https://doi.org/10.7735/ksmte.2026.35.5.389

J. Korean Soc. Manuf. Technol. Eng.
ISSN 2508-5107(Online)

Check for updates

실제 고장 라벨 부재 열처리 공정 온도 데이터를 위한 복합 점수 기반 비지도 이상탐지 프레임워크

모경인[a], 신규식[a,*]

Composite Score-Based Unsupervised Anomaly Detection Framework for Heat Treatment Process Temperature Data without Failure Labels

Kyeong In Mo[a], Kyoo Sik Shin[a,*]

[a] *Department of Manufacturing Artificial Intelligence, Hanyang University*

ARTICLE INFO

Article history:
Received 20 July 2026
Revised 28 July 2026
Accepted 3 August 2026

Keywords:
Heat treatment process
Label-free anomaly detection
Unsupervised anomaly detection
Alert prioritization
AutoEncoder
Forecasting residual

ABSTRACT

This study developed a composite score-based unsupervised temperature anomaly detection (CS-UTAD) framework for heat treatment process data without failure labels, maintenance records, or quality-defect labels. CS-UTAD reformulates the task as anomaly candidate screening and warning/alarm generation rather than supervised failure prediction. The framework combines forecasting residuals, autoencoder reconstruction, and statistical anomaly scores through rank fusion, supported by leakage-safe batch-wise splitting, train/validation-based threshold calibration, and a confirmed warning/alarm policy. On 345,390 one-second process records, CS-UTAD reduced the warning/alarm rate from 5.738% to 4.996% and the alarm rate from 2.971% to 2.531% compared with the guidebook autoencoder baseline; in the test interval, the corresponding rates were 4.986% and 2.613%. Its top-100 candidates included seven IQR weak-signal hits, whereas the baseline included none. The confirmed policy compressed 93,102 test samples into 51 alert episodes for expert review. The results support label-free anomaly monitoring rather than supervised failure prediction.

1. 서 론

열처리 공정은 금속 소재의 기계적 특성과 품질 안정성을 결정하는 핵심 제조 공정이며, 공정 중 온도 변화는 제품 품질과 설비 상태에 직접적으로 연결된다. 온도 제어가 불안정하거나 특정 구간에서 비정상적인 온도 패턴이 발생하면 품질 편차, 설비 부하 증가, 예기치 않은 정지 및 유지보수 비용 증가로 이어질 수 있다. 따라서 열처리 설비에서 수집되는 온도 시계열 데이터를 이용하여 이상징후 후보를 조기에 선별하는 것은 제조 현장의 예지보전 지원 관점에서 중요하다[1-4]. 그러나 실제 제조 현장 데이터에는 고장 발생 시점, 정비 이력, 품질 불량 여부와 같은 명확한 고장 라벨이 충분히 축적되지 않는 경우가 많다. 본 연구의 열처리 공정 데이터도 시간, batch, 온도 및 운전 조건 변수는 포함하지만 실제 고장 라벨은 포함되어 있지 않다[5]. 따라서 본 연구는 고장 발생 여부를 직접 예측하는 문제가 아니라, 고장 라벨이 없는 상황에서 먼저 확인해야 할 온도 이상징후 후보를 선별하고 warning/alarm으로 운영하는 문제로 정의한다[1-3]. 기존의 약한 지도 학습(weakly supervised) 또는 준지도 학습(semi-supervised) 접근법[6, 7]은 최

* Corresponding author. Tel.: +82-31-400-5245
E-mail address: norwalk87@hanyang.ac.kr (Kyoo Sik Shin).

소한의 의사 라벨(pseudo label) 생성에 의존한다는 한계가 있다. 본 연구는 이와 달리 어떠한 라벨도 없는 완전한 비지도 설정에서 복합 점수 기반 이상 후보 선별과 confirmed warning/alarm policy를 통해 전문가 검토를 구조화하는 접근 방식을 채택한다[5]. 본 연구는 가이드북 AutoEncoder reconstruction error 단독 방식의 한계를 보완하기 위해 forecasting residual, reconstruction score, classical anomaly score를 결합한 composite score-based unsupervised temperature anomaly detection (CS-UTAD) 프레임워크를 제안한다[8-14]. 제안 방법은 이상 점수를 실제 고장 판정으로 해석하지 않고, 전문가 검토가 필요한 이상 후보와 경보 운영 정보를 제공하는 데 목적을 둔다. 본 연구의 차별성은 실제 고장 라벨이 없는 제조 데이터에서 평가 가능한 label-free anomaly monitoring 절차를 정립하고, batch-wise split과 train/validation 기반 threshold calibration을 통해 데이터 누수 가능성을 줄이며, fused score를 곧바로 고장으로 판정하지 않고 confirmed warning/alarm episode로 압축하여 현장 전문가 검토 우선순위로 연결한 데 있다. ‘label-free anomaly monitoring 절차’란 명시적인 고장 라벨 없이도 데이터의 비정상적인 패턴을 탐지하고, 이상 후보를 선별 및 구조화하는 일련의 과정을 의미한다. 이를 위해 본 연구는 데이터 누수 방지를 위한 batch-wise split, 객관적인 기준 확보를 위한 train/validation 기반 threshold calibration, 그리고 confirmed warning/alarm episode로의 압축 절차를 정립하였다. 데이터 품질 진단 결과와 변수 구성은 각각 Tables 1 and 2에 제시하였다.

