

이상 고객 데이터에 강건한 냉장고 온도 예측용 데이터 수준 연합학습 기법

배진수¹ · 박소연¹ · 최지형¹ · 김지현¹ · 함성일² · 이보경² · 강필섭² · 김성범^{1*}

¹고려대학교 산업경영공학과 / ²삼성전자

Data-Level Robust Federated Learning for Refrigerator Temperature Prediction with Anomalous Customers

Jinsoo Bae¹ · Soyeon Park¹ · Jihyung Choi¹ · Jihyun Kim¹ · Seongil Ham² · Bokung Lee² · Pilsob Kang² · Seoung Bum Kim¹

¹School of Industrial and Management Engineering, Korea University

²Samsung Electronics

Refrigerator temperature prediction using neural networks can improve both energy efficiency and customer experience in smart home environments. Training high-quality neural networks, however, requires large-scale data that accurately reflects diverse customer usage patterns. In practice, collecting and centralizing such data is challenging because of data privacy concerns. To address this issue, we adopt federated learning (FL), which enables collaborative model training across customers while preserving data privacy. In this study, we propose a robust FL that incorporates a data-level anomalous customer detection using an autoencoder model. This component is crucial because a subset of customers may exhibit anomalous data from various factors such as hardware malfunctions or abnormal usage patterns, which degrade the efficacy of the FL. The proposed framework, named DrFL, identifies customers with anomalous data and excludes them in the FL, thus preventing their adverse impact on the global model. Experimental results demonstrate that DrFL reduces prediction error by over 40% on average compared to existing FL methods, and maintaining robust generalization performance for new customers who did not participate in federated training. These results indicate that the proposed method effectively mitigates the influence of anomalous customer data.

Keywords: Federated Learning, Refrigerator, Temperature Prediction, Anomalous Customer, Deep Learning

1. 서 론

최근 사물인터넷(internet of things, IoT) 기술의 급격한 발전과 함께, 네트워크에 연결되어 데이터를 수집하고 스스로 상태를 제어할 수 있는 스마트 가전(smart appliances) 보급이 빠르게 확대되고 있다. 이러한 환경 변화에 따라, 가정 내 에너지 관리를 최적화하고 사용자 경험을 극대화하기 위한 데이터 분석 기술 중요성 또한 날로 커지고 있다. 그 중에서도 냉장고는 365일 24시간 가동되며 가정 내 전체 전력 소비의 상당 부분을

차지할 뿐만 아니라, 식재료 신선도 유지와 직결되는 핵심 가전제품이다(Biglia *et al.*, 2020; Li *et al.*, 2025). 따라서 냉장고 내부 온도의 미래 변화를 예측하는 기술은 스마트 가전 서비스의 핵심 요소로 꼽힌다. 예를 들어, 정확한 미래 온도 예측이 가능하다면 냉장고는 압축기의 가동 시점을 최적화하여 불필요한 에너지 소모를 줄이거나, 사용자의 문 열림 패턴을 미리 파악하여 급격한 온도 상승 전에 선제적으로 냉각을 수행함으로써 식품의 신선도를 획기적으로 유지할 수 있다(Zhao *et al.*, 2024).

* 연락저자 : 김성범 교수, 02841 서울특별시 성북구 안암로 145 고려대학교 산업경영공학과, Tel : 02-3290-3397, Fax : 02-929-5888, Email : sbkim1@korea.ac.kr

2026년 1월 29일 접수; 2026년 3월 26일 수정본 접수; 2026년 4월 6일 게재 확정.

이러한 잠재력에 주목하여, 최근 산업계에서는 냉난방공조 (heating, ventilation, and air conditioning, HVAC) 시스템의 미래 온도 예측 모델 개발을 위하여 데이터 분석 기술을 활용하고 있다(Mtibaa *et al.*, 2020; Machacuay *et al.*, 2025). 예를 들어, Mtibaa *et al.*(2020)은 스마트빌딩 내 HVAC 시스템을 대상으로 다변량 센서 데이터를 학습한 장기-단기 기억 신경망(long short-term memory model, LSTM) 모델 사용을 제안하였으며, 이를 기반한 실내 온도 예측 프레임워크를 제시하였다. Machacuay *et al.*(2025)은 기존 LSTM 구조에 얕은 인공신경망(shallow neural network)을 결합한 ShallowLSTM 모델을 개발하여 냉장·냉동 시스템의 미래 온도 예측 정확도를 크게 향상시켰다. 이러한 선행 연구들은 LSTM과 같은 시계열 딥러닝 모델이 열역학적 동특성을 효과적으로 학습하여 기존 물리 기반 온도 예측 모델을 대체할 수 있는 기술적 가능성을 보여주었다.

이때 실제 주거 환경에서도 우수한 성능을 발휘할 수 있는 냉장고 온도 예측 모델 구축을 위해서는, 단일이 아닌 다수의 냉장고 사용 고객으로부터 수집된 대규모 학습 데이터가 필수적이다. 이는 가정마다 냉장고 설치 위치의 온도 및 습도 조건이 다를 뿐만 아니라, 문 열림 빈도, 시간대별 식재료 사용 패턴, 보관량 등 사용자별 행동 특성이 크게 다르기 때문이다. 이러한 다양한 사용 행태가 충분히 반영되지 않으면 모델이 특정 사용자나 특정 환경 조건에 편향될 가능성이 높아지며, 실제 서비스 환경에서의 성능 또한 크게 저하될 수 있다(Li *et al.*, 2019; Srivastava *et al.*, 2014). 이러한 문제를 해결하기 위하여 기업은 여러 고객에게서 발생하는 사용 이력을 통합해 대규모 데이터셋을 구축할 수 있다. 하지만 고객 사용 이력 데이터를 중앙 서버로 수집 및 관리하는 과정은 여러 현실적 제약 조건을 가진다. 예를 들어, 냉장고 사용 로그에는 식사 시간, 외출 여부, 생활 패턴 등 민감한 사생활 정보가 포함될 수 있어 중앙화 과정에서 개인정보 침해 우려 가능성이 매우 높다(Brand *et al.*, 2020; Zeng *et al.*, 2017). 또한, 수백만 대의 장치에서 발생하는 고빈도 시계열 데이터를 실시간으로 업로드하고 저장하기 위해서는 막대한 네트워크 대역폭 비용과 서버 스토리지 비용이 필요하다(Pouriyeh *et al.*, 2022; Shahid *et al.*, 2021).

이와 같은 문제 해결을 위한 방법으로 연합학습(federated learning, FL) 방법론이 주목받고 있다. 연합학습은 고객의 데이터를 중앙 서버로 직접 수집하지 않고 각 고객별 제품 내에서 로컬 모델(local model)을 개별적으로 학습한 뒤, 그 결과로 얻어진 모델 파라미터만을 서버에 전송하는 방식으로 동작한다(Yang *et al.*, 2019; Li *et al.*, 2020, 2021). 중앙 서버는 여러 고객들로부터 전달된 로컬 모델 정보를 집계(aggregation)하여 새로운 글로벌 모델(global model)을 생성하고, 이를 다시 각 고객들에게 배포한다. 이러한 과정은 여러 연합학습 라운드(round)에 걸쳐 반복 수행되며, 라운드가 진행될수록 글로벌 모델의 성능이 점진적으로 향상된다. 이러한 구조를 통해 기업은 다양한 고객의 실제 제품 사용 패턴을 글로벌 모델 학습

에 반영할 수 있으며, 동시에 개인정보 보호 요구사항도 충족한다. 그러나 현실 환경에서 연합학습을 적용할 경우, 추가적인 문제에 직면하게 된다. 실제로 일부 고객은 제품 센서 노후화나 부품 고장, 사용자 행동 다양성, 통신 및 소프트웨어 이슈 등 다양한 요인으로 인해 정상 상태와는 다른 이상 고객 데이터(anomalous customer data)를 생성할 수 있다. 이러한 이상 데이터가 로컬 학습에 사용되면, 해당 고객은 왜곡된 파라미터를 서버로 전달하게 되고, 이는 집계 과정에서 글로벌 모델의 품질을 심각하게 저하시킬 수 있다(Bhagoji *et al.*, 2019; Li *et al.*, 2023). 즉, 연합학습이 본래 목적과 달리 오히려 모델 품질을 떨어뜨리는 요인이 될 수 있어, 이를 사전에 감지하고 대응하는 기술이 필요하다.

