

# 불균형 데이터를 갖는 냉동 컨테이너 고장 판별 및 원인 분석을 위한 기계학습 모형 개발

이희원<sup>1</sup>, 박성호<sup>1</sup>, 이승현<sup>1</sup>, 이승재<sup>1</sup>, 이강배<sup>2\*</sup>  
<sup>1</sup>동아대학교 경영정보학과 학생, <sup>2</sup>동아대학교 경영정보학과 교수

## Development of machine learning model for reefer container failure determination and cause analysis with unbalanced data

Huiwon Lee<sup>1</sup>, Sungho Park<sup>1</sup>, Seunghyun Lee<sup>1</sup>, Seungjae Lee<sup>1</sup>, Kangbae Lee<sup>2\*</sup>

<sup>1</sup>Student, Dept. of MIS, Donga University

<sup>2</sup>Professor, Dept. of MIS, Donga University

**요약** 냉동 컨테이너의 고장은 큰 비용의 손실을 야기하지만, 현재 냉동 컨테이너의 알람 체계는 효율성이 떨어진다. 기존에 냉동 시스템의 시뮬레이션 데이터를 활용한 연구는 존재하지만, 냉동 컨테이너의 실제 운영 데이터를 활용한 연구는 부족하다. 이에 본 연구는 실제 냉동 컨테이너 운영 데이터를 활용하여 고장 원인을 분류하였다. 실제 데이터에서는 데이터 불균형이 발생하였으며 ENN-SMOTE, 클래스 가중치를 둔 Logistic 회귀분석과 본 연구에서 개발한 2-stage 알고리즘을 비교하여 데이터 불균형문제를 해결하였다. 2-stage 알고리즘은 XGboost, LGBost, DNN을 사용하여 첫 번째 단계에서는 고장 및 정상을 분류하고, 두 번째 단계에서는 고장의 원인을 분류하는 알고리즘이다. 2-stage 알고리즘에서 LGBost를 사용한 모델이 99.16%의 정확도로 가장 우수하였다. 본 연구는 데이터 불균형을 해결하기 위해 2-stage 알고리즘을 활용한 최종모형을 제안하며 이는 다른 산업에도 활용할 수 있을 것으로 사료된다.

**주제어** : 냉동 컨테이너, 머신러닝, 딥러닝, 고장진단, 고장 원인 분류, 다중 클래스, 고장 분류

**Abstract** The failure of the reefer container causes a great loss of cost, but the current reefer container alarm system is inefficient. Existing studies using simulation data of refrigeration systems exist, but studies using actual operation data of refrigeration containers are lacking. Therefore, this study classified the causes of failure using actual refrigerated container operation data. Data imbalance occurred in the actual data, and the data imbalance problem was solved by comparing the logistic regression analysis with ENN-SMOTE and class weight with the 2-stage algorithm developed in this study. The 2-stage algorithm uses XGboost, LGBost, and DNN to classify faults and normalities in the first step, and to classify the causes of faults in the second step. The model using LGBost in the 2-stage algorithm was the best with 99.16% accuracy. This study proposes a final model using a two-stage algorithm to solve data imbalance, which is thought to be applicable to other industries.

**Key Words** : Reefer Container, Machine Learning, DNN, Fault Diagnosis, Cause analysis, Multi-class, Fault type classification

\*This work was partly supported by Institute of Information & communications Technology Planning & Evaluation (IITP) grant funded by the Korea government(MSIT) (No.2020-0-02091, Development and Commercialization of IoT-based refrigerated container real-time monitoring and BigData / AI-based failure predictive service platform to strengthen competitiveness of shipping & logistics company).

\*Corresponding Author : Kangbae Lee(kanglee@donga.ac.kr)

Received October 16, 2021

Revised November 15, 2021

Accepted January 20, 2022

Published January 28, 2022

## 1. 서론

냉동 컨테이너는 신선 과일 및 채소, 냉동식품, 유제품, 화학제품 및 의약품, 냉동 농축물, 반도체 등을 운반하는 동안 변질을 멈추거나 늦추는 역할을 한다[1]. 세계 무역의 약 90% 가량이 컨테이너를 활용하여 이루어지고 있으며, 냉동 컨테이너의 물동량 또한 2011년 이후로 꾸준히 증가하고 있는 양상을 보이고 있다[2,3]. 이와 같은 상황이 지속됨에 따라 냉동 컨테이너의 고장은 큰 비용의 손실을 야기한다.

