

대기오염물질 농도에 따른 천식 응급환자 수 예측 연구

이한주*, 지민규**, 김청원***

목 차

요약	그렇지 않은 경우에 따른 예측 결과.
1. 서론	
2. 선행연구	4.2 천식 응급환자 수와 양의 상관관계를 가지는 단일 대기오염물질을 이용한 예측 결과
2.1 환경성 질환과 대기 오염 물질에 대한 선행연구	4.3 상관관계가 높았던 3개 대기오염물질의 조합별로 예측을 진행하였을 때 실험 결과
2.2 시계열 예측에 관한 선행 연구	4.4. 연구 결과 및 결론
3. 제안방법	References
3.1 이용모델	Abstract
3.2 데이터	
4. 결과	
4.1 응급환자 수를 이용한 경우와	

요약

산업이 발전하면서 대기오염물질에 대한 관심도는 높아졌다. 대기오염물질은 환경오염, 지구 온난화 등 다양한 분야에 영향을 미쳤다. 그 중 환경성 질환은 대기오염물질에 의해 영향을 받은 분야 중 하나이다. 대기오염물질은 분자의 크기가 작아 인체의 피부나 호흡기를 통해 영향을 미칠 수 있다. 이러한 점 때문에 대기오염물질과 환경성 질환에 대한 연구가 다양하게 진행됐다. 환경성 질환의 일환인 천식은 증상이 심해져 천식발작을 일으킬 경우 생명에 위협을 줄 수 있고 성인 천식의 경우 한번 발병을 하면 완치가 어렵다. 천식을 악화시키는 요인에는 황사, 대기오염이 포함된다. 전 세계적으로 천식은 유병률이 증가하고 있는 추세이다. 본 논문에서는 대기오염물질이 천식 환자의 응급실 입원 건수와 어떤 상관관계를 가지는지 연구하고 상관관계가 높은 대기오염물질을 이용하여 미래의 천식 환자 수를 예측했다. 대기오염물질은 이황산가스(SO_2), 일산화탄소(CO), 오존(O_3), 이산화질소(NO_2)와 미세먼지(PM_{10}) 5가지 대기오염물질의 농도를 이용하고 환경성 질환은 천식 환자의 응급실 입원 건수 데이터를 이용하였다. 대기오염물질과 천식의 응급환자 수에 대한 데이터는 2013년 1월1일 부터 2017년 12월 31일 까지 총 5년 치의 데이터를 이용하였다. 모델은 Informer와 LTSF-Linear의 두 가지 모델을 이용하여 예측을 진행하였고 모델의 성능을 측정하기 위해 MAE, MAPE, RMSE 의 성능지표를 이용했다. 천식의 응급환자 수 예측은 응급환자 수를 포함하여 예측을 진행한 경우와 포함하지 않고 진행한 두 가지 경우 모두 진행하여 결과를 비교했다. 본 논문은 Informer와 LTSF-Linear 모델을 이용한 천식 응급환자 수의 예측에 모델의 성능을 향상 시키는 대기 오염물질을 제시한다.

포제어: 환경성 질환, 대기오염물질, 미세먼지, 시계열 예측, 헬스케어

접수일(2023년 2월 03일), 수정일(1차: 2023년 2월 24일), 게재확정일(2023년 3월 1일)

이 논문은 환경부 “기후변화특성화대학원사업”의 지원에 의한 결과임.

* 제1저자, 세종대학교, 소프트웨어학과, 석사과정, jeffrey7108@gmail.com

** 공동저자, (주)셀버스, 과장, mk92@hanmail.net

*** 교신저자, 세종대학교, 소프트웨어학과, 교수, wikim@sejong.ac.kr

1. 서론

산업이 발전하면서 미세먼지와 같은 대기오염물질에 대한 관심도는 점점 높아졌다. 대기오염물질은 지구 온난화, 환경오염 등 실제 생활에 영향을 미칠 수 있는 분야와 직접적으로 연관되어 있다. 특히 인체에 직접적인 영향을 미치는 환경성 질환과 연관이 있으며, 대기오염물질이 실제 인체에 미치는 영향에 대한 관심이 높아지고 있다. 실제로 이러한 대기오염이 인체 질환에 미치는 영향에 대한 관심이 증가하면서 국제적으로 환경성 질환이라는 용어가 새롭게 제시되며, 국내외적으로 다양한 연구가 진행되고 있다. 대기오염물질의 가장 큰 문제점은 기체의 물질이 해당 지역의 모든 인원에게 노출될 수 있는 잠재적 위험을 가지고 있다는 것이다. 동시에 물질이 눈에 보이지 않아 어느 정도로 영향을 미치고 있는지 파악하기 어렵다. 대기오염물질은 사람에게 다양한 신체적, 정신적 문제를 초래할 수 있지만, 특히 비염, 천식 등 환경성 질환에 대한 영향이 크게 나타난다. 이러한 문제를 인식하여 세계보건기구(World Health Organization, 2006)에서는 환경성 질환을 일으킬 수 있는 주요 환경요인 14개와 주요 102가지 질환에서 환경요인에 의해 발생할 수 있는 환경성 질환 85개를 분류하여 제시하였다. 여기서 주요 환경 요인에는 대기, 실내 오염, 기후변화 등이 포함되며, 환경성 질환에는 천식, 만성폐쇄성 폐질환 등이 포함된다.(Pruss-Ustun et al., 2016) 국내에서는 환경성 질환 연구를 위해 국민건강보험에서 2013년부터 5년간의 환경성 질환 진료 데이터를 제공하고 있다. 해당 자료에서는 환경성 질환 중 비염, 천식, 아토피에 대한 일별 환자 수를 제공하고 있다. 또한 국내에서는 위와 같은 문제에 대응하기 위하여 질환 데이터뿐만 아니라 대기오염 정보에 대한 관측 및 경보를 지속적으로 실시하고 있다. 1995년에 미세먼지 기준을 마련하고 관측을 실시하고 있으며, 최근 크게 이슈가 되고 있는 초미세먼지 또한 2011

