Air Pollution and Cardiorespiratory Hospitalization, Modeling and Analysis Using Artificial Intelligence Techniques

Raja Sher Afgun Usmani (rajasherafgun@gmail.com)
Taylor's University School of Computer Science and Engineering https://orcid.org/0000-0003-2027-1425
Thulasyammal Ramiah Pillai
Taylor's University School of Computer Science and Engineering
Ibrahim Abaker Targio Hashem
University of Sharjah
Mohsen Marjani
Taylor's University School of Computer Science and Engineering
Rafiza Shaharudin
Ministry of Health Malaysia: Kementerian Kesihatan Malaysia
Mohd Talib Latif
Universiti Kebangsaan Malaysia

Research Article

Keywords: air pollution, health, air quality, predition, machine learning, hospitalization, hospital admissions

Posted Date: April 5th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-330603/v1

Noname manuscript No.
(will be inserted by the editor)

Air pollution and cardiorespiratory hospitalization, modeling and analysis using artificial intelligence techniques

Raja Sher Afgun Usmani · Thulasyammal Ramiah Pillai · Ibrahim Abaker Targio Hashem · Mohsen Marjani · Rafiza Binti Shaharudin · Mohd Talib Latif

Received: date / Accepted: date

Abstract Air pollution has a serious and adverse effect on human health, and it has become a risk to human welfare and health throughout the globe. In this paper, we present the modeling and analysis of air pollution and cardiorespiratory hospitalization. This study aims to investigate the association between cardiorespiratory hospitalization and air pollution, and predict cardiorespiratory hospitalization based on air pollution using the Artificial Intelligence (AI) techniques. We propose the Enhanced Long Short-Term Memory (ELSTM) model and provide a comparison with other AI techniques, i.e., Long Short-Term Memory (LSTM), Deep Learning (DL), and Vector Autoregressive (VAR). This study was conducted at seven study locations in Klang Valley, Malaysia. The prediction results show that the ELSTM model performed significantly better than other models in all study locations, with the best RMSE scores in Klang study location (ELSTM: 0.002, LSTM: 0.013, DL: 0.006, VAR: 0.066). The results also indicated that the proposed ELSTM model was able

Raja Sher Afgun Usmani
School of Computer Science and Engineering, Taylor's University, Selangor, Malaysia
ORCID: 0000-0003-2027-1425 E-mail: rajasherafgunusmani@sd.taylors.edu.my

Thulasyammal Ramiah Pillai
School of Computer Science and Engineering, Taylor's University, Selangor, Malaysia

Ibrahim Abaker Targio Hashem
College of Computing and Informatics, Department of Computer Science, University of Sharjah, 27272 Sharjah, UAE

Mohsen Marjani
School of Computer Science and Engineering, Taylor's University, Selangor, Malaysia

Rafiza Binti Shaharudin
Institute for Medical Research, Environmental Health Research Centre, Occupational Health Unit, Jalan Pahang, 50588, Kuala Lumpur, Malaysia

Mohd Talib Latif
Department of Earth Sciences and Environment, Faculty of Science and Technology, Universiti Kebangsaan Malaysia, 43600, Bangi, Selangor, Malaysia

to detect and predict the trends of monthly hospitalization significantly better than the LSTM and other models in the study. Hence, we can conclude that we can utilize AI techniques to accurately predict cardiorespiratory hospitalization based on air pollution in Klang Valley, Malaysia.

Keywords air pollution · health · air quality · predition · machine learning · hospitalization · hospital admissions

1 Introduction

Air pollution is among the main contributors to climate change throughout the world, and it is considered one of the most significant environmental challenges of the 21st century. According to WHO, Ninety-two per cent of the human population are breathing dirty air, and air pollution is attributed to 6.5 million deaths (11.6% of all deaths) worldwide (WHO et al., 2016). Air pollution, especially, urban air pollution has become a risk to human welfare and health throughout the globe as more than 50% of the world's population lives in urban areas and this number will increase to 70% by 2050 (Kampa and Castanas, 2008; Liu et al., 2013; Usmani et al., 2020e).

Air pollution is also associated with cardiorespiratory hospitalizations, i.e., respiratory hospitalizations are associated with nitrogen dioxide (NO_2) (Chen et al., 2020; Tajudin et al., 2019), nitrogen oxide (NO) (Barnett et al., 2005), sulfur dioxide (SO_2) (Tajudin et al., 2019), ozone (O_3) (Soleimani et al., 2019), and particulate matter (PM) (Chen et al., 2020), and cardiovascular hospitalizations are associated with PM (Sokoty et al., 2021; Wang et al., 2018), NO_2 (Wang et al., 2018), SO_2 (Amsalu et al., 2019), and (O_3 (Raza et al., 2019; Sokoty et al., 2021). Moreover, asthma hospitalizations are associated with NO_2, carbon monoxide (CO), and PM (Lee et al., 2002; Lin et al., 2002). However, there are very few studies available that predict hospitalizations based on air pollution. The daily hospital admissions resulting from air pollution are considered a leading issue in the field of environmental science, and an improvement in the available applied methodologies to estimate and predict air pollution based hospital admissions is essential (Araujo et al., 2020).

Asthma is one of the most common respiratory diseases and its attacks are affected by environmental factors such as meteorological factors and air pollutants. With the rise in air pollution around the world, asthma patients are at considerable risk. Alharbi and Abdullah (2019) argued that healthcare suppliers could adequately plan and deliver the services and provide resources to asthma patients if the need for their services is predicted (Alharbi and Abdullah, 2019). The authors use the linear regression model and the quantile regression model to predict the need for asthma care services in Polk, Iowa, USA. The results indicated that the weather variables provide better results, even with a low correlation between them. Chaves et al. (2017) conducted a similar study to accurately predict the number of asthma and pneumonia hospitalizations associated with air pollution in the city of São José dos Campos,

São Paulo State, Brazil, using a computational model with fuzzy logic based on Mamdani's inference method.

The recent studies in the prediction of hospitalizations based on air pollution are utilizing the Artificial Intelligence (AI) and Machine Learning (ML) techniques as AI and ML are found to be very helpful in solving numerous problems (Bellinger et al., 2017; Jones et al., 2016; Kang et al., 2018), particularly the problems related to a large amount of data (Bilal et al., 2020; Huck, 2019; Le and Cha, 2018; Ngiam and Khor, 2019; Zaree and Honarvar, 2018), which is usually the case in health impact studies. Araujo et al. (2020) (2020) showed significant improvements in estimating the respiratory hospital admissions due to meteorological variables and particulate matter in Campinas and São Paulo cities, Brazil using Artificial Neural Network (ANN) and ensemble methods . In another study to investigate the association between ambient air pollution and respiratory admissions, the Support vector regression (SVR) was applied to predict daily respiratory admissions based on air pollution (Zhou et al., 2019). The results show that air pollution was significantly associated with daily respiratory admissions, duration, and economic impact, and SVR successfully predicted daily respiratory admissions. Similarly, Abedi et al. (2020) conducted a time series study in Isfahan, Iran to investigate the validity of the relationship between air pollution and cardiovascular and respiratory hospitalization. The results show that the air quality index (AQI) significantly impacts cardiovascular hospitalization in both the long and short run. The results also indicate that the AQI has a stronger relationship with cardiovascular hospitalizations than respiratory hospitalizations.

To the best of our knowledge, a study of this variety and volume is not carried out before, especially in a Malaysian context. The closest attempts to predict air pollution's effects on hospitalization rates are presented in a few recent studies (Abedi et al., 2020; Alharbi and Abdullah, 2019; Araujo et al., 2020; Chaves et al., 2017; Zhou et al., 2019). The limitations of these studies can be categorized into i) limited scope (Abedi et al., 2020; Alharbi and Abdullah, 2019; Araujo et al., 2020; Chaves et al., 2017; Zhou et al., 2019) ii) limited parameters (Alharbi and Abdullah, 2019; Araujo et al., 2020; Chaves et al., 2017) iii) use averaged air pollution data from one location for other locations (Abedi et al., 2020; Chaves et al., 2017; Zhou et al., 2019). The study's aims include i) investigation in the associations between cardiorespiratory hospitalizations and air pollution, ii) prediction of the cardiorespiratory hospitalizations based on air pollution. To achieve these aims, we propose the Enhanced Long Short-Term Memory (ELSTM) model and provide a comparison with other AI techniques, i.e., Long Short-Term Memory (LSTM), Deep Learning (DL), and Vector Autoregressive (VAR). The ELSTM model performs the location-based prediction while utilizing the time-series nature of temporal parameters.

2 Methods and Materials

2.1 Study Location

This study is conducted in Klang Valley, Malaysia. Malaysia consists of two geographical regions divided by the South China Sea. East Malaysia is made up of the two largest states, Sabah and Sarawak, and West Malaysia consists of eleven states and three federal territories. Klang Valley is an urban region in West Malaysia that is centered in Kuala Lumpur. Klang Valley includes Kuala Lumpur's adjoining cities and towns in the state of Selangor. The study locations are presented in Figure 1. The study areas used in this study are scattered around the Klang valley, also known as Greater Kuala Lumpur, is an urban agglomeration of 7.564 million people as of 2018 (UN DESA, 2019). It is among the fastest growing metropolitan regions in Southeast Asia, in both population and economic development. The study areas include Klang (KLN), Shah Alam (SA), Putrajaya (PUJ), Petaling Jaya (PJ), Cheras, Kuala Lumpur (CKL), Batu Muda, Kuala Lumpur (BMKL) and Banting (BAN).

The first study location is Klang, officially Royal Town of Klang, which is a royal town and former capital of the state of Selangor, Malaysia. It is located within the Klang District. Port Klang, which is located in the Klang District, is the 12th busiest transshipment port and the 12th busiest container port in the world (Wikipedia, 2020c). Klang has a population of 879,867 (WorldoMeters, 2020).

