

# Improving the Performance of Machine Learning Models for Anomaly Detection based on Vibration Analog Signals

Jaehun Kim\* · Sangcheon Eom\*\* · Chulsoon Park\*<sup>†</sup>

\*Department of Industrial & Systems Engineering, Changwon National University

\*\*Department of Industrial Engineering, Pusan National University

## 진동 아날로그 신호 기반의 이상상황 탐지를 위한 기계학습 모형의 성능지표 향상

김재훈\* · 엄상천\*\* · 박철순\*<sup>†</sup>

\*국립창원대학교 산업시스템공학과

\*\*부산대학교 산업공학과

New motor development requires high-speed load testing using dynamo equipment to calculate the efficiency of the motor. Abnormal noise and vibration may occur in the test equipment rotating at high speed due to misalignment of the connecting shaft or looseness of the fixation, which may lead to safety accidents. In this study, three single-axis vibration sensors for X, Y, and Z axes were attached on the surface of the test motor to measure the vibration value of vibration. Analog data collected from these sensors was used in classification models for anomaly detection. Since the classification accuracy was around only 93%, commonly used hyperparameter optimization techniques such as Grid search, Random search, and Bayesian Optimization were applied to increase accuracy. In addition, Response Surface Method based on Design of Experiment was also used for hyperparameter optimization. However, it was found that there were limits to improving accuracy with these methods. The reason is that the sampling data from an analog signal does not reflect the patterns hidden in the signal. Therefore, in order to find pattern information of the sampling data, we obtained descriptive statistics such as mean, variance, skewness, kurtosis, and percentiles of the analog data, and applied them to the classification models. Classification models using descriptive statistics showed excellent performance improvement. The developed model can be used as a monitoring system that detects abnormal conditions of the motor test.

**Keywords :** Hyperparameter Optimization, Response Surface Method, Anomaly Detection, Vibration Measurement, Supervised Machine Learning, Descriptive Statistics

### 1. 서 론

진동 데이터를 이용한 모터 상태 모니터링에서 진동 데이

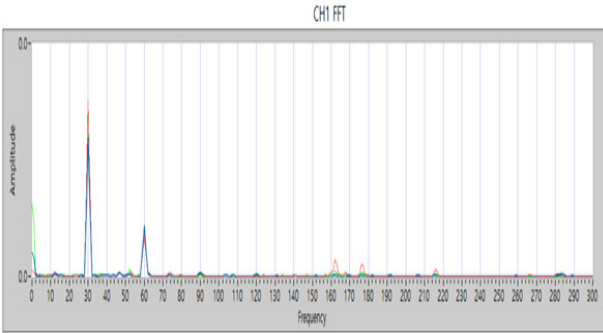
터가 수집되면 고장 진단을 위해서 일반적으로 FFT를 이용한 주파수분석, ISO 20816-1 기반의 기계진동 평가, 기계학습 등의 세 가지 방법을 사용할 수 있다. 먼저, <Figure 1>에 표시된 주파수분석은 FFT(Fast Fourier Transform)를 이용하여 신호에 대한 주파수 패턴을 도출하여 이상원인이나 모터 내부 베어링의 고장 원인을 찾아낼 수 있다[13]. 둘째,

Received 8 March 2024; Finally Revised 16 April 2024;

Accepted 17 April 2024

<sup>†</sup> Corresponding Author : cspark@changwon.ac.kr

ISO 20816-1에 기반한 기계진동 평가 방법은 <Figure 2>에 표시된 것과 같이 모터 용량에 따른 진동의 속도 값을 기준으로 모터의 이상 여부를 판단할 수 있다. 셋째, 기계학습을 이용한 진단에서는 아날로그 신호에 대해 샘플링을 통해 디지털화 및 특성치(Features)를 도출하여 이를 이용한 분류모형을 개발하여 분석할 수 있다[11, 23].



<Figure 1> FFT-based Frequency Analysis

ISO/DIS 20816-3	Medium-size Machines	Large Machines	
Advisor	Group 2	Group 1	
Velocity	Rated Power		
mm/s RMS	15 kW - 300 kW	300 kW - up	
11.0	DAMAGE OCCURS		
7.1	RESTRICTED OPERATION		
4.5	UNRESTRICTED OPERATION		
3.5			
2.8			
2.3			
1.4	NEWLY COMMISSIONED MACHINERY		
0.7			
0.0 Foundation	Rigid	Flexible	Rigid
			Flexible

<Figure 2> ISO 20816-1 Guideline [17]

최근 모터의 이상상황 또는 고장탐지를 위해서 기계학습 기반의 이상상황 및 고장탐지 방법이 많이 사용되고 있다[3, 7, 12, 22, 23]. 이 방법은 기존의 물리적 해석적인 방법과 비교하여 신호 해석 등을 위한 전문가가 필요하지 않으며, 수집된 데이터를 이용하면 실시간으로 모터의 이상상황을 탐지해 낼 수 있다.

하지만, 진동, 소음, 전류 등의 아날로그 신호의 경우 기계학습 모형을 만들기 위해서는 샘플링을 통해 아날로그 신호를 디지털화해야 한다. 이때, 아날로그 신호에 대한 샘플링 값의 경우 신호의 숨겨진 패턴을 반영하지 못할 수 있다. 따라서 신호에 대한 샘플링 값을 이용하여 학습 모형을 만드는 경우 만족스럽지 못한 성능을 보여주는 경향이 있다.