년에 기준을 마련하고, 2015년부터 적용하여 관측을 실시하고 있다. 그 외에도 아황산가스, 일산화탄소 등 다양한 대기오염물질에 대한 기준을 마련하고 농도를 시간별로 측정하여 제공하고 있다. 미세먼지와 같은 입자상 물질은 호흡곤란과 같은 호흡기계 질환 증가, 만성 기관지염의 발생, 천식 악화, 경미한 심장마비 등의 건강영향을 유발한다.(Kim and Lim, 2020) 환경성 질환의 하나인 천식은 만성적이고 재발이 잦은 질환으로 증상이 심해져 천식발작을 일으킬 경우 생명에 위협을 줄 수 있다.(Seoul national university hospital, 2023) 천식은 선진국형 질환으로 사회의 발전에 따라 유병률이 증가하고, 세계적으로 천식 유병률은 증가하고 있는 추세이다.(Lee, 2020) 천식을 악화시키는 요인에는 대기오염이 포함된다.(Park, 2018) 이번 연구의 목적은 천식 환자의 응급실 입원 건수와 대기오염물질의 단기적인 관련성을 확인하고 더 나아가 상관관계를 가지는 대기오염물질을 확인하고 해당 대기오염물질을 이용하여 예측을 진행한 결과를 제시하는 것이다. 시계열 예측은 순차적으로 기록된 데이터를 이용하여 미래를 예측하는 분야이다. 본 논문에서는 시계열 예측 모델인 Transfomer기반의 Informer 모델과 LTSF-Linear 모델을 이용하여 미래의 천식 응급환자 수를 예측할 때, 가장 높은 예측 성능을 나타내는 대기오염물질을 제시한다.

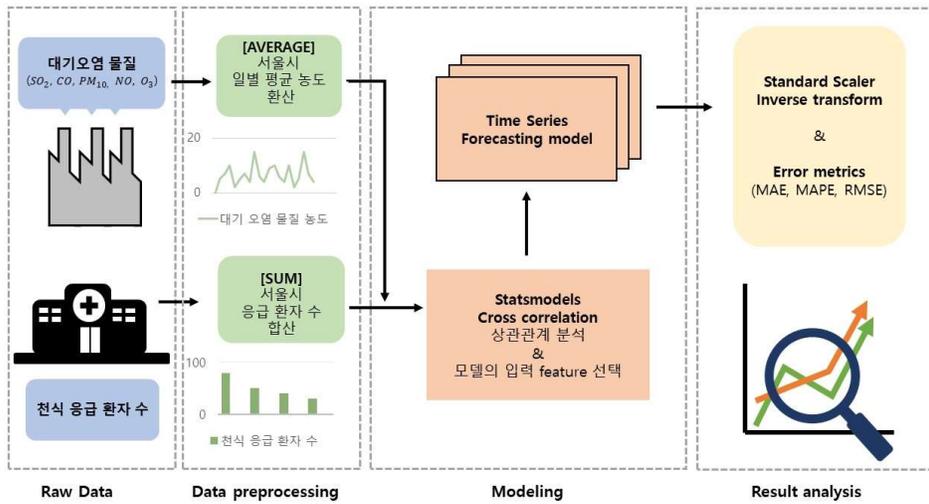


Fig 1-1. Overview of the paper

2. 선행 연구

2.1 환경성 질환과 대기오염물질에 대한 선행 연구

대기오염물질과 환경성 질환의 관계는 다양한 방법으로 연구가 진행됐다. Parker et al(2009)는 미국의 어린이를 대상으로 관계를 연구했다. 1999년부터 2005년까지 3-17세 어린이 7만 명의 건강조사 데이터와 대기오염물질 데이터를 활용했다. 논문은 관계 분석을 위해 Logistic Regression을 사용했으며, 오존과 초미세먼지가 호흡기 알레르기, 알러지성 비염에 영향을 미친다는 결과를 제시했다. Cho et al.(2013)는 서울시의 25개 자치구를 대상으로 환경성 질환 관계를 분석하는 연구를 진행했다. 연구는 2005년부터 2009년까지의 패널 데이터를 활용하였다. 자치구별 인구차이를 해결하기 위해 자치구 인구수 중 환경성 질환의 환자 수를 비율로 표현하였다. 모델은 패널데이터의 회귀 분석 모델을 활용하였다. 아황산가스와 오존은 알레르기 비염 환자 비율, 미세먼지는 천식 환자 비율 증가에 영향을

