



Multi-step short-term $PM_{2.5}$ forecasting for enactment of proactive environmental regulation strategies

Saba Gul · Gul Muhammad Khan ·
Sohail Yousaf

Received: 2 March 2022 / Accepted: 5 April 2022 / Published online: 21 April 2022
© The Author(s), under exclusive licence to Springer Nature Switzerland AG 2022

Abstract Particulate matter is one of the key contributors of air pollution and climate change. Long-term exposure to constituents of air pollutants has exerted serious health implications in both humans and plants leading to a detrimental impact on economy. Among the pollutants contributing to air quality determination, particulate matter has been linked to serious health implications causing pulmonary complications, cardiovascular diseases, growth retardation and ultimately death. In agriculture, crop yield is also negatively impacted by the deposition of particulate matter on stomata of the plant which is alarming and can cause food security concerns. The deleterious impact of air pollutants on human health, agricultural and economic well-being highlights the importance

of quantifying and forecasting particulate matter. Several deterministic and deep learning models have been employed in the recent years to forecast the concentration of particulate matter. Among them, deep learning models have shown promising results when it comes to modeling time series data and forecasting it. We have explored recurrent neural networks with LSTM model which shows potential to predict the particulate matter ($PM_{2.5}$) based on multi-step multi-variate data of two of the most polluted regions of South Asia, Beijing, China and Punjab, Pakistan effectively. The LSTM model is tuned using Bayesian optimization technique to employ the appropriate hyper-parameters and weight initialization strategies based on the dataset. The model was able to predict $PM_{2.5}$ for the next hour with root-mean-square error (RMSE) of 0.1913 (91.5% accuracy) and this error gradually increases with the number of time steps with next 24 hours steps prediction having RMSE of 0.7290. While in case of Punjab dataset with data recorded once a day, the RMSE for the next day forecast is 0.2192. These multi-step short-term forecasts would play a pivotal role in establishing an early warning system based on the air quality index (AQI) calculated and enable the government in enacting policies to contain it.

Gul Muhammad Khan and Sohail Yousaf contributed equally to this work.

S. Gul (✉)
School Of Electrical Engineering and Computer
Science, National University of Science and Technology,
Islamabad, Pakistan
e-mail: sabagul@uetpeshawar.edu.pk

S. Gul · G. M. Khan · S. Yousaf
National Center of Artificial Intelligence, University
of Engineering and Technology, Peshawar, Pakistan
e-mail: gk502@uetpeshawar.edu.pke-mail: syousaf@
uetpeshawar.edu.pk
e-mail: gk502@uetpeshawar.edu.pk

S. Yousaf
e-mail: syousaf@uetpeshawar.edu.pk

Keywords Air pollution · Particulate matter · AQI ·
Forecasting · LSTM · Early warning system

Introduction

Due to their detrimental impact on climate change, human, and plant health, fine dust particles known as particulate matter (PM) in the ambient atmosphere are of significant importance in the research domain. The major source contributing to the production of particulate matter is anthropogenic activities which results in complex mixture of fine particles and water droplets. South Asian countries in particular have the most polluted cities in the world with particulate matter concentrations well above the standards set by WHO ($PM_{2.5}$: 25 $\mu\text{g}/\text{m}^3$ 24-hour mean, PM_{10} : 50 $\mu\text{g}/\text{m}^3$ 24-hour mean).

Particulate matter has been observed to have an adverse effect on human health by causing cardiovascular, pulmonary diseases and increase the risk of lung cancer. Evidence suggests that deleterious health effects are attributed to long-term exposure to combustion-derived nano-particles which augments atherogenesis and causes vascular and acute adverse thrombotic effects (Mills et al., 2009). They have also been observed to impact food safety, since $PM_{2.5}$ acts as carrier for hazardous materials such as heavy metal, the accumulation of which causes organ damage (Noh et al., 2019). The metallic and other chemical elements in $PM_{2.5}$ can be attributed to cause health issues such as pneumonia, asthma, cardiovascular disease, and neurological diseases, the combined effect of which can be fatal (Park, 2021). Moreover, during the pandemic studies were carried out to find association between particulate matter ($PM_{2.5}$ and PM_{10}) with some suggesting that these fine particles act as transportation agent for SARS-CoV-2 in the COVID-19 pandemic (Nor et al., 2021). Negative binomial mixed effect models were employed by (Solimini et al., 2021) globally across 63 countries to observe the statistical correlation of climate, air pollution parameters on COVID-19 cases which were suggestive of a link between them. Thus, the quantification and analysis of trends of air pollutants is of prime importance due to their pivotal role on the human, plant life and economy. Forecasting the air pollution and air quality index (Gul & Khan, 2020) would enable the government and the respective environmental protection agencies in enacting policies to contain the outbreak of airborne diseases and educate the masses about the potential hazards associated with the concentration of a particular pollutant.

To study the trends of pollutants and their impact on different domains, several deterministic and statistical models were explored (Bai et al., 2018; Javadinejad et al., 2021; Ostad-Ali-Askari et al., 2017; Reddy et al., 2017). The conducive nature of meteorological parameters in the forecasting of air pollutants and their rapid change in concentration due to the former presents a challenge in accurately modeling the patterns. Deep learning models in particular were found to be promising due to the ability of recurrent neural networks (RNN) to emulate the trends in time series to predict pollutants.

In this article, we present the recurrent neural network, long short-term memory (LSTM) model tuned by Bayesian optimization strategy to effectively learn the trends and model them. The network is trained on two unique datasets of different terrain belonging to two of the most polluted regions in the world, Beijing China and Panjab, Pakistan. A multi-step multi-variate model of LSTM is introduced which emulates the pattern of particulate matter by effectively encapsulating the historical events and quantifies the air quality in a region. The proposed model would enable the air quality regulatory bodies to take timely decisions to control the emissions, inform the general public about the detrimental health implications and present solutions while monitoring the short-term trend of pollutants.

Literature survey

Worldwide Epidemiological and toxicological studies carried out have suggested a strong association between exposure to particulate matter and impending adverse health effects comprising of pulmonary disease, cardiovascular disease, lung cancer, and premature mortality (Dockery et al., 1994; Lelieveld et al., 2015; Pope et al., 1995).