따라서, 모터의 진동 샘플링을 이용한 이상상황 탐지 및 상태에 대한 예측 모델의 성능을 향상시키기 위해서는 하이퍼파라미터 최적화를 적용할 수 있다. 하이퍼파라미터 최적화는 모형의 최상의 성능을 가져오는 하이퍼파라미터 조합을 찾기 위해 탐색을 수행하는 것을 의미하며, 기계학습 모형을 구축하는 데 있어 중요한 프로세스 중 하나이다 [13]. 하이퍼파라미터 최적화를 위한 일반적인 기법으로는 그리드 탐색(Grid Search), 랜덤 탐색(Random Search), 베이저안 최적화(Bayesian Optimization) 등이 있으며, 또한 실험계획법에 기반한 반응표면분석(RSM)을 최적화에 적용할 수 있다.

그리드 탐색은 가장 간단하고 직관적인 방법이지만, 탐색 시간이 많이 소요되는 단점이 있다. Hwang[5], You et al.[21], Kwon et al.[9]은 학습 모델의 하이퍼파라미터 최적화를 위해 그리드 탐색을 적용하는 방안을 제시하였다. Park et al.[11], Lee et al.[14], Vulpe-Grigorasi et al.[18]은 예측 모델의 정확성을 확보하기 위해 랜덤 탐색을 적용하여 하이퍼파라미터를 최적화하는 방안을 검증하였다. Choi et al.[1], Kim et al.[8], Hwang et al.[4]은 분류모형에 베이저안 최적화를 적용한 예측 기법을 제안하였다.

한편, 하이퍼파라미터 최적화를 위해 실험계획법에 기반한 방안을 제시한 연구도 있다[2]. 먼저 요인 분석을 통해 주요 하이퍼파라미터를 찾아낸 후 반응표면분석을 이용하여 하이퍼파라미터를 최적화하는 방법을 사용한다. 또한 Lee et al.[10]은 이미지 분류 문제에서 합성곱 신경망의 하이퍼파라미터 최적값을 찾기 위한 실험계획을 적용하는 방안을 제안하였는데, 높은 성능을 갖는 값을 랜덤 탐색으로 찾아낸 후 그리드 탐색을 적용함으로써 정확도를 높일 수 있음을 보였다.

본 연구에서는 진동 센서를 통해 수집된 아날로그 데이터를 적용한 기계학습 모형을 개발하였다. 일반적으로 기계학습 모형을 적용하기 위해서는 95% 이상의 분류 정확도를 요구하고 있다. 하지만, 본 연구의 경우 분류 정확도가 93% 안팎의 값을 나타내고 있어 이를 개선할 방안이 필요하였다.

기계학습 모형에서는 일반적으로 Raw 데이터를 이용하여 학습을 수행한 후 분류 정확도 등의 지표값이 낮은 경우 하이퍼파라미터 조정이 필수적이다. 기계학습 알고리즘의 하이퍼파라미터를 조정하기 위한 여러 가지 방법이 있으며, 본 연구에서는 기계학습에서 제공하는 대표적인 하이퍼파라미터 최적화 방법인 그리드 탐색, 랜덤 탐색, 베이저안 최적화 방법을 수행한다. 본 연구에서는 추가로 직교배열표를 이용하여 효과가 큰 인자(하이퍼파라미터)를 선택하고 선택된 인자의 탐색 범위에서 반응표면분석 기반의 최적화를 시도한다. 직교배열표는 실험의 크기를 키우지 않고 실험에 많은 인자를 적용할 수 있으므로 효율적이다. 또한 불필요한 교호작용을 구하지 않음으로써 실

험 횟수를 적게 할 수 있다. 그러나, 아날로그 신호 데이터는 샘플링 데이터만을 통해서 패턴을 찾기가 어려우므로, 주기를 고려한 샘플링 및 분포를 반영할 수 있는 기술 통계치를 통해서 학습을 수행할 필요가 있다. 따라서 본 논문은 위와 같은 절차에 의해 연구를 수행하였다.

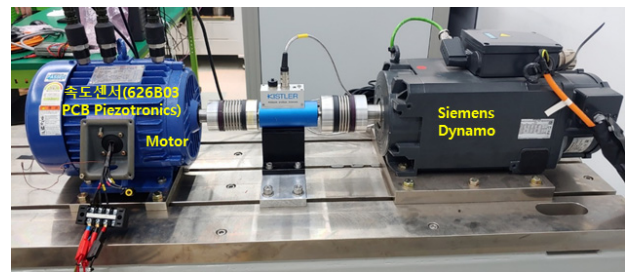
먼저, 진동 센서를 이용하여 진동 신호에 대한 샘플링을 통해 아날로그 데이터를 수집한다. 둘째, 수집된 데이터를 이용하여 분류모형을 개발한다. 셋째, 분류 정확도를 개선하기 위하여 그리드 탐색, 랜덤 탐색, 베이지안 최적화, 반응표면분석을 적용한다. 이때 위에 언급된 하이퍼파라미터 최적화 기법을 통해서 분류 정확도 개선에 한계가 있는 것으로 판명되었다. 따라서, 아날로그 신호 데이터의 숨겨진 특성치 패턴을 도출하기 위해 평균, 분산, 왜도, 첨도, 백분위수 등의 기술통계치를 구해서 이를 분류모형에 적용하였다.

