Global Air Pollutant Forecasting Using Sequential Transfer Learning: Addressing the Cold Start Problem

Masters Thesis Presentation

Date 26th May 2025

Presenter Christopher Daniel Ash

Advisor Professor Bubruyr Kim

Department Robotics and Smart Systems

Kyungpook National University

Presentation Overview

1. Introduction
2. Review of Related Works
3. Datasets and Data Preparation
4. Models and Methodology
5. Results and Discussion
6. Conclusion and Future Works

1 Introduction

Air pollution is known to be a major threat to public health

- Millions of people die yearly due to the impacts of air pollution
- Major pollutants include: SO_2, CO, NO_2, NO and Particulate Matter (PM)

Accurate prediction allows for insights into environmental trends

- Foreknowledge allows for better planning
- Also allows for better mitigation measures to be put in place

AI and Deep Learning (DL) models are used extensively for air pollution predictions

- Common models: LSTM, CNN, GRU
- Typically uses large amounts of regional historical data
- Therefore, data limitations (cold-start problem) and generalization are known challenges

1 Introduction – Research Objectives

1 Development of LSTM model trained via sequential transfer learning

- Typically, air pollution studies are carried out using data from one region
- This study uses the learnings from major global regions to predict air pollution globally

2 Ensuring data quality

- DL studies, including air pollution studies, are usually hindered by poor data quality
- Data pre-processing is applied to guarantee high data quality

3 Addressing the cold-start problem

- Data scarcity simulations are carried out to test the model's ability to handle data limitations
- Extensive testing on various datasets across the globe to assess the model's performance

4 Comparative analysis

- Assessing the effectiveness of the sequential transfer learning method
- Testing the applicability of the proposed method across multiple models

2 Review of Related Works - Overview

Deep learning models in recent research:

CNN

- Used for their feature extraction abilities
- Commonly used alongside other models
- References 23, 28, 30, 31 and 50

RNN

- The most popular being the LSTM model
  - References 13, 17, 31 and 37
- GRU models are also popular
  - References 28, 55 and 56

Hybrid models

- Models are commonly combined in order to improve prediction ability
- References 18, 21 and 40

Transfer Learning:

- ➢ Training a model on a dataset from a data-rich source and using the trained model on a data-scarce target
- ➢ Has been minimally applied in air pollution prediction studies (usually within one region) with some level of success
- ➢ Example study Yadav et al [48], which used satellite imagery and combined with ground data from data-rich regions to estimate pollution in data-poor regions

2 Review of Related Works – Challenges faced

1. Data quality issues

- The accuracy of DL model predictions is dependent on the quality of the data

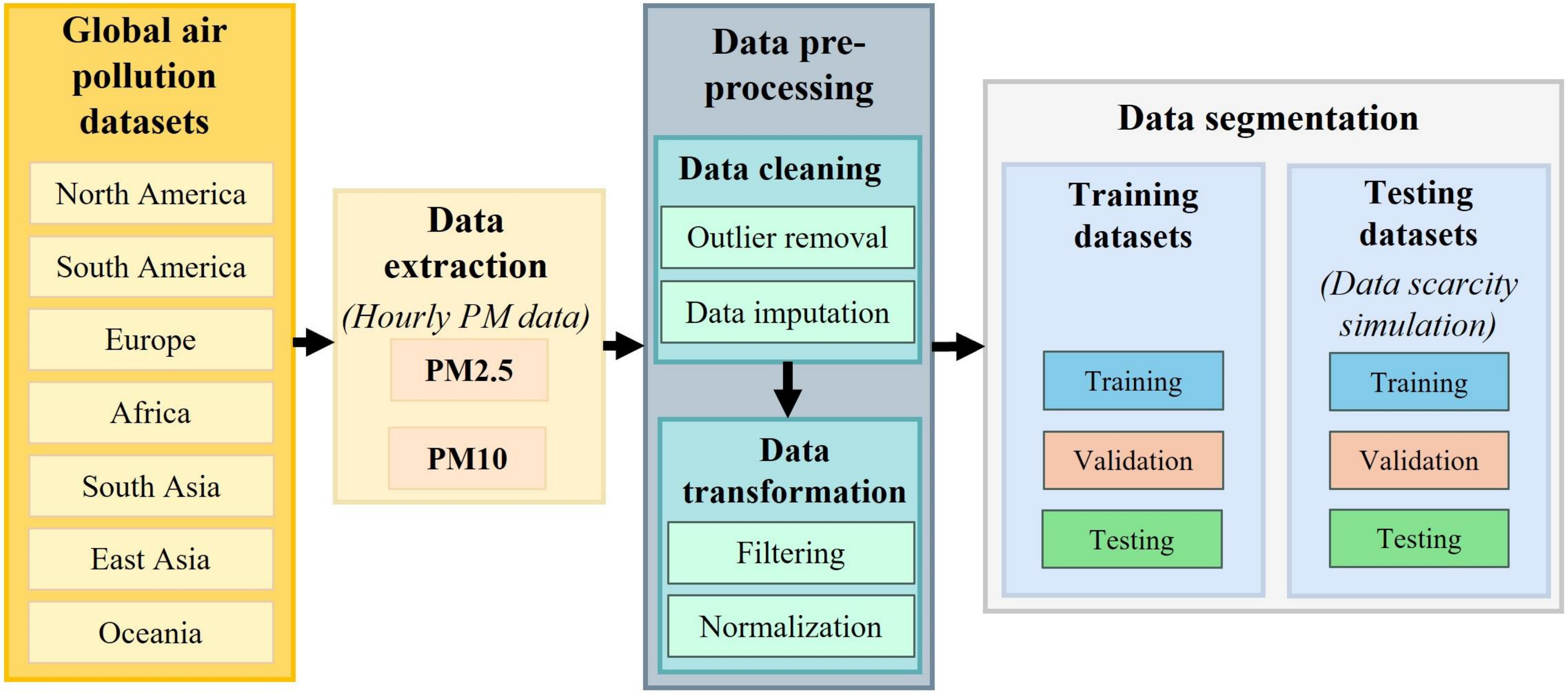
2. Lack of generalizability

- Commonly, datasets from only one region are used in air pollution studies
- This is mainly due to air pollution being affected by local factors, which can vary from region to region
- This leads to models being developed that can only be used in one region

3. Cold start problem

- Arises when there is little to no data available to train DL models
- In air pollution studies, regions with little air pollution data face challenges with the creation of prediction models

Overall Data Preparation Process



3 Datasets and Data Preparation

Datasets

Training datasets

Region	Representative country	Date range	Number of data points
North America	US	01/01/2013 – 31/08/2023	93481
South America	Colombia	01/01/2016 – 31/12/2021	52606
Europe	UK	01/01/2019 – 01/01/2024	43824
Africa	Uganda	21/11/2019 – 31/12/2020	9760
South Asia	India	01/01/2015 – 01/07/2020	48192
East Asia	China	02/01/2017 – 31/12/2020	35040
Oceania	Australia	01/01/2020 – 01/01/2024	35064

Testing datasets

Region	Representative country	Date range	Number of data points
East Asia	South Korea	01/01/2017 – 31/12/2019	26280
North America	Mexico	01/01/2016 – 31/12/2021	47592
Europe	Spain	01/01/2015 – 31/12/2021	61368

Data-preprocessing

Data cleaning

- *Outlier removal*
  - STL decomposition using Loess
- *Data imputation*
  - K-Nearest Neighbors

