



Abstract

- 인플루엔자는 계절성과 변동성이 큰 예측이 어려운 감염병으로, 정밀한 단기 예측은 공중보건 대응에 필수적
- 본 연구에서는 2014~2025년 국내 인플루엔자 의사환자분율 데이터를 기반으로 수리·통계·머신러닝·딥러닝 기반의 10개 예측 모델을 구축하고, 이를 13가지 앙상블 기법으로 결합함
- 그 결과, 앙상블 모델군의 RMSE와 MAE 평균은 개별 모델군 평균 대비 각각 36%, 33% 낮았으며, 앙상블 모델군의 평균 R<sup>2</sup>는 0.91 이상으로, 전반적으로 우수한 예측 성능을 유지함
- 이는 앙상블 기반 접근이 변동성이 큰 감염병 예측에 있어 강건한 대응 전략이 될 수 있음을 시사함

Introduction

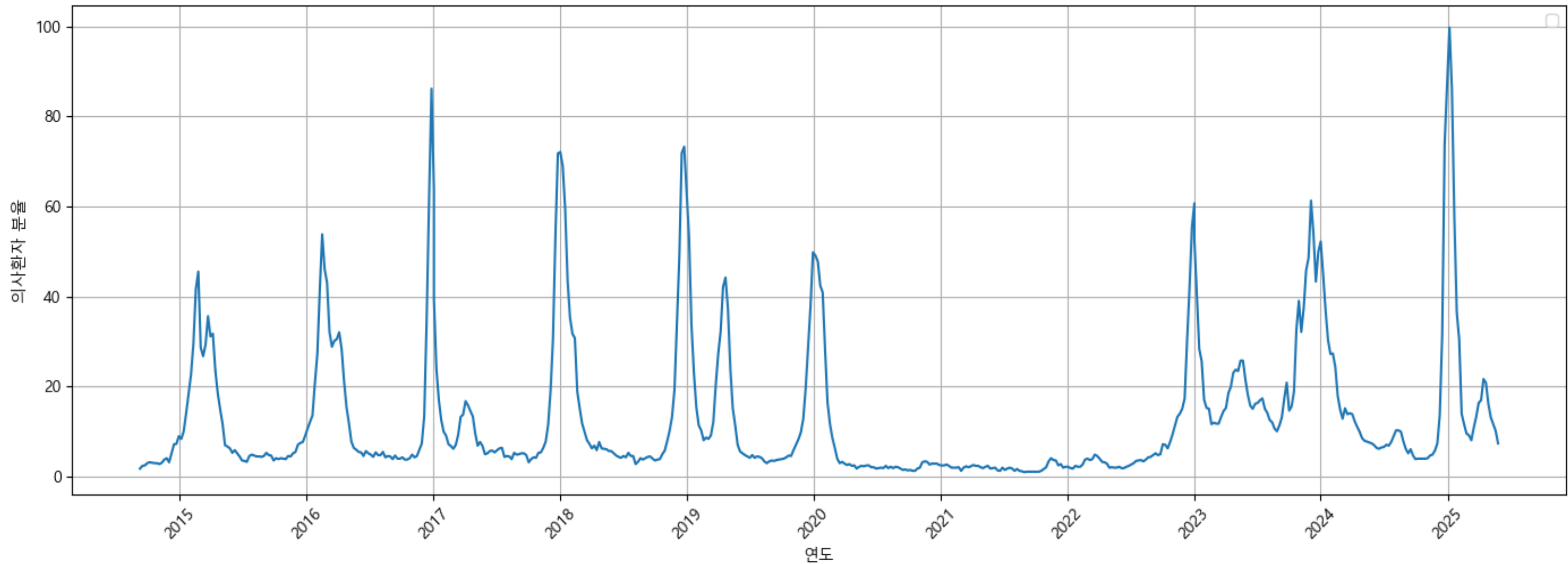
- 인플루엔자는 매년 11~4월에 유행하는 대표적인 호흡기 감염병
- 고위험군을 포함한 전 연령대에 많은 감염자를 발생시켜 사회·경제적 부담을 초래
- 최근 기후 변화와 국제 이동 증가로 감염 패턴이 더욱 복잡해지고 있지만 기존 방역 정책만으로는 한계가 있어, 정교한 예측 모델 개발이 필수적
- 수리, 통계, 머신러닝, 딥러닝 개별 모델 및 앙상블 모델로 예측 정확도를 높이고, 이를 통해 감염자 수 변동을 정밀하게 파악하여 방역 정책과 의료 자원 배분에 기여하고자 함