본 논문의 구성은 다음과 같다. 제2장에서는 모터 다이내모 시험을 통한 데이터 수집과 각 분류모형에 대한 하이퍼파라미터를 도출한다. 제3장에서는 하이퍼파라미터 최적화 과정을 통해 도출된 하이퍼파라미터 값을 제시하고, 그 결과를 논의한다. 제4장에서는 기술통계치를 적용한 분류모형에 관해 설명하고, 마지막으로 제5장에서는 제안된 방법론의 한계점과 추후 연구에 대한 제시와 함께 결론을 제시하도록 한다.

## 2. 데이터셋 및 하이퍼파라미터 선정

### 2.1 데이터셋 수집 및 분류모형 학습

본 연구에서는 데이터 수집을 위해 <Figure 3>과 같이 다이내모 시험장치를 이용하여 시료 모터 시험을 실시하였다. 먼저 테스트베드에 시료 모터를 느슨하게 고정하는 경우, 고정볼트를 과하게 조여서 misalignment가 발생하는 경우, 적절하게 고정한 정상적인 경우에 대한 진동 데이터를 수집하였다. 진동 센서는 시료 모터의 베어링 위치의 표면에 X, Y, Z 방향으로 부착하여 진동 데이터를 수집하였으며 수집된 데이터 중 일부는 <Table 1>과 같다.



<Figure 3> Testbed to Collect Vibration Data

<Table 1> Examples of collected data

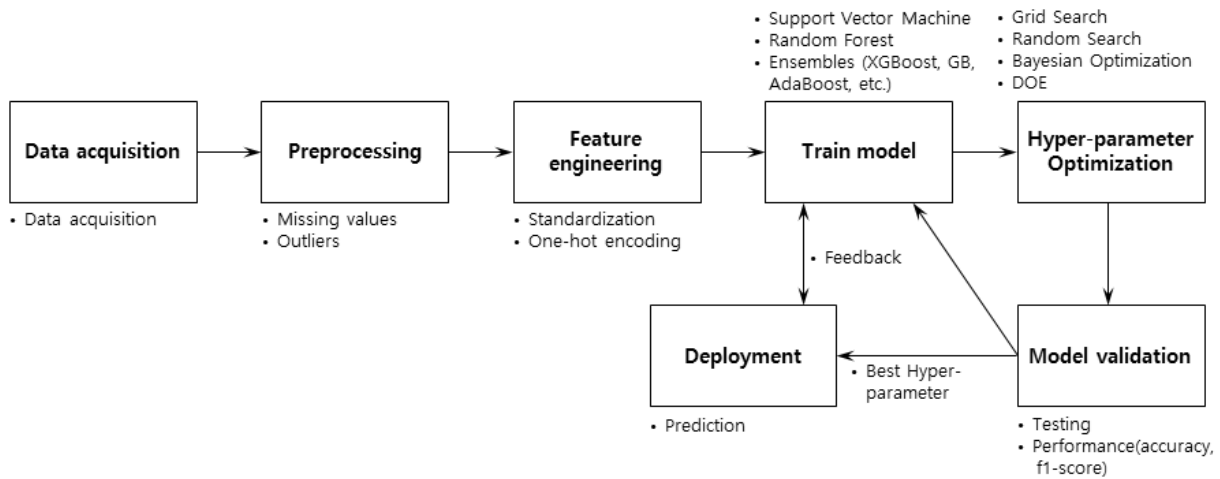
Time_sec	Vibration1_Velocity	Vibration2_Velocity	Vibration3_Velocity
0	0.1145	0.1343	0.2706
0.1	0.1188	0.1577	0.2077
0.2	0.0856	0.1450	0.1680
0.3	0.0978	0.1456	0.1680
0.4	0.1125	0.1345	0.1703
0.5	0.1366	0.1222	0.1679
0.6	0.1071	0.1611	0.2018
0.7	0.1932	0.1660	0.2832
0.8	0.1740	0.1441	0.2612
0.9	0.1575	0.2515	0.2382
1.0	0.1068	0.1329	0.1991
1.1	0.0956	0.1357	0.2441
1.2	0.2079	0.2563	0.1572
1.3	0.1738	0.2112	0.1840
1.4	0.1948	0.2022	0.1588
1.5	0.1572	0.2104	0.2122
1.6	0.1278	0.1526	0.1530
1.7	0.1317	0.1363	0.2049
1.8	0.1651	0.1281	0.1998
1.9	0.1391	0.1253	0.1569
2.0	0.1274	0.1712	0.1967
2.1	0.1671	0.1184	0.2839
2.2	0.1002	0.2026	0.2145

다이나모 시험 중 이상상황 탐지를 위해서 지도학습 (Supervised Learning) 방법 중 Support Vector Machine (SVM), Random Forest(RF), Adaptive Boosting(AdaBoost) 등의 분류모형을 사용하였다. 수집된 정상 데이터와 비정상 데이터에 대해 라벨링을 수행한 후 특성치 값들이 평균이 0, 분산이 1을 갖도록 표준화하였다. 모형 학습 단계에서는 수집 데이터의 70%를 학습데이터(train data)로 사용하여 분류모형을 만들고, 나머지 30%(test data)를 사용하여 학습모형을 검증하였다. 이때 분류모형의 정확도는 93% 안팎의 값을 보여주었기 때문에 정확도 향상을 위해 분류모형별 하이퍼파라미터의 최적화 작업을 진행할 필요

가 있었다.

