사물인터넷융복합논문지 Vol. 11, No. 6, pp. 233-240, 2025 https://doi.org/10.20465/KIOTS.2025.11.6.233

불량품 예측을 위한 이상치 탐지 기법 연구

이태오[1], 김태국[2*]
[1]국립부경대학교 데이터공학과 학생, [2]국립부경대학교 컴퓨터·인공지능공학부 교수

A Study on Anomaly Detection Methods for Defective Product Prediction

Tae-O Lee[1], Tae-Kook Kim[2*]
[1]Student, Department of Data Engineering, Pukyong National University
[2]Professor, School of Computer and Artificial Intelligence Engineering, Pukyong National University

요 약 본 연구는 설비 운전 데이터에 정확한 불량 레이블이 존재하지 않는 제조 환경에서 불량을 예측하기 위한 방법으로, 이상치(outlier)를 불량 발생의 잠재적 신호로 간주하여 대체 레이블(pseudo-label)을 생성하는 접근법을 제안한다. 이상치 탐지를 위해 K-Nearest Neighbor(KNN), Isolation Forest, Local Outlier Factor(LOF) 등 서로 다른 특성을 지닌 세 가지 알고리즘을 적용하였으며, 모델 신뢰성을 향상시키기 위해 다음의 두 가지 보강 전략을 비교·분석하였다: (1) 세 모델의 다수결 기반 이상치 판정 방식과 (2) 세 모델의 이상치 점수 평균을 활용한 방식이다. 원 데이터의 레이블 부재로 인한 직접적인 성능 평가의 한계를 보완하기 위해, 유사한 공정 특성을 갖는 타 설비에서 수집된 레이블 데이터에 모델을 적용하여 제안 방법의 타당성을 검증하였다. 실험 결과, 세 가지 접근법 중 이상치 점수 평균 방식이 가장 안정적이며 우수한 예측 성능을 보였다. 본 연구는 불량 레이블이 부족하거나 부재한 공정 데이터 환경에서도 실용적인 불량 예측이 가능함을 보여주며, 특히 압출 공정 기반 예지보전(predictive maintenance) 시스템 구축을 위한 실증적·방법론적 기초를 제공한다.

주제어 : 사물인터넷, 인공지능, 압출 공정, 이상치 탐지, 불량 예측, 예지보전, 비지도 학습

Abstract This study proposes an approach for predicting defects in manufacturing environments where accurate defect labels in equipment operation data do not exist. This approach considers outliers as potential signals of defect occurrence and generates pseudo-labels. For outlier detection, we employ three algorithms with different characteristics: K-Nearest Neighbor (KNN), Isolation Forest, and Local Outlier Factor (LOF). To enhance model reliability, we compare and analyze two reinforcement strategies: (1) a majority-vote-based outlier detection method across the three models and (2) an average of the outlier scores across the three models. To address the limitations of direct performance evaluation due to the lack of labeled raw data, we validated the proposed method by applying the model to labeled data collected from other equipment with similar process characteristics. Experimental results show that, of the three approaches, the average of the outlier scores demonstrated the most stability and superior predictive performance. This study demonstrates that practical defect prediction is possible even in process data environments where defect labels are insufficient or absent, and provides an empirical and methodological foundation for building a predictive maintenance system based on the extrusion process.

Key Words : Internet of Things(IoT), artificial intelligence(AI), Extrusion process; Anomaly detection; Defect prediction; Predictive maintenance; Unsupervised learning

본 연구는 2025년 부산빅데이터혁신센터의 데이터산업분야 지산학연협력 교과개설 및 PBL 사업 지원을 받아 수행되었음.
*교신저자 : 김태국(king@pknu.ac.kr)
접수일 2025년 11월 01일 수정일 2025년 11월 19일 심사완료일 2025년 11월 25일
www.kci.go.kr

1. 서론

1.1 연구 배경

최근 산업 현장에서는 센서 네트워크와 데이터 통신 기술이 발달함에 따라, 사물인터넷(Internet of Things, IoT)이 공장 자동화 및 공정 최적화에 광범위하게 활용되고 있다. 다양한 센서가 장비 및 생산 라인에 부착되어 압력, 온도, 진동, 속도 등의 정보를 실시간으로 수집하고, 이를 분석하여 공정 이상을 파악하거나 품질 변화를 예측하는 시스템이 구축되고 있다[1-5]. 또한 이러한 데이터를 축적해 빅데이터 분석 기법을 적용하면 공정 변수 간 상관관계를 파악하고, 원인 불명으로 여겨졌던 품질 문제나 장비 결함의 원인을 도출하는 등 보다 고도화된 제조 문제 해결이 가능해지고 있다. 이를 통해 제조 현장은 기존의 경험 중심 운영 방식에서 데이터 기반 의사결정 체계로 빠르게 전환되고 있다[6-8].

