Air Pollution Forecasting Using Deep Learning

Manal Alghieth(✉), Raghad Alawaji, Safaa Husam Saleh, Seham Alharbi
Department of Information Technology, College of Computer,
Qassim University, Buraydah, Saudi Arabia
mgietha@qu.edu.sa

Abstract—Nowadays, air pollution is getting an extreme problem that affects the whole environment. Due to the dangerous effects of air pollution on human's health, this study proposes an air pollution prediction system. Because of the high dust pollution in Saudi Arabia, and the fact that there is no system for predicting the percentage of air pollution in it, this study applies an air pollution prediction system to the most affected area in Saudi Arabia. This paper aims to forecast the concentrations of PM10 particles due to their dangerous effects. This study aims to align with the Saudi vision 2030 by having an ideal environment and act in an efficient way in case of a warning situation. It applies a deep learning technique, which called Long Short-Term Memory (LSTM) to predict the air pollution in Saudi Arabia and achieved exceptional results due to the low error rates that have been obtained by this study. The error rate of Mean Absolute Error (MAE) is 0.98, for Root Mean Square Error (RMSE) is 8.68 and 0.999 for R-Squared.

Keywords—air pollution, forecasting, deep learning, LSTM

1   Introduction

Air pollution is considered an environmental issue that affects human's health, whether in the long or short term. Around 7 million deaths per year are due to air pollution exposure and other atmospheric natural disasters [1]. The kingdom of Saudi Arabia is known for its desert environment. In a case when the speed of the wind increase, it would cause the dust to be raised. For instance, in certain seasons in Riyadh that is located in Saudi Arabia, the dust becomes a severe problem.

Based on the reports and research that have been done by the Saudi Arabia General Authority for Meteorology and Environmental Protection [2], it showed the main dangerous pollutants that can be found in Saudi Arabia and have the ability to make the quality of air in low levels. These pollutants are carbon monoxide (CO), nitrogen dioxide (NO2), ground ozone (O3), sulfur dioxide (SO2), and dust particles with its two types: (PM10 and PM2.5). The high concentrations of these pollutants can be harmful.

Due to the pollutants that the dust contains, it is considered one of the air pollutants. PM10 and PM2.5 are pollutants that can be found in the dust. The danger of these pollutants that it is fatal to humans. The proposed system aims to forecast the concentrations of PM10, so it can give and warn the people for the levels of air pollution.

Throughout history, there were so many studies that used several techniques to predict the levels of air pollution. Those studies are continues until the present time. The most used techniques were as follows:

- Statistical Methods.
- Machine Learning (ML).
- Deep Learning (DL).

In the current term, the most trending technique is DL, which is meant by simulating the human brain to make patterns that can be used in making a decision.

The proposed system will use one of the Deep Learning algorithms, which is Long Short Term Memory (LSTM). LSTM is a sort of Recurrent Neural Network (RNN).

This study applies LSTM due to its advantages. One of the reasons for choosing LSTM is LSTM's memory, and by that LSTM is capable of keeping relevant and essential information to be used in the prediction [3]. LSTM has the ability of predicting the long- and short-term time, which can be used in the prediction problems [3]. LSTM used for forecasting and classifying the data that are in the form of a time-series format.

The rest of this paper is organized as follows:

- Section 2 will outline the related work with different types of previous studies in separate sections.
- Section 3 will identify the used methodology in this research, and discuss the data collection and analysis steps.
- Section 4 will discuss the process of forecasting PM10 concentrations using the suggested model in ordered steps.
- Section 5 will outline the promising results obtained from the model, describing the difference between this study and other studies in the same field.
- Section 6 will discuss some important topics regarding this paper.
- Section 7 will show the conclusion of this paper.
- Section 8 will show the references.
- The last section, which is section 9, shows some information about the authors.

2 Related work

Many countries worldwide have developed a warning system for monitoring and forecasting air pollution data. The General Authority for Meteorology and Environmental Protection in Saudi Arabia has developed a monitoring network consisting of 79 stations that provide hourly measurements of air pollutants [2]. However, currently there is no system for forecasting air pollution in Saudi Arabia.