### 2.2 하이퍼파라미터 선정

본 연구에서는 <Figure 4>에 표시된 것과 같은 절차에 따라 학습모형을 개발하였다. 도출된 학습모형에 대해서는 성능지표 개선을 위해 그리드 탐색, 랜덤 탐색, 베이지안 최적화 등의 하이퍼파라미터 최적화 기법을 이용하여 모형 최적화를 진행하였다. 각 분류모형에 대해 주요 인자로 알려진 하이퍼파라미터를 <Table 2>와 같이 선정하였다[20].



<Figure 4> Architecture of Machine Learning

<Table 2> Hyperparameters

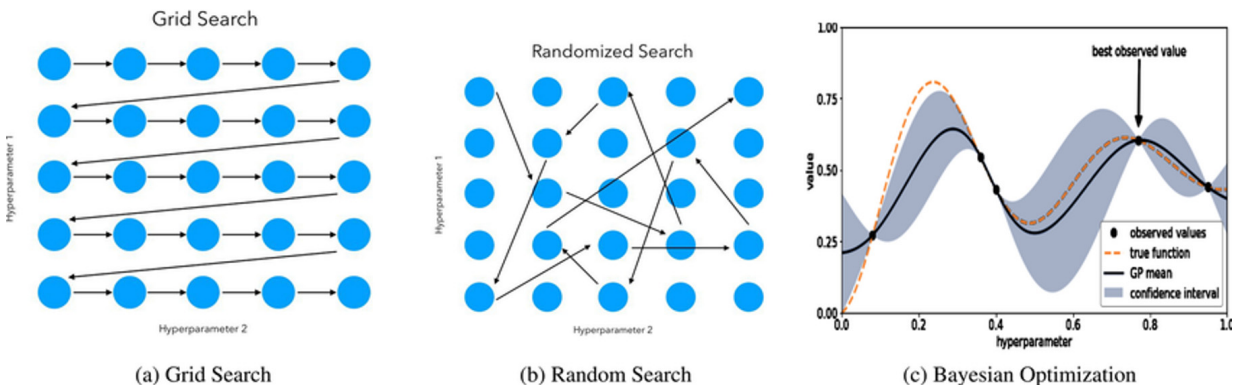
Model	Hyperparameter	
SVM	Kernel	Specifies the kernel type to be used in the algorithm.
	Cost	Regularization parameter. The strength of the regularization is inversely proportional to Cost.
	Gamma	How much influence a single training example has. The larger gamma is, the closer other examples must be to be affected.
RF	n_estimators	The number of trees in the forest.
	max_features	The number of features to consider when looking for the best split.
	max_depth	The maximum depth of the tree.
	min_samples_leaf	The minimum number of samples required to be at a leaf node.
	min_samples_split	The minimum number of samples required to split an internal node.
	bootstrap	Whether bootstrap samples are used when building trees.
AdaBoost	n_estimators	The maximum number of estimators at which boosting is terminated.
	base_estimator	The base estimator from which the boosted ensemble is built.
	learning_rate	Weight applied to each classifier at each boosting iteration.
	algorithm	If 'SAMME.R' then use the SAMME.R real boosting algorithm. The SAMME.R algorithm typically converges faster than SAMME, achieving a lower test error with fewer boosting iterations.

### 3. 하이퍼파라미터 최적화 수행

일반적으로 분류 정확도를 개선하기 위해서는 그리드 탐색, 랜덤 탐색, 베이지안 최적화 등의 방법을 사용한다. <Figure 5>는 그리드 탐색, 랜덤 탐색, 베이지안 최적화 방법의 탐색 구간을 진행하는 방법을 나타내고 있다. 그리드 탐색은 미리 정의 되어 있는 하이퍼파라미터 구간에서 일정 간격으로 하이퍼파라미터 값을 선택하여 성능지표를 확인하고, 가장 높은 성능지표를 갖는 하이퍼파라미터값을 최적해로 선택한다. 그리드 탐색과 달리 하이퍼파라미터 구간 내에서 임의로 값을 선택하는 랜덤 탐색은 반복 탐색 구간을 줄여 빠르게 최적의 하이퍼파라미터 값을 찾는 방법이다. 탐색 구간을 구성할 때 전 구간을 대상으로 하므로 초기해를 찾는 데 유리하다. 또한 랜덤 탐색 후 세분화하여 그리드 탐색으로 최적화 진행이 가능하다. 베이지안 최적화는 손실함수(Loss function)를 설정한 후 손실함수를 최소화하는 방향으로 하이퍼파라미터 값을 탐색한다[16].

