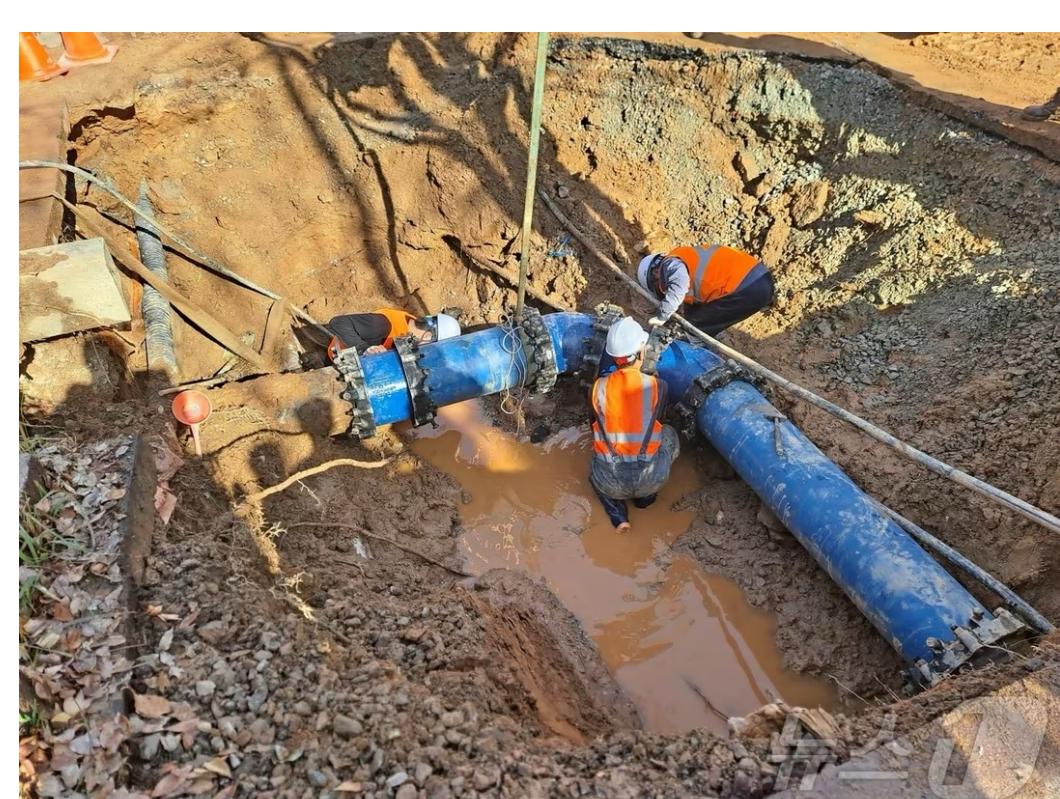


상수도관 누수 탐지를 위한 머신러닝 및 딥러닝 기법 비교 연구

이원준¹, 강대희¹, 김광주, 이연창[†]
 George Mason University, UNIST, ETRI
 wlee40@gmu.edu, {kangdaehhee, yeonchang}@unist.ac.kr, kwangju@etri.re.kr

연구 배경

- 상수도관 누수 탐지는 도시 인프라의 안정성과 회복 탄력성을 좌우하는 핵심기술
- 장기간 방치 시, 수자원 손실과 수익 저하, 심할 경우 토양 침하와 도로 파손 등으로 이어져 **인명 피해 야기**