Evidence from recent studies suggests that the most harmful effects of particulate matter are related to the size of the particle. Its exposure effectiveness level is greatly affected by weather, topography, concentration, and source. As the size of the particle decreases, there is an increase in their acidity and their ability to penetrate the lower pathway of respiratory system (Kim et al., 2015). The short-term impact of fine particles and meteorological extremes on human health was carried out in Seoul, South Korea by observing their correlation with mortality

due to cerebrovascular diseases. It was observed that the impact of particulate matter on human health like asthma, pneumonia, neurological diseases was more pronounced under extreme weather conditions which can often lead to death (Park, 2021). The impact of long-term exposure to $PM_{2.5}$ and its relation to mortality rate was explored in (Wang et al., 2020). The study comprised of 53 million senior Medicare beneficiaries living across America. It was observed that exposure to $PM_{2.5}$ is responsible for causing respiratory, cardiovascular, and certain variants of cancer. The data from the study showed that blacks, younger and urban beneficiaries were most vulnerable to the consequences of long-term exposure to $PM_{2.5}$. Meta-analysis was carried out in (Farhadi et al., 2020) to substantiate the relationship between exposure to $PM_{2.5}$ and myocardial infraction hospitalizations. The investigation found that $PM_{2.5}$ plays a key role in the development of myocardial infractions in humans. Airborne microorganisms are pervasive in the atmosphere and are vital constituent of particulate matter which can lead to wide range of diseases in microorganisms due to their pathogenic nature (Zhai et al., 2018). A database for toxicity score for source specific particulate matter was constructed by (Park et al., 2018) to get an insight of their role in triggering adverse health effects and this information can assist the decision makers take steps to create apposite $PM_{2.5}$ abatement policies. In (Gul & Khan, 2020) an LSTM inspired hazard level prediction system was developed using meteorological and pollutants data of two of the most polluted cities in the world. The hazard level of the next 24 hours of both cities was predicted with an average accuracy of 97%.

Due to the detrimental effect of $PM_{2.5}$ and PM_{10} on human health, the determination of their concentration is of great importance. In (Doreswamy et al., 2020), several machine learning models are evaluated to forecast the pollutant concentration of Taiwan with chronological data of 76 different stations recorded over a span of 5 years. It was observed that the gradient boosting regression algorithm was able to perform better in comparison with other regression models on the TAQMN dataset. The seamless modelling of air quality forecast system would assist the decision makers to improve the quality of air and its associated impact on human health, agriculture, transport, economy, climate, and ecosystems. A novel SVR-based model is introduced in (Hu et al., 2016) and trained on static and

dynamic data of air pollutants concentration in Sydney. In comparison with ANN model, it was observed that the SVR model developed was able to accurately forecast the hourly concentration of air pollutants. Due to dissemination of air pollutants through wind direction and speed, the concentration of $PM_{2.5}$ is strongly correlated with spatiotemporal characteristics. To leverage the spatial and temporal dependency of air pollutant for determination of air quality, weighted long short-term memory neural network extended model (WLSTME) was introduced in (Xiao et al., 2020). It was observed that based on the pollutant and meteorological data of Beijing, Tianjin, and Hebei over the period of 2015 to 2017, the network showed exceptional performance in comparison with STSVR, LSTME, and GWR. A forecasting model using LSTM is introduced in (Han et al., 2018) which uses sensory data of Aerosol Optical Depth (AOD), particulate matter and meteorological conditions. The network was observed to provide effective prediction of the concentrations of harmful gases with 80% $PM_{2.5}$ variability. The system was successfully installed in Beijing, China and these prediction statistics have helped in reducing the air pollution in Beijing by 23%. Due to temporal characteristics of air pollutants, recurrent neural network that is LSTM is employed by (Reddy et al., 2017) to estimate the pollutant concentration for 6 to 10 hours into the future. It was observed that the proposed network was able to predict the pollutant concentration for several future time steps with the same accuracy as forecast for a single future time step of 1 hour which exhibit the predictive robustness of the network. A hybrid deep learning model is proposed in (Du et al., 2021) which employs Bi-LSTM to capture the temporal trends in data and a 1-D CNN to learn the spatial characteristics. The model was able to learn the non-linear co-relationships and model interdependence of the multi-variate temporal data of pollutants and produce effective results in forecasting $PM_{2.5}$ on two real-time datasets from Beijing, China.

LSTM models were observed to capture and learn the non-linear co-relationships of the highly variable data of pollutants more effectively than other deterministic and statistical models. Thus, we propose an LSTM model tuned using optimization strategies to re-structure the network to learn the temporal characteristics of multivariate pollutant data. The model was analyzed and evaluated on two real-world air quality datasets to access its forecasting performance and ability to generalize.

Prediction model framework

Forecasting air pollutants through meteorological and pollutant data requires encapsulation of temporal trends and temporal characteristics are more accurately modelled by recurrent neural networks (RNN) (Reddy et al., 2017). A multi-step multi-variable LSTM model is introduced to capture the sequential trends in the data and effectively forecast particulate matter.

The prediction model framework comprises of an LSTM which is employed due to its ability to retain information over longer sequences (Gul & Khan, 2020; Park, 2021; Reddy et al., 2017). LSTM layer enables the model to capture the temporal trends in the data followed by two dense layers. The LSTM layer is provided with 128 nodes, an activation function, weight initializer and L2 regularizer (Fig. 1). Appropriate weight initialization is selected based on activation function to prevent the issue of vanishing and exploding gradients during back propagation. L2 regularization is used to deal with the over-fitting problem described by Eq. 1. The dense layer comprises of 64 nodes followed by an output layer for forecasting particulate matter $PM_{2.5}$. The output layer can be extended to multiple configurations depending on the sequences of predicted time steps.

$$L(x, y) \equiv \sum_{i=1}^n (y_i - h_{\theta}(x_i))^2 + \lambda \sum_{i=1}^n (\theta_i)^2 \tag{1}$$

Employed datasets

We have evaluated our network performance on two datasets of the most polluted regions in South Asia. The first dataset covers the region of Beijing, China and is available publicly at UCI website (Beijing $PM_{2.5}$ Data) (Liang et al., 2015) with the dataset further extended by (Reddy et al., 2017). The modified Beijing air quality dataset comprises of pollutant data of $PM_{2.5}$ and meteorological parameters such as dew pint, temperature, pressure, cumulative hours of snow, combined wind direction, cumulative wind speed, and cumulative hours of rain. The data is recorded over 35 different stations across the city of Beijing over a span of 7 years from 2010 to 2017 with 43,825 samples (Table 2).

Data cleansing is performed and the aberrant values (High spikes in samples) of the sensors due to some defect or anomaly are removed (Reddy et al., 2017). We have further modified the dataset by adding the columns of AQI and hazard level according the formula (Eq. 2) and standards set by EPA USA to visualize and understand the nature of the data (see Fig. 2). The formula for computing AQI is given by Eq. 2 (Kanchan et al., 2015) with $PM_{2.5}$ being used as a primary pollutant for quantifying air quality.

$$I_p = \frac{I_{Hi} - I_{Lo}}{BP_{Hi} - BP_{Lo}} * (C_p - BP_{Lo}) + I_{Lo} \tag{2}$$