본 연구에서는 하이퍼파라미터의 탐색 범위 설정은 수준간 차이를 크게 해서 성능에 영향을 미치는 주요 범위를 스크리닝하기 위해 <Table 3>과 같이 선정하였다[2]. 세 가지 최적화 방법으로 구한 최적 하이퍼파라미터 값은 <Table 4>와 같으며, 모형별 분류 정확도가 가장 높은 하이퍼파라미터 조합을 보여주고 있다. 최적 하이퍼파라미터를 이용한 분류모형의 정확도는 <Table 10>의 두 번째 열, 세 번째 열, 네 번째 열에 각각 표시되어 있으며, 여전히 93% 안팎에 머물러 있다. 따라서, 위와 같은 세 가지 방법에 따른 하이퍼파라미터 최적화로는 원하는 성능지표 값을 달성할 수 없으므로, 추가로 반응표면분석을 이용한 하이퍼파라미터 최적화를 수행하였다.

본 연구에서는 주효과와 2인자 교호작용을 검출하고 실험 횟수를 줄일 수 있는 실험계획법인 직교배열표를 적용하였으며, 유의한 요인 인자들을 판별하기 위해 분산분석(ANOVA)을 수행하였다. 분산분석으로부터 유의한 인자를 도출한 후 인자에 대한 최적치를 찾기 위해 반응표면분석을 수행하였다.



<Figure 5> Comparison of three hyperparameter search methods[6]

<Table 3> Hyperparameter Set for Optimization Process

Model	Hyperparameter	Grid Search	Random Search	Bayesian Optimization
SVM	Kernel	['rbf', 'linear']	['rbf', 'linear']	['rbf', 'linear']
	Cost	[0.1, 1, 100, 3000]	(0.1, 3000)	(0.1, 3000)
	Gamma	[0.001, 0.01, 0.1]	(0.001, 0.1)	(0.001, 0.1)
RF	n_estimators	[10, 100, 200, 500]	randint(10, 500)	(10, 500)
	max_features	[1, 3]	randint(1,3)	(1,3)
	max_depth	[10, 100]	randint(10, 100)	(10, 100)
	min_samples_leaf	[1, 10, 29]	randint(1,29)	(1,29)
	min_samples_split	[2, 5, 10]	randint(2, 10)	(2, 10)
	bootstrap	[True, False]	[True, False]	[True, False]
AdaBoost	n_estimators	[10, 100, 500, 1000]	stats.uniform(10, 1000)	(10, 1000)
	base_estimator	[1, 10, 29]	stats.uniform(1, 29)	(1, 29)
	learning_rate	[0.01, 0.05, 0.1, 1.0]	stats.uniform(0.01, 1.0)	(0.01, 1.0)
	algorithm	["SAMME", "SAMME.R"]	["SAMME", "SAMME.R"]	["SAMME", "SAMME.R"]

<Table 4> Result of Hyperparameter Optimization

Model	Hyperparameter	Grid Search	Random Search	Bayesian Optimization
SVM	Kernel	rbf	rbf	rbf
	Cost	100	23.57	51.07
	Gamma	0.1	0.022	0.065
RF	n_estimators	500	454	300
	max_features	1	1	1
	max_depth	10	16	96
	min_samples_leaf	1	8	26
	min_samples_split	2	5	8
	bootstrap	False	True	True
AdaBoost	n_estimators	500	444	440
	base_estimator	1	1	2
	learning_rate	1.0	0.63	0.19
	algorithm	SAMME.R	SAMME.R	SAMME

반응표면분석을 통한 최적화를 위해 먼저, 하이퍼파라미터와 그 범위를 선택한다. 둘째, 직교배열표를 이용하여 주요 인자(하이퍼파라미터)를 식별해 낸다. 셋째, 식별된 인자를 사용하여 반응표면분석을 수행하고, 반응의 변화가 원하는 목표치가 될 때까지 모델 최적화를 진행한다. 직교배열표에서 사용한 인자와 수준은 <Table 5>와 같다.

<Table 5> Hyperparameter Set of Orthogonal array

Model	Hyperparameter	Level
SVM	Kernel	['rbf', 'linear']
	Cost	[0.1, 3000]
	Gamma	[0.001, 0.1]
RF	n_estimators	[10, 500]
	max_features	[1, 3]
	max_depth	[10, 100]
	min_samples_leaf	[1, 29]
	min_samples_split	[2, 10]
	bootstrap	[True, False]
AdaBoost	n_estimators	[50, 1000]
	base_estimator	[1, 29]
	learning_rate	[0.01, 1.0]
	algorithm	["SAMME", "SAMME.R"]

본 연구에서는 2 수준계 직교배열표를 적용하였으며, 3 인자 이상의 교호작용은 없을 것으로 판단하여 제외시켰다. SVM, RF, AdaBoost 모형에 2수준계 직교배열표를 적용하였고, 그중에서 <Table 6>은 Random Forest에 사용되는 하이퍼파라미터에 대한 직교배치 요인과 수준을 보여주고 있다. <Table 6>의 인자와 수준을 적용한 2 수준계

직교배열 설계의 분산분석 결과는 <Table 7>과 같으며, 유의수준 0.05 기준으로 유의하지 않은 인자는 풀링(pooling)하였다. 주효과 B와 C는 유의수준 0.05에서 유의하지 않지만, 교호작용(B\*C)에서 유의하므로 인자 B, C, D, F를 사용하여 반응표면분석을 실시하였다. 본 연구에서 최적화할 응답 변수는 정확도(accuracy)이며, 정확도는 크면 클수록 좋은 망대특성이므로 반응표면분석을 통해 얻은 인자의 값은 <Table 8>과 같다.

