

## 연구 배경 및 목적

- 라돈(Rn)은 자연 방사성 기체로 무색·무취이며, 장기 고농도 노출 시 폐암 위험을 증가시키는 것으로 보고됨[1]. 서울과 같은 고밀도 도시에서는 계절·기상 변화에 따른 라돈 농도의 급변 가능성이 커, 선제적 예측과 관리가 필수적임[4].
- 본 연구는 라돈넷의 810대 측정망에서 확보한 장기간 라돈 자료와 기상청의 서울 기상 자료를 융합하여, 기온·기압과 라돈 농도의 관계를 딥러닝으로 학습하고 예측 성능을 체계적으로 검증함. 결과를 바탕으로 실시간 모니터링·위험경보 시스템으로의 적용 가능성을 논의함. 또한, 서울 지역에서 라돈과 기상 인자 간 상관성이 보고되어 왔으며[2], 본 연구는 이를 예측 문제로 확장하였다.
- 라돈넷은 (주)에프티랩에서 2018년도부터 운영중인 라돈모니터링 시스템에서 서울지역 810여대의 라돈측정센서(모델명 라돈아이P2/라돈아이PRO)에서 서버로 upload된 data를 사용하였다. 참고로 현재 라돈넷에 등록된 실시간 라돈측정센서는 5,600여개이며, 관계회사인 ECOSENSE.INC(USA)에서는 data 저장 가능한 라돈모니터링 장치(모델명 EcoQube) 47,000여대를 운영중에 있다[3].