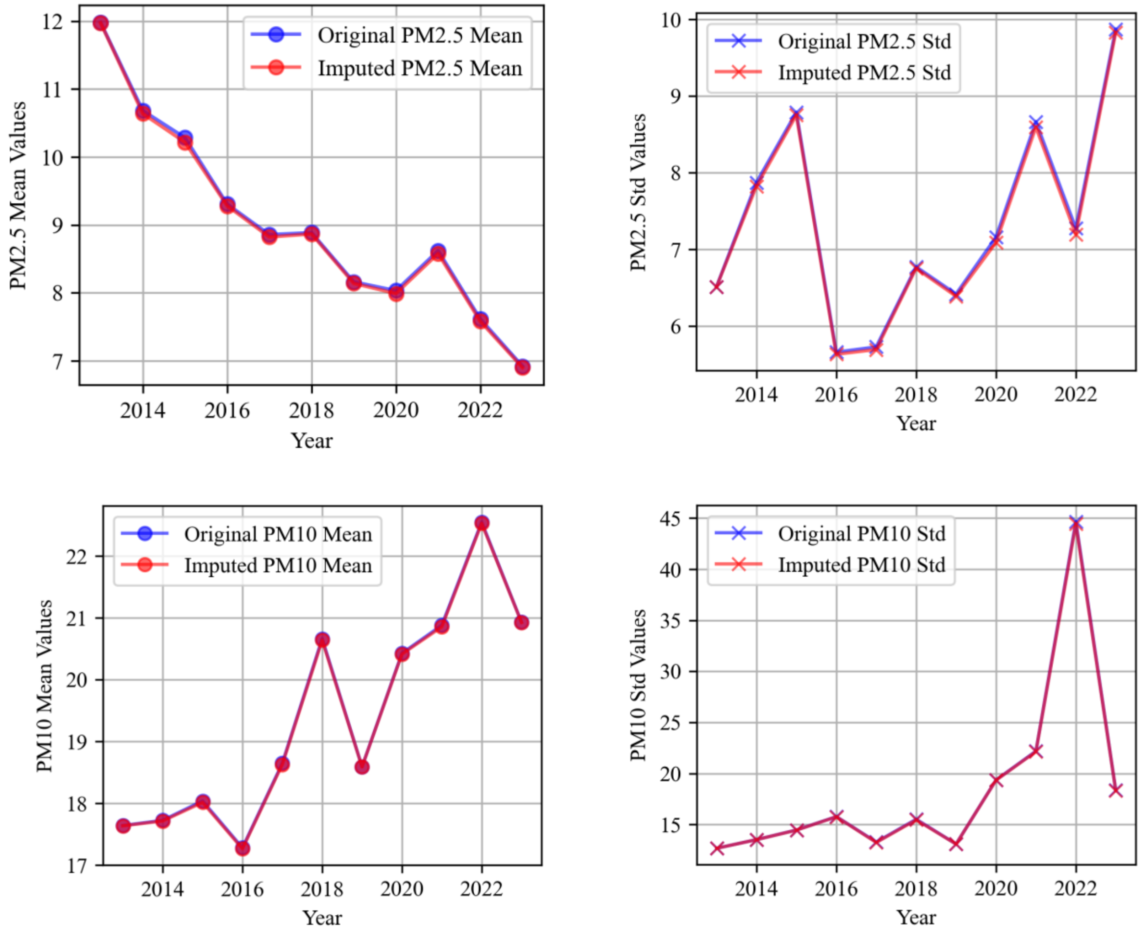
Data transformation

- *Filtering*
  - Butterworth low-pass filter
- *Normalization*

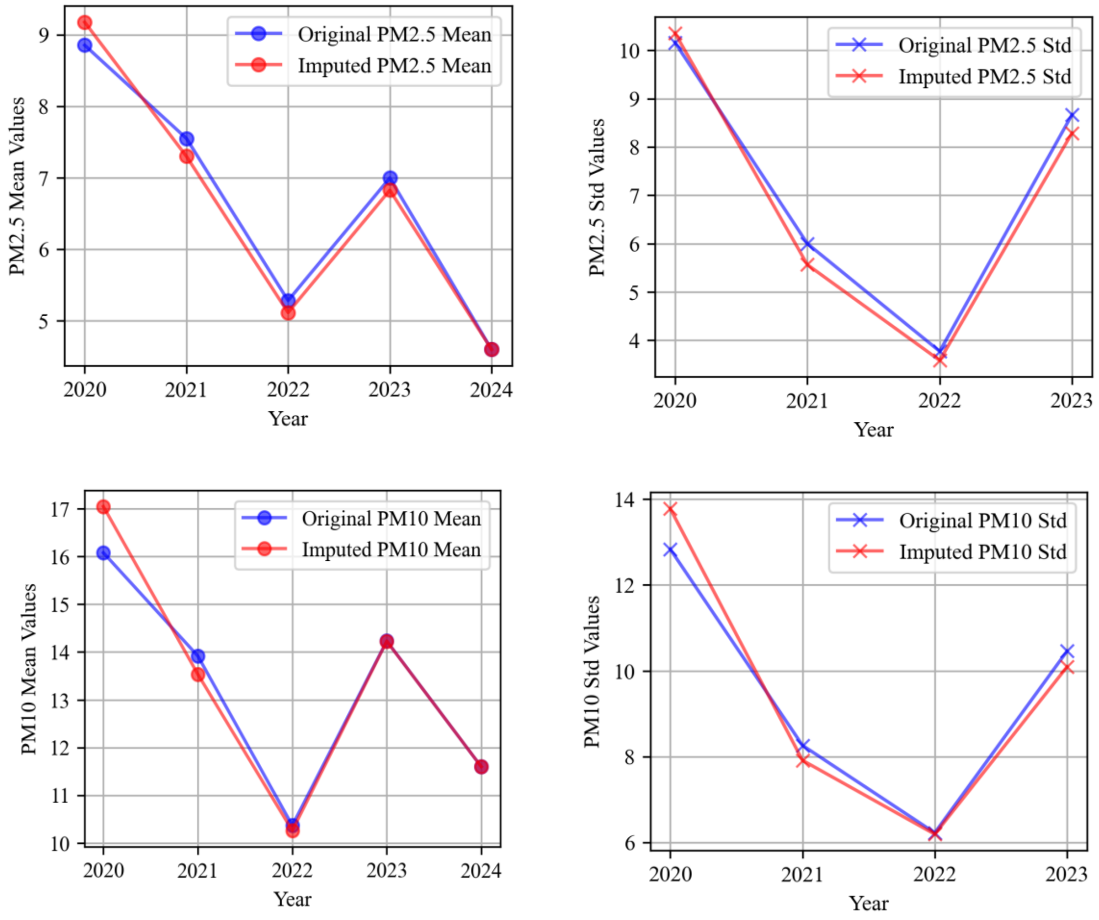
3 Data Preprocessing Results

Dataset characteristics comparison before and after imputation

American dataset



Australian dataset



3 Data Preprocessing Results

Dataset cleaning summary

Country	Change in PM2.5 mean (%)	Change in PM2.5 Std (%)	Change in PM10 mean (%)	Change in PM10 Std (%)
US	0.428	0.578	0.061	0.167
Columbia	0.732	1.579	0.465	2.281
UK	0.056	0.238	0.080	0.321
Uganda	1.196	1.164	1.382	0.972
China	0.449	1.143	0.314	0.799
India	0.460	0.149	0.839	0.513
Australia	0.430	1.747	1.310	2.486
South Korea	0.071	0.571	0.112	0.572
Mexico	2.185	0.982	0.800	2.330
Spain	0.127	1.638	0.081	1.346

3

Data Segmentation

Model Training

Training 60%

Validation 20%

Testing 20%

Model Testing – Data Scarcity Simulations

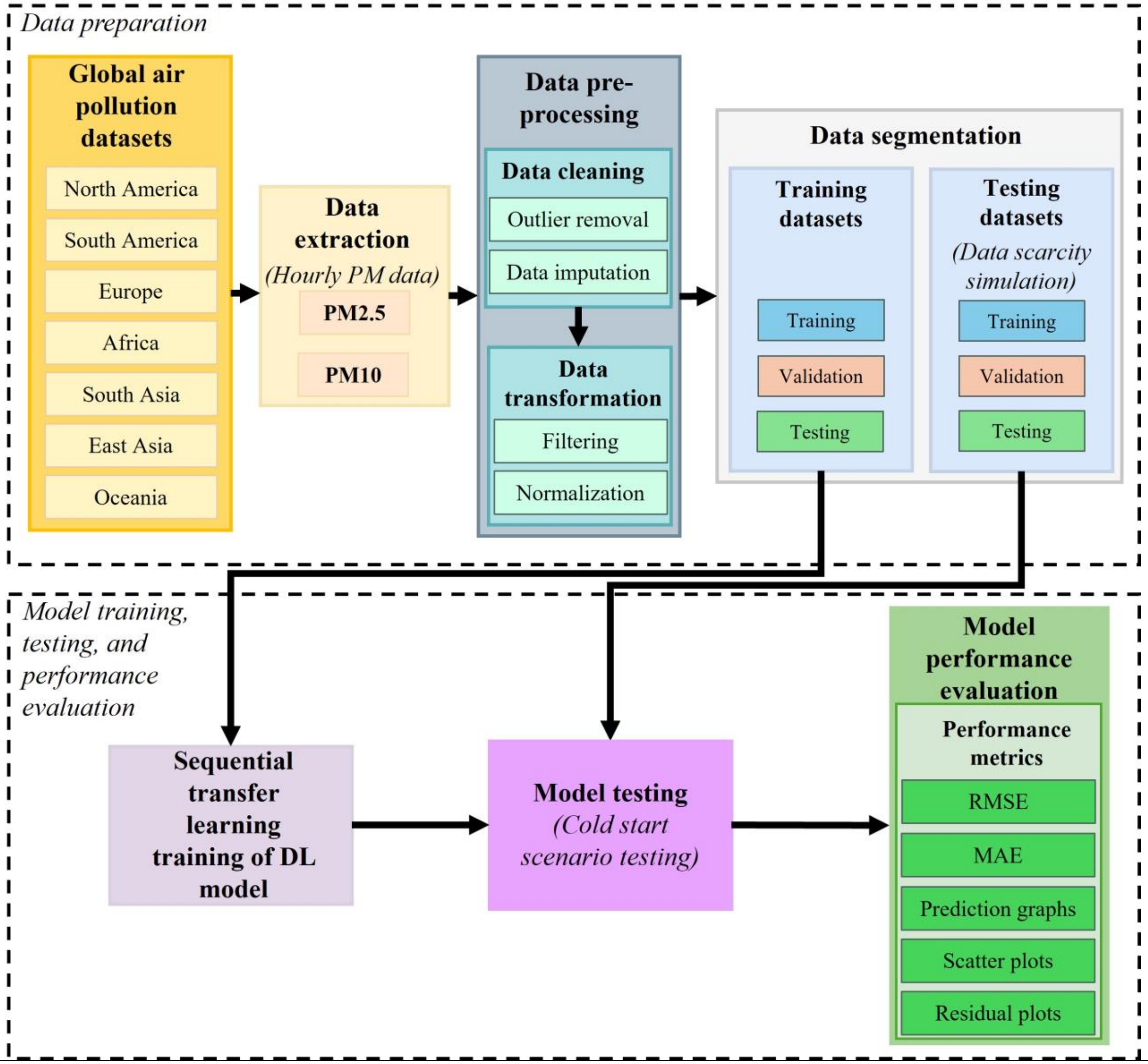
Data scarce conditions:

- Having days to weeks of data
- Datasets were split so that the training segment contained this limited amount of data

