IAIA Project #1

Drying Process RUL Estimation & Classification using ML

School of Mechanical and Control Engineering

Jin Kwak(21900031)

Sunwoo Kim(22000090)

24.10.22

Industrial AI and Automation

Prof. Young-Keun Kim

한동대학교
HANDONG GLOBAL UNIVERSITY

1. Introduction

1) Problem

- Errors throughout the process
  - → Classification of 12 states(normal and 11 errors)
  - → Prediction of Remaining Useful Life(RUL) of equipment

2) Goal of project

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

3) Specific Goals

- Prediction equipment lifespan with over 80% accuracy within the α bound of the predicted RUL after the train-test breakpoint.
- By using (90% F1 score) Classification models(KNN, SVM …etc), condition check for the equipment.

2. Dataset

1) AI dataset for early detection of equipment abnormalities
2) Data Measurement
   - PLC data
   - Sound data
3) Data Categories
   - Sound data: Normal / Error
   - PLC data: **Normal**(0) or **11 types of errors**(1-11)
4) Limitation on the data
   - Sampling Frequency ↓
   - 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

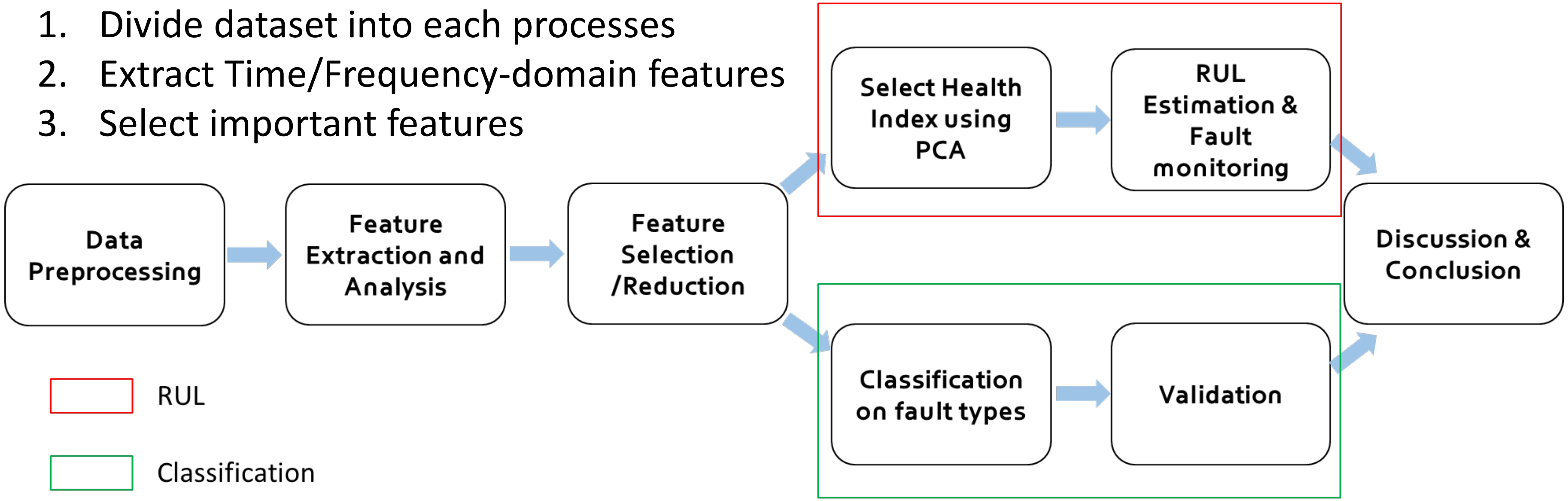
FAN_sound_OK
- FAN_sound_01.wav
- FAN_sound_02.wav
- FAN_sound_03.wav
- FAN_sound_04.wav
- FAN_sound_05.wav
- FAN_sound_06.wav
- FAN_sound_07.wav
- FAN_sound_08.wav
- FAN_sound_09.wav
- FAN_sound_10.wav
- FAN_sound_11.wav
- FAN_sound_12.wav
- FAN_sound_13.wav

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)

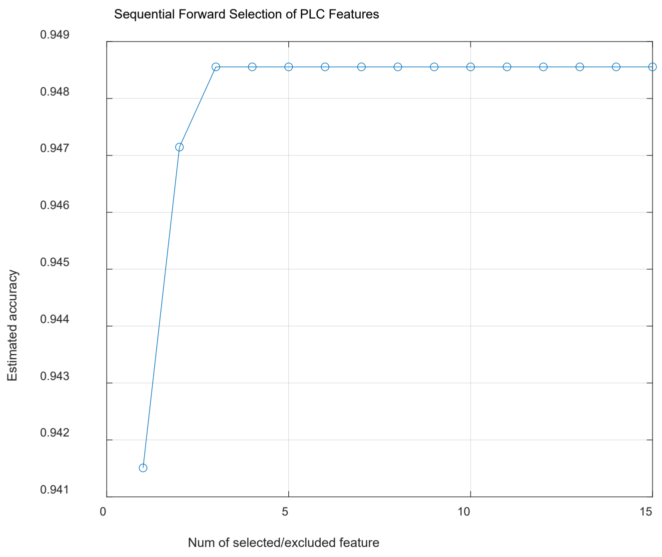
1) Time/ Frequency-domain based Feature extraction

