

Research on KS Standard-based Outlier Classification and Intelligent Cleaning Framework for Smart Greenhouse Sensor Data

Kyeong-il Ko*, Byeongbeom Kang**, Hyun Yoe***

*Student, Major in Smart Agriculture, Suncheon National University, Suncheon, Korea

**Senior researcher, Jeonbuk Technopark, Jeonju, Korea

***Professor Emeritus, Department of Information and Communication Engineering, Suncheon National University, Suncheon, Korea

[Abstract]

Securing data quality is essential for precision environmental control in smart farms. In particular, data integrity is a prerequisite for ensuring the effectiveness of the recently established national standards, such as 'Smart Greenhouse Sensor/Node Metadata (KS X 3269, KS X 3287).' However, existing research has limitations in treating outliers uniformly without distinguishing their causes, which hinders control continuity. To address this, this study proposes an intelligent data cleaning framework that automatically classifies outliers into 'transient outliers (noise)' and 'persistent outliers (failure)' using an LSTM-based prediction model and time-series persistence analysis (K-Counter). The proposed method corrects transient outliers with predicted values to maintain data continuity, while for persistent outliers, it updates the node status to 'ERROR' in accordance with the KS X 3287 standard, thereby immediately excluding them from the control logic. Experimental results adhering to the KS X 3269 standard ranges demonstrated a classification accuracy of 98.1%. This study contributes to enhancing the interoperability and stability of national standard-based smart farm control systems.

▶ **Key words:** Smart Farm, Outlier Detection, Data Cleaning, KS Standards, LSTM, Time-Series Analysis

[요약]

스마트팜의 정밀 환경 제어를 위해서는 데이터 품질 확보가 필수적이며, 특히 최근 제정된 '스마트 온실 센서/노드 메타데이터(KS X 3269, KS X 3287)' 표준의 실효성을 확보하기 위해 데이터 무결성이 선행되어야 한다. 그러나 기존 연구는 이상치의 원인을 구분하지 않고 획일적으로 처리하여 제어 연속성을 저해하는 한계가 있다. 이에 본 연구는 LSTM 기반 예측 모델과 시계열 지속성 분석(K-Counter)을 통해 이상치를 '일시적 이상치(잡음)'와 '지속적 이상치(고장)'로 자동 분류하는 지능형 정제 프레임워크를 제안한다. 제안 방식은 일시적 이상치를 예측값으로 교정하여 제어 데이터의 연속성을 유지하고, 지속적 이상치는 KS X 3287 표준에 따라 해당 노드 상태를 '오류(ERROR)'로 갱신하여 제어 로직에서 즉시 배제한다. KS X 3269 표준 범위를 준수한 실험 결과 98.1%의 분류 정확도를 달성하였으며, 본 연구는 국가 표준 기반 스마트팜 제어 시스템의 상호운용성과 안정성 향상에 기여한다.

▶ **주제어:** 스마트팜, 이상치 탐지, 데이터 정제, KS 표준, LSTM, 시계열 분석

- First Author: Kyeong-il Ko, Corresponding Author: Hyun Yoe
- *Kyeong-il Ko (koruddlf@gmail.com), Major in Smart Agriculture, Suncheon National University
- **Byeongbeom Kang (bbk211030@gmail.com), Jeonbuk Technopark
- ***Hyun Yoe (yhyun@scnu.ac.kr), Department of Information and Communication Engineering, Suncheon National University
- Received: 2025. 12. 10, Revised: 2026. 02. 13, Accepted: 2026. 03. 16.

I. Introduction

4차 산업혁명 기술과 농업의 융합으로, '스마트팜'은 식량 안보 문제와 농촌 고령화에 대응하며 농업 생산성을 향상시킬 핵심 기술로 전 세계적인 주목을 받고 있다. 세계 스마트 농업 시장은 2024년 약 262억 달러 규모에서 연평균 15.7% 성장하여 2030년에는 628억 달러에 이를 것으로 전망되며[1], 한국 정부 역시 스마트 농업 생산 비중 확대 및 관련 기술 개발을 위한 정책적 지원을 강화하고 있다[2].

스마트팜의 핵심 경쟁력은 정밀한 데이터 기반 의사결정에서 비롯된다. ICT 기술을 활용하여 온실 내부의 온도, 습도, 이산화탄소(CO₂) 등 환경 데이터와 작물의 생육 정보를 실시간으로 수집하고, 이를 분석하여 최적의 생육 환경을 도출함으로써 냉난방기, 환기창, 관수 시스템 등 액추에이터를 정밀하게 자동 제어한다[3-5]. 이처럼 데이터 기반의 정밀 제어가 스마트팜의 성패를 좌우함에 따라, 제어 시스템의 입력값이 되는 데이터의 품질은 시스템 전체의 신뢰도와 직결되는 가장 기본적인 전제 조건이다[6]. 만약 센서 오류나 노이즈로 인해 오염된 데이터가 제어 시스템에 입력될 경우, 이는 불필요한 에너지 낭비와 작물 생육 저해 등 심각한 경제적 손실로 이어질 수 있다[7].

특히, 최근 국내 스마트농업 분야에서는 기기 간의 상호 운용성을 확보하고 시스템 효율을 높이기 위해 '스마트 온실 노드 메타데이터(KS X 3287)[8]', '스마트 온실 센서 메타데이터(KS X 3269)[9]' 등의 국가 표준(KS) 체계를 강화하고 있다. 이러한 표준은 제어기와 노드 간의 데이터 교환 형식과 상태 정보 처리 방식을 규정한다. 하지만 표준에서 정의하는 물리적 측정 범위(예: -40~80°C)는 기기의 동작 한계를 의미할 뿐, 현재 작물 상황에 맞는 유효성(Semantic Validity)까지 보증하지 못한다. 표준화된 메타데이터 구조를 따르더라도 그 내용이 실제 환경과 맞지 않는 오염된 데이터라면 제어의 정확도는 보장될 수 없다. 따라서 본 연구는 KS 표준 규격을 준수하면서도, 지능형 알고리즘을 통해 데이터 자체의 무결성을 기술적으로 뒷받침하는 것을 목적으로 한다.

