



# LSTM Autoencoder와 Unsupervised Anomaly Detection 모델의 시간 해상도별 이상치 탐지 성능 분석

박준성<sup>1</sup>, 이정섭<sup>1</sup>, 유동건<sup>1</sup>, 이충호<sup>2</sup>, 허태욱<sup>2</sup>, \*이상금<sup>1</sup>

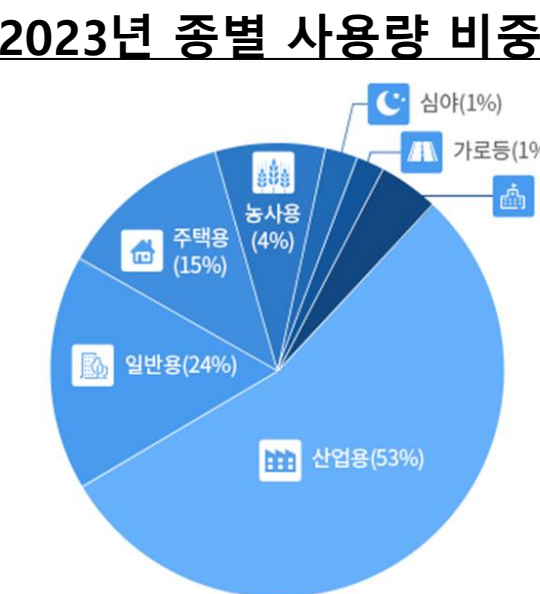
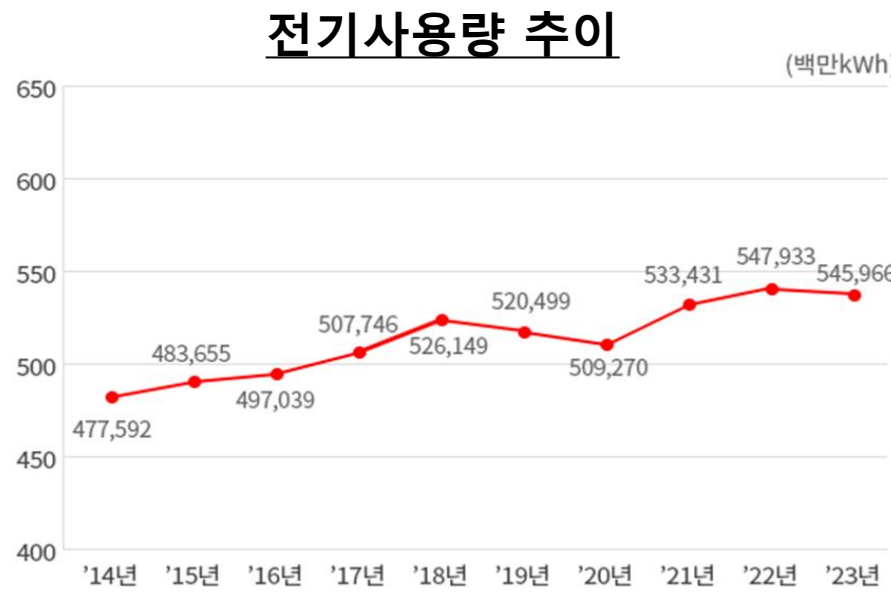
\*국립한밭대학교<sup>1</sup>, 한국전자통신연구원<sup>2</sup>



