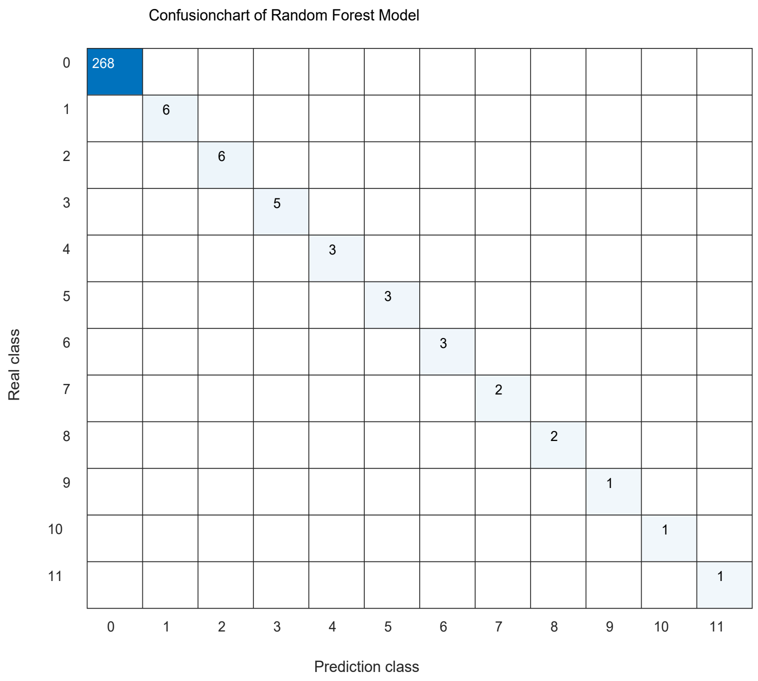
	SVM	KNN
Accuracy	0.97	0.94
Precision	0.53	0.32
Recall	0.58	0.39
F1-Score	0.55	0.35

4.2. Classification Result (2/2)

3) Decision Tree



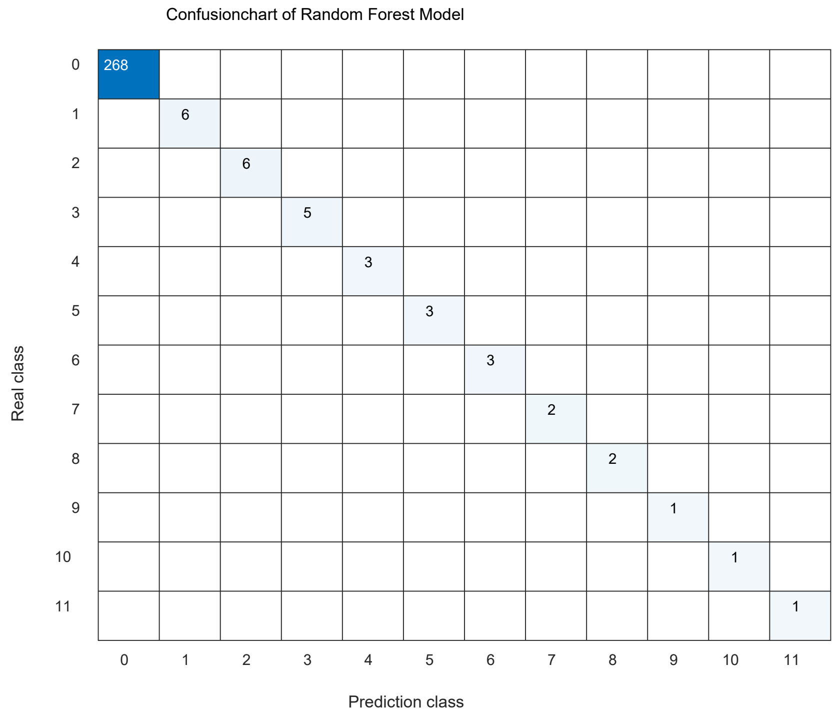
4) Random Forest (RF)



	Decision Tree	Random Forest
Accuracy	0.96	1.0
Precision	0.66	1.0
Recall	0.68	1.0
F1-Score	0.67	1.0

4.3. Classification Analysis

1) Performance of **Random Forest Models** is best among 4 ML models. Random Forest Model achieved **100% F1-Score**
2) Random Forest Model uses ensemble method, It combines multiple decision tree model and this prevents overfit.(Robust)
3) This data set has multiple classes, but not enough the number. In this case, multiple weak models for high accuracy as a result.



