<div align="center">

|  |  |  |
| --- | --- | --- |
| **2018** | **Report – 2018/ 12월 5주** | **Seong-gu** |

</div>

**I. Plan for this week**

***Personal Research Subject***

* 학회 제출용 논문 작성 완료

**II. Plan for next week**

* 논문 수정 완료 후 제출 (영어 워딩 검토 작업)
* Attention / Transformer에 대한 공부

**III. Response on Prof. Shin's feedback last week**

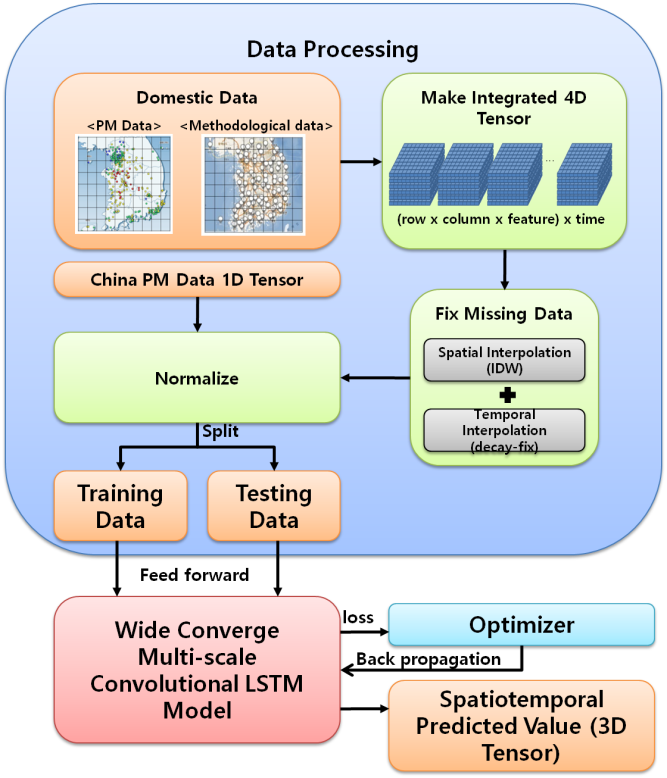
* 현재는 시간에 따른 예측이지만, 누락된 공간적인 데이터도 잘 예측하는 방법 검토.
* Unsupervised learning이 될텐데 방법을 찾아보자.