Where,

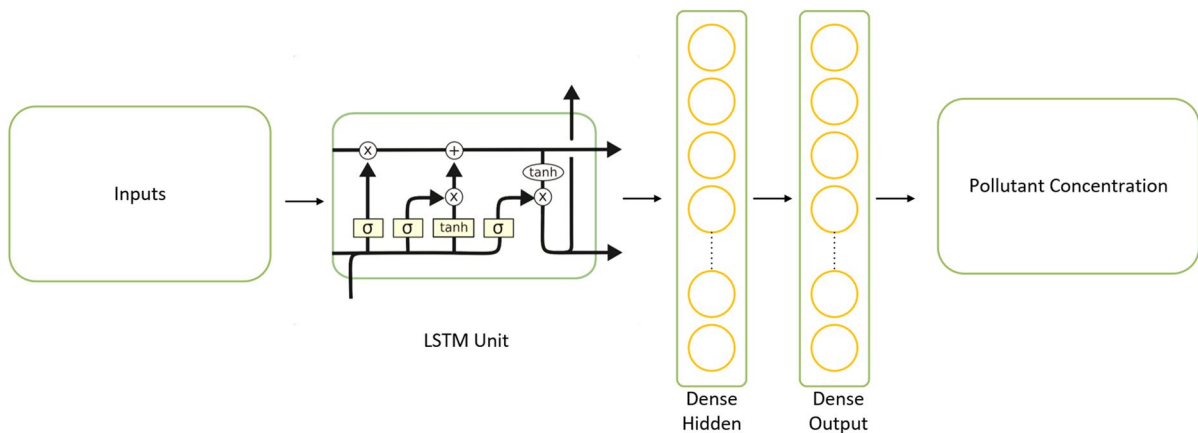


Fig. 1 Proposed Network Architecture

Table 1 Air Quality Index set by environment protection agency, US

O_3 (ppm)	PM_{10} (ug/m ³)	$PM_{2.5}$ (ug/m ³)	CO (ppm)	SO_2 (ppm)	NO_2 (ppm)	AQI Values	Level of Health Concern
0.000–0.059	0–54	0–12	0–4.4	0.000–0.034	-	0–50	Good
0.060–0.075	55–154	12.1–35.4	4.5–9.4	0.035–0.144	-	51–100	Moderate
0.076–0.095	155–254	35.5–55.4	9.5–12.4	0.145–0.224	-	101–150	Unhealthy for sensitive groups
0.096–0.115	255–354	55.5–150.4	12.5–15.4	0.225–0.304	-	151–200	Unhealthy
0.116–0.374	355–424	150.5–250.4	15.5–30.4	0.305–0.604	0.65–1.24	201–300	Very Unhealthy
-	425–504	250.5–350.4	30.5–40.4	0.605–0.804	1.25–1.64	301–400	Hazardous
-	505–604	350.5–500.4	40.5–50.4	0.805–1.004	1.65–2.04	401–500	Hazardous

I_p is index for pollutant p ,
 C_p is the rounded concentration of pollutant p ,
 BP_{Hi} is the breakpoint greater or equal to C_p ,
 BP_{Lo} is the breakpoint less than or equal to C_p ,
 I_{Hi} is the AQI corresponding to BP_{Hi} ,
 And, I_{Lo} is the AQI corresponding to BP_{Lo} .

The hazard levels are classified into seven categories according to the pollutant concentration as depicted in Table 1 with the level of health concern listed.

The t-distributed Stochastic Neighbor Embedding (t-SNE) is used for non-linear dimensionality reduction of the modified Beijing air quality dataset described in Fig. 3 which represents the distribution of the data with respect to hazard levels calculated using Table 1.

The second dataset covers some regions across Punjab with 4 stations across Lahore which is one of the most polluted cities of the world and one station each in Multan and Gujranwala respectively. The dataset comprises of meteorological parameters acquired from Pakistan Meteorological Department and pollutant concentration obtained from Environment Projection Department, Punjab described in Table 3 which is recorded over a span of 3 years over 6 different stations.

Data cleansing is performed by removing rows of data with sensor failure. The columns of AQI are added to the datasets by employing Eq. 2 and health hazard is categorized into seven levels according to the $PM_{2.5}$ concentration in Fig. 2.

The visual description of data distribution of Punjab dataset based on hazard intensity is shown in the t-SNE plot (Fig. 4).

Network hyper-parameters optimization

Hyper parameters of the deep neural network used to forecast particulate matter ($PM_{2.5}$) are optimized to deliver the best performance on a particular dataset with validation set used as a measure for evaluation. The search space is explored to find the optimal set of hyperparameters by using Eq. 3.

$$x^* = \operatorname{argmin}_{x \in X} f(x) \tag{3}$$

Where the score of the objective function that needs to be minimized is represented by $f(x)$; x^* represents the group of hyperparameters with the lowest score and x is any value from the range of X domain.

Several optimization strategies such as grid search, random search and Bayesian optimization

Table 2 Data-set Specifications Of Modified UCI Dataset

Modified Beijing air quality Database Details	
City/Region	Beijing, China
Time Span	2010–2017
Meteorological data	Dew Point, Temperature, Pressure, Combined wind direction, Cumulated wind speed, Cumulated hours of snow, Cumulated hours of rain
Pollutants Data	$PM_{2.5}$ concentration (ug/m ³)
Number of Recording Stations	35

PM 2.5 Concentration	0-12 ug/m ³	12.1-35.4 ug/m ³	35.5-55.4 ug/m ³	55.5-150.4 ug/m ³	150.5-250.4 ug/m ³	250.5-350.4 ug/m ³	350.5-500.5 ug/m ³
AQI Range	0-50	50-100	100-150	150-200	200-300	300-400	400-500

Fig. 2 *PM*_{2.5} Air Quality Index (AQI) Scale, EPA USA

are frequently employed for algorithmic optimization (Bergstra et al., 2011; Bergstra et al., 2013). In grid and random search, all the experiments are mutually exclusive of each other which doesn't help in exploring the search space in an effective manner making it computationally expensive and time consuming. While Bayesian optimization is a sequential-based model optimization algorithm which instead of blindly exploring the search-space makes decisions based on results from previous experiments. Thus, Bayesian optimization

makes use of intuition through historical evidence to narrow down the search space to an optimal set of hyperparameters for selection (Bergstra et al., 2013). The Bayesian search uses Bayes rule (Eq. 4) to utilize the knowledge of previously known priors to direct the search towards combinations of hyperparameter which has higher probability to improve model performance.