Dataset

감염병	출처	단위	값	기간
인플루엔자	감염병포털	주	의사환자분율	2014년 36주 ~ 2025년 21주

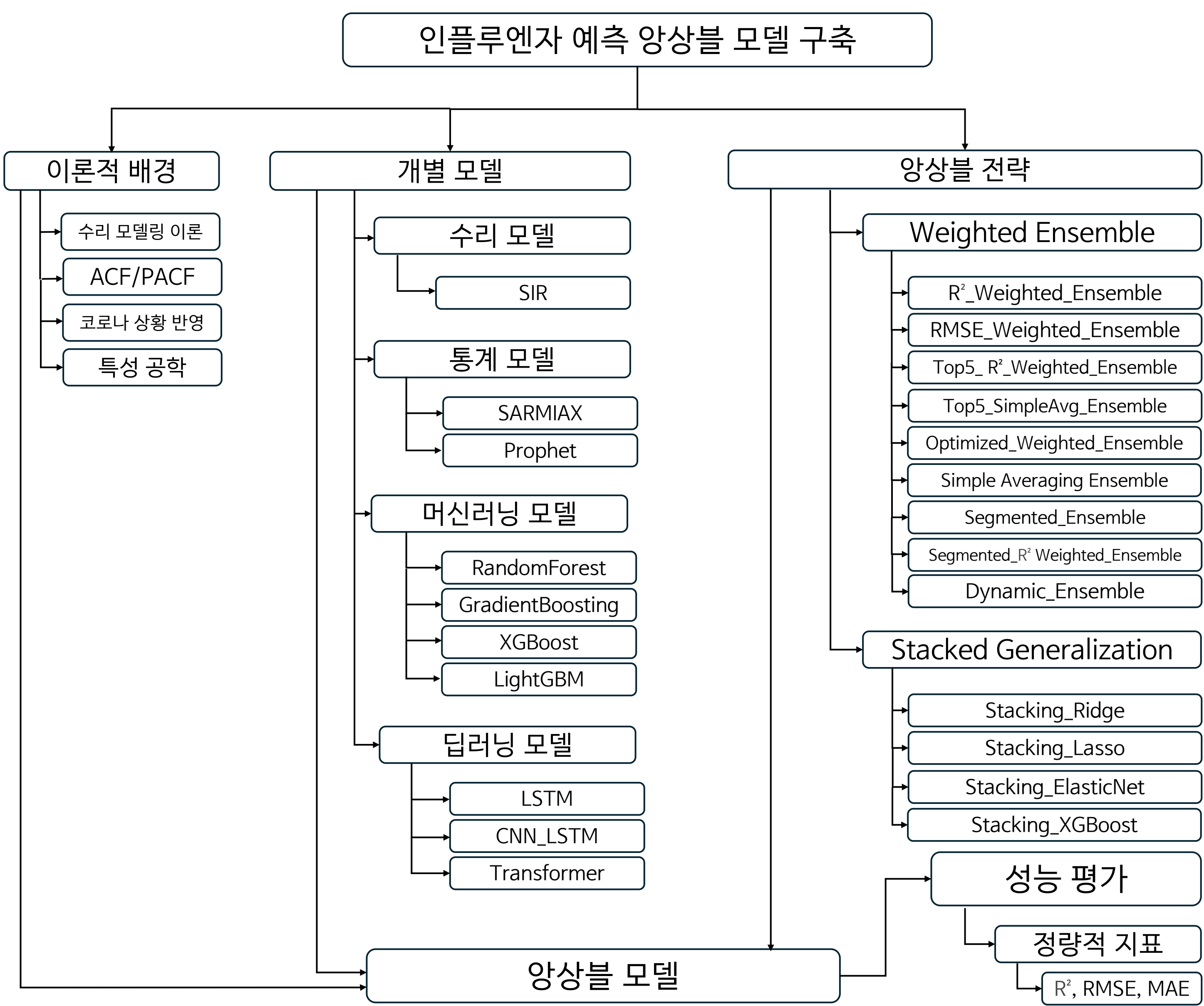
- \* 인플루엔자 의사환자: 38℃ 이상의 갑작스러운 발열과 더불어 기침 또는 인후통을 보이는 자
- \* 인플루엔자 의사환자분율: 외래환자 1,000명당 인플루엔자 의사환자의 비율