-노후화로 인한 누수-

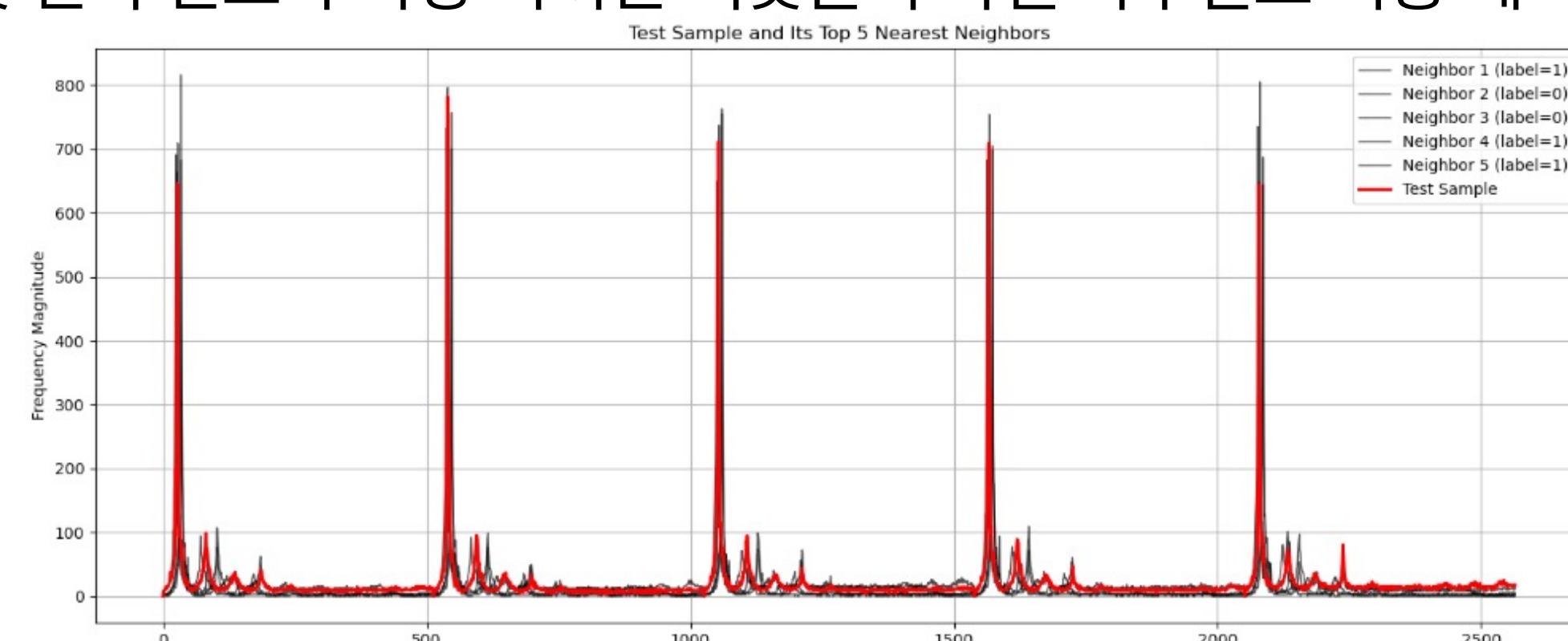


-수도관 이음새 이격-

머신 러닝 기반 누수 탐지 전략

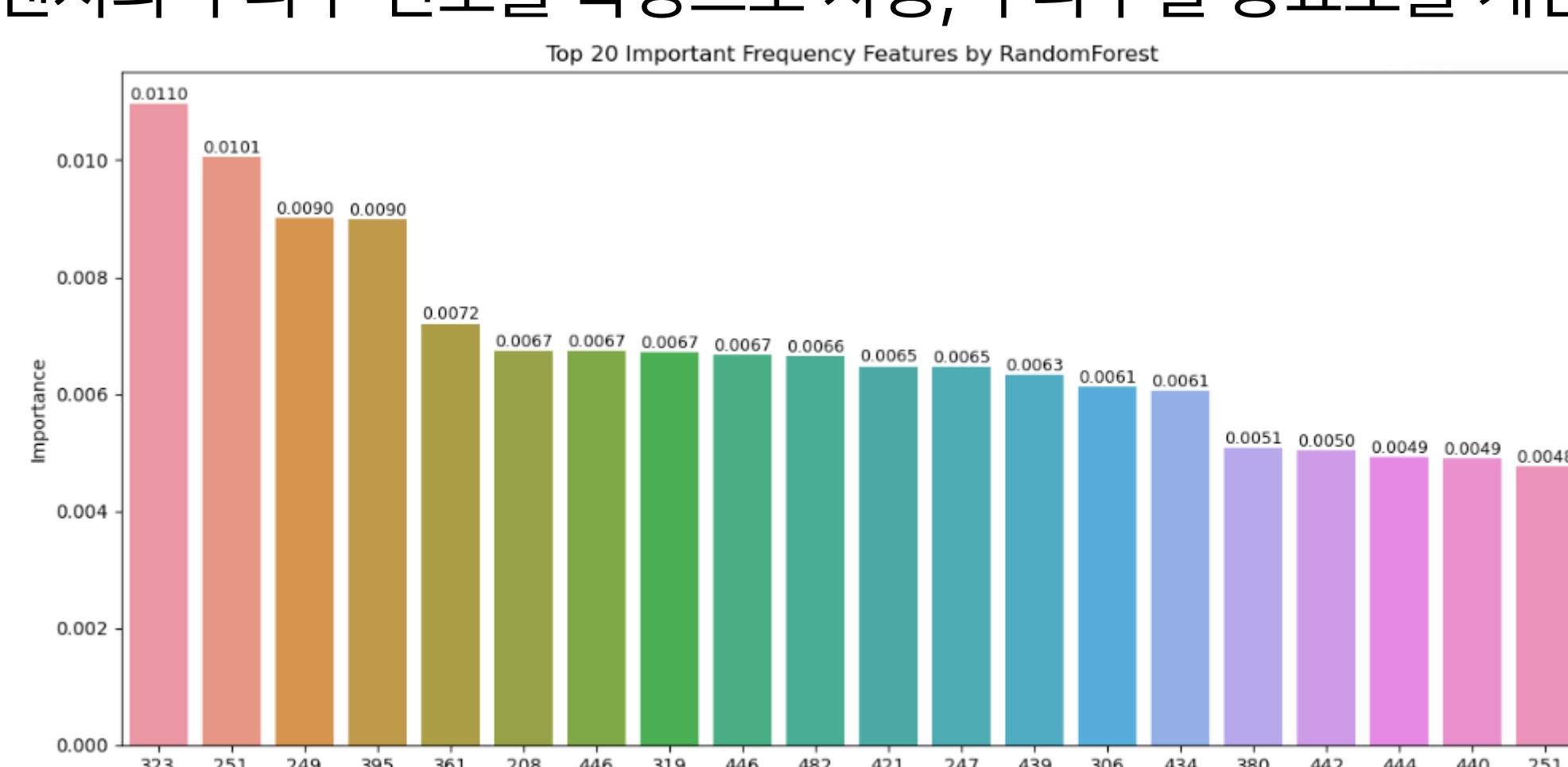
1. K-NN

- 주파수 신호 간 형태 차이를 Euclidean Distance로 계산
- 타겟 센서 신호와 가장 가까운 이웃들의 라벨 다수결로 최종 예측



2. Random Forest

- 각 센서의 주파수 신호를 특징으로 사용, 주파수별 중요도를 계산



실험 설계

◆ 데이터 집합의 통계

센서 인스턴스	주파수	정상 (0)	일반 누수 (1)	미세 누수 (2)
1,955	512	690	705	560

◆ 수집 데이터 예시*

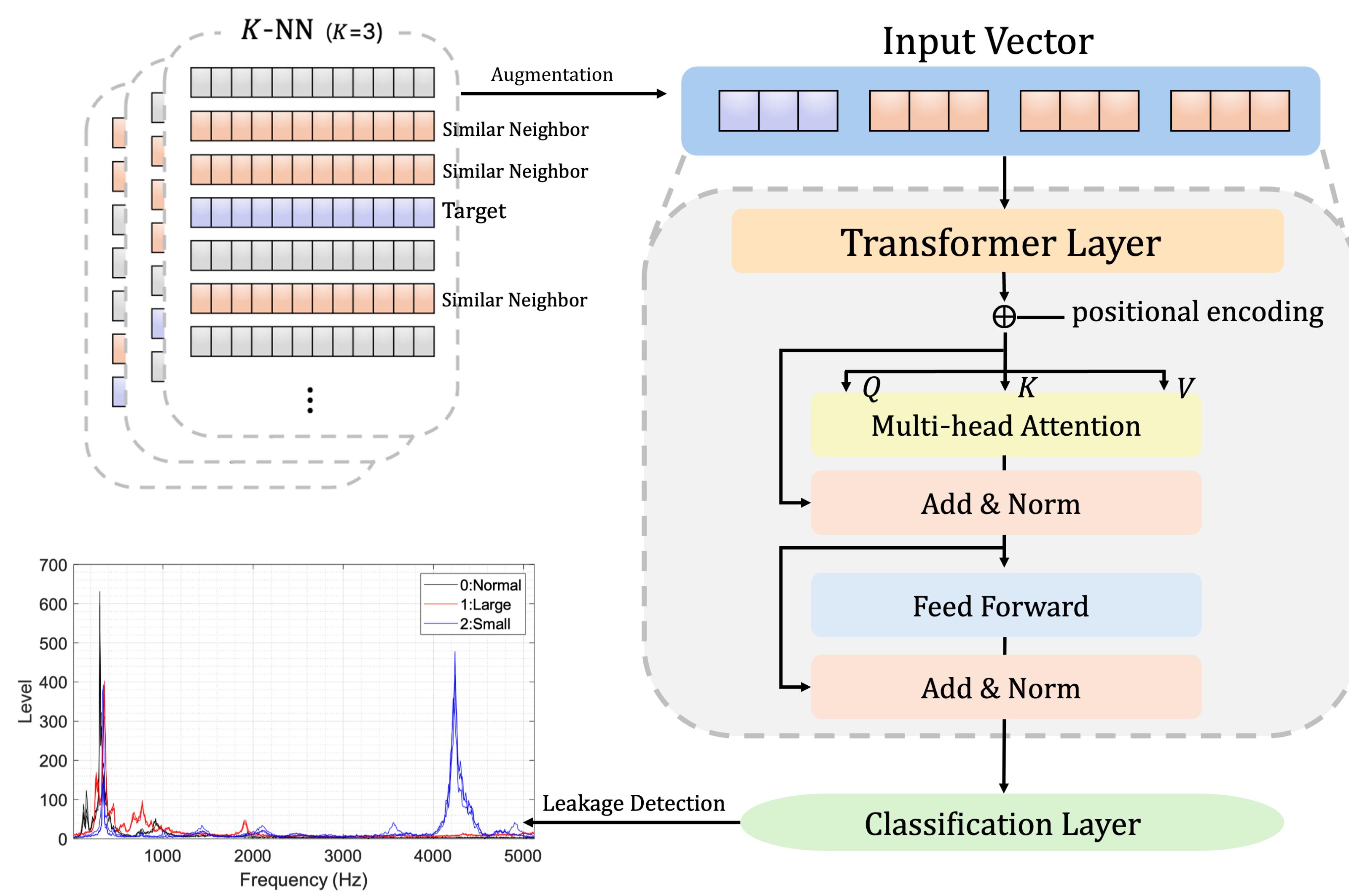
누수 Label	주파수 (512 × 5일)
일반 (1)	1.4, 14.2, 17.1, 22.7, 31.3, 28.4, 31.3, 45.5, 51.2, 56.9, 68.2, ...
일반 (1)	8.9, 8.9, 11.8, 14.8, 20.7, 29.6, 26.6, 35.5, 47.3, 47.3, 53.2, ...
미세 (2)	29.8, 32.3, 32.3, 44.8, 47.2, 69.6, 67.1, 184.5, 139.2, 634.0, ...
일반 (1)	11.3, 12.9, 14.5, 14.5, 16.1, 17.7, 16.1, 19.3, 19.3, 17.7, ...
정상 (0)	10.8, 13.5, 16.2, 16.2, 24.3, 18.9, 24.3, 48.6, 45.9, 27.0, 29.7, ...