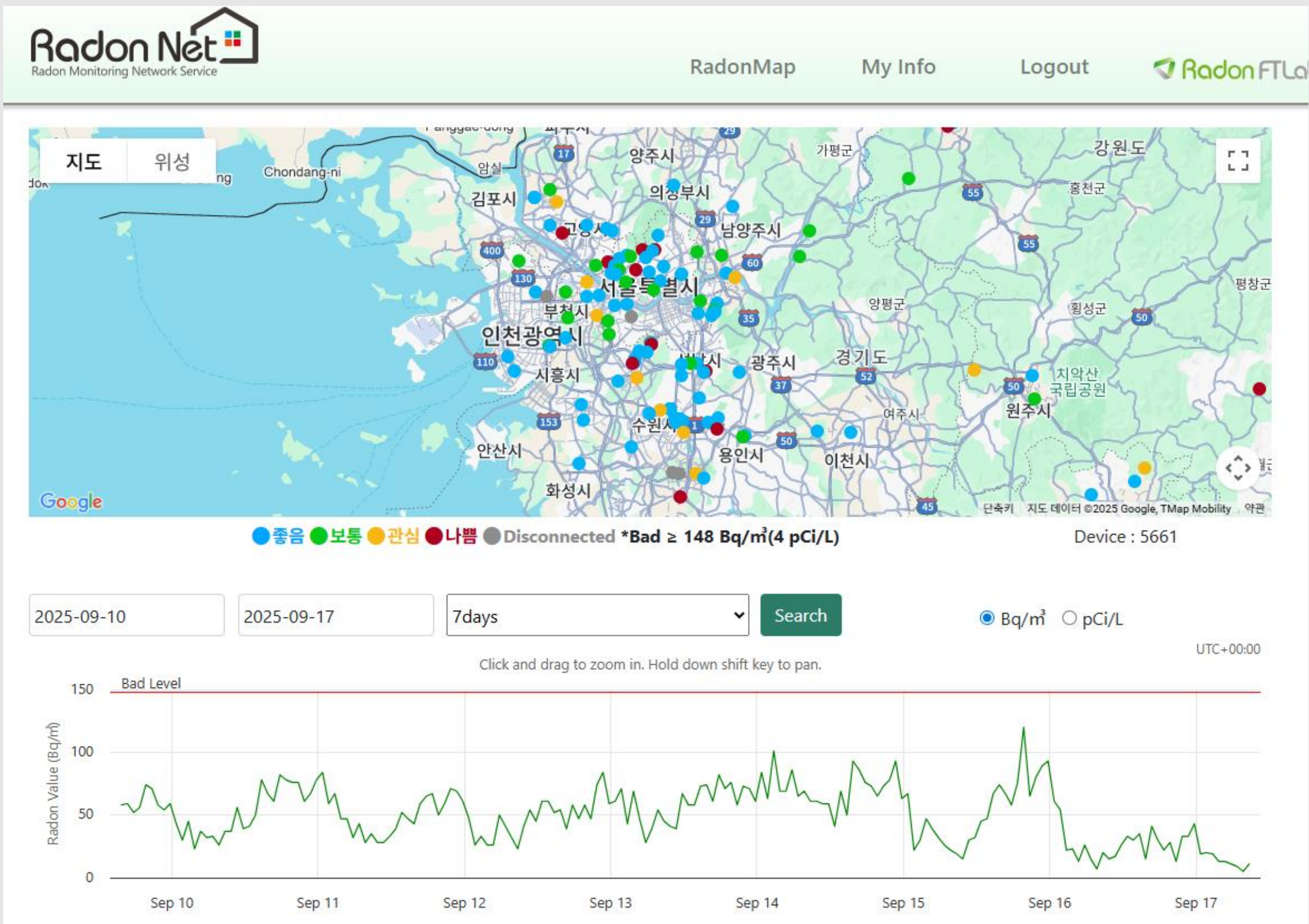


그림 1. 라돈넷 서울 권역 측정망 예시(전국망 중 서울 810대 활용).

## 데이터셋 구성·전처리 및 학습 환경

- 수집 : 기상청(서울) 일별 평균 기온(avgTa)·기압(avgPa), 라돈넷(서울) 라돈 농도(Bq/m³), 2018.05~2024.11. (본 연구 데이터 기준: 2018-05-05~2024-11-25, 총 2,374일; 라돈넷 서울권 측정망 810대 활용)
- 전처리 : 단기 변동성 완화와 장기 추세 반영을 위해 10일 이동평균 적용. 단위 일원화, 한국표준시(KST) 기준 정렬.
- 정합/품질관리 : 날짜 기준 병합, 결측치 제거, 센서 품질 점검 및 이상치(센서 오류·비정상 급변) 행 제거. 데이터 연속성과 표본 밀도(가동률·결측률)를 비교한 결과, 2019-10-01~2020-12-31 구간이 810대 센서에서 가장 치밀하고 연속적인 관측이 확보되어 대표성·학습 안정성이 높아 최종 학습·평가 원도우로 선정(총 454일). 다른 기간은 센서 증감·점검 등으로 유효 표본 수가 상대적으로 적음.
- 학습/실험 환경 : Ubuntu 22.04, JupyterLab, NVIDIA RTX 4070