압출 공정은 플라스틱이나 고분자를 원하는 형태로 만들기 위해 널리 사용되는 대표적인 제조 방법이다. 이 공정에서는 압력, 온도, 속도와 같은 다양한 변수들이 동시에 작용하며, 작은 변화만으로도 제품 품질이 크게 달라질 수 있다. 따라서 불량을 줄이고 안정적인 생산을 유지하는 것은 매우 중요한 과제이다. 최근에는 센서와 데이터 수집 장치가 발달하면서 공정 데이터를 기반으로 품질을 예측하고, 장비 고장을 미리 발견하는 '예지보전' 기술이 활발히 연구되고 있다[9, 10]. 그러나 실제 공정 데이터에는 불량 여부가 표시되지 않는 경우가 많다. 이 때문에 전통적인 지도학습 기반의 품질 예측 모델을 적용하기 어렵고, 라벨이 없는 데이터를 어떻게 활용할 것인가가 중요한 문제로 남아 있다[11]. 이러한 한계 때문에, 정상 데이터를 기준으로 이상치를 탐지하는 비지도 학습 방법이 대안으로 주목받고 있다. 이상치는 정상 공정에서 벗어난 데이터를 의미하며, 이는 불량 발생의 신호로 간주할 수 있다. 따라서 이상치 탐지는 불량 라벨이 없는 상황에서도 품질 관리와 예지보전에 활용할 수 있는 중요한 수단이 된다[12-14].

1.2 연구 목적 및 기여

본 연구는 라벨이 없는 압출 공정 데이터를 대상으로 비지도 이상치 탐지를 활용해 불량 가능성을 나타내는 대체 라벨을 생성하고 이를 기반으로 예지보전 모델을 구축하는 방법을 제시한다. 주요 기여는 다음과 같다. 첫째, 라벨 부재 상황에서 적용 가능한 대체 라벨 생성 방안을 제안한다. 둘째, Isolation Forest, Local Outlier Factor, K-Nearest Neighbor을 비교하고, 두 가지 다수결 방식을 적용해 성능 개선을 시도한다. 셋째, 생성된 라벨을 활용해 시계열 기반 예지보전 모델을 학습하고 불량 조기 탐지 가능성을 검증한다. 이를 통해 라벨 부족 환경에서도 신뢰성 있는 예지보전 시스템 설계의 가능성을 제시한다.

2. 관련 연구

이상치 탐지는 예지보전 및 공정 모니터링에서 핵심적인 역할이다. 대표적인 이상치 탐지 알고리즘으로는 Isolation Forest, Local Outlier Factor(LOF), 그리고 K-Nearest Neighbor (KNN)이 있다. Isolation Forest는 무작위 분할을 통해 데이터를 고립시키는 과정을 기반으로 하며, 계산 효율성과 대규모 데이터 처리 능력으로 널리 활용되고 있다[12, 13]. 이후 다양한 연구에서 Isolation Forest의 변형 및 다른 이상치 탐지 기법과의 비교가 이루어졌으며[14, 15], 실제 데이터 셋에 적용할 때 여러 알고리즘을 조합하거나 선택할 필요성이 강조되었다[16]. LOF는 데이터의 국소 밀도를 이웃과 비교하여 이상 여부를 평가하는 방법으로, 통계적 공정 모니터링 분야에서 Independent Component Analysis와 같은 차원 축소 기법과 함께 활용된 바 있다 [9]. 한편, KNN 기반 이상 탐지 기법은 직관적이고 간단한 구조로 인해 다양한 산업 현장에서 적용이 가능하며, 최근에는 펌프 시스템과 같은 설비의 예지보전에 활용된 사례도 보고되고 있다 [18]. 이러한 이상 탐지 기법은 제조 및 공정 산업의 예지보전 연구와 긴밀히 연결된다. 특히, 압출 공정에서는 머신러닝 기반으로 공정 중 발생하는 재료의 열화를 모니터링하거나[9], 물리 기반 모델과 딥러닝을 결합한 그레이박스 접근법으로 용융 점도를 추정하는 소프트 센서 연구가 이루어지고 있다[10]. 또한, 적층 제조와 같은 유사한 연속 공정에서도 공정 데이터 기반 품질 예측 및 불량 탐지가 활발히 이루어지고 있으며 [19], 이러한 흐름은 공정 신뢰성 확보 및 불량률 저감과 직접적으로 연결된다. 예지보전은 다양한 산업 영역에서 큰 주목을 받고 있으며, 항공우주 산업[20]과 자동차 산업 [21]에서도 빅데이터와 머신러닝 기반 접근이 확대되고 있다. 다만, Nunes et al. 의 리뷰에 따르면 실제 산업 데이터는 라벨이 부족하거나 불균형한 경우가 많아, 기존의 지도학습 기반 방법론을 적용하기 어렵다는 도전