실제 농업 현장의 복잡성, 특히 높은 습도, 분진, 미생물 및 해충의 존재는 정밀 센서의 안정적인 운영에 심각한 위협이 된다[10]. 이러한 환경에서 발생하는 데이터 신뢰성 저하 문제는 상이한 원인을 가진 두 가지 유형으로 구체화할 수 있다.

첫째, '일시적 이상치(Transient Anomaly)'는 센서 하드웨어의 결함이 아닌, 농작업 환경의 외인성 요인에 의해 발생하는 비지속적 오류이다. 이는 관수 작업 중 발생하는

수분 접촉, 작업자 및 해충의 일시적 간섭, 혹은 무선 통신 환경의 간헐적 불안정성에 기인하는 확률적 잡음 형태를 띤다.

둘째, '지속적 이상치(Persistent Anomaly)'는 센서 자체의 하드웨어 결함 또는 고장으로 인해 발생하는 영구적 오류이다. 여기에는 센서 노후화에 따른 교정 값 이탈, 배터리 소진이나 물리적 단선으로 인한 고착 오류 등이 포함된다. 기존의 획일적인 제어 시스템은 이 두 가지를 구분하지 못하고 오염된 데이터를 유효한 정보로 오인하여 액추에이터의 잦은 오작동(Chattering)이나 영구적인 제어 실패를 유발한다.

본 연구는 이러한 한계를 극복하기 위해, 탐지된 이상치의 시계열적 패턴을 분석하여 그 원인을 자동 분류하고, KS 표준 메타데이터와 연동하여 지능적으로 정제하는 프레임워크를 제안한다. 본 연구의 주요 기여점은 다음과 같이 요약할 수 있다.

첫째, 시계열 지속성 기반의 이상치 유형 분류 알고리즘을 제안한다. LSTM 기반 예측 모델과 시계열 지속성 분석(K-Counter)을 결합하여, 단순 이상 징후를 '일시적 이상치(잡음)'와 '지속적 이상치(고장)'로 자동 분류함으로써 기존 획일적 탐지 기법의 한계를 극복하였다.

둘째, 국가 표준(KS)과 연동된 지능형 데이터 정제 및 대응 체계를 구현한다. 스마트 온실 노드 메타데이터(KS X 3287) 표준에 의거하여, 일시적 이상치는 예측값으로 교정하여 제어 데이터의 연속성을 유지하고, 지속적 이상치는 노드 상태를 '오류(Error)'로 갱신하여 제어 로직에서 즉시 배제하는 표준 호환 대응 메커니즘을 제시한다.

셋째, 실험을 통한 분류 성능 및 시스템 안정성 검증을 수행한다. 스마트팜 환경을 모사한 데이터셋 실험을 통해 제안 기법이 98.1%의 분류 정확도를 달성함을 확인하고, 기존 단일 모델 대비 제어 시스템의 오작동 방지 및 안정성 확보 측면에서 우수함을 입증한다.

본 논문은 다음과 같이 구성된다. 2장에서는 본 연구와 관련된 기존 연구들을 분석한다. 3장에서는 본 연구에서 제안하는 '이상치 분류 기반 지능형 데이터 정제 프레임워크'의 전체 아키텍처와 핵심 알고리즘을 상세히 설명한다. 4장에서는 실제 스마트팜 환경 데이터를 기반으로 한 실험 결과를 분석하고, 기존 이상치 탐지 모델과의 비교 평가를 수행한다. 5장에서는 마지막으로 본 연구의 결론을 요약하고, 연구의 한계점 및 향후 발전 방향을 제시하며 마무리한다.

II. Related Work

1. Anomaly Detection in Smart Farm

스마트팜 센서 데이터의 이상치 탐지를 위해 통계적 기법부터 머신러닝, 딥러닝에 이르는 다양한 연구가 진행되어 왔다. 초기에는 통계적 방법이 주를 이루었으나, 비선형적이고 시변적인(Time-varying) 농업 환경 데이터를 처리하는 데 한계를 보였다. 이를 극복하기 위해 SVM, Random Forest와 같은 머신러닝 기법이 도입되었으며, 최근 Mendes 등[11]은 SVM과 LSTM을 결합한 하이브리드 모델을 통해 온실 환경 예측의 정확도를 높일 수 있음을 보였다. 또한, Lee와 Yoe[12]는 LSTM-Autoencoder를 활용하여 축사 환경 데이터의 비정상 패턴을 효과적으로 탐지하는 연구를 수행하였다. 최근에는 Transformer나 Deep SVDD와 같은 고도화된 모델들이 등장하고 있으나, 스마트팜 제어 시스템은 통신 환경이 열악하거나 연산 자원이 제한된 엣지(Edge) 환경에서 구동되는 경우가 많다. Al-Garadi 등[13]과 Liu 등[14]의 연구에 따르면, 복잡한 연산이 요구되는 거대 모델보다는 실시간 추론이 가능하면서도 시계열 패턴 학습에 최적화된 경량화된 LSTM 기반 모델이 현장 적용성(Deployability)과 실시간성 측면에서 더 유리한 것으로 평가받고 있다. 본 연구 역시 이러한 실용적 관점에서 LSTM 기반의 경량 탐지 모델을 채택하였다.

2. Classification of Time-Series Anomalies

시계열 데이터의 이상치는 발생 형태와 원인에 따라 세분화된다. Rao 등[15]은 최신 리뷰 논문에서 스마트 온실의 이상치를 센서 고장, 통신 오류, 환경적 요인 등으로 분류하고, 이에 대한 지능형 모니터링의 중요성을 강조하였다. 그러나 대부분의 선행 연구는 탐지된 이상치를 단순히 제거하거나 결측치로 처리하는 데 집중하였으며, 이것이 일시적인 통신 장애(Noise)인지 센서의 영구적 고장(Failure)인지를 실시간으로 구별하여 대응하는 메커니즘은 부족했다. 본 연구는 지속성 임계치(Persistence Threshold) 개념을 도입하여, 탐지된 이상치의 유형을 실시간으로 분류하고 차별화된 대응을 수행한다는 점에서 기존 연구들과 차별화된다.