현재 냉동 컨테이너 시스템은 고장이 발생했을 경우, 경고 알람이 울리는 구조로 이루어져 있다. 현업 업무 종사자에 따르면 알람의 범위가 넓고, 필요하지 않은 알람이 다수 발생하여 불편함을 겪고 있다는 것을 확인하였다. 주요 알람이 조기에 감지되지 않고, 방치되는 경우 유지관리에 많은 비용이 발생한다[4]. 운영 사항을 고려한 고장원인 진단 방법을 사용하면 유지 관리와 수리 및 운송 비용 절감하고, 운영의 효율성을 향상시킬 수 있다[5].

기존의 연구에서 냉동 컨테이너의 온도 상승 원인을 조사하기 위해 시뮬레이션을 통해 온도 변화를 모니터링하고, 그에 영향을 주는 변수들을 확인하였다[6]. 또한 냉동 컨테이너와 유사한 냉동 시스템, 콜드 체인을 분석하기 위해 시뮬레이션 데이터 혹은 실제 데이터와 차이가 있는 실험데이터들이 사용되고 있음을 확인하였다[7,8]. 특히 냉동 시스템과 관련된 연구들이 RP-1043 프로젝트로 도출된 시뮬레이션 데이터를 사용하여 고장 원인 진단에 대해 활발하게 연구되고 있다[9-12].

시뮬레이션 데이터, 실험에 의해 도출된 데이터는 고장 진단에 대한 다양한 문제를 다룰 수 있다. 그러나 결측치가 존재하지 않는 균일한 데이터로 실제 운영 상황을 반영하지 못한다는 한계가 있다. 실제 데이터로 분석한 연구의 경우 데이터가 불충분하여 다양한 변수를 사용하지 못해 온도 제어에만 연구 방향이 국한되어 있었다.

이에 본 연구는 실제 냉동 컨테이너 운영 데이터를 사용하여 분석을 진행하였다. 의도적으로 만들어진 데이터와 달리 실제 컨테이너 운영 데이터에는 데이터 결측치 문제와 불균형 문제가 존재하였다. 결측치 문제를 해결하기 위해 특성에 따라 다른 보간법을 사용하였고, 불균형 문제를 해결하기 위해 첫 번째 단계에서 고장 및 정상을 판단한 후, 두 번째 단계에서 고장 원인을 진

단하는 2-stage 알고리즘을 개발하였다.

이후 2장에서 분석을 위한 이론적 배경에 대해 설명하고, 3장에서는 데이터와 처리 방법 및 분석 방법에 대해 설명한다. 4장에서는 분석 결과를 확인하고, 5장에서는 연구의 최종 결론을 제시한다.

## 2. 이론적 배경

### 2.1 기계 학습

기계 학습은 컴퓨터 알고리즘을 연구하는 것으로 인공지능의 한 분야로 1949년에 Hebb이 Hebbian Learning 이론을 발표하는 것을 기점으로 시작되었다. 기계 학습은 경험을 통해 컴퓨터가 자동으로 알고리즘을 개선한다는 특징을 가지고 있으며 다양한 분야에서 연구되고 있다[13]. 본 연구에서는 기계 학습 중 지도 학습을 사용한 Tree 기반 학습 알고리즘인 XGBoost (Extreme Gradient Boosting; XGBoost)와 LGBost(Light Gradient Boosting; LGBost)과 더불어 최근 각광받고 있는 알고리즘인 딥러닝(Deep Neural Net; DNN)을 사용하였다[14,15]. XGBoost는 기존 부스팅 기법들과 비교하였을 때 학습과 분류가 빠르다는 장점이 있고, LGBost은 속도가 빨라 큰 사이즈의 데이터를 다룰 수 있고, 정확도 또한 우수하다는 장점이 있다. DNN은 다양한 선형조합을 통해 비선형 문제로 변환하는 알고리즘으로 2016년 구글의 알파고(AlphaGo)와의 바둑 대결 이후로 다시 한번 활발한 연구가 진행되고 있다. DNN은 특성을 직접 만들지 않아도 스스로 경우의 수를 줄이며 특성을 만든다는 장점이 존재한다[16]. 본 연구는 상황에 따라 2-stage 모델에서 3가지 알고리즘을 선택적으로 사용하였으며, 성능을 비교하여 최종 모델을 선정하였다.