[표1] 사용 데이터



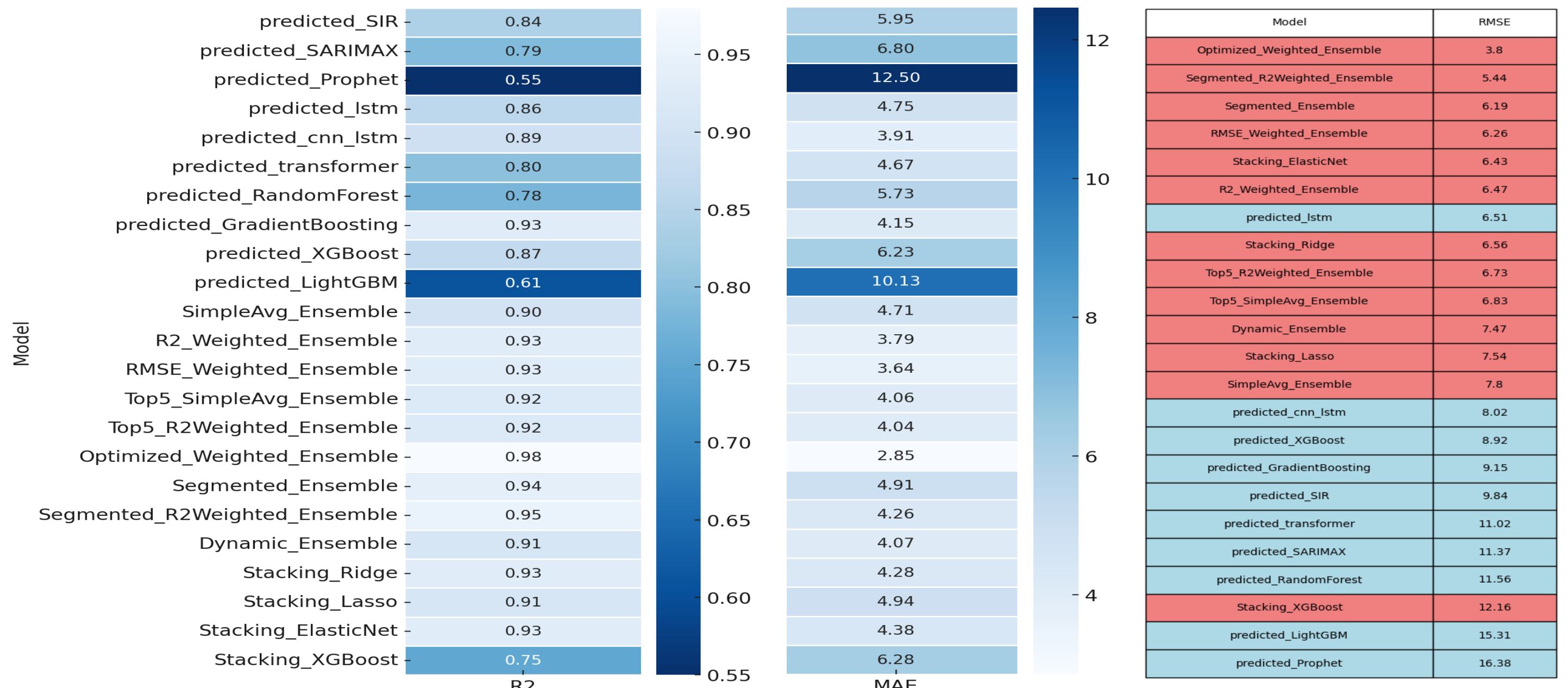
[그림1] 인플루엔자 의사환자분율 시계열 데이터

Methods



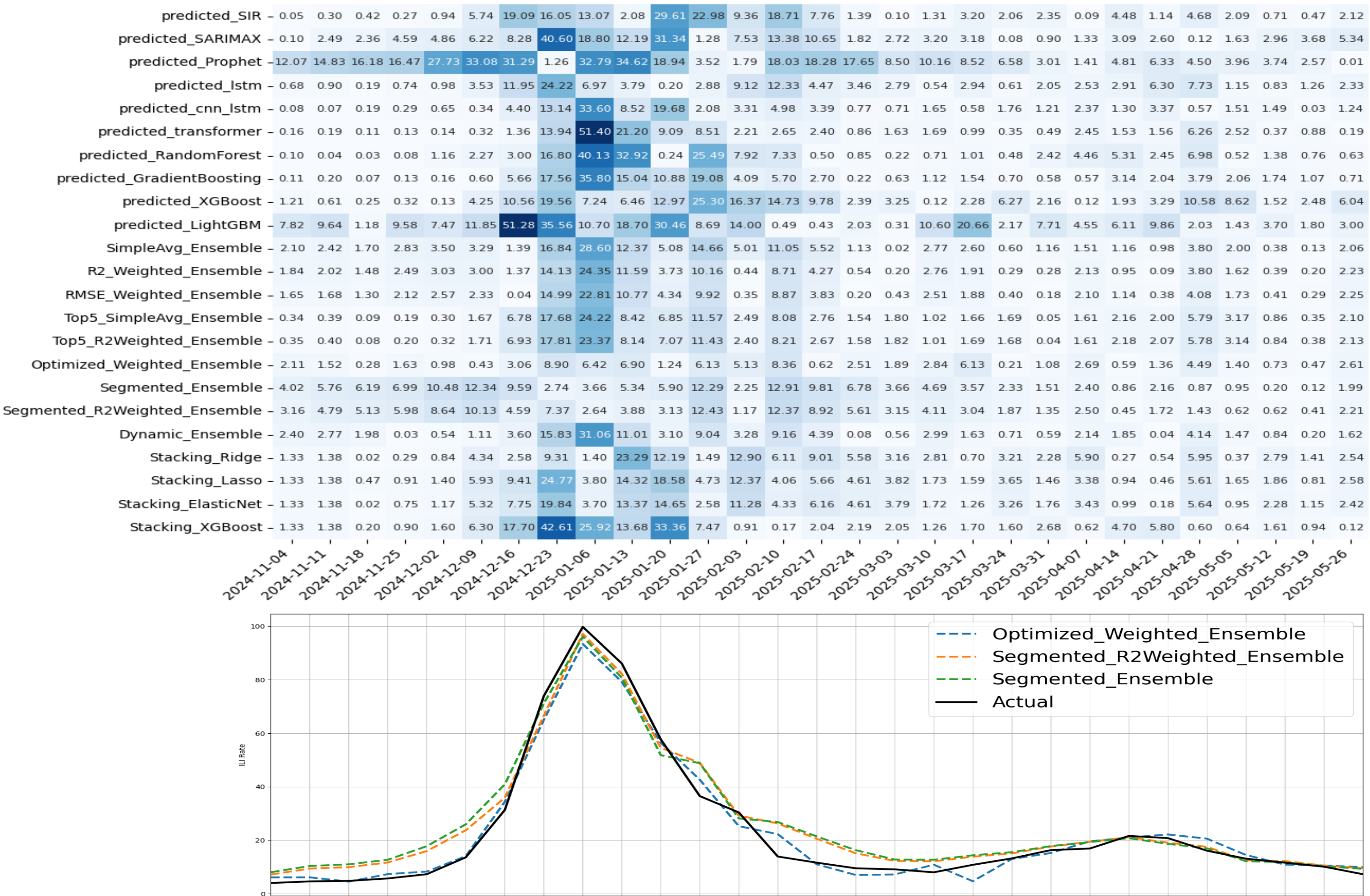
[그림2] 인플루엔자 예측을 위한 앙상블 모델링 프레임워크

Results



[그림3] 예측 모델별 성능 비교: 결정계수(R<sup>2</sup>), 평균절대오차(MAE), 평균제곱근오차(RMSE)

