

# 회전기계 고장진단을 위한 최적 환경 구축 방법

송영원 <sup>\*†</sup>, 박정우\*, 유진수\*, 홍세현\*

\*한동대학교 기계제어공학부

## Optimal Environment Construction Approach for Rotating Machine Fault Diagnosis

Yeong-won Song <sup>\*†</sup>, Jeong-woo Park\*, Jin-su Yu\*, Se-hyun Hong\*

\*Dept. of Mechanical and Control Eng., Handong Univ.

Key Words: 고장진단(Fault diagnosis), 회전기계(Rotating machine), 머신러닝(Machine learning), 통계적 특징(Statistical Feature), 진동센서(Vibration Sensor), 샘플링 주기(Sampling Period)

### Abstract

This paper presents an optimal environment construction and efficient approach for rotational machine fault diagnosis. As a research method for this, fault diagnosis is performed at various sampling frequencies based on data acquired by over-specification sensors. Statistical features of data are used for fault diagnosis, and statistical features extracted according to various sampling frequencies and data sample size are used as AI model learning variables. Several ensemble models such as Random Forest and XGBoost were used, and the accuracy of the models was compared and analyzed. As a result, the fault diagnosis performance did not significantly increase after the sampling frequency exceeded 10 times the rotating machine's maximum frequency. This means that even when a low-performance sensor with a low frequency is used, high performance may be expected. Therefore, this paper is expected to contribute to economically performing fault diagnosis.

### 1. 서 론

회전기계는 산업 전반에 걸쳐 흔히 사용되는 중요한 요소이다. 회전기계의 예상치 못한 고장은 공정 및 생산 중단으로 이어지고, 이는 곧 상당한 경제적 손실과 직결된다. 산업 현장에서 사용하는 기계의 약 90%가 유도 전동기, 즉 회전 기계로 구성되기 때문에<sup>(1)</sup> 회전기계에 대한 결함 검출을 통한 상태 진단은 매우 중요한 사항이다.

회전기계의 상태 진단을 통한 효율적인 공정 운영에 대한 관심 증가로, 그간 회전기계 고장 진단 분야에서 다수의 연구가 진행된 바 있다. Rauber, W. H.<sup>(2)</sup>는 취득한 진동 신호 데이터로부터 Time-Frequency domain에서의 통계적 특성을 추출하고, Complex Envelope Analysis & Wavelet Packet Analysis를 통해 얻은 Feature Table을 기반으로 머신 러닝을 수행함으로써 회전기계의 중요한 요소가 되는 베어링 고장 진단 분야에 기여하였다. 본 논문에서 사용한 데이터셋의 제공처인 KAMP<sup>(13)</sup>에서는, 진동 신호로부터 취득한 Raw Data에 대해 Linear Interpolation, Moving Average Filter 등을 적용한 데이터를 딥러닝에 적용함으로써 도메인 지식 없이 고장 진단을 수행할 수 있는 프레임워크를 제시한 바 있다.

본문에 언급한 연구 이외에도 다양한 선행 연구가 수행된 바 있는데, 이들 대부분은 고장진단 연구를 위해 오픈 데이터셋을 활용하거나, 테스트

베드를 만들어 데이터를 직접 취득하였다. 산업 현장에서 가동되는 기계 대부분은 정상 상태로 가동되는데, 학습에 필요한 고장 데이터를 얻기 위해 인위로 고장을 가하는 방식은 매우 비효율적이기에 이러한 선행 연구자들의 접근은 합리적인 것으로 보인다.