www.kci.go.kr

과제가 존재한다[22]. 이를 해결하기 위해 정상 데이터만을 활용하는 비지도 학습 기반 접근이 제안되고 있으며, Carrasco et al. 은 시계열 기반 비지도 이상 탐지 알고리즘 평가 프레임워크를 통해 이러한 방법론의 검증 가능성을 제시하였다[11]. 더 나아가, 대체 라벨(proxy label)이나 약한 라벨을 활용하여 모델을 학습 및 검증하려는 시도도 보고되고 있으며[16, 23], 이는 실제 산업 현장에서 데이터 라벨링 비용을 절감하고 적용성을 높이는 데 기여할 수 있다. 종합하면, 이상치 탐지 기법은 Isolation Forest, LOF, KNN 과 같은 대표 알고리즘을 중심으로 발전해 왔으며, 다양한 산업 응용 사례와 함께 데이터 라벨 부족 문제를 해결하기 위한 새로운 접근이 모색되고 있다. 이러한 연구들은 제조업 및 공정 산업 전반에서 예지보전 기술의 신뢰성과 적용성을 높이는 중요한 기반을 제공한다.

3. 연구 방법

3.1 연구 방법 소개

본 연구는 라벨이 존재하지 않는 압출 공정 데이터를 활용하여 불량 발생 가능성을 탐지하고, 이를 예지보전 모델로 확장하는 방법론 제시를 목표로 한다. 실제 산업 현장의 압출 공정 데이터는 대부분 불량 여부 라벨이 제공되지 않기 때문에, 전통적인 지도학습 기반 품질 예측 모델을 적용하기 어렵다. 따라서 본 연구에서는 비지도 학습 기반 이상치 탐지 기법을 활용하여 대체 라벨을 생성하고, 이를 바탕으로 예측 모델을 학습하는 접근을 취한다. 구체적인 연구 절차는 다음과 같다. (1) 확보된 압출 공정 데이터에 라벨이 존재하지 않기 때문에, 이상치 탐지 모델의 성능 검증에 어려움이 있다. 따라서 해당 데이터와 유사한 데이터를 준비한다. (2) Isolation Forest (iForest), Local Outlier Factor(LOF), K-Nearest Neighbor(KNN) 세 가지 알고리즘을 준비된 유사 데이터에 적용하여 성능을 비교해 본다. 본 연구에서 해당 세 모델을 선택하여 비교하는 이유는 각 모델은 이상치 탐지 기법에서 차이가 있기 때문이다. iForest은 트리 기반, LOF은 밀도 기반, 그리고 KNN은 거리 기반으로 각각 다른 기반의 알고리즘을 사용하여 이상치 점수를 측정한다. 특히, KNN의 이상치 점수는 데이터 포인트 x_i 가 이웃하는 k개의 최근접 점들과의 평균 거리를 통해 정의되며, 아래와 같이 계산된다.

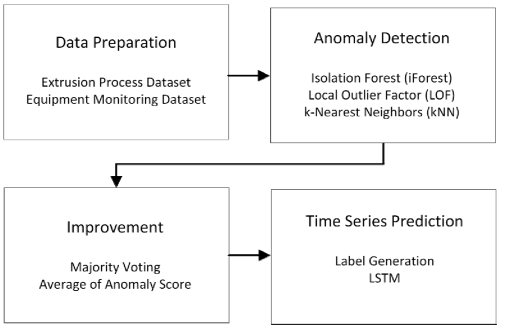
$$AnomalyScore(x_i) = \frac{1}{k}\sum_{j=1}^{k} d(x_i, x_j),$$

$$d(x_i, x_j) = \sqrt{\sum_{m=1}^{M} (x_{i,m} - x_{j,m})^2}$$

정상 데이터는 이웃과의 거리가 작아 낮은 점수를 갖는 반면, 이상치는 상대적으로 큰 거리를 나타내어 높은 점수를 가지게 된다.

(3) 라벨이 없는 상황에서 더욱 신뢰성 있는 모델을 만들기 위해 기존 모델을 앙상블·구조 변경하여 개선된 결과를 모색한다. 이는 다수결 방식, 이상치 점수 평균으로 두 가지 개선 방법을 제안한다. (4) 개선된 모델을 통해 생성한 대체 라벨을 시계열 예측 모델에 반영하여 불량 발생 가능성을 조기에 탐지하는 예지보전 모델을 설계한다.

[Fig. 1]은 본 연구의 진행 순서를 도식화한 것이다.