## Abstract

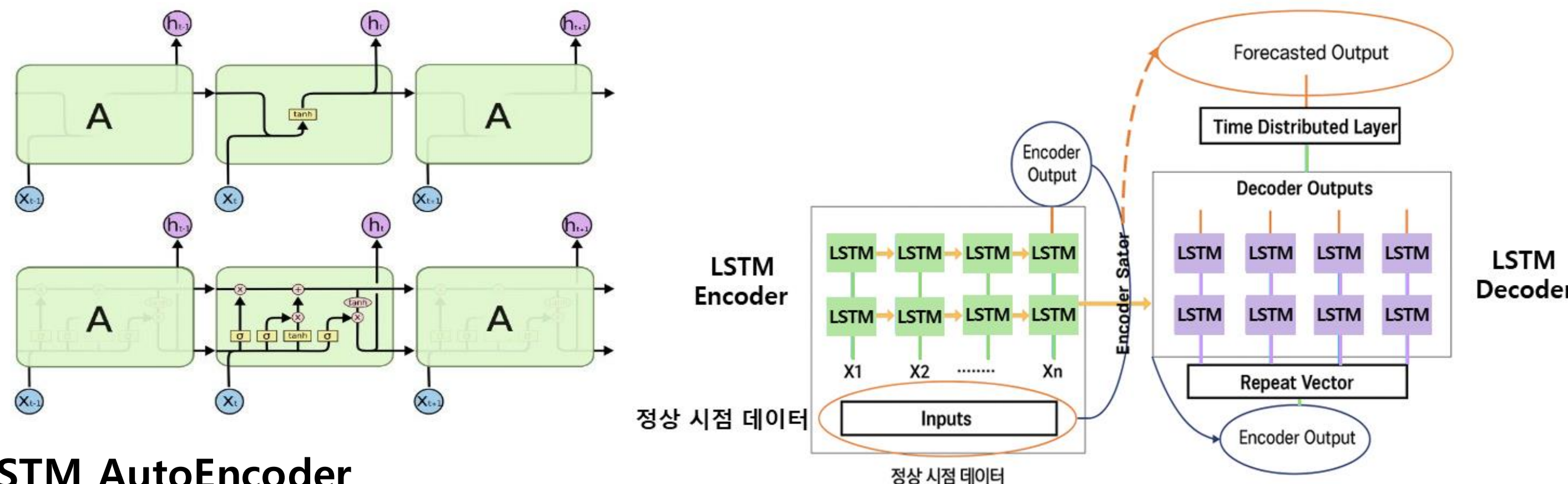
- 본 연구는 산업 전력 데이터 내 이상치를 탐지하여 에너지 낭비를 줄이고 전력 효율을 향상시키는 것을 목표로 하며, 이를 위해 시계열 기반 비지도 학습 모델의 성능을 분석.
- LSTM(Long Short-Term Memory) Autoencoder와 USAD(Unsupervised Anomaly Detection) 모델을 적용하여, 시간 해상도에 따른 이상 탐지 성능 차이를 비교.
- 실험에는 저해상도의 LEAD1.0, 고해상도의 KPI, 중해상도의 실제 산업 전력 데이터셋을 활용하고, F1-score, Precision, Recall 지표를 통해 정량적 성능 평가를 수행.
- USAD는 중저해상도 데이터에서 LSTM Autoencoder보다 우수한 성능을 보이며, 이는 실제 산업 환경에서의 센서 제약이나 저장 한계를 고려했을 때 실용적임을 시사.
- 시계열 데이터의 해상도에 따라 모델의 이상 탐지 성능이 달라질 수 있으며, 데이터 특성에 맞는 모델 선택 기준의 필요성을 강조.

## 개요



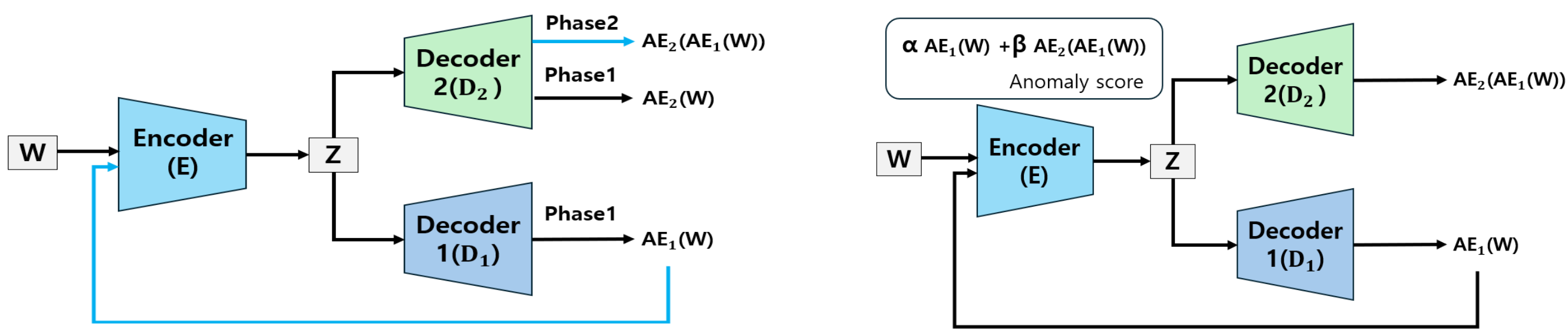
- 기후변화와 환경오염 우려 속, 에너지 효율 향상과 소비 절감의 중요성이 증가.
- 산업 부문은 2023년 기준 전체 전력 사용량의 53% 차지하며, 에너지 관리의 핵심 대상으로 부각.
- 최근 10년간 전력 사용량이 지속적으로 증가했으며, 산업용 전력이 전체 소비의 과반 이상을 차지.
- 산업 설비 운영 중 발생하는 이상치는 에너지 낭비와 시스템 불안정을 초래할 수 있음.
- 이상치 탐지를 통해 효율적인 전력 관리가 가능하므로, 이에 대한 필요성이 강조됨.