### 2.2 선행연구

냉동 컨테이너에 관한 선행연구가 부족하여 냉동 컨테이너와 비슷한 메커니즘을 가지고 있는 RP-1043 프로젝트의 냉동 시스템 시뮬레이션 데이터를 사용한 선행연구들을 참고하였으며, 해당 프로젝트에서 고장의 유형은 냉매 누출, 응축기 고장, 냉매 과충전, 응축기 유량 감소, 응축기 냉각수 감소, 증발기 냉수 감소, 오일 과충전으로 총 7개의 유형으로 분류되어 있다.

Yalan Wang은 실험 데이터를 사용하여 엔트로피에 기반의 베이지안 네트워크 방법을 사용한 고장 유형 진단 모델을 제안하였으며, 3 심각도 이상에서 고장 유형의 평균 정확도가 93%인 것을 확인하였지만, 1, 2 심각도에서는 90%가 되지 못하는 경우가 존재하였다[9].

G.N Li는 실험 데이터를 사용하여 PCA 기반의 PCA-R-SVDD 방법을 사용한 고장 유형 진단 모델을 제안하였으며 냉매 누출을 제외한 6가지 결함에 대해 약 92%로 분류한 것을 확인하였다[10].

Li는 실험 데이터를 사용하여 HVAC 시스템 오류를 진단하였다. 폴링 계층을 제외하고 컨볼루션 필터 커널의 크기를 1로 설정하였으며 Grad-Absolute-CAM 가중치를 적용한 DCNN을 방법을 제안하였다. 전체 고장에 대한 정확도는 98.6%를 도출하였지만 고장의 원인에 대해 판별하지 못하였다[11].

G.N Li는 DNN과 PCA를 여러 방법으로 결합하여 하이퍼파라미터를 도출하고 냉동 시스템의 고장 진단을 위한 분석을 수행한 결과 7개의 레이어를 사용한 DNN과 PCA를 결합한 모델이 가장 우수한 것을 확인하였으며 대체적으로 90% 이상의 정확도를 도출하였지만, r1에서는 정확도가 90%를 넘지 못하였다[12].

본 연구는 냉동 컨테이너에 관한 연구로 선행 연구에서 제시된 정제된 냉동 시스템의 시뮬레이션 데이터와 비교하였을 때 결측치 문제, 데이터 불균형의 문제 등 차이가 존재한다. 이에 따라 현업 업무 종사자와의 협업을 통해 중요한 고장 유형을 선정하였으며, 그 결과 냉매, 파워, 모터, 센서, 기타 총 5가지로 고장 유형을 분류하였다.

### 3. 데이터 및 연구 모델

#### 3.1 데이터 설명

본 연구는 A 해운사에서 제공받은 Starcool 사의 냉동 컨테이너 실제 운영 데이터를 사용하였다. 분석에는 총 111개의 Starcool 사의 냉동 컨테이너를 사용하였고, 수집된 데이터의 특성은 Table 1과 같다.

Table 1. Data Feature

no.	Variable	Full name	Unit
1	Time	-	-
2	container	container number	-

3	Type	container data Type	-
4	RH_set	relative Humidity	%
5	Tsup	Supply Air Temperature	℃
...	...	...	...
60	Mair Avg	Motor air Average Status	%
61	Mpump On Avg	motor pump on Average	%
62	Battery pack voltage	Battery pack voltage	V
63	Moumo service runtime	Motor pump sevice runtime	Hour

수집된 데이터는 총 63개의 특성으로 구성되어 있으며, 데이터 수집 주기는 1시간이다. 알람이 발생했을 경우 알람이 발생한 시간에 따라 데이터가 생성된다. 이 때, 알람에 대한 정보를 제외한 부분은 다른 특성 값들이 기록되지 않고 모두 결측치로 처리된다. 데이터는 총 299,576개가 존재하며, 그 중 고장에 대한 데이터는 총 31,919개가 존재한다.