2. 데이터 및 문제 정의

분석 데이터는 KAMP 열처리 예지보전 AI 데이터셋의 data.csv[5]이며, 2022년 1월 3일부터 1월 26일까지 약 1초 간격으로 수집된 345,390행, 24개 열로 구성된다. 데이터는 22개의 공정 변수와 시간 컬럼(TAG_MIN), batch 식별 컬럼이 함께 포함된다[5]. 이러한 구조는 시간적 의존성, 다변량 센서 구조, batch별 정상 분포 차이를 동시에 고려해야 하는 제조 공정 시계열 데이터의 특성을 갖는다. 각 공정 변수는 열처리로 내부의 온도 센서, 설정값 출력(output, OP), 구역별 실제 측정값(process variable, PV) 등을 포함하며, 단위는 주로 ºC이다. 예를 들어, ‘Drying zone 1 OP’는 건조 구역 1의 온도 설정 출력값으로, 실제 온도 편차를 보정하기 위한 제어 신호에 해당한다. 데이터 수집 빈도는 1초 간격이며 연속형(continuous) 수치 데이터로 구성된다. 데이터 품질 진단 결과, 완전성 품질 지수는 99.994%, row uniqueness, timestamp parse quality, timestamp uniqueness, timestamp monotonic consistency는 모

Table 1 Data quality diagnosis results

Quality metric	Value	Interpretation
Completeness quality index	99.994%	The dataset has very few missing values and high completeness
Row uniqueness quality index	100.000%	No duplicated rows were found
Timestamp parse quality index	100.000%	All timestamps were successfully parsed
Timestamp uniqueness quality index	100.000%	No duplicated timestamps were found
Timestamp monotonic consistency	100.000%	The time sequence increases consistently
Numeric validity quality index	99.994%	Numeric data validity is high
IQR non-outlier validity	96.505%	Some extreme values exist under the IQR criterion

Table 2 Heat treatment process data variable composition

Category	Variable examples	Meaning
Time information	TAG_MIN	Data collection timestamp
Batch information	Batch assignment number	Batch or work-unit identifier
Drying process operating condition	Drying zone 1 OP	Output value for maintaining the drying-zone temperature
Drying furnace temperature	Drying furnace temperature Zone 1, Zone 2	Temperature of each drying-furnace zone
Washer	Washer	Washer temperature
Quenching furnace operating condition	Quenching zone 1 OP to zone 4 OP	Output value for maintaining the quenching-zone temperature
CP-related variables	Quenching furnace CP value, CP monitor value	CP and monitoring values in the quenching furnace
Quenching furnace temperature	Quenching furnace temperature Zone 1 to Zone 4	Temperature of each quenching-furnace zone
Salt process operating condition	Salt zone 1 OP, Salt zone 2 OP	Output value for maintaining the salt-zone temperature
Salt-related temperature	Salt sludge removal, salt conveyor temperature Zone 1/2, salt bath temperature Zone 1/2	Temperature of the salt process and related equipment

Table 3 CS-UTAD framework processing steps

Step	Main procedure	Purpose
1	Data loading	Identify the dataset purpose, variables, and baseline
2	Data quality diagnosis	Confirm time-series suitability and absence of labels
3	Batch-wise split	Prevent data leakage and approximate an operational evaluation setting
4	Feature engineering	Reflect temperature variation, operating conditions, and regime characteristics
5	Forecasting model	Compute anomaly scores based on forecasting errors
6	LSTM AutoEncoder	Compute anomaly scores based on reconstruction errors
7	Classical anomaly score	Reflect statistical extremeness and distributional abnormality
8	Pseudo-normal refinement	Stabilize the normal reference distribution
9	Rank fusion and weight search	Derive an integrated anomaly score from multiple scores
10	Train/validation-based threshold setting	Set thresholds from train/validation score distributions
11	Multi-condition alert confirmation	Confirm operational alert states
12	Episode summary and expert review support	Support result interpretation and field-level review prioritization