The new capital of the state of Selangor is Shah Alam, which is our second study location. Shah Alam is situated within the Petaling District and a small portion of the neighbouring Klang District. Shah Alam is majorly an industrial area, with manufacturing playing a big role in its economy (Wikipedia, 2020f). Shah Alam was the first planned city in Malaysia after independence from Britain in 1957and it has a population of 740,750 (Wikipedia, 2020f).

The third study location is Putrajaya, a planned city and the Malaysian capital's federal administrative centre. The federal government's seat was shifted in 1999 from Kuala Lumpur to Putrajaya because of overcrowding and congestion in the former (Wikipedia, 2020e). Putrajaya's territory is entirely enclaved within the Sepang District of the state of Selangor. Putrajaya is also a part of the Multimedia Super Corridor (MSC), Malaysia, a special economic zone that covers Klang Valley, and it has a population of 91,900 (Theborneopost, 2018).

The fourth study location is Petaling Jaya, a city in Petaling District, in the state of Selangor, Malaysia. Developed initially as a satellite township for Kuala Lumpur, Malaysia's capital, it is part of the Greater Kuala Lumpur area. Petaling Jaya was granted city status on 20 June 2006. It has an area of approximately 97.2 square kilometres (37.5 sq mi) and a population of 638,516 (Wikipedia, 2020d). Kuala Lumpur surrounds Petaling Jaya to the east, Sungai Buloh to the north, Shah Alam, the capital of Selangor, and Subang Jaya to the west and Bandar Kinrara (Puchong) to the south.

The fifth and sixth study locations are from Kuala Lumpur, which is the biggest city in Malaysia. Kuala Lumpur is a federal territory and the capital city of Malaysia. It is the largest city in Malaysia, covering an area of 243 km^2 (94 sq mi) with an estimated population of 1.73 million as of 2016 (Wikipedia, 2020b). The study locations in Kuala Lumpur are situated in Cheras and Batu Muda. The last study location is Banting, which is an agricultural hub and the seat of Kuala Langat District, Selangor, Malaysia. Banting has a population of 93,497 (Wikipedia, 2020a). Banting will provide us with an interesting comparison with industrial study locations (Shah Alam), major port (Klang) and major cities (Cheras, Batu Muda, Petaling Jaya).

2.2 Data

In this study, the Monthly Air Pollution Hospitalization (MAPH) dataset is utilized. The dataset contains daily and monthly air pollutant readings from selected monitoring stations and associated count of cardiorespiratory hospitalization. The dataset is generated from the air quality dataset provided by the Department of Environment (DOE), Malaysia, and cardiorespiratory hospitalization statistics provided by the Ministry of Health (MOH), Malaysia. Figure 1 presents the air quality monitoring station locations, which present as our study locations' center point.

Fig. 1: Air Quality Monitoring Stations in Klang Valley, Malaysia

The dataset is generated using our previous work (Usmani et al., 2020c,d). The dataset cleaning and generation are carried out using the novel feature engineering algorithm (Usmani et al., 2020a). The association of patients with the AQM stations and MAPH dataset generation is carried out using the spatial feature engineering algorithm (Usmani et al., 2020c,d). The spatial

feature engineering algorithm contains the facility of specifying the radius parameter. The radius parameter plays an integral part in the inclusion criteria. A researcher can specify the maximum distance between the Air Quality Monitoring (AQM) station and the patient as a radius. The spatial feature engineering algorithm will ensure that those living far away from AQM stations are excluded from the dataset. The radius value of 10,000 meters was used to generate dataset for the current study.

There were twelve hospitals included in the study in total, with eight hospitals from Selangor and three hospitals from Kuala Lumpur and one hospital from Putrajaya. The hospitals from Selangor include the Hospital Ampang, Hospital Kajang, Hospital Sungai Buloh, Hospital Selayang, Hospital Orang Asli Gombak, Pusat Kawalan Kusta Negara, Hospital Serdang, and Hospital Tengku Ampuan Rahimah. The hospitals from Kuala Lumpur include the Hospital Kuala Lumpur, Hospital Rehabilitasi Cheras, Institut Perubatan Respiratori and Hospital Putrajaya was included from Putrajaya. In terms of hospitalization causes, the cardiovascular and respiratory hospitalization represented by the International Classification for Diseases (ICD) codes I00-I99 for cardiovascular diseases and J00-J99 for respiratory diseases are taken into consideration for this study (ICD, 2011).

2.3 Methods

In this section, we will discuss the models used to predict cardiorespiratory hospitalization in this study, the metrics used for comparison, and these models' results. In the end, we will discuss what the metrics indicate and how these models compare with each other. Table 1 shows the parameters of the dataset. The Date, Station, the air pollutants, i.e., O_3, particulate matter with diameter less than 10 micrometers PM_{10}, NO_x, NO_2, NO and SO_2 are used to predict the hospitalization count, which is the prediction or output parameter. For the prediction of cardiorespiratory hospitalization, the dataset is divided into two parts, 70% for training and 30% for testing.

2.3.1 Models

In this study, we have used four models 1) Long Short-Term Memory (LSTM) model, 2) Enhanced Long Short-Term Memory (ELSTM) model, 3) Deep Learning (DL) model, and 4) Vector autoregression (VAR) time series model. The first three models, i.e., ELSTM, LSTM, and Deep learning, are neural network-based models and the last model, i.e., VAR, is a time series model. We have included a time series model in our study for comparison due to our dataset's time series nature. The models were selected based on different criteria, i.e., i) LSTM model was selected for comparison as it is the most popular model available for multivariate time series prediction and classification, ii) the DL model was selected because it performed well in various studies for time series prediction and DL model is the basis for various DL-based time

series prediction models, including LSTM, iii) the VAR model was selected as it is the most popular traditional multivariate time series model. Furthermore, the methodology for VAR model is easier in practice (Simionescu, 2013), and the current study focused on the autoregressive properties of the time series dataset.

Long Short-Term Memory (LSTM): A special type of Recurrent Neural Network, LSTMs are able to learn long-term dependencies. Hochreiter & Schmidhuber (1997) developed LSTM, and many scholars working in the area of machine learning and data science have improved and popularized them (Bai et al., 2019; Hochreiter and Schmidhuber, 1997; Zhao et al., 2019). LSTMs perform exceedingly good on a wide range of problems and are commonly used in the modeling of time series (Bao et al., 2017; Hua et al., 2019; Malhotra et al., 2015). LSTMs were explicitly designed to prevent the long-term dependency problem. Remembering data for long periods is basically their default practice. Figure 2 demonstrates the three-gate LSTM architecture, i.e., 1) input gates that detect the value from the input can be used to change the memory, 2) Forget gate determines what data to discard from the block, and 3) Output gate generates the output depending on the block's input and memory.

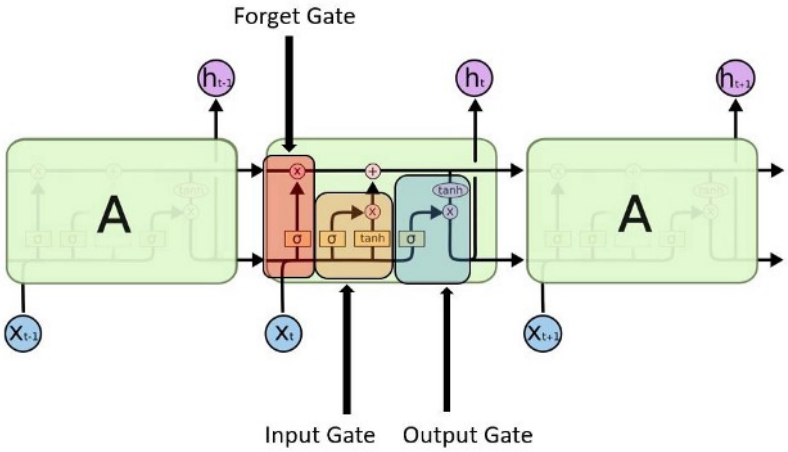


Fig. 2: Long Short-Term Memory

Enhanced Long Short-Term Memory (ELSTM): The proposed predictive spatial model, named Enhanced Long Short-Term Memory (ELSTM) is an enhanced form of LSTM, which uses spatial information and the previous data

associated with the model to learn and predict the outcome. The ELSTM model is enhanced to work with our datasets by optimizing the network's hyperparameters and learning mechanism. Figure 3 shows the ELSTM model's overview, with four distinct parts, i.e., 1) features, 2) the spatial layer, 3) the ELSTM units, and 4) the output parameter. The first part presents the fea-

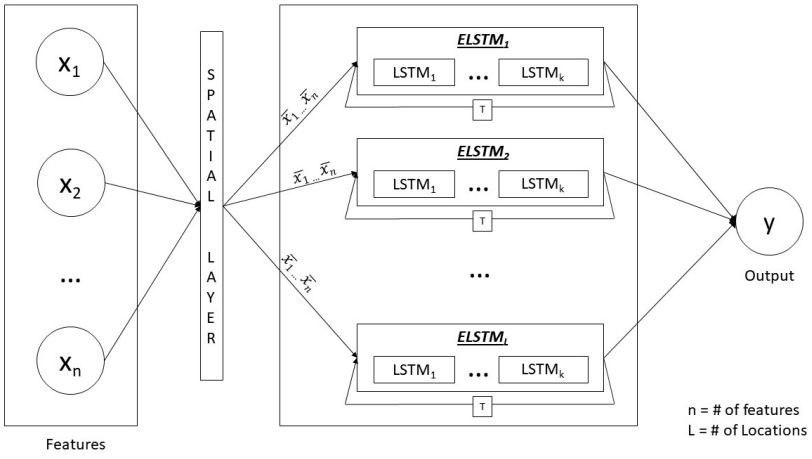


Fig. 3: Enhanced Long Short-Term Memory

tures from x_i to x_n, where n is the number of features. The second part is the spatial layer, which is responsible for remembering the state of the model according to the study location. The state of the model loaded in the spatial layer contains the best hyperparameters of the study location and the best number of time-steps. Using the best hyperparameters and time-step for each study location ensures that the ELSTM model performs to the best of its ability. The spatial layer is also responsible for the normalization of the features, with $\bar{x_i}$ to $\bar{x_n}$ representing the normalized values for the features.

