

Autoencoder Based N-Segmentation Frequency Domain Anomaly Detection for Optimization of Facility Defect Identification

Kichang Park[†] · Yongkwan Lee^{††}

ABSTRACT

Artificial intelligence models are being used to detect facility anomalies using physics data such as vibration, current, and temperature for predictive maintenance in the manufacturing industry. Since the types of facility anomalies, such as facility defects and failures, anomaly detection methods using autoencoder-based unsupervised learning models have been mainly applied. Normal or abnormal facility conditions can be effectively classified using the reconstruction error of the autoencoder, but there is a limit to identifying facility anomalies specifically. When facility anomalies such as unbalance, misalignment, and looseness occur, the facility vibration frequency shows a pattern different from the normal state in a specific frequency range. This paper presents an N-segmentation anomaly detection method that performs anomaly detection by dividing the entire vibration frequency range into N regions. Experiments on nine kinds of anomaly data with different frequencies and amplitudes using vibration data from a compressor showed better performance when N-segmentation was applied. The proposed method helps materialize them after detecting facility anomalies.

Keywords : Anomaly Detection, Prediction Maintenance, Autoencoder, Unsupervised Learning, Frequency Domain

설비 결함 식별 최적화를 위한 오토인코더 기반 N 분할 주파수 영역 이상 탐지

박기창[†] · 이용관^{††}

요약

제조 분야 설비 예지보전을 위해서 진동, 전류, 온도 등 물리 데이터를 기반으로 설비 이상을 탐지하는 인공지능 학습 모델이 활용되고 있다. 설비 결함, 고장 등 설비 이상 유형은 매우 다양하므로, 주로 오토인코더 기반 비지도 학습 모델을 이용한 이상 탐지 방법이 적용되고 있다. 설비 상태의 정상, 비정상 여부는 오토인코더의 재구성 오차를 이용해 효과적으로 분류할 수 있지만, 설비 이상의 구체적인 상태를 식별하는 데 한계가 있다. 설비 불균형, 정렬 불량, 고정 불량 등 설비 이상 상황 발생 시, 설비 진동 주파수는 특정 영역에서 정상 상태와 다른 패턴을 나타낸다. 본 논문에서는 전체 진동 주파수 범위를 N개 영역으로 나누어 이상 탐지를 수행하는 N 분할 이상 탐지 방법을 제시하였다. 압축기의 진동 데이터를 이용해 주파수와 강도를 달리한 9종의 이상 데이터를 대상으로 실험한 결과, N 분할을 적용하였을 때 더 높은 이상 탐지 성능을 나타냈다. 제안 방법은 설비 이상 탐지 이후, 설비 이상 구체화에 활용될 수 있다.

키워드 : 이상탐지, 예지보전, 오토인코더, 비지도 학습, 주파수 영역

1. 서론

설비 예지보전(predictive maintenance)은 정상적인 설비

상태를 지속해서 유지하는 개념으로 중단시간(downtime) 감소, 생산성 향상, 비용 절감, 근로자 안전 확보 등 제조 현장의 효율성 증대를 위해 중요하며, 이를 실현하기 위해 다양한 기계 학습 모델이 연구되고 있다[1]. 효과적인 예지보전을 위해서는 지속적인 설비 상태 감시(monitors)를 통해 설비 상태가 정상(normal) 상태인지 이상(anomaly) 상태인지 판단할 수 있어야 한다. 설비 이상 상태는 결함, 고장이 발생한 설비의 비정상적인 상태로 주로 진동, 전류, 온도 등 설비 운용 중 측정 가능한 물리 데이터를 분석해 이상 여부를 판단, 예측할 수 있다[2-4]. 하지만, 설비 이상의 발생 원인과 상태는 매우 다양하므로, 기계 학습을 위한 모든 이상 상황에 해당하는 충

※ 이 논문은 과학기술정보통신부의 재원으로 정보통신기획평가원의 지원을 받아 수행된 지역지능화혁신인재양성사업(IITP-2024-2020-0-01741, 50%)과 중소벤처기업부의 재원으로 중소기업기술정보진흥원의 지원을 받아 수행된 스마트제조혁신 R&D 지원사업 연구 결과로 수행되었음(RS-2022-0014 1076, 50%).

† 정 회 원 : (주)리베니에 제조지능화기술연구소 팀장

†† 정 회 원 : 한국공학대학교 그랜드ICT연구소 산학협력중점교수

Manuscript Received : March 10, 2023

First Revision : January 23, 2024

Accepted : February 26, 2024

* Corresponding Author : Yongkwan Lee(ivan.lee@tukorea.ac.kr)

분한 양의 데이터 확보는 현실적으로 어렵다. 따라서, 정상 상태의 설비 데이터를 학습하고, 이를 기반으로 이상 상태를 식별하는 비지도 학습방식이 주로 활용되고 있으며, 오토인코더(autoencoder)의 재구성 오차(reconstruction error)를 이용하는 방법이 대표적이다. 오토인코더를 이용한 이상 탐지(anomaly detection)는 정상 데이터를 학습하고, 복원하는 과정에서 발생하는 오차를 이용해 설비의 정상 상태와 이상 상태를 판별한다.

설비 이상 탐지에는 센서(sensor), 설비 제어 PC, PLC(Programmable Logic Controller)에서 수집된 시계열(timeseries) 데이터가 주로 활용되고 있으며, 이중 진동 데이터가 대표적이다[5]. 대부분 제조 설비는 모터 구동으로 운용되며, 이때 발생하는 진동 데이터를 이용해 설비 상태를 예측할 수 있다. 특히, 진동 데이터의 주파수 영역(frequency domain)은 설비 불균형(unbalance), 정렬 불량(misalignment), 고정 불량(looseness) 등 설비 이상 상태를 구체적으로 파악하는데 중요한 정보를 제공한다[6-8]. 따라서, 설비 이상 유형에 따라 특정 진동 주파수 영역에서 정상 상태와 다른 패턴을 나타내기 때문에, 주파수 영역별로 이상 탐지를 수행할 필요가 있다.

본 논문에서는 설비 이상 상태를 구체적으로 파악하기 위해 주파수 영역을 N개 영역으로 분리하고, 주파수 영역별 이상 탐지를 수행하는 방법을 제시하였다. 사례연구로 약 2개월간 운영된 압축기(compressor)의 진동 데이터를 오토인코더 모델로 학습하고, N 분할 이상 탐지를 수행함으로써 특정 주파수 영역 이상 탐지에 효과적으로 적용될 수 있음을 보인다.

본 논문의 구성은 다음과 같다. 2장에서는 관련 연구로 딥러닝 학습 유형별 설비 이상 탐지 모델을 소개한다. 3장에서는 오토인코더를 이용한 주파수 영역별 이상 탐지 방법을 제시하고, 4장에서는 사례연구와 적용 결과를 소개한다. 5장에서는 시사점을 요약하고, 향후 연구를 제시한다.