대부분의 선행 연구가 고장진단의 정확성을 향상시키는 것에 집중한 것과 달리, 본 논문에서는 실제 산업 현장에서의 고장진단을 위한 최적의 센서 선정 방법을 다룬다. 데이터 취득을 위해 사용하는 진동 센서가 over-specification일 경우, 비효율적인 측면이 다수 존재한다. 센서의 가격을 결정하는 주된 요인 중 하나인 Sampling Frequency가 고장 진단을 위해 요구되는 사양보다 필요 이상으로 높은 경우, 센서 관리 측면에서 과한 비용이 소모될 수 있다. 또한 Sampling Frequency가 높다는 것은 곧 동일 시간 대비 더욱 많은 양의 데이터를 수집하는 것인데, 이는 곧 관리해야 할 데이터베이스의 크기 증가 및 데이터 처리 과정에서의 추가적인 자원 소모를 야기한다.

따라서, 본 연구에서 우리는 일정한 속도로 회전하고 있는 테스트베드를 통해 수집된 데이터인 'KAMP 회전기계 고장유형 AI 데이터셋'과 Case Western Reserve University(CWRU) Bearing Dataset을 바탕으로 최적의 데이터 취득을 가능하게 하는 최소 사양의 센서 선정 방법론에 대해 제시하고자 한다. 제시한 방법론의 적합성과 타당성을 다양한 머신 러닝 기법으로 검증하였으며, 이 과정에서 고장 데이터와 같은 이상치 검출에 뛰어난 통계적 특징에 기반하여 Feature Table을 생성하였다.

† 학생회원, 한동대학교 기계제어공학부

E-mail : wonhg1446@hanmail.net

TEL : (054)260-1322

이 논문의 나머지 부분은 다음과 같이 구성되어 있다. 2절에서는 데이터 취득 및 통계적 특징에 대한 이론과 전처리 과정에 대해 설명하고, 3절에서는 본 연구에서 사용한 데이터셋에 대해 기술하였다. 이후 4절에서는 센서의 Sampling Frequency 및 통계적 특징 추출을 위한 데이터 샘플 집단의 크기에 따른 머신 러닝 모델 별 분류 정확도에 대해 제시하고, 5절에서는 본 논문에서 제시한 접근법에 대한 평가와 함께 결론을 도출하였다.

## 2. 고장진단 최적 환경구축 알고리즘

서론에서 언급한 바와 같이, 본 논문에서는 회전기계의 고장진단에 사용될 데이터를 취득하기 위해 필요한 최소 사양의 센서 선정 방법에 대해 제시하고자 한다.

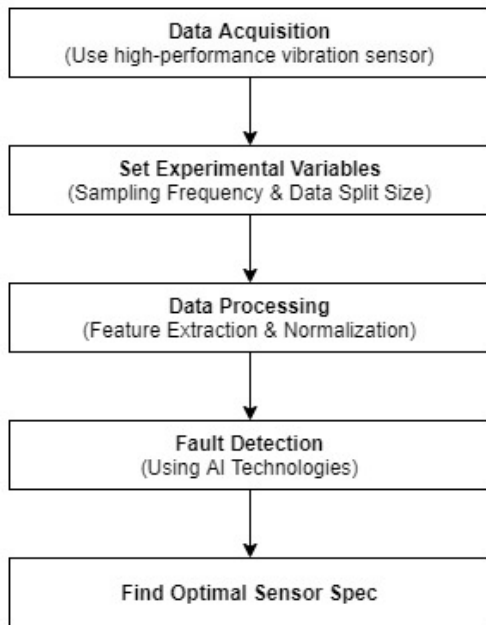


Fig. 1 Flowchart for Optimal Fault Detection

### 2.1 Data Acquisition

진동센서는 회전기계의 아날로그 신호를 디지털 신호로 변환한다. 진동센서로부터 수집되는 디지털 정보들을 바탕으로 신호를 계측하기 위해서는 회전기계의 원 신호, 즉 아날로그 신호를 완벽하게 기록할 수 있어야 한다. 이를 위해서는 Nyquist Theorem(5)에 근거하여, 아날로그 신호에 포함된 가장 높은 진동수의 최소 2배에 해당하는 주파수로 일정하게 샘플링 해야 한다.

