

# 군집화 기반 정상상태 식별을 활용한 시스템 에어컨의 냉매 충전량 분류 모델 개발

김 재 희<sup>1</sup>·노 유 정<sup>2†</sup>·정 종 환<sup>3</sup>·최 봉 수<sup>3</sup>·장 석 훈<sup>3</sup>

<sup>1</sup>부산대학교 기계공학부 석사과정, <sup>2</sup>부산대학교 기계공학부 교수, <sup>3</sup>LG전자 SAC개발실 책임연구원

## Development of Classification Model on SAC Refrigerant Charge Level Using Clustering-based Steady-state Identification

Jae-Hee Kim<sup>1</sup>, Yoojeong Noh<sup>2†</sup>, Jong-Hwan Jeung<sup>3</sup>, Bong-Soo Choi<sup>3</sup> and Seok-Hoon Jang<sup>3</sup>

<sup>1</sup>Graduate Student, School of Mechanical Engineering, Pusan Nat'l University, Busan, 46242, Korea

<sup>2</sup>Professor, School of Mechanical Engineering, Pusan Nat'l University, Busan, 46242, Korea

<sup>3</sup>Principal Research Engineer, SAC Research/Engineering Division, LG Electronics, Changwon, 51554, Korea

### Abstract

Refrigerant mischarging is one of the most frequently occurring failure modes in air conditioners, and both undercharging and overcharging degrade cooling performance. Therefore, it is important to accurately determine the amount of charged refrigerant. In this study, a support vector machine (SVM) model was developed to multi-classify the refrigerant mischarge through steady-state identification via fuzzy clustering techniques. For steady-state identification, a fuzzy clustering algorithm was applied to the air conditioner operation data using the difference between moving averages. The identification results using the proposed method were compared with those using existing steady-state determination techniques studied through the inversed Fisher's discriminant ratio (IFDR). Subsequently, the main features were selected using minimum redundancy maximum relevance (mRMR) considering the correlation among candidate features, and an SVM multi-classification model was devised using the derived features. The proposed method achieves satisfactory accuracy and robustness from test data collected in the new domain.

**Keywords** : system air conditioner, refrigerant mischarging, steady-state filter, moving average difference, fuzzy C-means clustering, mRMR

### 1. 서론

에어컨은 설계 기준에 따른 목표 냉방 성능을 유지하기 위해 에어컨 내외부에서 발생하는 다양한 형태의 고장모드를 감지하고 진단하는 작업이 요구된다. 그중 냉매 오충전은 에어컨 내부에서 빈번하게 발생하는 고장 모드 중 하나로, 적정 충전량 대비 증감 모두 제품의 성능 저하로 이어지기에 현재 충전량을 효과적으로 판단하는 것이 중요하다(Kim *et al.*, 2021; Kim *et al.*, 2022). 이에 따라, 에어컨의 냉매량에 따른 시스템 관측 및 고장진단에 관한 많은 연구가 진행되어 왔으며, Kim과 Braun(2012)은 냉매 누설에 따른 냉방 용량, COP(Coefficient

Of Performance) 그리고 SEER(Seasonal Energy Efficiency Ratio)의 변화를 관측하여 누설량 탐지 실패에 따른 운전 비용의 증가율을 연구하였다. 최근에는 RFE(Recursive Feature Elimination) 기반 랜덤 포레스트(random forest) 모델을 활용한 특징 선택 기법을 통해 VRF(Variable Refrigerant Flow) 시스템의 냉매 오충전 문제를 예측한 연구(Li *et al.*, 2020) 등 냉매 오충전 문제를 진단 혹은 예측하는 모델을 개발하는 연구도 진행되었다.

이러한 고장진단 모델의 개발에는 학습 과정에 사용될 데이터의 전처리 과정이 필수적으로 수반된다. 데이터 전처리에는 다양한 기법이 연구 및 개발된 만큼, 수집된 데이터의 형태 및 특성에 따라 적합한 전처리 기법을 선택하는 것이 필수적이다.

†Corresponding author:

Tel: +82-51-510-2308; E-mail: yoonoh@pusan.ac.kr  
Received October 18 2022; Revised November 15 2022;  
Accepted November 15 2022

© 2022 by Computational Structural Engineering Institute of Korea

This is an Open-Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License(<http://creativecommons.org/licenses/by-nc/3.0>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

에어컨의 경우, 설정 온도조건에 해당하는 냉방 성능을 출력하기 위해 과도상태를 지나 정상상태로 안정화되는 특성을 가진다. 안정화되지 않은 과도상태에서는 센서를 통해 측정되는 값의 편차가 크므로, 일반적으로 시스템의 고장진단을 위한 연구들은 과도상태와 정상상태를 구분하여 정상상태 데이터만을 활용하였다. Wan 등(2021)은 VRF 시스템의 냉매 유량을 판단하기 위해 전체 운전 구간을 과도 및 정상상태로 구분하여, 각 운전 상태에 대한 개별적인 유량 예측 모델을 개발하였다. 정상상태에 대해 20-Coefficient 모델, 효율 기반 모델 그리고 머신러닝 모델을 비교하였으며, CNN(Convolutional Neural Network)을 통해 과도상태에 대한 모델을 추가로 개발하였다. Yoo 등(2017)은 정상상태에 도달한 가정용 에어컨의 운전 데이터를 활용하여, 냉매 누설을 감지하는 주요 인자들을 선정 및 관측하였다. 하지만, 기존 연구에서는 냉매량 예측을 위한 전처리 방법을 제안하였지만, 각 방법에 대한 비교 연구가 없어 어떠한 전처리 방법이 효과적인지 판단하기 어렵다.