## 이상치 탐지 모델



### LSTM AutoEncoder

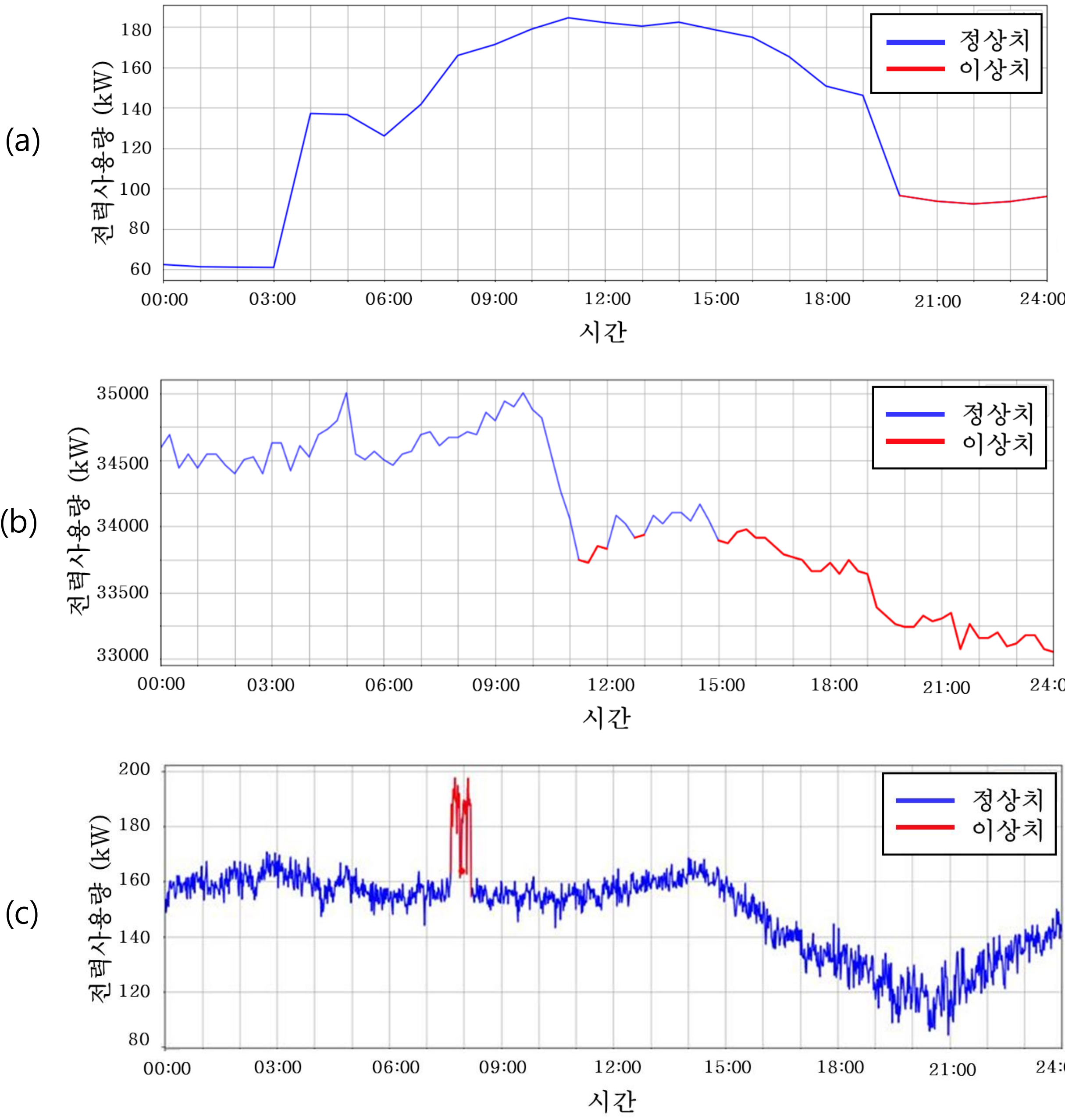
- LSTM 기반의 Autoencoder는 시계열 데이터의 시간적 패턴을 효과적으로 학습하며, 입력과 복원 값 간의 차이를 바탕으로 이상 여부를 판단.
- 시계열 데이터의 장기 의존성 반영에 강점이 있어, 정상 패턴이 반복되는 산업 설비와 같은 환경에서 효과적으로 활용 가능.