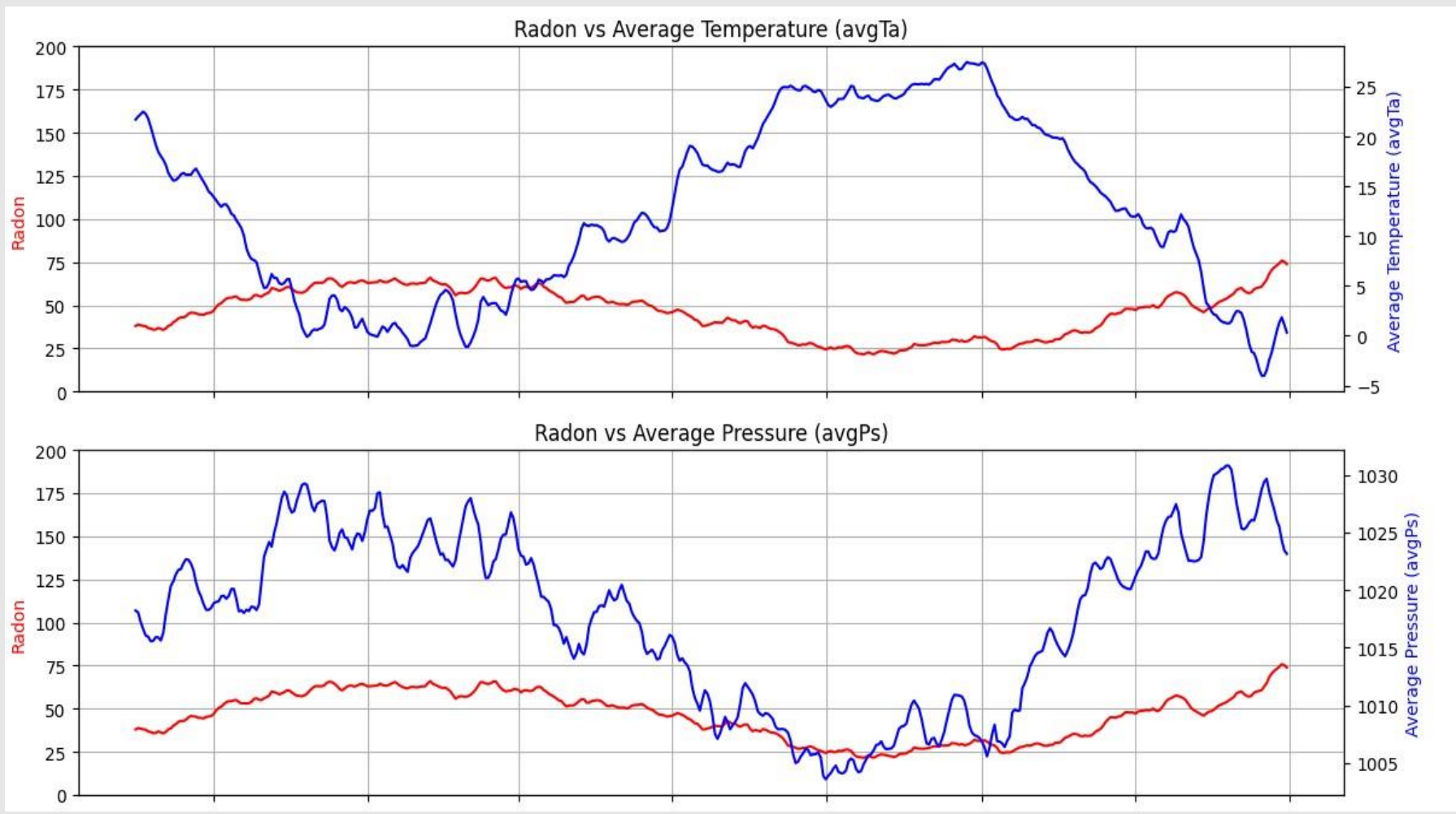
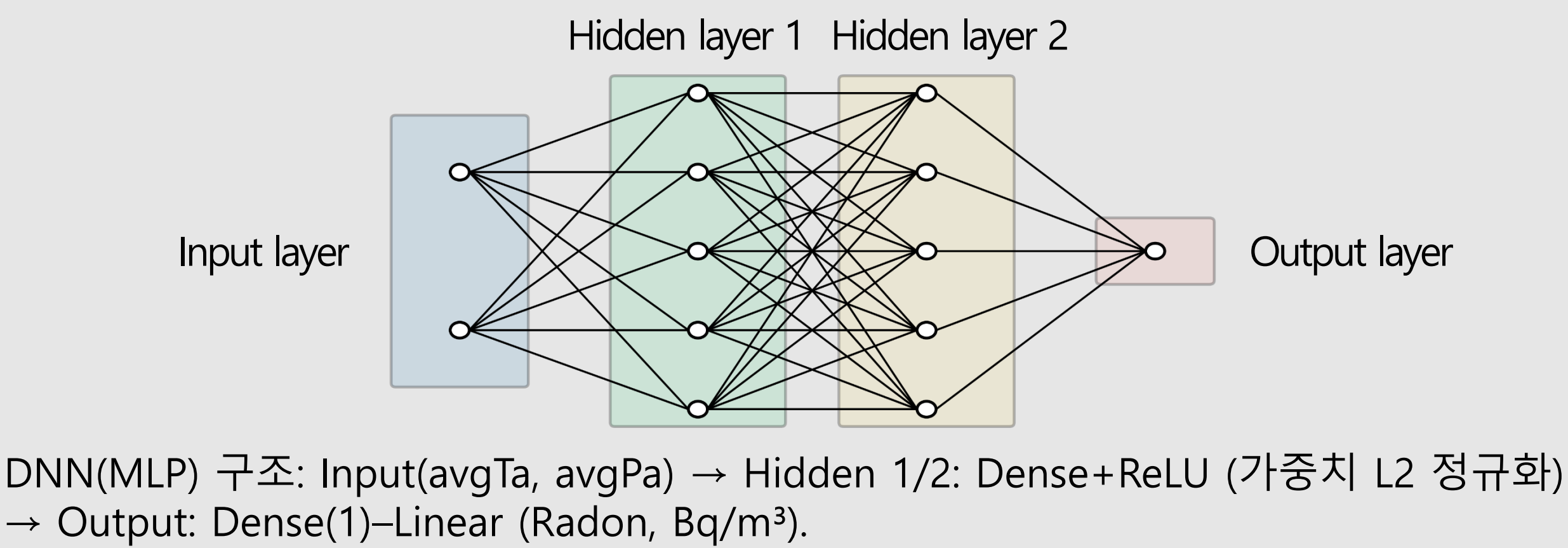