준다는 결과를 제시했다. Yara et al.(2016)는 환경성 질환의 입원환자 수를 예측하는 모델을 제안했다. 예측 모델 학습을 위해 2007년부터 2009년까지 브라질 Campinas의 기상 데이터와 환경성 질환의 병원 입원환자 수 데이터를 활용했다. 모델은 미세먼지, 온도, 습도를 입력으로 하였으며, Fully Connected Layer를 통과하여 입원환자 수를 출력하였다. Cao et al.(2017)는 중국 내 공간적 정보를 포함하여 대기오염물질과 환경성 질환의 관계성을 분석하였다. 데이터는 2004년부터 2008년까지 중국의 기상 데이터와 호흡기 질환 데이터를 사용하였다. 중국의 각 지역의 대기오염물질 및 호흡기 질환 데이터를 활용하면서, 각 지역의 주변 지역 정보를 확인하고, 관계성 분석에 고려하였다. 이를 위해 중국을 31개의 지역으로 구분하고 인접 지역에 대한 대기 데이터를 추가로 활용했다. Polezer et al.(2018)는 미세먼지뿐만 아니라 인체에 영향을 미칠 가능성이 있는 원소들의 농도 정보를 수집했다. 이 데이터를 활용하여 환경성 질환의 병원 입원환자 수를 예측하는 MLP, ESN, ELM 모델을 제시했다. Seo and Lee(2019)는 국민건강보험에서 제공하는

환경성 질환 환자 수와 기상 데이터를 분석했다. 해당 논문은 분석을 통해 오존과 미세먼지가 알레르기 비염에 정방향 영향을 주며, 이황산 가스가 천식에 정방향 영향을 주는 것으로 결과를 제시했다. 반대로 미세먼지와 오존은 천식에 역방향 영향을 주는 것으로 제시했다. Zeinab et al.(2020)는 기상 데이터와 대기오염물질 농도 데이터를 활용하여 환경성 질환 위험도를 예측하는 연구를 진행했다. 모델은 ANN을 기반으로 기상의 온도, 풍속, 습도 등과 대기오염물질의 농도를 입력으로 학습했다. 출력은 Low Risk, Medium Risk, High Risk의 세 가지 클래스로 분류했다. Park et al.(2020)는 대한민국의 대기오염물질과 환경성 질환 환자 수 데이터를 활용해 관계를 분석했다. 분석 시 데이터를 지역별, 일수별로 구분하여 분석하였다. 해당 논문은 5일 이동 평균 모형이 환경성 질환의 외래이용건수에 가장 높은 영향력을 보이는 것을 확인하였다. Qiu et al.(2012)는 과거 4일 동안의 평균 대기오염물질 농도에 초점을 맞추어 분석을 진행하였고 4일간의 미세먼지 평균 농도는 응급의료건수와 관계성을 가지는 것을 확인하였다.

2.2 시계열 예측에 관한 선행 연구

시계열 예측(Time-series forecasting)은 환경, 질병, 날씨, 금융 등 많은 분야에서 사용되고 있다. 최근 시계열 예측에 대한 연구는 활발하게 이루어지고 있고 통계 기법을 이용한 예측과 딥러닝을 이용한 예측 등의 방법들이 있다. 시계열 예측에 있어 Transformer 기반의 예측 모델들은 최근에 큰 주목을 받고 있다. Transformer는 자연어처리 뿐만 아니라 이미지 예측 그리고 시계열 데이터의 예측에 있어서도 좋은 성능을 보였고 Transformer 기반의 시계열 예측 모델들은 SOTA를 달성한 바 있다. Zhou et al(2021)는 기존 Transformer 모델을 이용한 시계열 데이터 예측에서의 문제점을

self-attention에 따른 이차계산 복잡도, 긴 시계열 데이터를 입력 값으로 받았을 시에 메모리 병목현상, 긴 기간을 예측할 때 속도 급락의 3가지 문제점을 제시하였다. 해당 논문에서는 이러한 문제점을 개선한 Informer라는 모델을 제시한다. Zeng et al.(2022)는 Transformer기반의 시계열 예측 모델들은 시간적 정보의 손실이 불가피하고 look-back window size를 늘려도 에러율이 줄어들지 않는다는 문제점을 제시하였다. 해당 논문에서는 LTSF-linear 라는 one-layer linear model를 제시한다. 논문에서 제시된 모델은 9개의 데이터 셋에서 Transformer 기반의 예측 모델 보다 낮은 에러율을 보인다는 결과를 제시하였다.