연합학습에서 이상 고객 데이터가 글로벌 모델의 품질을 저하시킨다는 문제가 알려지면서, 이를 완화하기 위한 몇 가지 강건한 연합학습(robust FL) 기법들이 제안되어 왔다(Cao *et al.*, 2021; Guo *et al.*, 2023). Cao *et al.*(2021)은 중앙 서버가 사전에 보유하고 있는 신뢰 가능 데이터셋(trusted dataset)을 사용하여 각 고객들이 전송한 로컬 모델 업데이트의 방향성(gradient)을 분석하고, 서버 gradient와 유사한 방향을 갖는 고객만을 정상으로 판단하는 FLTrust 기법을 제안하였다. 해당 기법은 서버 기준 정보를 활용한다는 점에서 효과적이나, 실제 냉장고 사용 환경에서는 고객별 온도 분포, 문 열림 빈도, 식재료 적재량 등 데이터 분포의 이질성이 매우 크다. 이로 인해 trusted dataset이 모든 정상 고객의 데이터 특성을 충분히 대표하기 어려우며, 결과적으로 정상 고객임에도 불구하고 이상 고객으로 오분류되는 오탐(false positive) 문제가 발생할 수 있다. 한편, Guo *et al.*(2023)은 오토인코더(autoencoder) 기반 모델 파라미터 복원 오차를 활용하여 이상 고객을 탐지하고, 동시에 인센티브 메커니즘을 결합하여 안정적인 연합학습 프레임워크(FL framework for detecting and incentivizing abnormal clients, FL-DIAC)를 제안하였다. 하지만 FL-DIAC는 모델 파라미터의 이상 패턴만을 분석하기 때문에, 이상 고객 데이터에서 나타나는 데이터 수준의 시계열 패턴 붕괴를 직접적으로 반영하지 못할 수 있다. 나아가 고객 간 데이터 분포 이질성이 심한 실제 냉장고 사용 환경에서는, 정상 고객 간에도 모델 파라미터 차이가 크게 나타나 이상 고객 판단 안정성이 떨어질 수 있다.

이러한 문제점을 해결하기 위해, 본 연구는 데이터 수준(data-level)의 이상 탐지를 연합학습에 통합한 강건 연합학습 프레임워크(data-level reconstruction-based FL, DrFL)를 제안한다. DrFL은 autoencoder 기반 복원 오차(reconstruction error)를 이용하여 고객 데이터가 정상 분포에서 벗어났는지를 평가한다. 이때, autoencoder는 중앙 서버의 trusted dataset 내 정상 고객 데이터를 학습한 모델로, 입출력 데이터 간의 차이인 복원 오차를 계산한다. DrFL은 이 오차 값을 기반으로 각 고객의 이상도(anomaly score)를 산출하고, 이를 사전에 설정된 분위수(quantile) 기준과 비교하여 이상 고객과 정상 고객을 구분한다.

이후 중앙 서버가 정상으로 판단된 고객들로부터 로컬 모델 업데이트만을 수집하여, 고객별 데이터 크기 기반 가중 평균 방식으로 새로운 글로벌 모델을 생성한다. 이는 데이터량이 많은 고객일수록 해당 고객의 로컬 업데이트가 글로벌 모델의 방향성을 더 신뢰성 있게 반영하기 때문이다(McMahan *et al.*, 2023). 제안 방법론의 효과 검증을 위해, 다양한 가정의 냉장고로부터 수집된 온도 시계열, 문 열림 이력 등 실제 운전 로그 데이터를 활용하였다. 실험 결과, DrFL은 기존 연합학습 기법 대비 냉장실 · 냉동실 · 변온실 전 구간에서 예측 오차를 평균 40% 이상 감소시켰다. 본 논문은 국내 냉장고 내부 온도 예측에 딥러닝 기반 robust FL을 체계적으로 시도한 최초의 연구이며, 향후에 이상 고객 데이터가 포함된 현실 상황에서의 스마트 가전 보급에 도움될 것으로 기대된다.

2. 제안 방법

2.1 데이터 수집 및 전처리

본 연구에서는 제안한 DrFL 방법론의 성능을 평가하기 위해 실제 삼성전자 냉장고 제품에서 수집된 센서 데이터를 활용하였다. 해당 데이터는 삼성전자 실험실 환경과 국내 1,011개 가정에서 사용 중인 냉장고로부터 수집되었으며, 28개의 시계열 센서 데이터 변수로 구성된다. 전체 데이터 중 978대의 냉장고는 정상적으로 동작한 제품이며, 나머지 33대는 비정상적인 온도 패턴을 보이는 이상 제품으로 분류된다. 이와 같이 실제 운용 환경에서 발생한 이상 제품 데이터를 활용함으로써, DrFL이 데이터 수준에서 나타나는 구조적 이상 시계열 패턴을 얼마나 효과적으로 탐지할 수 있는지 평가하고자 하였다. 또한, 이러한 이상 고객 데이터를 글로벌 모델 학습 과정에서 적절히 배제함으로써, 전체 연합학습 모델 성능을 안정적으로 유지할 수 있는지를 검증하였다.

본 연구의 온도 예측 모델은 과거 M 분 동안 1분 간격으로 수집된 28개의 시계열 데이터를 입력으로 사용하여, N 분 후의 냉장실 · 냉동실 · 변온실 내부 온도를 예측하는 것을 목표로

한다. 이를 위해 수집된 전체 시계열 데이터셋에 대해 슬라이딩 윈도우(sliding window) 절차를 적용하여 입력-출력 쌍을 생성하였다. 구체적으로, 길이 M 의 입력 윈도우를 1분 단위로 이동시키면서, 각 윈도우에 대응하는 미래 N 분 후의 온도를 예측 타겟(target) 값으로 구성하였다. 이러한 전처리 과정은 모델이 단기적인 온도 변화 패턴을 효과적으로 학습하도록 지원한다. 이때, 냉장고 센서 데이터는 고객별 사용환경 및 센서 특성에 따라 값의 절대 크기나 변동 폭이 다르다. 이러한 비정규화된 시계열 데이터를 그대로 모델이 학습할 경우, 스케일이 큰 변수 또는 특정 고객의 온도 구간이 모델 학습에 과도하게 영향을 미칠 수 있다. 이를 방지하기 위해 본 연구에서는 고객별 표준화(customer-wise standardization)를 적용하였다. 즉, 각 고객 단위로 시계열 데이터의 평균을 0, 표준편차를 1로 정규화함으로써, 모든 입력 변수가 유사한 스케일에서 학습되도록 하였다.

2.2 이상 고객 데이터에 강건한 연합학습 방법 기반 온도 예측 모델 학습

본 연구는 냉장고 시계열 센서 데이터로부터 유의미한 특징을 추출하여 N 분 후 미래 온도를 예측하기 위해 일반적인 1차원 합성곱 신경망(one-dimensional convolutional neural network, 1D-CNN) 구조를 기반으로 모델을 설계하였다. 설계된 모델은 특징 추출기(feature extractor) f_θ 와 예측기(regressor) g_θ 로 구성되며, 전체 모델의 연산 구조는 <Table 1>에 정리되어 있다. 이때, θ 는 설계된 전체 모델의 파라미터를 의미한다. 입력 데이터 x 는 7×28 크기의 다차원 시계열 센서 데이터이며, 출력 데이터 $\hat{y}$ 는 입력 데이터로부터 예측된 N 분 후의 냉장고, 냉동고, 변온실 온도를 의미하며, 3차원 실수 벡터이다.