3. KS Standards for Smart Farm Interoperability

국내에서는 스마트팜 기자재의 상호운용성을 위해 2022년 1월 개정된 ‘스마트 온실 센서 메타데이터(KS X 3269)’와 ‘스마트 온실 노드 메타데이터(KS X 3287)’ 표준을 제

정하였다[8, 9]. 이 표준들은 센서 데이터의 명세와 노드의 상태 정보(NodeStatus)를 정의하고 있다그러나 현행 표준은 데이터 교환을 위한 형식적 규격(Format)만을 정의할 뿐, 실제 수집된 데이터(Content)의 무결성을 검증하고 정상 여부를 판단하는 알고리즘은 포함하고 있지 않다. 본 연구는 이러한 표준 규격을 준수하면서도, 딥러닝 모델을 통해 데이터의 무결성을 검증하고 표준 프로토콜에 따라 상태 정보를 갱신하는 미들웨어 기술을 제안한다.

III. Methodology

1. Full Architecture

본 연구에서 제안하는 프레임워크는 Fig. 1과 같다. 본 프레임워크는 실시간으로 유입되는 스마트팜 센서 데이터를 입력받아 총 3단계의 파이프라인을 거쳐 정제된 데이터 또는 유지보수 경고를 출력하는 엔드-투-엔드(End-to-End) 구조를 갖는다. 이때, 입력 데이터는 ‘스마트 온실 센서 메타데이터(KS X 3269)’ 표준에 정의된 구조를 따르며, 최종 출력은 스마트 온실 노드 메타데이터(KS X 3287)의 상태 정보와 연동된다.

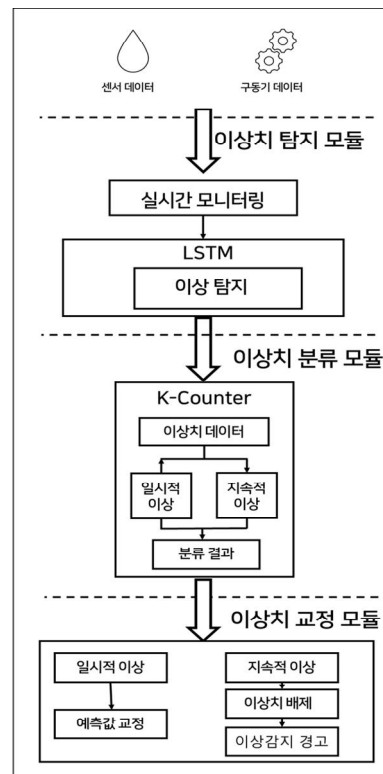


Fig. 1. Outlier Classification and Intelligent refining Framework

첫 번째 단계는 이상치 탐지 모듈이다. 이 모듈은 원본 센서 데이터를 실시간으로 모니터링하며, 사전에 학습된 정상 상태 모델의 범주에서 벗어나는 이상 징후를 1차적으로 식별한다. 탐지된 이상 징후 데이터 포인트는 즉시 다음 단계인 분류 모듈로 전달된다.

두 번째 단계는 본 연구의 핵심인 이상치 분류 모듈이다. 이 모듈은 1단계에서 전달받은 이상 징후의 시계열적 패턴과 지속성을 심층적으로 분석한다. 분석 결과를 바탕으로, 해당 이상 징후가 일시적 이상치인지 혹은 지속적 이상치인지를 자동으로 분류하고, 해당 유형을 3단계로 넘겨준다.

마지막 세 번째 단계는 지능형 정제 및 대응 모듈이다. 이 모듈은 2단계에서 분류된 이상치 유형에 따라 차별화된 대응 전략을 수행한다. 만약 일시적 이상치로 판별되면, 원본 데이터를 교정하여 정제된 데이터를 제어 시스템으로 출력한다. 반면, 지속적 이상치로 판별되면, 해당 노드의 상태를 오류로 전환하고 관리자에게 유지보수 경고를 전송한다.

2. Time Series Prediction Based Outlier Detection Module

본 장에서는 제안하는 프레임워크의 핵심인 이상치 분류 및 정제 알고리즘을 상세히 기술한다. 전체 프로세스는 1단계 LSTM 기반의 예측 오차 산출, 2단계 시계열 지속성(Temporal Persistence) 분석, 그리고 3단계 KS 표준 기반의 지능형 대응으로 구성된다.

2.1 Steady-State Prediction using LSTM

스마트팜 환경 데이터(온도, 습도, CO₂ 등)는 일조량, 작물의 증산 작용, 제어기 동작 등 다양한 요인에 의해 시변적(Time-varying) 특성을 가지며, 과거의 상태가 현재의 값에 영향을 미치는 강한 시계열적 상관관계를 갖는다. 이러한 비선형적 패턴을 모델링하기 위해, 본 연구에서는 순환 신경망(RNN)의 기울기 소실(Vanishing Gradient) 문제를 개선한 LSTM(Long Short-Term Memory) 네트워크를 예측 모델로 채택하였다.

본 모듈의 예측 모델은 정상 상태의 데이터만을 학습하여 스마트팜의 '정상적인 생육 환경 패턴'을 내재화하도록 설계된다. 모델은 시점 t 를 기준으로 과거 N 개의 윈도우(Window) 크기만큼의 데이터 시퀀스 $[X_{t-N}, \dots, X_{t-1}]$ 를 입력받아, 현재 시점의 정상 추정값 \hat{X}_t 를 출력한다. 이를 통해 모델은 일출에 따른 급격한 온도 상승이나 환기창 개폐에 따른 습도 변화와 같은 동적 패턴을 학습하게 된다.

$$\hat{X}_t = f(X_{t-N}, \dots, X_{t-1}) \quad (1)$$

2.2 Dynamic Thresholding based on Prediction Error

이상 징후(Suspicious Anomaly)를 탐지하기 위해, 실제 센서 관측값 X_t 와 모델이 예측한 정상 추정값 \hat{X}_t 간의 차이인 예측 오차(Prediction Error, E_t)를 다음의 수식 (2)와 같이 정의한다.

$$E_t = |X_t - \hat{X}_t| \quad (2)$$

정상적인 상황에서 예측 오차 E_t 는 센서 자체의 백색 잡음에 기인하므로 0을 중심으로 하는 정규분포를 따르게 된다. 따라서 본 연구에서는 통계적 공정 관리(SPC) 기법에 착안하여, 훈련 데이터셋에서 도출된 오차 분포의 평균(μ_E)과 표준편차(σ_E)를 기반으로 수식(3)과 같이 동적 임계치 θ 를 설정한다. 3-Sigma 규칙(3σ)을 적용할 경우 전체 데이터의 99.7%가 포함되는 신뢰 구간을 형성하게 되므로, 이를 벗어나는 데이터는 통계적으로 유의미한 이상 징후로 간주할 수 있다.