$$P(A | B) = \frac{P(B | A) * P(A)}{P(B)} \tag{4}$$

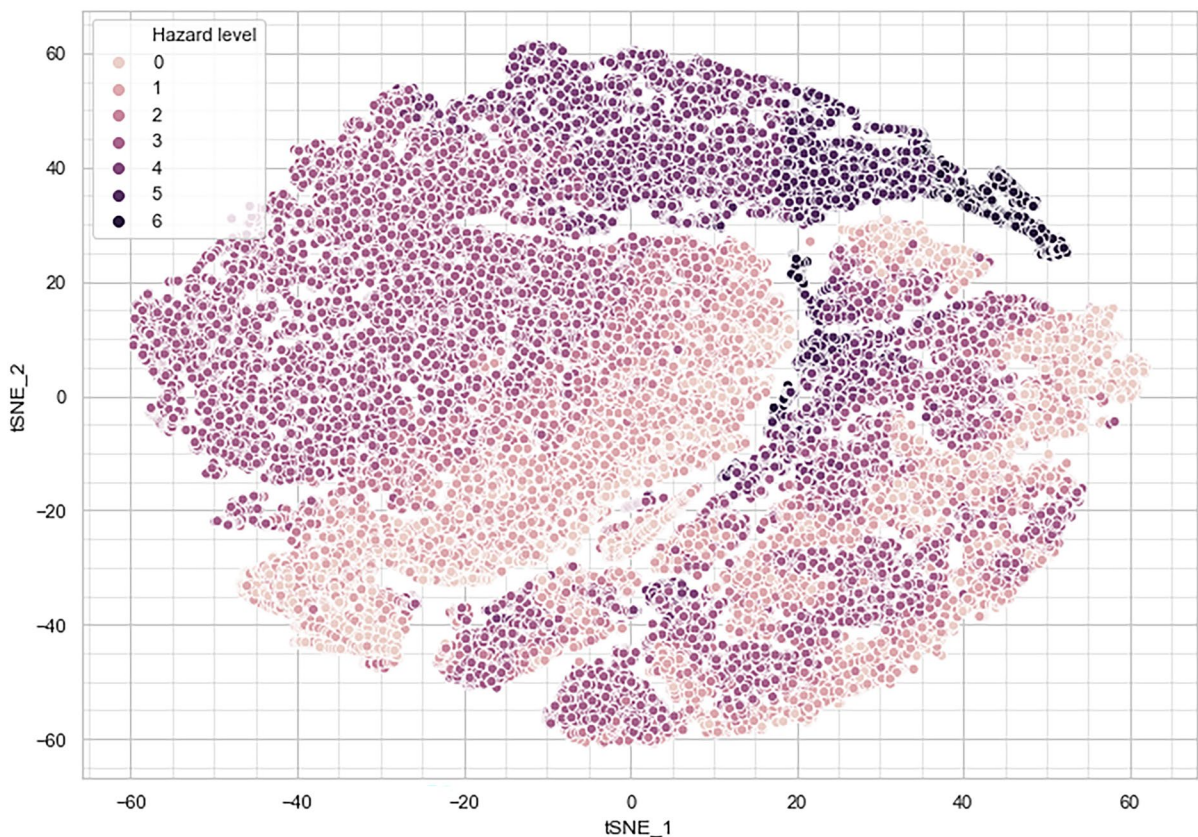


Fig. 3 t-SNE plot of the modified UCI Beijing air quality dataset

Table 3 Data-set Specifications Of Punjab Dataset

Punjab Database Details	
City/Region	Punjab, Pakistan
Time Span	2017–2020
Metrological data	wind direction, temperature, barometric pressure, humidity, and visibility
Pollutants Data	PM_{10} concentration (ug/m^3), $PM_{2.5}$ concentration (ug/m^3), Nitrogen dioxide (ppm), Sulphur dioxide (ppm) and surface ozone (ppm)
Number of Recording Stations	6

Where $P(A | B)$ is the posterior probability, $P(B | A)$ is the likelihood, $P(A)$ is the class prior probability and $P(B)$ is the predictor prior probability. Expected improvement (EI) is employed as an acquisition function which defines the criteria for selection of hyperparameters from the Gaussian process model which is used as a surrogate function for tuning (Eq. 5).

$$EI_{y^*}(x) = \int_{-\infty}^{y^*} (y^* - y)p(y | x)dy \tag{5}$$

Here $p(y | x)$ is the Gaussian surrogate probability model, x describes the hyperparameter, y is the true objective function score and y^* is the latest minimum score observed so far of the true objective function.

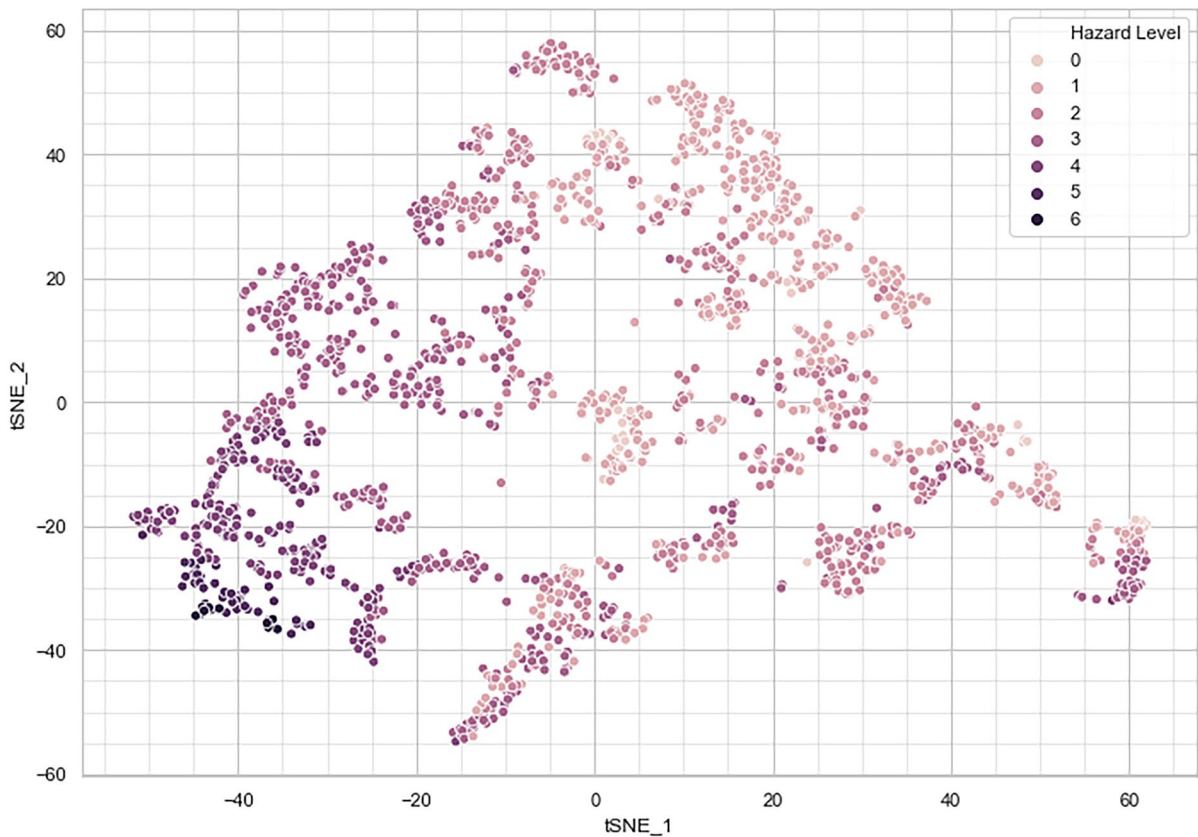


Fig. 4 t-SNE plot of Punjab dataset

To find the optimal hyperparameters under the surrogate model $P(y | x)$, the expected improvement should be maximized with respect to x .

Results and analysis

To capture the temporal characteristics of the air pollution parameters LSTM, a recurrent neural network is employed. The features from the dataset are pre-processed by normalizing the data with anomalous entries of sensor failure removed. The data is then rearranged in the form of cyclic packets with previous value of $PM_{2.5}$ feed in addition to the meteorological data to predict the next time stamp. We start with a single step prediction of 1 hour and extend it gradually to 24 hour steps while increasing the number of output steps which gives us an insight about the networks performance on multi-step data.

After selection of appropriate network, the hyperparameters such as learning rate, optimizer and activation function are tuned using Bayesian optimization. Expected improvement is used as a performance

metric by Bayesian optimization to evaluate the selection of a suitable hyperparameter. The network performance on each hyperparameter from the search space is gauged by the validation set which assists in guiding the search in an appropriate direction as depicted in Fig. 5.