두 100.0%로 나타났다. 이는 시계열 모델 학습과 평가를 수행하기 위한 기본 품질이 확보되어 있음을 의미한다. 다만 IQR non-outlier validity는 96.505%로 일부 극단값 후보가 존재하였다. 본 연구에서는 이러한 IQR 기반 신호를 실제 정답 라벨이 아니라 보조적인 weak signal로만 사용한다. 이 weak signal은 Ratner et al. (2017)[15]의 약한 감독(weak supervision) 개념에 기반하여, 전문가 라벨 없이 데이터 분포 기반의 규칙(IQR 기준 1.5배 초과 여부)으로 자동 생성된 보조 신호로 정의된다. 실제 고장 라벨과 달리 이 신호는 이상 여부를 확정하지 않으며, 모델 평가 및 후보 선별의 보조 근거로만 활용한다. 본 데이터에는 실제 고장 라벨, 정비 이력 컬럼, 품질 불량 라벨이 존재하지 않는다[5]. Precision-recall 계열 지표는 실제 양성·음성 라벨을 전제로 해석되어야 하므로[16], 본 데이터에서는 precision, recall, F1과 같은 지도학습 평가 지표를 실제 고장 예측 성능으로 해석하지 않는다. 이러한 label-free 운영 지표들은 실제 고장 라벨이 없는 상황에서 시스템의 실용적 유용성을 평가하기 위한 대체 지표로서, 경보 부담 감소, 상위 후보 집중도, 검토 우선순위화와 같은 실용적 관점의 지표가 현장 적용성 평가에 더욱 적합하다. 본 연구의 평가는 이러한 label-free 운영 지표를 중심으로 수행한다[16]. CS-UTAD의 전체 처리 절차는 Table 3에 정리하였다.

3. CS-UTAD 프레임워크

본 CS-UTAD는 고장 라벨이 없는 열처리 공정 데이터에서 정상 패턴과 다른 온도 변화 구간을 선별하기 위한 복합 점수 기반 비지도 이상탐지 프레임워크이다. 전체 절차는 데이터 품질 진단, feature engineering, leakage-safe batch-wise split, forecasting model 및 LSTM AutoEncoder 학습, classical anomaly score 산출, rank fusion, train/validation 기반 threshold calibration, confirmed warning/alarm policy 적용으로 구성된다. Feature engineering은 원시 온도값, 변화량, rolling 통계량, 운전 조건, batch 및 regime 정보를 포함한다. 데이터 분할은 동일 batch 정보가 학습과 평가에 동시에 포함되어 발생할 수 있는 leakage를 줄이기 위해 batch-wise split을 사용하였다. 구체적으로, batch-wise split은 각 batch ID 내의 모든 데이터 포인트를 시간 순서에 따라 train, validation, test 중 하나의 집합에만 배타적으로 할당하는 방식으로 수행하였다. 인접한 배치 간에 데이터가 혼합되지 않도록 배치 경계를 기준으로 분리하였으며, test 구간에 해당하는 배치 정보는 threshold 결정 및 가중치 탐색 과정에서 일절 사용하지 않았다. 이를 통해 동일 배치 정보가 학습과 평가에 동시에 포함되는 데이터 누수를 방지하였다. 이후 forecasting model은 과거 window와 운전 조건을 이용하여 다음 시점의 온도 패턴을 예측하고, 실제 관측값과 예측값의 차이를 forecasting residual score로 변환한다[8, 9]. Forecasting model은 LSTM 기반 시계열 예측 구조로, 입력 window 길이 60, 은닉층 2개(각 64 유닛), 드롭아웃 비율 0.2, 활성화 함수 tanh, 최적화 알고리즘 Adam(학습률 0.001), 배치 크기 256, 에포크 수 20으로 구성하였다. LSTM AutoEncoder는 입력 window의 정상적인 다변량 시계열 구조를 재구성하도록 학습되며, reconstruction error를 reconstruction score로 사용한다[10, 11]. LSTM AutoEncoder는 인코더 2층(각 64, 32 유닛)과 디코더 2층(각 32, 64 유닛) 구조로 구성하였으며, 동일하게 Adam 최적화(학습률 0.001), 배치 크기 256, 에포크 20으로 학습하였다. 구현 환경은 Python 3.9, TensorFlow 2.10을 사용하였다. Classical anomaly score는 robust z-score, percentile threshold, Isolation Forest 계열의 통계적 이상도를 포함하여 순간적 극단값과 국소적 분포 이탈을 반영한다[12-14]. 각 고전적 이상 점수는 탐지 대상 이상 유형에 따라 역할이 구분된다. Robust z-score는 중앙값과 IQR을 기반으로 개별 시점의 분포 극단값(점 이상, point anomaly)을 포착하며, 이동 윈도우(window size = 60)를 적용하였다. Percentile threshold는 상위 분위수(99.5 퍼센타일) 초과 여부를 기반으로 컨텍스트 이상(contextual