<Table 6> Levels of Orthogonal array(RF)

Factor		low level	high level
n_estimators	A	10	500
max_features	B	1	3
max_depth	C	10	100
min_samples_leaf	D	1	29
min_samples_split	E	2	10
bootstrap	F	True	False

<Table 7> Result of ANOVA(pooling)

Factor	Adj SS	Adj MS	F-Value	P-Value
B	0.000000	0.000000	0.18	0.676
C	0.000004	0.000004	2.71	0.113
D	0.000026	0.000026	16.96	0.000**
F	0.000023	0.000023	14.53	0.001**
B*C	0.000014	0.000014	9.07	0.006*
B*F	0.000009	0.000009	5.99	0.022*
C*D	0.000042	0.000042	27.04	0.000**
D*F	0.000009	0.000009	5.65	0.026*

\*\*p < 0.001 , \* p < 0.05.

<Table 8> Result of Hyperparameter

Model	Hyperparameter	Value
SVM	Kernel	rbf
	Cost	3000
	Gamma	0.1
RF	n_estimators	-
	max_features	1
	max_depth	100
	min_samples_leaf	25
	min_samples_split	-
	bootstrap	False
AdaBoost	n_estimators	500
	base_estimator	5
	learning_rate	-
	algorithm	SAMME.R

<Table 10>의 다섯 번째 열에 표시된 것과 같이 반응표면분석을 통한 최적화 역시 분류모형의 정확도 향상에는 한계를 보여주었다. 이는 아날로그 신호는 주기(주파수) 단위의 패턴을 근거로 정상/비정상 등 특성을 판단해야 한다. 따라서 신호에 대한 일반적인 샘플링을 통해서만 패턴

을 찾기가 어렵고, 주기를 고려한 샘플링 및 기술통계치와 같은 분포 정보를 통해서 아날로그 신호의 본래 특성을 파악할 수 있다. 따라서, 본 연구에서는 아날로그 데이터의 패턴 정보를 찾기 위해 신호에 대한 샘플링 데이터를 이용한 기술통계치를 도출하고, 이를 분류모형에 적용하였다.

#### 4. 기술통계치를 활용한 분류모형 개발

수집한 샘플링 데이터로 <Table 9>와 같이 평균(mean), 분산(variance), 표준편차(standard deviation), 왜도(skewness), 첨도(kurtosis), 백분위수(Percentiles) 등의 기술통계치를 도출하였다. 데이터의 전처리만 앞서 진행한 방법과 동일하게 적용하였으며, 기술통계치를 특성치로 사용하여 분류모형을 학습시켰다. 기술통계치를 이용한 분류모형의 정확도는 <Table 10>의 마지막 열에 표시된 것과 같이 99% 이상의 만족스러운 성능을 보였다.

<Table 10>은 각 분류모형에 대해서 검증 데이터(test data)를 사용한 Raw 데이터, 그리드 탐색, 랜덤 탐색, 베이직한 최적화, 반응표면분석, 기술통계치를 적용하였을 때

<Table 9> Descriptive Statistics of Analog Data

a1_mean	a1_var	a1_std	a1_skew	a1_kurt	a1_25_percent	a2_std	a2_skew	a2_kurt	a3_std	a3_skew	Y
2.2034	2.0761	1.9240	0.2369	-0.7052	1.9895	0.9017	-0.8053	-0.4964	0.6548	-1.8949	0
2.4558	3.7339	3.1138	2.0987	2.3238	1.6290	0.8055	-1.3157	0.2875	0.6329	-0.1517	0
2.3462	0.0153	0.1053	-0.9689	-0.5629	2.8527	0.2137	-0.5979	-0.9413	2.3199	-0.9748	0
3.7126	0.8750	0.9266	-2.0888	-0.6426	4.5686	0.5929	-0.9678	0.4732	1.6048	0.6373	0
3.6831	3.9959	3.2871	-0.6445	-0.7265	2.1673	0.4514	-1.3330	-0.7527	0.6927	-0.2620	0
2.2649	1.2685	1.2695	-0.9827	-0.8327	1.8891	1.4531	0.1026	-1.0650	1.4398	-0.0577	0
1.6012	3.1284	2.6992	1.1450	0.3227	-0.0452	0.2594	-1.6282	-0.7160	0.0501	-1.5522	0
0.2732	-0.2752	-0.2006	-1.2585	-1.0860	-0.4165	-0.0009	-1.1829	-1.0072	2.1392	1.5146	0
1.4752	0.6923	0.7610	0.0430	-0.3540	1.6567	-0.3604	-1.8984	-0.6939	-0.1881	0.3821	0
1.6318	-0.7928	-0.7946	0.8080	0.3917	1.7340	1.8819	-0.0170	0.3659	0.6814	-0.5463	0
1.1106	1.0245	1.0590	0.7767	0.8956	0.2183	0.0672	-1.2457	-0.0487	0.9982	0.1759	0
0.7451	1.5834	1.5319	0.5583	0.3983	-0.3143	0.1266	-0.7035	-0.9035	-0.2899	-1.0883	0
0.1398	0.4013	0.4879	-1.0976	-0.9181	0.0160	0.2610	-1.9656	-1.1351	-0.6573	-1.5146	0
2.5284	1.5221	1.4816	-0.8212	-0.8051	2.3004	1.1653	0.7081	0.4417	-0.3050	0.2304	0
-0.6687	-0.5008	-0.4510	0.7057	0.4267	-0.2178	0.7120	1.1710	0.2715	-0.6987	-0.6883	0
1.3670	1.0767	1.1046	0.1821	-0.5310	0.2761	1.2300	0.9358	-0.3864	0.2049	-0.3617	0
-0.0026	-0.6353	-0.6064	-0.7652	-0.6786	-0.2829	1.0461	0.0708	-0.7378	0.6631	-0.4505	0
1.4707	2.0928	1.9370	-0.6146	-0.6496	0.3034	1.1571	0.0431	-0.3998	0.3744	-0.1895	0
1.7196	0.1510	0.2428	-1.5986	-1.2509	0.6970	-1.1109	-0.7213	0.4611	-0.7271	0.2602	0
0.9817	0.9761	1.0164	-0.3589	-1.1626	0.2370	1.0945	-0.9360	-0.5050	-1.2976	0.0910	0
1.6107	1.5364	1.4934	1.0982	1.1758	0.2594	1.7206	1.2045	0.9920	0.3057	0.0539	0
-0.9432	-0.4820	-0.4297	2.2093	2.5980	-0.5289	0.7924	-1.9143	-1.3090	0.0741	0.7704	0
0.6534	0.2142	0.3057	-0.4008	-0.5450	0.3443	0.1911	0.1292	-0.1524	-0.0370	0.6201	0