**IV. Milestone**

1. 실험실 환경에서 미세먼지 농도 예측 시스템 구축   
   1. 머신러닝 기법을 활용한 미세먼지 농도 예측  
        
      **- 한국의 지리학적 특성 및 계절풍의 영향을 고려하여, 활용할 측정소 범위를 국내 전역으로 하고, 중국의 미세먼지 농도를 모델링 데이터로 사용함.**(중국데이터를 확보하는 방법 조사할 것)

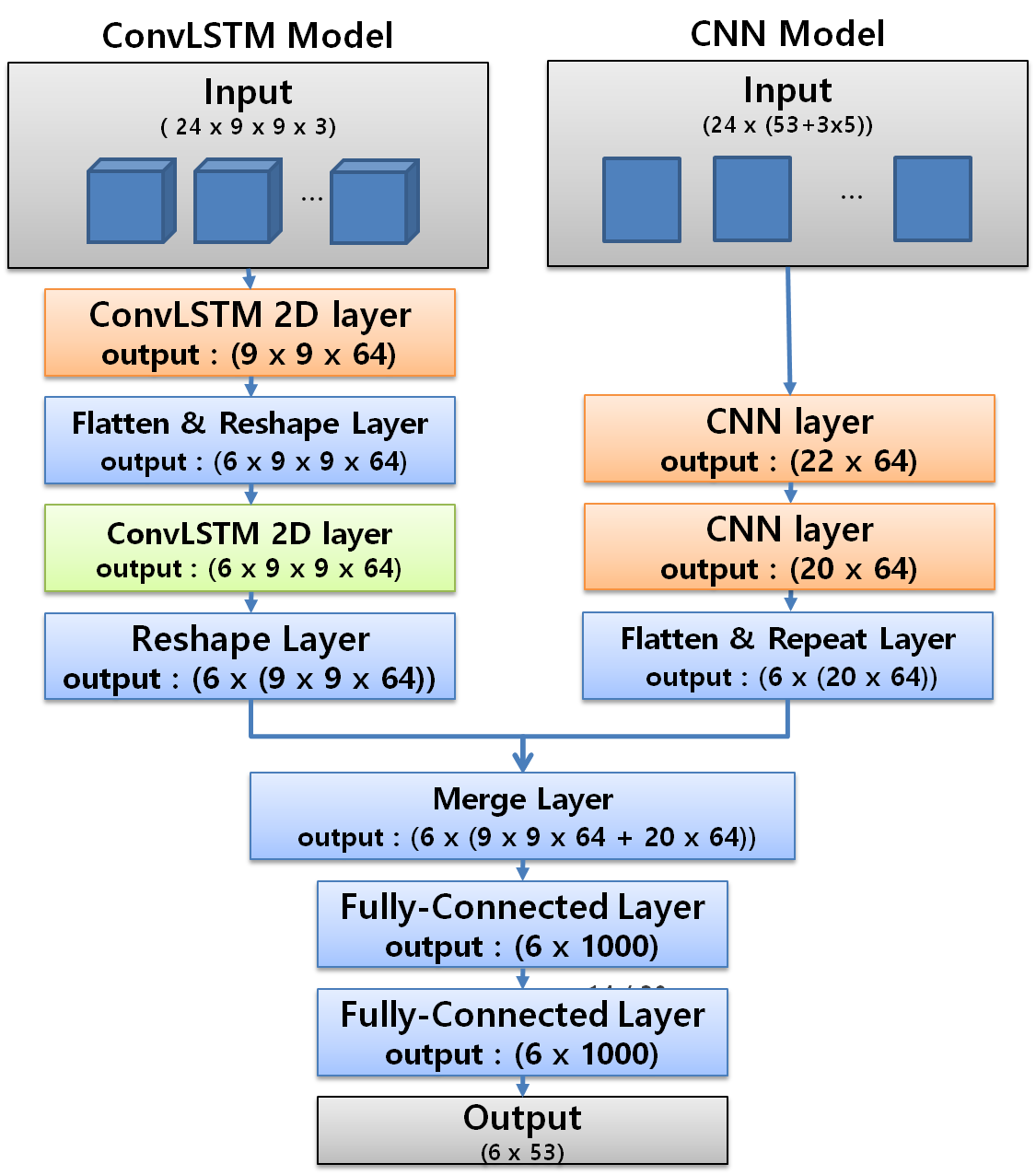
**- 결측치는 공간적 보간법인 IDW 기법과 시간적 보간법인 decay-fix 기법을 결합한 시공간 보간법을 사용하여 추정함.**