2. 관련 연구

딥러닝 모델은 학습 유형에 따라 크게 지도학습(supervised learning), 비지도 학습(unsupervised learning), 반지도 학습(semi-supervised learning) 3가지 유형으로 나뉘며, 제조 분야 설비 이상 탐지에서도 이들 유형이 모두 사용되고 있다. 이 중 지도학습은 설비 이상을 식별할 수 있는 레이블(label)에 따른 이상 탐지를 다룬다. Mohammed 외[9]는 기어 박스(gearbox)의 3가지 깊이(1mm, 2mm, 3mm) 균열 감지에 ANN(Artificial Neural Network)을 활용했다. Mey 외[10]는 4가지 유형(강도)의 불균형 인자(unbalance factor)를 적용한 모터(motor)에서 발생한 진동 데이터를 수집하고, CNN(Convolutional Neural Network), FCNN(Fully Connected Neural Network), Random Forest, HMM(Hidden Markov Model)을 이용해 고장 감지를 수행했으며, Bouisfi 외[11]는 정상 로터

바(rotor bar)와 4가지 유형의 고장 로터 바에서 전류 데이터를 수집하고, SVM(Support Vector Machine), Random Forest, LightGBM(Light Gradient Boosting Machine)을 이용해 고장 감지를 수행했다. 이와 같은 지도학습은 높은 이상 탐지 성능을 나타내지만, 이상 탐지 범위는 학습에 사용된 결함 유형으로 제한된다. 결과적으로, 설비 이상 진단을 위한 지도학습은 고장 유형별 충분한 데이터를 요구하지만, 제조 현장에서 설비 이상 발생 빈도는 매우 낮고 유형은 다양하므로 지도학습 적용은 매우 제한적일 수밖에 없다.

이에 대한 대안으로 대량의 정상 상태 설비 데이터 학습으로 설비 이상 여부를 판별하는 임계치(threshold)를 결정하고, 이상 탐지를 수행하는 다양한 유형의 오토인코더 기반 비지도 학습 방법이 연구되고 있다. Chen 외[12]는 산업용 로봇의 이상 탐지에 효과적인 SWCVA(Sliding-Window Convolutional Variational Autoencoder) 모델을 제안했으며, Givnan 외[13]는 회전기(rotary machine) 이상 탐지 시 매개변수 수를 개선한 RSAE(Reduced Stacked Autoencoder) 모델을 제안했다. Huang 외[14]는 모터의 이상 탐지와 특징 추출(feature extraction)에 효과적인 RNN(Recurrent Neural Network) 기반 변분 오토인코더(Variational Autoencoder) 모델을 연구했다. Tziolas 외[15]는 ANN, LSTM(Long Short Term Memory), CNN을 기반으로 한 3가지 유형의 오토인코더 모델을 이용해 엘리베이터 유압 동력 장치(HPU, Hydraulic Power Unit)의 이상 탐지 성능을 비교했다. 이러한 비지도 학습은 라벨링 데이터가 제한적인 제조 분야에 효과적으로 적용될 수 있지만, 정상 데이터만을 이용해 모델을 학습하므로 쉽게 과적합(overfitting) 될 수 있으며, 잘못된 탐지(false positive, false negative) 결과를 도출할 수 있다. 따라서, 소량의 레이블 데이터를 추가로 활용하는 반지도 학습을 통해 이상 탐지 성능을 높이는 연구가 수행되고 있다. Zhang 외[16]는 베어링(bearing) 이상 탐지에 VAE(Variational Autoencoder) 기반 생성 모델을 이용해 이상 탐지 정확도를 높이는 연구를 수행했으며, Thoidis 외[17]는 기계음(machinery sound)을 이용한 설비 상태 모니터링에 반지도 학습을 이용했다.

설비에서 발생하는 진동 데이터는 결함 유형에 따라 특정 주파수 영역에서 높은 진폭(amplitude)을 나타낸다[6-8, 18, 19]. 예를 들어, 정상 진동 신호(Fig. 1. (a))를 갖는 회전기기 설비에 설비 불균형(Fig. 1. (b)), 정렬 불량(Fig. 1. (c)), 고정 불량(Fig. 1. (d))이 발생한 경우, 불량유형과 설비 특성에 따라 특정 주파수 영역에서 이상 진폭이 발생한다.

설비 결함 탐지를 위해 진동 데이터를 이용한 기존 비지도 학습 적용 연구[20-22]는 설비 결함 발생 여부만을 예측한다. 본 논문에서는 설비 결함 발생뿐만 아니라, 결함 발생 원인을 구체적으로 파악하기 위해 Fig. 2와 같이 진동 데이터의 주파수 영역을 N개로 나눠 주파수 영역별로 이상 탐지를 수행하는 N 분할 이상 탐지 알고리즘을 제안하고, 사례연구로 공기압축기에서 수집된 진동 데이터를 이용해 수행했다.

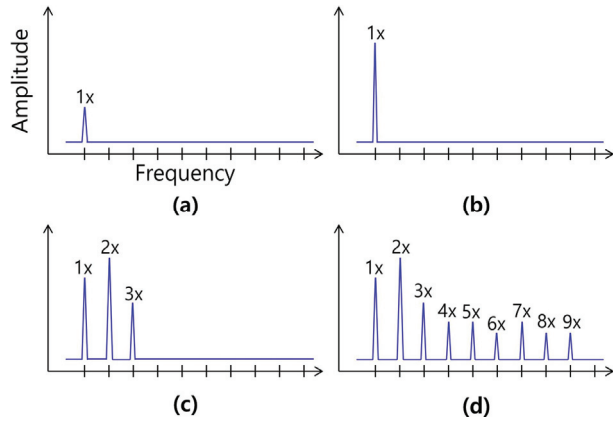


Fig. 1. Shape of Vibration Frequency by Fault Type

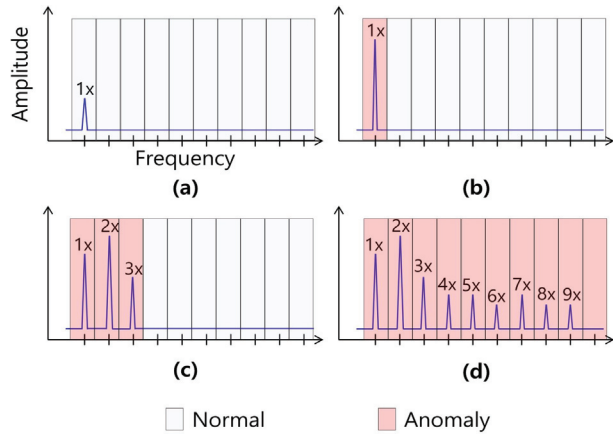


Fig. 2. N-segmentation Anomaly Detection

3. 주파수 영역별 이상 탐지

3.1 N 분할 이상 탐지 알고리즘

오토인코더 모델은 입력신호(input signal)를 더 작은 차원으로 축소하는 인코더(encoder)와 원래 차원으로 복원해 출력신호(output signal)를 생성하는 디코더(decoder)로 구성된다. 이때, 입력신호와 출력신호의 차이(difference)로 재구성 오차가 발생하며, 이를 이용해 정상 신호와 이상 신호를 구분할 수 있다. 오토인코더 모델은 정상 신호를 이용해 학습하기 때문에 정상 신호와 입력신호의 유사 정도에 따라 재구성 오차 크기가 결정된다. 이상 탐지를 위한 오토인코더 모델은 Fig. 3과 같다.