본 연구의 온도 예측 모델 학습 절차는 사전학습용 데이터와 실제 고객별 로컬 데이터로 구분되는 두 종류의 데이터셋을 기반으로 한다. 먼저, 중앙 서버는 실험실 환경과 소수의 정상 고객으로부터 수집된 trusted dataset을 이용하여 모델을 사전학습(pre-training)한다. 이 데이터셋은 냉장고의 정상 동작

Table 1. Architecture Details of Refrigerator Temperature Prediction Model

Model name	Layer name	Type	Details
f_θ (feature extractor)	Conv.1	1D convolutional layer	Filter size=3, in_channels=28, out_channels=16, and padding=1
	ReLU.1	Activation function	Rectified linear unit function
	Conv.2	1D convolutional layer	Filter size=3, in_channels=16, out_channels=8, and padding=1
	ReLU.2	Activation function	Rectified linear unit function
	Flatten	Flatten layer	Flattening the final two-dimensional feature map into a one-dimensional vector
g_θ (regressor)	Linear	Fully connected layer	Mapping the flattened vector to the three predicted values by applying a linear transformation

상태를 대표하며, 사전학습을 통해 모델이 기본적인 열역학적 동작 패턴을 학습하도록 한다. 이를 통해 고객 간 데이터 분포 차이로 인해 초기 학습이 불안정해지는 문제를 완화할 수 있다. 사전학습은 다음 최적화 문제를 통해 수행된다:

$$\theta^{(0)} = \argmin_{\theta} \frac{1}{L} \sum_{l=1}^L (g_{\theta}(f_{\theta}(x^{(l)})) - y^{(l)})^2. \quad (1)$$

사전학습에 사용되는 trusted dataset은 $D_{lab} = \{(x^{(l)}, y^{(l)}) \mid l = 1, 2, \dots, L\}$ 으로 정의된다. 여기서 $x^{(l)} \in R^{7 \times 28}$ 은 실험실 및 소수의 정상 고객 사용 환경에서 수집된 시계열 입력 데이터이며, $y^{(l)} \in R^3$ 은 N분 후 온도를 나타내는 출력 데이터이다. 또한 L 은 trusted dataset의 데이터 개수를 의미한다. 식 (1)에서 θ 는 전체 모델의 학습 파라미터를 나타내며, f_{θ} 와, g_{θ} 는 각각 모델의 특징 추출기와 예측기를 의미한다. 본 연구에서는 확률적 미니 배치 경사 하강법(stochastic mini-batch gradient descent, SGD)을 이용하여 최적화를 진행하였다.

사전학습이 완료된 초기 모델 파라미터 $\theta^{(0)}$ 는 모든 고객에게 배포되며, 이를 기반으로 본격적인 연합학습이 시작된다. 이후 각 연합학습 라운드에서 서버는 전체 고객 중 일부를 무작위로 선택하여 학습에 참여시키며, 선택된 고객 $i \in C^{(t)} \subseteq \{1, \dots, N\}$ 은 서버로부터 전달 받은 글로벌 모델을 초기값으로 하여 자신의 로컬 데이터셋을 이용해 독립적인 학습을 수행한다. 이러한 로컬 학습은 다음과 같이 정의된다:

$$\theta_i^{(t)} = \argmin_{\theta} \frac{1}{n_i} \sum_{k=1}^{n_i} (g_{\theta}(f_{\theta}(x_i^{(k)})) - y_i^{(k)})^2. \quad (2)$$

고객 i 가 학습하는 로컬 데이터셋은 $D_i = \{(x_i^{(k)}, y_i^{(k)}) \mid k = 1, \dots, n_i\}$ 으로 정의된다. 여기서 $x_i^{(k)} \in R^{7 \times 28}$ 는 고객 i 로부터 수집된 입력 시계열 데이터이며, $y_i^{(k)} \in R^3$ 는 해당 시점으로부터 N분 후 미래 온도값을 나타낸다, 또한 n_i 는 고객 i 의 데이터 개수를 의미한다. 식 (2)에서 $\theta_i^{(t)}$ 는 고객 i 가 연합학습 라운드 t 에서 중앙 서버로 전송하는 모델 파라미터를 의미하

며 f_{θ} 와, g_{θ} 는 각각 입력 시계열 데이터로부터 잠재 표현을 추출하는 특징 추출 함수와 미래 온도를 예측하는 회귀 함수를 의미한다. 그러나 연합학습에 참여 고객들 중 일부는 이상 데이터로 인해 정상적이지 않은 시계열 데이터 패턴을 학습할 수 있으며, 이러한 왜곡된 로컬 모델 업데이트가 집계 과정에 포함될 경우 글로벌 모델의 품질이 크게 저하될 수 있다.

이에 따라 본 연구에서는 로컬 모델의 파라미터를 중앙 서버에서 집계하기 전에 고객 데이터의 품질을 평가하는 data-level의 이상 탐지 절차를 도입하였다. 이를 위해 중앙 서버는 서버 자체가 사전에 확보한 trusted dataset, D_{lab} 을 활용하여 autoencoder를 사전 학습한다. Autoencoder는 인코더(encoder) h_{ϕ} 와 디코더(decoder) d_{ϕ} 로 구성되며, ϕ 는 학습에 의해 결정되는 모델 파라미터이다. Autoencoder의 자세한 모델 구조는 <Table 2>와 같다. Autoencoder는 입력 시계열 $x^{(l)}$ 에 대한 복원값 $\hat{x}^{(l)} = d_{\phi}(h_{\phi}(x^{(l)}))$ 을 기반으로 복원 오차를 최소화하도록 학습되며, 최적화 과정은 다음과 같이 정의된다:

$$\phi^* = \argmin_{\phi} \frac{1}{L} \sum_{l=1}^L \|d_{\phi}(h_{\phi}(x^{(l)})) - x^{(l)}\|^2. \quad (3)$$

여기서 ϕ 는 autoencoder의 전체 파라미터, h_{ϕ} 와 d_{ϕ} 는 각각 autoencoder의 인코더와 디코더, L 은 trusted dataset의 데이터 개수를 의미한다. 본 연구에서는 이 최적화 과정 또한 SGD 기법을 이용하여 수행하였다. Trusted dataset은 이상 패턴이 포함되지 않은 정상 데이터로 구성되어 있으므로, autoencoder는 정상적인 냉장고 시계열 패턴의 구조적 특성을 효과적으로 학습하게 된다. 이로 인해 autoencoder는 연합학습 과정에서 고객 데이터가 정상 분포에서 얼마나 벗어나는지를 평가하는 신뢰도 높은 기준 모델로 기능한다. Autoencoder가 사전 학습된 이후, 중앙 서버는 참여 고객 $i \in C^{(t)}$ 의 로컬 데이터에 대해 복원 오차를 계산함으로써 데이터 품질을 평가한다. 고객 i 의 anomaly score는 다음과 같이 정의된다:

$$r_i = \frac{1}{n_i} \sum_{k=1}^{n_i} \|d_{\phi}(h_{\phi}(x_i^{(k)})) - x_i^{(k)}\|^2. \quad (4)$$

Table 2. Architecture Details of Autoencoder Model

Model name	Layer name	Type	Details
h_{ϕ} (encoder)	LSTM.1	LSTM layer	Input_size=28, hidden_size=64, num_layers=1, bidirectional=false, and batch_first=true
	Last-step selection	Operation	Select the encoder's final time-step representation
Bottleneck	Repeat vector	Operation	Repeat this final representation along the temporal dimension to form the decoder input sequence
	LSTM.2	LSTM	Input_size=64, hidden_size=64, num_layers=1, bidirectional=false, and batch_first=true
d_{ϕ} (decoder)	Linear	Fully connected layer	In_features=64, out_features=28 (time-step-wise projection)