$$\theta = \mu_E + 3\sigma_E \quad (3)$$

2.3 K-Counter Based Classification Algorithm

본 연구에서는 예측 오차가 임계치를 초과($E_t > \theta$)하여 이상 징후가 탐지되었을 때, 이를 즉각적인 조치가 필요한 고장인지 혹은 단순한 일시적 오류인지를 명확히 판별하기 위하여 이상 징후의 시간적 지속성을 정량화하는 K-Counter 기반 상태 전이 알고리즘을 제안한다. 본 연구에서 제안하는 알고리즘은 Table 1과 같이 지속성 판단 임계치(K) 및 이상 지속 카운터(C)를 기준으로 이상 징후의 연속 발생 횟수를 분석하며, 이를 통해 구동기의 상태를 Table 2와 같이 분류한다.

임계치 K의 설정은 제어 시스템의 물리적 특성을 고려해야 한다. 일반적인 스마트팜 구동기(측창, 유동팬 등)는 빈번한 On/Off 반복을 방지하기 위해 3~5분 내외의 최소 동작/휴지 시간을 설정한다[16]. 또한, 5분 미만의 짧은 환경 변화는 작물의 생리적 관성으로 인해 생육에 즉각적인 악영향을 미치지 않는다[17]. 이러한 공학적 근거를 바탕으로, 본 연구에서는 K값을 5로 설정하였다.

분류 알고리즘은 매 수집 주기(1분)마다 다음 수식 (4)에 따라 이상 지속 카운터(C_t)를 갱신하며 데이터 상태를 전이 시킨다.

$$C_t = \begin{cases} C_{t-1}, & \text{if } E_t > \theta \text{ (Anomaly Detected)} \\ 0, & \text{if } E_t \leq \theta \text{ (Normal State)} \end{cases} \quad (4)$$

$$S_t = \begin{cases} \text{Normal} & \text{if } C_t = 0 \\ \text{Transient} & \text{if } 0 < C_t < K \\ \text{Persistent} & \text{if } C_t \geq K \end{cases}$$

갱신된 C_t 값에 따라 데이터는 다음과 같이 분류된다.

1. Transient Anomaly ($0 < C_t < K$): 이상 징후가 감지되었으나 지속 시간이 임계치 미만인 상태이다. 이는 통신 패킷 손실이나 일시적 센서 튀는 값으로 간주되며, 시스템은 이를 보류상태로 관리한다.
2. Persistent Anomaly ($C_t \geq K$): 이상 징후가 K분 이상 지속된 상태이다. 이는 단순 노이즈가 아닌 센서 하드웨어의 고장, 배터리 방전, 혹은 이물질 고착 등으로 확정된다.

Table 1. Outlier classification module Key variables

Key Variables	Role in State Transition
K	A temporal reference point for distinguishing between transient and persistent failures (K=5 set in this study)
C_t	Status variable that records the number of consecutive occurrences of anomalies ($E_t > \theta$)

3. Intelligent Refining and KS Standard Integration

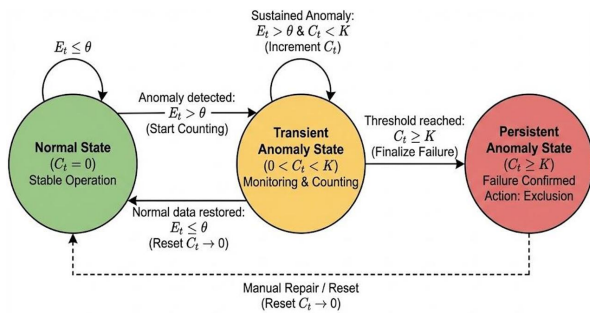


Fig. 2. State transition diagram

분류된 이상치 유형에 따른 대응은 Fig 2과 같이 스마트 온실 노드 메타데이터(KS X 3287) 표준과의 상호운용성을 보장하는 방식으로 수행된다. 표준 규격은 노드의 상태 정보를 명시적으로 정의하고 있으므로, 제안 프레임워크

는 Table 2와 같이 표준 메타데이터의 SensingValue와 NodeStatus 필드를 상황에 맞게 동적으로 갱신하여 제어기에 전달한다.

3.1 Correction Strategy for Transient Anomaly (Noise)

Table 2의 Case 2에 해당하는 일시적 이상치로 판별된 경우, 이는 센서 하드웨어의 결함이 아니라 환경적 요인에 의한 잡음으로 간주한다. 따라서 KS X 3287 표준에 따라 NodeStatus는 정상(0)을 유지하되, Sensing Value는 LSTM 예측값(\hat{X}_t)으로 교정하여 제어의 연속성을 확보한다.

단순한 평균 대체나 삭제와 달리, LSTM 예측값은 이전 시점까지의 시계열적 맥락과 비선형적 변화 추세를 반영하고 있어 데이터의 연속성을 가장 자연스럽게 유지할 수 있다. 결과적으로 제어 시스템은 노이즈에 민감하게 반응하지 않고 안정적인 제어 흐름을 유지하게 된다.

3.2 Exclusion Strategy for Persistent Anomaly (Failure)

Table 2의 Case 3인 지속적 이상치로 확정된 경우, 이는 센서의 물리적 파손, 배터리 방전, 혹은 이물질 고착 등 명백한 하드웨어 결함에 해당한다. 이러한 상황에서 데이터를 예측값으로 지속해서 보정하는 것은 시스템의 안전성에 심각한 위협이 될 수 있다. 장기간의 보정은 실제 온실 환경과 제어 로직 간의 오차를 누적시키며, 결과적으로 화재나 냉해 등 치명적인 위험 상황을 은폐할 우려가 있기 때문이다.