① Time-domain: e.g. Peak to Peak value, RMS, mean, kurtosis, standard deviation

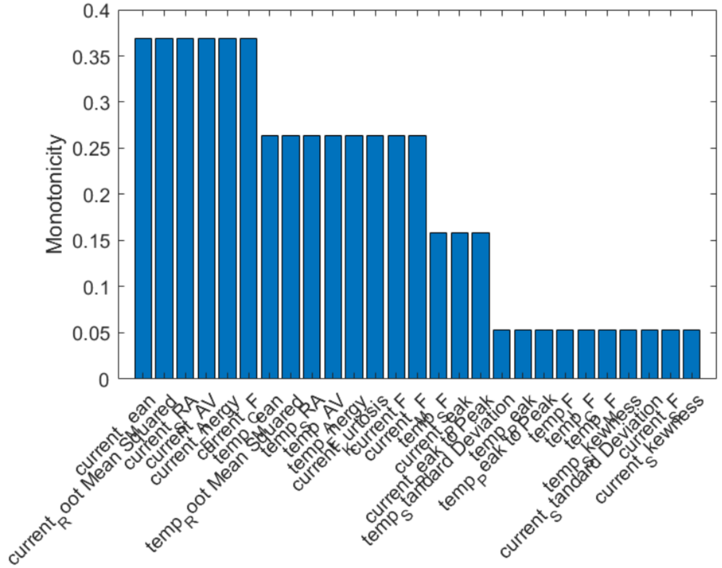
② Frequency-domain: e.g. Fast Fourier transform, RMSF