anomaly)을 반영한다. Isolation Forest(n_estimators = 100, max_samples = 'auto', random_state = 42)는 고차원 공간에서 집단 이상(collective anomaly) 및 국소 분포 이탈을 탐지하는 데 유리하다. 각 score는 scale이 서로 다르므로 원시값을 직접 더하지 않고 rank fusion을 적용하여 결합한다. Forecasting residual은 시간적 예측 불일치, reconstruction score는 다변량 구조 이탈, classical anomaly score는 분포 극단과 국소 희소성을 각각 반영한다[8-14]. 따라서 rank fusion은 상이한 이상 관점을 동일한 순위 척도에서 집계하여 단일 점수의 유형별 편향과 원시 score 척도 차이의 영향을 완화하고, detector agreement에 필요한 다원적 근거를 제공한다. 본 연구에서 이 결합의 장점은 실제 고장 탐지 정확도의 우월성이 아니라 label-free 운영 지표와 ablation 결과가 보여 준 후보 선별 및 운영 단계의 보완성으로 한정한다. 최종 fused score는 train/validation 구간에서 탐색한 fusion weight를 기반으로 산출된다. Rank fusion 가중치 탐색은 validation 데이터셋에서의 IQR weak signal 히트율을 최대화하는 목적 함수를 기준으로, 각 가중치를 0.00부터 1.00까지 0.05 단위 그리드 서치(grid search) 방식으로 수행하였다. 탐색 과정에서 test 구간 정보는 일절 사용하지 않았으며, 가중치 합이 1.00이 되는 조합만을 탐색 대상으로 한정하였다. 본 연구에서 선택된 weight는 reconstruction 0.20, forecast 0.40, trend 0.00, classical 0.40이었다.

이 값들은 이론적으로 고정된 보편값이 아니라 validation 구간에서 weak signal의 상위 순위 집중도를 최대화한 경험적 값이며, 가중치 크기는 각 점수의 물리적 중요도가 아니라 해당 데이터의 시간 변화, 다변량 재구성, 분포 이탈 패턴이 목적함수에 기여한 상대적 정도를 반영한다. 따라서 공정 regime, sampling interval, sensor 구성 또는 weak signal 정의가 달라지면 최적 가중치도 달라질 수 있으므로, 새로운 데이터에서는 test 정보를 제외한 train/validation 구간에서 동일한 leakage-safe 절차로 다시 탐색해야 한다. 본 연구는 단일 KAMP 데이터셋만 사용하였으므로 수치 가중치 자체의 외부 일반화 가능성은 아직 검증되지 않았다. trend score의 가중치가 0.00으로 결정된 이유는 가중치 탐색 과정에서 trend score가 본 데이터의 이상 징후 탐지에 유의미한 독립적 기여를 하지 못하거나, 다른 점수(특히 forecasting residual)와 높은 상관관계를 보여 중복 정보를 제공한다고 판단되었기 때문이다. 이는 특정 데이터셋 특성에 따라 trend score의 기여가 제한적일 수 있음을 시사하며, 가중치 탐색 실험 결과에 의해 도출된 경험적 결론이다. 이는 본 데이터에서 AutoEncoder reconstruction error 단독보다 forecasting residual과 classical anomaly score가 이상 후보 선별에 더 중요한 기여를 했음을 시사한다. Threshold calibration은 test 정보를 사용하지 않고 train/validation score 분포를 기준으로 임계값을 설정하였다. Confirmed policy는 여러 검출 조건을 함께 고려하는 다중 조건 기반 경보 확정 절차로 적용하여, 일시적 spike로 인한 과도한 경보를 줄이고 검토 후보를 episode 단위로 압축하였다. 구체적인 batch-wise split은 train 199,598 samples, validation 52,690 samples, test 93,102 samples로 구성하였다. 본 연구에서 fused score 기반 warning/alarm 후보와 confirmed alert profile은 서로 다른 단계의 결과로 구분된다. Fused score 기반 WOA rate와 alarm rate는 train/validation 분포에서 산출한 threshold를 test 구간에 적용했을 때의 1차 후보 비율이며, confirmed alert profile은 이 후보들에 다중 조건 기반 경보 확정을 적용한 운영 단계의 최종 상태이다. 따라서 fused score 기반 WOA/alarm 감소 결과는 전체 후보 부담의 변화로, confirmed alert profile은 실제 현장에서 검토해야 할 최종 경보 상태의 압축 정도로 해석해야 한다.