그러나, 실제 현장에서 수집되는 자연계의 신호는 광범위한 스펙트럼을 갖는다. 따라서, 이론에 근거하여 예상되는 최대 주파수의 2배로 Sampling Frequency를 설정할 경우 Aliasing이 발생할 수 있다. 이러한 이유로 산업 현장에서는 예상 최대 주파수의 5~10배 이상으로 Sampling Frequency를 선정한다. 또한 (2), (13) 등의 선행 연구에서는, 보다

시각적인 연구 결과를 제시하기 위해 Sampling Frequency를 이론치의 50배 이상으로 선정한 바 있다. 서론에서도 언급하였듯, 지나치게 높은 Sampling Frequency를 설정하는 것은 경제적으로 비효율적이다. 반대로, 지나치게 낮은 Sampling Frequency를 설정한다면, 계측 데이터에 대한 정확한 분석이 어려워져 분류 성능이 저하될 수 있다. 따라서, 최적의 Sampling Frequency를 선정하는 것은 매우 유의미한 작업이다.

산업 환경마다 진동 신호가 갖는 스펙트럼이 상이하기에, 최적의 Sampling Frequency를 찾는 과정은 실험적으로 접근할 수밖에 없다. 따라서, 본 논문에서는 먼저 over-specification의 진동 센서로 데이터를 취득하기를 권장한다. 이러한 이유에 대해서는, 2.2절에 후술하였다.

### 2.2 Set Experimental Variables

Over-specification의 진동 센서로 취득된 데이터, 즉 높은 Sampling Frequency로 취득된 데이터에 대해 일정한 간격으로 데이터를 re-sampling함으로써 그보다 낮은 Sampling Frequency의 진동 센서로 데이터를 취득한 것과 동일한 효과를 얻을 수 있다.

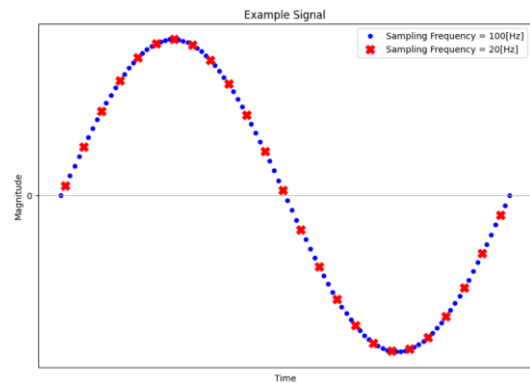


Fig. 2 Frequency Re-Sampling

Fig 2와 같이, 최초에 취득한 데이터를 시간 영역에서 indexing하고, 규칙적으로 re-sampling 함으로써  $1/2, 1/3, \dots, 1/n$  배의 sampling frequency로 데이터를 취득할 수 있음을 확인하였다. 본 논문에서는 회전기계 고장진단을 위한 최적의 진동 센서 선정을 목표로 하고 있기 때문에 진동 센서의 성능과 직결될 수 있는 sampling frequency, 즉 데이터 re-sampling 간격을 변수로 간주하였다.

또한, 통계적 특징을 바탕으로 Feature Table을 생성하고, 머신 러닝에 적용했을 때 우수한 분류 성능을 기대할 수 있음이 밝혀진 바 있다(2). 이를 수행하기 위해서는 통계치를 추출할 표본의 개수를 정의하여야 한다. 본 연구에서는, ‘시간’에 따라 통계치 추출을 위한 표본의 개수를 정의하였다.

3절에서는, Sampling Frequency와 데이터 split time의 변화에 따라 고장진단 결과가 어떻게 나타나는지를 기술하였다.

## 2.3 Data Processing

### 2.3.1 Feature Extraction

통계적 특징은 데이터 분포 및 이상치 탐지에 효과적이다. 따라서, 이를 기반으로 Feature Extraction을 진행하였으며, 추출한 통계적 특징은 Table 1 & Table 2에 기술하였다.