&lt;Table 10&gt; Classification Model Accuracy

Model	Raw data	Hyperparameter Optimization				Descriptive Statistics
		Grid Search	Random Search	Bayesian Optimization	RSM	
SVM	93.86	93.94	93.90	93.82	93.90	99.75
RF	93.55	93.88	93.89	94.19	93.85	99.86
AdaBoost	93.57	93.88	93.85	93.70	93.92	99.71

의 정확도를 비교한 결과를 보여주고 있다. 아날로그 신호 데이터를 이용한 모형과 이 모형에 대한 하이퍼파라미터 최적화를 수행한 모형의 경우 요구되는 95%를 달성할 수 없었지만, 기술통계치를 이용한 분류모형의 경우 분류 정확도를 99% 이상 달성할 수 있어 현장에 적용할 수 있다고 판단할 수 있다.

## 5. 결론

본 연구에서는 주기를 갖는 아날로그 신호 데이터를 이용한 이상상황 탐지 기계학습 알고리즘에서 성능향상을 위한 방법론들을 제시하였다. 아날로그 신호에 대해서는 <Table 10>에서 알 수 있듯이 그리드 탐색, 랜덤 탐색, 베이지안 최적화, 반응표면분석을 활용한 하이퍼파라미터 최적화의 경우 성능향상에 한계가 있음을 알 수 있었다. 이는 아날로그 신호에 대한 샘플링 데이터는 신호에 숨겨진 패턴을 반영하지 못하고 있기 때문이라고 판단된다. 아날로그 신호 데이터의 경우 신호에 숨겨진 패턴을 반영할 수 있는 기술통계치를 이용한 분류모형의 개발이 대안이 될 수 있음을 알 수 있다.

본 연구에서는 시료 모터 체결상태의 불량률 통틀어 이상 상황 판단에 초점을 맞추었다. 추후 연구로는 첫째, 모터 시험 중에 시료 모터의 베어링 파손 등과 같은 기계적인 결함이 발생할 수 있으므로 이러한 상황에 대한 추가적인 데이터 수집을 통해 분류모형에 반영할 필요가 있다. 둘째, 기술통계치를 사용하는 경우 과적합 문제가 발생할 수 있지만, 앙상블 학습을 사용할 때는 각 베이스 모형들을 학습한 후 대푯값을 취하므로, 과적합의 가능성을 줄일 수 있다고 알려져 있다. 따라서 과적합의 가능성은 낮은 것으로 예상되지만 이에 대한 추가 검증이 필요하다.

## Acknowledgement

This research was supported by Changwon National University in 2023~2024.

## References

- [1] Choi, Y.U., Yoon, D.U., Choi, J.H., and Byun, J.M., Hyperparameter Search for Facies Classification with Bayesian Optimization, *Geophysics and Geophysical Exploration*, 2020, Vol. 23, No. 3, pp. 157-167.
- [2] Gustavo, A., Lujan-Moreno, P.R., Howard, O.G., and Rojas, D.C., Montgomery, Design of experiments and response surface methodology to tune machine learning hyperparameters, with a random forest case-study, *Expert Systems with Applications*, 2018, Vol. 109, pp. 195-205.
- [3] Han, J.H., Choi, D.J., Hong, S.K., and Kim, H.S., Motor Fault Diagnosis Using CNN Based Deep Learning Algorithm Considering Motor Rotating Speed, *2019 IEEE 6th International Conference on Industrial Engineering and Applications (ICIEA)*, Tokyo, Japan, 2019, pp. 440-445.
- [4] Hwang, H.E., Cho, Y.S., Hwang, S.C., and Kim, S.B., Optimal Tire Design Using Machine Learning and Bayesian Optimization, *Journal of the Korean Institute of Industrial Engineers*, 2022, Vol. 48, No. 4, pp. 433-440.
- [5] Hwang, H.S., Hybrid Machine Learning Model for Predicting the Direction of KOSPI Securities, *Journal of the Korea Convergence Society*, 2021, Vol. 12, No. 6, pp. 9-16.
- [6] Iqbal, S., Qureshi, A.N., Khursheed, K., Alhussain, M., Haider, S.I., and Rida, I., AMIAC: adaptive medical image analyzes and classification, a robust self-learning framework, *Neural Comput & Applic*, 2023. <https://doi.org/10.1007/s00521-023-09209-1>.
- [7] Jung, H. and Kim, J.W.. A Machine Learning Approach for Mechanical Motor Fault Diagnosis, *Journal of Society of Korea Industrial and Systems Engineering*, Mar. 31, 2017.
- [8] Kim, S.W. and Lee, S.M., Fault Detection Technique for PVDF Sensor Based on Support Vector Machine,