그림 2. 입력 변수와 라돈의 동시 추세(상: 기온-라돈, 하: 기압-라돈).

## 입력/타겟·모델·검증

- 입력/타겟: 입력—avgTa, avgPa 표준화(평균 0, 분산 1); 타겟—radon\_avg(Bq/m³).
- 모델: 다층 퍼셉트론(MLP) 회귀, 은닉 2층(비선형 활성화 ReLU), 가중치 초기화 표준 방식, L2 정규화.
- 학습: 손실 함수—MSE, 최적화—Adam, Epoch 1,000(각 Fold), 미니배치 학습, Early-Stopping 적용
- 검증 설계: 시간 순서를 보존한 5-Fold Cross-Validation (CV)으로 일반화 성능을 추정한다. 각 Fold는 훈련 80% / 검증 20%로 사용하며, 분할 전 셔플(shuffle)을 적용하여 Fold 간 분포를 균등화. K-Fold 원리상 각 표본은 정확히 한 번 검증에, K-1번 훈련에 사용되어 편향을 줄임. 평가에는 Out-of-Fold(OOF) 예측을 사용해 공정한 평가를 수행하고, 연속 예측 시각화를 위해 전체 데이터 Refit(단일 모델) 예측선을 병기하였다.
- 지표 설명: Coefficient of Determination ( $R^2$ : 1에 가까울수록 우수), Mean Squared Error (MSE: 작을수록 우수).

## DNN(Deep Neural Network) 구조



## 시각화 및 해석

- 시계열(True / Refit / 교차검증 예측): 계절 추세를 안정적으로 추종하며, 교차검증 과정의 Fold 경계에서 나타날 수 있는 이음새는 가독성을 위해 Refit(단일 모델) 곡선을 병기하여 보완하였다. 산점도는  $y=x$  기준으로 분산·치우침을 확인한다.
- 계절 전환기와 강한 기압 변동 구간에서 단기 급변의 일부 미스매치가 관찰되나, 이는 일 단위 10일 이동평균 기반의 평활과 입력 변수 제한에서 기인한 것으로 해석된다. 연속 운영 관점에서는 Refit 곡선이 추세의 일관된 맥락을 제공하며, 교차검증 예측은 일반화 수준을 가늠하는 참고 지표로 제시한다.