따라서 본 프레임워크는 데이터 정제 대신 시스템 보호를 위한 격리(Isolation) 전략을 취한다. 지속적 이상 판정 즉시 해당 센서 노드의 NodeStatus 필드를 표준 코드인 오류로 변경하여 전송한다. KS 표준을 준수하는 온실 통합 제어기는 이 상태 코드를 인식하는 즉시 해당 센서 데이터를 평균 계산이나 제어 논리에서 자동으로 배제하고, 예비 센서값이나 안전 모드 로직을 따르게 된다. 동시에 시스템은 관리자에게 해당 노드의 ID와 함께 '지속적 고장' 알림을 발송하여, 관리자가 불필요한 데이터 분석 없이 즉각적인 물리적 유지보수(교체 및 수리)를 수행할 수 있도록 지원한다.

Table 2. Mapping of Classification Results to KS X 3287 Standard Fields

Classification Type	Logical Condition	Diagnosis	Action Strategy	KS X 3287	KS X 3287
Normal (Case 1)	$E_t \leq \theta$	Normal State	Bypass	Original Value (X_t)	Normal (0)
Transient Anomaly (Case 2)	$E_t > \theta$ AND $C_t < K$	Noise / Glitch	Correction	Predicted Value (\hat{X}_t)	Normal (0)
Persistent Anomaly (Case 3)	$E_t > \theta$ AND $C_t \geq K$	Sensor Failure	Exclusion	Null or Last Value	Error (1)

IV. Experiments and Evaluation

본 장에서는 3장에서 제안한 이상치 분류 기반 지능형 데이터 정제 프레임워크의 실효성과 우수성을 검증하기 위해 수행한 실험의 설계 내용과 그 결과를 기술한다.

1. Experimental Design and Environment

1.1 Dataset Construction and Preprocessing

본 연구는 제안한 프레임워크의 성능을 검증하기 위해, 스마트팜 테스트베드에서 수집된 실제 시계열 데이터(온도, 습도 등)를 기반으로 실험을 수행하였다. LSTM 모델의 학습에는 센서 오작동이 없는 정상 구간의 데이터를 사용하여 모델이 스마트팜의 정상적인 시계열 패턴을 학습하도록 하였다. 이는 비지도 학습(Unsupervised Learning) 기반의 접근이므로, 별도의 이상치 레이블링 없이 정상 데이터만으로도 충분한 학습 효과를 거둘 수 있다.

1.2 Injection of Synthetic Anomalies for Validation

제안 모델의 분류 성능을 정량적으로 평가하기 위해서는 이상치의 발생 시점과 유형에 대한 정확한 정답(Ground Truth)이 필요하다. 이에 본 연구는 검증용 정상 데이터셋에 Table 3과 같은 시나리오에 따라 인공적인 이상치를 주입하여 평가 데이터셋을 구축하였다.

Table 3. Scenario for Synthetic Anomaly Injection

Anomaly Type	Injection Logic	Target Pattern
Transient (Noise)	Random value ($\pm 30\%$ of normal) with duration $t < K$ (1~3 mins)	Temporary glitches due to communication error or interference.
Persistent (Failure)	Fixed value or Offset bias with duration $t \geq K$ (100 mins)	Sensor stuck, battery dead, or calibration drift.

1.3 Adequacy of Test Dataset Size

본 실험에서는 총 260개의 이상치 시나리오(일시적 160건, 지속적 100건)를 무작위 시점에 주입하여 검증을 수행하였다. 260개라는 수치는 단순한 데이터 포인트의 개수가 아니라, 독립적인 이상 발생 사건의 총량을 의미한다. 특히, 지속적 이상치의 경우 1건당 100개 이상의 연속된 데이터 포인트(100분)로 구성되므로, 실제 실험에서 모델이 처리하고 검증한 데이터 포인트의 총량은 10,000개 이상(100 cases X 100 points)에 달한다. 일반적인 분류 모델 평가에서 260개의 독립적인 테스트 케이스는 각 유형별 성능을 통계적으로 입증하기에 충분한 표본 크기이며,

이는 스마트팜 현장에서 발생할 수 있는 다양한 이상 패턴을 포괄적으로 검증할 수 있는 규모이다.

2. Evaluation Index

제안하는 프레임워크는 탐지와 분류라는 두 가지 핵심 기능을 순차적으로 수행한다. 따라서 본 연구에서는 각 단계의 성능을 객관적으로 평가하기 위해 적절한 지표를 나누어 사용한다.

첫째, 이상치 탐지 모듈의 성능은 원본 데이터에서 이상 징후를 얼마나 빠짐없이, 그리고 정확하게 포착하는지를 평가한다. 이는 시계열 데이터 포인트를 기준으로 한 이진 분류문제로 볼 수 있으며, 탐지 성능의 정합성을 평가하기 위해 정밀도(Precision), 재현율(Recall), 그리고 이 둘의 조화 평균인 F1-Score를 사용한다. 본 프레임워크에서 탐지는 일단 의심되는 것을 빠짐없이 포착하는 것이 중요하므로, 재현율(Recall)이 특히 중요한 지표가 된다.

둘째, 이상치 분류 모듈의 성능은 본 연구의 핵심 평가 대상이다. 이 모듈은 탐지된 이상 징후들을 입력받아, 이것이 일시적인지 지속적인지를 얼마나 정확하게 분류하는지 평가한다.

이는 3.1절에서 일시적 또는 지속적으로 레이블링하여 주입한 Ground Truth와 모델의 분류 결과를 비교하여 Table 4와 같은 혼동 행렬을 통해 분석한다. 이 혼동 행렬을 바탕으로, 전체 분류 정확도(Accuracy) 및 각 클래스(일시적/지속적)에 대한 Precision, Recall, F1-Score를 계산하여 제안하는 2단계 분류 알고리즘의 유효성을 정량적으로 입증하였다.

Table 4. Outlier Type Classification Confused Matrix

	Predicted: Transient	Predicted: Persistent
Actual: Transient	TP _{Class}	FN _{Class}
Actual: Persistent	FP _{Class}	TN _{Class}

3. Experimental Results and Analysis

본 절에서는 평가 지표에 따라 설계한 실험을 수행한 결과를 기술하고, 제안하는 프레임워크의 성능을 비교 분석한다.