진동 센서에서 수집된 원시 데이터(raw data)는 대부분 시간 흐름에 따른 진동 세기(amplitude)를 측정된 시계열 데이터이다. 이러한 시계열 데이터는 고속 푸리에 변환(Fast Fourier Transform, FFT)을 거쳐 주파수 영역으로 변환될 수 있으며, 하나의 신호(signal)에 존재하는 주파수별 중요도(magnitude)를 분석할 수 있다. 진동 신호의 주파수 영역 데이터를 오토인코더 모델의 입력 데이터로 사용할 경우, 전체 주파수 영역에 대해 재구성 오차가 계산되므로 진동 신호의 이상 발생 여부만 판별할 수 있다. 하지만, 설비 이상 유형에

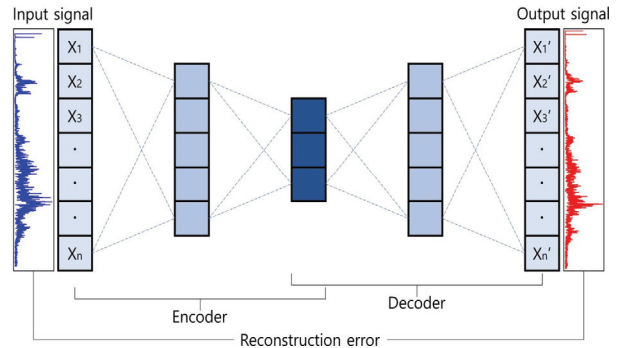


Fig. 3. Autoencoder Model for Anomaly Detection

Input:

- Autoencoder model M
- Frequency range $F = [f_0, \dots, f_{q-1}]$
- Vibration data $V = [v_0, \dots, v_{q-1}]$
- Training dataset $X = [V_0, \dots, V_{m-1}]$
- MAE loss $E = [E_0, \dots, E_{m-1}]$
- Number of divisions N

Output:

- Threshold $T = [T_0, \dots, T_{N-1}]$

```

01: begin
02:    $X' = M.predict(X)$ 
03:    $L = Len(F) / N$ 
04:   for  $i \in [0:N-1]$  do
05:      $NS = i \times L$ 
06:      $NE = NS + L - 1$ 
07:     for  $j \in [0:m-1]$  do
08:        $E[j][i] = Mean(Abs(X'[j][NS:NE] - X[j][NS:NE]))$ 
09:     end for
10:   end for
11:   for  $i \in [0:N-1]$  do
12:      $NS = i \times L$ 
13:      $NE = NS + L - 1$ 
14:      $T_i = Mean(E[:,NS:NE]) + 3 \times Std(E[:,NS:NE])$ 
15:   end for
16:   return  $T$ 
17: end
    
```

Fig. 4. Algorithm of Determining the Thresholds

따라 특정 주파수 영역에서 이상 패턴이 발생하므로 주파수 영역별로 이상 탐지를 수행할 필요가 있다. 주파수 영역을 N 개로 나눠 이상 탐지를 수행하는 N 분할 이상 탐지를 위한 임계치 결정 알고리즘과 이상 탐지 알고리즘 의사코드(pseudo code)는 Fig. 4, 5와 같다.

Fig. 4는 오토인코더 모델(M), 학습 데이터셋(X), 분할 수(N)를 이용해, 분할 주파수 영역의 N 개 임계치(T) 결정 과정을 나타낸다. 2, 3라인은 학습 데이터셋에 대한 모델 예측 결과(X')와 N 분할에 따른 주파수 영역 단위 길이(L)을 구한다. 4~10라인은 L 을 이용해 각 분할에서의 시작 주파수(NS), 끝 주파수(NE)에 따라, 분할 영역별로 재구성 오차(E)를 계산한다. 이때, E 는 학습 데이터 수(m , 행) \times 분할 수(N , 열)의 2차원

```

Input:
Autoencoder model M
Frequency range F = [f0,...,fq-1]
Vibration data V = [v0,...,vq-1]
Test dataset X = [V0,...,Vm-1]
MAE loss E = [E0[],...,Em-1[])
Number of divisions N
Threshold T = [T0,...,TN-1]

Output:
Anomaly signals AS = [AS0[],...,ASN-1[])
Normal signals NS = [NS0[],...,NSN-1[])

01: begin
02: X' = M.predict(X)
03: L = Len(F) / N
04: for i ∈ [0:N-1] do
05:   NS = i × L
06:   NE = NS + L - 1
07:   for j ∈ [0:m-1] do
08:     E[j][i] = Mean(Abs(X'[j][NS:NE] - X[j][NS:NE]))
09:   end for
10: end for
11: for i ∈ [0:N-1] do
12:   for j ∈ [0:m-1] do
13:     if E[j][i] > T[i] then
14:       AS[i].append(X[j])
15:     else
16:       NS[i].append(X[j])
17:     end if
18:   end for
19: end for
20: return AS, NS
21: end

```

Fig. 5. Algorithm of Anomaly Detection

배열 형태를 취한다. 11~15라인은 E의 전체 학습 데이터를 대상으로, N 영역별로 3 시그마 연산을 통해, N 개 임계치를 결정한다.

Fig. 5는 M, 평가 데이터셋(X), N, T를 이용해, 각 주파수 영역에서의 이상 신호(AS), 정상 신호(NS) 결정 과정을 나타낸다. 4~10라인은 X, X'를 이용해 E를 계산하는 과정이고, 11~19라인은 각 데이터 주파수 영역별 E와 학습 데이터를 통해 결정된 주파수 영역별 T를 비교해, E가 T보다 크면 AS, 작으면 NS로 결정한다. 결과적으로 AS와 NS에는 N개 주파수 영역에서의 이상, 정상 결과가 저장되며, 이때, 각 데이터의 각 주파수 영역은 이상, 정상 중 하나로 결정된다.

Fig. 4, 5의 입력 데이터와 출력(복원) 데이터의 재구성 오차는 식 (1) 평균절대오차(Mean Absolute Error, MAE)를 이용한다.

$$MAE = \frac{1}{n} \sum_{i=1}^n |X_i - X'_i| \quad (1)$$

3.2 이상 탐지 프로세스

제안 알고리즘을 이용한 전체 이상 탐지 프로세스는 Fig. 6과 같이 크게 4단계를 거쳐 수행된다.

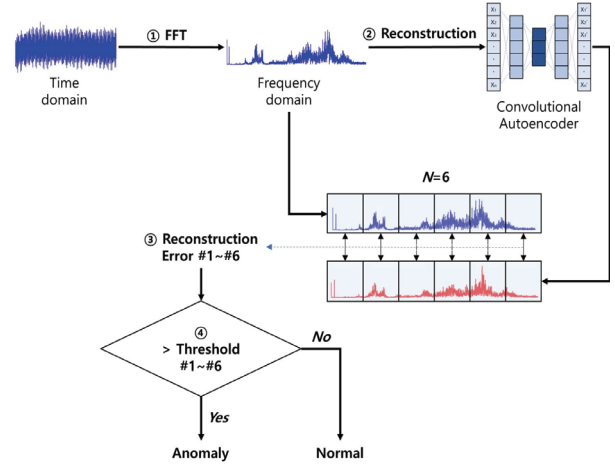


Fig. 6. The process of N-division anomaly detection

① 단계는 진동 데이터의 시간 영역 데이터를 FFT를 거쳐 주파수 영역으로 변환한다. 이때, 고역 통과 필터(high-pass filter), 저역 통과 필터(low-pass filter)를 적용해 데이터 특성이 반영된 특정 주파수 대역으로 제한할 수 있다. ② 단계는 정상 신호로 학습이 완료된 오토인코더 모델을 이용해 데이터의 주파수 영역을 재구성한다. ③ 단계는 식 (1)을 이용해 오토인코더 모델의 입력신호와 출력신호의 재구성 오차를 N 분할 영역별로 계산한다. Fig. 6은 6분할(N=6)을 적용한 예이다. ④ 단계는 ③ 단계에서 계산된 영역별 재구성 오차(#1~#6)를 영역별 임계치(#1~#6)와 각각 비교해 해당 주파수 영역에서 이상 여부를 결정한다.