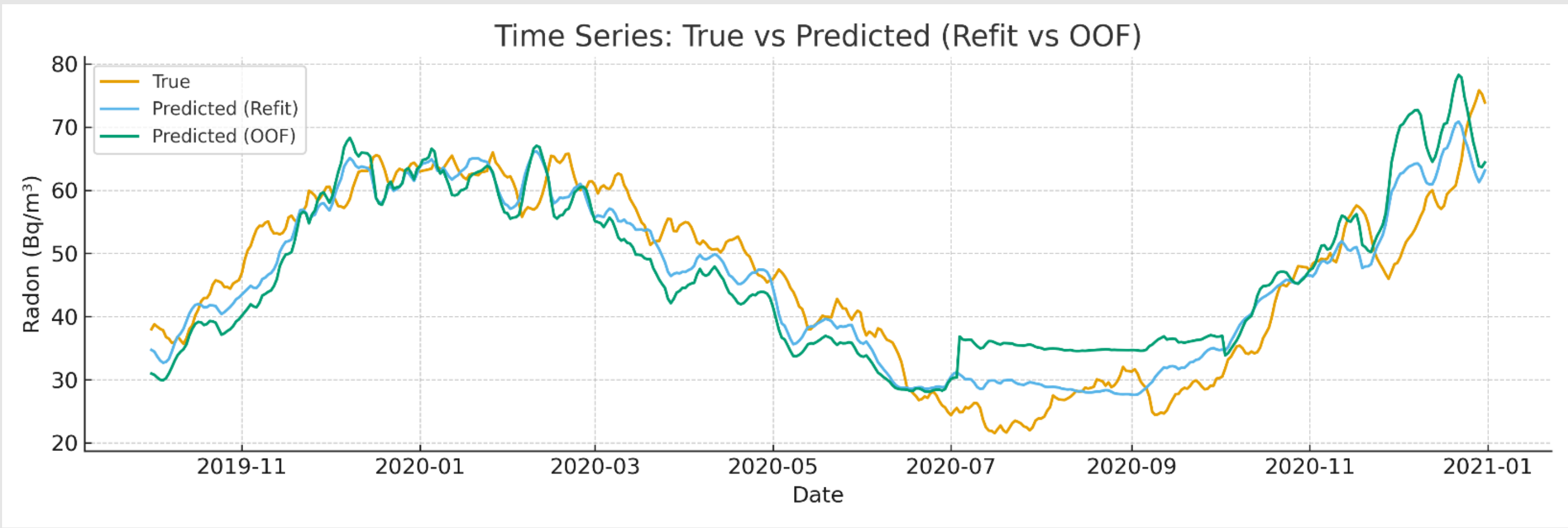


그림 3. 시계열 비교(예측 vs 실제, 2019-10-01~2020-12-31). 주황=실측(True), 파랑=Refit, 녹색=교차검증 예측(OOF). (단위: Bq/m³)