- The Journal of the Korea Institute of Electronic Communication Sciences*, 2023, Vol. 18, No. 5, pp. 785-796.
- [9] Kwon, K.B., Choi, H.S., Oh, J.Y., and Kim, D.K., A Study on EPB Shield TBM Face Pressure Prediction Using Machine Learning Algorithms, *Journal of Korean Tunnelling and Underground Space Association*, 2022, Vol. 24, No. 2, pp. 217-230.
- [10] Lee, J.E., Kim, Y.B., and Kim, J.N., Hyperparameter Optimization for Image Classification in Convolutional Neural Network, *Journal of the Institute of Convergence Signal Processing*, 2020, Vol. 21, No. 3, pp. 148-153.
- [11] Lee, S.H., Ko, S.K., and Lee, S.A., Fault Classification Model Based on Deep Learning Using Vibration Data of Mechanical Equipment, *The Journal of Korean Institute of Next Generation Computing*, 2022, Vol. 18, No. 2, pp. 36-46.
- [12] Lee, T.C., Cho, H.J., and Choi, J.S., A Study on Single Motor failure Diagnosis and Fault Tolerant Control by Using Open Source and Machine Learning, *Journal of The Korean Society for Aeronautical & Space Sciences*, Nov. 2022, pp. 810-812.
- [13] Lee, Y.K., Hong, J.K., and Hong, S.C., A study on the anomaly prediction system of drone using big data, *Journal of Internet Computing and Services (JICS)*, Vol. 21, No. 2, Apr. 2020, pp. 27-37.
- [14] Lee, Y.S., Yoo, J.H., and Kim, B.H., Multi-grid Application and Optimal Hyper-parameter Selection in AI Urban Flood Prediction, *Korean Society of Civil Engineers*, 2022, Vol. 2022, No. 10, pp. 301-302.
- [15] Park, M.J., Lee, Y.S., and Kim, J.Y., Comparing Accuracy of Predicting PM2.5 Concentration according to Hyperparameter Optimization Methods of Long Short-Term Memory (LSTM) Model, *Proceeding of Korean Society for Atmospheric Environment*, 2021, Vol. 2021 No. 10, pp. 371-371.
- [16] Snoek, J., Larochelle, H., and Adams, R.P., Practical bayesian optimization of machinelearning algorithms, *Proceeding of NeuralInformation Processing Systems*, 2012, pp. 2951-2959.
- [17] Vermesan, O., Coppola, M., Bahr, R., Bellmann, R.O., Martinsen, J.E., Kristoffersen, A., Hjertaker, T., Breiland, J., Andersen, K., Sand, H.E., and Lindberg, D., An Intelligent Real-Time Edge Processing Maintenance System for Industrial Manufacturing, Control, and Diagnostic, *Frontiers in Chemical Engineering*, 2022.
- [18] Vulpe-Grigorasi, A. and Grigore, O., Convolutional Neural Network Hyperparameters optimization for Facial Emotion Recognition, 2021, *12th International Symposium on Advanced Topics in Electrical Engineering (ATEE)*, Bucharest, Romania, pp. 1-5.
- [19] Won, J.J., Shin, J.M., Kim, J.H., and Lee, J.W., A Survey on Hyperparameter Optimization in Machine Learning, *The Journal of Korean Institute of Communications and Information Sciences*, 2023, Vol. 48, No. 6, pp. 733-747.
- [20] Yang, L. and Abdallah, S., On hyperparameter optimization of machine learning algorithms: Theory and practice, *Neurocomputing*, 2020, Vol. 415, pp. 295-316.
- [21] You, H.J. and Kim, C.H., Machine Learning Modeling and Hyper-Parameter Optimization for Weld Nugget Formation and Failure Behavior of Resistance Spot Welds, *Journal of Welding and Joining*, 2021, Vol. 39, No. 6, pp. 658-665.
- [22] Yuk, D.G. and Sohn, J.W., Motor Anomaly Detection Using Deep Learning, *Korean Society of Mechanical Engineers*, Nov. 2022, pp. 421-422.
- [23] Yu, S.H., *A Study on the Method of Diagnosing Failure and Detecting Failure Signals of the Motor and Inverter of Electric Vehicle Using Machine Learning*, Graduate School of Soonchunhyang University, Master's thesis, 2023.

#### ORCID

Jaehun Kim | <https://orcid.org/0009-0009-3280-0444>

Sangcheon Eom | <https://orcid.org/0009-0002-9873-6202>

Chulsoon Park | <https://orcid.org/0000-0003-3147-7613>