3. 제안 방법

3.1 이용모델

대기오염물질 데이터를 이용한 응급환자 수의 예측에 LTSF-linear 와 Informer 모델의 성능을 각각 비교하여 제안한다. Fig. 3-1은 Informer 모델의 전반적인 구조를 나타내는 그림으로 해당 모델은 Transformer 기반의 모델로 기존 모델의 self-attention이 아닌 Multi-head ProbSparse Self-attention을 이용한다. LTSF-Linear 모델의 전반적인 구조는 Fig. 3-2의 그림과 같고 해당 모델은 여러 layer의 linear 모델로 구성된다. LTSF-Linear 모델은 D-Linear 모델과 N-Linear 모델이 존재하는데 본 논문에서는 D-Linear 모델을 이용하여 실험을 진행한다. Informer 모델과 LTSF-Linear 모델은 단변량(univariate)과 다변량(multivariate) 데이터에 대한 예측이 가능하다. 단변량은 종속변수 즉, 결과변수가 하나인 경우를 의미하며 다변량은 결과변수가 하나 이상인 경우를 의미한다. 실험은 2013년 1월 1일부터 2017년 12월 31일 까지 총 1826개 데이터에 대하여 train set,

validation set, test set의 비율을 7:1:2로 설정하고 100 epoch 동안 학습한 모델을 기준으로 실험을 진행한다. 또한, 입력 값이 2개 이상일 경우 천식환자에 대한 단변량 예측과 다변량 예측을 함께 진행한다.

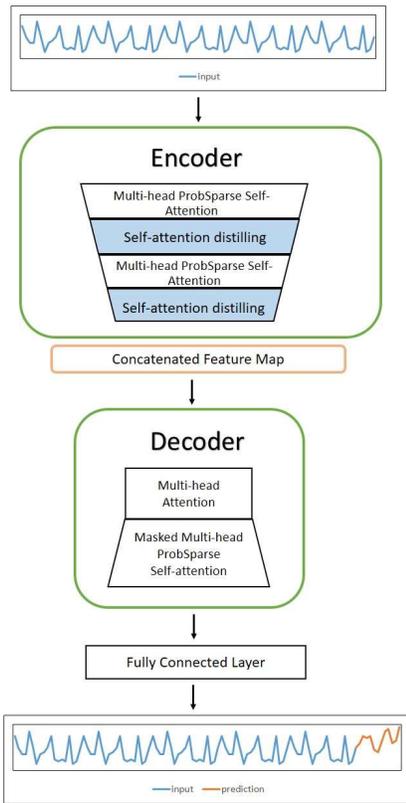


Fig. 3-1 The overall architecture of Informer model

3.2 데이터

연구에 필요한 분석 모델을 위해 미세먼지를 포함한 대기오염물질 및 환경성 질환의 환자 수 데이터를 수집했다. 대기오염물질은 한국 환경공단 에어코리아에서 제공하는 데이터를 활용했다.

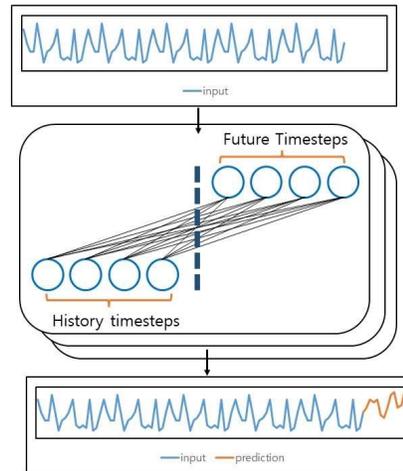


Fig. 3-2 The overall architecture of LTSF-Linear model

3.2.1 대기오염물질 데이터

에어코리아의 대기오염물질 데이터는 각 관측소에서 관측한 대기오염물질 6개에 대한 시간별 농도 값을 제공한다. 6개의 대기오염물질은 미세먼지(PM_{10}), 초미세먼지($PM_{2.5}$), 아황산가스(SO_2), 일산화탄소(CO), 오존(O_3)과 이산화질소(NO_2)로 구성된다. 여기서 미세먼지 및 초미세먼지는 정수형 농도 정보를 제공하며 단위는 $\mu g/m^3$ 를 사용한다. 나머지 대기오염물질은 실수형 농도 정보를 제공하며 단위는 Parts-per Million(PPM)을 사용한다. 다만, 관측소 장비 및 상황에 따라 일부 정보가 없는 Row가 존재하며, 초미세먼지는 2015년부터 관측을 시작하여 데이터를 제공하고 있다. 에어코리아는 493개의 관측소로부터 대기오염물질 관측을 시작한 1990년부터의 데이터를 모두 제공하고 있다. 하지만 분석을 위한 환경성 질환의 환자 수 데이터가 2013년부터 2017년까지 5년간의 데이터만 제공하고 있어, 대기 오염물질 농도도 그에 맞게 2013년부터 2017년까지 데이터를 활용했다. 대기오염물질 농도는 각 서울시내의 측정소별로 일별 평균 농도를 구한 이후 서울특별시의 평균 대기오염물질 농

도를 구하여 입력 값으로 이용한다.