[Fig. 1] Flowchart

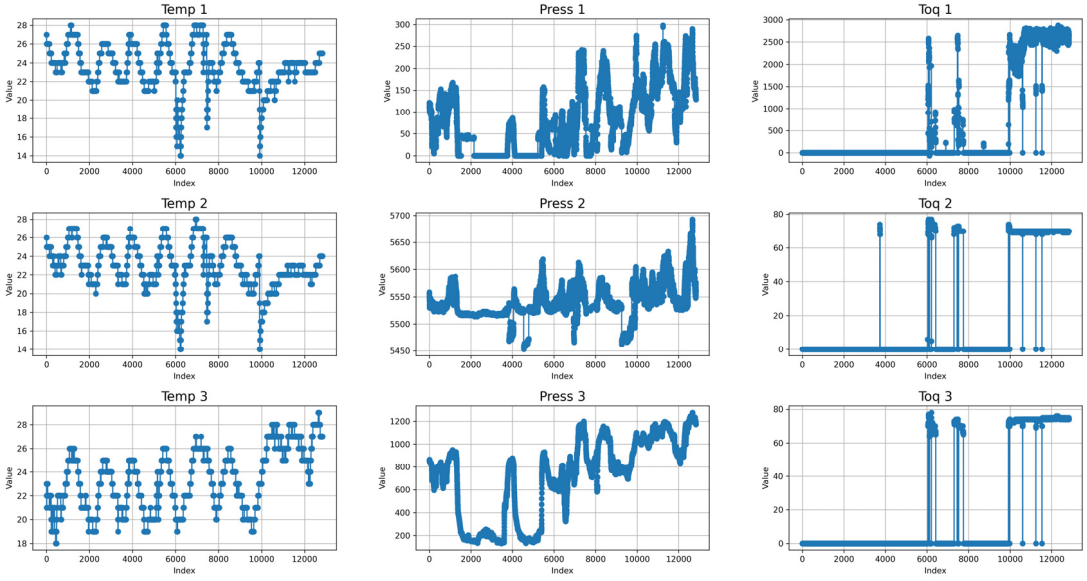
이와 같은 접근을 통해 본 연구는 라벨이 없는 산업 데이터 환경에서도 신뢰성 있는 불량 예측이 가능함을 입증하고, 향후 압출 공정 예지보전 시스템 구축의 기초를 마련하고자 한다.

3.2 데이터

본 연구에서 사용된 데이터는 두 가지로 구분된다. 첫째, 압출 공정 데이터는 시계열 형태로 구성되어 있으며, 예지보전 시스템 구축을 위한 주요 분석 대상으로 활용된다. 둘째, 장비 모니터링 데이터는 신뢰성 있는 라벨 생성 모델을 구축하기 위한 보조 데이터로 사용된다.

[Fig. 2]는 국내 OO 산업에서 제공받은 압출 공정 데이터의 일부로, (a)는 냉각수의 온도, (b)는 설비의 압력, (c)는 토크를 나타낸다. 각 데이터는 변수가 시간 순서대로 기록된 시계열 데이터이다. 해당 데이터는 다양한 외