**Table 1** Time Domain Statistical Features

Statistical Feature	abbreviation	Formula
Root Mean Square	$X_{rms}$	$\left(\frac{1}{N} \sum_{i=1}^N x_i^2\right)^{1/2}$
Square Root Average	$X_{sra}$	$\left(\frac{1}{N} \sum_{i=1}^N \sqrt{ x_i }\right)^2$
Kurtosis Value	$X_{kv}$	$\frac{1}{N} \sum_{i=1}^N \left(\frac{x_i - \bar{x}}{\sigma}\right)^4$
Skewness Value	$X_{sv}$	$\frac{1}{N} \sum_{i=1}^N \left(\frac{x_i - \bar{x}}{\sigma}\right)^3$
Peak to Peak Value	$X_{ppv}$	$\max(x_i) - \min(x_i)$
Crest Factor	$X_{cf}$	$\frac{\max( x_i )}{X_{rms}}$
Impulse Factor	$X_{if}$	$\frac{\max( x_i )}{\frac{1}{N} \sum_{i=1}^N  x_i }$
Margin Factor	$X_{mf}$	$\frac{\max( x_i )}{X_{sra}}$
Kurtosis Factor	$X_{kf}$	$\frac{X_{kv}}{(X_{rms})^4}$
Skewness Factor	$X_{sf}$	$\frac{X_{sv}}{\frac{1}{N} \sum_{i=1}^N  x_i }$

**Table 2** Frequency Domain Statistical Features

Statistical Feature	abbreviation	Formula
Freq. Center	$X_{fc}$	$\frac{1}{N} \sum_{i=1}^N f_i$
RMS Freq.	$X_{rmsf}$	$\left(\frac{1}{N} \sum_{i=1}^N f_i^2\right)^{1/2}$
Root Variance Freq.	$X_{rvf}$	$\left(\frac{1}{N} \sum_{i=1}^N (f_i - x_{fc})^2\right)^{1/2}$

$x_i$ 는 진동 센서로부터 수집된 데이터이며,  $f_i$ 는  $x_i$ 에 대해 Fast Fourier Transform(FFT)을 적용하여 얻은 데이터이다. 또한,  $N$ 은 수집된 데이터의 개수,  $\sigma$ 는 수집된 데이터의 표준편차를 나타낸다.

### 2.3.2 Normalization

일반적으로, 진동 센서로부터 취득한 데이터에는 여러 노이즈가 포함된다. 취득되는 환경에 따라 노이즈의 유형 또한 상이하고, 이를 제거하기 위한 기법도 다양하다.

본 논문은 범용적인 산업 환경에 적용 가능한 ‘고장진단 최적 환경구축 알고리즘’을 제시하고 이것을 검증하는 것을 목적으로 한다. 따라서, 분류 성능을 높이기 위해 적용되는 노이즈 제거 기법을 사용하지 않고, 이상치의 영향을 최소화하기 위해 널리 사용되는 Robust Scaler만을 활용하였다.

## 3. 실험 데이터

본 논문에서 제시한 고장진단 최적 환경 구축 알고리즘을 검증하기 위해 국내 인공지능 중소벤처 제조 플랫폼(KAMP)에서 제공하는 ‘회전기계 고장유형 AI 데이터셋(13)과, Case Western Reserve University(CWRU) Bearing Dataset을 사용했다. 해당 데이터셋은 회전기계의 다양한 고장 유형에 대한 다중 분류가 가능하다는 특징을 갖는다. 그러나 본 연구에서는 단순 고장 발생 유무에만 집중하여 단일 고장 진단을 수행하였다.

### 3.1 KAMP Dataset

‘KAMP 회전기계 고장유형 AI 데이터셋’은 시그널링크㈜의 테스트베드에, 4개의 진동 센서를 부착하여 취득된 데이터이다. 모터가 1,500RPM으로 회전하는 동안 테스트베드의 상태를 임의로 조정함으로써 4가지(정상, 질량 불균형, 지지 불량, 질량 불균형과 지지 불량) 조건의 데이터를 취득하였다. 사용된 진동 센서의 사양 및 Sampling frequency는 Table 3에 정리하였다.