4. 실험 결과 및 분석

4.1 실험 설정 및 학습 수렴

실험은 실제 고장 라벨이 없는 조건을 반영하여 모델 학습, threshold 결정, profile 선택 과정에서 고장 라벨을 사용하지 않는 방식으로 수행하였다. Forecasting model과 LSTM AutoEncoder는 20 epochs의 단일 학습에서 train/validation loss가 지속적으로 감소하였다. 두 모델의 loss 함수는 평균 제곱 오차(MSE: Mean Squared Error)를 사용하였다. 두 모델의 train/validation loss는 학습 초기에 빠르게 감소한 후 완만하게 수렴하는 공통 추세를 보였다. 이 MSE 감소는 모델이 정상적인 온도 변화 패턴을 더욱 정확하게 학습하였음을 의미하며, 이후 이상 후보 선별 시 정상 패턴 기준이 더욱 안정화되어 미묘한 이상 징후의 감지 민감도가 높아지는 기반이 된다. 가이드북 AutoEncoder baseline은 reconstruction error 단독 기준으로 warning/alarm 후보를 산출하는 비교 기준으로 사용하였다. 가이드북 AutoEncoder baseline은 일반적인 재구성 오차 기반 AutoEncoder 방식[10, 11]을 따르며, threshold 초과 여부만으로 warning/alarm을 결정하는 단순화된 운영 구조를 갖는다. CS-UTAD는 동일 데이터 조건에서 복수 score 결합과 confirmed policy를 적용하여 경보 부담과 후보 집중도를 비교하였다. 모든 weak-label 기반 지표는 실제 고장 검증 성능이 아니라 guidebook IQR 기반 보조 신호와의 일치 정도로 한정하여 해석하였다[16]. Fig. 1은 개별 loss 값보다 두 모델의 감소 및 수렴 추세를 중심으로 제시하였다.

4.2 Guidebook AutoEncoder 대비 성능

전체 구간 기준으로 CS-UTAD fused score는 Guidebook

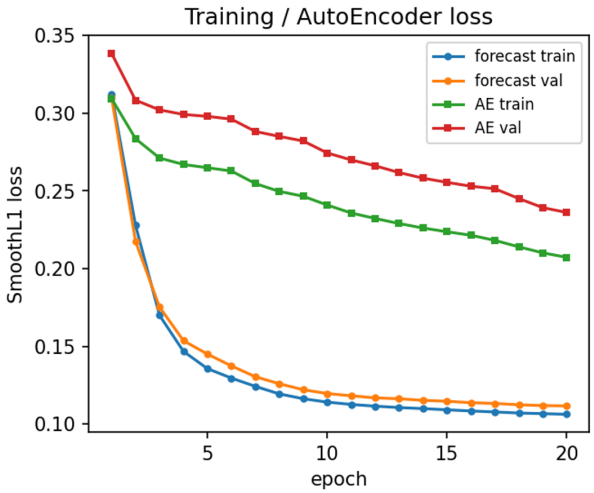


Fig. 1 Train/validation loss curves for each model

AutoEncoder baseline 대비 warning-or-alarm rate를 5.738%에서 4.996%로, alarm rate를 2.971%에서 2.531%로 낮추었다. Test 구간에서도 baseline의 WOA rate 7.738%, alarm rate 4.246%에 비해 CS-UTAD는 각각 4.986%, 2.613%로 낮은 경보 부담을 보였다. Test 구간 93,102개 sample에서 warning-or-alarm 후보는 7,204건에서 4,642건으로 2,562건(약 35.6%) 감소하였다. 이러한 경보 감소는 현장 운영자의 '경보 피로도(alarm fatigue)'를 크게 줄여 중요한 이상 징후에 대한 집중도를 높이고 적시에 대응할 수 있도록 한다. Top-N enrichment 분석에서는 IQR weak signal과 중첩되는 후보를 상위에 배치하는 경향을 보였다. 전체 데이터 Top-100 후보 기준으로 Guidebook AutoEncoder baseline은 IQR weak signal을 포함하지 못한 반면, CS-UTAD fused score는 7건을 포함하였다. 이는 top-100 hit rate 7.0%, weak signal coverage 9.72%, enrichment 335.80배에 해당한다. 다만 이 결과는 실제 고장 탐지 성능이 아니라 고장 라벨 부재 조건에서 상위 검토 후보가 weak signal을 얼마나 집중적으로 포함하는지를 나타내는 보조 결과이다. 추가 ablation 분석에서 reconstruction-only는 test WOA rate 7.741%, alarm rate 4.246%로 Guidebook AutoEncoder baseline과 유사한 수준을 보였으나, forecast-only는 4.926%/2.429%, classical-only는 4.737%/2.398%, forecast+classical 조합은 4.577%/2.271%로 낮은 경보 부담을 보였다. Full CS-UTAD는 4.986%/2.613%로, 단순 경보율 최소화보다 forecasting residual, reconstruction score, classical anomaly score를 통합하여 다원적 detector agreement와 episode-level review 가능성을 동시에 확보하는 '운영형 구성'으로 해석된다. '운영형 구성'이란 다양한 이상 점수를 통합적으로 활용하여 현장 전문가가 이상 징후를 직관적으로 이해하고 효과적으로 검토할 수 있도록 지원하는 실용적이고 통합적인 구조이다. Ablation 결과는 forecast-only가 급격한 이상 징후에 더 민감한 반면, CS-UTAD의 복합 점수는 개별 모델들이 놓칠 수 있는 다양한 이상 유형을 포괄적으로 탐지함을 보여준다. 따라서 본 결과는 고장 라벨 부재 조건에서 복합 score와 confirmed policy가 경보 부담을 줄이고 전문가 검토 후보를 구조화할 수 있음을 보여준다. 개선 폭을 현장 관점에서 보면, 1,000개 온도 기록 기준으로 WOA 후보가 약 77건에서 50건으로, alarm 후보가 약 42건에서 26건으로 감소하였으며, Top-100 후보에서는 baseline이 IQR weak signal을 포함하지 못한 반면 CS-UTAD는 7건을 포함하였다. 전체 구간의 AE baseline 비교 결과와 Top-100 weak signal hit 여부는 각각 Tables 4 and 5에 제시하였다.