#### 3.2 데이터 전처리

3.1의 설명과 같이 주기적으로 기록되는 기록 데이터 및 알람 발생 시에만 기록되는 알람 데이터에는 결측치가 존재한다. 본 연구에서는 해당 결측치들을 Table 2와 같은 방법으로 보간하였다.

Table 2. Variable interpolation method

Variable Type	Method
Setting values	Using the preceding value
Numerical variables	linear interpolation
Categorical variables	Create a new category

설정 값과 관련된 변수들은 컨테이너의 온도가 정해지면 설정 값이 변하지 않기 때문에 앞의 값을 사용하여 보간하였고, 수치형 변수들은 선형 보간법을 사용하여 보간하였다. 선형 보간법은 Fig. 1과 같이 두 끝점이 주어졌을 경우 그 사잇값을 추정하기 위해 두 점 사이에 직선을 긋고 비례식에 따라 결측치를 보간하는 방법이다. 데이터와 결측치 사이의 시간 차가 존재하고, 시간 차 동안 값들이 점점 증가 혹은 감소할 수 있기 때문에 해당 방법을 사용하였다. 또한 범주형 변수들은 결측치도

의미가 있을 수 있어 새로운 카테고리를 만들어 레이블 인코딩을 진행하였다.

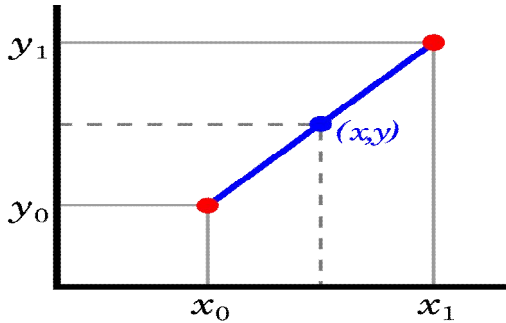


Fig. 1. Linear interpolation

또한 수치형 변수들의 범위와 크기가 다르다는 문제를 해결하기 위해 표준화를 진행하였다.

3.3 불균형 데이터 처리

본 연구에 사용된 정상 데이터의 개수는 267,657개 이고, 고장 데이터의 개수는 31,919개 이다. 각 고장 원인에 대한 데이터의 개수는 Table 3과 같다.

Table 3. Fault Diagnosis number

Fault Diagnosis	Number
Refrigerant(1)	161
Power(2)	9,868
Sensor(3)	20,576
Motor(4)	247
Ect(5)	1,067

고장에 대한 데이터의 수는 분석을 하기에 충분한 반면, 원인별 데이터 수는 데이터 수가 적어 불균형하다. 전체 데이터의 1%가 되지 않는 Motor, Refrigerant와 같은 고장 원인의 경우 분류하기 어려움이 있다. 기존에 존재하였던 Resampling을 활용한 방법, 로지스틱 회귀분석의 클래스 가중치를 이용한 방법, 본 연구에서 제안하는 2-stage를 이용한 방법을 비교하여 불균형 문제를 해결하였다. Resampling에는 과소 표집과 과대 표집 두 가지 방법이 있다. 과소 표집은 불균형한 데이터 셋에서 비율이 높은 클래스의 데이터를 줄이는 방법으로 전체 데이터의 수가 감소한다는 단점이 존재하고, 과대 표집은 비율이 작은 클래스의 데이터의 수를

늘리는 방법으로 과적합을 초래할 수 있다는 단점이 존재한다. ENN-SMOTE는 과소 표집과 과대 표집을 조합한 방법으로 다른 조합 방법에 비해 성능이 가장 우수하였다[17].

본 연구는 ENN-SMOTE를 활용한 Resampling, 클래스 가중치를 활용한 로지스틱 회귀분석, 본 연구에서 제안하는 2-stage 알고리즘 3가지 방법론을 비교하여 가장 성능이 우수한 방법론을 최종 모델로 선정하였다.

3.4 최종 프로세스

본 연구는 데이터 불균형 문제를 해결하기 위해 3가지 방법론을 사용하였으며, 가장 성능이 우수한 2-stage 알고리즘을 최종 모델로 선정하였다. 2-stage 알고리즘은 개별 고장 원인에 대한 데이터가 불균형한 문제를 해결하기 위해 고안된 알고리즘으로 Fig. 2와 같이 첫 번째 단계에서 정상/고장을 판단한 후, 두 번째 단계에서 원인을 분류하는 알고리즘이다.