3.2.2 환경성 질환자 수 데이터

환경성 질환 데이터는 국민건강보험공단에서 제공하는 환경성 질환 진료 데이터를 활용했다. 진료 데이터는 2013년부터 2017년까지 총 5년간의 데이터를 제공한다. 데이터는 지역 구분을 위해 시, 구, 동 단위까지 구분하였으며, 총 3649개의 동 단위 지역이 존재한다. 데이터는 환자 구분을 세분화하기 위해 나이 대와 성별을 추가로 제공한다. 따라서 환자 수의 구분은 각 시, 구, 동의 인구 중 나이 대 및 성별에 따른 일별 환자 수를 제공한다. 데이터는 환경성 질환 중 아토피, 천식과 비염의 환자 수를 제공한다. 이 때, 환자 수는 외래 건수, 입원 건수, 응급의료 이용 건수를 나타낸다. 본 논문에서는 서울시의 천식 환자의 응급의료 이용건수를 예측한다.

3.2.3 데이터 전처리

데이터 분석을 위해서 대기오염물질 데이터와 환경성 질환자 수 데이터의 전처리가 필요하다. 두 데이터를 제공하는 기관이 다르기 때문에 위치, 측정 간격 등이 모두 다르다. 에어코리아 데이터의 경우 관측소 번호와 주소를 제공하며, 대기오염물질 농도를 시간별로 제공하고 있다. 국민건강보험공단 데이터의 경우 지역 코드와 시, 구, 동 정보를 제공하며, 환자 수를 일별로 제공하고 있다. 따라서 에어코리아의 관측소 및 주소와 국민건강보험공단의 지역코드 및 시, 구, 동 정보를 일치시키는 작업이 필요하다. 이를 위해 에어코리아의 주소 정보를 시, 구, 동 단위로 분할하고, 국민건강보험공단의 지역 정보와 일치하도록 데이터를 연결시켰다. 에어코리아의 데이터는 시간별, 국민건강보험공단의 데이터는 일별 정보이기 때문에, 두 데이터를 통합하기 위해 측정 기간을 통일시켰다. 일별 측정 데이터를 시간별로

분할할 수 없기 때문에 에어코리아의 시간별 정보를 일별 정보로 변환했다. 에어코리아에서 일별로 획득한 00시부터 23시까지 측정 데이터를 최저, 최고, 평균 농도로 합치고 일별로 통합했다. 따라서 시간별 대기오염물질의 농도로 구성됐던 에어코리아 데이터를 일별 대기오염물질의 평균 농도로 변환했다.

3.2.4 천식 응급환자 수와 대기오염물질간의 상관관계

응급실의 입원 날짜를 기준으로 전날의 대기오염물질의 일별 평균 농도와 상관관계를 확인한다. 대기오염물질은 미세먼지(PM_{10}), 아황산가스(SO_2), 일산화탄소(CO), 오존(O_3)과 아황산가스(NO_2)의 5가지 오염물질별 일별 평균농도를 이용하여 실험을 진행한다. 응급환자 수와 높은 상관관계를 가지는 대기오염물질이 시계열 예측 모델을 이용할 때, 결과에 영향을 미치는지의 여부를 확인한다. 상관관계를 판단하는 기준은 아래의 Tab. 3-1를 기준으로 판단하고 대기오염물질에 따른 상관관계는 Tab.3-2와 같았다. 응급환자 수와의 상관관계가 가장 높았던 전날의 대기오염물질은 일산화탄소(CO)와 아황산가스(SO_2)로 상관관계수가 0.23으로 상관관계가 있으나 낮음을 확인하였다. 초미세먼지($PM_{2.5}$)와 아황산가스(NO_2)는 천식 응급환자 수와 상관관계가 없음을 확인했다. 오존(O_3)는 응급환자 수와 음의 상관관계가 있으나 그 정도가 낮음을 확인했다. 이에 따라 본 논문에서는 상관관계가 높은 물질을 이용하여 예측을 진행할 때 Informer와 LTSF-Linear 두 모델의 예측 결과에 영향을 미치는지 확인한다.

Tab. 3-1 Correlation judging criteria

상관계수(r)	상관관계
±0.9 이상	상관관계가 아주 높음.
±0.7 ~ 0.9	상관관계가 높음.
±0.4 ~ 0.7	상관관계가 있음.
±0.2 ~ 0.4	상관관계가 있으나 낮음.
±0.2미만	상관관계가 없음.

Tab. 3-2 Correlation between the number of emergency patients and air pollutants (Round to the third decimal place)

대기오염물질	응급환자 수와의 상관계수
CO	0.23
SO ₂	0.23
PM ₁₀ (미세먼지)	0.21
PM _{2.5} (초미세먼지)	0.12
NO ₂	0.12
O ₃	-0.2

4. 연구 결과 및 결론

4.1 응급환자 수를 이용한 경우와 그렇지 않은 경우에 따른 예측 결과

상관관계가 높은 대기오염물질을 이용하여 예측하는 것이 예측 성능을 향상 시키는지 확인하기 위해 에어코리아에서 제공하는 대기오염물질 데이터 중 2015년 이후의 데이터만 존재하는 초미세먼지(PM_{2.5})를 제외한 미세먼지(PM₁₀), 아황산가스(SO₂), 일산화탄소(CO), 오존(O₃)과 이산화질소(NO₂)를 모두 이용하여 예측한 결과를 확인한다. 뿐만 아니라, 해당 실험은 천식 응급환자 수의 입력 값 포함여부에 따른 예측 결과를 확인하기 위해 포함 여부에 따라 2가지 경우로 나누어 예측을 진행하였고 결과는 Tab. 4-1의 표와 같았다. LTSF-Linear 모델은 응급환자 수를 포함하지 않은 경우 MAPE가 약 1.38% 증가했다. Informer 모델