**Table 3** Vibration Sensor Information of KAMP Dataset

Frequency Span	672 [Hz]
Frequency Resolution	1.3125 [Hz]
Shaft Frequency	= 1,500 [RPM] = 25 [Hz]
Sampling Frequency	≈ 1.3 [kHz]
Total Sampling Time	140 [sec]

테스트베드의 모터가 25Hz로 회전하고 있기에, Nyquist Theorem<sup>(5)</sup>에 근거하여 센서의 Sampling Frequency는 최소 50Hz 이상의 값으로 설정되어야 한다. 본 데이터셋은 약 1.3kHz, 즉 최소 요구 진동수의 약 25배에 달하는 Sampling Frequency로 데이터가 취득되었기에 본 논문에서 제시하고자 하는 접근법 검증에 적합하다고 판단하였다.

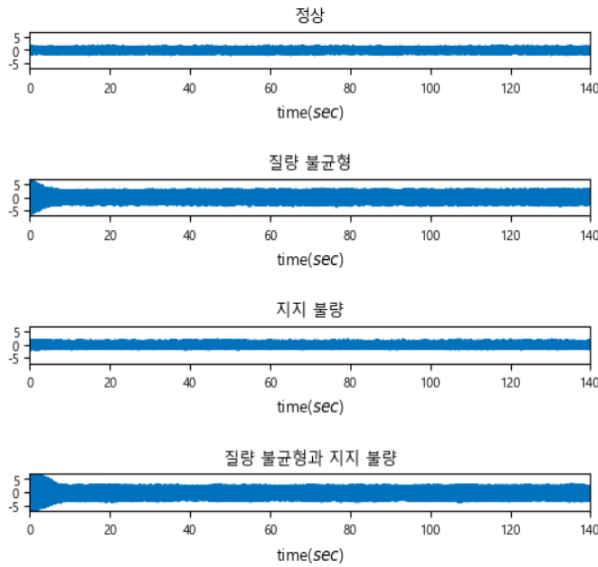


Fig. 3 Visualization of KAMP Dataset

한편, Fig 3에서 확인할 수 있듯이, 테스트베드 가동 초기 과도 상태가 발생한 것으로 판단된다. 따라서, 정상 상태의 데이터만을 AI 학습에 이용하기 위해 transient response 구간으로 간주되는 0~30초 구간의 데이터를 제거하였다.

### 3.2 CWRU Bearing Dataset

CWRU 데이터는 베어링 고장진단 연구에서 널리 사용되는 오픈 데이터셋이다. 모터가 약 1,700RPM으로 회전하는 가운데, 테스트베드의 베어링의 상태를 임의로 조정함으로써 4 가지(정상, 내륜 고장, 외륜 고장, 볼 고장) 조건의 데이터를 취득하였다. 데이터 취득 환경과 관련된 정보는 Table 4에 정리하였다.

Table 4 Acquisition Environment of CWRU Dataset

Bearing Model	6205-2RS JEM SKF
Shaft Frequency	$\approx 1,740$ [RPM] $\approx 29$ [Hz]
Sampling Frequency	12 [kHz]
Total Sampling Time	10 [sec]

테스트베드의 모터가 29Hz로 회전하고 있기에, Nyquist Theorem<sup>(5)</sup>에 근거하여 센서의 Sampling Frequency는 최소 58Hz 이상의 값으로 설정되어야 한다. 본 데이터셋은 약 12kHz, 즉 최소 요구 진동수의 약 200배에 달하는 Sampling Frequency로 데이터가 취득되었기에 본 논문에서 제시하고자 하는 접근법 검증에 적합하다고 판단하였다.