5.1. PHM (1/2)

1) Data with errors is selected and Time-domain features are extracted

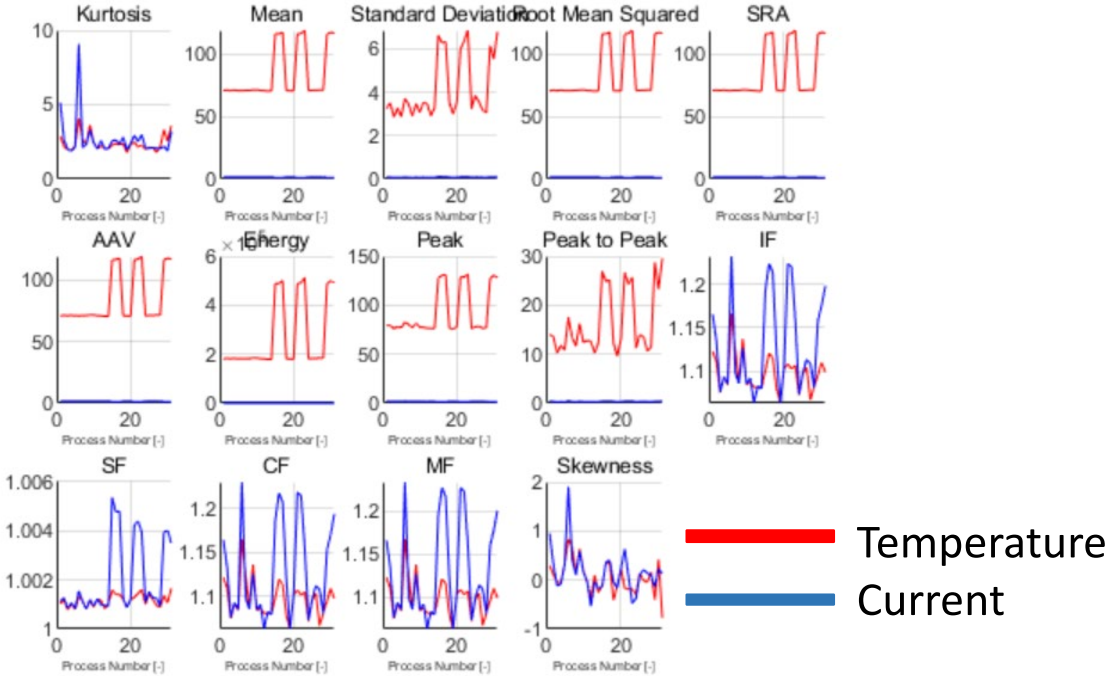
2) Earliest process with error is selected as break-point

   Last process with error is assumed to be the end of useful life

	1	2	3	4	5	6	7	8	9	10	11
2021-09-06	32	33	20	21	22	31					
2021-09-07	32	33	34								
2021-09-08											
2021-09-09	15	16	17	21	22	23	29	30	31		
2021-09-10	32	28	29	30	31						
2021-09-13	27	28	29								
2021-09-14											
2021-09-15	40	41	39								
2021-09-16	2	35	3	34	36						
2021-09-17	12	13	14	16	17	18	28	29			
2021-09-23	8	9	6	7							