Data-scarcity simulation dataset segmentation

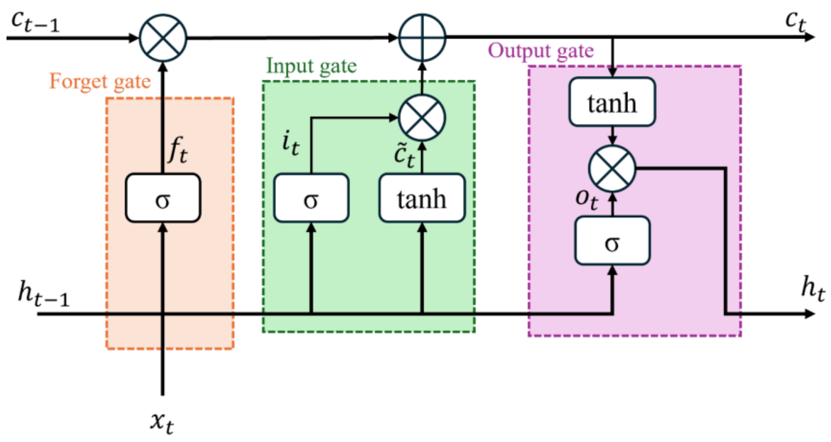
Dataset	Data scarcity period	Percentage of dataset used for training and validation (%)	Percentage of dataset used for Testing (%)	Size of testing dataset
South Korea	0.5 weeks (84 hours)	0.3196	99.6804	26196
	1 week (168 hours)	0.6393	99.3607	26112
	2 weeks (336 hours)	1.2785	98.7215	25944
	1 month/ 30 days (720 hours)	2.7397	97.2603	25560
	2 months (1440 hours)	5.4795	94.5205	24840
	3 months (2160 hours)	8.2192	91.7808	24120
Mexico	0.5 weeks (84 hours)	0.1765	99.8235	47508
	1 week (168 hours)	0.3530	99.647	47424
	2 weeks (336 hours)	0.7060	99.294	47256
	1 month/ 30 days (720 hours)	1.5129	98.4871	46872
	2 months (1440 hours)	3.0257	96.9743	46152
	3 months (2160 hours)	4.5386	95.4614	45432
Spain	0.5 weeks (84 hours)	0.1369	99.8631	61284
	1 week (168 hours)	0.2738	99.7262	61200
	2 weeks (336 hours)	0.5475	99.4525	61032
	1 month/ 30 days (720 hours)	1.1733	98.8267	60648
	2 months (1440 hours)	2.3465	97.6535	59928
	3 months (2160 hours)	3.5197	96.4803	59208

Overall Methodology



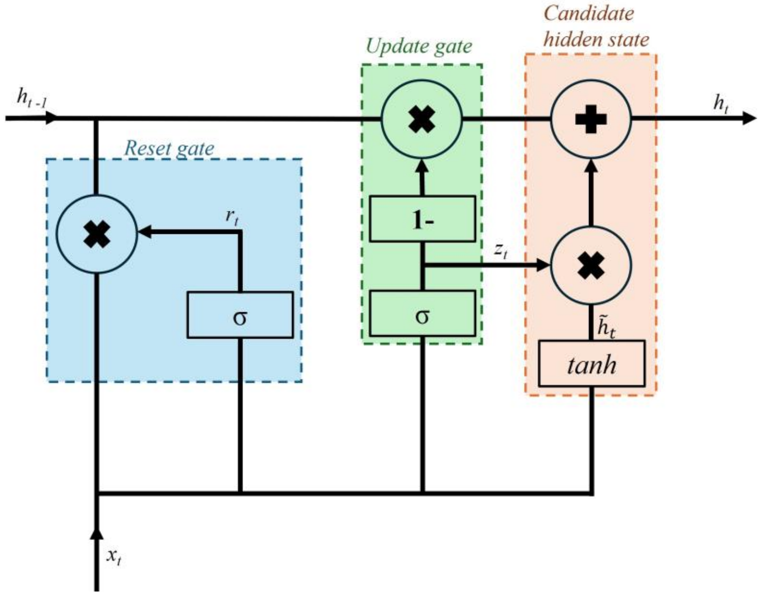
4 Models and Methodology – DL models

LSTM model

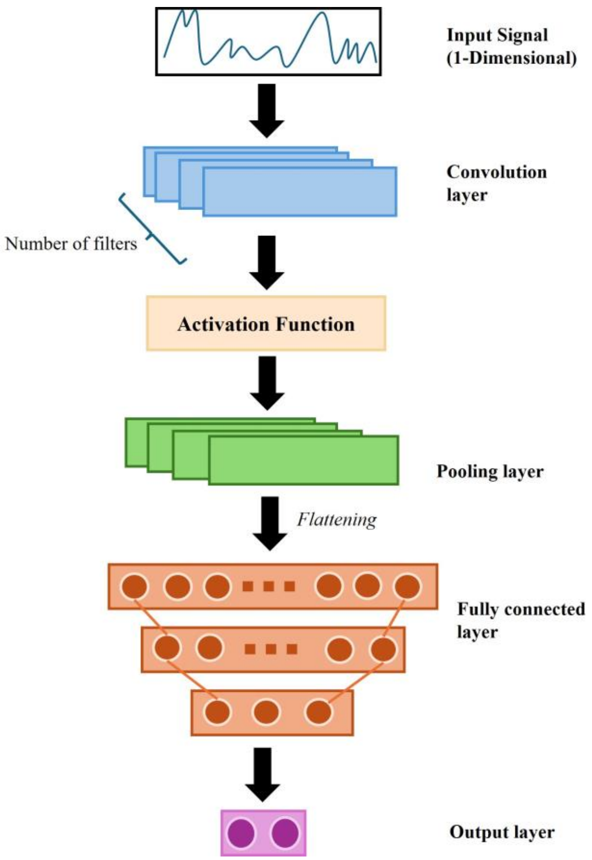


- The LSTM model is the focus of the study
- The 1D-CNN and GRU models are used for comparative analysis

GRU model



1D-CNN



4 Models and Methodology – Model Parameters

Random search hyperparameter tuning results

LSTM model

Test	Hyperparameter	Hyperparameter range	Optimal value	Best validation loss	Mean validation loss
a	Hidden size	50 – 200	119	0.00038032	0.002709
	Number of layers	2 – 4	2		
	Weight decay	0.0001 – 0.01	0.00003		
	Dropout	0.1 – 0.5	0.19		
	Learning rate	0.00001 – 0.001	0.00797		
b	Hidden size	50 – 120	94	6.80547e-06	0.000023
	Number of layers	1 – 2	1		
	Weight decay	0	0		
	Dropout	0	0		
	Learning rate	0.00001 – 0.001	0.00161		

4 Models and Methodology – Model Parameters

Random search hyperparameter tuning results

GRU model

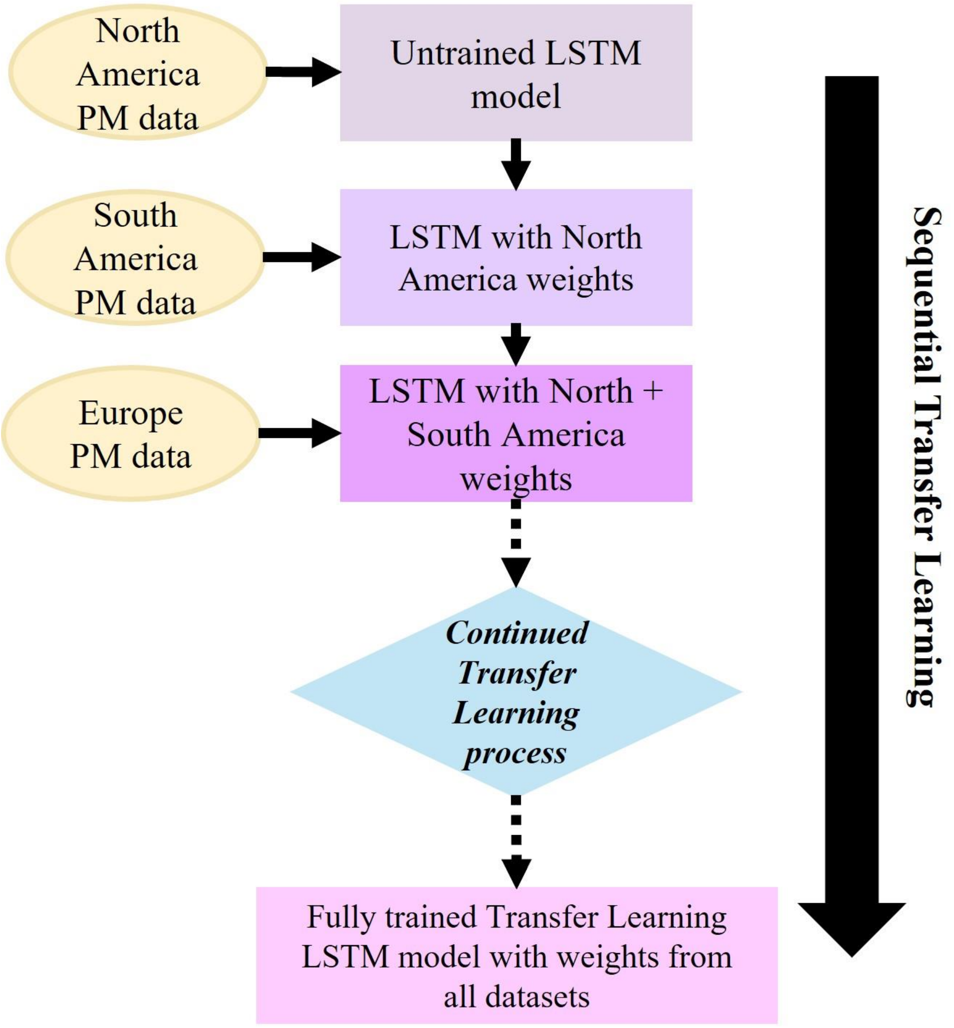
Test	Hyperparameter	Hyperparameter range	Optimal value	Best validation loss	Mean validation loss
a	Hidden size	50 – 200	125	2.06077 e-05	0.001386
	Number of layers	2 – 4	2		
	Weight decay	0.0001 – 0.01	0.00065		
	Dropout	0.1 – 0.5	0.18		
	Learning rate	0.00001 – 0.001	0.00036		
b	Hidden size	50 – 120	85	9.79405 e-06	0.000187
	Number of layers	1 – 2	1		
	Weight decay	0	0		
	Dropout	0	0		
	Learning rate	0.00001 – 0.001	0.00187		