www.kci.go.kr



(a) Temperature (b) Press (c) Toque

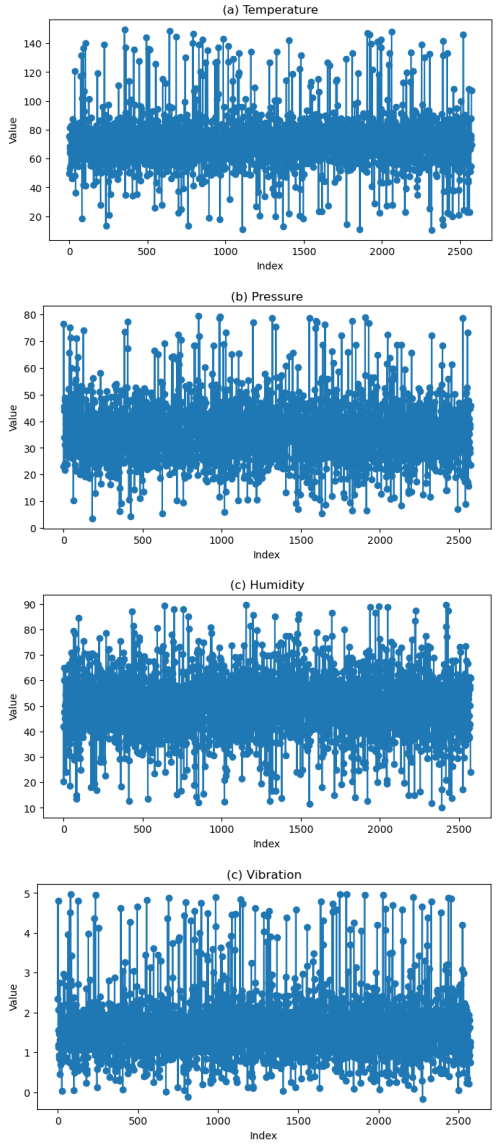
[Fig. 2] Data of Extrusion Process

부 요인으로 인해 변동성을 가지며, 이들이 복합적으로 작용해 제품 품질에 영향을 미친다. 주기적 변동, 스파이크, 드리프트 등 여러 패턴이 나타나지만, 이러한 변동이 실제 불량과 직결되는지는 단순 관찰이나 기본 통계 분석만으로는 파악하기 어렵다. 즉, 정상처럼 보이는 값이 잠재적 이상일 수도 있고, 급격한 변동이 반드시 불량을 의미하지 않을 수도 있다. 따라서 전통적인 규칙 기반 관리에는 한계가 있으며, 다변량 시계열의 복잡한 상관관계를 반영할 수 있는 머신러닝 기반 이상치 탐지가 필요하다. 본 연구는 이러한 이상치 탐지 결과를 불량 발생의 대체 라벨로 활용하고, 시계열 예측 모델과 결합해 예지보전 시스템을 설계한다.

[Fig. 3]은 본 연구에서 사용된 두 번째 데이터로, 캐글 (Kaggle)에서 제공하는 장비 모니터링 데이터이다 [24]. 이 데이터에는 온도, 압력, 진동, 습도 등 다양한 운전 상태 변수가 포함되어 있으며, 특히 실제 오작동 여부가 라벨로 제공된다는 점에서 중요한 특징을 가진다. 이것으로 인해 불량 라벨이 존재하지 않는 압출 공정 데이터와 달리, 본 데이터는 이상치 탐지 모델의 성능을 정량적으로 검증할 수 있는 기준 자료로 활용될 수 있다. 데이터의 구조적 특성도 차이를 보인다. [Fig. 2]의 압출 공정 데이터가 시계열 기반으로 연속성이 강하게 나타나는 반면, [Fig. 3]의 장비 모니터링 데이터는 개별 관측값 중심으로 기록되어 연속성이 상대적으로 약하다. 이러한 차이로 인해 압출 공정 데이터는 실제 예지보전 시스템 구축에 직접 적용되고, 장비 모니터링 데이터는 모델의 탐지 성능을 평가하고 대체 라벨의 신뢰성을 확보하는 보조적 역할을 수행한다. 더 나아가, [Fig. 3]의 (a), (b), (d)는 원 데이터를 제시한 [Fig. 2]의 (a), (b), (c)와 동일한 종류의 데이터로 구성되어 있어 상호 대체가 가능하다고 판단하였다. 이를 통해 원 데이터에 레이블이 부재하여 성능 평가가 어려웠던 문제를 효과적으로 보완할 수 있다. 본 연구에서는 이를 활용하여 Isolation Forest, LOF, KNN 모델의 이상치 탐지 성능을 비교·분석하였다.

3.3 실험 환경

본 연구는 Intel Core i9-14900K CPU와 NVIDIA RTX 4090 GPU 기반의 데스크톱 환경에서 수행되었다. 해당 환경은 불량품 예측 모델 구현에 필요한 이상치 탐지 및 시계열 분석을 원활히 처리할 수 있는 충분한 계산 자원을 제공한다. 이상치 탐지 실험에는 scikit-learn 라이브러리를 사용하여 iForest, LOF, KNN 알고리즘을 구현하였으며, 시계열 기반 예측 실험은 PyTorch 라이브러리를 이용하여 LSTM 모델로 구성하였다. 또한, 실험 전반은 Python 3.13.7 환경에서 Jupyter Notebook을 통해 수행하여 알고리즘을 구현, 시각화 및 결과 검증을 하였다. 다음 〈Tablel 1〉은 상세한 실험 환경을 보여준다.



[Fig. 3] Kaggle equipment monitoring data

〈Table 1〉 실험 환경

Category	Item	Specification
Hardware	CPU	Intel Core i9-14900K
	GPU	NVIDIA RTX 4090 24GB
	Memory	128GB RAM
	Storage	SSD 2TB
SoftWare	OS	Windows 11
	Python	3.13.7
	Scikit-learn	1.7.1
	PyTorch	2.8.0+cu129

4. 연구 결과

본 장에서는 앞서 제시한 연구 방법을 기반으로 수행한 결과를 제시한다. 먼저, 라벨이 존재하는 장비 모니터링 데이터를 이용하여 Isolation Forest(iForest), Local Outlier Factor(LOF), K-Nearest Neighbor(KNN) 모델의 기본 성능을 비교한다. 이후, 세 가지 성능 개선 방안(다수결, 이상치 점수 평균을 적용하여 성능 변화를 분석한다. 마지막으로, 성능 평가에서 가장 우수한 방법을 압출 공정 데이터에 적용하여 대체 라벨을 생성하고, 이를 활용한 시계열 예측 모델 학습 결과를 제시한다.