Table 4 Overall performance comparison between AE and CS-UTAD

Metric	Guidebook AutoEncoder baseline	CS-UTAD fused Score	Interpretation
WOA count	19,819	17,257	CS-UTAD reduced warning/alarm candidates
WOA rate	5.738%	4.996%	CS-UTAD reduced alert burden
Alarm count	10,262	8,741	CS-UTAD reduced alarm candidates
Alarm rate	2.971%	2.531%	CS-UTAD reduced alarm burden
Weak-positive count	72	72	Based on the IQR weak signal

Table 5 Top-100 weak signal hit comparison

Method	Top-100 weak signal hit
Guidebook AutoEncoder baseline	No
CS-UTAD fused score	Yes

4.3 Confirmed alert 및 batch/episode 분석

최종 confirmed alert profile은 test 구간에서 normal 99.59%, warning 0.10%, alarm 0.31%를 보이며 매우 보수적으로 작동하였다. 이는 모든 높은 fused score를 즉시 경보로 전환하지 않고, 다중 조건 기반 경보 확정 절차를 만족하는 제한된 구간만 warning/alarm으로 확정했기 때문이다. 이러한 정책은 현장 경보 피로를 줄이는 데 유리하지만, 실제 이상을 놓칠 가능성도 있으므로 현장 전문가 검토와 함께 조정되어야 한다. Batch/equipment 별 alert burden 분석 결과, 경보는 test 구간 전체에 무작위로 분

Table 6 Final confirmed alert profile state ratio

Normal	Warning	Alarm	WOA count	Alarm count
99.59%	0.10%	0.31%	382	289

Table 7 Top three batches by WOA rate

Batch ID	WOA rate	Alarm rate
103675	1.254%	1.013%
102585	0.587%	0.391%
107402	0.400%	0.293%

포하지 않고 특정 batch 및 episode에 집중되었다. Confirmed policy는 검토 대상을 WOA 382건과 51개 alert episode로 압축하였으며, batch 103675는 가장 높은 WOA rate와 alarm rate를, batch 107402는 가장 많은 alert episode를 보였다. Batch 103675의 경보 집중은 온도 설정값과 실제값 편차가 운전 범위를 반복적으로 벗어나는 조건이 존재했을 가능성을, batch 107402의 많은 episode는 장기 운전(38,237 samples)과 장비 미세 노후화로 인한 소규모 이상 징후 지속 발생 가능성을 시사한다. 이러한 batch별 특성 분석은 향후 현장 전문가가 심층적인 원인 분석과 예방적 조치를 계획하는 출발점이 된다. CS-UTAD 결과는 어느 batch와 episode를 우선 확인해야 하는지 구조화된 정보를 제공한다. 이러한 우선순위 지정은 현장 검토의 효율성을 극대화하지만, 우선순위는 실제 설비 고장을 확정하는 근거가 아니며, 정비 이력과 품질 검사 결과가 확보된 이후에는 threshold 및 confirmed policy를 재보정해야 한다. 현장 담당자는 93,102개 sample 전체를 처음부터 훑기보다 51개 alert episode와 경보 부담이 높은 batch를 먼저 확인할 수 있다. 최종 상태 비율과 WOA/Alarm rate 상위 3개 batch의 핵심 결과는 각각 Tables 6 and 7에 제시하였다.