1D-CNN model

Test	Hyperparameter	Hyperparameter range	Optimal value	Best validation loss	Mean validation loss
a	Kernel size	3 – 6	4	0.000381	0.001089
	Number of filters	32 – 128	68		
	Weight decay	0.0001 – 0.01	0.00001		
	Dropout	0.0 – 0.5	0.35		
	Learning rate	0.00001 – 0.001	0.00262		
b	Kernel size	1 – 3	1	6.86935 e-05	0.000131
	Number of filters	8 – 32	22		
	Weight decay	0	0		
	Dropout	0	0		
	Learning rate	0.00001 – 0.001	0.00765		

4 Sequential Transfer Learning Method

Sequential Transfer Learning Process



Order of addition of datasets

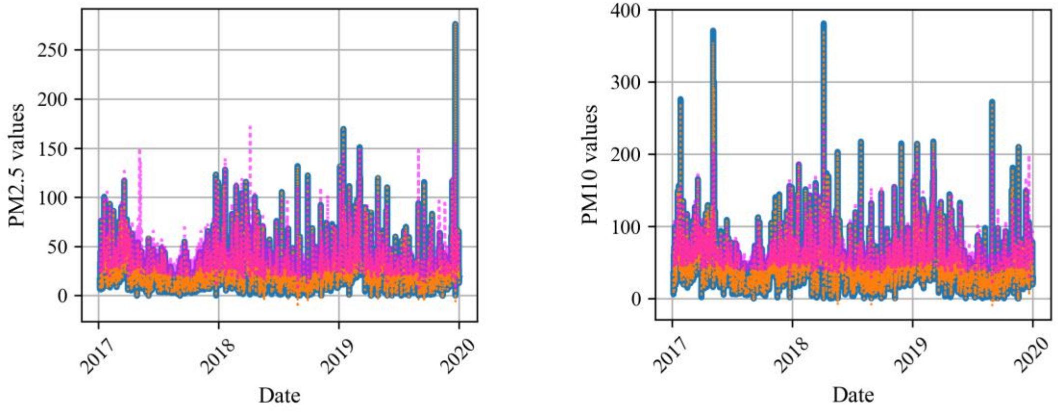
Dataset	Order of addition
North America	1st
South America	2nd
Europe	3rd
Africa	4th
South Asia	5th
East Asia	6th
Oceania	7th

5 Results and Discussion – South Korean dataset

Fully trained sequential transfer learning LSTM model performance: South Korea dataset

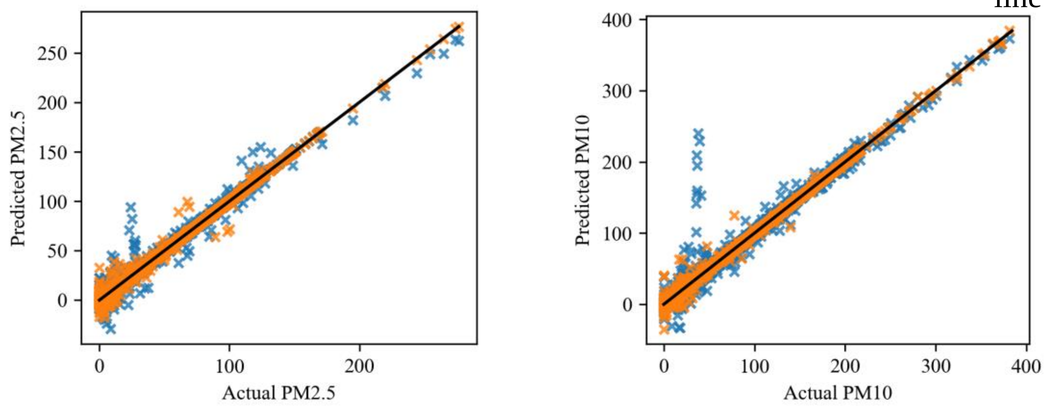
Prediction graphs at 0.5 weeks data scarcity period

Actual values | TL model predicted values | Base model predicted values



Scatter plots at 3 months data scarcity period

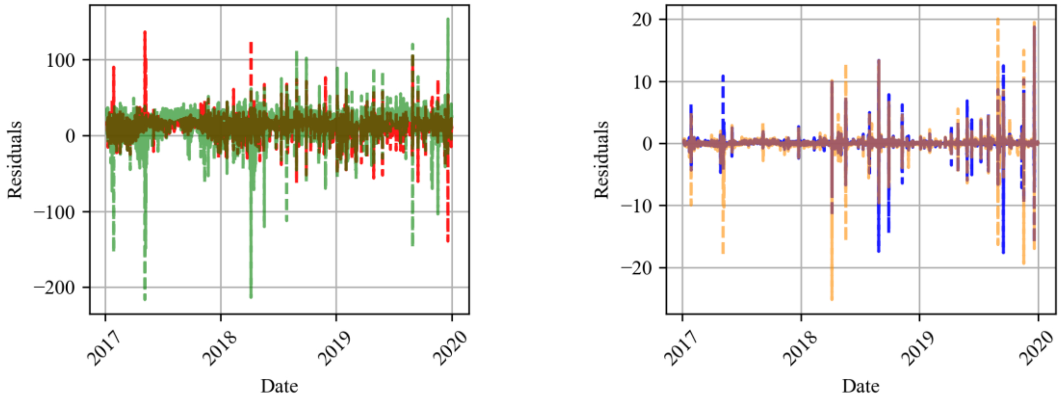
X Base model predicted values | X TL model predicted values | Ideal prediction line



Residual plots at 3 months data scarcity period

Base model PM2.5 residuals | TL model PM2.5 residuals

Base model PM10 residuals | TL model PM10 residuals

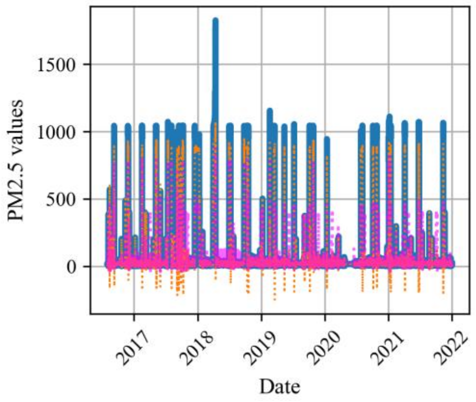


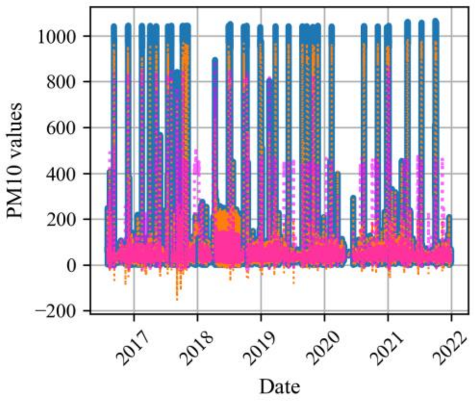
5 Results and Discussion – Mexican dataset

Fully trained sequential transfer learning LSTM model performance: Mexican dataset

Prediction graphs at 0.5 weeks data scarcity period

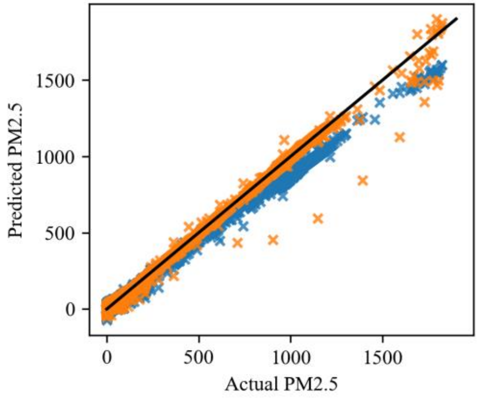
Actual values — TL model predicted values — Base model predicted values

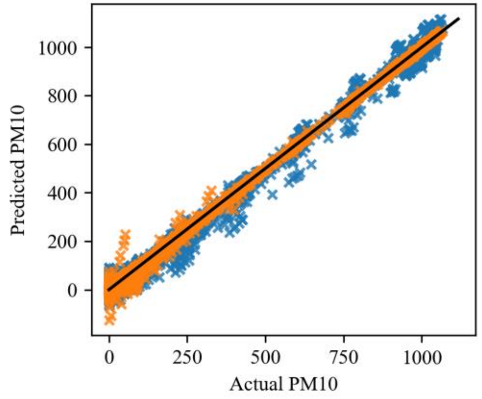




Scatter plots at 3 months data scarcity period

X Base model predicted values — X TL model predicted values — Ideal prediction line