본 연구에서는 MAD(Moving Average Difference)를 활용한 퍼지 군집화 알고리즘을 통해 정상상태 식별을 진행함과 동시에, 식별 결과에 대해 군집 간의 상이한 정도를 정량적으로 나타내는 IFDR(Inversed Fisher's Discriminant Ratio)을 산출함으로써, 기존 기법들 간의 정상상태 식별 결과를 비교하였다. 이후, 판단된 정상상태 데이터에 대해 상관성을 고려한 특징 선택법을 적용함으로써, 에어컨의 냉매 충전량을 효과적으로 다중 분류하는 머신러닝 모델을 개발하였다.

## 2. 배경 이론

### 2.1 Fuzzy C-Means 군집화 알고리즘

군집화 알고리즘은 데이터 마이닝, 시계열 분석, 혹은 패턴 인식과 같이 다양한 연구 분야에 사용되는 비지도 학습 기법으로, 데이터 간의 거리를 측정하여 유사한 위치의 데이터를 하나의 그룹으로 정의하는 기법이다. 군집화 알고리즘에는 분할(partition), 계층(hierarchy), 퍼지 이론(fuzzy theory) 그리고 분포(distribution) 등에 기반한 다양한 기법들이 존재하며, FCM 군집화 알고리즘은 Dunn(1973)이 제안한 퍼지 이론 기반의 소프트 군집화(soft clustering) 기법이다. 이는 데이터가 다수의 군집에 동시에 속할 수 있다는 점을 바탕으로 데이터 간 거리와 군집에 속할 확률을 동시에 고려하여 군집 중심을 갱신하는 기법으로, 확률론적 접근에 따라 현실적으로 군집화를 진행한다는 점과 높은 군집 정확도를 가진다는 장점을 지닌다(Xu and Tian, 2015).

FCM 군집화 알고리즘은 판단하고자 하는 군집의 개수( $c$ )

를 사전에 결정하여야 하며, 식 (1)과 같이  $j$ 번째 데이터가 각 군집에 속할 가중치인  $\mu_{ij}$ 를 설정한다. 이때, 각 군집에 속할 가중치는 0과 1 사이의 값을 가지게 되며, 가중치의 합은 1이 된다.

$$\sum_{i=1}^c \mu_{ij} = 1 \tag{1}$$

각 데이터의 초기 가중치는 임의로 설정되며, 군집화 알고리즘은 각 군집 중심에 대해 데이터의 거리와 가중치를 모두 고려한 채, 각 군집 내 유사성을 최대화하기 위해 식 (2)의 목적함수를 최소화하는 방향으로 값을 갱신하게 된다.

$$O(\mu_{ij}, a_i) = \sum_{j=1}^n \sum_{i=1}^c \mu_{ij}^m \|x_j - a_i\|^2 \tag{2}$$

여기서,  $m$ 은 퍼지 알고리즘 내에서 가중치( $\mu_{ij}$ )를 반영하는 정도를 의미하는 하이퍼파라미터(hyperparameter)이다. 군집 중심( $a_i$ )과 가중치( $\mu_{ij}$ )는 식 (2)의 편미분을 통해 계산할 수 있으며, 각각 식 (3)과 식 (4)와 같은 형태로 표현할 수 있다.

$$a_i = \frac{\sum_{j=1}^n (\mu_{ij})^m x_j}{\sum_{j=1}^n (\mu_{ij})^m} \tag{3}$$

$$\mu_{ij} = \left( \frac{\sum_{k=1}^c \|x_j - a_i\|^{2/(m-1)}}{\sum_{k=1}^c \|x_j - a_k\|^{2/(m-1)}} \right)^{-1} \tag{4}$$

이후, 군집 중심과 각 데이터의 가중치 값이 1.0E-5 이하로 더 이상 변하지 않을 때까지 식 (2)부터 식 (4)까지의 과정을 반복하여 값을 계산한다.

### 2.2 Inverse Fisher's Discriminant Ratio

FDR(Fisher's Discriminant Ratio)은 두 1차원 데이터 그룹이 주어졌을 때, 통계적 특성을 비교하여 두 그룹 간의 상이한 정도를 판단하는 지표로써 활용된다(Webb, 2003). FDR은 식 (5)와 같이 그룹 간의 평균과 분산을 통해 계산되며, FDR 값이 클수록 그룹 간 거리가 멀며, 동시에 그룹 내 분산이 작음을 나타낸다.

$$FDR = \frac{(\mu_1 - \mu_2)^2}{\sigma_1^2 + \sigma_2^2} \tag{5}$$

그럼에도 FDR은 적용되는 정규화 방법 혹은 데이터의 형태에

따라 스케일이 일정하지 않으며, 상이한 정도를 판단하기 위한 명확한 임계값이 존재하지 않는다는 한계가 있다. 따라서 식 (6)과 같이 FDR에 대해 역수를 취한 IFDR을 활용하여, 0을 분리 정도에 대한 최대값으로 설정함으로써 기존의 FDR의 한계를 개선하였다.

$$IFDR = \frac{\sigma_1^2 + \sigma_2^2}{(\mu_1 - \mu_2)^2} \quad (6)$$

### 2.3 mRMR 특징 선택

특징 선택은 데이터 분석 및 머신러닝 모델 개발에 있어, 모델의 출력 변수와 상관성이 높은 조합을 선택함으로써 모델의 과대적합 문제를 방지함과 동시에 차원 축소를 통해 연산 속도를 향상시키는 과정이다. 일반적으로, 특징 선택 과정은 필터(filter), 래퍼(wrapper) 그리고 임베디드(embedded) 기반으로 구분할 수 있으며, mRMR특징 선택 기법은 Peng 등(2005)이 제시한 필터 기반 특징 선택 기법이다. 이는 출력 변수에 대한 최대 관련성을 가짐과 동시에 나머지 입력 변수들에 대한 최소한의 중복성을 가지는 특징을 선택하는 기법으로, 빠른 연산 속도와 함께 입력 변수들 간의 상관성을 최소화함으로써 학습된 모델의 설명력을 확보할 수 있다는 강점이 있다(Liu et al., 2022).