2.1 Machine learning approach

ML is one of the Artificial Intelligence (AI) fields. As the name suggests, ML made the computers able to learn and act just like a human. These computers were provided with information to process them and learn by themselves.

In recent years, there are many research papers focused on traditional methods to predict air pollution. Most of these studies depend on statistical equations such as linear regression or classification. There was a proposed model based on stepwise Multiple Linear Regression (MLR) to predict concentrations of (NOx) and PM10 [4]. However, with an ever-increasing air pollution ratio, and the factors that affect the forecasting process, using a regression-based model with nonlinear characteristics leads to a complex system. For that, it is important to develop a model with both high accuracy and high computational efficiency. In [5], a proposed model was used to predict ozone in the USA by Bayesian inference framework. The method combined observational air monitoring data with a numerical model output forecast to create a statistical model that could give accurate forecast maps for the current 8-hours average and the largest 8-hours average ozone concentration levels on the next day. Three types of classifiers have been selected for estimating the ozone concentration level in Valencia, Spain [6]. The performance of C4.5 classification algorithm was 73.65%, for Back-propagation neural network was 82.05%, and 83.14% for the σ–FLNMAP classifier. It concluded that σ–FLNMAP is simple and fast since the rules extracted from the model are only three rules (low, mid, and high) and represented as hyperboxes. However, the mechanism by which regression tasks are transformed into classification tasks is problematic, since it is time-consuming and produces inaccurate data.

To improve the performance of prediction of air pollution, several types of Artificial Neural Networks (ANNs) have been successfully developed. In [7], showed that ANN used the L-M BP algorithm to predict PM2.5 that has better performance than that obtained by Multivariate Methods (MVM) and the correlation coefficient using trained neural network was increased from 0.58 to 0.80. Moreover, in [8], it showed that Multi-layer Perceptron (MLP) model for early warning system in China suffer from different drawbacks such as local minima, overfitting and poor generalization. Alternative model was suggested to overcome this drawback and it was concluded with that the Least Squares Support Vector Machine (LS-SVM) produced low error while the error of MLP was triple. Furthermore, four air pollution indicator levels: carbon monoxide, sulfur dioxide, nitrogen oxides and Ozone were predicted by Adaptive Neuro-Fuzzy Inference System (ANFIS). The results of the mean absolute error of the ANFIS was less than 15% [9]. A similar approach used in [10], a study that was made to compare between ANNs and ANFIS to forecast PM2.5 for the next coming hours. The data collected from the Munich Station. The results showed that Root Mean Square Error (RMSE) in ANN was less than RMSE in ANFIS. RMSE in ANN was 3.1931 µg/m3 and in ANFIS was 3.2089 µg/m3, which was better for prediction.

2.2 Deep learning approach

DL is a subfield of ML. As the name implies, DL consists of multiple hidden layers instead of one single layer. The brain of humans has inspired the algorithms of DL. These algorithms will have the ability to learn and solve complex problems. Nowadays, and with the advance of technologies, DL is one of the trending models to predict air pollution.

Many studies have been done in DL to forecast air pollution. This section would focus on the most important studies that have been done with DL.

One of the studies is [11], which made an experimental study to forecast air pollution. It made a comparison between two models: LSTM and Long Short Term Memory Multivariate Regression (LSTM-MVR) model. It showed that LSTM-MVR Model had improved the operation of forecasting air pollution when compared to LSTM. Furthermore, another study forecasted air pollution based on machine learning techniques [12], it suggested two models: MLP and LSTM models. It confirmed that the LSTM model worked better than the MLP model. Moreover, one of the studies that were further done to forecast air pollution is [13], where it used Deep Feed-forward Neural Network (DFNN) that were composed of many layers of LSTM. To deal with the missing value collected from Jingjinji area, it used processing algorithms. The study showed that LSTM outperformed baseline models such as DFNN.