The third part contains the study location based ELSTM units. Every ELSTM unit contains hidden layers of stacked LSTM, making the proposed model deeper and more accurately described as a deep learning technique. Stacked LSTMs allow the hidden state to operate at different timescales at each level, as the LSTM layer above provides a sequence output to the next LSTM layer. For both input time steps, precisely, one output per input time step, rather than one output time step. Each stacked LSTM contains multiple LSTMs, and each LSTM is designed according to the description provided in Section **??**. The last part of the design contains the output parameter, which is denoted by y. The output parameter is calculated by the ELSTM model based on the normalized values of features provided by the spatial layer.

Deep Learning (DL): Deep learning (DL) represents a larger family of machine learning approaches based on artificial neural networks. It is also known as deep structured learning. In DL, it is possible to use supervised, semi-supervised, or unsupervised learning and DL is used in different fields such as natural language processing (Young et al., 2018), voice recognition (Bae et al., 2016), computer vision (Voulodimos et al., 2018), social network filtering (Nguyen et al., 2017), among others, to generate results equal to and in some cases exceeding the human expert performance (Abou Jaoude et al., 2020; Lu et al., 2020; Williams, 2020). Figure 4 presents the architecture of deep learning, where $\overrightarrow{x}$ and $\overrightarrow{y}$ presenting the input and output parameters. $\overrightarrow{h_1}, \overrightarrow{h_2}, \overrightarrow{h_3}$ present the hidden layers, and W_1, W_2,W_3, W_4 are the weights of the deep learning network.

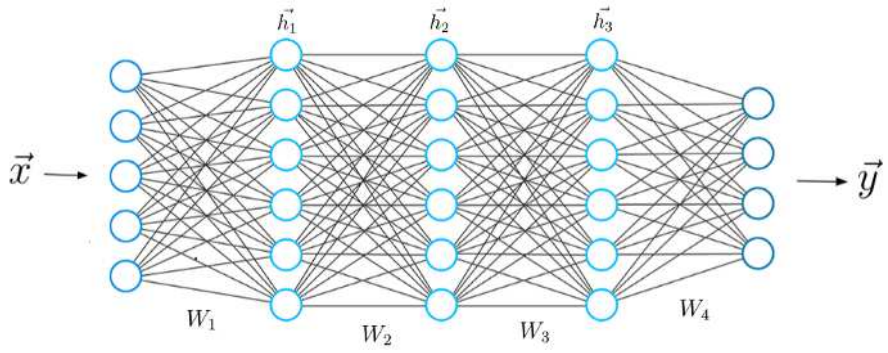


Fig. 4: Deep Learning

Vector autoregression (VAR): Vector autoregression (VAR) is an algorithm used for multivariate forecasting of two or more time series that impact each other (Zivot et al., 2003). That is, the relation is two-way between the time series concerned. Since each variable (Time Sequence) is represented as a function of past values, it is called an Autoregressive model, which means the predictors are none other than the set of lags (time-delayed value). The key difference between VAR and other Autoregressive models such as AR, ARMA, or ARIMA models is that VAR is not uni-directional. However, AR, ARMA, ARIMA models are uni-directional. The predicted variable is influenced by the predictor variables and not vice versa. Vector Auto Regression (VAR), meanwhile, is bi-directional. That is, the variables influence each other. Each variable in a VAR model is a linear function of itself's past values and all the other variables' past values.

2.3.2 Comparison Metrics

In this section, we will define the metrics used to compare the results of prediction models. We are using two metrics,i.e., 1) Root Mean Squared Error (RMSE) and 2) Mean Absolute Error (MAE).

Root Mean Squared Error (RMSE): RMSE is a rule for quadratic scoring that also calculates the average error magnitude. It is the square root of the forecast and actual data average of square differences. RMSE expresses the average error in the predictions of a model in units of the prediction variable. The equation 1 presents the mathematical form of RMSE.

$$RMSE = \sqrt{\frac{1}{n}\sum_{j=1}^{n}(y_j - \widehat{y}_j)^2} \tag{1}$$

The RMSE measure will vary from 0 to ∞, and does not take into account the direction of errors. As they are negatively-oriented scores, lower values for RMSE are considered better. RMSE gives major errors a comparatively high weight. This means that while significant errors are especially unacceptable, the RMSE can be more useful.

Mean Absolute Error (MAE): MAE calculates the average magnitude of the errors, without taking their direction into account. MAE is the average of the absolute differences between actual and prediction observation over the test sample, where all differences have same weight. The equation 2 presents the mathematical form of RMSE.

$$MAE = \frac{1}{n}\sum_{j=1}^{n}|y_j - \widehat{y}_j| \tag{2}$$

Like RMSE, MAE expresses an average prediction error of the model in the units of the predicted variable. MAE is favoured over RMSE, as RMSE does not explain average error alone, and has other interpretations that are tough to figure out. Same as RMSE, The metric of MAE will vary from 0 to ∞ and lower values for MAE are considered better as they are negatively-oriented scores.

3 Results and discussion

Table 1 presents the parameters and descriptive statistics of the MAPH dataset. The MAPH dataset is a time series dataset with a fixed time interval between reading is one month. The Station parameter contains the geocoded location for the AQM station. The second parameter is the date of the air quality readings. In the second column, Table 1 shows the date range for each study

location. The last parameter presents the monthly cardiorespiratory hospitalizations count, provided by the MOH, Malaysia and all other parameters contain the monthly average readings of air quality variables, i.e., O_3, PM_{10}, NO_x, NO_2, NO, and SO_2.

Table 1: Descriptive statistics of MAPH dataset

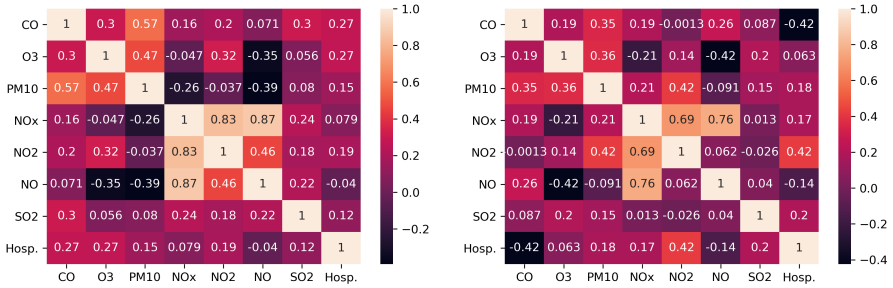
Station	Date	Statistic	PM_{10}**	O_3*	CO*	NO_x*	NO_2 *	NO*	SO_2*	Hospitalization count
KLN	Jan 2006 - Dec 2016	Mean	64	0.018	0.99	0.037	0.021	0.016	0.004	767
		Std	19	0.004	0.23	0.007	0.004	0.004	0.001	302
		Min	41	0.010	0.52	0.019	0.009	0.007	0.002	15
		Max	159	0.031	1.64	0.067	0.032	0.035	0.008	1377
SA	Jan 2006 - Dec 2016	Mean	52	0.021	0.79	0.037	0.020	0.017	0.003	301
		Std	16	0.004	0.19	0.006	0.004	0.004	0.001	107
		Min	34	0.014	0.46	0.021	0.009	0.010	0.001	71
		Max	147	0.034	1.33	0.058	0.038	0.026	0.006	622
PUT	Jan 2006 - Dec 2016	Mean	44	0.021	0.59	0.020	0.014	0.006	0.002	316
		Std	16	0.005	0.14	0.004	0.003	0.002	0.001	153
		Min	23	0.010	0.27	0.011	0.006	0.001	0.001	23
		Max	133	0.036	1.31	0.035	0.023	0.016	0.005	641
PJ	Jan 2006 - Dec 2016	Mean	49	0.015	1.29	0.062	0.029	0.033	0.004	141
		Std	15	0.003	0.23	0.008	0.004	0.007	0.001	64
		Min	26	0.009	0.86	0.044	0.017	0.019	0.002	19
		Max	126	0.026	1.94	0.081	0.039	0.050	0.007	288
CKL	Jan 2006 - Dec 2016	Mean	49	0.02	0.86	0.037	0.021	0.017	0.002	583
		Std	14	0.004	0.18	0.006	0.003	0.005	0.001	213
		Min	30	0.011	0.50	0.019	0.010	0.005	0.001	83
		Max	116	0.034	1.77	0.058	0.030	0.029	0.004	1076
BAN	Apr 2010 - Dec 2016	Mean	56	0.021	0.59	0.019	0.012	0.008	0.003	54
		Std	20	0.005	0.19	0.004	0.002	0.002	0.001	23
		Min	36	0.013	0.33	0.012	0.006	0.003	0.001	9
		Max	144	0.034	1.32	0.027	0.017	0.013	0.006	124
BMKL	Jan 2009 - Dec 2016	Mean	48	0.020	0.78	0.029	0.017	0.012	0.002	817
		Std	17	0.004	0.13	0.004	0.003	0.003	0.001	423
		Min	28	0.011	0.56	0.019	0.011	0.005	0.001	59
		Max	131	0.031	1.50	0.041	0.026	0.021	0.005	1727

Note: ** $\mu g/m^3$, *ppm, Std=Standard Deviation, Min=Minimum, Max=Maximum
Numbers highlighted in red represent the highest values.