연속된 출력 변수에 대한 특징의 관련성은 식 (7)과 같이 F-통계량(F-statistics)을 통해 계산된다.

$$F(x_i, h) = \left[ \sum_k n_k (\bar{x}_k - \bar{x}) / (K - 1) \right] / \sigma^2 \quad (7)$$

$n_k$ 는  $k$ 번째 클래스에 해당하는 데이터의 개수,  $\bar{x}_k$ 는 해당 클래스 내 데이터의 평균을 의미한다.  $\bar{x}$ 는 전체 데이터의 평균,  $K$ 는 전체 클래스( $h$ )의 개수를 의미한다.  $\sigma^2$ 는 식 (8)과 같이 다수의 클래스에 대해 공통된 분산을 확인하기 위한 합동 분산(pooled variance)이며,  $\sigma_k^2$ 는 각 클래스에 해당하는 분산을 의미한다.

$$\sigma^2 = \left[ \sum_k (n_k - 1) \sigma_k^2 \right] / (n - K) \quad (8)$$

만면, 특징 간 중복성은 식 (9)와 같이 피어슨(Pearson) 상관계수를 통해 계산된다.

$$c(x, y) = \frac{\sum_1^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_1^n (x_i - \bar{x})^2} \sqrt{\sum_1^n (y_i - \bar{y})^2}} \quad (9)$$

$x$ 와  $y$ 는 서로 다른 입력 변수를 의미하며, 피어슨 상관계수는 두 변수의 공분산을 각각의 표준편차로 나눈 값으로 두 변수 간의 선형 관계를 나타내는 지표로서 활용된다. 이후, 식 (7)과 식 (9)를 통해 각 특징에 대한 식 (10)의 목적함수를 계산하며, 최소 중복성과 최대 관련성을 판단하기 위해 목적함수를 최대 로 하는 변수를 주요 특징으로써 고려한다. 이때,  $S$ 는 데이터 내 특징 공간(feature space)을 의미한다.

$$O(F, c) = F(i, h) / \left| \sum_{j \in S} |c(i, j)| \right| \quad (10)$$

### 2.4 Support Vector Machine

SVM(Support Vector Machine)은 회귀 및 분류 분석에 활용되는 대표적인 지도학습 기반의 머신러닝 모델로써, 각 클래스에 해당되는 데이터를 구분하는 초평면(hyper plane)을 생성하는 기법이다. 초평면을 기준으로 최근접 거리에 위치한 각 클래스의 데이터를 서포트 벡터(support vector)로 설정 후 두 서포트 벡터 간의 거리인 마진(margin)을 최대화하는 것을 목표로 하며(Cortes and Vapnik, 1995), 기존의 머신러닝 모델에 비해 높은 일반화 성능을 가지고 있다. 또한, 커널(kernel) 트릭을 활용함으로써 고차원의 데이터에도 높은 모델 성능을 확보할 수 있다는 장점을 가지고 있다.

### 3. 제안된 방법

제안된 방법은 MAD 기반 퍼지 군집을 통해 데이터의 정상 상태를 식별하는 과정과 mRMR을 활용한 최적 특징 조합과 SVM을 결합한 냉매 충전량에 대한 다중 분류 모델 생성 과정으로 구분된다. Fig. 1에서 보는 바와 같이 먼저 도메인 지식을 기반으로 냉매량 예측에 중요한 인자를 도출하는 특징 엔지니어링(feature engineering) 과정을 거친 후, 각 특징 데이터에 대한 MAD를 계산하고 이를 정규화한 후, 퍼지 C-평균 클러스터링 기법을 이용해 정상상태가 시작되는 지점을 찾아 정상상태 데이터만을 추출한다.

정상상태 데이터는 정규화 과정과 중복 데이터 제거 과정을 통해 전처리 과정이 수행되고, mRMR 알고리즘을 이용해 특징 간 상관성이 가장 낮으면서도 분류 정확도에 큰 영향을 미치는 특징 조합을 최종적으로 도출한다. 학습과 검증 데이터 셋을 나눈 후, 하이퍼파라미터 최적화를 통해 데이터 적합도가 높은 SVM 모델을 생성하였다. 실제 냉매 데이터를 이용하여 모델을 검증한 결과는 4~6장에 상세히 설명되어 있다.

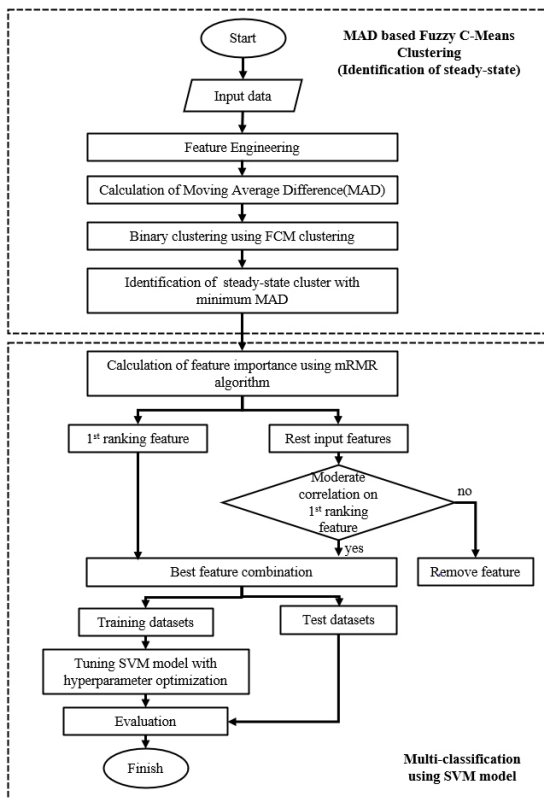


Fig. 1 Flowchart of the proposed method

#### 4. 데이터 수집

##### 4.1 데이터 수집 환경

냉매 오충전 실험은 냉매량 변화 및 운전 부하와 같은 다양한 운전 조건을 설정하기 위해 대형 챔버(chamber)가 설치된 환경에서 모사되었다. 데이터 수집에 사용된 시스템 에어컨은 Fig. 2와 같이 압축기, 팽창밸브, 응축기 그리고 4개의 증발기로 구성되어 있으며, 각 위치에 설치된 온도, 압력, 전기 그리고 제어 센서를 통해 2초 간격으로 각 실험 조건에 해당하는 에어컨 운전 데이터를 수집하였다.