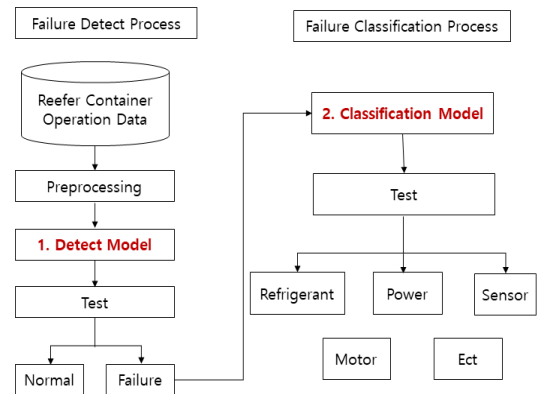


Fig. 2. 2-Stage Fault Diagnosis Process

냉동 컨테이너의 운영 데이터는 3.2에서 설명한 보간법과 표준화를 이용하여 전처리를 진행한 뒤, Train과 Test를 8:2 비율로 나누는 데이터가 Detect 모델의 Input 값으로 사용된다.

첫 번째 모델인 Detect 모델은 정상과 고장을 구분하는 모델이다. Detect 모델은 전체 데이터를 사용하여 때문에 처리 속도가 느린 DNN을 사용하지 않고, XGBoost와 LGBost를 사용하여 비교하였다.

첫 번째 모델에서 Test를 수행한 결과 고장으로 식별된 데이터들은 두 번째 모델인 Classification 모델

의 Input 값으로 사용된다. Classification 모델은 고장 원인을 냉매, 파워, 센서, 모터, 기타 고장으로 분류하는 모델이다. 두 번째 모델은 고장으로 식별된 데이터들만 input 데이터로 들어가기 때문에 데이터의 수가 적어 DNN과 XGBoost, LGBost를 모두 사용하여 그 결과를 비교하였다.

### 3.5 성능지표

본 연구는 정상/고장이 불균형한 데이터를 사용하여 분석을 진행하였다. 불균형한 데이터 사용 시, Accuracy만을 사용하여 성능을 도출할 경우 데이터가 많은 쪽으로 결과가 치우쳐 편향된 결과를 야기할 수 있다. 따라서 본 연구는 불균형 데이터에서 성능을 확인할 때 사용되는 Precision, Recall, F1-score를 추가적인 성능지표로 사용하였다. Fig. 3은 해당 성능지표들을 설명하기 위한 Confusion Matrix를 나타낸 그림이고, Table 4는 성능 지표에 대한 수식을 나타낸 표이다.

		Predict values	
		Positive	Negative
Actual values	Positive	TP	FN
	Negative	FP	TN

Fig. 3. Confusion matrix

Table 4. Performance indicators

indicator	formula
accuracy	$(TP+TN)/(TP+TN+FP+FN)$
precision	$TP/(TP+FP)$
recall	$TP/(TP+FN)$
F1-score	$2*Precision*Recall / Precision + Recall$

Precision은 예측한 값이 맞은 것에 대한 확률을 나타내고, Recall은 실제 값을 정확히 예측할 확률을 나타낸다. 그리고 F1-score는 Precision, Recall의 조화평균을 의미하고, Accuracy는 전체 모델의 정확도를 나타낸다. 본 연구에서 4가지 성능 지표를 모두 사용하였으며, 최종 결과에서 전체 모델의 정확도를 나타내는 Accuracy와 Precision, Recall을 모두 설명할 수 있는 F1-score를 성능지표로 사용하였다.

## 4. 분석 결과

### 4.1 1단계 : Detect Model

XGBoost와 LGBost에 초모수(Hyper Parameter) 조정을 통해 가장 우수한 모델을 도출한 결과 두 가지 방법 모두 우수한 성능을 보여주었다. Table 5는 두 모델의 성능을 나타낸 표이다.