4.1 기본 모델 (iForest, LOF, KNN)

라벨이 존재하는 장비 모니터링 데이터를 활용하여 iForest, LOF, KNN의 성능을 평가한다. 성능 평가는 혼돈행렬과 Precision, Recall, F1-score를 기준으로 수행한다. 기본 모델들의 성능 평가 결과, 세 모델은 모두 동일한 성능을 보였다. 정상 샘플에 대해서는 99% 이상의 높은 정확도를 달성했으며, 불량 샘플에 대해서도 약 93%의 탐지율을 기록한다. 전체 정확도는 98.6%로 매우 높은 수준이다. 그러나 세 모델의 결과가 모두 동일하게 나타난다는 점은 개별 모델의 성능 차이를 논하기 어렵다는 한계를 보여준다. 따라서 단일 모델의 결과보다는, 모델 간 결합을 통한 성능 향상 가능성이 연구의 중요한 다음 단계로 제시된다. 〈Table 2〉는 기본 모델들의 성능을 보여준다.

〈Table 2〉 기본 모델 평가 지표

Index	Precision	Recall	F1-score
Normal (0)	0.993	0.992	0.992
Anomaly (1)	0.930	0.934	0.932
Accuracy			0.986

4.2 다수결 방식

세 모델의 예측 결과를 결합하여, 두 모델 이상이 동일하게 이상치로 판정한 경우 이를 최종 이상치로 결정하는 다수결 방식을 적용한다. 세 모델이 동일한 성능을 보였음에도 불구하고, 다수결 방식을 적용한 결과 Recall이 0.934에서 0.938로 상승하며, 전체 정확도 또한 98.6%에서 98.8%로 개선되었다. 이는 개선 폭이 크지는 않지만, 결합 전략이 단일 모델 대비 안정적이고 일관적인 성능 향상을 제공할 수 있음을 보여준다. 특히 불

www.kci.go.kr

량 탐지율(Recall)의 소폭 향상은 산업 현장에서 불량을 놓치지 않는다는 점에서 실질적 의미를 가진다. 〈Table 3〉은 다수결 방식의 성능을 나타낸다.

〈Table 3〉 다수결 방식 평가 지표

Index	Precision	Recall	F1-score
Normal (0)	0.993	0.993	0.993
Anomaly (1)	0.938	0.938	0.938
Accuracy			0.988

4.3 이상치 점수 평균

앞선 다수결 방식은 모델의 결과를 단순히 이진 판정으로 결합하기 때문에, 모델 간 신뢰도의 정도를 충분히 반영하지 못한다. 따라서 각 모델의 이상치 점수를 합산하고 평균을 내는 방식을 통해 이러한 한계를 보완한다. 그 결과, 불량 탐지율은 0.942로 단일 모델(0.934) 대비 0.8% 향상되었다. 정확도는 다수결 방식과 동일하게 98.8%이나, F1-score는 소폭 향상(0.2%)되었다. 이는 단순한 투표 방식보다 점수 통합 방식이 모델 간 정보를 더 효과적으로 반영하여, 소수 클래스인 불량 샘플의 탐지 성능을 개선했음을 시사한다. 〈Table 4〉는 이상치 점수 평균 방식의 성능을 나타낸다.

〈Table 4〉 이상치 점수 평균 방식 평가 지표

Index	Precision	Recall	F1-score
Normal (0)	0.994	0.993	0.993
Anomaly (1)	0.938	0.942	0.940
Accuracy			0.988

4.4 LSTM 모델 불량품 예측

〈Table 5〉는 LSTM을 이용하여 다음 공정에서 불량 발생 여부를 예측한 결과를 나타낸다. 불량에 대한 여부는 이상치 점수 평균 방법을 사용해서 생성했으며, 전체 데이터의 5%를 이상치로 설정하였다. 정상 클래스에 대한 Recall은 약 0.99로 매우 높은 수준을 기록한다. 반면, 비정상 클래스에 대한 Recall은 0.5294로, 실제 불량의 절반 정도만 탐지할 수 있음을 보여준다. 이는 LSTM 기반 이진 분류기가 불균형 데이터 환경에서 정상 클래스에 과도하게 최적화되는 경향을 보이며, 불량 탐지 성능에는 한계가 있음을 나타낸다. 이는 LSTM 기반 이진 분류기가 불량 패턴의 빈도가 정상 대비 극히 낮은 데이터 특성상 모델이 정상 패턴을 우선적으로 학습하는 방향으로 최적화되는 경향을 보였기 때문이다. 본 연구에서 생성된 불량 라벨은 실제 산업 공정의 잠재적 이상 패턴을 반영하고 있지만, 해당 패턴이 데이터 전반에 비해 매우 희소하게 분포되어 있다. 이러한 불균형 데이터 환경에서 비정상 패턴은 정상 데이터에 비해 극히 희소하게 나타나기 때문에, 딥러닝 기반 분류 모델은 정상 클래스에 편향된 결정 경계를 학습하는 경향이 있다. 향후 연구에서는 데이터 증강, 클래스 가중치 적용, 손실 함수 개선, 모델 파라미터 수 증가 등과 같은 불균형 완화 전략을 도입함으로써 비정상 패턴에 대한 민감도를 향상시킬 필요가 있다.