실험은 Table 1에 나타난 표준 및 과부하 온도 조건에서 진행되었으며, 압축기가 정지된 상태에서부터 최대 4068초 간 운전하였다. 각 실험은 적정 충전량 기준 90에서 110%의 충전 구간인 정상 조건과 더불어, 정상 대비 120, 130%의 과충전 조건과 70, 80%의 부족 조건을 모사하여 총 5가지의 충전 조건에 대해 진행하였다.

##### 4.2 특징 엔지니어링

사이클 내에서 냉매량이 변하는 경우, 냉매가 지나가는 모든 사이클 구성요소의 운전 특성에 변화가 발생한다. 이러한 변화

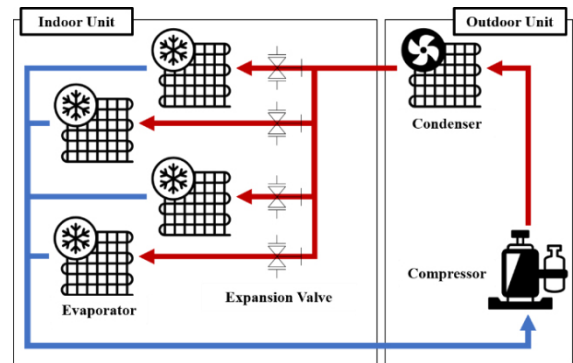


Fig. 2 Schematic diagram of air conditioner

Table 1 Experiment temperature condition

Condition	Indoor		Outdoor	
	DB (°C)	WB (°C)	DB (°C)	WB (°C)
Standard	27	19	35	24
Overload	32	23	43	32

를 나타내는 지표들은 냉매의 오충전량에 대해 민감하게 반응하므로 기존의 센서로부터 수집된 정보를 직접적으로 활용하거나 물리적 관계를 바탕으로 생성 후 활용하는 것이 중요하다.

냉매가 오충전되는 경우, 전체 냉동시스템 내부에 존재하는 냉매의 유량이 변하게 되며 이로 인해 사이클 전체의 압력이 변하게 된다. 압력의 증감은 사이클 내부에서 다양한 형태의 변화를 일으키지만, 그중 공기와 열교환을 직접적으로 수행하는 응축기와 증발기의 경우 냉매의 상변화에 필요한 열량이 변하게 되어 열교환기 내에서 수행하는 열교환 성능에 영향을 주게 된다. Yoo 등(2017)의 연구에 따르면 과열도(superheating), 과냉도(subcooling), 증발기 온도차(evaporator TD) 그리고 응축기 온도차(condenser TD)는 오충전 정도에 따라 열교환기 내부에서 발생하는 열교환 변화를 효과적으로 나타내는 인자들이며, 이들의 변화는 압축기와 팽창밸브 내에서의 토출과열도(discharge superheating), 밸브 개도(expansion valve opening) 그리고 팽창밸브 과열도(expansion valve superheating) 등의 변화를 추가로 유발하게 된다. 이에 따라, 기존 센서로부터 수집된 특징들을 활용하여 Table 2과 같이 핵심 인자를 정의하였고, 최종적으로 사용한 특징들은 Table 3에 나타나 있다.

#### 5. 퍼지 군집 기반 정상상태 식별

##### 5.1 정상상태 식별

일반적으로 과도상태와 정상상태는 시간에 따라 값이 변하는 정도로 정의되며, 사전적으로 정상상태는 변화율이 0인 상태를 의미한다. 하지만, 현실에서 센서 값이 항상 일정한 경우

Table 2 Generated features

No.	Features	Formula
1	EEV opening	$\frac{\sum_{i=1}^n EEV_i \cdot capacity_i}{\sum_{i=1}^n capacity_i}$
2	EEV superheating	$\frac{\sum_{i=1}^n (T_{evap,out} - T_{evap,in})_i \cdot capacity_i}{\sum_{i=1}^n capacity_i}$
3	Evaporator TD	$T_{air,in} - T_{evap}$
4	Condenser TD	$T_{cond} - T_{air,out}$

EEV(Electric Expansion Valve): 팽창밸브 개도,  $T_{evap,in}$ : 증발기 흡입 온도,  $T_{evap,out}$ : 증발기 토출온도,  $T_{evap}$ : 증발온도,  $T_{cond}$ : 응축온도,  $T_{air,in}$ : 실내 유입 공기온도,  $T_{air,out}$ : 실외 유입 공기온도, Capacity: 실 내기 용량

Table 3 Engineered features

No.	Features
1	EEV opening
2	EEV superheating
3	Evaporator TD
4	Condenser TD
5	Subcooling
6	Superheating
7	Discharge superheating
8	Input current

는 드물며, 이에 따라 정상상태를 현실적으로 정의하기 위한 기준이 필요하다. 이에 따른 다양한 연구가 진행되어 왔으며, 본 연구에서는 기존의 기법들이 가지는 한계와 더불어, 각 기법에 따른 정상상태 식별 결과를 비교하였다.