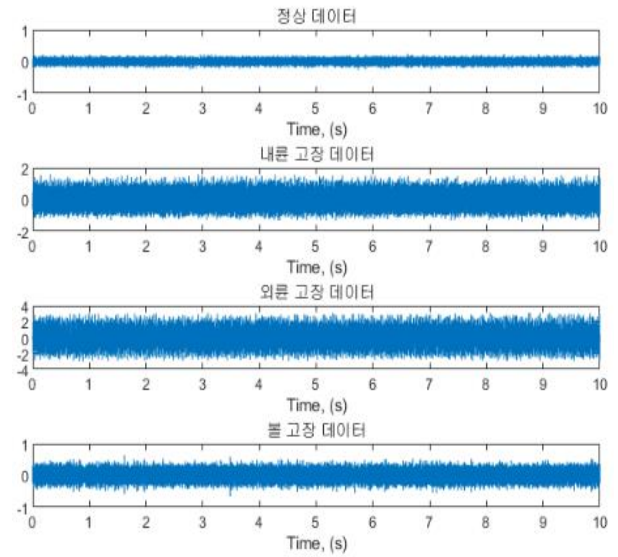


Fig. 4 Visualization of CWRU Dataset

Fig 4에서 확인할 수 있듯이, CWRU 데이터셋은 transient response 구간이 거의 관측되지 않았다. 그러나, 본 논문은 범용적인 환경에 적용 가능한 ‘회전기계 고장진단을 위한 최적 환경 구축 방법’ 제시를 목적으로 하므로, 앞서 검증에 사용한 KAMP 데이터셋에서 적용한 것과 유사하게, transient response 구간으로 간주되는 0~3초 구간의 데이터를 제거하였다.

## 4. 실험 결과

KAMP의 ‘회전기계 고장유형 AI 데이터셋’과, ‘CWRU Bearing Dataset’에 대해 본 논문에서 제안한 접근법을 적용하였다.

고장 분류를 위해 본 논문에서는 머신 러닝에 널리 사용되는 기법 SVM과, 앙상블 기법인 Random Forest, Bagging, XGBoost를 적용하였다. 평가 지표는, 실제 산업 현장에서 정상-비정상 데이터 불균형이 흔히 발생하는 것을 고려하여, 정확도(accuracy)와 재현률(recall)의 조화 평균인 F1 score를 사용하였다. KAMP 데이터셋에 대한 평가 결과는 Table 6~9를 통해 확인할 수 있고, CWRU 데이터셋에 대한 평가 결과는 Table 11~14를 통해 확인할 수 있다.

### 4.1 KAMP Dataset Experiment Result

최초 1.3kHz Sampling Frequency로 수집된 데이터에 대해 일정한 간격으로 데이터를 re-sampling 하였다. 해당 과정을 통해 구현한 Sampling Frequency는 Table 5에 정리하였다.

**Table 5** Re-Sampling Result about KAMP Dataset

Indexing interval	Sampling Frequency	Multiplier of Shaft Frequency
26	50 [Hz]	= 2
13	125 [Hz]	= 5
7	185.71 [Hz]	≈ 7.5
5	260 [Hz]	≈ 10
1	1,300 [Hz]	≈ 50

**Table 6** F1 Score of Random Forest

Split Time Sampling Frequency	1.00[s]	0.50[s]	0.25[s]	0.10[s]
1300 [Hz]	0.973	0.971	0.952	0.938
260 [Hz]	0.960	0.940	0.910	0.867
185.71 [Hz]	0.938	0.897	0.879	0.854
100 [Hz]	0.919	0.883	0.873	0.844
50 [Hz]	0.887	0.872	0.846	0.842

**Table 7** F1 Score of Bagging

Split Time Sampling Frequency	1.00[s]	0.50[s]	0.25[s]	0.10[s]
1300 [Hz]	0.959	0.956	0.943	0.923
260 [Hz]	0.927	0.923	0.889	0.853
185.71 [Hz]	0.928	0.895	0.858	0.832
100 [Hz]	0.917	0.881	0.856	0.820
50 [Hz]	0.870	0.847	0.823	0.811