여기서 h_ϕ 와 d_ϕ 는 각각 autoencoder의 인코더와 디코더, n_i 는 고객 i 의 데이터 개수를 의미한다. 중앙 서버는 계산된 anomaly score를 사전에 설정된 분위수 기반 임계값 τ 와 비교하여 정상 고객 집합 $N^{(t)} = \{i \in C^{(t)} | r_i \leq \tau\}$ 을 결정한다. 이후 서버는 정상 고객들로부터 전달된 로컬 모델 파라미터만을 활용하여 새로운 글로벌 모델을 갱신한다. 고객별 데이터 크기에 비례한 가중치를 적용하여 다음과 같이 새로운 글로벌 모델 $\theta_i^{(t+1)}$ 을 계산한다:

$$\theta^{(t+1)} = \sum_{i \in N^{(t)}} \frac{n_i}{\sum_{j \in N^{(t)}} n_j} \theta_i^{(t)}. \quad (5)$$

업데이트된 글로벌 모델 파라미터 $\theta^{(t+1)}$ 는 다시 모든 고객에게 전달되며, 동일한 절차가 반복됨에 따라 모델은 점진적으로 수렴하고 성능이 안정적으로 향상된다. 이와 같이 제안된 DrFL은 부분 고객 샘플링, 로컬 학습, 데이터 수준 이상 탐지, 정상 고객 기반 글로벌 집계로 구성된 순환 구조를 가지며, 전체 알고리즘은 Algorithm 1에 요약되어 있다. 이러한 설계는 실제 냉장고 사용 환경에서 발생하는 센서 오류와 데이터 분포 이질성에 효과적으로 대응할 수 있으며, 결과적으로 높은 품질의 글로벌 온도 예측 모델을 구축하는 데 기여한다.

Algorithm 1: DrFL

Require: Trusted lab dataset D_{lab} , customer local datasets $\{D_i\}_{i=1}^N$, global rounds T , client sampling ratio ρ , local epochs E , learning rate η , pretrain epochs E_0 , autoencoder pretrain epochs E_{ae} , and quantile level q

Step 0: Server pretraining of refrigerator temperature prediction model on trusted dataset

Initialize θ (feature extractor f_θ and regressor g_θ)

for $e_0 = 1$ to E_0 do

Update θ by SGD on D_{lab} for solving Equation (1)

end for

$\theta^{(0)} \leftarrow \theta$

Step 1: Server pretraining of autoencoder on trusted dataset

Initialize ϕ (encoder h_ϕ and decoder d_ϕ)

for $e_{ae} = 1$ to E_{ae} do

Update ϕ by SGD on D_{lab} for solving Equation (3)

end for

$\phi^* \leftarrow \phi$

Step 2: Federated training with data-level anomaly detection

for $t = 1$ to T do

Sample $C^{(t)} \subset \{1, \dots, N\}$ with $|C^{(t)}| = \lceil \rho N \rceil$.

Broadcast $\theta^{(t)}$ to all $i \in C^{(t)}$.

for each $i \in C^{(t)}$ in parallel do

$\theta_i^{(t)} \leftarrow \theta^{(t)}$.

for $e = 1$ to E do

Update $\theta_i^{(t)}$ by SGD on D_i

for solving Equation (2)

end for

Compute anomaly score r_i using ϕ^* and

Equation (4)

Upload $\theta_i^{(t)}$ and r_i to the server

end for

$\tau \leftarrow \text{Quantile}_q(\{r_i\}_{i \in C^{(t)}})$

$N^{(t)} \leftarrow \{i \in C^{(t)} \mid r_i \leq \tau\}$

Update $\theta^{(t+1)}$ using Equation (5)

end for

Return: $\theta^{(T+1)}$

3. 성능 검증

3.1 실험 셋팅

본 연구는 제안 방법론인 DrFL의 성능 검증을 위해 사전학습용 데이터, 연합학습 참여 고객 데이터, 신규 고객 평가용 데이터셋을 사용하였다. 사전학습용 데이터는 trusted dataset을 의미하며, 실험실 환경에서 수집한 데이터와 실제 정상 고객 32명이 사용한 데이터로 구성된다. 연합학습 참여 고객 데이터는 정상 냉장고 846대와 이상 냉장고 33대를 포함한 총 879대로 구성되며, 각 고객의 로컬 모델 학습에 활용된다. 다만, 실제 이상 냉장고 33대만으로는 현실 세계의 다양한 이상 상황을 반영하기 어렵다고 판단하여, 삼성전자 실무진의 자문을 통해 실제 냉장고 센서의 이상 패턴을 파악하고, 연합학습 참여 879대 중 20%를 무작위로 선택하여 특정 센서 값에 이상 패턴을 인위적으로 주입하였다. 구체적으로, 이상 패턴은 센서 통신 단절로 인해 같은 값이 계속 반복되는 경우와 일시적으로 센서 값이 기록되지 않는 경우 두 가지 유형으로 구성되며, 각각 50% 확률로 2~3 타임스텝(timestep) 구간에 무작위로 주입하였다. 연합학습 참여 고객 데이터는 각 고객 별로 전체 데이터의 20%를 평가용으로 분리한 후, 나머지 80% 중 20%를 검증용, 나머지를 모두 학습용으로 사용하였다. 단, 실제 이상 냉장고 33대의 데이터와 이상 패턴이 주입된 시뮬레이션 이상 데이터의 경우 학습용으로만 활용하였다. 신규 고객 평가용 데이터는 연합학습에 참여하지 않는 냉장고 100대로 구성되며, 모델의 일반화 성능을 평가하는 데 사용하였다. 해당 데이터는 모델 학습 과정에 포함되지 않으므로, 별도의 분할 없이

모두 평가용으로만 활용하였다.

이와 같은 데이터 구성 및 분할을 바탕으로, 본 연구는 두 가지 평가 시나리오를 설정하였다. 첫 번째는 연합학습 참여 고객 평가 시나리오로, 제안 방법론이 연합학습 과정에 직접적으로 참여한 고객들의 냉장고 온도 예측 성능을 효과적으로 향상시킬 수 있는지를 검증한다. 두 번째는 신규 고객 평가 시나리오로 실제 서비스 환경에서 신규 고객이 추가되었을 때, 별도의 추가 학습 없이도 학습된 글로벌 모델을 즉시 적용할 수 있는지를 확인하기 위함이다. 각 평가 시나리오에서는 사전학습 단계에서 얻어진 초기 모델 $g_{\theta^{(0)}} \circ f_{\theta^{(0)}}$, 제안 방법론으로 얻어진 모델 $g_{\theta^{(T+1)}} \circ f_{\theta^{(T+1)}}$, 그리고 대표적인 연합학습 기법 FedAvg(McMahan *et al.*, 2023), 고객 데이터간 이질성을 완화한 FedProx(Li *et al.*, 2020), 강건 연합학습 기법인 FLTrust(Cao *et al.*, 2021) 및 FL-DIAC(Guo *et al.*, 2023)을 적용하여 학습된 모델들의 예측 성능을 비교 분석하였다. 이때 공정한 비교를 위해 모든 연합학습 방법론들은 <Table 1>에서 제시된 1D-CNN 기반 온도 예측 모델을 공통적으로 사용하였다. 본 연구에서 예측 대상으로 설정한 온도 변수는 냉동실 온도(freezer temperature), 냉장실 온도(refrigerator temperature), 변온실 온도(convertible compartment temperature)이며, Y1, Y2, Y3로 표기하였다. 2.1절에서 언급한 입력 윈도우 길이 M과 예측 시점 N은 각각 7분, 5분으로 설정하였으며, 이는 냉장고의 압축기가 구동되어 내부 온도를 낮추는 데까지 평균적으로 걸리는 시간을 반영한 것이다.