Cao와 Rhinehart(1995)는 EWMA(Exponentially Weighted Moving Average)를 활용하여 MSD(Mean Square Deviation)와 MSSD(Means Square differences of Successive Data) 간의 F-통계량으로 정상상태를 판단하는 기법을 제안하였다. Wan 등(2021)은 냉매유량을 제외한 모든 특징에 대해 1분간 측정된 데이터를 하나의 그룹으로 하여, 해당 그룹 내 상대 편차가 1% 미만이 되는 경우를 정상상태로 판단하였다. 두 기법의 경우, 다변량 데이터에 적용 시 각각의 특징에 대해 판단기법을 적용 후 교차하는 구간을 최종적으로 정상상태로 판단하므로 특징의 개수가 늘어나는 경우 정상상태를 판단하는 기준이 엄격해지는 문제로 이어지게 된다. Fig. 3은 정상상태 식별을 위해 사용된 특징에 따라 판단된 정상상태 구간의 크기를 나타낸 그래프이며, 특징 수가 증가함에 따라 최종적으로 판단되는 정상상태의 구간이 짧아지는 문제를 확인하였다. Kim 등(2008)은 ASHRAE Standard 37-2005에 근거한 안정화된 시점에서 변화량이 가장 큰 특징의 평균과 표준편차를 활용하여 정상상태 판단을 위한 임계값을 설정하였다. 이후, 전체 운전 구간에 대해

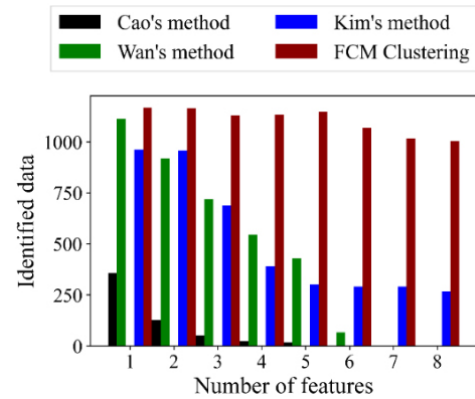


Fig. 3 Comparison of steady state filter results

해당 특징이 임계값 범위 내로 수렴하는 영역을 정상상태로 판단하였다. 본 연구에 사용된 시스템 에어컨의 경우, KS B ISO15042에 기재된 안정화 구간의 판단 조건을 적용하였다. 해당 규정의 경우, 5분간 측정된 평균 능력이 이전 35분간 측정된 평균 능력에 대해 2.5% 이내의 변화율을 보이는 경우 안정화 구간으로 판단한다. 이와 같은 규정에 근거한 임계값 설정은 장시간의 최소 운전시간이 요구된다는 문제가 발생한다. 또한, 이전 두 기법과 마찬가지로 확인하고자 하는 특징 수가 증가함에 따라 정상상태가 적게 식별되는 문제가 발생하였음을 확인하였다.

이에 따라, 본 연구에서는 다수의 특징을 동시에 고려하며, 시간에 따른 특징의 변화율과 데이터 간의 거리를 활용하여 정상상태를 판단하고자 한다. 먼저 시간에 따른 변화율을 판단하기 위해 각 특징의 이동 평균(moving window average)을 계산한 후, 식 (11)과 같이 인접한 두 윈도우(window) 평균 간의 차이를 의미하는 MAD를 계산하였다.

$$MAD_{i+1} = |MWA_{i+1} - MWA_i| \tag{11}$$

$MWA_i$ 는  $i$ 번째 윈도우에 해당하는 평균으로써 전문가의 경험에 따라, 에어컨 내에서 유의미한 변화를 확인할 수 있는 180 초를 평균 계산을 위한 최소 윈도우 크기로 설정하였다. 변환된 MAD 데이터에 최소-최대 정규화(min-max normalization)를 적용함으로써 특징별로 MAD의 스케일이 달라지는 현상을 방지하였으며, 이를 Fig. 4에 나타내었다.

정규화된 MAD 데이터로부터 데이터 간 거리에 따라 전체 데이터를 과도 및 정상상태로 이진 군집하기 위해 군집 개수를 사전에 지정하는 알고리즘을 채택하였다. 군집 개수를 지정하는 알고리즘으로는 대표적으로 K-평균 군집화(K-means clustering)와 FCM 군집화 기법이 존재하며, 각 기법은 주어진 데이터의 형태에 따라 채택 여부를 결정할 필요가 있다. 에어컨 운전 데

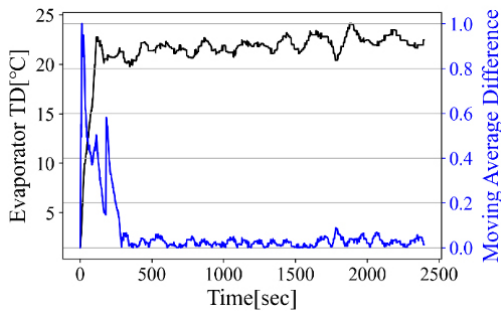


Fig. 4 MAD graph of Evaporator TD

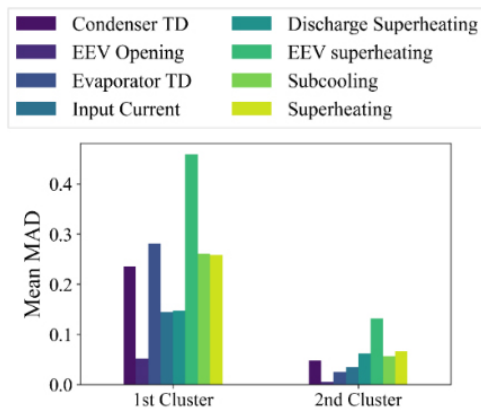


Fig. 5 Boxplot of MAD in each cluster

이터의 경우 연속된 시계열 데이터라는 특성이 있기에, 과도 상태에서 정상상태로 변하는 구간을 효과적으로 군집화하기 위해 하나의 데이터가 두 군집 모두에 속할 수 있다는 FCM 군집화 알고리즘을 적용하였다. 최종적으로, 알고리즘에 따라 판단된 두개의 군집 중 정상상태에 해당하는 군집을 판단하기 위해 각 군집에 속한 특징별 평균 MAD를 계산하였다.