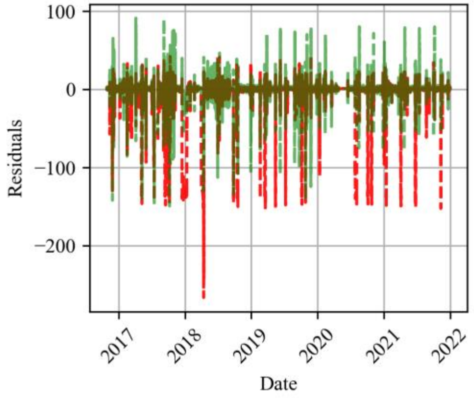


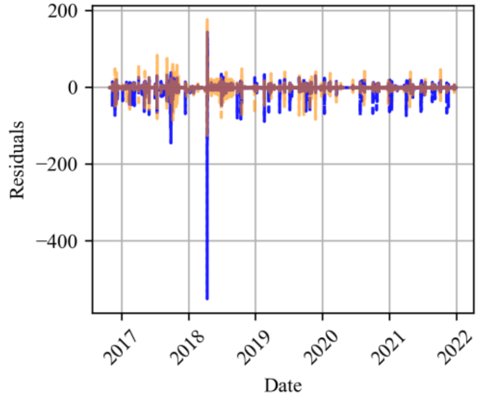


Residual plots at 3 months data scarcity period

Base model PM2.5 residuals — TL model PM2.5 residuals

Base model PM10 residuals — TL model PM10 residuals



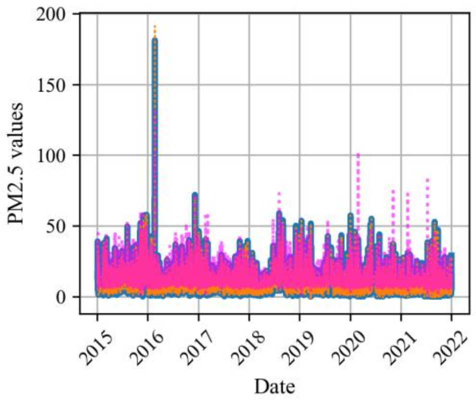


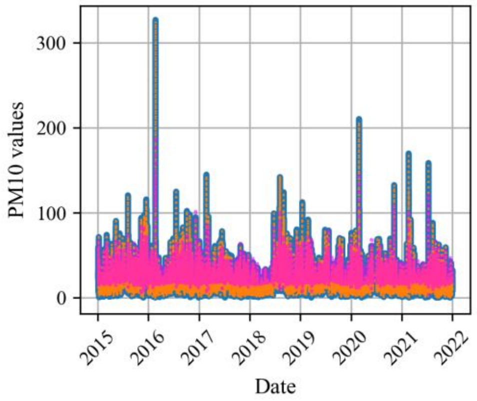
5 Results and Discussion – Spanish dataset

Fully trained sequential transfer learning LSTM model performance: Spanish dataset

Prediction graphs at 0.5 weeks data scarcity period

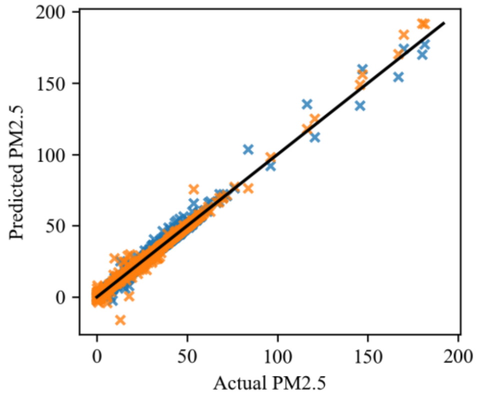
Actual values — TL model predicted values — Base model predicted values

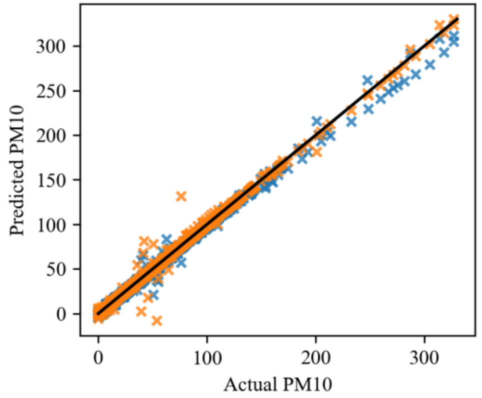




Scatter plots at 3 months data scarcity period

X Base model predicted values — X TL model predicted values — Ideal prediction line

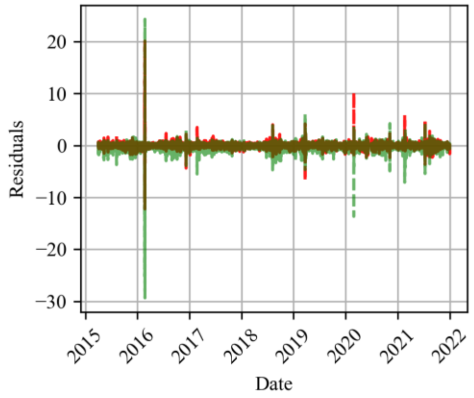


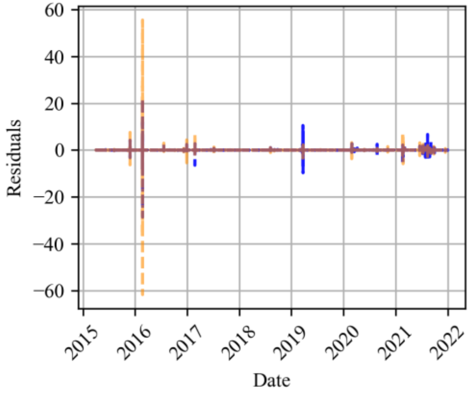


Residual plots at 3 months data scarcity period

Base model PM2.5 residuals — TL model PM2.5 residuals

Base model PM10 residuals — TL model PM10 residuals



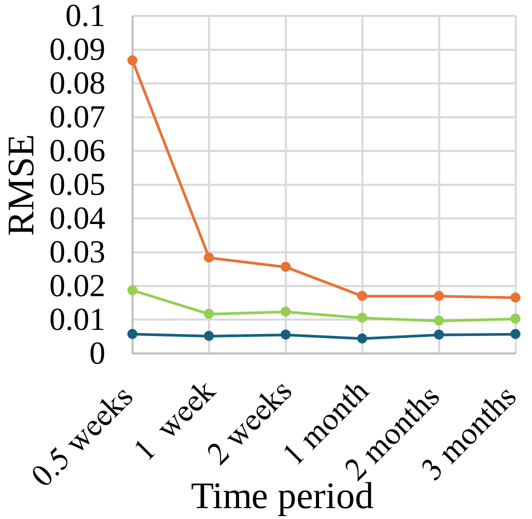


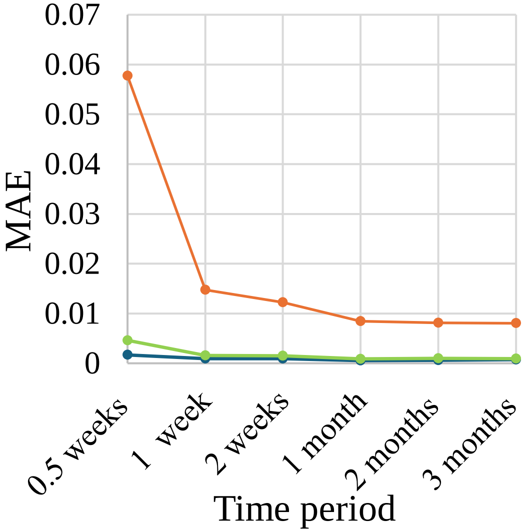
5 Results and Discussion – Comparative analysis