5. 결 론

본 연구는 실제 고장 라벨이 없는 열처리 공정 온도 데이터를 대상으로 복합 점수 기반 비지도 이상탐지 프레임 워크인 CS-UTAD를 제안하였다. 제안 방법은 Guidebook AutoEncoder baseline의 reconstruction error 단독 방식을 확장하여 forecasting residual, reconstruction score, classical anomaly score를 rank fusion으로 결합하고, batch/regime threshold calibration과 confirmed warning/alarm policy를 통해 실무 검토 가능한 경보 후보로 변환하였다. 실험 결과 CS-UTAD는 Guidebook AutoEncoder baseline보다 낮은 경보 부담과 상위 후보에서 더 많은 weak signal 중첩을 보였다. 전체 구간과 test 구간 모두에서 WOA rate와 alarm rate가 감소하였으며, Top-100 후보에서는 baseline이 weak signal을 포함하지 못한 반면 CS-UTAD는 7건을 포함하였다. 또한 confirmed policy는 test 구간의 검토 대상을 51개 alert episode로 압축하여 batch 및 episode 기반 전문가 검토 우선순위화를 지원할 수 있음을 보였다. 본 연구의 결과는 실제 고장 라벨로 검증된 고장 예측 성능이 아니다. Guidebook IQR weak signal은 데이터 분포 기반의 pseudo/weak signal이며, test 구간에는 weak-positive가 존재하지 않으므로 test precision, recall, F1은 해석할 수 없다. 이는 IQR weak signal이 train/validation 구간 분포에서 생성된 보조 신호로, test 구간에서 전통 정량 지표의 분모 역할을 하는 positive 샘플 자체가 없기 때문이다. 따라서 본 연구의 의의는 고장 라벨 부재 제조 데이터에서 이상징후 후보 선별, 경보 부담 감소, 전문가 검토 우선순위화가 가능한 운영 중심의 분석 절차를 제시한 데 있다. IQR 기반 weak signal에 의존하는 현재 평가의 한계를 보완하기 위해, 향후에는 실제 정비 이력, 고장 이벤트, 품질 검사 데이터 및 현장 전문가 검토 label과 결합하여 warning/alarm과 실제 설비 상태 간의 관계를 검증해야 한다. 특히 CS-UTAD 경보 에피소드를 전문가에게 제공하고 3개월 내 실제 고장 발생 여부를 추적하는 파일럿 연구를 통해 오경보 감소의 타당성을 검증할 계획이다. 라벨링 비용이 과도한 현장에서는 약한 라벨(예: 비정상 공정 보고서) 또는 능동 학습(Active Learning) 기법을 활용하여 모델 검증 및 개선을 위한 현실적인 방법론을 발전시킬 수 있다. 이와 함께, 본 연구의 방법론적 기여와 실용적 가치는 ablation 분석을 통해 더욱 명확히 드러난다. 또한 ablation 결과는 reconstruction error 단독 방식이 본 데이터에서 충분하지 않으며, forecasting residual과 classical anomaly score가 경보 부담 감소에 실질적으로 기여함을 보여준다. 요약하면, 본 연구의 실무적 의미는 라벨이 없는 대량의 온도 데이터를 그대로 제시하는 것이 아니라, 현장 전문가가 먼저 확인할 수 있는 작은 후보 목록과 episode 단위의 검토 순서로 바꾸어 준다는 데 있다. CS-UTAD의 모듈형 구성은 화학 공정, 발전소, 반도체 제조 등 시계열 특성을 가진 다른 산업 분야에도 각 도메인의 데이터 특성에 맞추어 가중치 및 threshold를 조정함으로써 적용 가능하다. 다만 본 연구는 단일 데이터셋(KAMP)에 한정되었으므로, 다른 설비 및 이종 산업 환경에서의 외부 유효성 검증과 다변량 센서 데이터로의 확장이 향후 연구 과제로 남는다. 또한 rank fusion 가중치(0.20, 0.40, 0.00, 0.40)는 특정 데이터셋에 최적화된 경험적 값이므로, 다른 환경에서는 메타 학습(meta-learning) 기반의 적응형 가중치 조정 방식 도입을 검토할 필요가 있다.

References

[1] Carvalho, T. P., Soares, F. A. A. M. N., Vita, R., Francisco, R. P., Basto, J. P., Alcalá, S. G. S., 2019, A Systematic Literature Review of Machine Learning Methods Applied to Predictive Maintenance, Comput. Ind. Eng., 137 106024, https://doi.org/10.1016/j.cie.2019.106024.

[2] Zonta, T., da Costa, C. A., da Rosa Righi, R., de Lima, M. J., da Trindade, E. S., Li, G. P., 2020, Predictive Maintenance in the Industry 4.0: A Systematic Literature Review, Comput. Ind. Eng., 150 106889, https://doi.org/10.1016/j.cie.2020.106889.

[3] Zhang, W., Yang, D., Wang, H., 2019, Data-driven Methods for Predictive Maintenance of Industrial Equipment: A Survey, IEEE Syst. J., 13:3 2213-2227, https://doi.org/10.1109/JSYST.2019.2905565.