2) Feature Selection/Reduction

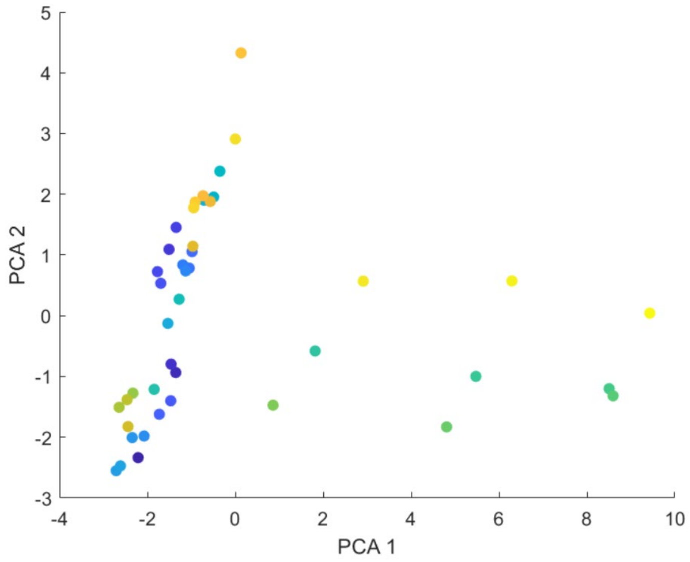
① Sequential Forward Selection



② Monotonicity



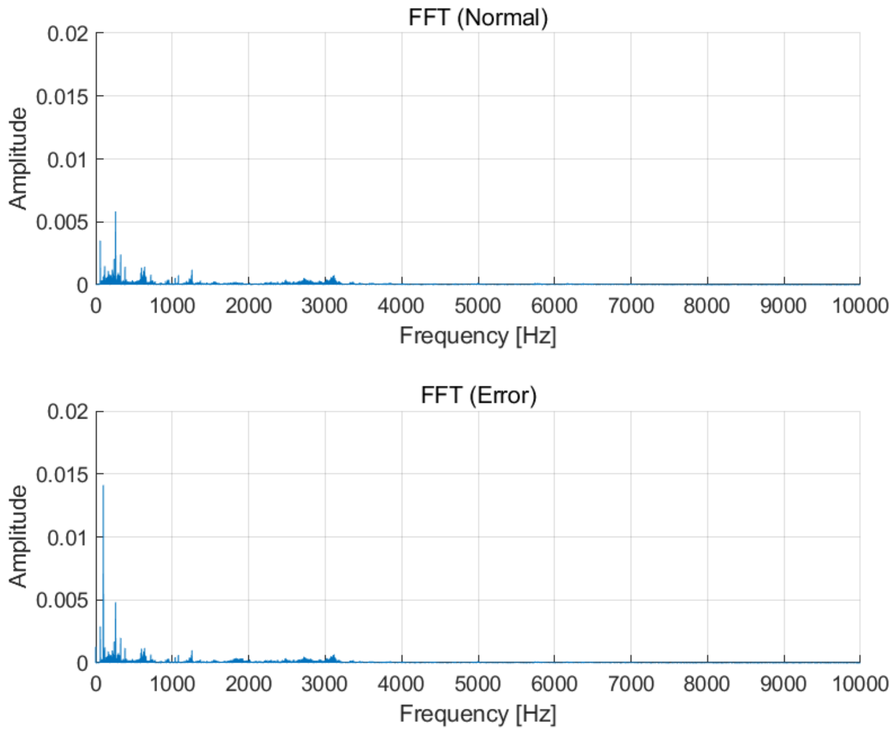
③ PCA



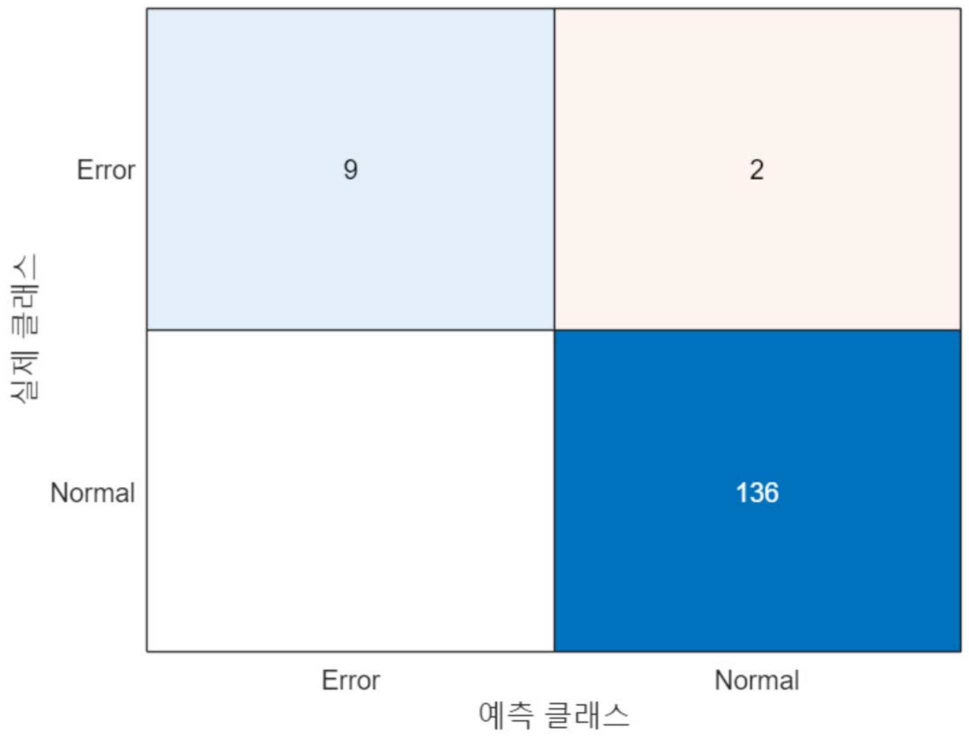
3.1. Classification (Sound Data)

3.1.1 Binary classification of two states (normal/error) using sound data

1) Binary Classification by sound data



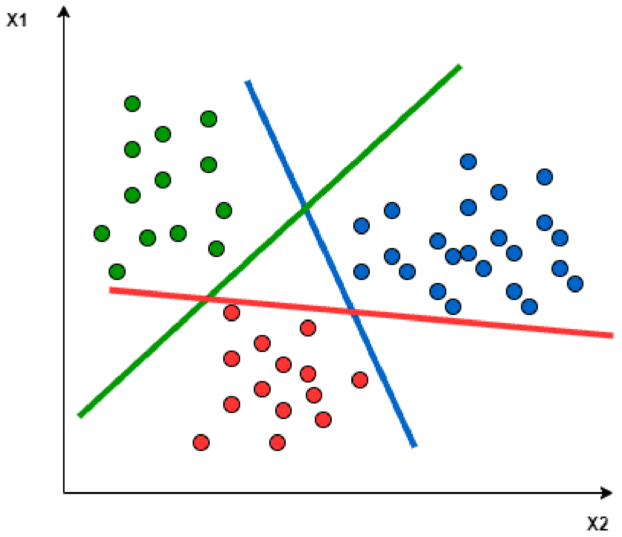
2) Confusion Matrix of binary classification(Logistic Regression Model)



3.1. Classification (PLC Data)

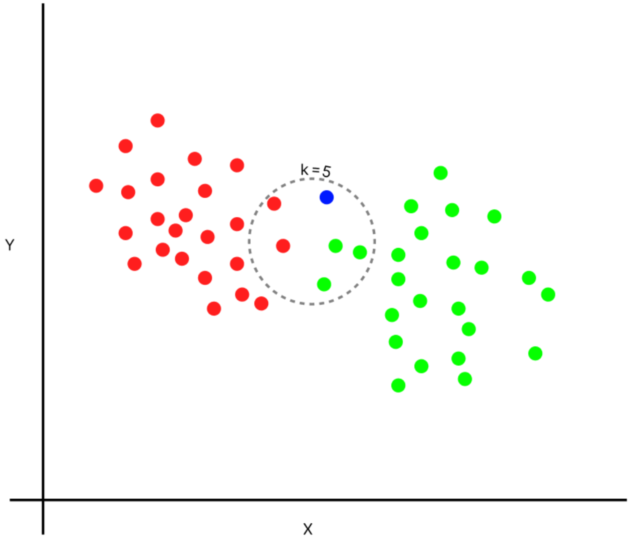
3.1.2 Multi-class classification of PLC data

1) Support vector machines (SVM)



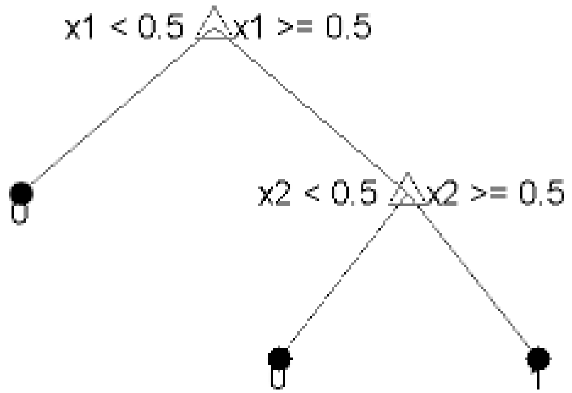
· Model: Multi-class SVM model
· Dimension: 5
· Method: Classification data based on hyperplane

2) K-Nearest Neighbor (KNN)



· Model: KNN
· Neighbor: 3
· Method: Classification based on data from the K nearest neighbors

3) Decision Tree



· Model: Decision Tree
· Branch: 24
· Method: Classification based on node setting by feature property