Fig. 5는 이진 군집된 두 그룹 내 특징들이 가지는 평균 MAD를 나타낸 그림이며, 평균 MAD가 최소가 되는 두 번째 클러스터(2nd cluster)를 정상상태에 해당하는 군집으로 판단하였다. 본 기법의 경우, 군집화 알고리즘의 학습 과정에서 전체 특징을 동시에 고려하므로 기존 정상상태 판단 기법에서의 특징 수에 따른 정상상태 식별 오류가 발생하지 않았다는 것을 Fig. 3를 통해 알 수 있다.

### 5.2 IFDR기반 정상상태 판단 정도 비교

본 연구에서 적용한 정상상태 식별 기법의 유효성과 기존 기법 대비 개선 정도를 판단하기 위해 식별된 과도상태와 정상상태 데이터 간의 분리 정도를 확인하고자 한다. 이에 따라, IFDR을 각 기법에 따른 식별 결과에 적용 후 비교하였다. 선택한 8개의 특징을 모두 고려하는 경우, Cao와 Wan의 기법은 전체 운전 구간을 과도상태로 식별하였으므로, Kim의 기법에 따

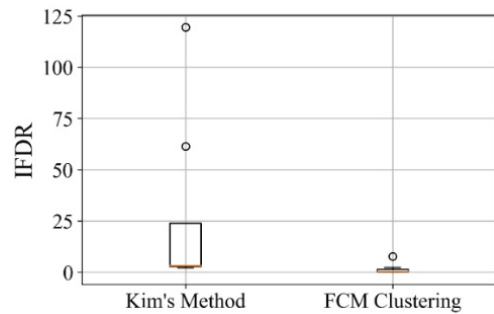


Fig. 6 IFDR for each identification method

Table 4 Feature ranking based on mRMR feature selection

Rank	Features	r	Selected
1	Condenser TD	1.000	Yes
2	Superheating	0.079	No
3	Expansion valve opening	0.477	Yes
4	Subcooling	0.486	Yes
5	Discharge superheating	0.640	Yes
6	Evaporator TD	0.608	Yes
7	EEV superheating	0.043	No
8	Input current	0.415	Yes

른 결과만을 비교하였다. Fig. 6은 각 특징의 IFDR 값을 사분위수로 나타낸 그림이며, 본 연구에서 제시한 FCM 군집화 기반 정상상태 식별 기법이 0에 근접한 IFDR 값으로, 운전 구간 간의 상이성이 최대가 되는 식별 결과를 확보하였음을 확인하였다.

## 6. 분류 모델의 개발

### 6.1 mRMR 알고리즘 기반 특징 선택

냉동사이클은 냉매를 압축, 팽창 그리고 열교환하는 다양한 요소들로 구성되어 있다. 각 요소들은 시스템 내 배관을 통해 직접적으로 연결이 되어있는 만큼, 개별 요소에서 발생하는 다양한 현상들은 나머지 요소에도 직접적인 영향을 주게 된다. 사이클 요소 간의 영향력은 특징들 간의 높은 상관성으로 이어지며, 이는 분류 모델 개발에 있어 모델의 설명력을 저하시키는 문제로 이어진다.

따라서, 냉매량 변화를 효과적으로 감지함과 동시에, 특징들 간의 상관성을 고려한 조합을 선정하기 위해 mRMR 알고리즘 기반 특징 선택 기법을 적용하였다. 냉매량 변화에 민감하게 반응하는 8개의 특징들에 대해 적용함으로써, 특징 간의 중요도 순위를 계산하여 Table 4에 나타내었다. 이후, Sun 등 (2016)의 연구에서 제시한 기법에 따라 1 순위 특징을 랜드마크 (landmark) 특징으로 지정 후, 랜드마크 특징과  $0.3 \leq |r| < 0.8$



Table 5 Confusion matrix of classification result

Refrigerant (%)		Predicted value				
		70	80	Normal	120	130
True value	70	429	1	4	0	0
	80	0	501	4	1	0
	Normal	0	4	1486	1	4
	120	0	0	1	434	0
	130	0	0	0	0	394
Accuracy		99.4%				

Table 6 Classification evaluation metrics

Refrigerant (%)	Accuracy	Precision	Recall	F1 Score
70	1.000	1.000	0.988	0.994
80	1.000	0.990	0.990	0.990
Normal	0.999	0.994	0.994	0.994
120	0.998	0.995	0.998	0.997
130	1.000	0.990	1.000	0.995

의 스피어만(Spearman) 상관계수를 가지는 특징들을 최종적으로 선택하였다.

### 6.2 SVM 모델 학습 및 분류 결과

mRMR 알고리즘 기반 특징 선택법으로 최종 선택한 특징 조합에 대해 SVM 모델 학습을 진행하였다. 다중 분류를 위해 SVM 모델의 결정경계(decision boundary)는 One-Versus-One 으로 결정하였으며, 추가적인 하이퍼파라미터 결정을 위해 격자 탐색(grid search)과 k-fold 교차 검증(k=5)을 활용하여 c,  $\gamma$  그리고 kernel 형태를 최적화하였다. 모델의 최적화에는 계층적 샘플링(stratified sampling)을 통해 7:3의 비율로 데이터를 학습 및 검증용 데이터 세트로 구분하여 사용하였다.

최적화가 완료된 SVM 모델을 통해 검증용 데이터에 대한 모델의 성능을 확인하였으며, Table 5와 같이 분류 결과에 대한 혼동행렬(confusion matrix)을 통해 99.4%의 분류 정확도(accuracy)를 가지는 것을 확인하였다. 또한, Table 6을 통해 분류 모델의 성능 지표인 정밀도(precision), 재현율(recall) 그리고 F1 점수(F1 score) 모두 1에 근접한 값을 가짐으로써, 학습된 모델이 냉매 충전 정도를 효과적으로 분류함을 확인하였다.