Some studies were made to forecast specific pollutants because of their dangerous effects on human being's health. In [14], for instance, forecasted the concentration of PM2.5, Where it used LSTM and Long Range (LoRa) techniques. It finished with that the system was able to predict the values of PM2.5 for the following few hours. Moreover, in [15], it predicted the PM2.5 concentration using LSTM and self-organizing algorithm. It achieved that due to the memory capability of LSTM, it is proper to be applied when the data are in the form of time-series. Though, the problem occurs when the time interval increase, the exactness of forecast is getting lower. In addition to the previous two studies, another study made the analysis and forecasted PM2.5 due to its correlation with air pollution [16]. It used three models: Random Forest, Encoder-Decoder, and LSTM model. It concluded that if the city has a concentration of PM2.5 as low value, then those three models can fit and have the same performance otherwise, the LSTM model has a better prediction compared with the two models. Additionally, in [17], RNN with LSTM was used to predict PM2.5 for the next four hours. Data was collected from 77 stations to measure air quality in Taiwan from 2012 to 2016. The study showed that the average RMSE values in each region were not significantly different, but the RMSE between the single stations was different.

Other studies predicted the dust levels as it is one of the pollutants in the air. For instance, in [18], the proposed model in the paper was LSTM-based dust prediction model with TensorFlow. The data collected from the Korean Meteorological Administration in Seoul and the results showed that RSME is 8.966 for fine dust levels expected after 1 hour. Besides, in South Korea, another paper predicted the fine dust, and it used RNN with Gated Recurrent Unit (GRU) [19]. The data was used from Air Korea website and the Korea Meteorological Agency website from 2015 to 2017. The paper has shown that the use of GRU is capable of predicting the PM10 and PM2.5 concentrations in the next one month. Furthermore, in South Korea, Seoul, another paper has shown the hourly concentration of fine dust of 25 different stations, which belong to distinctive districts [20]. The model successfully predicted the future dust with Mean Squared Error (MSE) less than 10.7%. There is a study that proposed a model to forecast the values of NO2, PM10 and PM2.5 for the next 24 hours [21]. The data was collected from the meteorological in Delhi for five regions. It used BiLSTM-A, LSTM, Bidirectional Long Short Term Memory (BiLSTM), LSTM-A, and Random Forest. The result showed that BiLSTM-A has better performance for forecasting NO2 and PM10 than the other models.

2.3 Comparisons

After having a full analysis of the previous studies, an important part of the analysis is to compare these studies with the proposed method that is LSTM.

Despite the previous two sections, this section will show some of the studies that were made with LSTM and other techniques.

A study was made in the Bay of Algeciras, Spain to produce precise forecasts of the NO2 concentrations [22]. In order to create the forecasting models, ANN and LSTM were used in the forecasting process. Moreover, the study proposed a new forecasting method based on the LSTM technique that combines LSTM with a cross-validation procedure for time-series (LSTM-CVT). LSTM-CVT models were the best performing models, which internally uses LSTM.

Another study was made to forecast the PM2.5 concentrations in eight Korean cities by using deep learning models. The used deep learning models were RNN, LSTM, and BiLSTM [23]. Principal components analysis (PCA) was used to solve the dimensionality problem. The results showed that both of LSTM and BiLSTM have better performance than RNN when applying PCA since RMSE and MAE values decreased of up to 16.6% and 33.3%.

In [24], a study was done to predict the PM2.5 concentrations by combining and applying Convolutional Neural Network (CNN) and LSTM to the PM2.5 forecasting system. Moreover, traditional ML methods were used to forecast PM2.5 concentrations such as MLP, Support Vector Machine (SVM), Random Forest (RF), Decision Tree (DT), CNN, and LSTM. The results confirmed that the combination of CNN and LSTM is very useful for the forecast of PM2.5 even though, the performances of CNN and LSTM are both great.

The last study to compare with is [25], which used LSTM and deep autoencoder (DAE) methods to forecast the fine PM concentrations. The models showed an effective prediction of fine PM concentrations. However, LSTM model showed better performance than the other models.

From the previous four studies, we can conclude that the LSTM model has the potential to perform predictions with reliable outcomes especially in time-series issues.