모델 예측 성능은 회귀 분석 기반 예측에서 널리 사용되는 평가지표들을 활용하였다. 첫 번째 평가지표는 결정계수(coefficient of determination, R^2)로, 모델이 실제 온도 변동을 얼마나 잘 설명하는지를 나타내며, 모델의 전반적인 설명력을 평가하는 데 사용하였다. 두 번째 평가지표는 평균 절대 백분율 오차(mean absolute percentage error, MAPE)로, 예측값과 실제값의 차이를 실제값 대비 백분율로 나타낸 지표이다. 다만, 변온실 온도인 Y3의 경우 실제값이 0에 가까운 샘플이 다수 존재하여 MAPE 값이 과도하게 커지는 왜곡이 발생하므로, 해당 출력 변수에 대해서는 MAPE를 평가 지표에서 제외하였다. 세 번째 평가지표는 평균 제곱근 오차(root mean squared error, RMSE)로, 예측값과 실제값 차이의 제곱 평균에 제곱근을 취한 값이며, 오차를 제곱하여 계산하므로 큰 오차에 더 민감하게 커지는 특성이 있다. 본 연구에서 사용한 평가 지표 중 R^2 는 1에 가까울수록, MAPE와 RMSE는 값이 낮을수록 우수한 예측 성능을 나타낸다. 한편, 연합학습 환경에서는 각 고객이 보유한 데이터의 분포, 양, 품질이 서로 상이한 특성을 가지므로, 글로벌 모델의 평균적인 예측 성능만으로는 실제 서비스 품질을 충분히 평가하기 어렵다. 평균 성능이 우수하더라도, 특정 고객에 대해서는 예측 성능이 크게 저하될 수 있기 때문이다. 이에 본 연구에서는 고객별 예측 성능의 표준편차(standard deviation across customers)를 추가적인 평가지표로 활용하였다. 이는 연합학습에 참여한 각 고객의 R^2 , MAPE, RMSE 값을 기

준으로 계산되며, 값이 낮을수록 글로벌 모델이 고객 간 데이터 이질성에도 불구하고 안정적인 예측 결과를 제공함을 의미한다. 실험에 사용한 구체적인 하이퍼파라미터 값들은 <Table 3>에 정리되어 있으며, 모든 실험은 5회 반복 수행하였고, 그 평균값을 최종 성능 지표로 제시하였다.

Table 3. Hyperparameter Details Used in this Study

Method	Hyperparameter	Notation	Value
Shared	Customer sampling ratio	ρ	0.1
	Prediction model pretrain epochs	E_0	50
	Local learning rate	η	0.0005
	Local training epochs	E	1
FedAvg	Local batch size	-	10
	Global communication rounds	T	20
Fedprox	Proximal term weight	-	0.01
	Global communication rounds	T	20
FLTrust	Server learning rate	η_{server}	1
	Anchor dataset size	-	300
	Local batch size	-	32
	Global communication rounds	T	100
FL-DIAC	Anomaly threshold quantile	q	0.1
	Autoencoder pretrain epochs	E_{ae}	20
	Autoencoder batch size	-	16
	Global communication rounds	T	20
DrFL (proposed)	Autoencoder pretrain epochs	E_{ae}	80
	Autoencoder learning rate	η_{ae}	0.0005
	Autoencoder batch size	-	64
	Anomaly threshold quantile	q	0.25
	Global communication rounds	T	20

3.2 온도 예측 결과

<Table 4>는 이상 고객 데이터가 포함된 환경에서 연합학습에 참여한 고객을 대상으로 한 온도 예측 성능 결과를 보여준다. 실험 결과, DrFL은 모든 온도 변수에 대해 비교 방법론 대비 가장 우수한 예측 성능을 달성하였으며, 특히 냉장실 온도(Y2) 예측에서 그 효과가 가장 두드러졌다. 구체적으로, 기본 연합학습 알고리즘인 FedAvg와 FedProx의 R^2 가 약 0.95 수준에 머무른 반면, DrFL은 0.99를 기록하며 우수한 예측 성능 향상을 보였다. 동시에 오차 지표인 RMSE는 60% 이상 감소하였는데, 이는 DrFL이 단순한 평균적 경향성을 넘어 실제 온도 변동의 미세한 패턴까지 정밀하게 포착하고 있음을 의미한다. 이러한 성능 우위는 냉장실(Y1)과 변온실(Y3) 온도 변수에서도 일관되게 관찰되었으며, DrFL이 이상 고객 데이터에 포함된 왜곡된 시계열 패턴을 효과적으로 배제하고, 정상 데이터에 내재된 열역학적 특성을 글로벌 모델에 안정적으로 반영했음을 보여준다.

기존 강건 연합학습 기법과의 비교에서도 DrFL의 우수성이 확인되며, 각 기법이 갖는 한계는 <Figure 1>에 제시된 이상치 점수 분포를 통해 직관적으로 파악할 수 있다. 먼저, FLTrust는 pretrain 모델에 비해 성능이 저하되는 결과를 보였는데, 이는 FLTrust의 이상 탐지 방식 고객 간 이질성이 큰 냉장고 사용 환경에서는 효과적으로 작동하지 못했음을 의미한다. FLTrust는 서버에 저장된 trusted dataset을 기준으로 서버 gradient를 계산하고, 각 고객의 로컬 업데이트가 해당 gradient와 유사한 방향을 가지는지를 기준으로 이상 여부를 판단한다. 그러나 고객 간 사용 행태가 다르고 데이터 분포 이질성이 큰 냉장고 데이터 환경에서는, 정상 고객이라 하더라도 서버 gradient와 다른 방향을 가질 수 있다. 이로 인해 FLTrust는 다수의 정상 고객을 이상 고객으로 오판하여 학습 과정에서 배제하게 되며, 결과적으로 글로벌 모델이 사전학습 성능을 넘어 서는 데 실패하게 된다. 이러한 현상은 <Figure 1(a)>에서 FLTrust의 정상 고객과 이상 고객의 이상치 점수 분포가 상당 부분 중첩되어 나타나는 것으로 확인할 수 있으며, FLTrust의 구조적 한계를 시각적으로 뒷받침한다.

한편, autoencoder 기반 이상 탐지를 활용하는 FL-DIAC는

FedAvg와 거의 동일한 성능을 보이는 데 그쳤다. 이는 FL-DIAC가 모델 파라미터 수준에서 이상을 탐지하는 방식이기 때문에, 이상 고객 데이터에서 나타나는 미세한 시계열 패턴 변화를 직접적으로 반영하지 못하기 때문이다. 실제로 <Figure 1(b)>에서도 정상 고객과 이상 고객의 분포가 상당 부분 겹치는 양상을 보이며, 이로 인해 이상 고객이 정상으로 오판되어 글로벌 모델 집계 과정에 포함될 가능성이 있음을 확인할 수 있다. 반면, 제안한 DrFL은 데이터 수준에서 직접적으로 이상을 감지하고 차단하는 방식을 활용함으로써, <Figure 1(c)>에서 나타난 바와 같이 정상 고객과 이상 고객을 높은 정확도로 구분한다. 이러한 정밀한 고객 필터링은 연합학습 집계 단계에서 왜곡된 업데이트의 영향을 효과적으로 차단하며, RMSE 기준 평균 70% 이상 예측 오차 감소라는 뚜렷한 성능 개선으로 이어졌다.

<Table 5>은 연합학습에 참여하지 않은 신규 고객을 대상으로 한 일반화 성능 평가 결과를 보여준다. 해당 시나리오에서도 DrFL은 전반적으로 우수한 성능을 유지하였으며, 이는 학습된 글로벌 모델이 실제 서비스 환경에 배포되었을 때도 안정적인 성능을 제공할 수 있음을 의미한다.

Table 4. Performance comparison of temperature prediction models on participating customers in FL using , MAPE, and RMSE. The best performance among compared models is highlighted in **bold**. Values in parentheses represent the standard deviation across customers. CF denotes the convertible compartment.