4. 사례 연구

4.1 데이터셋

인쇄공정에 사용되는 압축기(compressor)에서 수집된 진동 데이터를 이용해 사례연구를 수행했다. 해당 데이터셋은 약 2달간 IEPPE(Integrated Electronics Piezo-Electric)타입 진동 센서를 이용해 3,906Hz 샘플링 속도(sampling rate)로 수집됐다. 압축기의 분당 회전수(Revolutions Per Minutes, RPM)는 1,740rpm으로 기본 주파수(fundamental frequency)는 29Hz를 나타낸다.

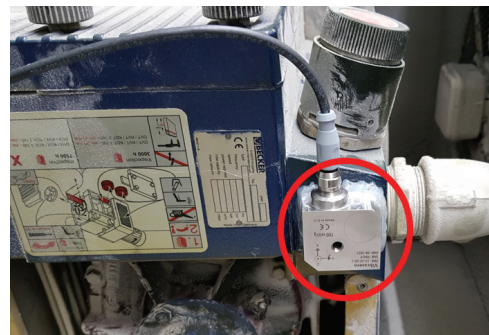


Fig. 7. Compressor and Sensor

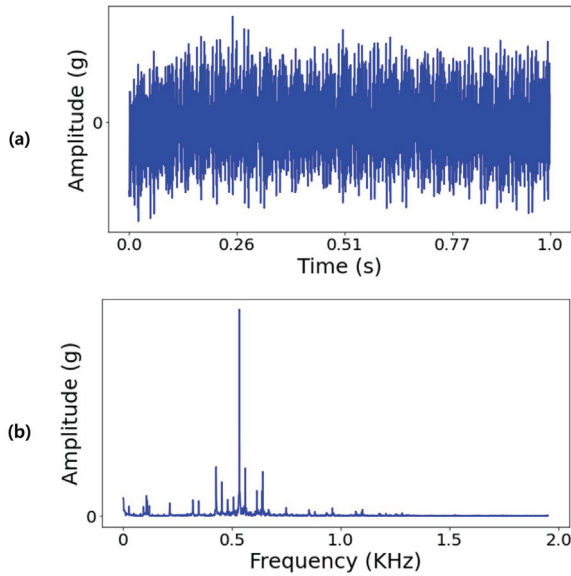


Fig. 8. Time Domain and Frequency Domain

데이터셋 규모(volume)는 4,102(row)×3,906(column) 이고, 이 중 2,401개(row)는 학습데이터로 사용하고 나머지 1,701개는 평가데이터로 활용했다. 수집된 데이터는 주파수 분석을 위해 Fig. 8과 같이 시간 영역(Fig. 8 (a)) 데이터를 FFT를 수행해 0~1,953Hz까지 주파수 영역(Fig. 8 (b))으로 변환하였으며, 고역 통과 필터를 적용해 3Hz 이상 주파수 신호만 추출해 사례연구에 사용했다.

4.2 합성곱 오토인코더

1차원 합성곱층(1-dimensional convolution layer)과 전치 합성곱층(transposed convolution layer)을 인코더, 디코더로 구성하는 합성곱 오토인코더(convolutional auto-encoder) 모델을 이용해 학습했으며, 모델 구성은 Fig. 9와 같고, 주요 초 매개변수(hyperparameter)는 Table 1과 같이 설정되었다.

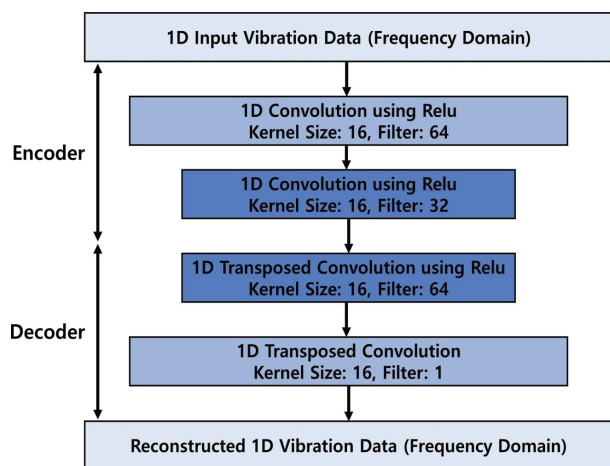


Fig. 9. Structure of the Model

Table 1. Hyperparameters of the Model

Hyperparameter	Value
Filters per layer	64, 32, 64, 1
Kernel size	16
Activation function	Relu
Strides	3
Epochs	100, Early stopping
Loss	MAE
Optimizer	Adam
Input shape	1,953×1

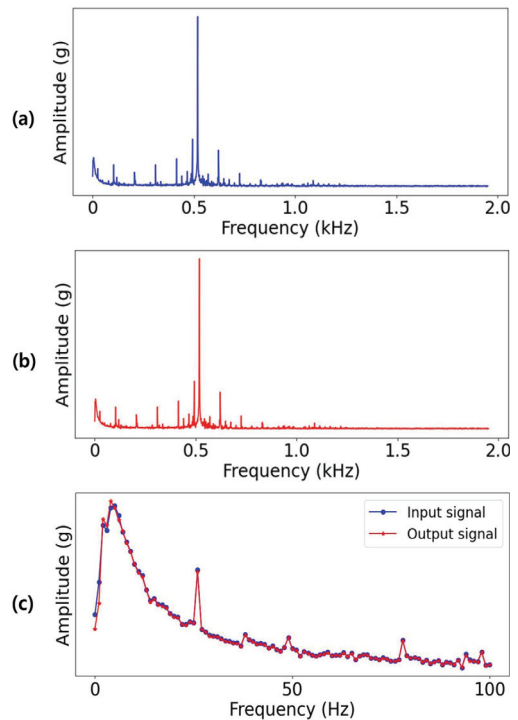


Fig. 10. Reconstructed Signal by the Model

학습이 종료된 합성곱 오토인코더 모델을 이용한 데이터 복원 결과 예는 Fig. 10과 같다. Fig. 10 (a)는 모델에 입력된 진동 주파수 영역 데이터이고 Fig. 10 (b)는 이를 모델이 예측(재구성)한 결과다. Fig. 10 (c)는 0~100Hz까지 재구성 결과를 상세히 가시화한 결과다.

2,401개 학습데이터와 1,701개 평가데이터를 대상으로 재구성 오차를 계산하고, 3 시그마 규칙을 적용해 임계치를 결정한 결과, 학습데이터의 2.6%(63개, Fig. 11 (a)), 평가데이터의 6.8%(116개, Fig. 11 (b))가 임계치보다 높은 재구성 오차를 나타냈다.

4.3 평가 데이터 생성

총 1,701개 평가데이터 중 임계치보다 높은 재구성 오차를 갖는 116개 데이터를 제외한 1,585개 데이터를 N 분할 이상 탐지 평가 데이터로 사용했다. 이상 데이터는 실제 현장에서

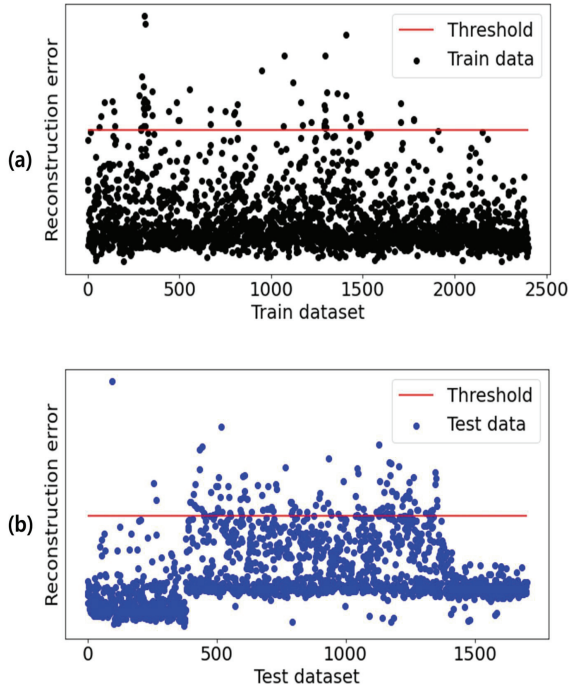


Fig. 11. Reconstruction Plots of the Model

충분한 수집이 어려운 경우, 설비 입력값 조절[23], 신호 변조 [24] 등 인위적으로 생성한 가상 데이터(synthetic data)를 모델 검증에 활용한다. 본 연구에서는 선택된 1,585개 정상 데이터에 대해 주파수와 강도를 달리한 9개 유형의 이상 데이터를 생성해 N 분할 이상 탐지를 수행했다. 결과적으로 1,585 × 9개 이상 데이터를 생성하고 실험에 활용하였다.