3 Methodology approach

Figure 1 represents the general model of Long Short Term Memory (LSTM) for predicting PM10 concentrations. A typical LSTM architecture is composed of an input layer, hidden layers, dense layer, and output layer. Both of PM10 historical data and meteorological factors will be represented as sequential data using vector feature and then goes through the LSTM layer. Each LSTM unit has three gates:

- Input gate:

  Controls the flow of input.

- Output gate:

  Controls the output flow of the cell into the rest of the network.

- Forget gate:

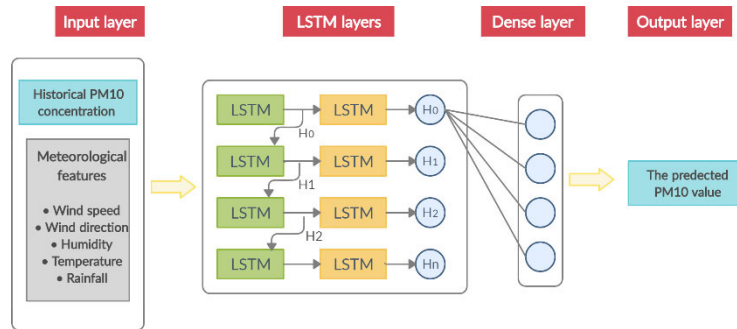 To decide which important information from a prior stage need to be kept or deleted.

Fig. 1. The long short term memory (LSTM) general model

Through replacing RNN neurons with LSTM units, the network can learn the temporal dependencies within the data even though notable gaps occur between events. The output of LSTM layers is passed through a fully connected hidden layer. The output layer produces the concentrations of pollutant [3, 26, 27].

To assess the pollution level, the PM10 concentration will be measured against the Air Quality Index scale (AQI) provided by the general authority of meteorology and environmental protection as shown in Figure 2. The AQI is a simplified method for identifying the status of air quality and it is based on data received from air quality monitoring and control stations, where the concentrations of pollutants are converted into simple numbers that can be understood by the public.

AQI	Air Pollution Level
0-50	Good
51-100	Moderate
101-150	Unhealthy for Sensitive Groups
151-200	Unhealthy
201-300	Very Unhealthy
301-500	Hazardous

Fig. 2. Air quality index (AQI) level

3.1   Type of selected method

After being covered with the theoretical part of the existed technologies to predict air pollution and being done with analyzing the previous studies, this proposed study will focus on the practical part of the proposed method. The proposed method in this study is LSTM, it is not a new method, and because of that, this method will be applied with new data.

This study is a time-series study. Time-series is meant by having a change in data over a specific time [3]. These data need to be analyzed in order to get knowledge from them, and they can be used in many applications [3]. One of these applications is the prediction of air pollution. LSTM can ideally make the prediction based on the analysis of time-series, which was noticed during the analysis of the previous studies.

3.2   Data collection

The used data in this study was obtained from the General Authority of Meteorology and Environmental Protection in Saudi Arabia, which includes data set for the PM10 concentrations in Riyadh and meteorological factors affecting the PM10 prediction process. The data set is collected every hour from the station during the period from 2016 to 2020. The meteorological parameters are Wind speed that expressed in meter per second (m/s), wind direction in degree from the north, the temperature in degree Celsius (oC), rainfall in millimeter per hour (mm/hr), and relative humidity in percentage (%) [2].

3.3   Data analysis

PM10, So2, and No2 samples collected from several locations in Saudi Arabia from 2010 to 2018. Average annual acceptable concentration limit for PM10, So2, and No2 is equal to 80 μg/m3, 80 μg/m3 and 100 μg/m3 and there is no such limits standard for O3 and Co.