은 응급환자 수를 포함하지 않은 경우 MAPE가 약 0.33% 더 낮아졌다. 해당 결과 내에서는 5개의 대기오염물질을 모두 이용한 예측에서 MAPE는 15.12%, MAE는 10.16이 가장 낮은 값으로 나타났다. 실험의 test set은 2015년 1월 1일부터 2017년 12월 31일까지의 데이터로 해당 기간에서 서울시의 천식 응급환자 수의 평균값은 약 72명이었다. MAE에서 10.16의 수치는 서울시의 일별 평균 천식 응급환자 수 72명을 기준으로 실제 값보다 평균적으로 10명을 적게 예측하거나 10명을 많게 예측했다는 것을 의미한다.

4.2 응급환자 수를 이용한 경우와 그렇지 않은 경우에 따른 예측 결과

천식 응급환자 수와 양의 상관관계를 가지는 대기오염물질 아황산가스(SO₂), 일산화탄소(CO), 미세먼지(PM₁₀) 3가지 대기오염물질을 Informer 및 LTSF-Linear 모델에 입력 값으로 이용하고 응급환자 수를 포함한 경우와 포함하지 않은 경우에 대해 각각 예측을 진행하고 결과를 확인했다. 고농도의 아황산가스(SO₂)는 옥외활동이 많고 천식에 걸린 어른과 어린이에게 일시적으로 호흡장애를 일으킬 수 있으며, 고농도에 폭로될 경우 호흡기계 질환을 일으키고 심장혈관 질환을 악화시킨다. (AirKorea, 2023) 이러한 아황산가스(SO₂)는 천식의 응급환자 수를 예측하는 데에 있어 가장 낮은 MAPE를 나타냈고 결과는 Tab. 4-2의 표와 같았다. 5개 대기오염물질을 이용한 실험결과 가장 낮았던 MAPE는 15.12 % 이었는데, 아황산가스(SO₂)만을 이용하여 예측을 실험한 결과에서 MAPE가 13.54%로 약 1.58% 더 낮았다. 약 1.58% 더 낮아진 MAPE는 서울시의 천식 응급환자 수의 예측에 있어 test set의 평균 응급환자 수 약 72명을 기준으로 이전 예측 결과보다 평균적으로 약 1명의 응급환자 수를 더 정확하게 예측했음을 나

타낸다.

4.3 상관관계가 높았던 3개 대기오염물질의 조합 별로 예측을 진행하였을 때 실험 결과

해당 실험에서는 천식 응급환자 수와 양의 상관 관계를 가지는 3가지 대기오염물질 아황산가스(SO_2), 일산화탄소(CO), 미세먼지(PM_{10})의 조합 별로 Informer 및 LTSF-Linear 모델에 입력 값으로 이용하여 예측을 진행했다. 예측 결과는 Tab. 4-3의 표와 같았다. 아황산가스(SO_2)와 미세먼지를 이용한 예측이 MAPE가 13.30%로 가장 낮았다. 단일 대기오염물질만을 이용하여 예측한 결과 중 가장 MAPE가 낮게 나타났던 결과보다 약 0.24% 낮아진 수치로써, 아황산가스와 미세먼지를 이용하여 천식 응급환자 수를 예측 하는 것이 가장 적은 오차를 보이는 것으로 나타났다.

4.4 결론

본 논문에서는 천식 응급환자 수 예측에 대한 실험을 Informer모델과 LTSF-Linear모델을 이용해 진행하였다. 예측 결과 내에서 가장 낮은 MAPE 값이 나왔던 경우는 아황산가스(SO_2)와 미세먼지(PM_{10})를 이용하여 예측을 진행한 경우로 MAPE가 13.30% MAE가 9.52인 것으로 나타났다. 해당 결과는 평균 약 72명인 test set에 대하여 평균적으로 약 10 명의 오차가 있고 이 값을 퍼센트로 환산하면 약 13.30%의 오차를 보인다는 것을 나타낸다. 단일 대기오염물질만을 이용한 경우는 아황산가스를 이용하여 예측을 진행한 경우로 MAPE가 13.54%인 것으로 나타났다. 예측 결과를 통해 천식의 응급환자 수와 가장 상관관계가 높았던 아황산가스와 일산화탄소 중 아황산가스를 이용한 예측이 가장 낮은 MAPE 값을 보이고 일산화탄소는 천식의 응급환자 수의 예측 결과인 MAPE에 영향을 미치지 않고 오

차는 줄어들지 않는 것으로 나타났다. 본 연구는 아황산가스와 미세먼지가 천식의 응급환자 수 예측에 영향을 미친다는 것을 시사한다.