3.2 PHM

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

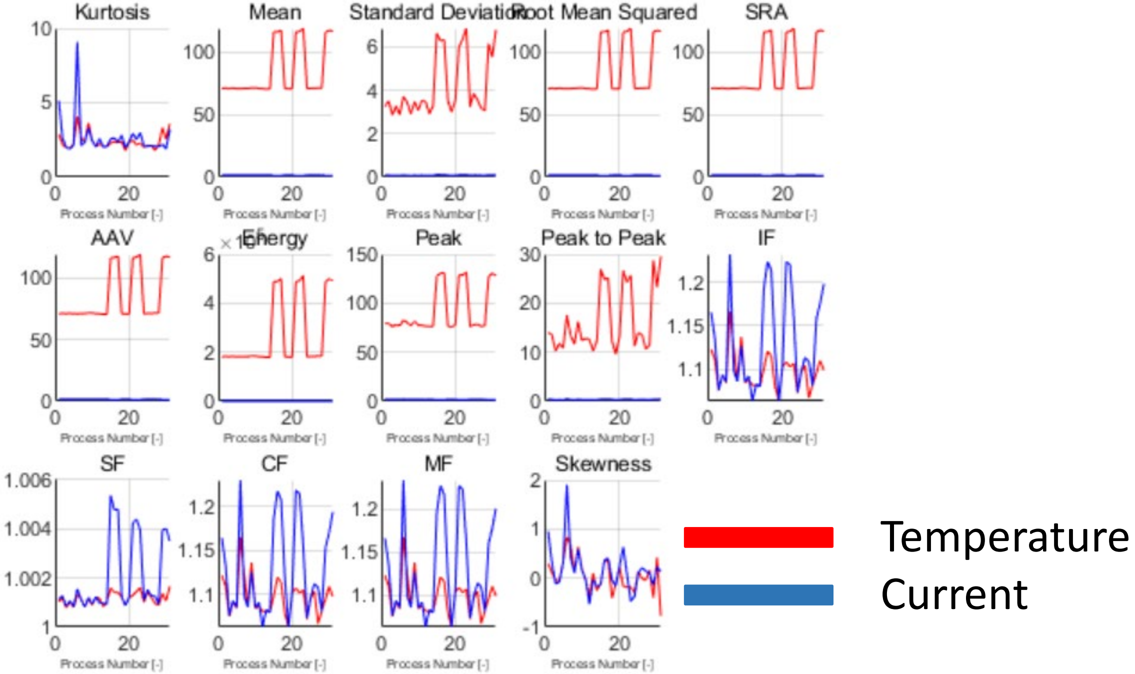
2) Earliest process with error is selected as break-point