**Table 8** F1 Score of XGBoost

Split Time Sampling Frequency	1.00[s]	0.50[s]	0.25[s]	0.10[s]
1300 [Hz]	0.979	0.959	0.946	0.931
260 [Hz]	0.935	0.936	0.902	0.873
185.71 [Hz]	0.945	0.884	0.876	0.862
100 [Hz]	0.922	0.892	0.869	0.837
50 [Hz]	0.885	0.884	0.852	0.845

**Table 9** F1 Score of SVM

Split Time Sampling Frequency	1.00[s]	0.50[s]	0.25[s]	0.10[s]
1300 [Hz]	0.865	0.854	0.876	0.890
260 [Hz]	0.860	0.853	0.858	0.857
185.71 [Hz]	0.853	0.855	0.858	0.857
100 [Hz]	0.850	0.851	0.857	0.857
50 [Hz]	0.854	0.857	0.855	0.857

## 4.2 CWRU Bearing Dataset Experiment Result

최초 12kHz Sampling Frequency 로 수집된 데이터에 대해 일정한 간격으로 데이터를 re-sampling 하였다. 해당 과정을 통해 구현한 Sampling Frequency 는 Table 10 에 정리하였다.

**Table 10** Re-Sampling Result about CWRU Dataset

Indexing interval	Sampling Frequency	Multiplier of Shaft Frequency
200	60 [Hz]	≈ 2
80	150 [Hz]	≈ 5
53	226.42 [Hz]	≈ 7.5
40	300 [Hz]	≈ 10
1	12,000 [Hz]	≈ 400

**Table 11** F1 Score of Random Forest

Split Time Sampling Frequency	1.00[s]	0.50[s]	0.25[s]	0.10[s]	0.05[s]
12000 [Hz]	1.0	1.0	1.0	1.0	1.0
300 [Hz]	1.0	1.0	1.0	1.0	0.964
226.42 [Hz]	1.0	1.0	1.0	0.989	0.934
150 [Hz]	1.0	1.0	1.0	0.950	0.933
60 [Hz]	1.0	0.933	0.938	0.916	0.861

**Table 12** F1 Score of Bagging

Split Time Sampling Frequency	1.00[s]	0.50[s]	0.25[s]	0.10[s]	0.05[s]
12000 [Hz]	1.0	1.0	1.0	1.0	1.0
300 [Hz]	1.0	1.0	1.0	1.0	0.941
226.42 [Hz]	1.0	1.0	1.0	0.989	0.976

150 [Hz]	1.0	1.0	0.970	0.941	0.903
60 [Hz]	1.0	1.0	0.944	0.914	0.900

Table 13 F1 Score of XGBoost

Split Time Sampling Frequency	1.00[s]	0.50[s]	0.25[s]	0.10[s]	0.05[s]
12000 [Hz]	0.889	1.0	1.0	1.0	1.0
300 [Hz]	0.889	1.0	1.0	1.0	0.982
226.42 [Hz]	1.0	1.0	1.0	0.976	0.936
150 [Hz]	1.0	1.0	1.0	0.964	0.949
60 [Hz]	0.909	0.933	0.941	0.918	0.881

Table 14 F1 Score of SVM

Split Time Sampling Frequency	1.00[s]	0.50[s]	0.25[s]	0.10[s]	0.05[s]
12000 [Hz]	1.0	1.0	1.0	0.977	0.966
300 [Hz]	1.0	0.889	0.971	0.988	0.928
226.42 [Hz]	1.0	0.941	1.0	0.911	0.943
150 [Hz]	0.909	0.941	1.0	0.909	0.905
60 [Hz]	0.909	0.941	0.944	0.884	0.871