[4] Gopaluni, R. B., Tulsyan, A., Chachuat, B., Huang, B., Lee, J. M., Amjad, F., Damarla, S. K., Kim, J. W., Lawrence, N. P., 2020, Modern Machine Learning Tools for Monitoring and Control of Industrial Processes: A Survey, IFAC-PapersOnLine, 53:2 218-229, https://doi.org/10.1016/j.ifacol.2020.12.126.

[5] Ministry of SMEs and Startups, n.d., 23 Dec. 2022, Heat Treatment Predictive Maintenance AI Dataset and Practice Guidebook, <https://www.kamp-ai.kr/aidataDetail?AI_SEARCH=%EC%97%B4%EC%B2%98%EB%A6%AC&page=1&DATASET_SEQ=43&DISPLAY_MODE_SEL=CARD&EQUIP_SEL=&GUBUN_SEL=&FILE_TYPE_SEL=&WDATE_SEL=>.

[6] Liu, F., Zhang, L., Li, H., Zhou, S., Zhou, Y., 2024, A Weakly Supervised Time Series Anomaly Detection Method with Dual-association Discrepancy, SIViP, 18 4831-4839, https://doi.org/10.1007/s11760-024-03119-y.

[7] Kim, T. H., Jin, K,-H., 2024, Manufacturing Time Series Data Anomaly Labeling and Semi-Supervised Learning based Anomaly Detection, Journal of the Korea Institute Of Information and Communication Engineering, 28 5 643-649, https://doi.org/10.6109/jkiice.2024.28.5.643.

[8] Malhotra, P., Vig, L., Shroff, G., Agarwal, P., 2015, Long Short Term Memory Networks for Anomaly Detection in Time Series, Long Short Term Memory Networks for Anomaly Detection in Time Series, 89-94.

[9] Hundman, K., Constantinou, V., Laporte, C., Colwell, I., Soderstrom, T., 2018, Detecting Spacecraft Anomalies Using LSTMs and Nonparametric Dynamic Thresholding, Proc. ACM SIGKDD Int. Conf. Knowl. Discov. Data Min., 387-395, https://doi.org/10.1145/3219819.3219845.

[10] Sakurada, M., Yairi, T., 2014, Anomaly Detection Using AutoEncoders with Nonlinear Dimensionality Reduction, Proc. Workshop Mach. Learn. Sens. Data Anal., 4-11, https://doi.org/10.1145/2689746.2689747.

[11] Pota, M., De Pietro, G., Esposito, M., 2023, Real-time Anomaly Detection on Time Series of Industrial Furnaces: A Comparison of AutoEncoder Architectures, Eng. Appl. Artif. Intell., 124 106597, https://doi.org/10.1016/j.engappai.2023.106597.

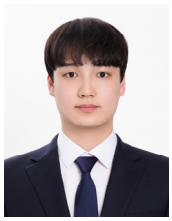
[12] Breunig, M. M., Kriegel, H. P., Ng, R. T., Sander, J., 2000, LOF: Identifying Density-Based Local Outliers, Proc. ACM SIGMOD Int. Conf. Manag. Data, 93-104, https://doi.org/10.1145/342009.335388.

[13] Schölkopf, B., Platt, J. C., Shawe-Taylor, J., Smola, A. J., Williamson, R. C., 2001, Estimating the Support of a High-Dimensional Distribution, Neural Comput., 13:7 1443-1471, https://doi.org/10.1162/089976601750264965.

[14] Liu, F. T., Ting, K. M., Zhou, Z. H., 2008, Isolation Forest, Proc. IEEE Int. Conf. Data Min., 413-422, https://doi.org/10.1109/ICDM.2008.17.

[15] Ratner, A., Bach, S. H., Ehrenberg, H., Fries, J., Wu, S., Ré, C., 2017, Snorkel: Rapid Training Data Creation with Weak Supervision, Proc. VLDB Endow., 11:3 269-282, https://doi.org/10.14778/3157794.3157797.

[16] Saito, T., Rehmsmeier, M., 2015, The Precision-Recall Plot Is More Informative than the ROC Plot When Evaluating Binary Classifiers on Imbalanced Datasets, PLoS ONE, 10:3 e0118432, https://doi.org/10.1371/journal.pone.0118432.



Kyeong In Mo
M.Sc. Candidate in the Department of Manufacturing Artificial Intelligence, Hanyang University. His research interests include artificial intelligence.
E-mail: kimo2020@hanyang.ac.kr



Kyoo Sik Shin
Professor in the Department of Manufacturing Artificial Intelligence, Hanyang University. His research interests include physical AI.
E-mail: norwalk87@hanyang.ac.kr