* 센서로부터 1일 동안 측정된 512개의 신호 강도 값

연구 동기

- 상수도관 누수는 시간이 지날수록 재정 손실, 기반시설 훼손, 안전사고 위험으로 확대되므로, 이를 조기에 진단할 수 있는 기술이 중요
 - 노후관로 증가 및 기후 변화로 인해 지속적인 모니터링과 자동 탐지 체계의 필요성이 커짐
 - 그러나, 비용 문제, 주파수 센서 고장 등으로 인한 데이터 수집의 어려움과 데이터 상 **환경 잡음, 실 사용자 수요에 따른 시간대별 패턴 이동의 도전과제** 존재
 - 현장에서 쉽게 활용할 수 있는 K-NN, Random Forest 등은 구축 용이성이 높으나, **데이터의 비정상성을 극복하고 누수를 진단하는데 한계**를 가짐
- 본 연구는 실제 상수관 주파수 데이터에 기반, **머신러닝·딥러닝 기법의 누수 탐지 성능을 비교하고, 딥러닝 고도화 전략의 적용 가능성을 탐구**하고자 함

Graph Transformer 기반 누수 탐지 전략



분석 결과

방법	Label	평가 지표		
		Precision	Recall	Accuracy
K-NN	정상(0)	0.9571	0.9710	0.9745
	일반(1)	0.9714	0.9577	
	미세(2)	1.0000	1.0000	
Random Forest	정상(0)	0.9385	0.8841	0.9337
	일반(1)	0.9167	0.9896	
	미세(2)	0.9492	1.0000	
MLP	정상(0)	0.9474	0.7714	0.855
	일반(1)	0.8000	0.9143	
	미세(2)	0.8413	0.8833	
Graph Transformer	정상(0)	0.9752	0.9316	0.9638
	일반(1)	0.9311	0.9714	
	미세(2)	1.0000	1.0000	

- K-NN**이 전체에서 정확도가 가장 높은 성능을 기록하며, 단순 구조임에도 일반·미세 누수에서 강함
- Graph Transformer**는 정상·미세 누수 Precision지표 1위를 포함해 상위 수준을 유지하며 **주변 센서 정보 활용의 강한 잠재력을 시사**
- 머신 러닝 기법을 근소한 차이로 추격하며, 데이터양 증가, 다양한 입력 feature를 활용할 경우, 고도화 측면에서 우위 확보 가능