**- 모델링은 미세먼지 유입의 공간적인 특성을 적극 반영하기 위해서 최근에 연구되고 있는 Convolutional LSTM 모델을 사용.**(모델링 최적화 필요)

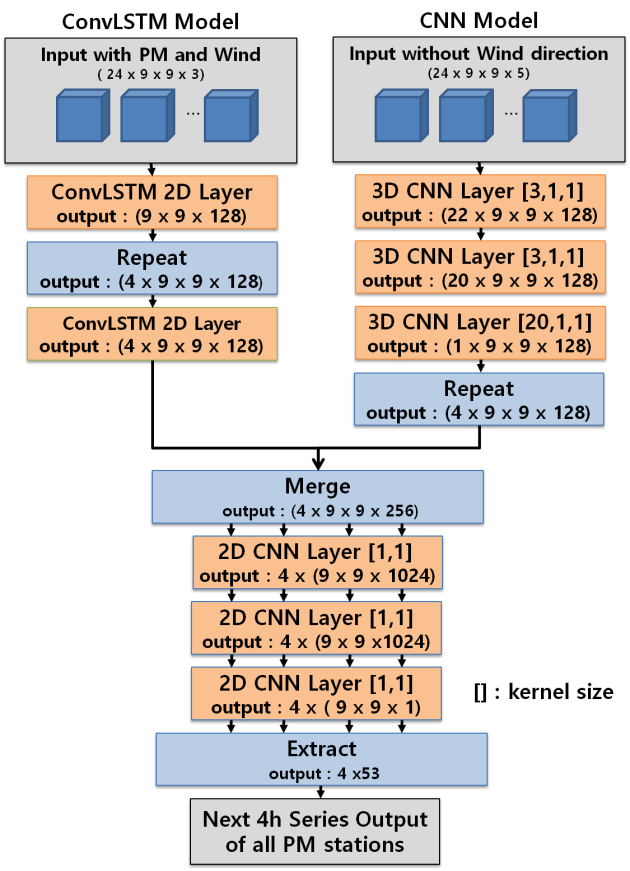


**V. Reports**

* 개인연구 진행사항
* **모델링 관련 – 작성중인 연구 논문 결과 관련 정리**

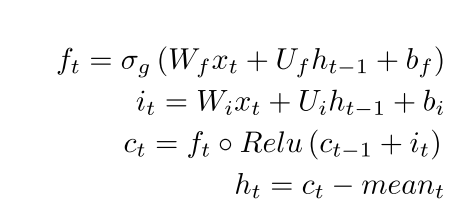


이전 작성 모델 - Merge과정에서 격자(위치) 정보 소실

  
최신 작성 모델 – 격자(위치)정보 유지

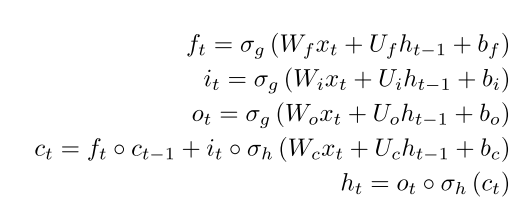
| Model | Prediction hour | | | | Average |
| --- | --- | --- | --- | --- | --- |
| 1h | 2h | 3h | 4h |
| **Hybrid 3x3 (이전 모델)** | **13.41** | **15.44** | **17.50** | **19.29** | **16.41** |
| Hybrid 1x1 | 10.49 | 14.36 | 17.04 | 19.13 | 15.25 |
| **Hybrid 3x3** | **9.76** | **13.36** | **16.38** | **18.86** | **14.59** |
| Hybrid 5x5 | 10.44 | 13.72 | 16.59 | 18.99 | 14.93 |
| ConvLSTM 3x3 | 9.85 | 13.61 | 16.69 | 19.19 | 14.84 |
| CNN-LSTM | 13.91 | 15.41 | 17.33 | 19.26 | 16.48 |
| LSTM | 13.90 | 16.00 | 17.99 | 19.67 | 16.89 |

* 공간정보를 끝까지 유지한 최신 Hybrid 모델에서 1-3시간 뒤 예측 결과는 확연한 차이를 보였습니다만, 4시간 부터는 성능이 비슷해지는 결과를 보여주었습니다.  
  따라서, 좀 더 정확한 예측을 하기 위한 모델링 개선이 추가로 필요합니다.

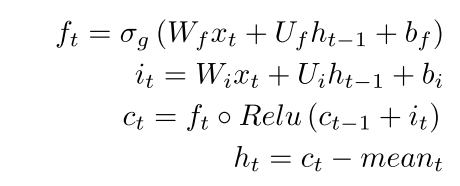
**- 공간정보를 유지하며, 미세먼지를 예측한 관련 논문  
(AIRNET: A MACHINE LEARNING DATASET FOR AIR QUALITY FORECASTING – ICLR accept 실패 논문)**