## 5. 결 론

다양한 사양의 센서를 직접 구비하여 데이터를 취득하는 번거로움 없이, over-specification의 센서로 획득된 하나의 데이터셋을 바탕으로 다양한 Sampling Frequency를 구현함으로써 고장 진단을 수행하기 위한 최적의 환경 구축 방법을 제안하였다. Sampling Frequency가 높을수록 전반적으로 우수한 고장 진단이 가능함을 확인했으나, ‘AI 모델 선정 및 통계치 추출을 위한 데이터 표본의 크기’가 검증 결과에 큰 영향을 미치는 것을 확인했다. 또한, Sampling Frequency가 기계 회전 주파수의 10 배를 초과한 후에는 고장 진단 성능이 크게 증가하지 않음을 확인하였다.

주기적인 센서 교체 및 데이터베이스 관리 등으로 소모될 비용을 고려한다면, 최적의 센서 사양을 선택하는 것은 매우 합리적인 태도이다. 본 논문의 접근법을 적용함으로써, 보다 효과적인 고장 진단이 가능할 것으로 보인다.

하지만 제안된 접근법은 통계치 추출을 위해 데이터를 ‘시간’기준으로 묶었기 때문에, 데이터 취득 시간이 매우 짧을 경우 Under Fitting을 야기할 수 있다. 뿐만 아니라, 범용적인 적용을 고려한

결과, 다양한 신호 처리 기법을 활용한 데이터 정제 과정이 부족했다는 한계를 가진다.

충분한 시간 동안 데이터를 수집하고, 데이터 취득 환경에 따라 적절한 데이터 정제 과정을 거치며, AI 모델의 하이퍼 파라미터를 조절하는 등의 과정을 수행함으로써 성능 개선을 기대할 수 있을 것으로 보인다.

## 참고문헌

- (1) SK Gundewar, PV Kane, (2020), “Condition Monitoring and Fault Diagnosis of Induction Motor,” *Journal of Vibration Engineering & Technologies*, Vol. 9, pp. 643-674
- (2) Rauber, W. H., Boldt, A. F. and Varejão, M. F., (2015), “Heterogeneous Feature Models and Feature Selection Applied to Bearing Fault Diagnosis,” *IEEE Transaction on Industrial Electronics*, Vol. 62, Issue 1, pp. 637~646
- (3) Kim, S., An, D., & Choi, J. H. (2020). Diagnostics 101: A tutorial for fault diagnostics of rolling element bearing using envelope analysis in matlab. *Applied Sciences*, 10(20), 7302.
- (4) Sassi, S., Badri, B., & Thomas, M. (2007). A numerical model to predict damaged bearing vibrations. *Journal of Vibration and Control*, 13(11), 1603-1628.
- (5) Nyquist, H. (1928). Certain topics in telegraph transmission theory. *Transactions of the American Institute of Electrical Engineers*, 47(2), 617-644.
- (6) Cooley, J. W., & Tukey, J. W. (1965). An algorithm for the machine calculation of complex Fourier series. *Mathematics of computation*, 19(90), 297-301.
- (7) Hunter, J. D. (2007). Matplotlib: A 2D graphics environment. *Computing in science & engineering*, 9(03), 90-95.
- (8) Van Der Walt, S., Colbert, S. C., & Varoquaux, G. (2011). The NumPy array: a structure for efficient numerical computation. *Computing in science & engineering*, 13(2), 22-30.
- (9) Breiman, L. (2001). Random forests. *Machine learning*, 45(1), 5-32.
- (10) Chen, T., & Guestrin, C. (2016, August). Xgboost: A scalable tree boosting system. In Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining (pp. 785-794).
- (11) Cortes, C., & Vapnik, V. (1995). Support-vector networks. *Machine learning*, 20(3), 273-297.
- (12) Breiman, L. (1996). Bagging predictors. *Machine learning*, 24(2), 123-140.
- (13) 회전기계 고장유형 AI 데이터셋[KAMP 인공지능 중소벤처 제조 플랫폼], (2022. 07. 24). <https://www.kamp-ai.kr/front/dataset/AiData.jsp>