정상 데이터를 이용한 이상 데이터 생성과정은 Fig. 12와 같다. Fig. 12 (a)는 정상 신호, Fig. 12 (b), (d), (f)는 이상 데이터를 만들기 위한 주파수 신호, Fig. 12 (c), (e), (g)는 정상 신호와 이상 신호를 합성한 신호로 각각 이상 데이터로 사용했다. 이때, 정상 데이터는 시험 조건에 따른 주파수 신호와 합성하기 위해 식 (2)를 이용해 최소-최대 정규화(min-max normalization)을 수행하고, 식 (3)을 이용해 생성된 사인파(sine wave)를 최소 최대 정규화가 적용된 정상 데이터와 합성해 이상 데이터를 생성했다. 식 (3)에서 A는 주파수 강도(amplitude), f는 주파수(frequency), t는 시간(time)을 의미하며, A, f 값을 조절해 9종의 이상 신호를 생성하고 활용했다.

N 분할 이상 탐지 성능 평가 척도(metric)로 정상, 이상 신호에 대한 TP(True Positive), FN(False Negative)를 이용해 이상 신호 전체 시험 데이터 중 이상 신호로 예측한 비율인 TPR(True Positive Rate)[25]을 적용해 이상 신호 탐지 성능을 측정했다.

4.4 N 분할 이상 탐지 결과

주파수 30Hz에 대해서 수집된 진동 데이터의 최고 주파수 강도(amplitude) 대비 30% 수준 신호를 합성한 이상 데이터

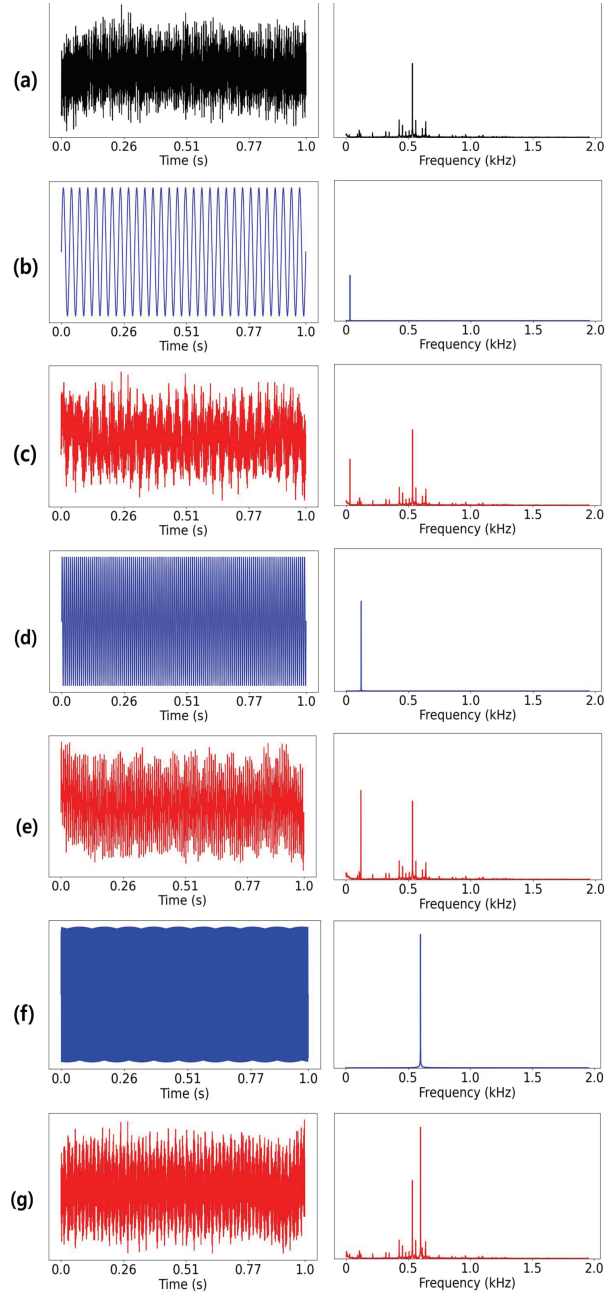


Fig. 12. Anomaly Signal Generation

$$x' = \frac{x - x_{min}}{x_{max} - x_{min}} \tag{2}$$

$$y(t) = A \sin(2\pi ft) \tag{3}$$

를 이용해 전체 주파수 영역에 대해 이상 탐지를 수행한 결과 TPR은 0.11로 나타났으며, 결과는 Fig. 13과 같다. 4.3절에서 설명한 바와 같이, 이상 신호는 정상 신호에 특정 주파수 신호를 합성한 신호로 모든 이상 신호는 정상 신호와 일대일 쌍(pair)을 이룰 수 있으며, Fig. 13에서는 이 쌍을 같은 인덱스(index)를 사용해 시각화(visualization)하였다. 따라서,

전체 이상 신호의 재구성 오차 분포는 정상 신호의 재구성 오차 분포와 유사한 형태를 취하면서 좀 더 높은 재구성 오차값을 갖는 형태로 나타났다. 결과적으로, 이상 신호 1,585개 중 TP(이상 신호를 이상 신호로 예측) 181, FP(이상 신호를 정상 신호로 예측) 1,404로 분할 이상 탐지를 수행하지 않은 경우, 정상 신호와 이상 신호를 효과적으로 구분하지 못하였다.

같은 조건에서 2, 4, 8, 16분할을 적용해 이상 탐지를 수행한 결과, 각 분할의 해당 주파수 영역에서의 TPR은 0.5(Fig. 14 (a)), 0.78(Fig. 14 (b)), 0.91(Fig. 14 (c)), 1(Fig. 14 (d))로 나타냈으며, 분할 수가 증가할수록 평가에 사용된 정상 신호와 이상 신호를 더 잘 구분하는 결과를 나타냈다.