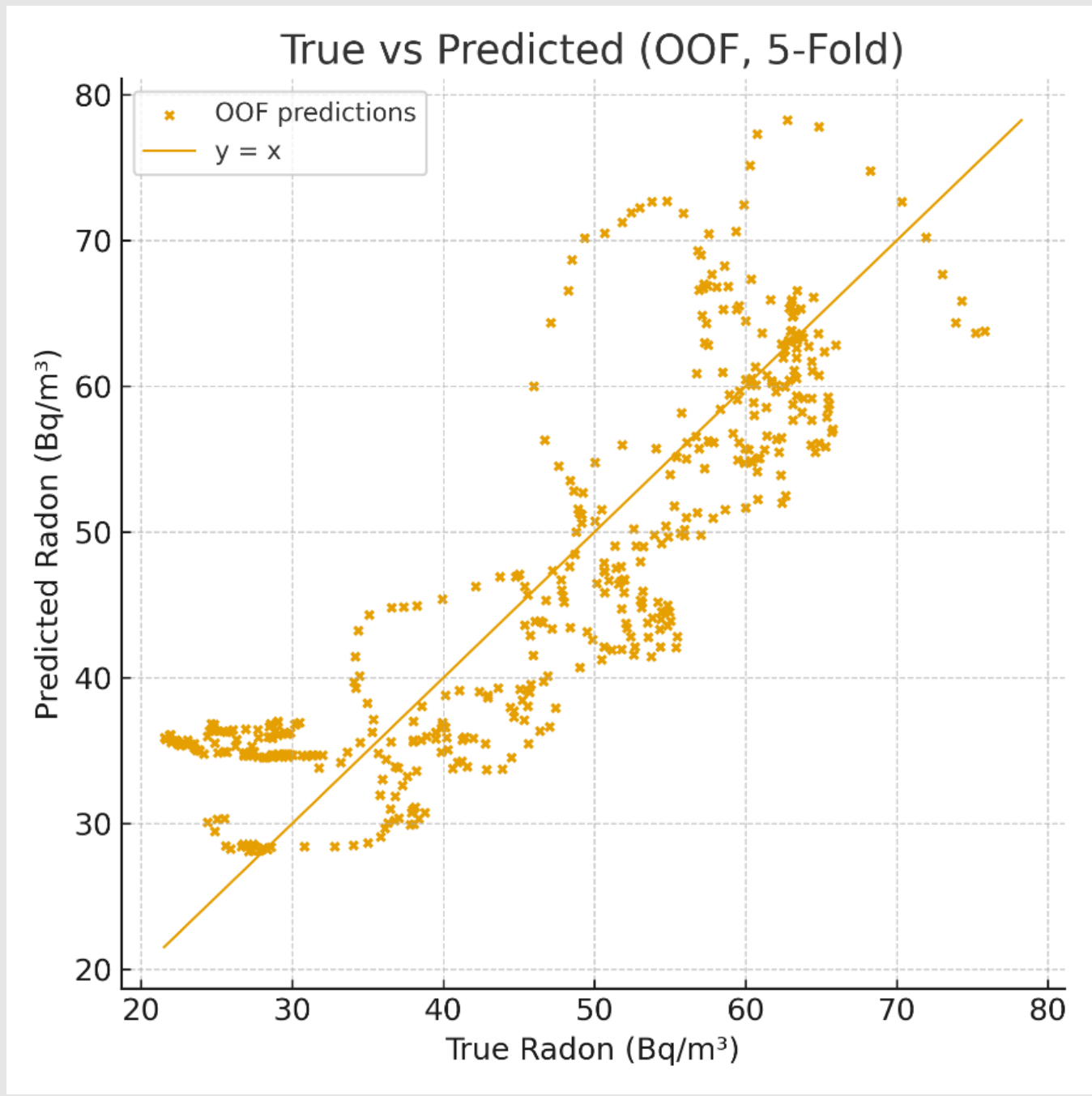


그림 4. 산점도(예측 vs 실제). 점선은 완전일치선 ( $y = x$ )\*\*이며, 각 점은 예측과 해당 시점 실측의 한 쌍을 의미한다. 분산과 치우침을 직관적으로 확인하기 위해 OOF만 별도로 도식화하였다. (단위: Bq/m³)