Error Number

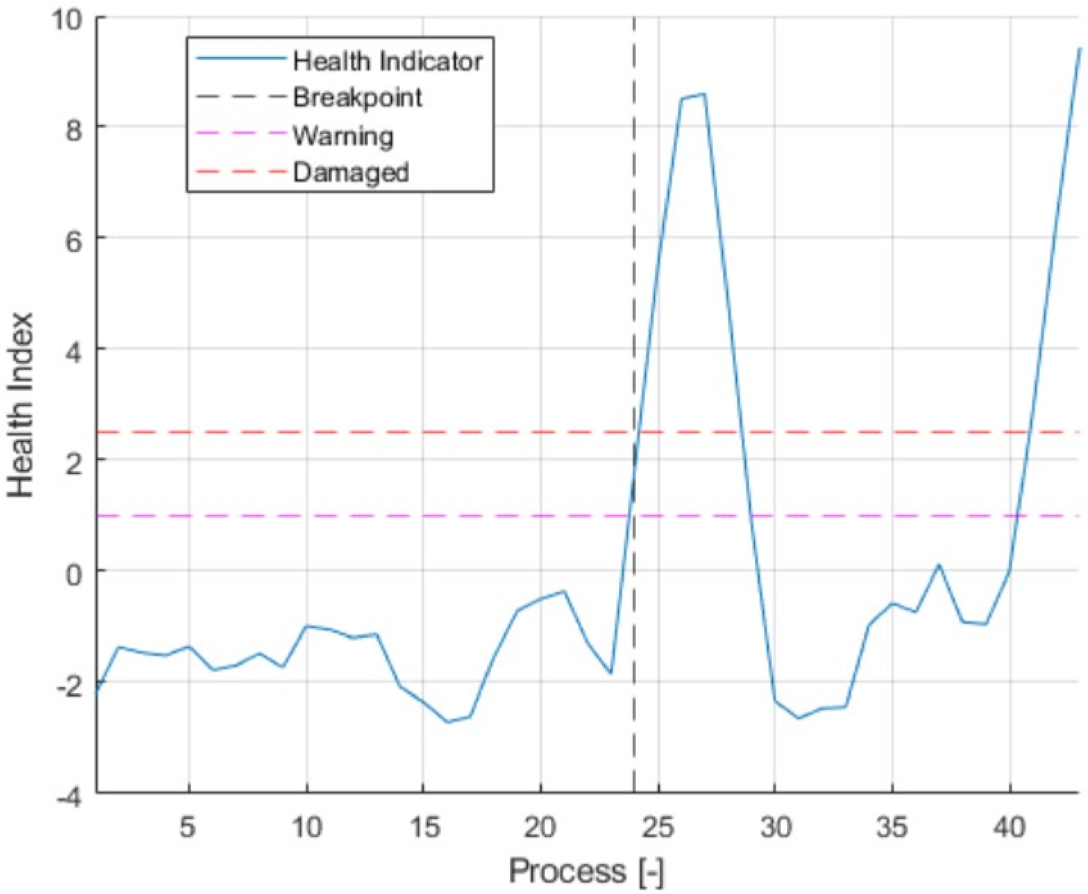
Process Number



5.1. PHM (2/2)

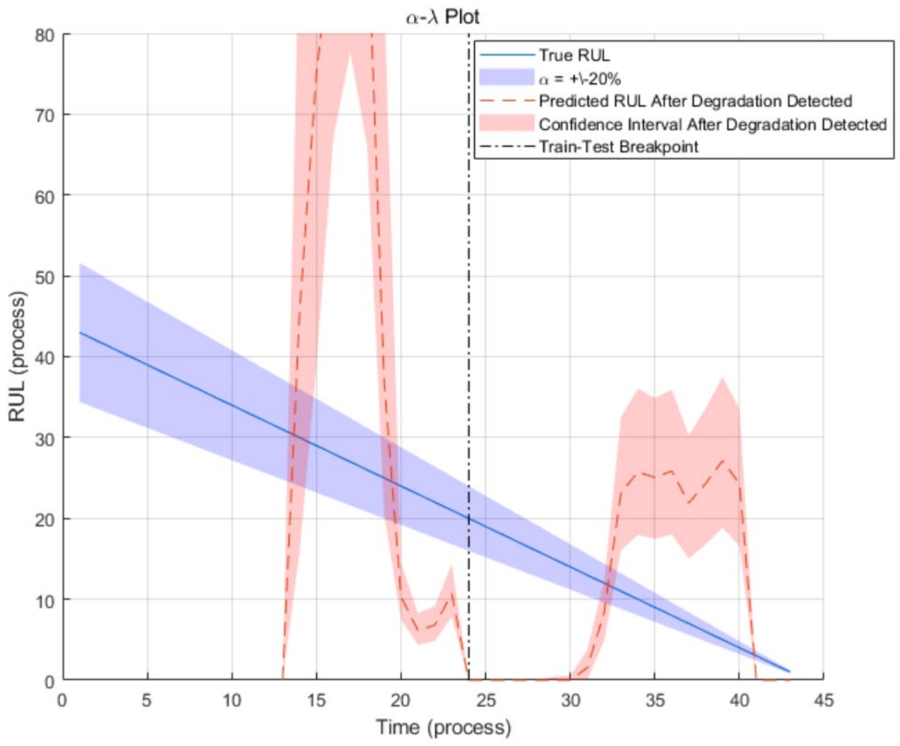
1) Health Indicator is selected for RUL estimation
2) Indicator tends to increase until the end
3) Limitation: Processes with error shows peak values of health indicator which disturbs degradation model