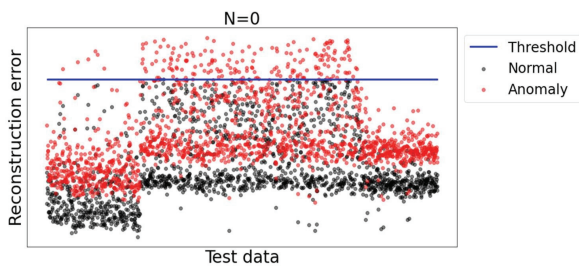


Fig. 13. Result of Anomaly Detection (N=0)

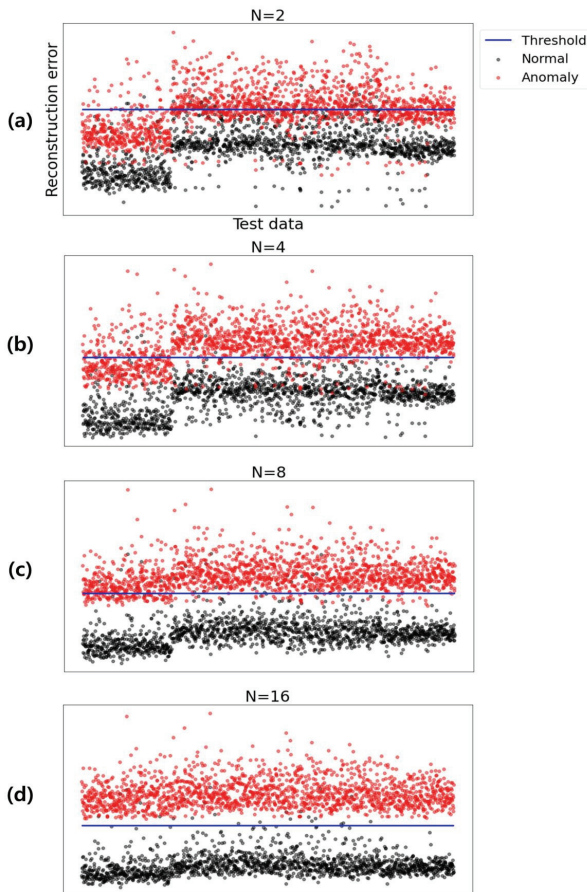


Fig. 14. Results of Anomaly Detection (N=2, 4, 8, 16)

이상 신호를 30Hz(1x), 120Hz(4x), 600Hz(20x) 3가지 주파수 신호와 최대 강도 대비 30%, 60%, 90% 조합으로 구성된, 총 9가지 평가 조건에서 기존 방식(N=0)과 N(2, 4, 8, 16) 분할 이상 탐지 성능은 Table 2와 같다. 이때, 해당 주파수가 포함된 영역에서 TPR을 측정했다. 예를 들어, Table 2의 Test 9, N=2일 때, TPR 0.98은 전체 주파수 0~1,953Hz에서 0~975Hz(Fig. 15 (a)), 976~1,953Hz(Fig. 15 (b)) 두 영역 중 600Hz가 포함된 0~975Hz 영역에서의 이상 탐지 결과를 의미한다.

N 분할 이상 탐지는 분할된 주파수 영역에서 이상 여부를 판단할 수 있다. 예를 들어, Test 3, N=4 전체 주파수 영역 4개 구간에 대한 이상 탐지 결과는 Fig. 16과 같으며, 이상 주파수 30Hz가 포함된 첫 번째 구간(Fig. 16 (a))에서 시험에 사용된 이상 신호를 모두 이상 신호로 예측했다. 특정 신호가 첫 번째 구간(Fig. 16 (a))에서 이상 신호로 예측되고, 나머지 구간(Fig. 16 (b), (c), (d))에서 정상으로 예측되었다면, 이 신호는 0~487Hz 영역(Fig. 17 (a))에서 이상이 발생했으며, 이외 영역(Fig. 17 (b), (c), (d))에서는 정상인 것으로 판단할 수 있다.

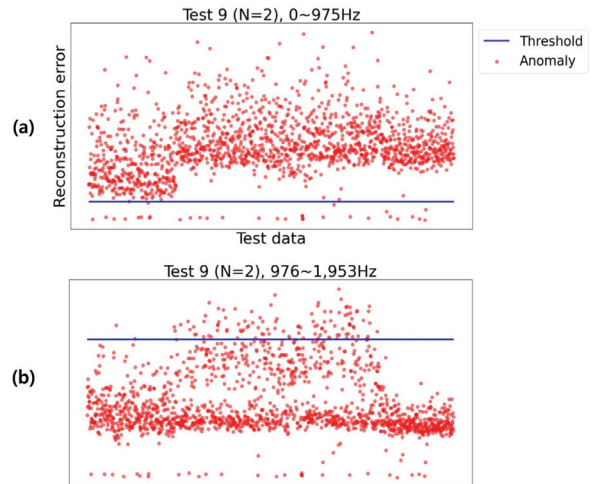


Fig. 15. Result of Test 9 (N=2)

Table 2. TPR of N-segmentation Method

Test #	Added frequency (Anomaly signal)		N				
	Frequency (Hz)	Amplitude (%)	0	2	4	8	16
1	30	30	0.11	0.50	0.78	0.91	1.0
2	30	60	0.30	0.99	1.0	1.0	1.0
3	30	90	0.92	1.0	1.0	1.0	1.0
4	120	30	0.03	0.09	0.06	0.05	0.04
5	120	60	0.07	0.20	0.25	0.24	0.16
6	120	90	0.13	0.46	0.74	0.81	0.47
7	600	30	0.07	0.20	0.54	0.98	1.0
8	600	60	0.20	0.81	0.99	1.0	1.0
9	600	90	0.42	0.98	1.0	1.0	1.0

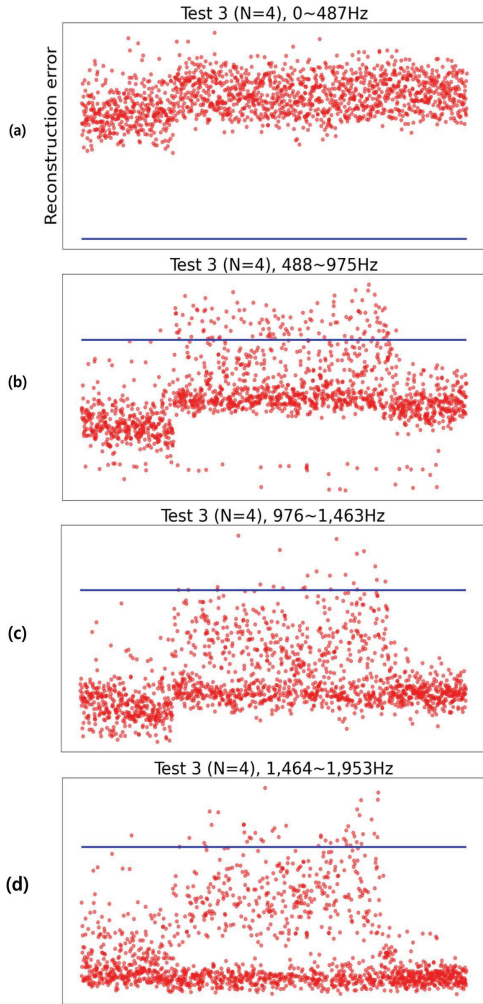


Fig. 16. Result of Test 3 (N=4)

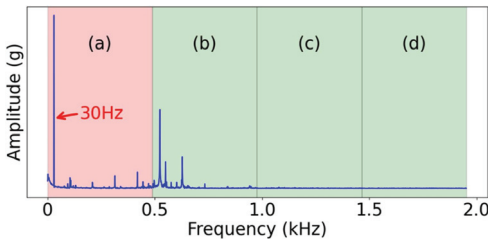


Fig. 17. Anomaly Detection by Frequency Region

Test 1~9에서 이상 주파수 신호가 포함된 영역을 이상으로 예측하고, 이외 나머지 주파수 영역은 정상으로 예측한 결과는 Table 3과 같다. 테스트 사례별 결과 편차가 심하게 나타났으며, 분할 수가 가장 높은 N=16 경우, 높은 TPR을 나타냈음에도, 나머지 영역에서 이상 신호로 판단된 경우가 많아 낮은 결과를 나타냈다. 또한, 테스트에 적용된 신호의 동일 주파수에서 진폭이 클수록, 분할 수가 증가할수록 더 낮은 결과가 나왔다. 이러한 결과는 이상 주파수의 강도가 커질수록 주변 주파수 진폭도 영향을 받아 모델이 주변 영역을 이상 신호로 예측한 결과로 판단된다.