Table 5. Detect Model performance indicators

indicator	Detect	XGB	LGBM
Precision	normal	0.9960	0.9966
	fault	0.9729	0.9689
Recall	normal	0.9968	0.9963
	fault	0.9666	0.9715
F1-Score	normal	0.9964	0.9964
	fault	0.9697	0.9702
Accuracy	-	0.99357	0.99364

XGBoost와 LGBost의 성능 지표 확인결과 Precision, Recall 성능지표에서는 두 모델이 비슷하였지만, 2가지 성능지표를 조합한 고장에 대한 F1-Score에서 0.0005만큼의 차이로 LGBost가 우수하였으며 Accuracy에서도 LGBost가 0.0001의 차이로 우수하였다. 또한 속도 측면에서도 LGBost가 약 4분가량 더 우수하였다. 따라서 첫 번째 단계인 Detect Model에서 LGBost 방법론을 채택하였다.

### 4.2 2단계 : Classification Model

Classification Model 단계는 고장으로 분류된 데이터의 원인을 분류하는 단계로 고장 데이터만 사용하여 원인을 학습시켰다. Detect Model과 다르게 데이터 불균형이 존재하며 클래스도 5개로 분류되어 미세한 차이로도 원인이 다르게 측정될 수 있다. 첫 번째 모델에 사용하였던 XGBoost, LGBost 방법론에 DNN 방법론을 추가적으로 분석하였으며 세 모델 모두 초모수(Hyper parameter) 조정을 통해 가장 우수한 모델을 도출하였다. Table 6은 세 가지 방법론의 성능을 비교한 표이다.

전체적인 성능을 확인한 결과 LGBost가 XGboost와 DNN에 비해 Precision, Recall, F1-Score, Accuracy 우수한 모습을 보여주었으며, 속도 측면에서 약 30초 가량 우수하였다. 따라서 두 번째 모델의 방법론으로도 첫 번째 모델과 동일하게 LGBost를 채택하였다.

**Table 6. Classification Model performance indicators**

indicator	Classification	XGB	LGBM	DNN
Precision	Refrigerant	0.84	0.86	0.78
	power	0.99	0.99	0.98
	sensor	1.00	1.00	1.00
	motor	0.88	0.86	0.82
	ect	0.86	0.87	0.83
	Recall	Refrigerant	0.84	0.97
power		0.99	0.99	0.99
sensor		1.00	1.00	0.99
motor		0.73	0.78	0.82
ect		0.87	0.88	0.79
F1-Score	Refrigerant	0.84	0.91	0.78
	power	0.99	0.99	0.98
	sensor	1.00	1.00	0.99
	motor	0.80	0.82	0.82
	ect	0.87	0.88	0.81
Accuracy	-	0.98857	0.98951	0.98261

**4.3 최종 결과**

여러 가지 방법론을 사용하여 성능을 비교해본 결과 LGBost가 가장 우수한 모습을 보여주었다. 본 연구에서는 LGBost 방법론을 이용한 2-stage 모델을 최종적으로 제안한다. ENN-SMOTE에 XGBoost, LGBost, DNN을 사용한 결과 XGBoost를 사용한 모델이 가장 우수하였으며, Logistic Regression의 경우 데이터에 맞게 클래스 가중치 및 하이퍼파라미터를 조정하였다. 2-stage 방법 사용하여 분석을 진행한 결과 ENN\_SMOTE, 클래스 가중치를 활용한 One-Stage 방법들과 비교하였을 때, 모든 유형의 고장에서 Precision, Recall, F1-Score, Accuracy 모두 우수한 성능을 보여주었다. Table 7은 세 방법의 차이에 대한 성능을 비교한 표로 F1-Score 와 Accuracy를 사용하여 최종 결과를 정리하였다.

**Table 7. Final Result**

Classification	ENN-SMOTE	Class weight Logistic Regression	2-Stage Model
normal	F1-score : 0.99	F1-score : 1.00	F1-score : 1.00
Refrigerant	F1-score : 0.53	F1-score : 0.59	F1-score : 0.84
power	F1-score : 0.93	F1-score : 0.90	F1-score : 0.96
sensor	F1-score : 0.99	F1-score : 0.99	F1-score : 0.99
motor	F1-score : 0.62	F1-score : 0.40	F1-score : 0.62
ect	F1-score : 0.56	F1-score : 0.31	F1-score : 0.67
	Accuracy : 0.9897	Accuracy : 0.9886	Accuracy : 0.9916