Error performance vs Time period

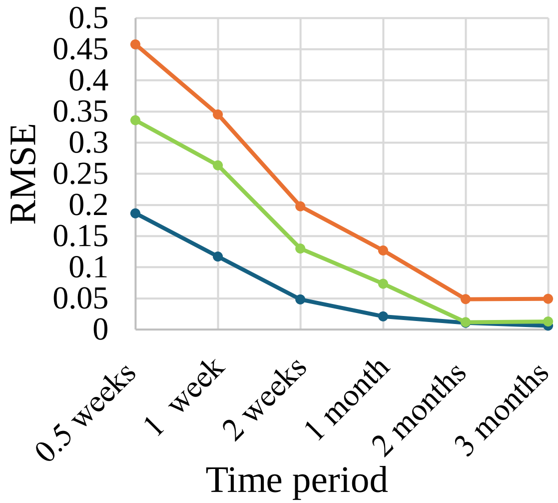
Korean dataset

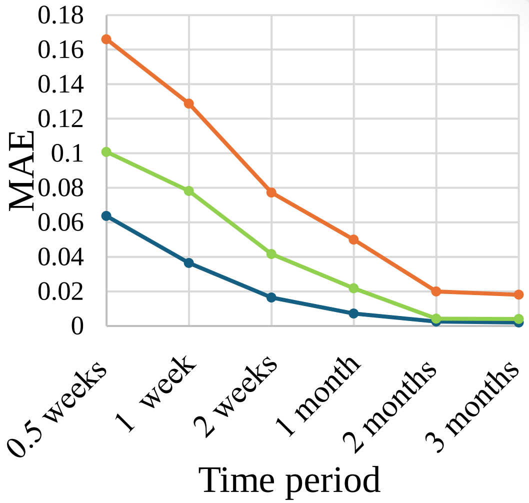
1D-CNN model GRU model LSTM model



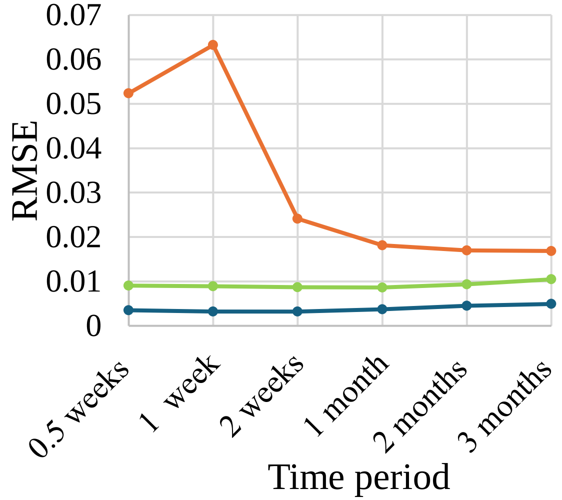


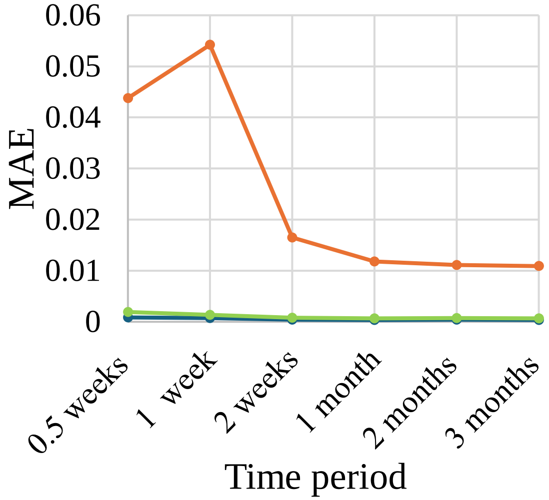
Mexican dataset





Spanish dataset

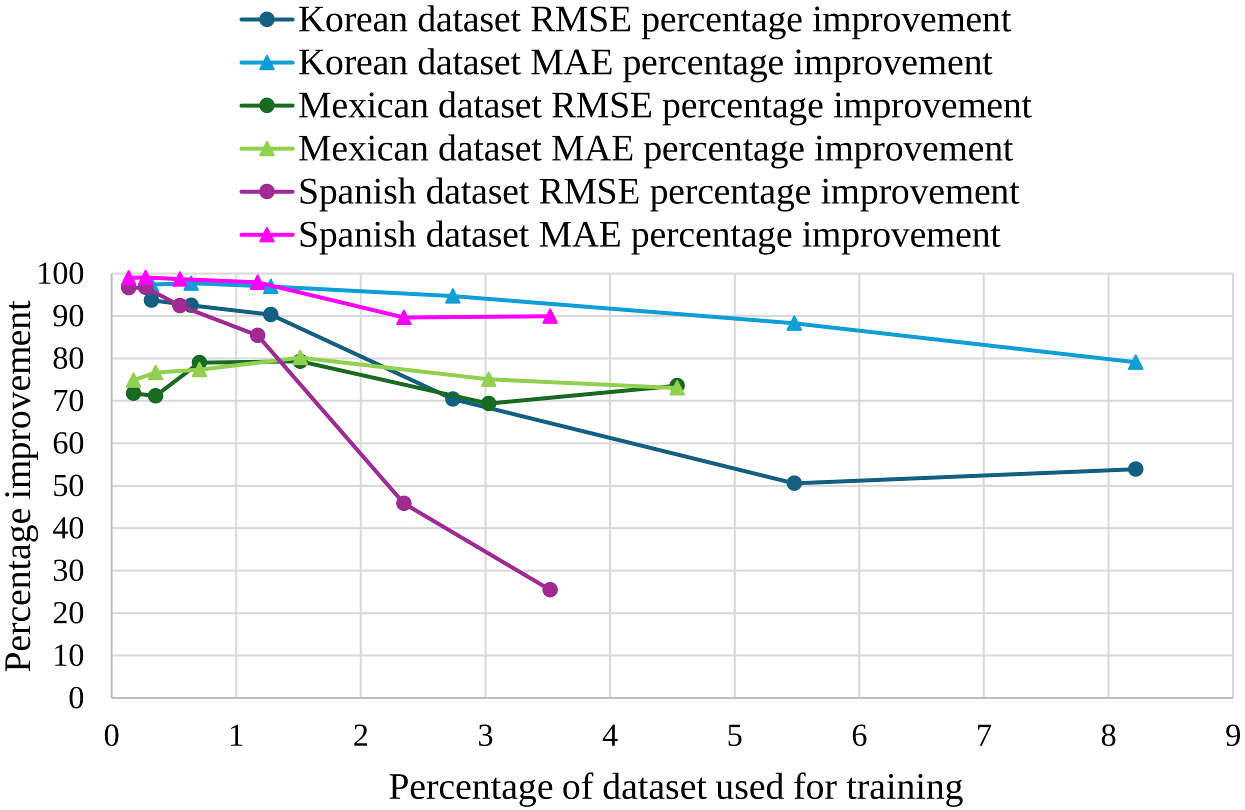




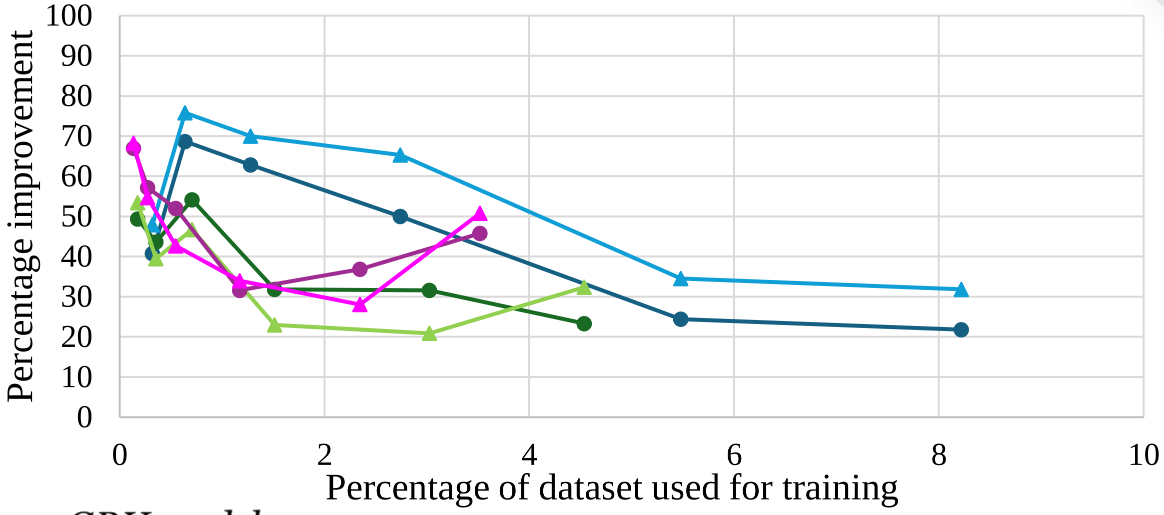
5 Results and Discussion – Comparative analysis