Tab. 4-1 Prediction results based on the use of the number of emergency patients and not (Round to the third decimal place)

error	LTSF-Linear		Informer	
	응급환자 수 포함	응급환자 수 제외	응급환자 수 포함	응급환자 수 제외
MAE	11.58 명	12.27 명	10.31 명	10.16 명
MAPE	17.55 %	18.93 %	15.45 %	15.12 %
RMSE	14.76 명	15.57 명	13.45 명	13.17명

Tab. 4-2 Prediction results based on highly correlated air pollutants (Round to the third decimal place)

학습데이터	error	LTSF-Linear		Informer	
		응급환자 수 포함	응급환자 수 제외	응급환자 수 포함	응급환자 수 제외
CO	MAE	11.50 명	12.90 명	11.01 명	11.28 명
	MAPE	17.45 %	19.46 %	16.93%	17.56 %
	RMSE	14.64 명	16.73 명	13.67 명	14.02 명
SO ₂	MAE	11.52 명	11.86 명	9.60 명	9.58 명
	MAPE	17.48 %	17.42%	13.64 %	13.54 %
	RMSE	14.68 명	15.31 명	12.82 명	12.96 명
PM ₁₀	MAE	11.58 명	12.27 명	9.70 명	9.65 명
	MAPE	17.61 %	18.93 %	14.71 %	14.70 %
	RMSE	14.73 명	15.57 명	12.9 명	12.94 명

Tab. 4-3 Prediction results by combination of highly correlated air pollutants (Round to the third decimal place)

학습데이터	error	LTSF-Linear		Informer	
		응급환자 수 포함	응급환자 수 제외	응급환자 수 포함	응급환자 수 제외
CO, SO ₂	MAE	11.55 명	12.90 명	11.12 명	10.22 명
	MAPE	17.52 %	19.46 %	16.92 %	15.25 %
	RMSE	14.73 명	16.73 명	14.15 명	13.21 명
CO, PM ₁₀	MAE	11.56 명	12.27 명	11.07 명	10.31 명
	MAPE	17.55 %	18.93 %	17.16 %	15.85 %
	RMSE	14.73 명	15.57 명	14.2 명	13.13 명
SO ₂ , PM ₁₀	MAE	11.58 명	12.27 명	9.88 명	9.52 명
	MAPE	17.58 %	18.93 %	13.83 %	13.30 %
	RMSE	14.77 명	15.57 명	13.40 명	12.97 명

References

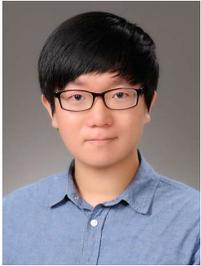
- [1] Airkorea (2023), "아황산가스", 2023년 02월 24일 접속,
https://www.airkorea.or.kr/web/airMatter?pMENU_NO=130
- [2] Cao, Q., Liang, Y., and Niu, X. (2017), China's air quality and respiratory disease mortality based on the spatial panel model, *International Journal of Environmental Research and Public Health*, 14(9), 1081.
- [3] Cho, Il H., Ju, H. J., and Kwon, G. H. (2013), The Impact Analysis of Air Pollutants on Increasing Environmental Disease : Focusing on Allergic Rhinitis and Asthma in Seoul Metropolitan City, *Seoul Studies*, 14(2), 97-114. (조일형, 주희진, 권기현 (2013), 서울시 대기오염물질이 환경성 질환자 증가에 미치는 영향 분석: 알레르기 비염과 천식을 중심으로. *서울도시연구*, 14(2), 97-114)
- [4] Kim, M.J. and Lim, G.L. (2020), Bigdata Analysis of Fine Dust Theme Stock Price Volatility According to PM10 Concentration Change, *Journal Of Service Research and Studies*, 10(1), 55-67. (김무정, 임규진 (2020), PM10 농도변화에 따른 미세먼지 테마주 주가변동 빅데이터 분석, *서비스 연구*, 10(1), 55-67)
- [5] Lee, K.H. (2020), "Effects of Psychological and Physiological Factors on Asthma in Korean Adults", *Science of Emotion & Sensibility*, 23(2), 13-22. (이경희 (2020), "심리 및 생리적 요인이 성인천식에 미치는 영향", *감성과학*, 23(2), 13-22)
- [6] Park, H.K. (2018), Air pollution and climate change: Effects on asthmatic patients, *Allergy, Asthma & Respiratory Disease*, 6(2), 79-84.
- (박혜경 (2018), 대기오염과 기후 변화: 천식에 미치는 영향, *Allergy, Asthma & Respiratory Disease*, 6(2), 79-84)
- [7] Park, J.H., Park, Y.Y., Lee, E.J., and Lee, K.S. (2020), Analysis on the effects of particular matter distribution on the number of outpatient visits for allergic rhinitis, *Health Policy and Management*, 30(1), 50-61. (박주현, 박영용, 이은주, 이광수 (2020), 지역별 미세먼지 농도의 알레르기비염 외래이용에 대한 분석, *보건행정학회지*, 30(1), 50-61)
- [8] Parker, J.D., Akinbami, L.J. and Woodruff, T.J. (2009), Air pollution and childhood respiratory allergies in the United States, *Environmental Health Perspectives*, 117(1), 140-147
- [9] Polezer, G., Tadano, Y.S., Siqueira, H.V., Godoi, A.F., Yamamoto, C.I., de Andre, P.A....., and Godoi, R.H. (2018), Assessing the impact of PM2.5 on respiratory disease using artificial neural networks, *Environmental Pollution*, 235, 394-403
- [10] Pruss-Ustun, A., Wolf, J., Corvalan, C., Bos, R., and Neira, M. (2016), Preventing disease through healthy environments: a global assessment of the burden of disease from environmental risks, World Health Organization
- [11] Qiu, H., Yu, I., Tian, L. T., Wang, X., Tse, L. A., Tam, W., and Wong, T. (2012), Effects of coarse particulate matter on emergency hospital admissions for respiratory diseases: a time-series analysis in Hong Kong, *Environmental Health Perspectives*, 120(4), 572-576.
- [12] Seo, H. J. and Lee, H. S. (2019), How air pollutants influence on Environmental diseases