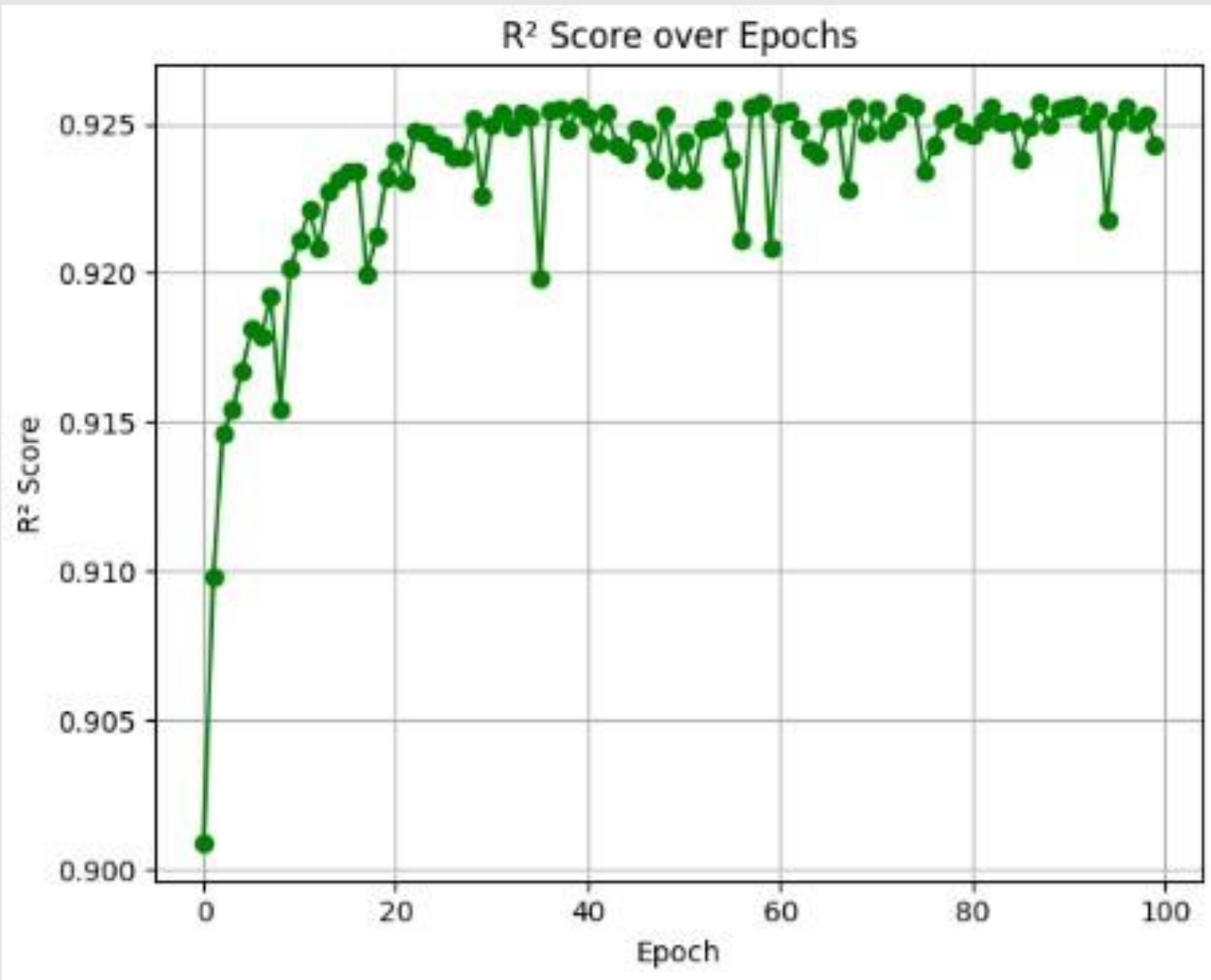
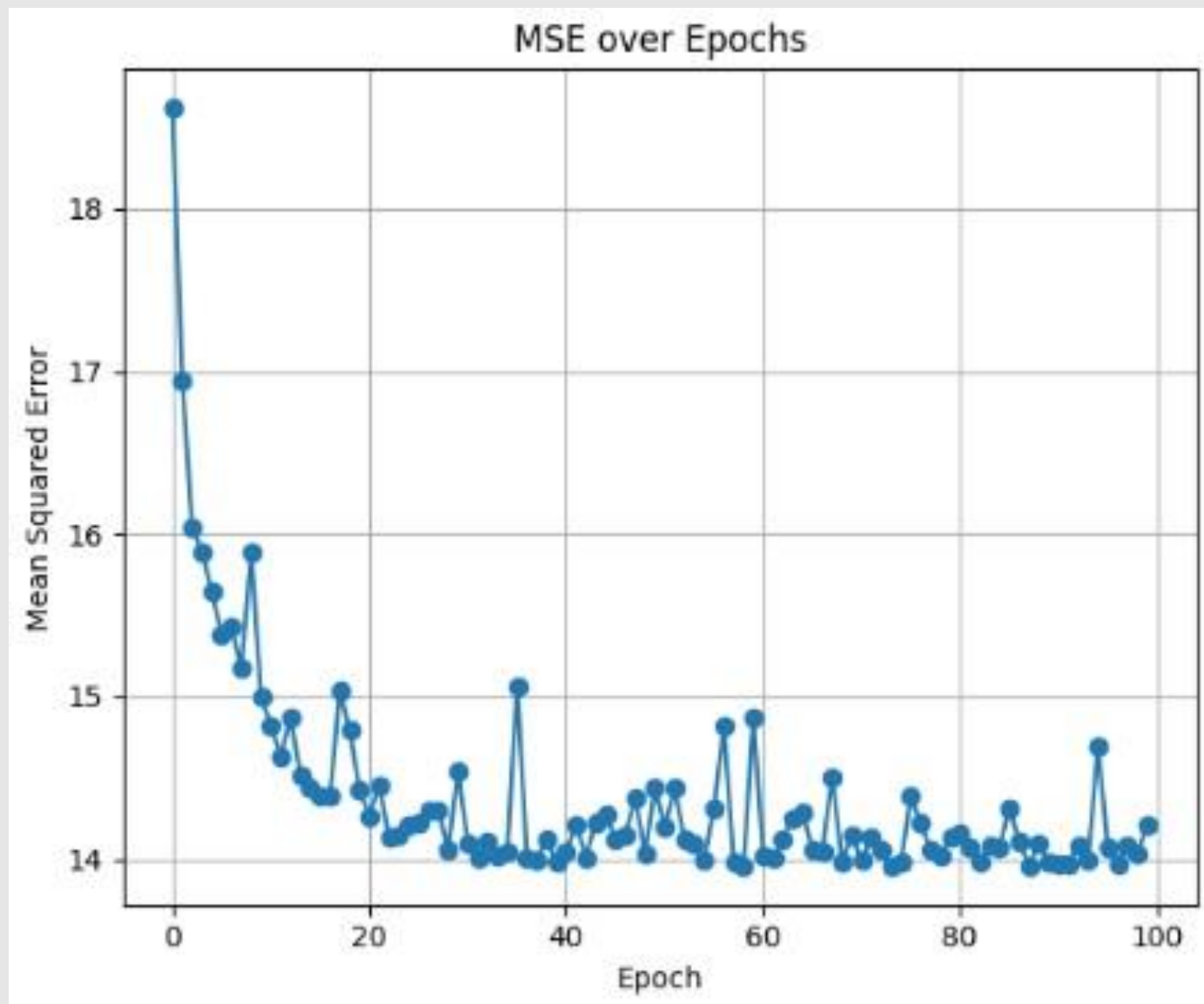