The MAPH dataset contains the data from January 2006 to December 2016 for five study locations, i.e., Klang (KLN), Shah Alam (SA), Putrajaya (PUJ), Petaling Jaya (PJ), and Cheras, Kuala Lumpur (CKL). Banting (BAN) and Batu Muda, Kuala Lumpur (BMKL) AQM stations are relatively new stations. The dataset for Banting contains data from April 2010 to December 2016, and the data for Batu Muda, Kuala Lumpur, contains data from January 2009 to December 2016.

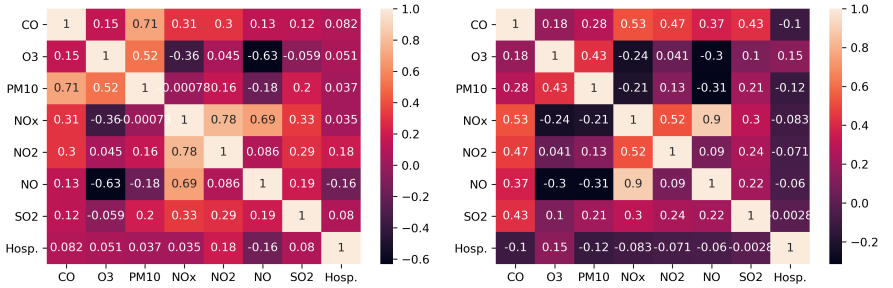
The descriptive statistics show the varying readings for air pollutants and hospitalizations in all study locations. The highest maximum values for each parameter is highlighted in Table 1. The highest mean hospitalizations (817) and second-highest mean hospitalizations (767) are found in BMKL and KLN. BMKL is located in Kuala Lumpur, which is the biggest city in Malaysia. The same pattern is found for maximum hospitalization count, with BMKL and KLN leading the statistics with 1727 and 1377 maximum hospitalization counts in a single month. The lowest mean hospitalization is found in BAN (54) and PJ (141).

Figure 5 shows the correlation matrices of all seven study locations. The correlation matrices show an interesting combination of correlations. It also



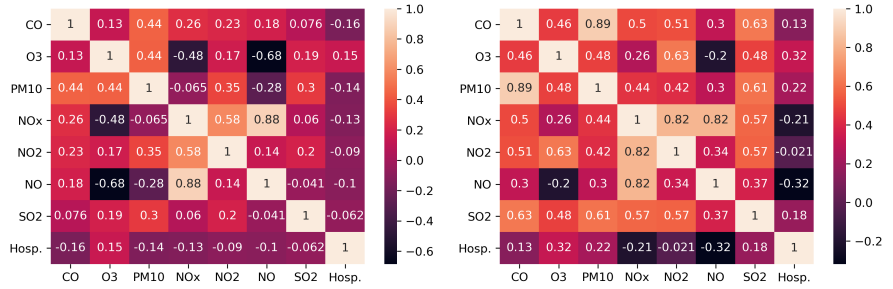
(a) Klang

(b) Shah Alam



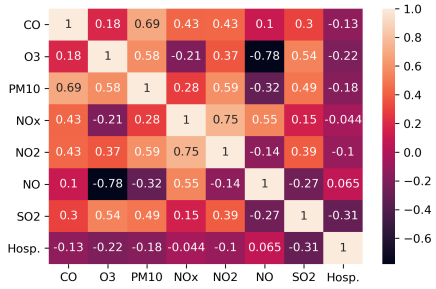
(c) Putrajaya

(d) Petaling Jaya



(e) Cheras, Kuala Lumpur

(f) Banting



(g) Batu Muda, Kuala Lumpur

Fig. 5: Correlation Matrices: Air quality parameters & Hospitalization

highlights the variations between the correlations with respect to the study locations. The air quality variable O_3 show a positive correlation with the hospitalization count in six of the seven study locations. The nitrogen-related air quality parameters also have positive correlations with hospitalization counts, i.e., NO_x has three positive correlations, NO_2 has three positive correlations, and NO has one positive correlation. The PM_{10}, SO_2 and CO showed four positive correlations each. The most positive correlations were found in the Klang, Shah Alam and Putrajaya study locations, with all seven air pollutants showed a positive correlation with hospitalization count in most study areas. The least positive correlations were found in the Petaling Jaya, Cheras and Batu Muda study areas, with only one air quality variable showing positive correlations with hospitalization count.

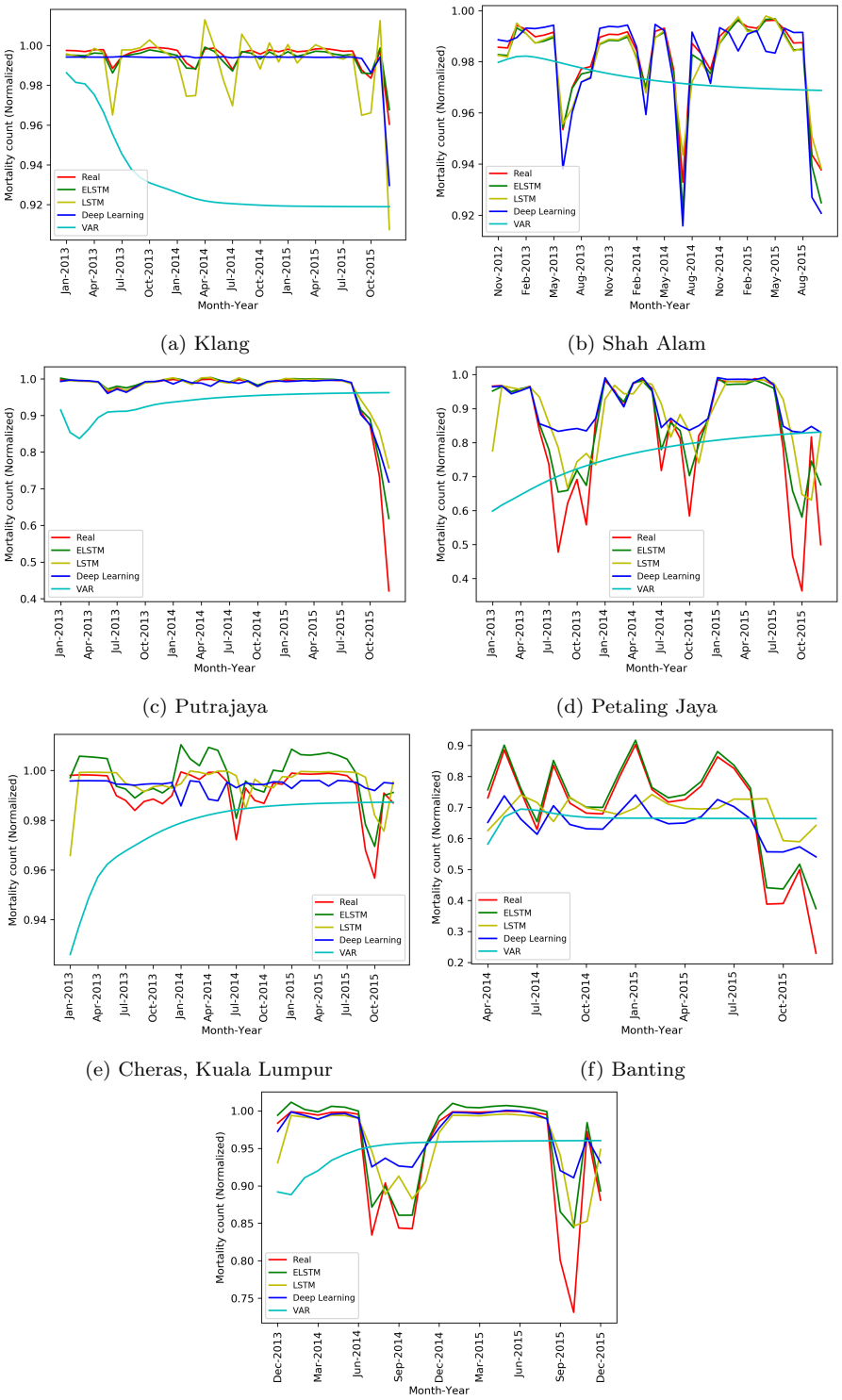
After observing the correlations for all parameters and the descriptive statistics of the dataset, it could be concluded that there no patterns are found between positive or negative correlations and increased or decreased cardiorespiratory hospitalization. Subsequently, no pattern was observed for the highest or lowest mean hospitalization with any air pollutant's highest or lowest mean in any study locations.

3.1 Model Results and Comparisons

The ELSTM model uses the stacked LSTM architecture, which enables the proposed model to describe more complex input patterns at every LSTM layer, creating a deep network. The ELSTM model is also equipped with the spatial layer to tune the model's hyperparameters for each study location. The results presented in Table 2 and Figure 6 show that these enhancements help the proposed ELSTM model to achieve the best results in comparison with all models, especially the LSTM model. The comparison of the base LSTM and ELSTM model showed a significant difference of results in terms of RMSE and MAE scores. As presented in Figure 6, the ELSTM model was able to detect and predict the trends of monthly hospitalization significantly better than the LSTM and other models in the study.

Figure 6 presents the hospitalization predictions by the models used in the study. Every graph represents a study location. The predictions were completed on the testing dataset (70% data), i.e., 40 months for five study locations (Klang, Shah Alam, Putrajaya, Petaling Jaya, Cheras, Kuala Lumpur), 24 months for Banting, and 29 months for Batu Muda, Kuala Lumpur. The prediction results demonstrated significant prediction power of ELSTM with respect to other models in every study location. The LSTM and DL models also performed well in some study locations, but the VAR model was the weakest model in every study location.