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



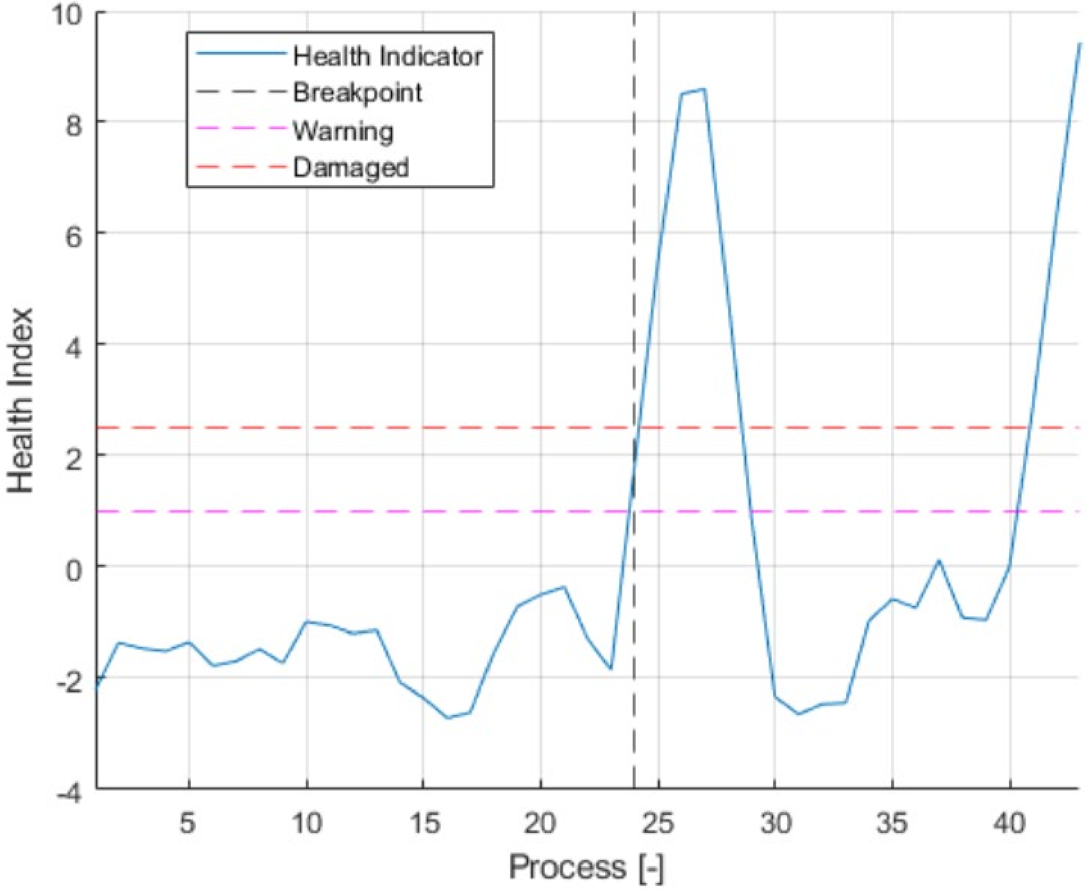
3.2 PHM

1) Health Indicator is selected for RUL estimation

2) Indicator tends to increase until the end

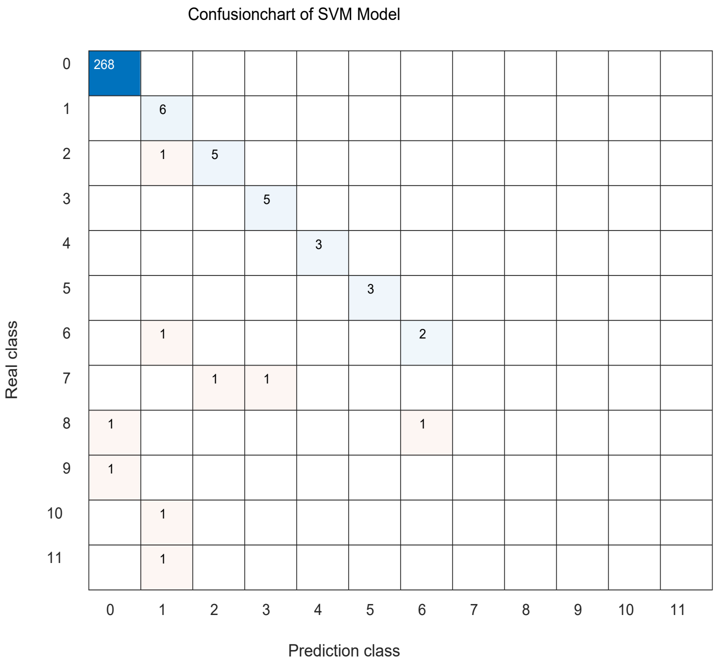
3) Limitation: Processes with error shows peak values of health indicator which disturbs degradation model

   → Use Moving Average Filter for smoothing



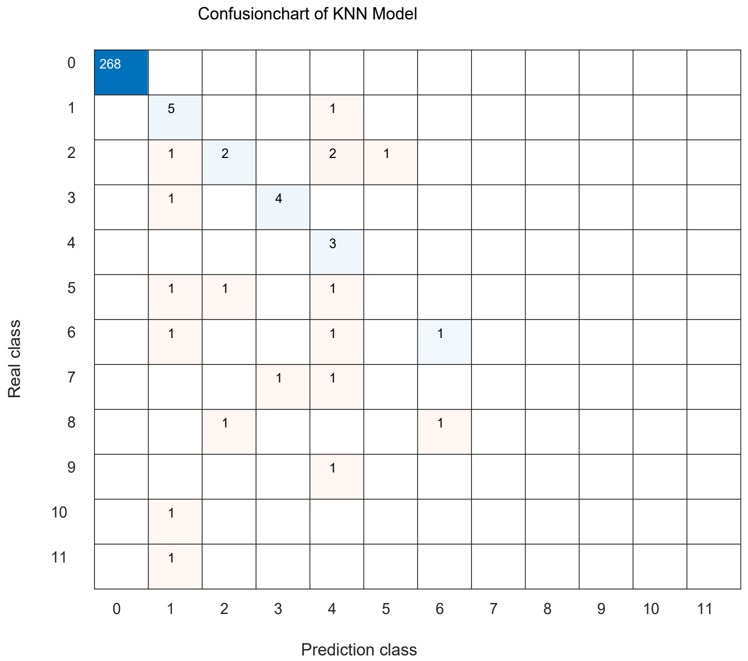
4.1. Classification Analysis

1) Support vector machines (SVM)



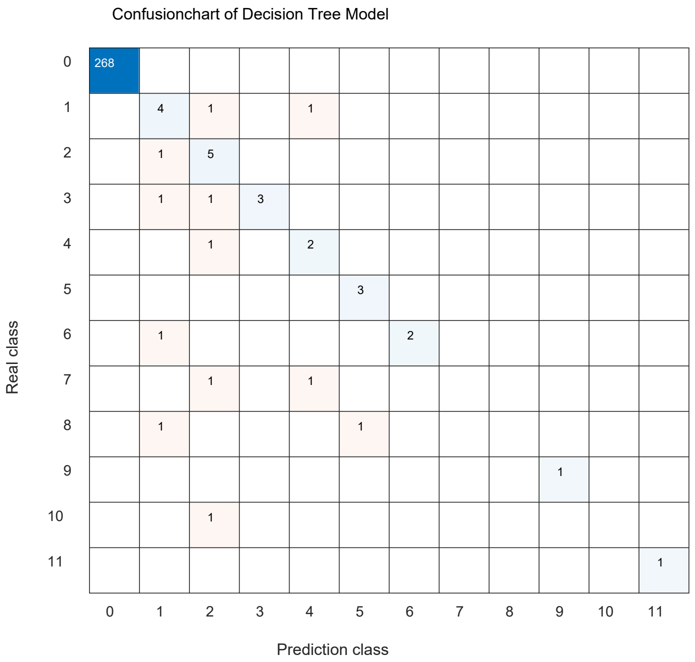
- $\text{Accuracy} = 0.97$
- $\text{Precision} = 0.53$
- $\text{Recall} = 0.58$
- $\text{F1} - \text{score} = 0.55$

2) K-Nearest Neighbor (KNN)



- $\text{Accuracy} = 0.94$
- $\text{Precision} = 0.32$
- $\text{Recall} = 0.39$
- $\text{F1} - \text{score} = 0.35$