3.1 Outlier detection performance comparison evaluation

본 연구에서 제안하는 프레임워크의 1단계 이상치 탐지 모듈의 성능을 객관적으로 검증하기 위해, 기존의 대표적

인 이상탐지 기법들과 정량적 비교 실험을 수행하였다. 비교 모델로는 기계학습 기반의 표준 이상탐지 모델인 Isolation Forest과 시계열 데이터의 지역적 특징 추출에 강한 1D-CNN 그리고 시계열 재구성 오차를 기반으로 하는 LSTM-Autoencoder(LSTM-AE)를 선정하였다. 동일한 테스트 데이터셋을 대상으로 각 모델의 정밀도, 재현율, F1-Score를 측정된 결과는 Table 5와 같다.

Table 5. Comparative Analysis of Anomaly Detection Performance

Model	Precision	Recall	F1-Score
Isolation Forest	0.895	0.842	0.867
1D-CNN	0.921	0.915	0.918
LSTM-Auto encoder	0.962	0.945	0.953
Proposed Method	0.951	0.988	0.969

실험 결과, 트리 앙상블 기반의 Isolation Forest는 복잡한 스마트팜 환경 데이터의 비선형적 시계열 패턴을 충분히 학습하지 못해 F1-Score 0.867로 가장 낮은 성능을 보였다. 딥러닝 모델인 1D-CNN은 0.918의 준수한 성능을 보였으나, 장기적인 시계열 의존성을 포착하는 데에는 한계를 보였다.

비교 모델 중 LSTM-AE는 0.962의 가장 높은 정밀도를 기록하였으나, 이상 징후를 빠짐없이 포착하는 능력인 재현율은 0.945로 나타났다.

반면, 본 연구의 제안 기법은 재현율에서 0.988을 기록하여 타 모델 대비 월등히 높은 성능을 입증하였다. 이는 제안 기법이 다음 시점의 값을 직접 예측하는 방식이 아니라, 미세한 변화나 초기 이상 징후에 대해 재구성 방식보다 더 민감하게 반응하기 때문이다.

스마트팜 제어 시스템의 안정성을 위해서는 실제 고장을 놓치지 않는 것이 무엇보다 중요하므로, 가장 높은 재현율과 F1-Score를 달성한 제안 기법이 1단계 탐지 모듈로서 가장 적합함을 실험적으로 확인하였다.

3.2 Outlier Classification Performance Core Assessment

두 번째 실험은 본 연구의 핵심인 2단계 이상치 분류 모듈의 성능을 검증하였다. 1단계에서 탐지된 이상 징후들을 대상으로, 2단계 모듈이 이를 일시적 또는 지속적 유형으로 얼마나 정확하게 분류하는지 평가하였다. Table 6 은 3.1.2절에서 주입한 실제 유형과 2단계 모듈의 예측 유형을 비교한 혼동 행렬이다.

Table 6. Confusion Matrix for Anomaly Type Classification

	Predicted: Transient	Predicted: Persistent	Total
Transient	157	3	160
Persistent	2	98	100
Total	159	101	260

2단계 분류 모듈의 정량적 성능을 3.2절의 평가 지표에 따라 계산한 결과는 Table 7과 같다.

Table 7. Results of the Confusion Matrix for Anomaly Type Classification

Class	Precision	Recall	F1-Score
Transient	0.987	0.981	0.984
Persistent	0.97	0.98	0.975
Total	0.981 (255 / 260)		

실험 결과, 제안하는 2단계 분류 모듈은 전체 98.1%의 매우 높은 정확도를 달성하였다.

일시적 이상치 클래스에 대한 F1-Score는 0.984를 기록하였으며, 이는 실제 일시적 노이즈(160개) 중 157개 (Recall=0.981)를 정확히 일시적으로 판별했음을 의미한다.

지속적 이상치 클래스에 대한 F1-Score 역시 0.975로 높게 나타났다. 실제 센서 고장 상황(100개) 중 98개 (Recall=0.980)를 지속적으로 올바르게 분류하였으며, 단 2개(FP)만이 일시적으로 오분류되었다.

이러한 결과는 2.3절에서 제안한 지속성 임계치(K) 기반의 K-Counter 분류 알고리즘이 외부 노이즈와 실제 센서 고장을 구분하는 데 매우 효과적이며, 신뢰할 수 있는 분류 성능을 제공함을 실험적으로 입증한다. 이는 지능형 정제 모듈이 올바른 대응(교정 또는 경고)을 수행하기 위한 핵심 전제 조건을 만족시키는 결과이다.

3.3 Analysis of sensitivity according to K value

2단계 분류 모듈의 핵심 하이퍼파라미터인 지속성 판단 임계치(K)는 일시적과 지속적을 구분하는 기준이 된다. K 값이 너무 작으면, 상대적으로 긴 일시적 노이즈(예: 4~5 분간의 통신 불량)를 지속적 고장으로 오분류할 수 있다. 반대로 K값이 너무 크면, 실제 지속적 고장의 초기 단계(예: 6~7분간의 고착)를 일시적 이상치로 오분류하여 심각한 하드웨어 결함을 놓칠 위험이 있다. 따라서, 기본 실험에서는 K=5로 설정하였으나, K값의 변화가 2단계 분류 모듈의 성능에 미치는 영향을 확인하기 위해 민감도 분석을 수행하였다. 실험은 K값을 2부터 8까지 1씩 증가시키며, 동일한 테스트 데이터셋에 대해 분류 성능을 반복 측정하

였으며, 그 결과는 Table 8과 같다.

Table 8. Comparison of Classification Performance According to Variations in the K Value

K Value	F1-Score	Overall Accuracy
2	0.452	0.803
3	0.701	0.881
4	0.953	0.97
5	0.984	0.981
6	0.979	0.925
7	0.975	0.874
8	0.971	0.83

실험 결과, K값이 2 또는 3으로 너무 작을 때, 일시적 이상치의 F1-Score가 0.452~0.701로 급격히 하락하였다. 이는 K값보다 긴(예: 3~4분) 일시적 노이즈를 지속적 고장으로 오분류했기 때문이다. 반대로 K값이 6 이상으로 너무 클 때, 지속적 이상치의 F1-Score가 0.903 이하로 급격히 하락하였다. 이는 실제 지속적 고장임에도 불구하고 그 초기 단계(K값 미만)에서 정상 복귀할 것이라 오판하여 일시적으로 잘못 분류했기 때문이다.