The values presented in Table 2 show a location-wise RMSE and MAE for each model by study location. RMSE is considered the primary criterion for predictive models. The values of RMSE indicate that the proposed ELSTM model performed better than the other models in all study locations. Among



(a) Klang

(b) Shah Alam

(c) Putrajaya

(d) Petaling Jaya

(e) Cheras, Kuala Lumpur

(f) Banting

(g) Batu Muda, Kuala Lumpur

Fig. 6: Comparison of hospitalization predictions

Table 2: Comparison of MAE and RMSE - Hospitalization prediction

Station	ELSTM		LSTM		DL		VAR	
	MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE
KLN	0.002	0.002	0.008	0.013	0.004	0.006	0.063	0.066
SA	0.003	0.004	0.003	0.004	0.006	0.008	0.016	0.018
PUJ	0.010	0.034	0.016	0.060	0.014	0.051	0.077	0.117
PJ	0.012	0.033	0.092	0.140	0.081	0.152	0.181	0.212
CKL	0.007	0.007	0.007	0.011	0.006	0.010	0.018	0.024
BAN	0.025	0.038	0.115	0.159	0.11	0.126	0.142	0.175
BMKL	0.016	0.028	0.034	0.055	0.029	0.054	0.068	0.084

Note: The lowest values for MAE and RMSE are underlined

other models, LSTM and deep learning performed better than VAR, owing to the conclusion that VAR is not suitable for long term predictions. Table 2 also presents the MAE values for the models used in this study. It is well known that RMSE will always be equal or greater to MAE values as RMSE gives more importance to the biggest errors. Hence, it is useful to compare these two matrices when large errors are undesirable, which is our study's case. The comparison for RMSE and MAE for ELSTM shows that the proposed model does not have large residual errors and is able to follow the trend of hospitalization and predict accordingly. The DL, LSTM, and VAR models do not show large residual errors, but a comparison of these models with ELSTM shows that the difference between RMSE and MAE is the smallest in the ELSTM model. Also, the MAE score shows that the ELSTM model performs better than other models and, LSTM and DL models perform better than the VAR model.

Hospitalization is associated with NO_2 (Barnett et al., 2005; Lee et al., 2002; Tajudin et al., 2019), CO (Lee et al., 2002), PM (Lee et al., 2002), SO_2 (Goudarzi et al., 2016; Lee et al., 2002; Tajudin et al., 2019) and O_3 (Lee et al., 2002; Luong et al., 2018; Tajudin et al., 2019). In our study, the readings for $PM_{2.5}$ are not utilized, but PM_{10} readings also include the readings for $PM_{2.5}$. The models presented in our study, especially the ELSTM model were able to predict the monthly hospitalization using the air pollutants. Consequently, it is concluded that these air pollutants impact human hospitalization, and steps should be taken to ensure their low future levels.

Our findings suggest that air quality can be used to accurately predict the cardiorespiratory hospitalization of Klang Valley residents, Malaysia. An interesting aspect of this research is that we have used different predictive models, and most of them can predict hospitalization relatively well. Subsequently, we can recommend that, the air quality warning systems can be of great use for controlling air pollutant emissions (Usmani et al., 2020b) and the efforts to control air quality should continue and be reinforced in the future as it has a clear impact on hospitalization.

To the best of our knowledge, a study to predict the cariorespiratory hospitalizations of this variety and volume is not done before, especially in a Malaysian context. The closest attempts to predict air pollution's effects on hospitalization rates are presented in few recent studies (Abedi et al., 2020; Alharbi and Abdullah, 2019; Araujo et al., 2020; Chaves et al., 2017; Zhou et al., 2019). These studies also have limitations which can be categorized into i) limited scope (Abedi et al., 2020; Alharbi and Abdullah, 2019; Araujo et al., 2020; Chaves et al., 2017; Zhou et al., 2019) ii) limited parameters (Alharbi and Abdullah, 2019; Araujo et al., 2020; Chaves et al., 2017) iii) use averaged AP data from one location for other locations (Abedi et al., 2020; Chaves et al., 2017; Zhou et al., 2019).

The limitations of the current study include the use of residential address as a patient's exposure association parameter. The data with residence location as an indicator for air pollution exposure is easy to collect, monitor, and it applies to various studies (Jie, 2017; Mabahwi et al., 2018; Tajudin et al., 2019). The use of residence location as an indicator for air pollution exposure is argued against due to aggregate exposures from the point, non-point, and mobile sources, as well as multiple possible exposure pathways (Huang and Batterman, 2000). For future study, we plan further to investigate the association of air pollutants and hospitalization and quantify a decrease in air pollution's impact on cardiorespiratory hospitalization. This study's future work can also include individual level, location-specific, quantitative evaluation of air pollution exposures to mitigate the risk of exposure misclassification by verifying the levels of exposure among the various sections of the cohort. Our study also opens up an avenue of hospitalization prediction for daily air quality and association of lags and hospitalization, whether in daily or monthly predictions.

4 Conclusion

The study suggests an association between air pollution and cardiorespiratory hospitalization. The air pollutants O_3, PM_{10}, NO_x, NO_2, NO and SO_2 are used to predict cardiorespiratory hospitalization. The study areas chosen for the study ranged from densely populated areas to relatively smaller cities. Although there was no correlation seen between patterns of air pollutants with cardiorespiratory hospitalization by descriptive analysis, however, the results showed that the proposed ELSTM model was suitable for predicting cardiorespiratory hospitalization based on the RMSE scores. The ELSTM model also performed better than the other AI models in all study locations with good predictive power. The LSTM and DL models also performed relatively well but VAR time series model had the least prediction power and performed poorly for long-term cardiorespiratory hospitalization predictions in all study locations. Therefore, we conclude that there was an association between air pollution and cardiorespiratory hospitalization as we could predict cardiorespiratory hospitalization quite accurately using air pollutant parameters (O_3,

PM_{10}, NO_x, NO_2, NO and SO_2) and we recommend continuous efforts in controlling ambient air quality as it has a clear impact on cardiorespiratory hospitalization among the population in the Klang Valley, Malaysia.

Declarations

All co-authors have seen and agree with the contents of the manuscript and there is no financial interest to report. We certify that the submission is original work and is not under review at any other publication.

Ethical Approval

Not applicable as the research is using secondary datasets provided by DOE and MOH, Malaysia.

Consent to Participate

Not applicable as the research is using secondary datasets provided by DOE and MOH, Malaysia.

Consent to Publish

The consent to publish is granted from the DG of health, Malaysia.

Authors Contributions

This research work is part of a project entitled as "Modeling and Visualization of Air-Pollution and its Impacts on Health". R.S.A. Usmani conducted the experiments and wrote the manuscript. TR Pillai conceived and designed the research. IAT Hashem and M Marjani are responsible for feature engineering and analysis. MT Latif and R Shaharuddin are responsible for data collection, statictical analysis, and discussion. All authors reviewed the final manuscript.

Funding

This research is funded by Taylor's University under the research grant application ID (TUFR/2017/004/04) entitled as "Modeling and Visualization of Air-Pollution and its Impacts on Health".

Competing Interests

The authors declare no competing interests.

Availability of data and materials

The datasets used for the experimentation and analysis are part of MOU between Taylor's university, DOE, Malaysia and MOH, Malaysia. These datasets are not available to publish publicly according to the MOU.

Acknowledgements We would like to thank the Director General of Health, Malaysia for his permission to publish this article. We would also like to extend our thank to Department of Environment and Ministry of Health, Malaysia for providing datasets of air quality and hospitalization respectively.

References

Abedi A, Baygi MM, Poursafa P, Mehrara M, Amin MM, Hemami F, Zarean M (2020) Air pollution and hospitalization: an autoregressive distributed lag (ARDL) approach. Environmental Science and Pollution Research DOI 10.1007/s11356-020-09152-x

Abou Jaoude M, Sun H, Pellerin KR, Pavlova M, Sarkis RA, Cash SS, Westover MB, Lam AD (2020) Expert-level automated sleep staging of long-term scalp EEG recordings using deep learning. Sleep DOI 10.1093/sleep/zsaa112

Alharbi E, Abdullah M (2019) Asthma attack prediction based on weather factors. Periodicals of Engineering and Natural Sciences DOI 10.21533/pen.v7i1.422

Amsalu E, Guo Y, Li H, Wang T, Liu Y, Wang A, Liu X, Tao L, Luo Y, Zhang F, Yang X, Li X, Wang W, Guo X (2019) Short-term effect of ambient sulfur dioxide (SO2) on cause-specific cardiovascular hospital admission in Beijing, China: A time series study. Atmospheric Environment DOI 10.1016/j.atmosenv.2019.03.015

Araujo LN, Belotti JT, Alves TA, Tadano YdS, Siqueira H (2020) Ensemble method based on Artificial Neural Networks to estimate air pollution health risks. Environmental Modelling and Software DOI 10.1016/j.envsoft.2019.104567

Bae HS, Lee HJ, Lee SG (2016) Voice recognition based on adaptive MFCC and deep learning. In: Proceedings of the 2016 IEEE 11th Conference on Industrial Electronics and Applications, ICIEA 2016, DOI 10.1109/ICIEA.2016.7603830

Bai Y, Zeng B, Li C, Zhang J (2019) An ensemble long short-term memory neural network for hourly PM2.5 concentration forecasting. Chemosphere DOI 10.1016/j.chemosphere.2019.01.121

Bao W, Yue J, Rao Y (2017) A deep learning framework for financial time series using stacked autoencoders and long-short term memory. PLoS ONE DOI 10.1371/journal.pone.0180944