   → Use Moving Average Filter for smoothing



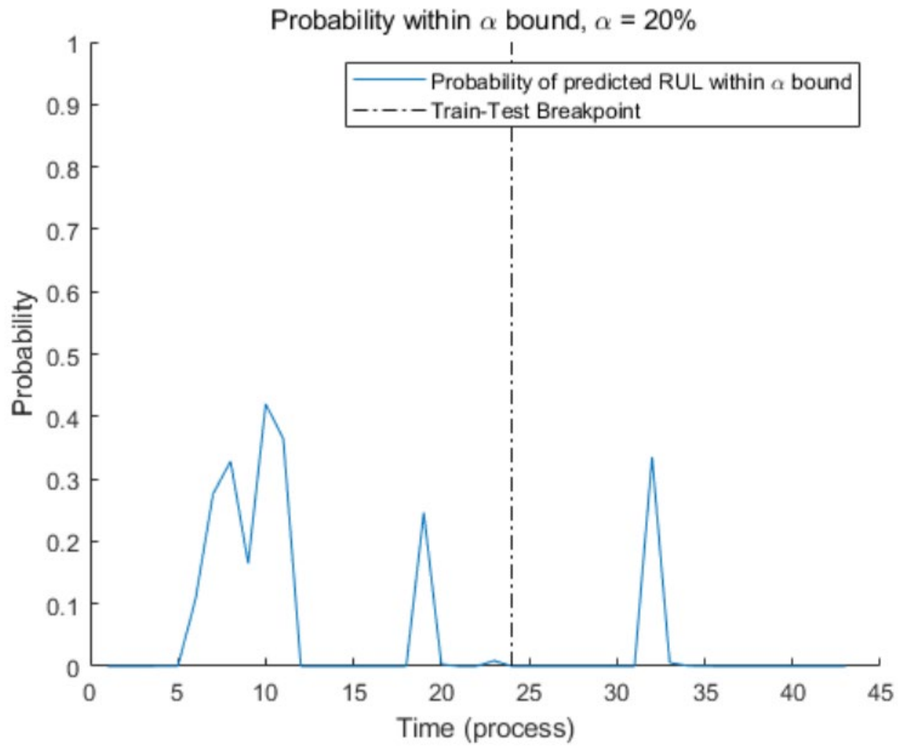
5.2. PHM Analysis

1) The $\alpha - \lambda$ plot



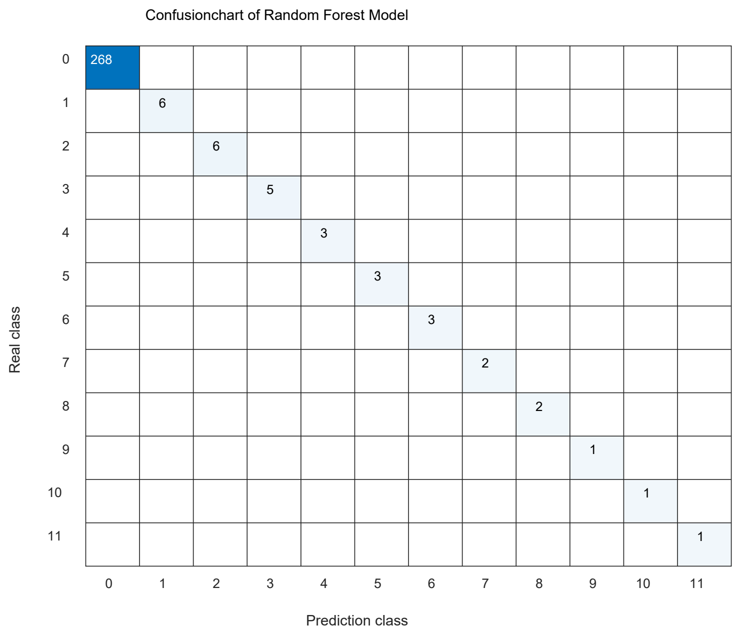
→ α band overlaps by at maximum about 30%