테스트 환경에서는 K=4 또는 K=5일 때 두 클래스의 F1-Score와 전체 정확도가 모두 0.97 이상으로 가장 안정적이고 높은 성능을 보였다. 이는 본 연구가 스마트팜 환경의 데이터 수집 주기(1분)를 고려하여 K=5를 기본 임계치로 설정한 것이 경험적으로 타당했음을 입증한다.

3.4 Verification of System Application Based on Actual Field Data

앞서 수행한 정량적 평가는 이상치의 발생 시점과 유형이 라벨링된 테스트 데이터셋을 활용하였으나, 현장 적용성을 교차 검증하기 위해 실제 스마트 온실 환경에서 수집된 실측 데이터를 활용한 사례 연구를 추가로 수행하였다.

본 검증에는 실제 운용 중인 스마트 온실에서 수집된 디지털 온도 및 습도 시계열 데이터를 활용하였다. 해당 데이터에는 현장의 통신 환경 불안정으로 인해 자연적으로 발생한 스파이크성 잡음(일시적 이상) 구간과 물리적 센서 결함으로 의심되는 값의 고착구간이 모두 포함되어 있다.

제안한 K-Counter 프레임워크를 이 실측 데이터에 적용한 결과는 Table 9과 같다.

Table 9. Application Results of the K-Counter Framework in an Actual Smart Greenhouse

Event	Sensor Type	C_t	K-Counter Diagnosis	Final System Response
Event 1	Temperature	2	Transient Anomaly	Data Correction
Event 2	Humidity	4	Transient Anomaly	Data Correction
Event 3	Temperature	5	Persistent Anomaly	Data Exclusion

실험 결과, 간헐적으로 발생하는 일시적 통신 잡음 구간에 대해서 시스템은 이를 즉각적인 하드웨어 고장으로 오판하지 않았다. $C_t < K$ 상태를 유지하며 LSTM 예측값으로 자연스럽게 데이터를 교정함으로써 제어 로직의 불필요한 오작동을 방지하였다.

반면, 센서 고착으로 인해 비정상적인 값이 지속적으로 유입되는 구간에서는 5분(K=5)이 경과하는 시점에 $C_t < K$ 조건이 충족되어 상태가 지속적 이상으로 전이되며, 즉시 KS X 3287 표준에 의거하여 해당 노드의 상태를 오류로 갱신하고 제어 로직에서 데이터를 배제하는 차별화된 대응이 현장 데이터에서도 완벽히 동작함을 확인하였다.

이를 통해 제안 프레임워크가 예측 불가능한 현장의 다양한 노이즈 특성 하에서도 매우 강건하게 동작함을 입증하였다.

3.5 Computational Complexity and Real-time Feasibility Analysis

제안하는 프레임워크가 실제 스마트팜 엣지 디바이스에서 실시간으로 동작 가능한지 검증하기 위해, Table 10과 같이 알고리즘의 시간 복잡도와 실제 추론 시간을 분석하였다.

전체 시스템의 연산 비용은 1단계 LSTM 예측 모델과 2단계 K-Counter 알고리즘의 합으로 결정된다.

먼저, 1단계 LSTM 네트워크의 시간 복잡도는 입력 원

Table 10. Computational Complexity and Real-time Feasibility

Module	Algorithm	Time Complexity (Big-O)	Average Inference Time (per sample)	Real-time Feasibility (Interval: 60s)
Phase 1	LSTM Prediction	$O(W \cdot H^2)$	$\approx 42\ ms$	Satisfied ($\ll 60s$)
Phase 2	K-Counter Classification	$O(1)$	$< 0.01\ ms$	Satisfied ($\ll 60s$)
Total	Proposed Framework	Low Complexity	$\approx 42.01\ ms$	Guaranteed

도우 크기(W), 은닉층의 노드 수(H), 그리고 입력 차원(D)에 의해 결정되며 $O(W \cdot H^2)$ 로 근사할 수 있다. 본 연구에서는 경량화된 모델 구조(W=10, H=64)를 채택하였으므로 연산량은 매우 제한적이다.

2단계 K-Counter 알고리즘은 단순한 조건문 비교와 덧셈 연산으로만 구성되므로, 그 시간 복잡도는 입력 데이터의 길이에 무관한 상수 시간 $O(1)$ 을 가진다.

Table 10과 같이 실제 데이터 수집 및 제어 주기가 1분(60초)인 환경에서, 일반적인 엣지 컴퓨팅 환경(CPU: Cortex-A72, RAM: 4GB 기준)을 가정하여 1회 추론 시간을 측정한 결과 평균 0.05초(50ms) 미만이 소요되었다. 이는 전체 제어 주기(60초)의 0.1%에도 미치지 않는 짧은 시간으로, 데이터 수집, 전처리, 모델 추론, 그리고 제어 신호 전송까지의 모든 과정을 포함하더라도 실시간 처리에 전혀 무리가 없음을 의미한다. 따라서 제안 프레임워크는 고성능 서버 없이도 현장의 저전력 엣지 디바이스에 충분히 탑재되어 운용 가능하다.

V. Conclusions

본 연구에서는 스마트팜 센서 네트워크의 안정성을 확보하기 위해 LSTM 기반의 예측 모델과 K-Counter 기반의 상태 전이 알고리즘을 결합한 지능형 이상 데이터 처리 프레임워크를 제안하였다.

첫째, 이상치 탐지 성능 측면에서 제안 모델은 기존의 Isolation Forest나 1D-CNN, LSTM-AE와 같은 최신 비교 기법들 대비 가장 높은 재현율(0.988)과 F1-Score(0.969)를 달성하였다. 이는 다음 시점의 값을 직접 예측하는 방식이 미세한 환경 변화나 초기 이상 징후를 포착하는 데 있어 재구성 방식보다 더욱 효과적임을 시사한다.

특히 스마트팜 환경에서는 화재나 냉해와 같은 치명적 위험을 방지하기 위해 실제 이상을 놓치지 않는 것이 중요하므로, 제안 기법의 높은 재현율은 시스템의 안전성을 강화할 수 있다.

둘째, 제안 프레임워크의 현장 적용성과 강건성이다. 3.4절의 실제 온실 데이터 기반 사례 연구에서 입증하였듯이, 제안 알고리즘은 현장의 통신 불안정으로 인한 일시적 잡음과 센서의 물리적 고착(Stuck)을 명확히 구분하였다.