**5. 결론**

본 연구는 Starcool 사의 실제 냉동컨테이너 운영 데이터를 사용하여 고장을 진단하고, 고장 원인에 대해 분석하였다. 냉동 컨테이너의 고장 발생 시, 신속하게 원인을 알아내고 수리하는 것은 유지 관리 및 수리, 운송 비용 절감에 큰 도움이 된다. 이에 높은 정확도로 조기에 필요한 고장을 진단하고, 해당 원인을 파악하기 위한 2-stage 알고리즘을 제시하였다. 2-stage 알고리즘은 ENN-SMOTE와 클래스 가중치를 활용한 단일 방법에 비해 높은 성능지표로 데이터 불균형 문제를 해결할 수 있고, 그에 따라 더 높은 정확도를 도출하였다. 여러 가지 모델을 사용하여 초모수(Hyper parameter)를 조정된 후, 성능을 비교하여 고장 진단 및 원인 분류를 하기 위한 최적의 모델을 도출하였다. 최종적으로 LGBost의 성능이 가장 우수하였으며, LGBost를 활용한 2-stage 모델을 제안한다. 제안된 모델은 기존에 성능이 우수했던 고장 원인들에 대한 성능은 유지하면서 성능이 떨어진 고장 원인들에 대한 성능은 향상시킬 수 있었다. 특히, 기존에 낮은 성능을 가졌던 냉매 (Refrigerant)는 ENN-SMOTE, 로지스틱 회귀분석과 비교했을 때 각각 29%, 25%가량 향상하였으며 주요 고장을 제외한 기타 고장들에 대한 분류도 각각 11%, 36% 가량 향상하였다. 전체적인 정확도 또한 99%로 향상하였으며 기존 방법론과 비교하였을 때, 우수한 성과를 확인했다. 본 연구는 냉동 컨테이너와 유사한 냉동 시스템 혹은 불균형한 데이터를 활용하여 기계에 대한 고장을 진단 및 원인을 분석하는 연구에 활용될 수 있을 것으로 기대한다.