Percentage Improvement vs Percentage of dataset used for testing

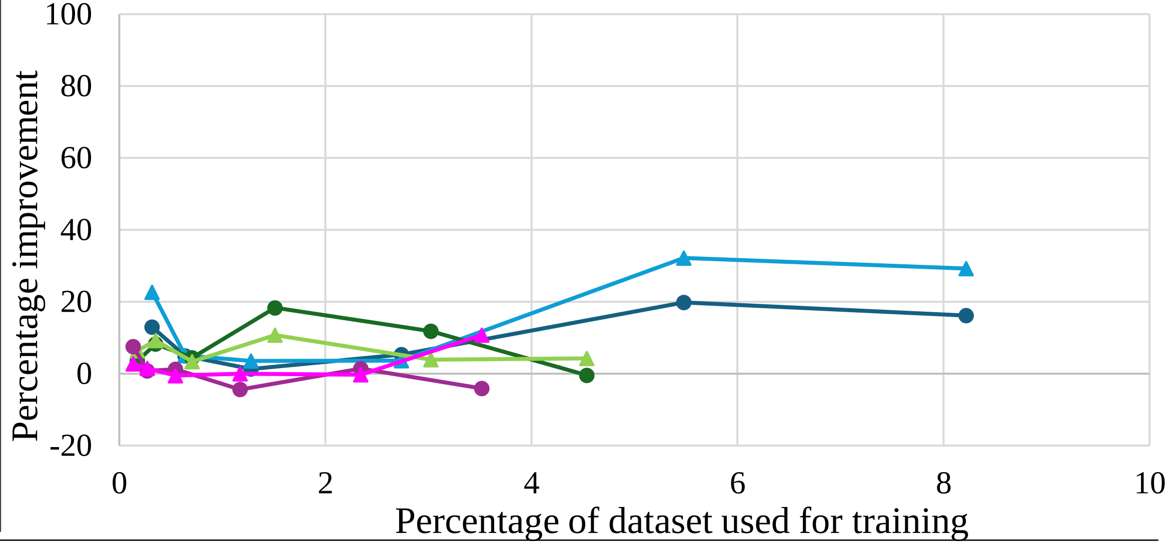
LSTM model



1D-CNN model

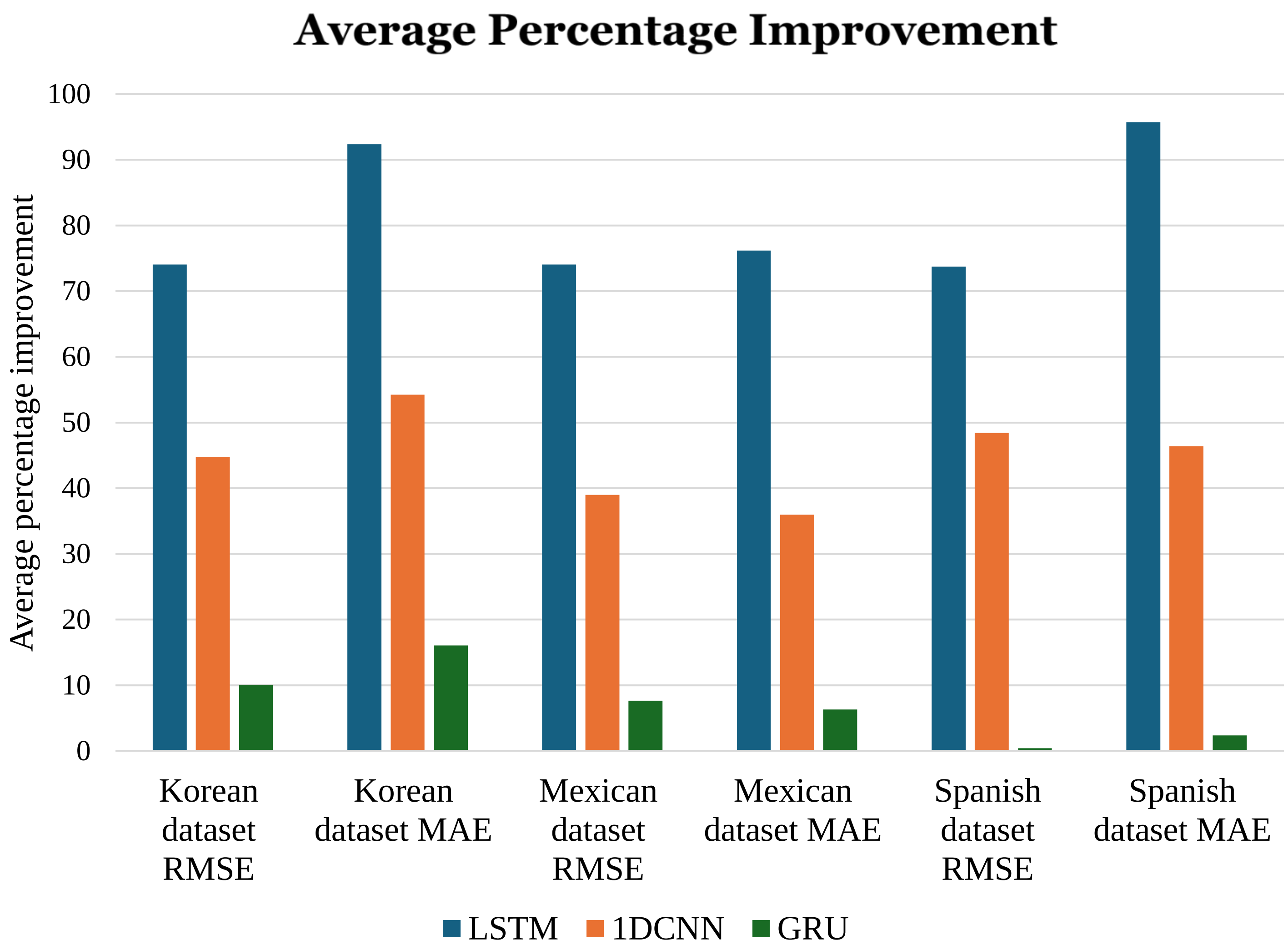


GRU model



5

Results and Discussion – Comparative analysis



5 Comparative analysis – Partial TL models' performance

Percentage Improvement vs Percentage of dataset used for testing

Korean Dataset

Legend: 0.15 | 0.12 | 0.09 | 0.06 | 0.03 | 0

Model	0	1	2	3	4	5	6	7
GRU RMSE at 3 months	0.0123	0.0063	0.0071	0.0048	0.0052	0.0050	0.0077	0.0103
GRU RMSE at 0.5 weeks	0.0215	0.0215	0.0235	0.0238	0.0525	0.0196	0.0177	0.0187
1DCNN RMSE at 3 months	0.0211	0.0186	0.0215	0.0219	0.0214	0.0188	0.0193	0.0165
1DCNN RMSE at 0.5 weeks	0.1464	0.100	0.0984	0.0968	0.113	0.106	0.0948	0.0867
LSTM RMSE at 3 months	0.0127	0.0176	0.0109	0.0036	0.0071	0.0065	0.0079	0.0057
LSTM RMSE at 0.5 weeks	0.091	0.0081	0.0077	0.0088	0.0314	0.0099	0.0054	0.0057

Legend: 0.11 | 0.09 | 0.07 | 0.04 | 0.02 | 0

Model	0	1	2	3	4	5	6	7
GRU MAE at 3 months	0.0013	0.0009	0.0009	0.0007	0.0007	0.0007	0.0008	0.0009
GRU MAE at 0.5 weeks	0.0046	0.0054	0.0024	0.0055	0.0177	0.0043	0.0022	0.0060
1DCNN MAE at 3 months	0.0118	0.0096	0.0126	0.0101	0.0105	0.0100	0.0103	0.0080
1DCNN MAE at 0.5 weeks	0.111	0.0704	0.0698	0.0660	0.0847	0.0790	0.0649	0.0577
LSTM MAE at 3 months	0.0034	0.0007	0.0008	0.0006	0.0007	0.0008	0.0007	0.0007
LSTM MAE at 0.5 weeks	0.0642	0.0039	0.0025	0.0036	0.0150	0.0033	0.0019	0.0017

0 – Base Model| **1** – Model with North America Weights| **2** – North + South America Weights| **3** – North and South America + Europe Weights| **4** – North and South America, Europe + Africa Weights| **5** - North and South America, Europe, Africa + South Asia Weights| **6** - North and South America, Europe, Africa, South + East Asia Weights | **7** – Fully Trained Transfer Learning Model/ All Weights

5 Comparative analysis – Partial TL models' performance

Percentage Improvement vs Percentage of dataset used for testing

Mexican Dataset

Legend: 0.9 | 0.72 | 0.54 | 0.37 | 0.19 | 0.01

Model	0	1	2	3	4	5	6	7
GRU RMSE at 3 months	0.0126	0.0081	0.0057	0.0059	0.0065	0.0088	0.0118	0.0127
GRU RMSE at 0.5 weeks	0.348	0.446	0.299	0.383	0.288	0.254	0.199	0.336
1DCNN RMSE at 3 months	0.0267	0.0239	0.0194	0.0174	0.0270	0.0282	0.0193	0.0181
1DCNN RMSE at 0.5 weeks	0.905	0.234	0.393	0.403	0.340	0.395	0.364	0.458
LSTM RMSE at 3 months	0.0226	0.0058	0.0039	0.0049	0.0073	0.0053	0.0834	0.006
LSTM RMSE at 0.5 weeks	0.662	0.277	0.207	0.298	0.234	0.188	0.195	0.1867

Legend: 0.36 | 0.29 | 0.21 | 0.14 | 0.07 | 0

Model	0	1	2	3	4	5	6	7
GRU MAE at 3 months	0.0042	0.0031	0.0025	0.0023	0.0026	0.0025	0.0028	0.0040
GRU MAE at 0.5 weeks	0.108	0.134	0.0975	0.128	0.0928	0.0825	0.0620	0.101
1DCNN MAE at 3 months	0.0267	0.0239	0.0194	0.0174	0.0270	0.0282	0.0193	0.0181
1DCNN MAE at 0.5 weeks	0.356	0.0901	0.142	0.138	0.119	0.150	0.138	0.166
LSTM MAE at 3 months	0.0076	0.0025	0.0022	0.0025	0.003	0.0025	0.0171	0.002
LSTM MAE at 0.5 weeks	0.253	0.0874	0.0657	0.0949	0.0769	0.0588	0.063	0.0637

0 – Base Model| **1** – Model with North America Weights| **2** – North + South America Weights| **3** – North and South America + Europe Weights| **4** – North and South America, Europe + Africa Weights| **5** - North and South America, Europe, Africa + South Asia Weights| **6** - North and South America, Europe, Africa, South + East Asia Weights | **7** – Fully Trained Transfer Learning Model/ All Weights