Robust initialization techniques are proposed in this section that removes the obstacles of training deep neural networks by solving the problem of vanishing and exploding gradients. Thus, weight initialization practices like Lecun, Glorot and He initialization are employed to effectively train the proposed network.

The network was optimized to improve the performance by tuning the best activation function and optimizer for each dataset followed by learning rate. The results of tuned hyperparameters based on their performance on validation set for modified Beijing air quality dataset are described in Tables 4 and 5. MSE is used as a metric by the recurrent network during training and validation to gauge the performance of the network in the search space to find an optimal hyper parameter.

After selection of optimal hyperparameters, the network is reconstructed and trained with a boost in

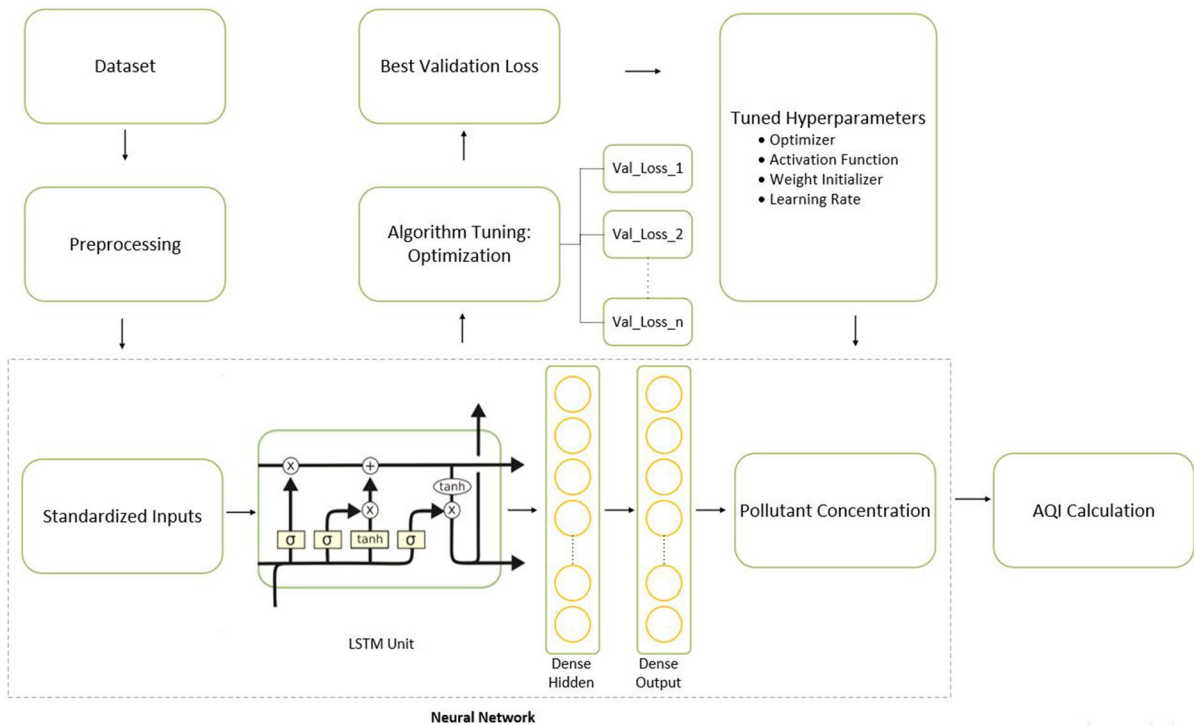


Fig. 5 Hyper-parameter tuning using Bayesian optimization

Table 4 Selection of Optimizer and Activation using Bayesian optimization on modified Beijing air quality dataset

Optimizer	Activation	MSE
<i>Adagrad</i>	<i>tanh</i>	<i>0.041617</i>
Adagrad	linear	0.0416348
Adagrad	linear	0.0416676
Adagrad	tanh	0.0419074
Adadelta	relu	0.0419169
Adagrad	linear	0.0419371
Adagrad	softsign	0.0420134
Adagrad	tanh	0.04209329

The bold and italic emphasis is used to highlight the best parameter in each table

performance observed. For weight initialization He/ Kaiming Initialization is selected (He et al., 2015) based on the nature of the activation function. Since Relu is a non-differentiable function, Kaiming He (He et al., 2015) proposed a weight initialization scheme that was tailored for deep neural networks that employ asymmetric and non-linear activation functions. The He normal initialization method is calculated as a random number with a normal probability distribution (U) having a mean of 0.0 and a standard deviation of $(\sqrt{\frac{2}{f_{in}}})$, where f_{in} describes the number of inputs to the node. While the He uniform comprises of weight samples taken from a uniform distribution (U) between the range $(-\sqrt{\frac{6}{fan_{in}}})$ and $(\sqrt{\frac{6}{fan_{in}}})$, where fan_{in} defines the number of input nodes.

$$weight = U\left(-\sqrt{\frac{6}{f_{in}}}, \sqrt{\frac{6}{f_{in}}}\right) \tag{6}$$

Table 5 Optimal learning rate selection using Bayesian optimization on modified Beijing air quality dataset

Learning rate	MSE
<i>0.03</i>	<i>0.0411263</i>
0.013885	0.0413445
0.00665	0.0415720
0.01271	0.0416139
0.009799	0.0416274
0.005592	0.0417381
0.011722	0.0418148
0.003004	0.0418748

The bold and italic emphasis is used to highlight the best parameter in each table

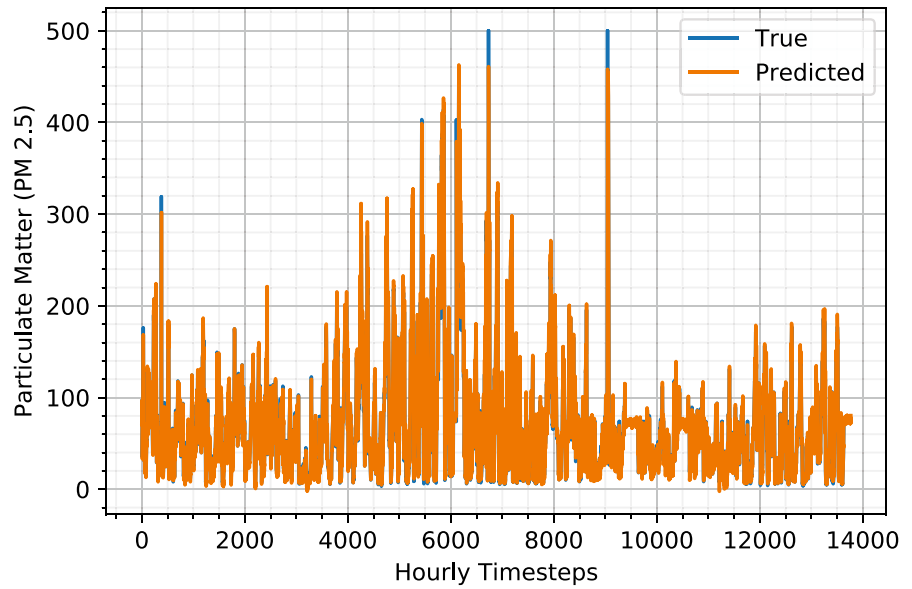
Upon employing He, Glorot and Lecun weight initializers to the network, it was observed that He uniform initialization was able to help the network in achieving appropriate weights for learning. Figures 6 and 7 describes the training and validation performance of the network with ability to forecast $PM_{2.5}$ level for the next hour and next day with RMSE of 0.1913 and 0.6341 respectively on the test-set.