- 단순 평균 앙상블이 대부분의 개별 모델보다 좋은 성능을 보여, 다양한 앙상블 기법을 추가로 구축함
  - R<sup>2</sup> 및 MAE 히트맵에서 앙상블 모델들이 전반적으로 밝게 나타나는 것으로 보아, 개별 모델에 비해 더 높은 설명력과 낮은 오차를 보인다는 점을 확인할 수 있음
  - RMSE 지표표를 기준으로 정렬한 표에서도 **앙상블 모델**이 대부분 상위권에 분포하여, 예측 정확도 측면에서 **개별 모델**보다 뛰어난 성능을 보임
- 앙상블 기법은 개별 모델보다 더 높은 정확성과 안정성을 갖춘 효과적인 예측 방법임



[그림4] 시계열 예측 성능 비교: 모델별 주차 단위 RMSE 히트맵 및 상위 3개 모델 예측 결과

- 방역 대응 및 의료 자원 관리를 위해 감염병 예측에서 가장 중요한 부분 중 하나는 유행 시기의 급격한 변화를 정확히 예측하는 것
  - RMSE 히트맵을 보면, 상단에 위치한 개별 모델들보다 하단의 앙상블 모델들이 피크 구간에서 전반적으로 더 밝은 색을 띠고 있으며, 이는 유행 시기의 예측 오차가 더 작다는 것을 의미함
  - 시계열 그래프에서는 RMSE 기준 상위 3개 모델 모두가 앙상블 기법에 기반한 모델임을 확인할 수 있으며, 이들의 예측값은 실제값과 유사한 추세를 나타내어 예측 성능이 뛰어난임을 알 수 있음
- 앙상블 기법은 인플루엔자 유행 시점을 더 정확히 예측함으로써, 선제적 방역 대응과 의료 자원 준비에 실질적인 도움을 줄 수 있을 것으로 기대됨

Discussion

Conclusion

본 연구는 수리·통계·머신러닝·딥러닝 기반의 다양한 개별 예측 모델을 바탕으로, 이들의 성능을 결합한 여러 형태의 앙상블 모델을 구성하여 인플루엔자 의사환자분율의 예측 정확도를 비교함. 실험 결과, 앙상블 모델 전반이 개별 모델 대비 우수한 예측 성능을 보였으며, 특히 RMSE, MAE, R<sup>2</sup> 모든 지표에서 평균적으로 더 낮은 오차와 높은 설명력을 기록함. 개별 모델 중 가장 성능이 우수했던 딥러닝 기반 모델(LSTM)의 경우에도, 다수의 앙상블 모델이 이를 상회하는 결과를 나타냄. 앙상블 모델은 급격한 유행 구간이나 예측 난이도가 높은 시점에서도 상대적으로 안정적인 추세 예측 능력을 보임. 이러한 결과는 다양한 모델의 특성을 통합한 예측 전략이 인플루엔자와 같은 비정형적 감염병의 단기 예측에 효과적임을 시사함.

Futureworks

- 고위험군을 대상으로 한 개별 모델 구축
- 유아와 노인층은 면역력이 약해 감염 시 중증으로 발전할 가능성이 높기 때문에 별도의 예측 체계를 마련하는 것은 공중보건 측면에서 큰 의의가 있음. 고위험군 대상 맞춤형 예측 모델 개발을 통해 의료 자원의 효과적인 분배와 사전 개입 전략 수립이 가능할 것으로 기대됨.
- 지역별 데이터를 활용한 정교한 예측 모델 개발
- 지역마다 인구 구조, 보건 인프라, 유행 양상이 상이하기 때문에 동일한 예측 모델을 일괄적으로 적용하는 데는 한계가 존재함. 지역별 데이터로 지역 특성에 맞춘 정교한 예측이 가능해지면 이를 바탕으로 지역 보건당국이 선제적으로 대응할 수 있는 기반을 마련할 수 있을 것으로 기대됨.