- : Focused on Seoul Metropolitan Area, *Seoul Studies*, 20(3), 39–59 (서형준, 이형석 (2019), 대기오염물질이 환경성질환에 미치는 영향: 수도권 지역의 자치구 및 시·군을 중심으로. *서울도시연구*, 20(3), 39–59)
- [13] Seoul national university hospital, 천식의 증상, 2023년 02월 24일 접속,
<http://www.snuh.org/health/nMedInfo/nView.do?category=DIS&medid=AA000398>
- [14] World Health Organization (2006), Preventing disease through health environments
- [15] Yara D. S.T., Siqueira, H.V., and Alves, T.A. (2016), Unorganized machines to predict hospital admissions for respiratory diseases, *IEEE Latin American Conference on Computational Intelligence (LA-CCI)*, 1–6
- [16] Zeinab, N. S., Karimi, M., and Alesheikh, A. (2020), Environmental and infrastructural effects on respiratory disease exacerbation: a LBSN and ANN-based spatio-temporal modelling, *Environmental Monitoring and Assessment*, 192, 1–17
- [17] Zeng, A., Chen M., Zhang, L. and Xu, Q. (2022), Are transformers effective for time series forecasting?, arXiv preprint.



Han Joo Lee (jeffrey7108@gmail.com)

Han Joo Lee is a M.S. student of Department of Software in Sejong University. His research interests are big data analysis, intelligent systems, and deep learning.



Min Kyu Jee (mk92@hanmail.net)

Min Kyu Jee is the manager of Selvers Inc. He received his M.S. in Department of Software Convergence in Sejong University. His research interests are text mining, machine learning, and deep learning.



Cheong Won Kim (wikim@sejong.ac.kr)

Cheong Won Kim is a professor of Department of Software in Sejong University. He received his Ph.D. in Syracuse University of New York with Computer and Information Science. His research interests are Deep learning, Bioinformatics, Polymer chemistry and Text mining.

A prediction study on the number of emergency patients with ASTHMA according to the concentration of air pollutants

Han Joo Lee^{*}, Min Kyu Jee^{**}, Cheong Won Kim^{***}

ABSTRACT

Due to the development of industry, interest in air pollutants has increased. Air pollutants have affected various fields such as environmental pollution and global warming. Among them, environmental diseases are one of the fields affected by air pollutants. Air pollutants can affect the human body's skin or respiratory tract due to their small molecular size. As a result, various studies on air pollutants and environmental diseases have been conducted. Asthma, part of an environmental disease, can be life-threatening if symptoms worsen and cause asthma attacks, and in the case of adult asthma, it is difficult to cure once it occurs. Factors that worsen asthma include particulate matter and air pollution. Asthma is an increasing prevalence worldwide. In this paper, we study how air pollutants correlate with the number of emergency room admissions in asthma patients and predict the number of future asthma emergency patients using highly correlated air pollutants. Air pollutants used concentrations of five pollutants: sulfur dioxide(SO_2), carbon monoxide(CO), ozone(O_3), nitrogen dioxide(NO_2), and fine dust(PM_{10}), and environmental diseases used data on the number of hospitalizations of asthma patients in the emergency room. Data on the number of emergency patients of air pollutants and asthma were used for a total of 5 years from January 1, 2013 to December 31, 2017. The model made predictions using two models, Informer and LTSF-Linear, and performance indicators of MAE, MAPE, and RMSE were used to measure the performance of the model. The results were compared by making predictions for both cases including and not including the number of emergency patients. This paper presents air pollutants that improve the model's performance in predicting the number of asthma emergency patients using Informer and LTSF-Linear models.

Keywords: Environmental diseases, Air pollutants, Particulate matter, Time series forecasting, Healthcare

This work is financially supported by Korea Ministry of Environment(MOE) as 「Graduate School specialized in Climate Change」.

* First Author, M.S. Student of Sejong University, Department of Software

** Co-Author, Manager, Selvers Inc.

*** Corresponding Author, Professor of Sejong University, Department of Software