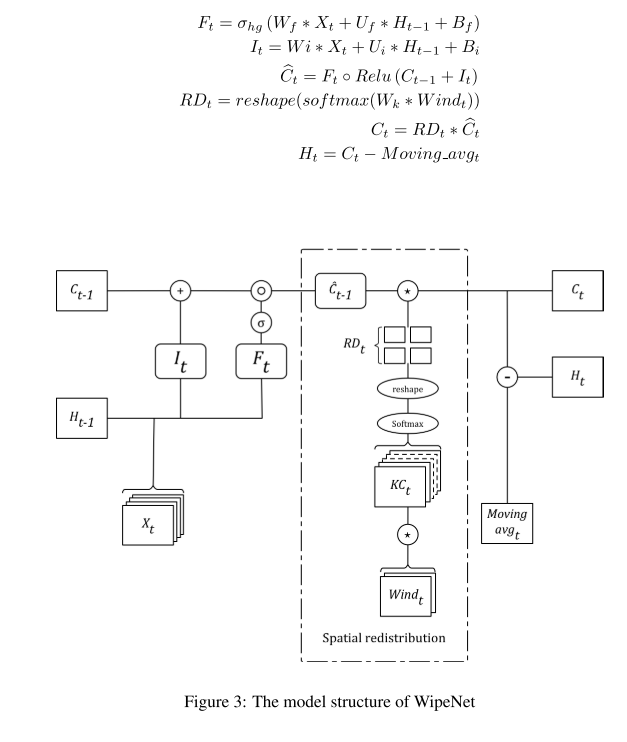
* 이 논문은 기존 논문들과 다르게, 직접 기존 모델링 식을 수정하여, 미세먼지 예측에 맞는 모델링식을 만들었다고 하여 리뷰중입니다.

기존 LSTM 식



Reduced LSTM 식



WipeNet 식  


It : 날씨에 의한 공기 오염 축적 효과,

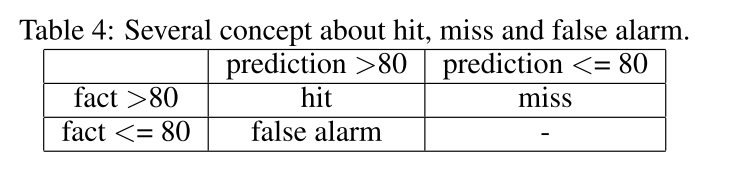
Ct-1 : 이전 공기 오염 축적량,

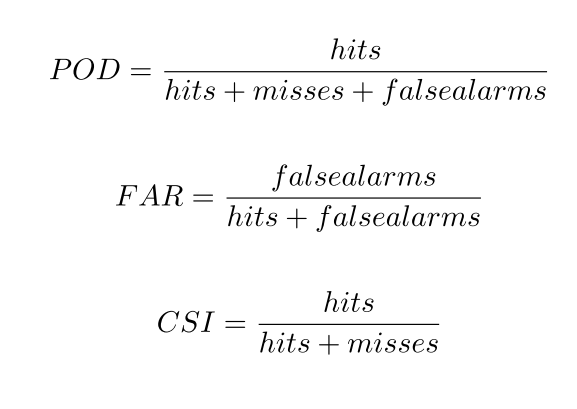
Ft : 날씨에 의한 공기 오염 감소 효과

RD : 바람에 의한 공기오염 transport 효과

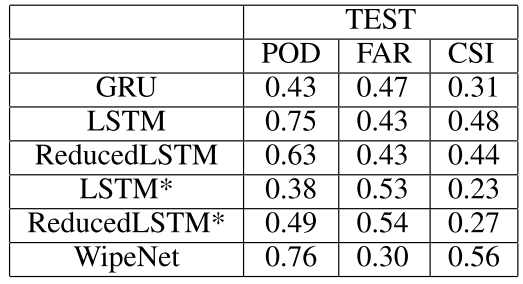
결과

본 논문에서는PM2.5가 80um/m3를 넘는지를 맞추는 것을 기준으로 삼고, hit, false alarm, miss를 정의하고, Probability of Detection (POD), False Alarm Rate(FAR), Critical Success Index (CSI) 지표를 정의했습니다.





매 3시간 단위 측정값에 대해 앞으로 5일간을 예측 시 POD, FAR, CSI 지표 비교.



(WipeNet의 kernel size는 5x5를 사용함)

* 모든 지표에서 WipeNet이 가장 좋은 결과를 보였으나, POD는 LSTM과 비슷한 결과를 보였으나, 기존RMSE, MAE 와 같은 지표로 결과를 보여주지 않기 때문에, 기존 연구 논문과는 비교하기가 어려운 점이 있습니다.  
    
  이 논문에서는 현재 작성되는 논문과 비슷하게, 공간적인 정보와 바람의 영향을 고려하여 모델링을 한 점이 있으며 이 논문에서 제시한 방법이 좋은 방법인지는 검토가 어려우나, 접근 방법에 대해서는 참고할 수 있다고 생각됩니다.