Method	Y1 (freezer temperature)			Y2 (refrigerator temperature)			Y3 (CF temperature)	
	R^2	MAPE	RMSE	R^2	MAPE	RMSE	R^2	RMSE
Pretrain	0.93 (0.15)	12.61 (8.97)	2.00 (0.78)	0.94 (0.14)	21.70 (15.48)	0.52 (0.31)	0.96 (0.14)	1.75 (0.76)
FedAvg	0.94 (0.12)	10.14 (10.62)	2.13 (0.73)	0.95 (0.11)	19.47 (14.69)	0.50 (0.25)	0.95 (0.12)	1.76 (0.11)
FedProx	0.95 (0.10)	8.46 (10.27)	1.70 (2.27)	0.96 (0.10)	18.38 (12.43)	0.49 (0.21)	0.95 (0.06)	1.75 (0.53)
FLTrust	0.91 (0.17)	13.80 (9.45)	2.10 (0.87)	0.93 (0.12)	25.18 (12.83)	0.62 (0.26)	0.91 (0.10)	2.27 (1.10)
FL-DIAC	0.97 (0.10)	7.17 (7.35)	1.06 (0.38)	0.96 (0.09)	17.42 (13.82)	0.45 (0.17)	0.95 (0.08)	1.74 (0.63)
DrFL (proposed)	0.99 (0.08)	6.89 (6.73)	0.78 (0.35)	0.97 (0.06)	12.74 (11.83)	0.39 (0.14)	0.99 (0.08)	0.65 (0.53)

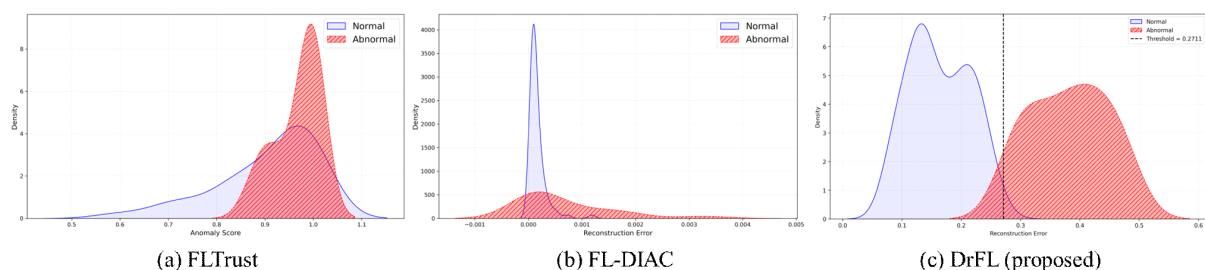


Figure 1. Anomaly Score Distributions of Robust FL Methods for Normal (blue) and Abnormal (red) Customers.

Table 5. Performance Comparison of Temperature Prediction Models on Nonparticipating Customers in FL Using R^2 , MAPE, and RMSE. The best performance among compared models is highlighted in **bold**. Values in parentheses represent the standard deviation across customers. CF denotes the convertible compartment.

Method	Y1 (freezer temperature)			Y2 (refrigerator temperature)			Y3 (CF temperature)	
	R^2	MAPE	RMSE	R^2	MAPE	RMSE	R^2	RMSE
Pretrain	-0.03 (19.84)	12.04 (7.32)	4.66 (5.02)	0.74 (0.57)	32.93 (21.79)	0.89 (0.31)	0.87 (1.05)	3.65 (2.77)
FedAvg	0.92 (0.12)	8.46 (5.13)	1.33 (0.54)	0.91 (0.16)	18.88 (11.82)	0.51 (0.14)	0.97 (0.08)	1.81 (0.80)
FedProx	0.93 (0.12)	7.62 (4.68)	1.19 (0.49)	0.91 (0.18)	20.16 (13.05)	0.53 (0.14)	0.97 (0.06)	1.71 (0.75)
FLTrust	0.31 (22.73)	10.95 (5.50)	4.77 (5.44)	0.86 (11.20)	30.38 (21.43)	0.62 (0.21)	0.85 (6.26)	3.54 (2.53)
FL-DIAC	0.94 (0.13)	7.04 (4.68)	1.09 (0.51)	0.91 (0.17)	19.78 (12.95)	0.51 (0.15)	0.97 (0.08)	1.83 (0.82)
DrFL (proposed)	0.96 (0.27)	6.20 (2.66)	1.05 (0.55)	0.93 (0.23)	17.79 (11.63)	0.47 (0.15)	0.97 (0.06)	1.72 (0.78)

이러한 안정적인 성능은 DrFL이 이상 데이터에 편향되지 않고 다양한 고객 환경에서 공통적으로 적용되는 냉장고의 본질적인 온도 변화 메커니즘을 효과적으로 학습했음을 의미한다. 반면, 사전학습 단계에서만 학습된 초기 모델 $g_{\theta^{(0)}}$ o $f_{\theta^{(0)}}$ 은 신규 고객 환경에서 현저한 성능 저하를 보였다. 특히 냉동실 온도(Y1)에 대해 R^2 가 음수인 -0.03을 기록하였고, 이는 제한된 실험실 환경에서 수집된 데이터만을 이용한 사전학습 모델만으로는 충분한 일반화 성능을 확보하기 어렵다는 점을 보여준다. FLTrust 또한 신규 고객 시나리오에서 저조한 성능을 보였는데, 이는 gradient 기반 비교 방식으로 인해 정상 고객을 과도하게 배제하여 연합학습 과정에서 글로벌 모델이 유의미하게 개선되지 못했기 때문이다. <Table 4>와 <Table 5>에서 모두 일관되게 우수한 성능을 갖는 방법론은 DrFL이 유일하였으며, 이러한 결과는 이상 고객 데이터가 존재하는 현실적인 환경에서 DrFL이 실제 서비스 적용에 가장 안정적인 연합학습 프레임워크임을 의미한다.

3.3 이상 고객 탐지 결과

본 절에서는 DrFL의 온도 예측 성능 개선의 핵심 요인인 이상 고객 탐지의 효과를 구체적으로 확인하고자 한다. 먼저 이상 고객 탐지 성능 평가를 위해서는 area under the receiver operating characteristic curve(AUROC)를 주요 지표로 사용하였으며, AUROC는 1에 가까울수록 정상 고객과 이상 고객을 안정적으로 구분할 수 있음을 의미한다(Hanley and McNeil, 1982). 먼저, FLTrust는 AUROC 0.705를 기록하였다. 이는 일정 수준의 구분 능력은 확보하고 있으나, 정상 고객과 이상 고객 간 경계가 충분히 명확하지 않음을 의미한다. 특히 gradient

기반 이상 탐지 방식은 고객 간 데이터 분포 이질성이 큰 냉장고 사용 환경에서 정상과 이상을 일관되게 구분하는 데 한계를 보인다. 한편, FL-DIAC는 AUROC 0.675로 세 기법 중 가장 낮은 값을 기록하였다. 이는 분류 임계값 전반에 걸쳐 정상 고객과 이상 고객을 구분하는 판별 능력이 부족함을 나타낸다. 이러한 결과는 모델 파라미터 수준에서의 복원 오차가 데이터 수준의 시계열 구조 붕괴를 충분히 반영하지 못하여, 이상 고객임에도 정상 고객과 유사한 패턴을 보이는 경우를 효과적으로 식별하지 못하기 때문으로 해석할 수 있다. 반면, 제한한 DrFL은 AUROC 0.999를 기록하며 거의 완벽에 가까운 판별 성능을 보였다. 이는 DrFL이 산출하는 이상치 점수가 특정 임계값 설정에 크게 의존하지 않으면서도, 정상 고객과 이상 고객을 매우 안정적으로 분리할 수 있음을 시사한다. 이러한 결과는 DrFL의 이상 고객 탐지 메커니즘이 데이터 분포 이질성과 시계열 구조 변화를 효과적으로 반영함으로써, 기존 방법 대비 현저히 향상된 판별 성능을 달성했음을 정량적으로 입증한다. 추가적으로, DrFL은 이상치 점수에서 상위 25%에 해당하는 고객을 이상 고객으로 분류하였으며, 이때 모델이 이상 고객으로 판별한 고객 중 실제 이상 고객 비율은 97.0%였고, 실제 이상 고객 중 99.8%를 정확히 탐지하였다. 이러한 결과는 DrFL이 이상 고객을 거의 놓치지 않으면서도 정확히 판별함을 정량적으로 입증한다.