Forecasting air pollutants can helps in identifying the long-term trends, impact of exposure and can enable the relevant bodies to devise strategies to contain the exponential growth of the pollutants and warn the sensitive groups. To cater for studying long-term patterns of particulate matter, the model is further modified to predict future time steps. It can be observed from Table 6 that the evaluation metrics used to gauge the performance degrades as the number of future hours increases. The root-mean-square error (RMSE) of the model increases gradually while the variance score R^2 drops with prediction of additional steps into the future. The gradual degradation in performance is due to increase in number of future steps to predict while keeping the past steps constant which shows the robustness of the proposed LSTM model when it comes to multi-step prediction of long sequences.

The deep air model proposed by (Reddy et al., 2017) was used to predict the particulate matter ($PM_{2.5}$) of Beijing, China and the model was trained and evaluated on the modified UCI Beijing air quality dataset. On comparison with our tuned model, we observed that for the same dataset, we were able to predict the particulate matter at single and multi-step more effectively as described in Table 7. The low root-mean-square error (RMSE) and high variance score in Table 7 shows the robustness of our model in comparison with (Reddy et al., 2017).

The network is then re-tuned for Punjab dataset by observing the trend of MSE of the validation set. Bayesian optimization is used to intuitively go through the search space to find the learning rate, activation function and optimizer. Since Tanh is selected as an optimized activation function which is non-linear in nature, the best practice of weight initialization used to prevent vanishing gradients is Xavier or Glorot initialization as proposed by (Glorot & Bengio, 2010).The Glorot uniform initialization is calculated as a random number with a uniform probability distribution (U) between the range

Fig. 6 Actual Vs. Predicted $PM_{2.5}$ values of employed architecture on Hourly data of modified UCI Beijing air quality dataset



$-\left(\sqrt{\frac{6}{fan_{in}+fan_{out}}}\right)$ and $\left(\sqrt{\frac{6}{fan_{in}+fan_{out}}}\right)$, where fan_{in} defines the number of input nodes and fan_{out} is the number of output nodes in the weight tensor.

$$weight = U \left[-\left(\sqrt{\frac{6}{fan_{in} + fan_{out}}}\right), \left(\sqrt{\frac{6}{fan_{in} + fan_{out}}}\right) \right] \tag{7}$$

While in-case of Glorot normal distribution, samples are drawn from truncated normal distribution which

is centered on zero and has standard deviation of $\left(\sqrt{\frac{2}{fan_{in}+fan_{out}}}\right)$.

The Lecun uniform initialization is calculated by drawing random samples from a uniform probability distribution (U) between the range $-\left(\sqrt{\frac{3}{fan_{in}}}\right)$ and $\left(\sqrt{\frac{3}{fan_{in}}}\right)$, where fan_{in} defines the number of input nodes and fan_{out} is the number of output nodes in the weight tensor.

Fig. 7 Actual Vs. Predicted $PM_{2.5}$ values of employed architecture on 24 hour data of modified UCI Beijing air quality dataset

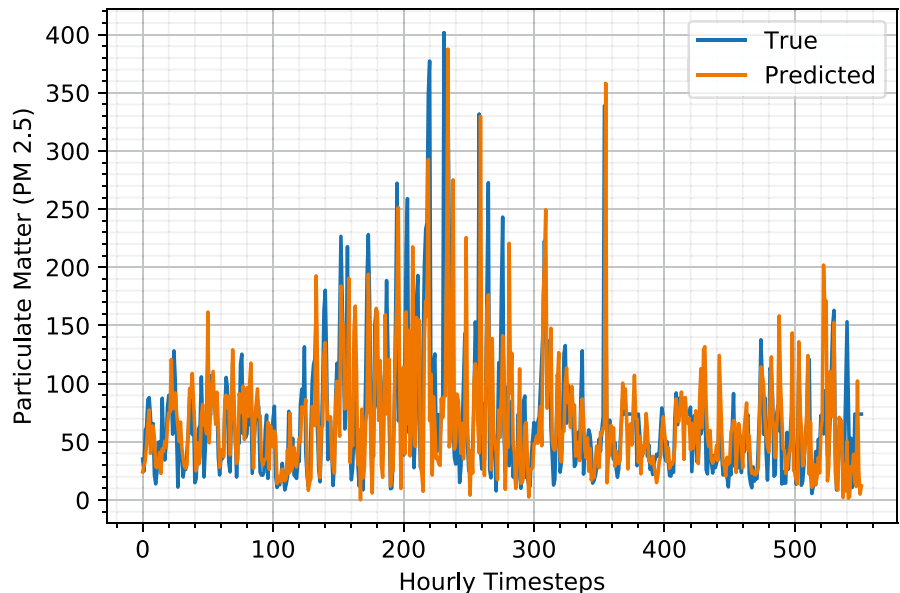


Table 6 Test RMSE for multi-step prediction of $PM_{2.5}$ on modified Beijing air quality dataset

Model	RMSE	R^2
3 future hours	0.2598	0.932526
6 future hours	0.3475	0.879228
9 future hours	0.4096	0.832258
12 future hours	0.5578	0.688855
15 future hours	0.5195	0.730129
18 future hours	0.5020	0.748015
21 future hours	0.5408	0.707541
24 future hours	0.7290	0.468538

$$weight = U \left[- \left(\sqrt{\frac{3}{fan_{in}}} \right), \left(\sqrt{\frac{3}{fan_{in}}} \right) \right] \tag{8}$$

While in-case of Lecun normal distribution, samples are drawn from truncated normal distribution centered around zero and has standard deviation of $(\sqrt{\frac{1}{fan_{in}}})$. Though according to (Glorot & Bengio, 2010), Glorot initialization performs better for Relu activation, but according to our validation set performance based on MSE, Lecun normal initialization was able to achieve superior performance.

Tables 8 and 9 describes the results of tuned hyper-parameters. These parameters are employed to retrain the network with a significant improvement in performance observed. The performance of the network on the training and validation set is described in Fig. 8. The network can forecast the $PM_{2.5}$ with RMSE of 0.2192 respectively on the test set for 24 hour future time step. The reason for the higher RMSE in-case of Panjab dataset can be attributed to the data employed which is recorded daily due to which it shows high variance making it difficult for the network to capture the trends as compared to modified UCI dataset where parameters are recorded hourly.