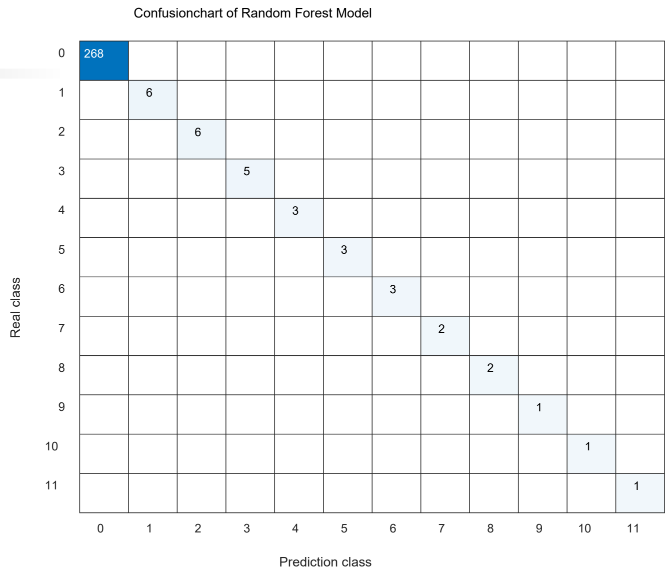
3) Decision Tree



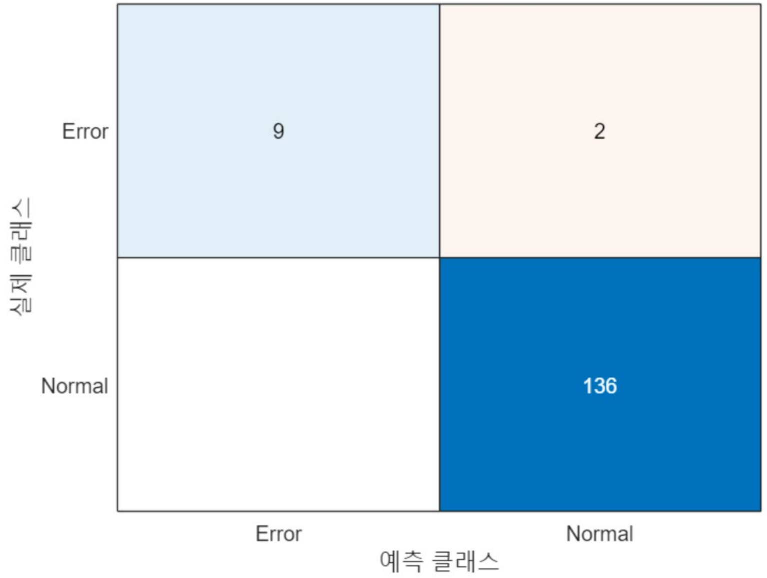
- $\text{Accuracy} = 0.96$
- $\text{Precision} = 0.66$
- $\text{Recall} = 0.68$
- $\text{F1} - \text{score} = 0.67$

4.1. Classification Analysis

1) Performance of **Decision Tree** is best among 3 ML models.
2) Refer to the paper that shows that R**andom Forest Models** are effective for classification and predictive preservation.
3) This is improved using Random Forest Model and achieved **100% F1-Score**