이상 고객 탐지의 효과는 연합학습의 수렴 과정에도 직접적인 영향을 미친다. <Figure 2>는 세 가지 온도 변수 각각에 대한 R^2 값을 연합학습 라운드에 따라 비교한 수렴 곡선을 보여준다. 이상 고객 탐지를 적용하지 않은 경우(w/o anomaly detection)와 DrFL의 데이터 수준 탐지를 적용한 경우(w/ anomaly detection)를 비교하였다. 세 가지 온도 변수 모두에서, 이상 고객 탐지를 적용한 경우가 적용하지 않은 경우에 비해 일관되

게 높은 R^2 값을 달성하며, 라운드가 진행될수록 두 조건 간의 성능 격차가 점진적으로 확대되는 공통적 경향이 관찰된다. 특히 냉장실 온도(Y1)의 경우, 초기 라운드에서 두 조건의 R^2 값이 0.93으로 동일하나, 라운드 10 부근에서 약 0.98로 빠르게 수렴하는 반면, 미적용 시에는 0.93 부근에서 정체되며 오히려 라운드 후반부에서 성능이 소폭 하락하는 불안정한 양상을 보인다. 뚜렷한 성능 격차가 발생한다. 이러한 수렴 양상의 차이는 이상 고객 데이터가 글로벌 모델 집계 과정에 지속적으로 포함될 경우, 왜곡된 로컬 업데이트가 누적되어 모델의 수렴 속도와 최종 성능을 모두 저해할 뿐만 아니라, 학습 과정의 안정성 자체를 훼손함을 보여준다. 반면 제안한 DrFL의 이상 고객 필터링은 매 라운드마다 정상 고객의 업데이트만을 선별적으로 집계함으로써, 세 가지 온도 변수 모두에서 글로벌 모델이 빠르고 안정적으로 수렴할 수 있도록 지원한다.

나아가, <Figure 3>에서는 이상 고객 탐지가 개별 고객 수준의 예측 성능 분포에 미치는 영향을 보여준다. <Figure 3>은 연합학습 참여 고객들의 Y1, Y2, Y3 예측 성능을 이상 고객 탐지 적용 여부(w/o anomaly detection vs. w/ anomaly detection)에 따

라 R^2 지표로 비교한 시각화 결과이다. 세 가지 온도 변수 모두에서, 이상 고객 탐지를 적용한 경우 R^2 의 중앙값이 뚜렷하게 상승하고, 사분위 범위가 현저히 좁아지는 공통적 양상이 관찰된다. 이는 이상 고객 필터링을 통해 대다수 고객에서 안정적으로 높은 예측 성능이 달성됨을 의미한다. 특히 주목할 점은 이상치 변화인데, 이상 고객 탐지를 적용하지 않은 경우, 세 변수 모두에서 R^2 가 0.60~0.75 이하로 떨어지는 극단적 이상치가 관찰되며, 이는 이상 고객 데이터에 의해 글로벌 모델이 왜곡됨에 따라 일부 정상 고객에서도 예측 품질이 심각하게 저하될 수 있음을 보여준다. 반면, 이상 고객 탐지를 적용한 경우 이러한 극단적 이상치가 크게 줄어들고, 대부분의 고객이 0.90 이상의 높은 R^2 를 달성하는 것을 확인할 수 있다. 즉, 이상 고객 탐지 적용 시 고객 간 예측 성능 편차가 줄어드는 것이며, 이는 글로벌 모델이 특정 고객 환경에 편향되지 않고 다양한 사용 패턴에 대해 고르게 높은 성능을 제공할 수 있음을 나타낸다. 이러한 고객 간 예측 성능 편차 개선은 실제 서비스 배포 시 모든 고객에게 균일한 품질의 온도 예측 서비스를 제공할 수 있다는 점에서 실용적으로도 중요한 의미를 가진다.

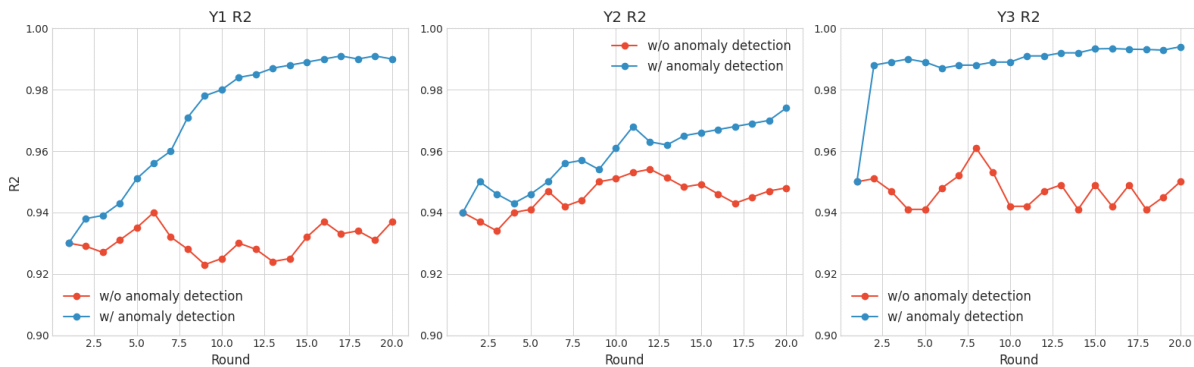


Figure 2. Convergence curves of the global model with and without anomaly detection over communication rounds. R^2 scores are reported for Y1 (freezer temperature), Y2 (refrigerator temperature), and Y3 (convertible compartment temperature). “w/ anomaly detection” denotes the proposed DrFL framework, while “w/o anomaly detection” denotes the same federated learning without anomaly detection.

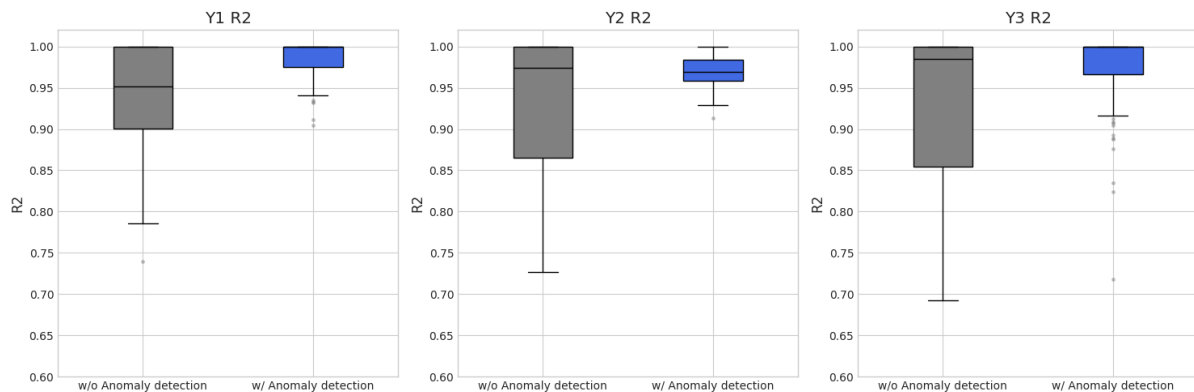


Figure 3. Box plots for demonstrating client-wise R^2 distributions with and without anomaly detection for Y1 (freezer temperature), Y2 (refrigerator temperature), and Y3 (convertible compartment temperature).