Barnett AG, Williams GM, Schwartz J, Neller AH, Best TL, Petroeschevsky AL, Simpson RW (2005) Air pollution and child respiratory health: A case-crossover study in Australia and New Zealand. American Journal of

Respiratory and Critical Care Medicine 171(11):1272–1278, DOI 10.1164/rccm.200411-1586OC

Bellinger C, Mohomed Jabbar MS, Zaïane O, Osornio-Vargas A (2017) A systematic review of data mining and machine learning for air pollution epidemiology. BMC Public Health 17(1):907, DOI 10.1186/s12889-017-4914-3

Bilal M, Usmani RSA, Tayyab M, Mahmoud AA, Abdalla RM, Marjani M, Pillai TR, Targio Hashem IA (2020) Smart Cities Data: Framework, Applications, and Challenges BT - Handbook of Smart Cities. Springer International Publishing, Cham, pp 1–29, DOI 10.1007/978-3-030-15145-4_6-1, URL https://doi.org/10.1007/978-3-030-15145-4{_}6-1

Chaves LE, Nascimento LFC, Rizol PMSR (2017) Fuzzy model to estimate the number of hospitalizations for asthma and pneumonia under the effects of air pollution. Revista de saude publica DOI 10.1590/S1518-8787.2017051006501

Chen C, Liu X, Wang X, Qu W, Li W, Dong L (2020) Effect of air pollution on hospitalization for acute exacerbation of chronic obstructive pulmonary disease, stroke, and myocardial infarction. Environmental Science and Pollution Research DOI 10.1007/s11356-019-07236-x

Goudarzi G, Geravandi S, Idani E, Hosseini SA, Baneshi MM, Yari AR, Vosoughi M, Dobaradaran S, Shirali S, Marzooni MB, Ghomeishi A, Alavi N, Alavi SS, Mohammadi MJ (2016) An evaluation of hospital admission respiratory disease attributed to sulfur dioxide ambient concentration in Ahvaz from 2011 through 2013. Environmental Science and Pollution Research 23(21):22001–22007, DOI 10.1007/s11356-016-7447-x

Hochreiter S, Schmidhuber J (1997) Long Short-Term Memory. Neural Computation DOI 10.1162/neco.1997.9.8.1735

Hua Y, Zhao Z, Li R, Chen X, Liu Z, Zhang H (2019) Deep Learning with Long Short-Term Memory for Time Series Prediction. IEEE Communications Magazine DOI 10.1109/MCOM.2019.1800155, 1810.10161

Huang YL, Batterman S (2000) Residence location as a measure of environmental exposure: A review of air pollution epidemiology studies. Journal of Exposure Analysis and Environmental Epidemiology 10(1):66–85, DOI 10.1038/sj.jea.7500074

Huck N (2019) Large data sets and machine learning: Applications to statistical arbitrage. European Journal of Operational Research DOI 10.1016/j.ejor.2019.04.013

ICD (2011) International Statistical Classification of Diseases and Related Health Problems. In: Encyclopedia of Clinical Neuropsychology, DOI 10.1007/978-0-387-79948-3_3055

Jie Y (2017) Air pollution associated with sumatran forest fires and mortality on the malay peninsula. Polish Journal of Environmental Studies 26(1):163–171, DOI 10.15244/pjoes/64642

Jones DE, Ghandehari H, Facelli JC (2016) A review of the applications of data mining and machine learning for the prediction of biomedical properties of nanoparticles. Computer Methods and Programs in Biomedicine 132:93–103, DOI 10.1016/j.cmpb.2016.04.025, URL http://

www.sciencedirect.com/science/article/pii/S016926071630027X

Kampa M, Castanas E (2008) Human health effects of air pollution. Environmental pollution 151(2):362–367

Kang GK, Gao JZ, Chiao S, Lu S, Xie G (2018) Air Quality Prediction: Big Data and Machine Learning Approaches. International Journal of Environmental Science and Development 9(1):8–16, DOI 10.18178/ijesd.2018.9.1.1066

Le VD, Cha SK (2018) Real-time Air Pollution prediction model based on Spatiotemporal Big data. arXiv preprint arXiv:180500432 URL http://arxiv.org/abs/1805.00432, 1805.00432

Lee JT, Kim H, Song H, Hong YC, Cho YS, Shin SY, Hyun YJ, Kim YS (2002) Air pollution and asthma among children in Seoul, Korea. Epidemiology 13(4):481–484, DOI 10.1097/00001648-200207000-00018

Lin M, Chen Y, Burnett RT, Villeneuve PJ, Krewski D (2002) The influence of ambient coarse particulate matter on asthma hospitalization in children: Case-crossover and time-series analyses. Environmental Health Perspectives 110(6):575–581, DOI 10.1289/ehp.02110575

Liu HY, Skjetne E, Kobernus M (2013) Mobile phone tracking: In support of modelling traffic-related air pollution contribution to individual exposure and its implications for public health impact assessment. DOI 10.1186/1476-069X-12-93

Lu D, Polomac N, Gacheva I, Hattingen E, Triesch J (2020) HUMAN-EXPERT-LEVEL BRAIN TUMOR DETECTION USING DEEP LEARNING WITH DATA DISTILLATION AND AUGMENTATION. 2006.12285

Luong LM, Phung D, Dang TN, Sly PD, Morawska L, Thai PK (2018) Seasonal association between ambient ozone and hospital admission for respiratory diseases in Hanoi, Vietnam. PLoS ONE 13(9), DOI 10.1371/journal.pone.0203751

Mabahwi NA, Leh OLH, Musthafa SNAM, Aiyub K (2018) Air quality-related human health in an urban region. Case study: State of Selangor, Malaysia. EnvironmentAsia 11(1):194–216, DOI 10.14456/ea.2018.15

Malhotra P, Vig L, Shroff G, Agarwal P (2015) Long Short Term Memory networks for anomaly detection in time series. In: 23rd European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning, ESANN 2015 - Proceedings

Ngiam KY, Khor IW (2019) Big data and machine learning algorithms for health-care delivery. DOI 10.1016/S1470-2045(19)30149-4

Nguyen DT, Alam F, Ofli F, Imran M (2017) Automatic image filtering on social networks using deep learning and perceptual hashing during crises. In: Proceedings of the International ISCRAM Conference, 1704.02602

Raza A, Dahlquist M, Jonsson M, Hollenberg J, Svensson L, Lind T, Ljungman PL (2019) Ozone and cardiac arrest: The role of previous hospitalizations. Environmental Pollution DOI 10.1016/j.envpol.2018.10.042

Simionescu M (2013) The use of varma models in forecasting macroeconomic indicators. Economics and Sociology DOI 10.14254/2071-789X.2013/6-2/9

Sokoty L, Rimaz S, Hassanlouei B, Kermani M, Janani L (2021) Short-term effects of air pollutants on hospitalization rate in patients with cardiovascular disease: a case-crossover study. Environmental Science and Pollution Research DOI 10.1007/s11356-021-12390-2, URL https://doi.org/10.1007/s11356-021-12390-2

Soleimani Z, Boloorani AD, Khalifeh R, Teymouri P, Mesdaghinia A, Griffin DW (2019) Air pollution and respiratory hospital admissions in Shiraz, Iran, 2009 to 2015. Atmospheric Environment DOI 10.1016/j.atmosenv.2019.04.030

Tajudin MABA, Khan MF, Mahiyuddin WRW, Hod R, Latif MT, Hamid AH, Rahman SA, Sahani M (2019) Risk of concentrations of major air pollutants on the prevalence of cardiovascular and respiratory diseases in urbanized area of Kuala Lumpur, Malaysia. Ecotoxicology and Environmental Safety 171:290–300, DOI 10.1016/j.ecoenv.2018.12.057

Theborneopost (2018) Malaysia's population stood at 32.6 million in Q4 2018. URL http://www.theborneopost.com/2019/02/13/malaysias-population-stood-at-32-6-million-in-q4-2018/

UN DESA (2019) World Urbanization Prospects, The 2018 Revision. URL https://population.un.org/wup/Publications/Files/WUP2018-Report.pdf

Usmani RSA, Azmi WNFBW, Abdullahi AM, Hashem IAT, Pillai TR (2020a) A novel feature engineering algorithm for air quality datasets. Indonesian Journal of Electrical Engineering and Computer Science 19(3)

Usmani RSA, Hashem IAT, Pillai TR, Saeed A, Abdullahi AM (2020b) Geographic Information System and Big Spatial Data. International Journal of Enterprise Information Systems (IJEIS) 16(4)

Usmani RSA, Pillai TR, Hashem IAT, Jhanjhi NZ, Saeed A (2020c) A Spatial Feature Engineering Algorithm for Creating Air Pollution Health Datasets. URL https://www.techrxiv.org/articles/preprint/A{_}Spatial{_}Feature{_}Engineering{_}Algorithm{_}for{_}Creating{_}Air{_}Pollution{_}Health{_}Datasets/12376427/2

Usmani RSA, Pillai TR, Hashem IAT, Jhanjhi NZ, Saeed A (2020d) A Spatial Feature Engineering Algorithm for Creating Air Pollution Health Datasets. DOI 10.1016/j.ijcce.2020.11.004, URL https://linkinghub.elsevier.com/retrieve/pii/S2666307420300115https://www.techrxiv.org/articles/preprint/A{_}Spatial{_}Feature{_}Engineering{\ _ }Algorithm{\ _ }for{\ _ }Creating{\ _ }Air{\ _ }Pollution{_}Health{_}Datasets/12376427/2

Usmani RSA, Saeed A, Abdullahi AM, Pillai TR, Jhanjhi NZ, Hashem IAT (2020e) Air pollution and its health impacts in Malaysia: a review. Air Quality, Atmosphere & Health DOI 10.1007/s11869-020-00867-x, URL https://doi.org/10.1007/s11869-020-00867-xhttp://link.springer.com/10.1007/s11869-020-00867-x