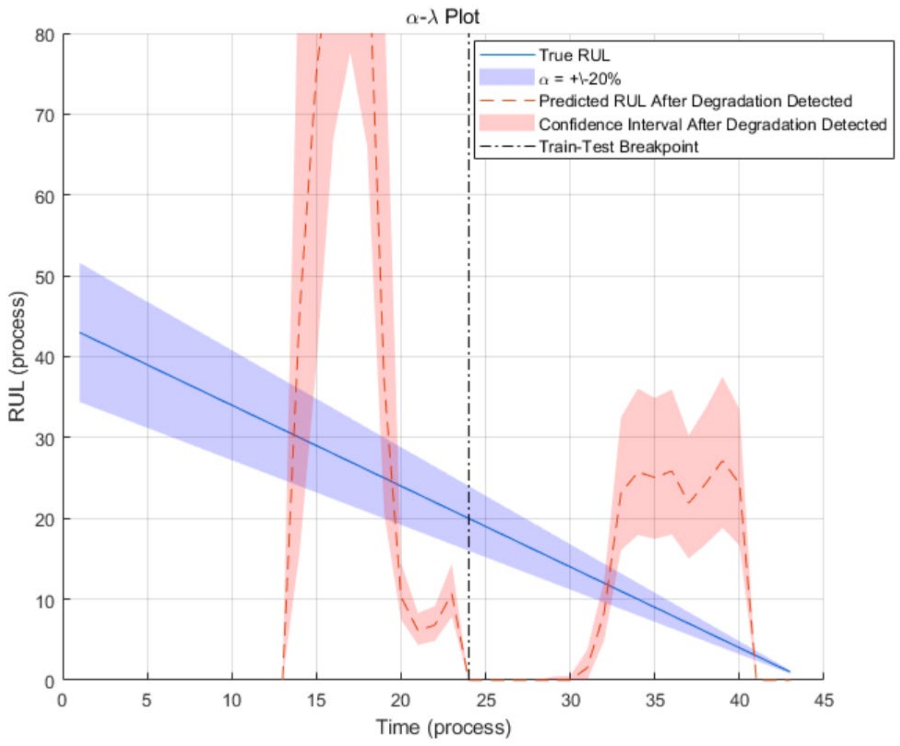


4) Sound Sensor Classification shows 95% of F1-Score

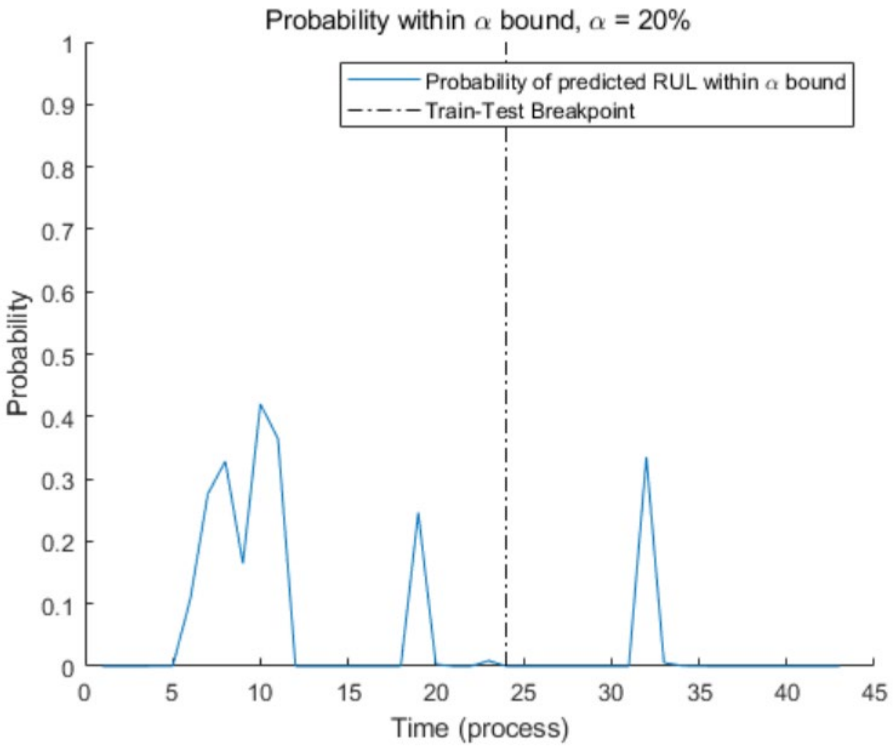


4.2. PHM Analysis

1) The $\alpha - \lambda$ plot



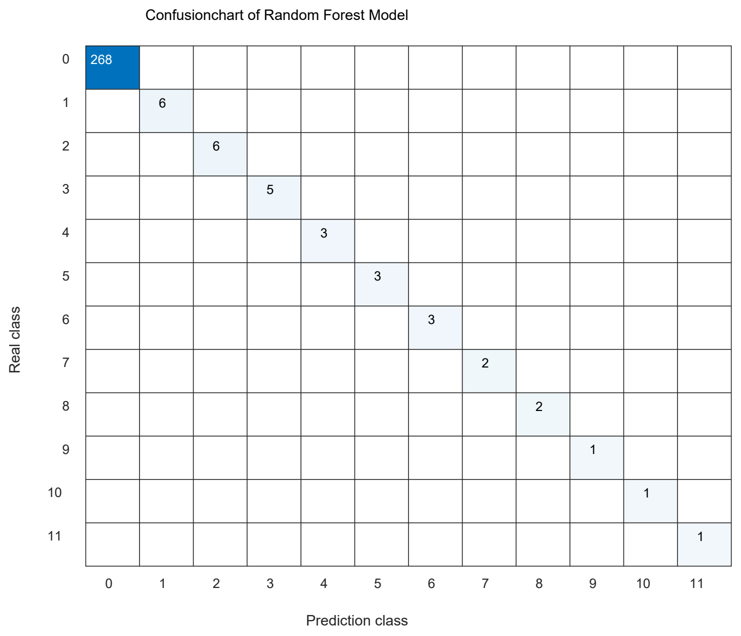
2) Probability of α bound



→ α band overlaps by at maximum about 30%

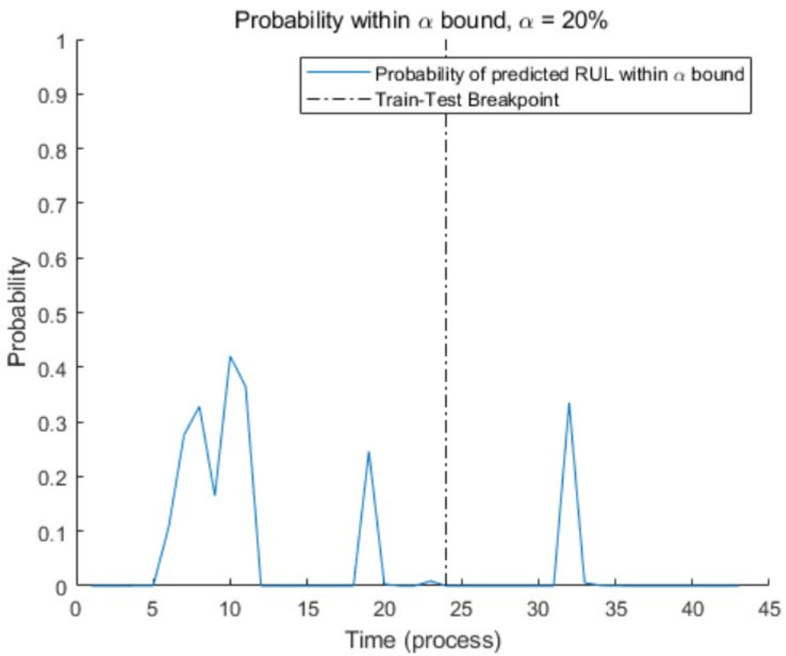
5. Result

1) Classification model



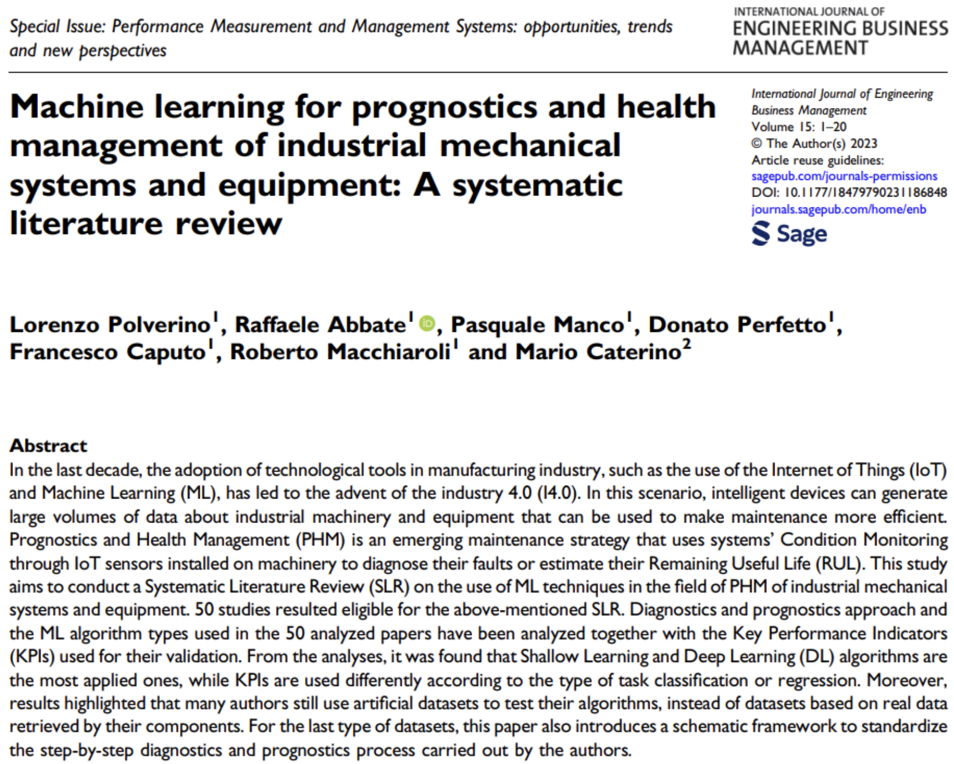
Random forest model
→ F1-score : 1
→ Expected Result $F1\text{-}score > 0.9$

2) RUL model



RUL model
→ α bound does not overlap
→ Prediction equipment lifespan with over 80% accuracy

6. Paper Review



Special Issue: Performance Measurement and Management Systems: opportunities, trends and new perspectives

INTERNATIONAL JOURNAL OF ENGINEERING BUSINESS MANAGEMENT

Machine learning for prognostics and health management of industrial mechanical systems and equipment: A systematic literature review

International Journal of Engineering Business Management
Volume 15: 1–20

DOI: 10.1177/18479790231186848
journals.sagepub.com/home/enb
Sage

Lorenzo Polverino[1], Raffaele Abbate[1], Pasquale Manco[1], Donato Perfetto[1], Francesco Caputo[1], Roberto Macchiaroli[1] and Mario Caterino[2]

Abstract
In the last decade, the adoption of technological tools in manufacturing industry, such as the use of the Internet of Things (IoT) and Machine Learning (ML), has led to the advent of the industry 4.0 (I4.0). In this scenario, intelligent devices can generate large volumes of data about industrial machinery and equipment that can be used to make maintenance more efficient. Prognostics and Health Management (PHM) is an emerging maintenance strategy that uses systems' Condition Monitoring through IoT sensors installed on machinery to diagnose their faults or estimate their Remaining Useful Life (RUL). This study aims to conduct a Systematic Literature Review (SLR) on the use of ML techniques in the field of PHM of industrial mechanical systems and equipment. 