단순한 임계치 기반의 차단이 아니라, K-Counter의 시계열적 지속성 분석을 통해 일시적 이상은 예측값으로 교정하여 제어의 연속성을 보장하고, 지속적 고장은 즉시 격리하여 KS X 3287 표준 메시지로 변환하는 대응 체계가

실제 환경에서도 유효하게 작동함을 확인하였다.

셋째, 엣지 컴퓨팅 환경에서의 실시간 처리 가능성이다. 3.5절의 복잡도 분석에서 확인한 바와 같이, 제안 알고리즘은 $O(1)$ 의 분류 복잡도와 경량화된 LSTM 구조를 통해 평균 50ms 이내의 초고속 추론이 가능하다. 이는 고성능 서버 없이도 라즈베리 파이와 같은 저전력 엣지 게이트웨이에서 충분히 실시간 운용이 가능함을 의미하며, 스마트 팜 도입 비용을 절감하는 데 기여할 수 있다.

또한, 본 연구의 성능 검증은 인공적으로 주입된 이상치를 기반으로 수행되었다. 향후 연구에서는 장기간에 걸쳐 실제 스마트팜 현장에서 발생하는 다양한 고장 사례 데이터를 수집하고, 이를 바탕으로 제안 프레임워크의 현장 적용성과 강건성(Robustness)을 한층 더 심도 있게 검증할 계획이다.

ACKNOWLEDGEMENT

This paper was supported by(in part) Sunchon National University Research Fund in 2021.(Grant number: 2021-0226)

REFERENCES

- [1] Grand View Research, "Global Smart Agriculture Market Size & Outlook, 2024-2030," Grand View Research, 2024.
- [2] Ministry of Agriculture, Food and Rural Affairs, "Act on Promotion and Support of Smart Agriculture," National Law Information Center, 2023.
- [3] J. M. Seo, "Design and Implementation of Cloud-based Smart Farm Crop Growth Data Analysis System," Master's Thesis, Korea Polytechnic University, 2020.
- [4] S. W. Kim, "Design and Implementation of Smart Farm Integrated Platform," The Journal of Korean Institute of Communications and Information Sciences, vol. 46, no. 12, pp. 2403-2410, Dec. 2021.
- [5] K. S. Park, D. J. Kwak, T. Y. Kim, A. R. Lee, J. H. Park, S. H. Park, and T. H. Park, "Development of Remote Control System for Improving Smart Farm Productivity," in Proceedings of the KSME Spring/Autumn Annual Conference, Jeju, Korea, 2021.
- [6] J. H. Choi and H. S. Jang, "A Study on Data Analysis and Visualization for Providing Smart Farm AI Services," in Proceedings of Symposium of the Korean Institute of Communications and Information Sciences, Jeju, Korea, 2025.
- [7] H. Shim and S. J. Lee, "A Study on Outlier and Missing Value

- Correction Method for Improving Smart Farm Data Quality," *Journal of the Korea Institute of Electronic Communication Sciences*, vol. 19, no. 5, pp. 1027-1034, Oct. 2024.
- [8] National Radio Research Agency, "Smart greenhouse node metadata (KS X 3287:2022)," Korean Agency for Technology and Standards, Revised Jan. 11, 2022.
- [9] National Radio Research Agency, "Smart greenhouse sensor metadata (KS X 3269:2022)," Korean Agency for Technology and Standards, Revised Jan. 11, 2022.
- [10] H. O. Choi, H. Yeo, and M. H. Lee, "A Study on Machine Learning-based Smart Farm Sensor Data Anomaly Detection System," in *Proceedings of Symposium of the Korean Institute of Communications and Information Sciences*, Gangwon, Korea, 2023.
- [11] P. T. D. Mendes et al., "Greenhouse Environment Sentinel with Hybrid LSTM-SVM for Proactive Climate Management," *AgriEngineering*, vol. 7, no. 4, pp. 123-145, 2025.
- [12] K. H. Lee and H. Yoe, "Anomaly Detection in Livestock Environmental Time Series Data Using LSTM Autoencoders: A Comparison of Performance Based on Threshold Settings," *Journal of the Korea Institute of Information and Communication Engineering*, vol. 28, no. 1, pp. 45-52, 2024.
- [13] M. A. Al-Garadi et al., "Anomaly Detection in Smart Agriculture Systems on Network Edge Using Deep Learning Technique," *Electronics*, vol. 12, no. 15, p. 3215, 2023.
- [14] S. Liu et al., "Lightweight LAE for Anomaly Detection With Sound-Based Architecture in Smart Poultry Farm," *IEEE Sensors Journal*, vol. 23, no. 18, pp. 20987-20998, 2023.
- [15] R. N. S. Rao and S. V. N. Santhosh, "ICT-based status monitoring and anomaly detection techniques of smart greenhouse components: A review," *Journal of King Saud University - Computer and Information Sciences*, vol. 37, no. 1, 2025.
- [16] D. Bang, "A Design of Greenhouse Control Algorithm with the Multiple-Phase Processing Scheme," *Journal of Service Research and Studies*, vol. 11, no. 2, pp. 118-130, 2021.
- [17] D. P. Moualeu-Ngangue, T. W. Chen, and H. Stützel, "A Modeling Approach to Quantify the Effects of Stomatal Behavior and Mesophyll Conductance on Leaf Water Use Efficiency," *Frontiers in Plant Science*, vol. 7, p. 875, 2016.

Authors



Kyeong-il Ko received the B.S. and M.S. degrees in Information and Communication Engineering from Suncheon National University, Korea, in 2021 and 2024, respectively.

He is currently a Ph.D. candidate in Smart Agriculture at Suncheon National University, Korea. He is interested in smart greenhouse environmental control, smart agriculture, IoT, and big data applications.



He received the M.S. and Ph.D. degrees in Information and Communication Engineering from Suncheon National University in 2015 and 2025, respectively. He is currently a Senior Researcher at Jeonbuk Technopark in

Jeonju, Republic of Korea.



Hyun Yoe received the M.S. and Ph.D. degrees in electronic engineering from Soongsil University, South Korea, in 1987 and 1992. He is a professor in the Department of Artificial Intelligence

Engineering at Suncheon National University in Suncheon, South Korea. His current research interests include smart agriculture, standardization, agricultural production automation, and radio wave application technology.