### USAD(Unsupervised Anomaly Detection)

- 하나의 Encoder와 두 개의 Decoder로 구성되어, 두 단계의 학습을 통해 정상 데이터의 패턴을 효과적으로 학습하고 입력과 복원 값의 차이를 이용해 이상치를 탐지.
- 정상과 비정상 사이의 경계에 위치한 애매한 이상치 구간까지 효과적으로 구분할 수 있음.

## 데이터 구성 및 전처리



데이터셋의 시간별 전력사용량 그래프 (a) LEAD1.0 데이터셋 (b) 실제 산업 전력 데이터셋 (c) KPI 데이터셋

- 이상치 탐지 성능 평가를 위해, 두 공개 데이터셋 LEAD1.0(저해상도, 1시간 간격)과 KPI-Anomaly-Detection(고해상도, 1분 간격)과, 실제 산업 환경에서 15분 간격으로 수집된 중해상도 산업 전력 데이터를 활용.
- LEAD와 KPI 데이터셋에는 수작업 또는 전문가에 의해 라벨링 된 이상치 정보가 포함되어 있으며, 실제 산업 데이터는 라벨 없이 모델 성능을 간접 평가하는 기준으로 사용.

### 데이터 전처리

- LEAD1.0 : 산업 전력과 유사한 건물을 선택
- KPI-Anomaly-Detection : 산업 전력의 정상적인 분포를 보이는 항목을 선별
- 세 데이터셋 모두 : 전력 사용량은 Min-Max 정규화를 적용해 입력 특성 간 스케일 차이 조정.

## 실험 방법 및 결과 분석

	LEAD1.0	Industrial Power	KPI
Train	5241	10354	63849
Validation	1747	3451	21283
Test	1747	3452	21284

표 1) 세 데이터셋을 훈련/검증/테스트 6:2:2의 비율로 분할한 표.

	LEAD1.0		Industrial Power		KPI	
Model	LSTM Autoencoder	USAD	LSTM Autoencoder	USAD	LSTM Autoencoder	USAD
Recall	0.6210	0.8179	0.7777	0.8465	0.8376	0.7952
F1 Score	0.7013	0.8636	0.8135	0.8612	0.8800	0.7685
Precision	0.8053	0.9147	0.8526	0.8764	0.9111	0.7435
AUC	0.6665	0.8629	0.8532	0.9190	0.9625	0.8906

표 2. LEAD1.0, KPI와 실제 산업 전력 데이터셋에서 LSTM Autoencoder와 USAD 이상치 탐지 성능 비교

- (표1) 이상치 탐지 성능을 평가하기 위해 산업 전력 데이터 구조를 반영한 세 가지 시계열 데이터셋을 훈련/검증/테스트 6:2:2 비율로 활용.
- (표2) USAD 모델이 중저해상도 데이터에서는 더 높은 탐지 성능을 보이고, KPI와 같은 고해상도 데이터에서는 상대적으로 낮은 성능을 보임.
- 변화 폭이 작고 이상치가 드물게 발생하는 환경에서 USAD의 적대적 학습 구조가 미세한 이상치에도 민감하게 반응.
- 실제 산업 환경에서 자주 발생하는 중저해상도 데이터 수집 환경을 고려할 때, USAD의 Decoder 구조를 통해 미세한 이상치 변화에도 민감한 이상치 탐지에 강점을 지닌 모델로 평가.

## 결론 및 향후 연구

- LSTM Autoencoder와 USAD를 활용하여 서로 다른 시간 해상도를 가진 산업 전력 데이터셋에서의 이상치 탐지 성능을 비교.
- 실험 결과, 데이터셋의 해상도에 따라 두 모델의 성능 차이가 나타나며, 중저해상도 환경에서는 USAD가 높은 탐지 성능을 나타내고, LSTM Autoencoder는 상대적으로 낮은 탐지 성능을 보임.
- USAD는 복잡한 LSTM 구조 없이 이상치에 민감하게 반응할 수 있어 중저해상도 데이터가 주를 이루는 산업 환경에 적합한 대안.
- 향후 두 모델의 강점을 결합한 하이브리드 이상치 탐지 모델을 설계하여 해상도 변화나 실시간 환경에서도 강건한 이상치 탐지 성능을 유지하는 시스템을 구축할 예정.
- 다양한 해상도의 산업 전력 데이터의 신뢰성을 향상시키고, 에너지 효율 최적화 및 탄소 배출 저감 목표 달성에 기여하고자 함.