5 Comparative analysis – Partial TL models' performance

Percentage Improvement vs Percentage of dataset used for testing

Spanish Dataset

Legend: 0.14 | 0.11 | 0.08 | 0.06 | 0.03 | 0

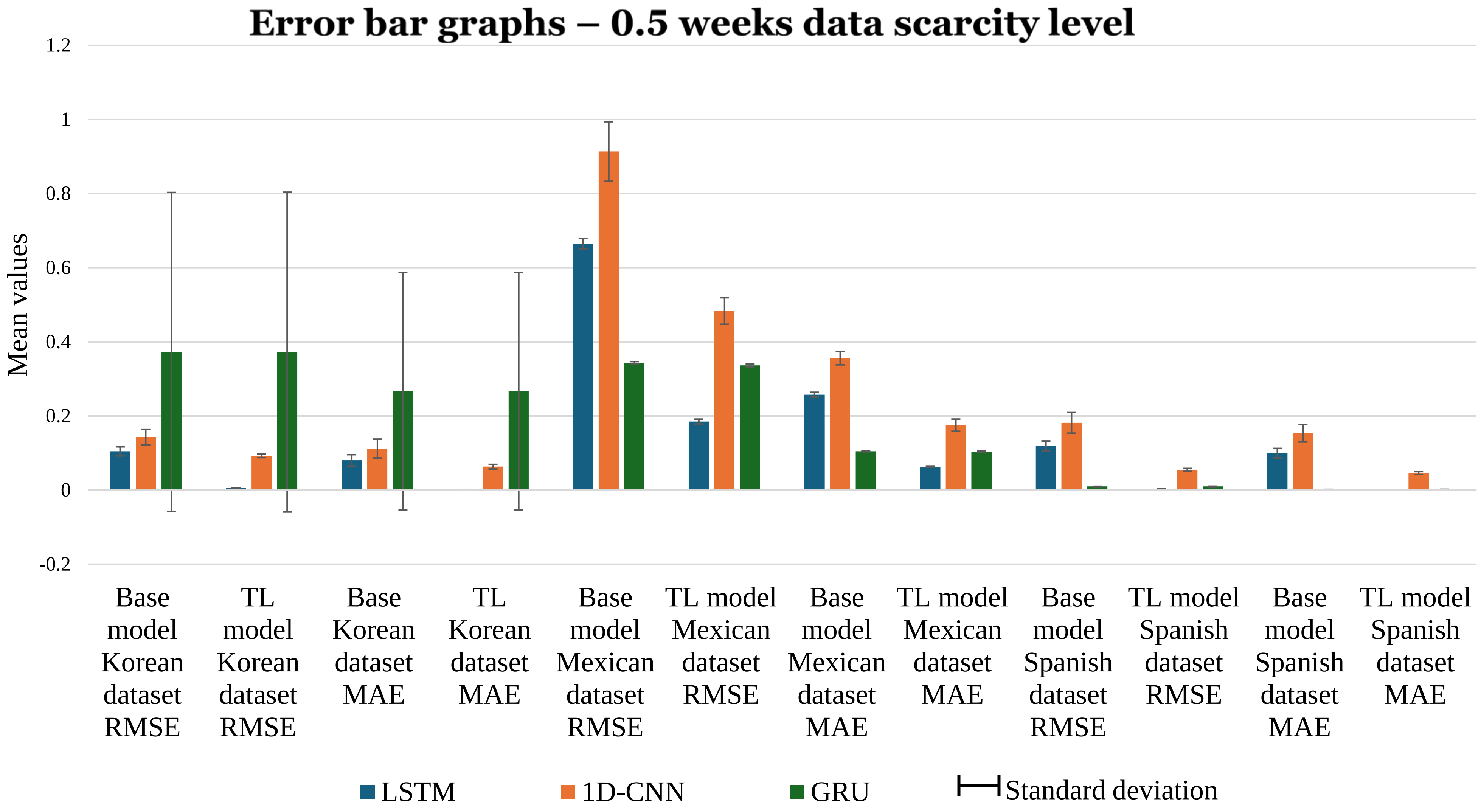
Model	0	1	2	3	4	5	6	7
GRU RMSE at 3 months	0.0101	0.0277	0.0131	0.0124	0.0135	0.0140	0.0138	0.0105
GRU RMSE at 0.5 weeks	0.0091	0.0186	0.0231	0.0217	0.0141	0.0143	0.0129	0.0098
1DCNN RMSE at 3 months	0.0222	0.0175	0.0174	0.0162	0.0165	0.0167	0.0179	0.0169
1DCNN RMSE at 0.5 weeks	0.138	0.0641	0.0662	0.0605	0.0683	0.0639	0.0721	0.0524
LSTM RMSE at 3 months	0.0066	0.0365	0.0062	0.0042	0.0052	0.0047	0.0035	0.0049
LSTM RMSE at 0.5 weeks	0.103	0.0104	0.0050	0.0067	0.0058	0.0047	0.0030	0.0035

Legend: 0.16 | 0.13 | 0.1 | 0.06 | 0.03 | 0

Model	0	1	2	3	4	5	6	7
GRU MAE at 3 months	0.0008	0.0011	0.0007	0.0007	0.0007	0.0005	0.0006	0.0007
GRU MAE at 0.5 weeks	0.0020	0.0030	0.0010	0.0020	0.0019	0.0009	0.0008	0.0019
1DCNN MAE at 3 months	0.0310	0.0116	0.0119	0.0106	0.0108	0.0107	0.0125	0.0109
1DCNN MAE at 0.5 weeks	0.159	0.0559	0.0551	0.049	0.0572	0.0502	0.0641	0.0438
LSTM MAE at 3 months	0.0035	0.0014	0.0005	0.0006	0.0004	0.0004	0.0003	0.0003
LSTM MAE at 0.5 weeks	0.0842	0.0028	0.0009	0.0018	0.0017	0.0007	0.0006	0.0009

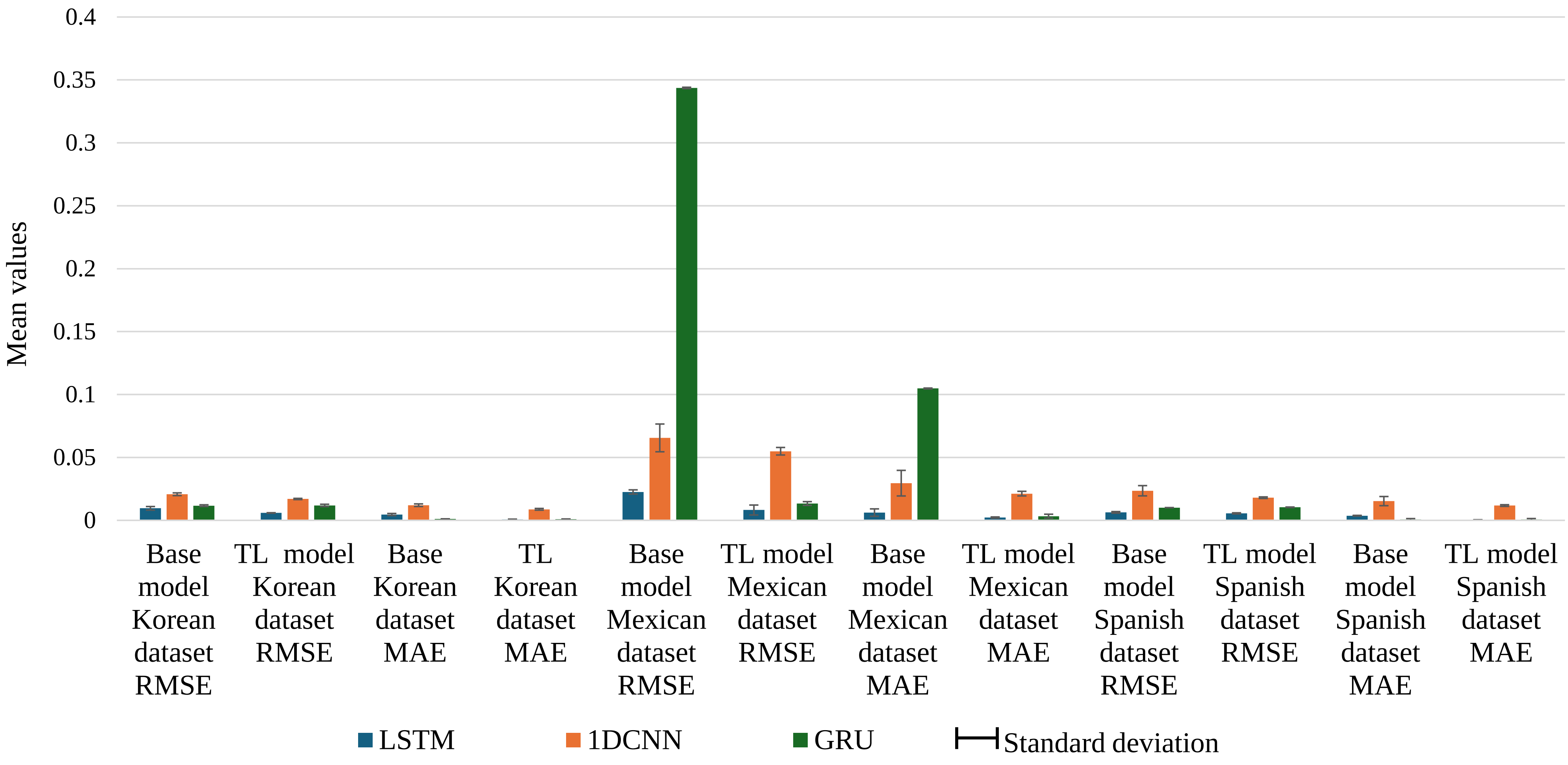
0 – Base Model| **1** – Model with North America Weights| **2** – North + South America Weights| **3** – North and South America + Europe Weights| **4** – North and South America, Europe + Africa Weights| **5** - North and South America, Europe, Africa + South Asia Weights| **6** - North and South America, Europe, Africa, South + East Asia Weights | **7** – Fully Trained Transfer Learning Model/ All Weights

5 Comparative analysis – Statistical Analysis



5 Comparative analysis – Statistical Analysis

Error bar graphs – 3 months data scarcity level



6 Conclusion and Future Works

- ✓ **Creation of a sequential transfer model that can predict air pollution in cold-start conditions across the globe**
  - ➢ Models were trained on datasets from seven major regions around the globe
  - ➢ Percentage improvement from 25% to over 99% over the base LSTM model

- ✓ **The proposed method is also effective on other models**
  - ➢ LSTM was best-suited for the proposed method
  - ➢ 1D-CNN and GRU also saw improvements in performance

- ✓ **Limitations and future works:**
  - ➢ Further optimization can be done
  - ➢ Exploration of ensemble models
  - ➢ Accounting for local environmental factors

THANK YOU!