Table 3. Results of Anomaly Detection

Test #	N			
	2	4	8	16
1	0.48	0.67	0.71	0.71
2	0.97	0.89	0.80	0.71
3	0.97	0.89	0.79	0.69
4	0.09	0.04	0.03	0.01
5	0.19	0.19	0.15	0.01
6	0.44	0.64	0.62	0.0
7	0.19	0.47	0.84	0.26
8	0.77	0.85	0.81	0.0
9	0.91	0.74	0.69	0.0

Table 4. Average TPR by Model

Model #	Hyperparameter	Avg. TPR				
	Number of layer (Filters per layer)	0	2	4	8	16
1	4 (64-32-64-1)	0.25	0.58	0.71	0.78	0.74
2	6 (128-64-32-64-128-1)	0.43	0.73	0.84	0.88	0.86
3	8 (256-128-64-32-64-128-256-1)	0.56	0.80	0.88	0.91	0.90

Table 1의 초 매개변수를 적용한 4개 층(Layer)으로 구성된 합성곱 오토인코더 모델(Table 4. Model 1)의 층을 6개(Table 4. Model 2), 8개(Table 4. Model 3)로 증가시켜, 주파수 영역 분할 수(N)별 평균 TPR을 측정된 결과는 Table 4와 같다. Model 3, 8분할 적용 시, 평균 0.91로 가장 좋은 성능을 나타냈다.

사례연구 결과는 다음과 같이 요약할 수 있다.

첫째, 동일 주파수 조건에서 강도가 30%, 60%, 90%로 높아질수록 TPR도 높게 나타났다. 이는 이상 신호의 주파수 강도가 높아질수록 재구성 오차가 더 크게 나타나, 결과적으로 이상 신호 탐지 성능이 향상된 것으로 해석된다. 둘째, N 분할을 하지 않은 경우(N=0)보다 N 분할을 적용한 이상 탐지 성능이 높게 나타났다. 이러한 결과는 N 분할 적용 시, 식 (1)에 따라 재구성 오차 범위(n)가 상대적으로 줄어들어 이상 신호에 대한 재구성 오차 변별력이 향상된 것으로 해석된다. 셋째, 최적 분할 수 N은 시험 사례마다 다르게 나타났다. 이는 시험에 사용된 주파수(30Hz, 120Hz, 600Hz)와 재구성 오차 계산에 사용되는 이웃 주파수의 분산(variance) 크기가 영향을 미치는 것으로 해석된다. 120Hz 시험 사례(Test 4~6) 경우, N 분할 이상 탐지 성능이 낮게 나타났으며, 이는 120Hz 근처 분산이 Fig. 18과 같이 상대적으로 크기 때문으로 해석된다. 가장 큰 분산을 나타낸 539Hz 경우, 강도 30%를 적용한 N=2일 때 TPR은 0.07로 매우 저조한 성능을 나타냈다.

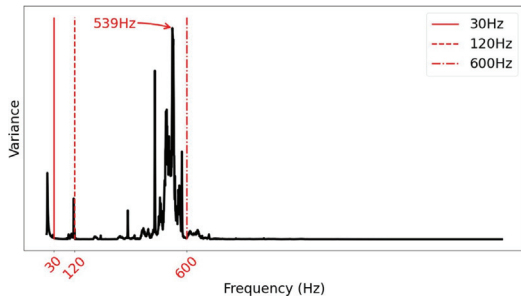


Fig. 18. Variance of Each Frequency

5. 결론 및 향후 연구

제조 설비의 효율적인 예지보전을 위해서는 설비 상태변화를 감지할 수 있어야 하며, 이를 위해 다양한 오토인코더 모델을 활용한 비지도 학습기반 이상 탐지 방법이 적용되고 있다. 대부분 제조 설비는 진동 데이터를 이용해 설비 상태를 효과적으로 예측할 수 있으며, 주파수 분석을 통해, 더욱 구체적인 설비 이상을 파악할 수 있다. 본 논문에서는 주파수 영역별 이상 탐지를 위해 N 분할 이상 탐지 알고리즘과 오토인코더 모델을 활용한 이상 탐지 프로세스를 제안하고, 압축기 진동 데이터와 9가지 주파수 신호를 합성한 이상 데이터를 대상으로 제안 방법을 평가했다.

본 연구의 결론은 다음과 같다. 첫째, N 분할 이상 탐지는 전체 주파수 영역을 대상으로 이상 탐지를 수행한 경우(N=0) 보다 전반적으로 더 나은 이상 탐지 성능을 나타냈다. 표 2의 9가지 테스트 데이터에 대해 N 분할을 적용하지 않았을 때 평균 TPR 0.25로 나타났지만, N 분할의 경우, 전체 평균 TPR 0.70 이상을 기록했다. 특히 낮은 주파수 강도(최대 강도 대비 30%, 60% 수준)에서도 2~3배 이상 높은 성능을 나타냈다 (Test 1, 2, 4, 5, 7, 8). 둘째, 분할 수(N) 증가가 반드시 이상 탐지 성능으로 이어지지는 않는다. 시험 조건에 따라 N 증가 시 이상 탐지 성능이 향상하는 경우(Test 1, 2, 7, 8, 9), 감소하는 경우(Test 4), 혼재하는 경우(Test 5, 6)로 나타났다. 셋째, N 분할 이상 탐지 성능은 주파수별 분산에 영향을 받는다. 주파수별 분산에 따라 이상 탐지 성능 차이가 발생했으며, 낮은 강도의 이상 신호에서 그 차이가 더 크게 발생했다(Test 1과 Test 4 비교). 분산이 작고 이상 신호 강도가 높은 경우 (Test 3)와 분산이 크고 이상 신호 강도가 낮은 경우(Test 4), N 증가에 따른 성능 변화는 크지 않았다.

본 연구의 한계와 향후 연구는 다음과 같다. 첫째, 가상의 단일 주파수 신호를 합성한 이상 데이터를 평가에 활용했다. 향후, 설비 불균형, 정렬 불량, 고정 불량 등 설비 이상 환경에서 수집된 데이터를 활용할 필요가 있다. 둘째, IEPF 타입 센서의 x축 진동에 해당하는 단일 데이터만 활용했다. 따라서, 추가적인 제조 데이터를 활용한 다변량 분석으로 확장될 수 있다. 셋째, 이미 장기간에 거쳐 운영 중인 압축기의 진동 데이터 사용으로 정상 상태에 대한 신뢰성이 확보되지 않았다.

설비 구축 후, 정상 상태에 대한 참조 데이터를 확보하고 활용할 필요가 있다. 넷째, 본 연구의 주파수 영역별 이상 탐지 결과를 바탕으로 최종적인 설비 상태 진단을 위해서는 도메인 지식이 추가로 요구된다.