〈Table 5〉 LSTM 평가 지표

Index	Precision	Recall	F1-score
Normal (0)	0.9901	0.9557	0.9726
Anomaly (1)	0.1957	0.5294	0.2857
Accuracy			0.9472

5. 결론

본 연구에서는 라벨이 존재하지 않는 압출 공정 데이터를 대상으로 불량 발생 가능성을 예측하기 위한 방법론을 제안하였다. 우선, Isolation Forest (iForest), Local Outlier Factor(LOF), K-Nearest Neighbor(kNN)과 같은 비지도 이상치 탐지 기법을 적용하여 대체 라벨을 생성하고, 라벨이 존재하는 장비 모니터링 데이터를 활용하여 모델 신뢰성을 검증한다. 실험 결과, 세 모델의 기본 성능은 유사하게 나타나나, 다수결 및 점수 평균 방식은 불량 탐지 성능을 소폭 향상되었다. 이를 통해 기본 이상치 탐지 모델들(LOF, kNN, iForest)의 이상치 탐지의 특성이 다른 점을 보여주며, 앙상블 기법을 통해 효과적으로 성능을 향상시킬 수 있다는 것을 보여준다. 또한, 생성된 라벨을 기반으로 LSTM 시계열 예측을 수행한 결과, 정상 클래스에 대해서는 높은 성능을 보이나 불량 클래스에 대해서는 Recall이 0.5294에 그쳐 불균형 데이터 환경에서의 한계를 드러낸다. 이상의 결과는 라벨이 부재한 공정 데이터에서도 불량 예측이 가능함을 보여주며, 압출 공정 예지보전 시스템 구축을 위한 기초적 가능성을 확인했다는 점에서 의의가 있다. 향후 연구에서는 데이터 불균형 문제를 해소하기 위한 민감 학습 등의 기법을 적용하고, 대체 라벨 생성 과정의 정교화를 통해 불량 탐지 성능을 더욱 개선할 필요가 있다. 또는 Yanxia

www.kci.go.kr

Li 가 수행한 연구에서 사용한 방법인 노이즈 제거 후 시계열 예측을 통해 성능 향상을 기대할 수 있을 것이다 [25].

REFERENCES

[1] T.K.Kim, "Technology Trends of Indoor Localization Scheme," *Journal of Internet of Things Convergence*, Vol.11, No.2, pp.173-177, 2025.

[2] E.J.Yun, D.E.Kim, D.H.Kim, M.J.Park, T.K.Kim, "Implementation of Smart Waste Medicine Collection Box Based on IoT," *Journal of Internet of Things Convergence*, Vol.11, No.4, pp.1-9, 2025.

[3] I.R.Numonov, B.Peng, Y.Li, Y.I.H.Ugli, T.O.Lee, T.K.Kim, "Development of an AI-Based Energy Management System for Factory Power Saving," *Journal of Internet of Things Convergence*, Vol.10, No.6, pp.49-55, 2024.

[4] J.W.Kim, B.S.Geum, T.K.Kim, "Implementing a Lecture Summary Application Using STT API and ChatGPT API," *Journal of Internet of Things Convergence*, Vol.11, No.3, pp.167-174, 2025.

[5] E.S.Oh, S.R.Gwon, J.M.Oh, B.Peng, T.K.Kim, "Implementation of a real-time public transportation monitoring system," *Journal of Internet of Things Convergence*, Vol.10, No.4, pp.9-19, 2024.

[6] B.Peng and T.K.Kim, "YOLO-HF: Early Detection of Home Fires Using YOLO," *JIEEE Access*, Vol.13, pp.79451-79466, 2025.

[7] T.Kim, T.-K. Kim, "Systemic Analysis of the QS International Research Network Indicator Using Big Data: Regional Inequalities and Recommendations for Improved University Rankings," *JIEEE Access*, Vol.13, pp.111335-111353, 2025.

[8] T.Kim, T.-K. Kim, "Global Inequality in Research: A Quantitative Analysis of 1628 Institutions in THE World University Rankings 2025," *JIEEE Access*, Vol.13, pp.161262-161278, 2025.