Voulodimos A, Doulamis N, Doulamis A, Protopapadakis E (2018) Deep Learning for Computer Vision: A Brief Review. DOI 10.1155/2018/7068349

Wang X, Wang W, Jiao S, Yuan J, Hu C, Wang L (2018) The effects of air pollution on daily cardiovascular diseases hospital admissions in Wuhan from 2013 to 2015. Atmospheric Environment DOI 10.1016/j.atmosenv.2018.03.036

WHO, Osseiran N, Chriscaden K, WHO (2016) WHO releases country estimates on air pollution exposure and health impact. URL https://www.who.int/en/news-room/detail/27-09-2016-who-releases-country-estimates-on-air-pollution-exposure-and-health-impacthttp://www.who.int/mediacentre/news/releases/2016/air-pollution-estimates/en/

Wikipedia (2020a) Banting. URL https://en.wikipedia.org/wiki/Banting

Wikipedia (2020b) Kuala Lumpur. URL https://en.wikipedia.org/wiki/Kuala{_}Lumpur

Wikipedia (2020c) List of busiest container ports. URL https://en.wikipedia.org/wiki/List{_}of{_}busiest{_}container{_}ports

Wikipedia (2020d) Petaling Jaya. URL https://en.wikipedia.org/wiki/Petaling{_}Jaya

Wikipedia (2020e) Putrajaya. URL https://en.wikipedia.org/wiki/Putrajaya

Wikipedia (2020f) Shah Alam. URL https://en.wikipedia.org/wiki/Shah{_}Alam

Williams DP (2020) On the Use of Tiny Convolutional Neural Networks for Human-Expert-Level Classification Performance in Sonar Imagery. IEEE Journal of Oceanic Engineering DOI 10.1109/JOE.2019.2963041

WorldoMeters (2020) Malaysia Population 2020. URL https://www.worldometers.info/world-population/malaysia-population/

Young T, Hazarika D, Poria S, Cambria E (2018) Recent trends in deep learning based natural language processing [Review Article]. DOI 10.1109/MCI.2018.2840738

Zaree T, Honarvar AR (2018) Improvement of air pollution prediction in a smart city and its correlation with weather conditions using metrological big data. Turkish Journal of Electrical Engineering and Computer Sciences 26(3):1302–1313, DOI 10.3906/elk-1707-99

Zhao J, Deng F, Cai Y, Chen J (2019) Long short-term memory - Fully connected (LSTM-FC) neural network for PM2.5 concentration prediction. Chemosphere DOI 10.1016/j.chemosphere.2018.12.128

Zhou H, Wang T, Zhou F, Liu Y, Zhao W, Wang X, Chen H, Cui Y (2019) Ambient Air Pollution and Daily Hospital Admissions for Respiratory Disease in Children in Guiyang, China. Frontiers in Pediatrics DOI 10.3389/fped.2019.00400

Zivot E, Wang J, Zivot E, Wang J (2003) Vector Autoregressive Models for Multivariate Time Series. In: Modeling Financial Time Series with S-Plus®, DOI 10.1007/978-0-387-21763-5_11

Figures

Figure 1

Air Quality Monitoring Stations in Klang Valley, Malaysia Note: The designations employed and the presentation of the material on this map do not imply the expression of any opinion whatsoever on the part of Research Square concerning the legal status of any country, territory, city or area or of its authorities, or concerning the delimitation of its frontiers or boundaries. This map has been provided by the authors.