References

- [1] Z. M. Çinar, A. Abdussalam Nuhu, Q. Zeeshan, O. Korhan, M. Asmael, and B. Safaei, "Machine learning in predictive maintenance towards sustainable smart manufacturing in industry 4.0," *Sustainability*, Vol.19, No.19, 2020. DOI: 10.3390/su12198211
- [2] S. Namuduri, B. N. Narayanan, V. S. P. Davuluru, L. Burton, and S. Bhansali, "Review-deep learning methods for sensor based predictive maintenance and future perspectives for electrochemical sensors," *Journal of The Electrochemical Society*, Vol.167, No.3, 2020. DOI: 10.1149/1945-7111/ab67a8
- [3] K. Park and Y. Lee, "Anomaly detection in combined driving system based on unsupervised learning," *Journal of the Korean Society for Precision Engineering*, Vol.40, No.11, 2023. DOI: 10.7736/JKSPE.023.068
- [4] Y. K. Lee, S. Lee, and S. H. Kim, "Real-time defect monitoring of laser micro-drilling using reflective light and machine learning models," *International Journal of Precision Engineering and Manufacturing*, Vol.25, 2024. DOI: 10.1007/s12541-023-00849-w
- [5] M. Pech, J. Vrchota, and J. Bednár, "Predictive maintenance and intelligent sensors in smart factory," *Sensors*, Vol.21, No.4, 2021. DOI: 10.3390/s21041470
- [6] N. Azeem and Y. Xiaoqing, "Experimental study on the Condition Monitoring of Shaft Unbalance by using Vibrations Spectrum and phase Analysis," in *Proceedings of the Condition Monitoring and Diagnosis*, 2018. DOI: 10.1109/CMD.2018.8535737
- [7] N. Azeem, X. Yuan, H. Raza, and I. Urooj, "Experimental condition monitoring for the detection of misaligned and cracked shafts by order analysis," *Advances in Mechanical Engineering*, Vol.11, No.5, 2019. DOI: 10.1177/1687814019851307
- [8] D. Mohamed, B. Rania, and H. Meftah, "Detection of unbalance and looseness faults in a ventilation turbine using vibration signature analysis," *IOP Conference Series: Materials Science and Engineering*, Vol.1244, No.1, 2022. DOI: 10.1088/1757-899X/1244/1/012022
- [9] S. A. Mohammed, N. M. Ghazaly, and J. Abdo, "Fault diagnosis of crack on gearbox using vibration-based approaches," *Symmetry*, Vol.14, 2022. DOI: 10.3390/sym14020417

- [10] O. Mey, W. Neudeck, A. Schneider, and O. Enge-Rosenblatt, "Machine learning-based unbalance detection of a rotating shaft using vibration data," *arXiv preprint*, 2020. DOI: 10.48550/arXiv.2005.12742
- [11] R. E. Bouisfi, L. E. Menzhi, and X. Chimentin, "Supervised classification of induction motors faults," in *Proceedings of the International Conference on Energy and Green Computing*, 2021. DOI: 10.1051/e3sconf/202233600051
- [12] T. Chen, X. Liu, B. Xia, W. Wang, and Y. Lai, "Unsupervised anomaly detection of industrial robots using sliding-window convolutional variational autoencoder," *IEEE Access*, Vol.8, pp.47072-47081, 2020. DOI: 10.1109/ACCESS.2020.2977892
- [13] S. Givnan, C. Chalmers, P. Fergus, S. Ortega-Martorell, and T. Whalley, "Anomaly detection using autoencoder reconstruction upon industrial motors," *Sensors*, Vol.22, 2022. DOI: 10.3390/s22093166
- [14] Y. Huang, C. -H. Chen and C. -J. Huang, "Motor fault detection and feature extraction using RNN-Based variational autoencoder," *IEEE Access*, Vol.7, pp.139086-139096, 2019. DOI: 10.1109/ACCESS.2019.2940769
- [15] T. Tziolas, K. Papageorgiou, T. Theodosiou, E. Papageorgiou, T. Mastos, and A. Papadopoulos, "Autoencoders for anomaly detection in an industrial multivariate time series dataset," *Engineering Proceedings*, Vol.18, 2022. DOI: 10.3390/engproc2022018023
- [16] S. Zhang, F. Ye, B. Wang, and T. G. Habetler, "Semi-supervised learning of bearing anomaly detection via deep variational autoencoders," *arXiv preprint*, 2019. DOI: 10.48550/arXiv.1912.01096
- [17] I. Thoidis, M. Giouvanakis, and G. Papanikolaou, "Semi-supervised machine condition monitoring by learning deep discriminative audio features," *Electronics*, Vol.10, 2021. DOI: 10.3390/electronics10202471
- [18] Ö. Yilmaz, M. Aksoy, and Z. Kesilmiş, "Misalignment fault detection by wavelet analysis of vibration signals," *International Advanced Researches and Engineering Journal*, Vol.3, No.3, 156-163, 2019. DOI: 10.35860/iarej.451528
- [19] B. Cahyono, D. Priyanta, and F. R. F. Ramadhan, "Vibration spectrum analysis for indicating damage on turbine and steam generator amurang unit 1," *International Journal of Marine Engineering Innovation and Research*, Vol.2, 2017. DOI: 10.12962/j25481479.v2i1.2688
- [20] S. Ahmad, K. Styp-Rekowski, S. Nedelkoski, and O. Kao, "Autoencoder-based condition monitoring and anomaly detection method for rotating machines," in *Proceedings of IEEE International Conference on Big Data*, 2020.
- [21] I. Ahmed, M. Ahmad, A. Chehri, and G. Jeon, "A Smart-Anomaly-Detection System for Industrial Machines Based on Feature Autoencoder and Deep Learning," *Micro-machines*, Vol.14, 2023. DOI: 10.3390/mi14010154
- [22] M. Rao, M. J. Zuo, and Z. Tian, "A speed normalized auto-encoder for rotating machinery fault detection under varying speed conditions," *Mechanical Systems and Signal Processing*, Vol.189, 2023. DOI: 10.1016/j.ymsp.2023.110109
- [23] O. Serradilla, E. Zugasti, J. Ramirez de Okariz, J. Rodriguez, and U. Zurutuza, "Adaptable and explainable predictive maintenance: Semi-supervised deep learning for anomaly detection and diagnosis in press machine data," *Applied Sciences*, Vol.11, 2021. DOI: 10.3390/app11167376
- [24] T. Tziolas, K. Papageorgiou, T. Theodosiou, E. Papageorgiou, T. Mastos, and A. Papadopoulos, "Autoencoders for anomaly detection in an industrial multivariate time series dataset," *Engineering Proceedings*, Vol.18, 2022. DOI: 10.3390/engproc2022018023
- [25] M. Z. Naser and A. Alavi, "Insights into performance fitness and error metrics for machine learning," *arXiv preprint*, 2020. DOI: 10.48550/arXiv.2006.00887

박기창



<https://orcid.org/0000-0002-6122-2331>

e-mail : kc.park@reshenie.co.kr

2001년 전남대학교 기초과학부(학사)

2003년 전남대학교 전산학과(석사)

2013년 전남대학교 전산학과(박사)

2008년 ~ 2018년 동신대학교 산학협력단
책임연구원

2019년 ~ 2022년 연세대학교 산학협력단 연구교수

2022년 ~ 현 재 (주)리쎬이에 제조자동화기술연구소 팀장

관심분야 : 소프트웨어공학, 인공지능

이용관



<https://orcid.org/0000-0002-6795-6723>

e-mail : ivan.lee@tukorea.ac.kr

1994년 금오공과대학교 정밀기계공학(학사)

1996년 금오공과대학교 기계공학과(석사)

2004년 St.Petersburg State

Polytechnic Univ. 시스템해석
및 제어학과(박사)

1996년 ~ 2001년 (주)현대로템 중앙연구소 주임연구원

2004년 ~ 2020년 (주)삼성전기 글로벌기술센터 수석연구원/그룹장

2020년 ~ 현 재 한국공학대학교 그랜드ICT연구센터

산학협력중점교수

관심분야 : 제조인공지능, 설비진단기술 및 스마트제조 솔루션