[9] N.Munir, R.McMorrow, K.Mulrennan, D.Whitaker, S.McLoone, M.Kellomaki, E.Talvitie, I.Lyyra and M.McAfee. "Interpretable machine learning for extrusion-induced degradation monitoring," Polymers, Vol.15, No.4, pp.915, 2023.

[10] Y.S.Perera, J.Li and C.Abeykoon. "Machine learning enhanced grey box soft sensor for melt viscosity prediction in polymer extrusion processes," Scientific Reports, 15, pp.68262, 2025.

[11] J.Carrasco, D.López, I.A.Martos, D.G.Gil, I.Markova, M.G.Barzana, M.A.Rodil, J.Lueng and F.Herrera. "Anomaly detection in predictive maintenance: A new evaluation framework for temporal unsupervised anomaly detection algorithms," Neurocomputing, 461, pp.657-670, 2021.

[12] F.T.Liu, K.M.Ting and Z.H.Zhou. "Isolation Forest," ICDM, pp.413-422, 2008.

[13] F.T.Liu, K.M.Ting and Z.H.Zhou. "Isolation-based anomaly detection," ACM Transactions on Knowledge Discovery from Data (TKDD) 6.1, pp.1-39, 2012.

[14] Z.Cheng, C.Zou and J.Dong. "Outlier detection using isolation forest and local outlier factor," Proceedings of the conference on research in adaptive and convergent systems, pp.161-168, 2019.

[15] Y.Cao, H.Xiang, H.Jhang, Y.Zhu and K.M.Ting. "Anomaly detection based on isolation mechanisms: A survey," Machine Intelligence Research, 1-17, 2025.

[16] R.Vouman, Z.Bukhsh and T.Heskes. "Unsupervised anomaly detection algorithms on real-world data: how many do we need," Journal of Machine Learning Research 25.105. pp.1-34, 2024.

[17] J.M.Lee, C.K.Yoo and I.D.Lee. "Statistical process monitoring with independent component analysis," Journal of process control 14.5, pp.467-485, 2004.

[18] J.P.S.Silva and A.L.Maitelli. "K-Nearest Neighbors for Anomaly Detection and Predictive Maintenance in Water Pumping Systems," Sensors, Vol.25, No.11, pp.3532, 2025.

[19] A.Oleff, B.Küster, M.Stonis and L.Overmeyer. "Process monitoring for material extrusion additive manufacturing: a state-of-the-art review," Progress in Additive Manufacturing, Vol.6, No.4, pp.705-730, 2021.

[20] J.Daily and J.Peterson. "Predictive maintenance: How big data analysis can improve maintenance," Supply chain integration challenges in commercial aerospace: A comprehensive perspective on the aviation value chain. Cham: Springer International Publishing, pp.267-278, 2016.

[21] A.Theissler, J.P.Velázquez, M.Kettelgerdes and G.Elger. "Predictive maintenance enabled by machine learning: Use cases and challenges in the automotive industry," Reliability engineering & system safety 215, pp.107864, 2021.

[22] P.Nunes, J.Santos and E.Rocha. "Challenges in predictive maintenance–A review," CIRP Journal of Manufacturing Science and Technology 40, pp.53-67, 2023.

[23] S.Wenshi. "Explainable Anomaly Detection in Predictive Maintenance Using Shapelet Transform——Application in heavy vehicle components failure prediction," 2024.

[24] Kaggle, Industrial Equipment Monitoring Dataset [Internet], https://www.kaggle.com/datasets/dnkumars/industrial-equipment-monitoring-dataset.

[25] Y.Li, I.N.R.Ugli, Y.I.H.Ugli, T.O.Lee and T.K.Kim. "Optimizing Models and Data Denoising Algorithms for Power Load Forecasting," MDPI energies, vol.17, no.21, pp.5513, 2024.

www.kci.go.kr

이 태 오(Tae-O Lee) [준회원]

■ 2020년 2월 ~ 2025년 9월 : 국립부경대학교 컴퓨터공학과 (공학사)
■ 2025년 9월 ~ 현재: 국립부경대학교 데이터공학과 석사

〈관심분야〉
머신러닝, 딥러닝

김 태 국(Tae-Kook Kim) [종신회원]

■ 2004년 8월 : 고려대학교 전기전자전파공학부(공학사)
■ 2006년 8월 : 고려대학교 메카트로닉스학과(공학석사)
■ 2014년 8월 : 고려대학교 모바일솔루션학과(공학박사)
■ 2016년 3월 ~ 2022년 2월 : 동명대학교 AI학부 교수
■ 2022년 3월 ~ 현재 : 국립부경대학교 컴퓨터·인공지능 공학부 교수

〈관심분야〉
사물인터넷(IoT), 콘텐츠 전송 네트워크(CDN), 이동성, 인공지능(AI), 빅데이터(big data), 모바일 서비스

www.kci.go.kr