2) Probability of α bound



6. Result

1) Classification model

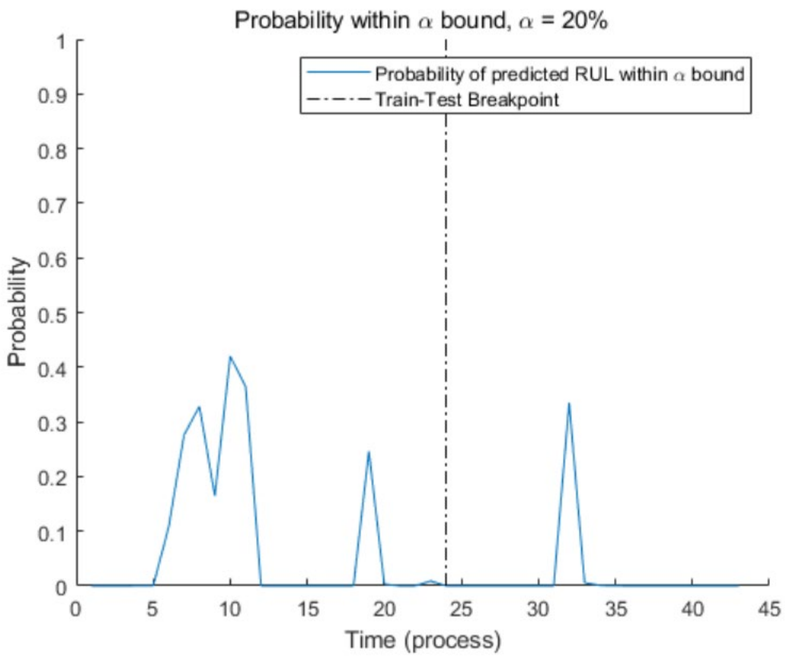


Random forest model

→ F1-score : 1

→ Expected Result $F1 - score > 0.9$

2) RUL model



RUL model

→ α bound does not overlap

→ Prediction equipment lifespan with over 80% accuracy

7. Conclusion

1. Classification model

   1) Decision tree model showed best performance in F1-score 0.67
   2) Followed by Trends in ML development research, Random Forest Model is improved.
   3) Improved model(Random Forest) satisfies expected goal ($F1 - score \geq 0.9$)
   4) More PLC data required → Model underfitting problem.
   5) 2 Classification Models combined gives robustness to the classification system.

2. RUL model

   1) Did not meet the expected result from proposal
   2) Sampling Time, Data size should be big enough for better performance