Table 7 Comparison of test RMSE with Deep Air for various future time lags on modified Beijing air quality dataset

Model	Time Steps	RMSE	R^2
Our Model	1 future hour/single step	0.1913	0.963402
Deep Air (Reddy et al., 2017)	1 future hour/single step	12.78	0.96
Our Model	5 future hours	0.3229	0.895758
Deep Air (Reddy et al., 2017)	5 future hours	44.15	0.689
Our Model	10 future hours	0.4086	0.833078
Deep Air (Reddy et al., 2017)	10 future hours	74.8	0.588

Table 8 Selection of Optimizer and Activation using Bayesian optimization on Punjab dataset

Optimizer	Activation	MSE
<i>Nadam</i>	<i>relu</i>	<i>0.6396949</i>
Nadam	relu	0.7275196
Nadam	relu	0.7335685
Nadam	relu	0.7343104
Nadam	relu	0.7576756
Nadam	relu	0.7613520
Nadam	relu	0.7674339
Nadam	relu	0.7830696

The bold and italic emphasis is used to highlight the best parameter in each table

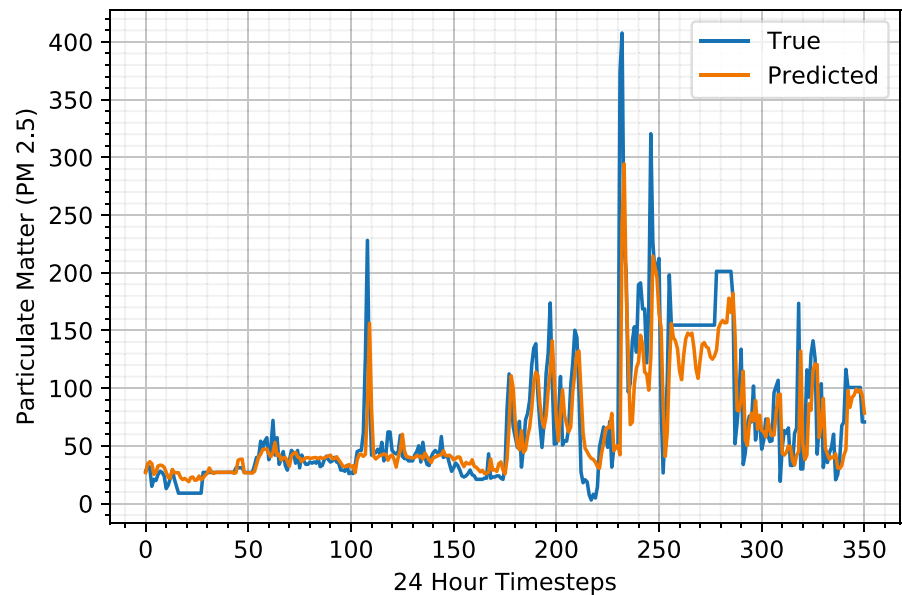
Table 9 Optimal learning rate selection using Bayesian optimization on Punjab dataset

Learning rate	MSE
<i>0.0199939</i>	<i>0.5433736</i>
0.0032243	0.6428966
0.0035615	0.6451763
0.0145944	0.6503993
0.0046863	0.6523746
0.0013331	0.6587814
0.0013885	0.6672297
0.0129494	0.6792039

The bold and italic emphasis is used to highlight the best parameter in each table

Since, the data of pollutants is highly variable and changes rapidly with meteorological parameters, thus the forecasting capability of the model drops gradually for prediction of extended future time steps as observed. From Tables 4 to 9, the performance of the model shows an ability to generalize well and emulate the trend of particulate matter $PM_{2.5}$ for multi-step multivariate data of different regions.

Fig. 8 Actual Vs. Predicted $PM_{2.5}$ values of employed architecture on 24 hour data of Punjab dataset



Conclusion

A deep learning model was proposed for quantification of hazard level by predicting the particulate matter concentration and evaluated on two of the most polluted regions in the world: Beijing, China and Punjab, Pakistan. Hyper-parameter tuning and weight initialization strategies were adopted to tune the network by exploring the search space effectively. The tuned LSTM model was used to learn and effectively model the temporal trends of the particulate matter and meteorological data to predict concentration of future instances of $PM_{2.5}$. Hourly concentration of $PM_{2.5}$ of Beijing was predicted with an RMSE of 0.1913 and based on the average 24-hour data, the RMSE drops to 0.6341 with state of the art performance observed. While the 24-hour $PM_{2.5}$ prediction of Punjab has an RMSE of 0.2192, this degradation in performances of the model can be attributed to drastic variance in the recorded data over a span of 24 hours. By feeding the historic data in hourly time stamps, the degradation in performance was observed to be subtle. The forecasting model using LSTM helps in mapping the AQI level and identifying the health concerns associated with it. This would enable the general public, government and environmental protection agencies to quantify the risk associated with air quality index and enable the authorities to take effective measures to minimize the consequences and assist the environment

protection agencies to enact policies towards reducing the health and economic risk associated with high concentration of particulate matter.

Author contributions Material preparation, data collection, analysis and first draft were written by Saba Gul. Supervision, writing—review and editing were done by Dr Gul Muhammad Khan and Dr. Shoail Yousaf

Funding Not Applicable.

Data availability The modified UCI Beijing dataset analyzed in this study is available at; <https://github.com/vikmreddy/deep-air>, while the Panjab dataset is available from the corresponding author on reasonable request.

The datasets during and/or analyzed during the current study available from the corresponding author on reasonable request.

Code availability The code repository generated during the study can be obtained from the corresponding author on reasonable request.

Declarations

Ethics approval Not Applicable.

Consent to participate Not Applicable.

Consent for publication All authors agreed with the content and gave explicit consent to submit and they have obtained consent from the responsible authorities at the institute/organization where the work was carried out.

Conflicts of interest The authors have no competing interests to declare that are relevant to the content of this article.

References

Bai, L., Wang, J., Ma, X., & Lu, H. (2018). Air pollution forecasts: An overview. *International Journal of Environmental Research and Public Health*, 15. <https://doi.org/10.3390/ijerph15040780>. Retrieved from <https://www.mdpi.com/1660-4601/15/4/780>

Bergstra, J., Bardenet, R., Bengio, Y., & Kégl, B. (2011). Algorithms for hyper-parameter optimization. In *Advances in Neural Information Processing Systems 24: 25th Annual Conference on Neural Information Processing Systems 2011, NIPS 2011* (pp. 1–9)

Bergstra, J., Yamins, D., & Cox, D. D. (2013). Making a science of model search: Hyperparameter optimization in hundreds of dimensions for vision architectures. In *Proceedings of the 30th International Conference on International Conference on Machine Learning - Volume 28*. (pp. I–115–I–123). [JMLR.org](https://doi.org/10.1145/2485855)

Dockery, D., Pope, C., Xu, X., Spengler, J., Ware, J., Fay, M., et al. (1994). An association between air pollution and mortality in six U.S. cities. *The New England Journal of Medicine*, 329, 1753–1759. <https://doi.org/10.1056/NEJM199312093292401>

Doreswamy, S. H. K., Yogesh, K. M., & Gad, I. (2020). Forecasting air pollution particulate matter (pm2.5) using machine learning regression models. *Procedia Computer Science*, 171, 2057–2066. Retrieved from <https://doi.org/10.1016/j.procs.2020.04.221>. <https://www.sciencedirect.com/science/article/pii/S1877050920312060>

Du, S., Li, T., Yang, Y., & Horng, S. J. (2021). Deep air quality forecasting using hybrid deep learning framework. *IEEE Transactions on Knowledge and Data Engineering*, 33, 2412–2424. <https://doi.org/10.1109/TKDE.2019.2954510>

Farhadi, Z., Gorgi, H. A., Shabaninejad, H., Delavar, M. A., & Torani, S. (2020). Association between pm2.5 and risk of hospitalization for myocardial infarction: a systematic review and a meta-analysis. *BMC Public Health*, 20, 314. Retrieved from <https://doi.org/10.1186/s12889-020-8262-3>

Glorot, X., & Bengio, Y. (2010). Understanding the difficulty of training deep feedforward neural networks. *Journal of Machine Learning Research - Proceedings Track*, 9, 249–256.