4. 결 론

본 연구는 autoencoder 기반 데이터 수준 복원 오차를 활용하여, 이상 고객 데이터가 포함된 현실적인 연합학습 환경에서도 우수한 냉장고 온도 예측 모델을 학습시킬 수 있는 새로운 연합학습 프레임워크를 제시하였다. 제한한 DrFL은 다른 robust FL 기법들과의 비교 실험을 통해, 이상 고객으로 인한 연합학습 성능 저하를 가장 효과적으로 완화할 수 있음을 확인하였다. 또한, 온도 예측 모델의 성능 검증은, 연합학습에 직접적으로 참여하고 있는 기존 고객과, 학습 과정에 참여한 적 없는 신규 고객들을 구분 지어서 진행하였다. 이는 제품 사용 기간에 따라 로컬 학습 데이터가 충분히 축적된 고객뿐만 아니라, 제품을 갓 구매한 신규 고객 환경에서도 제안 방법론이 안정적인 예측 성능을 제공할 수 있는지를 검증하기 위함이다. 두 가지 평가 시나리오에 대한 실험 결과를 종합적으로 분석한 결과, 제한한 DrFL은 모든 경우에서 전반적으로 가장 우수한 예측 성능과 높은 안정성을 보였다. 이러한 결과는 본 연구에서 제시한 방법론이 이상 고객 데이터가 필연적으로 존재하는 실제 냉장고 사용 환경에서도 높은 실용성을 가질 수 있음을 시사한다. 나아가 본 연구의 결과는 향후 대한민국 스마트 냉장고 지능화 및 연합학습 기반 예측 모델 개발 분야에서 실질적인 기술적 기반으로 활용될 수 있을 것으로 기대된다. 다만, 이상 고객 탐지를 위한 분위수 기반 임계값이 사전에 고정된 하이퍼파라미터로 설정되어 있어 실제 운용 환경에서 이상 고객 비율이 변동할 경우 최적의 성능을 보장하기 어려운 한계가 있다. 따라서 이를 해결하기 위해 향후 연구에서는 이상 고객 비율 변화에 적응적으로 대응할 수 있는 동적 임계값 설정 기법을 개발하고자 한다. 나아가 에어컨, 세탁기 등 다른 스마트 가전에서 본 프레임워크를 적용함으로써, 제안 알고리즘의 범용성과 안정성을 검증하고자 한다.

참고문헌

- Brand, D., DiGennaro Reed, F. D., Morley, M. D., Erath, T. G., and Novak, M. D. (2020), A Survey Assessing Privacy Concerns of Smart-Home Services Provided to Individuals with Disabilities, *Behavior Analysis in Practice*, **13**(1), 11-21.
- Cao, X., Fang, M., Liu, J., and Gong, N. Z. (2021), FLTrust: Byzantine-robust Federated Learning via Trust Bootstrapping, *28th Annual Network and Distributed System Security Symposium, NDSS*.
- Guo, H., Mao, Y., He, X., Zhang, B., Pang, T., and Ping, P. (2023), Improving Federated Learning through Abnormal Client Detection and Incentive, *Computer Modeling in Engineering and Sciences*, **139**(1), 383-403.
- Hanley, J. A. and McNeil, B. J. (1982), The meaning and use of the area under a receiver operating characteristic (ROC) curve, *Radiology*, **143**(1), 29-36.
- Li, H., Li, J., Guan, X., Liang, B., Lai, Y., and Luo, X. (2019), Research on Overfitting of Deep Learning, *Proceedings of 15th International*

- Conference on Computational Intelligence and Security, CIS*.
- Li, Q., Wen, Z., Wu, Z., Hu, S., Wang, N., Li, Y., Liu, X., and He, B. (2021), A Survey on Federated Learning Systems: Vision, Hype and Reality for Data Privacy and Protection.
- Li, T., Sahu, A. K., Zaheer, M., Sanjabi, M., Talwalkar, A., and Smith, V. (2020), Federated Optimization in Heterogeneous Networks, *Proceedings of Machine Learning and Systems*, **2**, 429-450.
- Machacuay, J., Manrique-Silupu, J., and Ipanaque, W. (2025), LSTM with Shallow NNs for Indoor Temperature Long-term Predictions in Refrigeration Systems, *Neural Computing and Applications*, **37**(2).
- McMahan, H. B., Moore, E., Ramage, D., Hampson, S., and Arcas, B. A. Y. (2023), Communication-Efficient Learning of Deep Networks from Decentralized Data, *In Artificial Intelligence and Statistics*, 1273-1282.
- Mtibaa, F., Nguyen, K. K., Azam, M., Papachristou, A., Venne, J. S., and Cheriet, M. (2020), LSTM-based Indoor Air Temperature Prediction Framework for HVAC Systems in Smart Buildings, *Neural Computing and Applications*, **32**(23), 17569-17585.
- Pouriyeh, S., Shahid, O., Parizi, R. M., Sheng, Q. Z., Srivastava, G., Zhao, L., and Nasajpour, M. (2022), Secure Smart Communication Efficiency in Federated Learning: Achievements and Challenges, *Applied Sciences* (Switzerland), **12**(18).
- Shahid, O., Pouriyeh, S., Parizi, R. M., Sheng, Q. Z., Srivastava, G., and Zhao, L. (2021), Communication Efficiency in Federated Learning: Achievements and Challenges, *Archive*.
- Srivastava, N., Hinton, G., Krizhevsky, A., and Salakhutdinov, R. (2014), Dropout: A Simple Way to Prevent Neural Networks from Overfitting, *In Journal of Machine Learning Research*, **15**.
- Yang, Q., Liu, Y., Chen, T., and Tong, Y. (2019), Federated Machine Learning: Concept and Applications, *ACM Transactions on Intelligent Systems and Technology*, **10**(2), 1-19.
- Zeng, E., Mare, S., and Roesner, F. (2017), End User Security and Privacy Concerns with Smart Homes, *Proceedings of the Thirteenth Symposium on Usable Privacy and Security*, 65-80.
- Zhao, T., Liu, G., He, G., Yan, G., Liu, P., and Wang, C. (2024), Effect of Structural Improvement of Gaskets on the Heat Leakage Load and Energy Consumption of the Refrigerator, *Energy*, 300.

저자소개

배진수: 건국대학교 수학과에서 2020년 학사학위를 취득하고 고려대학교 산업경영공학과에서 2026년 박사학위를 취득하였다. 연구분야는 semi-supervised learning와 model calibration이다.

박소연: 숙명여자대학교 통계학과에서 2024년 학사학위를 취득하고 고려대학교 산업경영공학과에서 석사과정에 재학 중이다. 연구분야는 test-time adaptation이다.

최지형: 고려대학교 바이오의공학부에서 2024년 학사학위를 취득하고 고려대학교 산업경영공학과에서 석사과정에 재학 중이다. 연구분야는 federated learning이다.

김지현: 홍익대학교 경영학과에서 2022년 학사학위를 취득하고 고려대학교 산업경영공학과에서 석박통합과정에 재학 중이다.

연구분야는 domain adaptation이다.

함성일: 서울대학교 컴퓨터공학부에서 2009년 박사학위를 취득하고, 이후 삼성전자 Samsung Research에 합류하여 무선통신, 실시간 운영체제, IoT 분야에서 연구개발을 수행하면서 Tizen Lite라는 새로운 IoT향 경량 S/W Platform을 만들었고, 2020년부터 삼성전자 Digital Appliances 사업부에 합류하여 Software Platform 개발을 책임지고 있다. 주요 연구 분야는 Software Platform, On-device AI, IoT 이다.

이보경: 삼성전자 Digital Appliances 사업부의 스마트 가전을 위한 AI 시스템 개발에 참여하고 있다. 주요 업무는 Federated Learning을 통한 On-Device Learning 시스템 설계, MLOps 파이프라인 구축 및 개발이다.

강필섭: 삼성전자 Digital Appliances 사업부에서 스마트 가전 제품에 적용되는 임베디드 온디바이스 AI 기술 개발을 하고 있다. 주요 업무로는 삼성 스마트 가전 플랫폼의 On-Device AI 서비스 아키텍처 설계, 데이터 전처리 최적화, 그리고 딥러닝 모델 경량화 및 최적화이다.

김성범: 고려대학교 산업경영공학부 교수로 2009년부터 재직하고 있으며, 인공지능공학연구소 소장, 기업산학협력센터 센터장, 한국데이터마이닝학회 회장을 역임했다. 미국 University of Texas at Arlington 산업공학과에서 교수를 역임하였으며, 한양대학교 산업공학과에서 학사학위를 미국 Georgia Institute of Technology에서 산업시스템공학 석사 및 박사학위를 취득하였다. 인공지능, 머신러닝, 최적화 방법론을 개발하고 이를 다양한 공학, 자연과학, 사회과학 분야에 응용하는 연구를 수행하고 있다.