그림 5. Epoch-wise Test 성능(MSE,  $R^2$ )과 Fold별 결과. Test  $R^2$  평균 $\approx$ 0.925, Test MSE 평균 $\approx$ 14.22.

- K-Fold 최종 평균 성능(대표 실행): Train  $R^2$  0.9280, Train MSE 13.5448, Test  $R^2$  0.9251(범위 0.912~0.944), Test MSE 14.2238.
- 해석: 기온·기압만으로도 변동성의 큰 부분을 설명하며, 예측 신뢰도가 계절 전환기 전후에서 특히 높게 나타남. 운영 적용 시 경보 임계값 설정 및 일별 위험지수 산출에 활용 가능.

## 결론 및 고찰

기상 인자(기온·기압)와 라돈 관측을 결합한 딥러닝 회귀모델로 서울 지역 라돈의 일 단위 추세 예측이 가능함을 보였다. 시간 순서를 보존한 5-Fold OOF로 일반화 성능을 추정했고, 평균  $R^2=0.9303$ , MSE=14.7465로 높은 설명력을 확인하였다. 운영 관점의 연속성은 Refit(단일 모델)로 제시하였다. 요약하면, 본 연구의 핵심은 기상·라돈 데이터를 이용한 ‘예측’ 자체에 있으며, 교차검증은 이를 공정하게 평가하기 위한 절차일 뿐이다. 제시된 정량 지표( $R^2$ , MSE)와 시각적 검증(시계열, 학습곡선)은 모델의 설명력·안정성을 동시에 뒷받침한다. 향후에는 습도·풍속·풍향·강수·지면온도·기압 변화율 등 기상 입력을 확대하고, 모델 구조·하이퍼파라미터를 고도화하여 예측 성능을 개선할 예정이다.

### 참고 문헌

- [1] UNSCEAR (2000). Sources and Effects of Ionizing Radiation. UN, New York.
- [2] Cho, B.W., Kim, Y.J., & Park, J.H. (2019). Journal of Environmental Radioactivity, 205, 17–25.
- [3] Carmona, M. et al. (2017). Intercomparison of commercially available active radon measurement devices in a ‘discovered’ radon chamber. Consortium for Verification Technology, Research Poster.
- [4] WHO (2009). WHO Handbook on Indoor Radon: A Public Health Perspective. World Health Organization..