Gul, S., & Khan, G. M. (2020). Forecasting hazard level of air pollutants using LSTM’s. In *Artificial Intelligence Applications and Innovations* (pp. 143–153). Springer International Publishing. https://doi.org/10.1007/978-3-030-49186-4_13

Han, Y., Lam, J. C. K., & Li, V. O. K. (2018). A Bayesian LSTM model to evaluate the effects of air pollution control regulations in China. In *2018 IEEE International Conference on Big Data (Big Data)* (pp. 4465–4468). <https://doi.org/10.1109/BigData.2018.8622417>

He, K., Zhang, X., Ren, S., & Sun, J. (2015). Delving deep into rectifiers: Surpassing human-level performance on imagenet classification. In *2015 IEEE International*

Conference on Computer Vision (ICCV) (pp. 1026–1034). <https://doi.org/10.1109/ICCV.2015.123>

Hu, K., Sivaraman, V., Bhrugubanda, H., Kang, S., & Rahman, A. (2016). SVR based dense air pollution estimation model using static and wireless sensor network. In *2016 IEEE Sensors* (pp. 1–3). <https://doi.org/10.1109/ICSENS.2016.7808827>

Javadinejad, S., Eslamian, S., & Ostad-Ali-Askari, K. (2021). The analysis of the most important climatic parameters affecting performance of crop variability in a changing climate. *International Journal of Hydrology Science and Technology*, 11, 1–25. <https://doi.org/10.1504/IJHST.2021.112651>

Kanchan, K., Gorai, A., & Goyal, P. (2015). A review on air quality indexing system. *Asian Journal of Atmospheric Environment*, 9, 101–113. <https://doi.org/10.5572/ajae.2015.9.2.101>

Kim, K. H., Kabir, E., & Kabir, S. (2015). A review on the human health impact of airborne particulate matter. *Environment International*, 74, 136–143. <https://doi.org/10.1016/j.envint.2014.10.005>. Retrieved from <https://www.sciencedirect.com/science/article/pii/S0160412014002992>

Lelieveld, J., Evans, J. S., Fnais, M., Giannadaki, D., & Pozzer, A. (2015). The contribution of outdoor air pollution sources to premature mortality on a global scale. *Nature*, 525, 367–371. Retrieved from <https://doi.org/10.1038/nature15371>

Liang, X., Zou, T., Guo, B., Li, S., Zhang, H., Zhang, S., Huang, H., & Chen S. X. (2015). Assessing Beijing’s pm2.5 pollution: severity, weather impact, APEC and winter heating. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 471(2182). <https://doi.org/10.1098/rspa.2015.0257>

Mills, N. L., Donaldson, K., Hadoke, P. W., Boon, N. A., MacNee, W., Cassee, F. R., ... & Newby, D. E. (2009). Adverse cardiovascular effects of air pollution. *Nature Clinical Practice Cardiovascular Medicine*, 6, 36–44. <https://doi.org/10.1038/ncpcardio1399>

Noh, K., Thi, L. T., & Jeong, B. R. (2019). Particulate matter in the cultivation area may contaminate leafy vegetables with heavy metals above safe levels in Korea. *Environmental Science and Pollution Research*, 26, 25762–25774. Retrieved from <https://doi.org/10.1007/s11356-019-05825-4>

Nor, N. S. M., Yip, C. W., Ibrahim, N., Jaafar, M. H., Rashid, Z. Z., Mustafa, N., ... & Nadzir, M. S. M. (2021). Particulate matter (pm2.5) as a potential SARS-COV-2 carrier. *Scientific Reports*, 11, 2508. Retrieved from <https://doi.org/10.1038/s41598-021-81935-9>

Ostad-Ali-Askari, K., Shayannejad, M., & Ghorbanzadeh-Kharazi, H. (2017). Artificial neural network for modeling nitrate pollution of groundwater in marginal area of Zayandeh-Rood River, Ifahan, Iran. *KSCE Journal of Civil Engineering*, 21, 134–140. <https://doi.org/10.1007/S12205-016-0572-8>

Park, M., Joo, H. S., Lee, K., Jang, M., Kim, S. D., Kim, I., ... & Park, K. (2018). Differential toxicities of fine particulate matters from various sources. *Scientific Reports*, 8, 17007. Retrieved from <https://doi.org/10.1038/s41598-018-35398-0>

Park, S. K. (2021). Seasonal variations of fine particulate matter and mortality rate in Seoul, Korea with a focus on the short-term

- impact of meteorological extremes on human health. *Atmosphere*, 12. <https://doi.org/10.3390/atmos12020151>. Retrieved from <https://www.mdpi.com/2073-4433/12/2/151>
- Pope, C., Bates, D. V., & Raizenne, M. E. (1995). Health-effects of particulate air-pollution - time for reassessment. *Environmental Health Perspectives*, 103, 472–480. <https://doi.org/10.2307/3432586>
- Reddy, V., Yedavalli, P., Mohanty, S., & Nakhat, U. (2017). Deep air: forecasting air pollution in Beijing, China. *Environmental Science*
- Solimini, A., Filippini, F., Fegatelli, D. A., Caputo, B., Marco, C. M. D., Spagnoli, A., & Vestri, A. R. (2021). A global association between covid-19 cases and airborne particulate matter at regional level. *Scientific Reports*, 11, 6256. Retrieved from <https://doi.org/10.1038/s41598-021-85751-z>
- Wang, B., Eum, K. D., Kazemiparkouhi, F., Li, C., Manjourides, J., Pavlu, V., & Suh, H. (2020). The impact of long-term pm2.5 exposure on specific causes of death: exposure-response curves and effect modification among 53 million U.S. Medicare beneficiaries. *Environmental Health*, 19, 20. Retrieved from <https://doi.org/10.1186/s12940-020-00575-0>
- Xiao, F., Yang, M., Fan, H., Fan, G., Al-qaness, M. A. A. (2020). An improved deep learning model for predicting daily pm2.5 concentration. *Scientific Reports*, 10, 20988. Retrieved from <https://doi.org/10.1038/s41598-020-77757-w>
- Zhai, Y., Li, X., Wang, T., Wang, B., Li, C., & Zeng, G. (2018). A review on airborne microorganisms in particulate matters: Composition, characteristics and influence factors. *Environment International*, 113, 74–90. <https://doi.org/10.1016/j.envint.2018.01.007>. Retrieved from <https://www.sciencedirect.com/science/article/pii/S016041201732055X>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.