### 6.3 시험 데이터에 대한 모델 검증

일반적으로 머신러닝 모델을 최적화 및 학습하기 위해 하나의 데이터를 일정한 비율로 학습과 검증용으로써 구분하여 활용한다. 비록 검증용 데이터가 모델 학습 과정에 직접적으로 포함되지 않음에도 불구하고, 학습용 데이터와 동일한 도메인

Table 7 Description of test data

	Charge level	Load condition	Label
Test Data #1	90%	Overload	Normal
Test Data #2	100%	Standard	Normal

Table 8 Classification evaluation metrics of validation

	Accuracy	Precision	Recall	F1 Score
Test Data #1	0.944	1.000	0.944	0.971
Test Data #2	0.999	1.000	0.999	1.000
Total	0.972	1.000	0.972	0.986

에서 수집된 데이터라는 점에서, 일반적으로 학습용 데이터로 최적화된 모델에서 높은 분류 정확도를 보이게 된다. 이에 따라, 모델의 강건성을 확인하기 위해서는 새로운 도메인으로부터 수집된 시험 데이터에서 모델의 성능을 유지할 수 있어야 한다.

모델 검증을 위한 시험 데이터는 학습용 데이터가 수집된 제품과 상이한 용량을 가진 실내기가 설치된 제품으로부터 수집되었다. 실내기 용량이 달라짐에 따라, 에어컨 내 필요로 하는 총 냉매량 또한 달라지게 되며, 학습용 데이터의 경우 5.45kg을 적정 충전량으로 설계된 반면, 시험 데이터의 경우 5.24kg을 적정 충전량으로 설계된 제품이다. 이와 같이 상이한 실내기 용량 및 냉매 요구량 조건에서 수집된 운전 데이터에 대해, 학습된 모델의 분류 성능 변화를 확인하고자 한다. 두 시험 데이터는 Table 7과 같이 각각 정상 조건에 해당되는 정량 대비 90%와 100%로 충전된 상태이며, 서로 다른 온도 조건에서 운전된 데이터이다. 각 데이터에 대해 모델의 분류 결과를 확인하였으며, Table 8과 같이 시험 데이터에 대해 각각 94.4%와 99.9%의 분류 정확도를 가지는 것을 확인함으로써, 새로운 도메인에서 수집된 운전 데이터에도 강건한 분류 성능을 가지는 것을 확인하였다.

마지막으로, 정상상태 식별 기법이 모델의 분류 정확도에 미치는 영향을 확인하기 위해 Kim의 기법을 이용해 식별된 정상상태 데이터로 다중 분류 모델을 학습 후, 시험 데이터에 대한 분류 정확도를 비교 분석하였다. Fig. 7과 같이, Kim의 기법을 적용하여 학습한 모델은 시험 데이터에 대한 64.8%의 평균 분류 정확도를 가지는 반면, 본 연구에서 적용한 FCM 군집화 기법을 이용해 식별된 정상 데이터로 모델을 학습 시 32.4%만큼 분류 성능이 향상되었음을 확인하였다.

첫 번째 시험 데이터에 대해 Kim의 방법이 정확도가 낮은 이유는 엄격한 정상상태 식별 기준으로 인해, Fig. 8과 같이 정상상태로 식별된 특징들의 분포가 좁아짐에 따라 발생한 문제로 볼 수 있다. Fig. 8은 각 기법에 따라 식별된 학습용 정상상태 데이터와 시험용 정상상태 데이터의 분포를 커널 밀도 함수

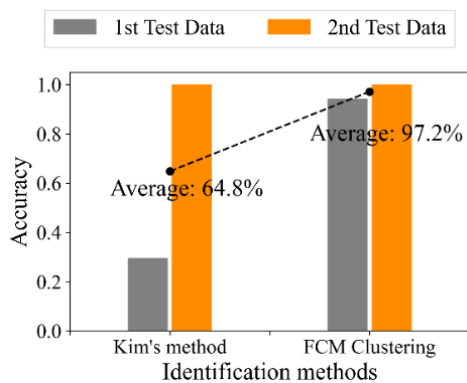


Fig. 7 Test accuracies for each identification method

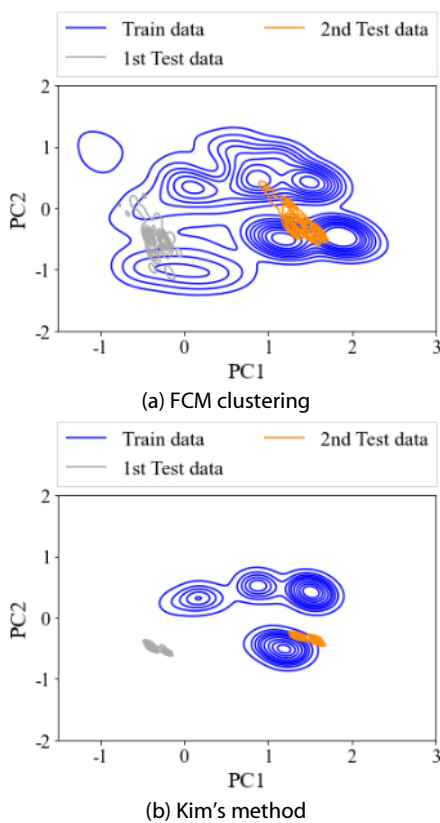


Fig. 8 Data distributions

로 시각화한 그림으로, 각 축은 주성분 분석(principle component analysis)을 통해 연산된 주성분들을 나타낸 것이다. Fig. 8(a)와 Fig. 8(b) 간의 분포 변화와 같이 엄격한 식별 기준에 따라 학습에 사용되는 특징 분포가 좁아졌으며, 이로 인해 첫 번째 시험 데이터에 대해 오분류가 증가하였음을 확인할 수 있다. 반면, 두 번째 시험 데이터에 대해서는 Kim의 기법 또한 높은 분류 정확도를 가짐을 확인하였다. 이는 90% 충전 조건으로 결정 경계에 근접하게 위치한 첫 번째 시험 데이터와 달리, 100% 충전 조건으로 정상 데이터의 중앙에 위치하기에 특징 분포가 좁아짐에 따른 영향이 적게 나타난 결과이다.