8. Paper Review (1/2)

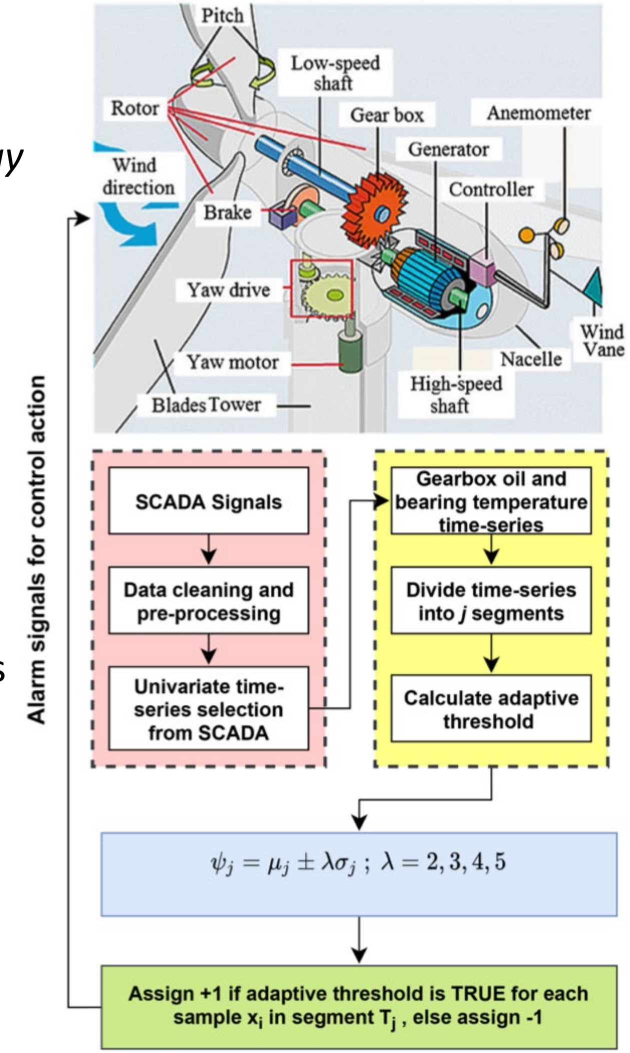
1) Paper: Dhiman, H. S., Deb, D., Muyeen, S. M., & Kamwa, I. (2021). Wind turbine gearbox anomaly detection based on adaptive threshold and twin support vector machines. *IEEE Transactions on Energy Conversion*, *36*(4), 3462-3469.

2) Background of the research

   1) Reduce the cost of O&M in wind turbine system, condition monitoring is important.

   2) Wind turbine shows error in the bearings(Rotary machine fault)

3) A condition monitoring (CM) system generally tracks the temporal behavior of variables in time-series such as wind speed, gearbox temperature, generator bearing temperature, nacelle temperature, and ambient temperature.



8. Paper Review (2/2)

4) Adaptive Thresholding

- Univariate time-series segmentation for analyzing the physical system's temporal behavior.

5) Twin SVM

- Two hyperplanes that satisfies each equations (Optimality Condition Problem
  → KKT, Chebyshev's inequality and Wolfe duality problem)

- $\min_{(w_1,b_1,p_1)} \frac{1}{2}\|X_+w_1 + e_1b_1\|_2 + C_1e_2^Tp_1$

- $\min_{(w_2,b_2,p_2)} \frac{1}{2}\|X_-w_2 + e_2b_2\|_2 + C_2e_1^Tp_2$

Classifier	Without Adaptive threshold		
	Accuracy	Precision	F1 score
SVM	86.52	87.55	88.07
MLPNN	87.17	88.21	88.73
KNN	89.28	90.35	90.88
DT	65.09	65.87	66.26
TWSVM	**94.26**	95.39	95.95
Classifier	**With Adaptive threshold**		
	Accuracy	**Precision**	**F1 score**
SVM	90.00	91.08	91.62
MLPNN	89.02	90.08	90.62
KNN	89.11	90.17	90.71
DT	90.21	91.29	91.88
TWSVM	**95.94**	97.13	97.74

6) Result of **Twin SVM** with/without **Adaptive Threshold**

By using Adaptive Thresholding and Twin SVM method,

the paper improved time-efficiency and accuracy for real-time approach

Classifier	Without **AT**	With **AT**	Computation time (sec)
SVM	88.47	90.55	2.05
MLPNN	87.01	90.12	1.93
KNN	89.36	89.68	2.36
DT	75.64	88.62	2.81
TWSVM	**94.27**	**95.84**	**1.63**

9. Reference

[1] ㈜ KEMP, Innozinc 세라믹 아연도금, https://www.kampai.kr/

[2] Dhiman, H. S., Deb, D., Muyeen, S. M., & Kamwa, I. (2021). Wind turbine gearbox anomaly detection based on adaptive threshold and twin support vector machines. *IEEE Transactions on Energy Conversion*, *36*(4), 3462-3469.