50 studies resulted eligible for the above-mentioned SLR. Diagnostics and prognostics approach and the ML algorithm types used in the 50 analyzed papers have been analyzed together with the Key Performance Indicators (KPIs) used for their validation. From the analyses, it was found that Shallow Learning and Deep Learning (DL) algorithms are the most applied ones, while KPIs are used differently according to the type of task classification or regression. Moreover, results highlighted that many authors still use artificial datasets to test their algorithms, instead of datasets based on real data retrieved by their components. For the last type of datasets, this paper also introduces a schematic framework to standardize the step-by-step diagnostics and prognostics process carried out by the authors.

. Polverino, Lorenzo, et al. "Machine learning for prognostics and health management of industrial mechanical systems and equipment: A systematic literature review." *International Journal of Engineering Business Management* 15 (2023): 18479790231186848

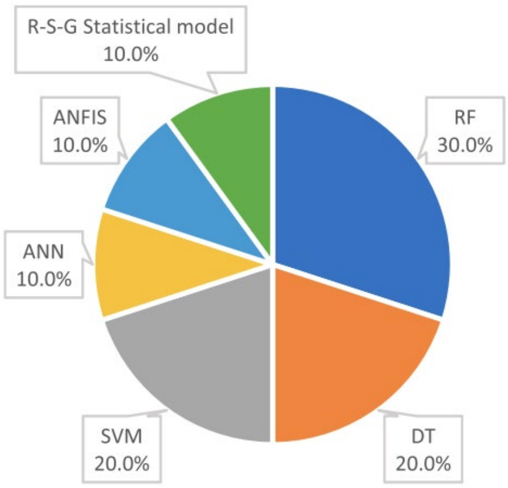


Table 1. Machine learning & deep learning models used in the 50 papers in recently

1) This paper analyzes the impact of the machine learning model on the manufacturing industry.
2) Expected for improvement ML model by get a recent trend of study for industrial ML model
3) Random forest models have good performance in classification and life prediction tasks → Improve classification model

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 ($F1\text{-}score \geq 0.9$)
   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. Reference

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

[2] Polverino, Lorenzo, et al. "Machine learning for prognostics and health management of industrial mechanical systems and equipment: A systematic literature review." *International Journal of Engineering Business Management* 15 (2023): 18479790231186848.