## 7. 결론

본 연구에서는 에어컨 운전 데이터 내 정상상태를 효과적으로 식별하기 위해 MAD 기반 FCM 군집화 알고리즘 기법을 적용하였다. 이를 통해, 특징 수의 증가에 따라 정상상태 식별이 엄격해지는 기존 기법들의 한계를 해결하였으며, 식별된 정상상태와 과도상태 간의 상이성을 IFDR 값을 통해 확인함으로써 두 구간을 효과적으로 구분하였음을 확인하였다. 이후, 식별된 정상상태 데이터에 대해 특징 간 상관성을 고려한 mRMR 특징 선택법을 적용하는 과정을 통해 냉매 충전량을 다중 분류하는 머신러닝 모델을 개발하였다. 개발된 모델은 주어진 냉매 오충전 문제를 99.4%의 정확도로 분류하는 것을 확인하였으며, 새로운 도메인에서 수집된 시험 데이터에 대해 97.2%의 분류 정확도를 가지는 것을 확인함으로써 모델의 강건성을 검증하였다.

## 감사의 글

이 논문은 부산대학교 기본연구지원사업(2년)에 의하여 연구되었음.

## References

- Cao, S., Rhinehart, R.R. (1995) An Efficient Method for On-line Identification of Steady State, *J. Process Control*, 5(6), pp. 363~374.
- Cortes, C., Vapnik, V. (1995) Support-vector Networks, *Mach. Learn.*, 20(3), pp.273~297.
- Dunn, J. C. (1973) A Fuzzy Relative of the ISODATA Process and Its Use in Detecting Compact Well-Separated Clusters, *J. Cybern.*, 3(3), pp.32~57.
- Kim, M., Yoon, S.H., Domanski, P.A., Payne, W.V. (2008) Design of a Steady-state Detector for Fault Detection and Diagnosis of a Residential Air Conditioner, *Int. J. Refrig.*, 31(5), pp.790~799.
- Kim, S., Kang, Y.-J., Noh, Y., Park, S., Ahn, B. (2021) Fault Classification Model based on Time Domain Feature Extraction of Vibration Data, *J. Comput. Struct. Eng. Inst. Korea*, 34(1), pp.25~33.
- Kim, S., Noh, Y., Kang, Y.-J., Park, S., Lee, J.W., Chin, S.W. (2022) Hybrid Data-scaling Method for Fault Classification of Compressors, *Meas.*, 201, p.111619.
- Kim, W., Braun, J.E. (2012) Evaluation of the Impacts of Refrigerant Charge on Air Conditioner and Heat Pump Performance, *Int. J. Refrig.*, 35, pp.1805~1814.



- Li, Z., Shi, S., Chen, H., Wei, W., Wang, Y., Liu, Q., Liu, T.** (2020) Machine Learning based Diagnosis Strategy for Refrigerant Charge Amount Malfunction of Variable Refrigerant Flow System, *Int. J. Refrig.*, 110, pp.95~105.
- Liu, A.J., Mukherjee, A., Hu, L., Chen, J., Nair, V.N.** (2022) Performance and Interpretability Comparisons of Supervised Machine Learning Algorithms: An Empirical Study, arXiv preprint arXiv :2204.12868.
- Peng, H., Long, F., Ding, C.** (2005) Feature Selection based on Mutual Information Criteria of Max-Dependency, Max-Relevance, and Min-Redundancy, *IEEE Trans. Pattern Anal. & Mach. Intell.*, 27, pp.1226~1238.
- Sun, K., Li, G., Chen, H., Liu, J., Li, J., Hu, W.** (2016) A novel Efficient SVM-based Fault Diagnosis Method for Multi-Split Air Conditioning System's Refrigerant Charge Fault Amount, *Appl. Therm. Eng.*, 108, pp.989~998.
- Wan, H., Cao, T., Hwang, Y., Chang, S.D., Yoon, Y.J.** (2021) Machine-Learning-based Compressor Models: A Case Study for Variable Refrigerant Flow Systems, *Int. J. Refrig.*, 123, pp.23~33.
- Webb, A.R.** (2003) *Statistical Pattern Recognition*. John Wiley & Sons.
- Xu, D., Tian, Y.** (2015) A Comprehensive Survey of Clustering Algorithms, *Ann. Data Sci.*, 2, pp.165~193.
- Yoo, J.W., Hong, S.B., Kim, M.S.** (2017) Refrigerant Leakage Detection in an EEV Installed Residential Air Conditioner with Limited Sensor Installations, *Int. J. Refrig.*, 78, pp. 157~165.

## 요 지

냉매 오충전은 에어컨에서 빈번하게 발생하는 고장 모드 중 하나로, 적정 충전량 대비 부족 및 과충전 모두 냉방 성능의 저하를 유발하므로 충전된 냉매량을 정확하게 판단하는 것이 중요하다. 본 연구에서는 퍼지 군집화 기법을 통한 정상상태 식별을 통해 냉매 오충전량을 다중 분류하는 모델을 개발하였다. 정상상태 식별을 위해 에어컨 운전 데이터에 대해 이동 평균 간의 차이를 활용한 퍼지 군집화 알고리즘을 적용하였으며, IFDR를 통해 기존 연구된 정상상태 판단 기법들과 식별 결과를 비교하였다. 이후, 시스템 내 상관성을 고려한 mRMR을 이용해 특징을 선택하였으며, 도출된 특징을 이용해 SVM 기반의 다중 분류 모델이 생성되었다. 제안된 방법은 시험 데이터를 통해 만족할 만한 분류 정확도와 강건성을 도출하였다.

**핵심용어** : 시스템에어컨, 냉매 오충전, 정상상태 필터, 이동평균차, 퍼지 C-평균 군집화, 최소중복최대상관