REFERENCES

- [1] S. K. Park, Y. G. Park & Y. R. Shin. (2012). A Study on the Improvement of Damage to Reefer Container Cargo. *Journal of Navigation and Port Research*, 36(10), 803-810. DOI : 10.5394/KINPR.2012.36.10.803
- [2] P. Tang, O. A. Postolache, Y. Hao & M. Zhong. (2019). Reefer Container Monitoring System. In *2019 11th International Symposium on Advanced Topics in Electrical Engineering (ATEE) (pp. 1-6)*. IEEE. DOI : 10.1109/ATEE.2019.8724950
- [3] S. B. Yang (2018). *A Study on the Monitoring Systems of Reefer Containers*. thesis dissertation. KOREA MARITIME & OCEAN UNIVERSITY. Busan.http://kmou.dcollection.net/common/orgView/2000000016889
- [4] B. Jin, D. Li, S. Srinivasan, S. K. Ng, K. Poolla & A. Sangiovanni-Vincentelli. (2019). Detecting and diagnosing incipient building faults using uncertainty information from deep neural networks. In *2019 IEEE International Conference on Prognostics and Health Management (ICPHM)*. 1-8. DOI : 10.1109/ICPHM.2019.8819438
- [5] A. Kan, T. Wang, W. Zhu & D. Cao. (2021). The characteristics of cargo temperature rising in reefer container under refrigeration-failure condition. *International Journal of Refrigeration*, 123, 1-8. DOI : 10.1016/j.ijrefrig.2020.12.007
- [6] J. Loisel, S. Duret, A. Cornuejols, D. Cagnon, M. Tardet, E. Derens-Bertheau & O. Laguerre. (2021). Cold chain break detection and analysis: Can machine learning help?. *Trends in Food Science & Technology*, 112, 391-399. DOI : 10.1016/j.tifs.2021.03.052
- [7] Y. Fan, X. Cui, H. Han & H. Lu. (2020). Feasibility and improvement of fault detection and diagnosis based on factory-installed sensors for chillers. *Applied Thermal Engineering*, 164, 114506. DOI : 10.1016/j.applthermaleng.2019.114506
- [8] Y. Wang, Z. Wang, S. He & Z. Wang. (2019). A practical chiller fault diagnosis method based on discrete Bayesian network. *International Journal of Refrigeration*, 102, 159-167. DOI : 10.1016/j.ijrefrig.2019.03.008
- [9] G. Li, Y. Hu, H. Chen, L. Shen, H. Li, M. Hu & K. Sun. (2016). An improved fault detection method for incipient centrifugal chiller faults using the PCA-R-SVDD algorithm. *Energy and Buildings*, 116, 104-113. DOI : 10.1016/j.enbuild.2015.12.045
- [10] X. Liu, Y. Li, X. Liu & J. Shen. Fault diagnosis of chillers using very deep convolutional network. In *2018 Chinese Automation Congress (CAC) . IEEE*. 1274-1279. DOI : 10.1109/CAC.2018.8623749
- [11] G. Li, Q. Yao, C. Fan, C. Zhou, G. Wu, Z. Zhou & X. Fang. (2021). An explainable one-dimensional convolutional neural networks based fault diagnosis method for building heating, ventilation and air conditioning systems. *Building and Environment*, 108057. DOI : 10.1016/j.buildenv.2021.108057
- [12] K. B. Lee, S. H. Park, S. H. Sung & D. M. Park. (2019). A Study on the Prediction of CNC Tool Wear Using Machine Learning Technique. *Journal of the Korea Convergence Society*, 10(10), 15-21. DOI : 10.15207/JKCS.2019.10.11.015
- [13] K. B. Lee, S. H. Park, H. W. Lee, S. H. Lee & S. J. Lee. (2021) A study on the 3-step classification algorithm for the diagnosis and classification of refrigeration system failures and their types. *Journal of the Korea Convergence Society*, 12(8), 31-37. DOI : 10.15207/JKCS.2021.12.8.031
- [14] Y. D. Yun, Y. Y. Yang, H. S. Ji & H. S. Lim. (2017) Development of Smart Senior Classification Model based on Activity Profile Using Machine Learning Method. *Journal of the Korea Convergence Society*, 8(1), 25-34. DOI : 10.15207/JKCS.2017.8.1.025
- [15] H. J. Park. (2020). Trend Analysis of Korea Papers in the Fields of 'Artificial Intelligence', 'Machine Learning' and 'Deep Learning'. *Korea Information Electron Communication Technology*, 13(4), 283-292. DOI : 10.17661/jkiieect.2020.13.4.283
- [16] A. More. (2016). Survey of resampling techniques for improving classification performance in unbalanced datasets. *arXiv preprint arXiv:1608.06048*.

이 희 원(Hui-Won Lee)

[학생회원]



- 2021년 2월 : 동아대학교 경영정보학과(학사)
- 2021년 3월 ~ 현재 : 동아대학교 경영정보학과 석사과정
- 관심분야 : 머신러닝, 딥러닝
- E-Mail : slcw@naver.com

박 성 호(Sungho Park)

[정회원]



- 2017년 2월 : 동아대학교(학사)
- 2019년 2월 : 동아대학교(석사)
- 2019년 2월 ~ 현재 : 동아대학교 경영정보학과 박사과정
- 관심분야 : 머신러닝, 딥러닝
- E-Mail : psh2975@donga.ac.kr

이 승 현(Seung-hyun Lee)

[학생회원]



- 2021년 2월 : 동아대학교 경영정보학과(학사)
- 2021년 3월 ~ 현재 : 동아대학교 경영정보학과 석사과정
- 관심분야 : 머신러닝, 딥러닝
- E-Mail : hyunwow1263@naver.com

이 승 재(Seung-jae Lee)

[학생회원]



- 2021년 2월 : 동아대학교 경영정보학과(학사)
- 2021년 3월 ~ 현재 : 동아대학교 경영정보학과 석사과정
- 관심분야 : 머신러닝, 딥러닝
- E-Mail : sj2170497@donga.ac.kr

이 강 배(Kangbae Lee)

[정회원]



- 1991년 3월 ~ 1995년 8월 : 한국과학기술원 산업공학(박사)
- 2001년 3월 ~ 2004년 8월 : 부산가톨릭대학교 경영정보학과 교수
- 2008년 2월 ~ 현재 : 동아대학교 경영정보학과 교수
- 관심분야 : 머신러닝, 딥러닝
- E-Mail : kanglee@daonga.ac.kr