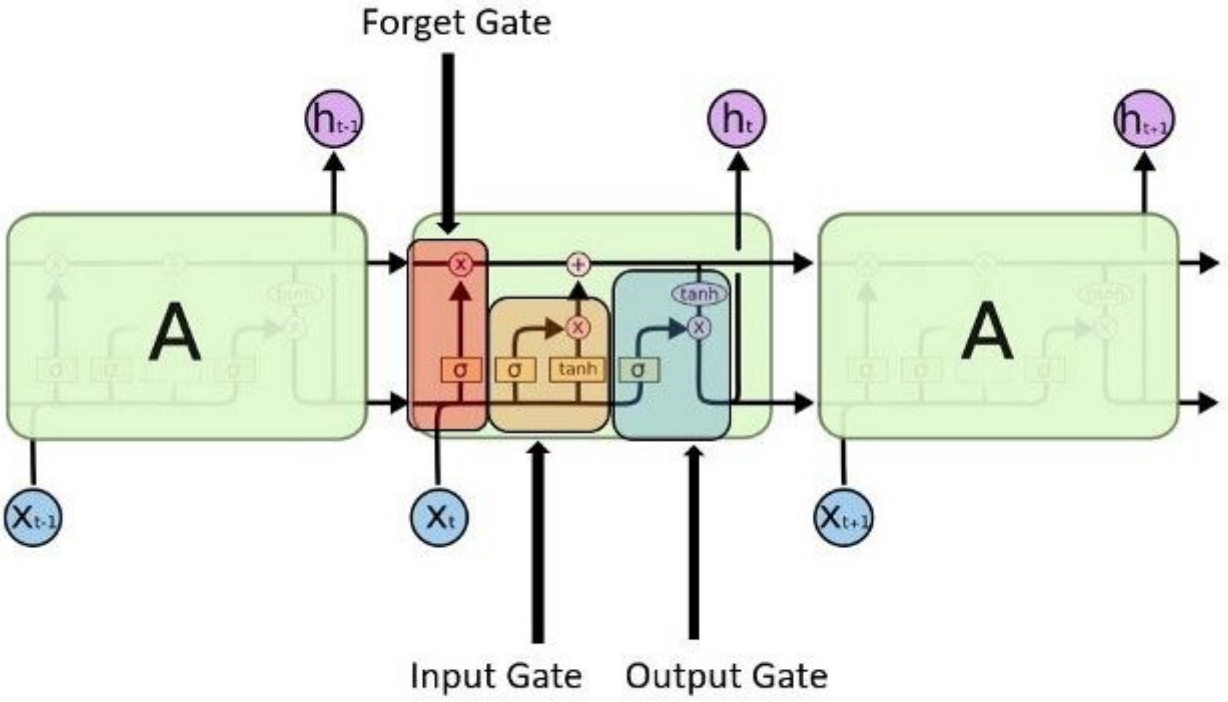


Figure 2

Long Short-Term Memory

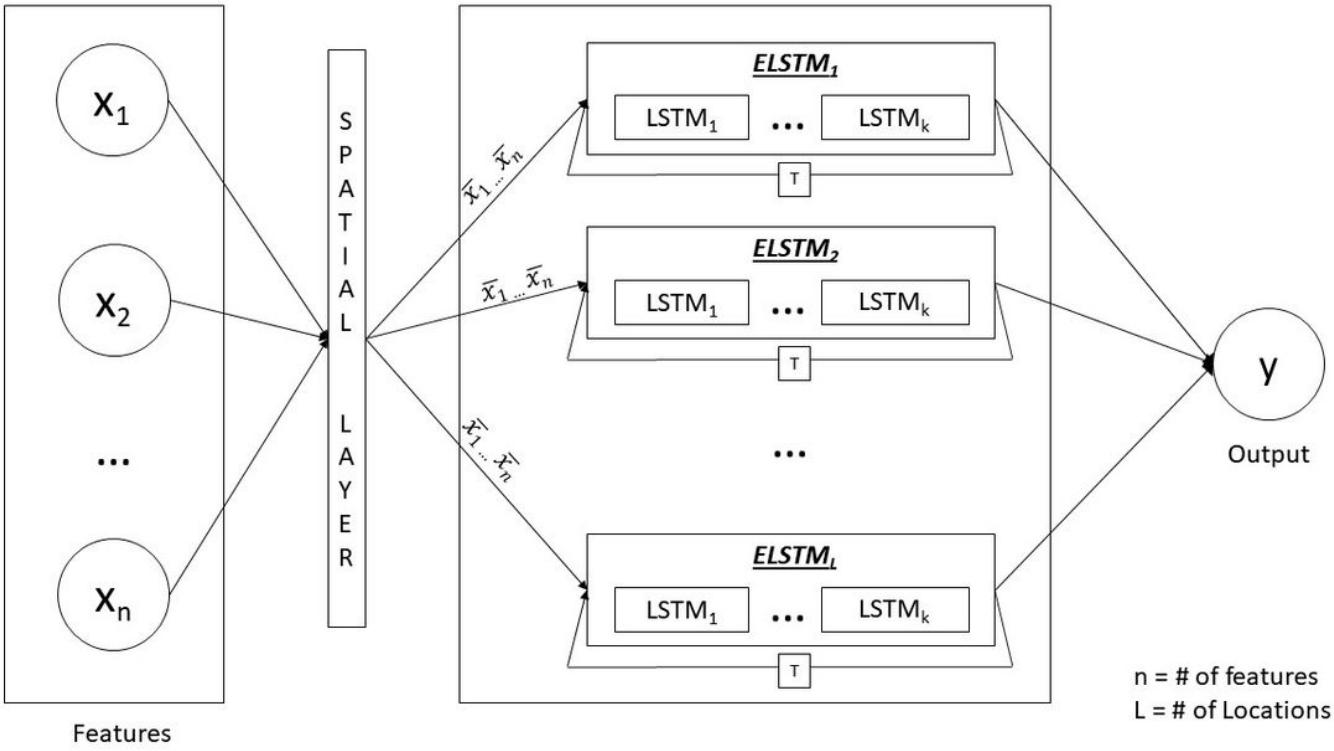


Figure 3

Enhanced Long Short-Term Memory

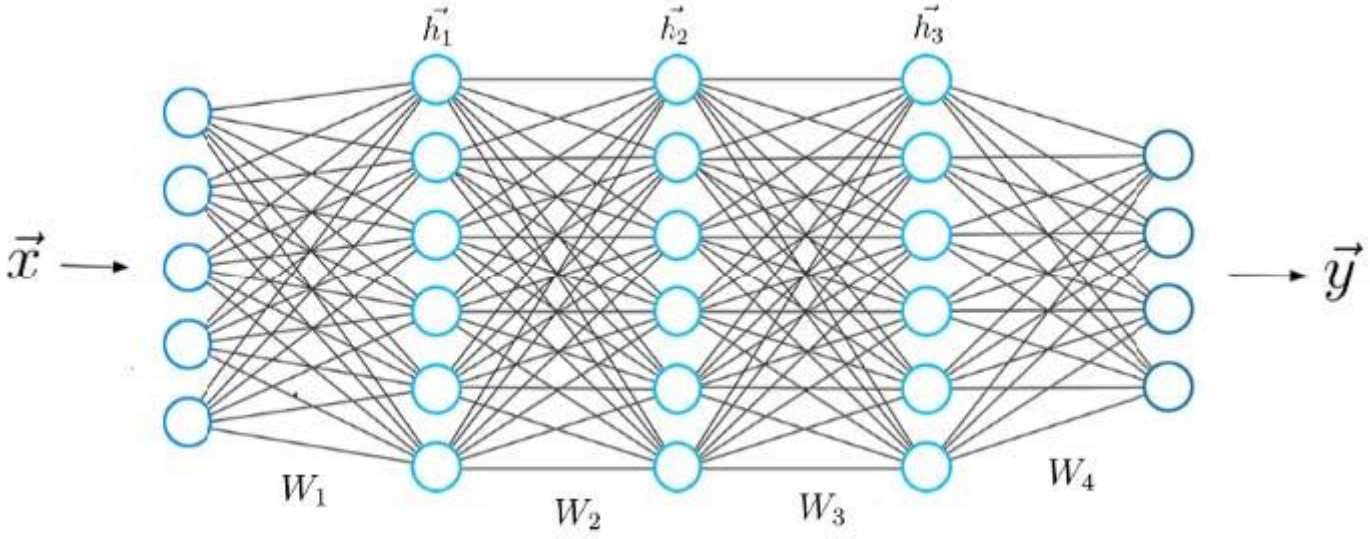


Figure 4

Deep Learning

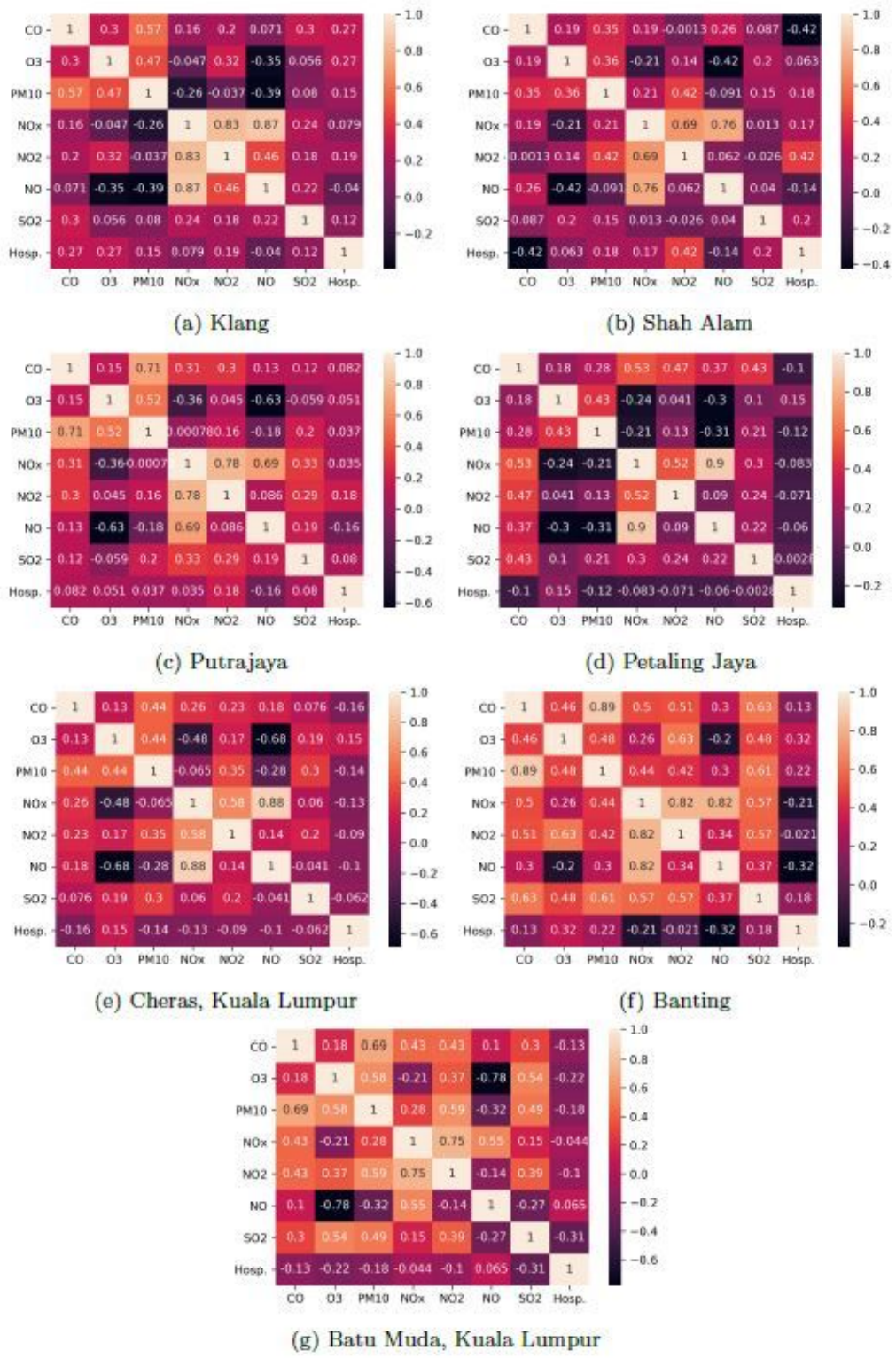


Figure 5

Correlation Matrices: Air quality parameters & Hospitalization

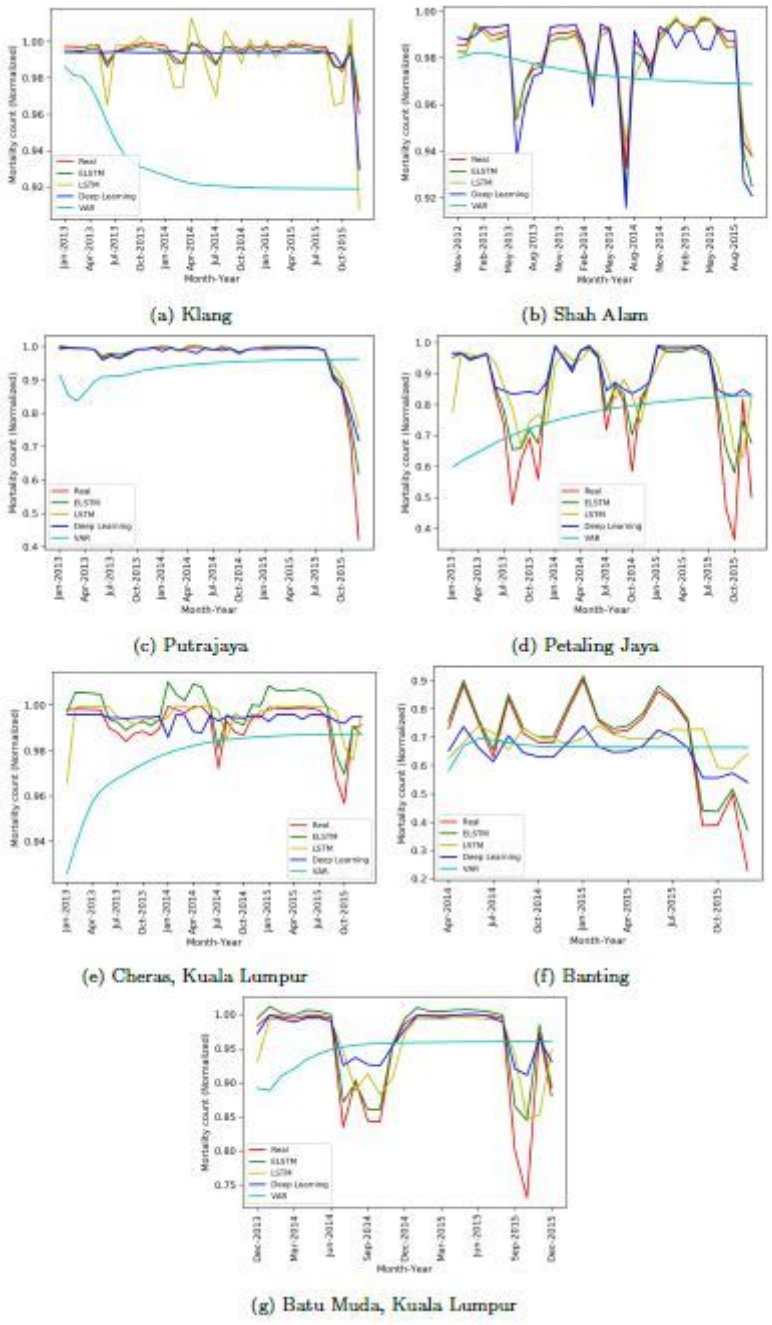


Figure 6

Comparison of hospitalization predictions