Figure 3 shows that PM10 have exceeded the limits for 26 times during this period. The upward trends in PM10 concentration increased tremendously from 2013 to 2015 in Riyadh city ranged from 208 μg/m3 to 181 μg/m3. Excesses occurred during all the years except 2010 to 2012 because there is no documented data according to the General Authority of Meteorology and Environmental Protection. However, there are no extreme excesses of No2 and So2 due to the climatic and geographical nature of Saudi Arabia, which is affected mostly by the occurrence of sandstorms/dust storms. The only excess of No2 was from 2013 to 2015 years in Riyadh. Based on these observations, PM10 is considered the main pollutant in Riyadh [2].
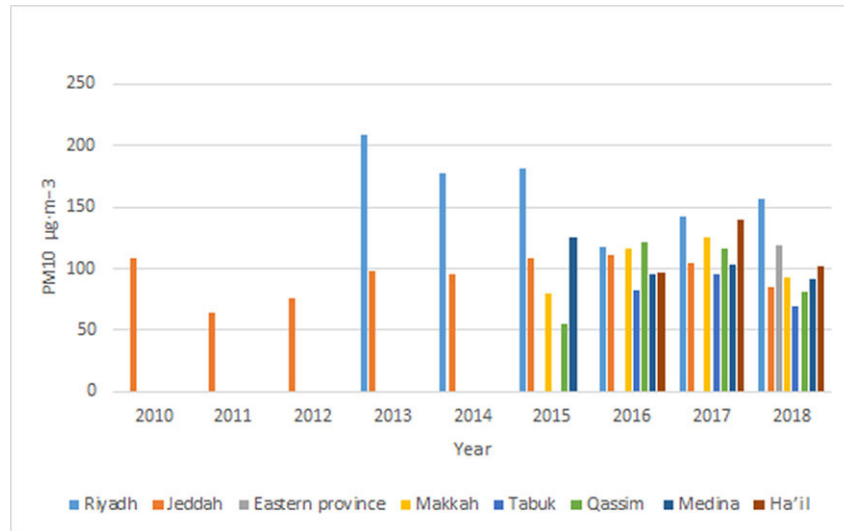
Fig. 3. Annual average concentrations of PM10

4 PM10 forecasting model using LSTM

4.1 Data preprocessing

The data processing step is the most significant step due to its impact on the other remaining steps. This step is the first step since if it was done professionally, then the entire steps will be correct and easier to implement.

The data processing step included these several steps:

- Fill the missing data:

This was the first issue that was noticed and it is a very critical issue. One suggested approach is to ignore this missing data and discard it from the data set [28]. However, this approach is not valid since the data set is in time-series format. Another recommended approach is to use an interpolate function [28]. Interpolate function was used in this study due to its capabilities in accurate predicting the desired values when compared to other techniques, like moving windows.

- Convert data to a supervised learning problem:

LSTM learns to depend on sequence data. Therefore, a key function has been used to convert time-series data into a supervised learning problem. The columns divided into input sequence of PM10 and factors variables and output variable, such that if the current observation of input from var1(t - 1) to var6(t - 1) is provided as input and only the var1(t) is considered as the predicted output.

- Split the data into input and output variables:

Before entering the data into the model, the division of columns into input sequence of variables and output variable is concerned to be done.

4.2   Enhancements

This study made several improvement steps on data before entering them into the model. These improvements contributed to help the LSTM method to be used efficiently to gain better prediction results.

The improvement steps are:

- The normalization step:

The goal of normalization is to bring all the variables to the same range in case of having different ranges in features of the data set. Based on the experience, this step made a huge improvement in the outcome. The used technique was MinMaxScaler, which is a simple technique where the data are fitted into a pre-defined boundary [29]. The data was scaled (normalized) in the range between zero and one.

- The reshape step:

The goal of reshaping the data is to make them suitable to be entered into the LSTM model. The reshape step was done specifically to transform the data to the 3D format [samples, time steps, features], which is expected by the LSTM model to be entered into it.

4.3   Training the model

Three layers have been used in the LSTM model. Each layer in the model contained 50 neurons. The model is trained for 100 epochs using the Adam optimizer. It took nearly 10 to 15 minutes to train the LSTM model. The number of layers was increased in this study after many tries to get more accurate results of prediction.

4.4   Implementation procedure

This experiment in Figure 4 used seven input features retrieved from Khalidia station. The arrangement of the data set was in the appropriate time-series format where each record is maintained hourly in chronological order. After that, when there are missing values in the data, the interpolation function was used to solve this problem. Next, all the data are normalized using MinMaxScaler function to bring all the values to the same range, since LSTM model performance is increased when the values are in zero to one range. The last step of preprocessing is to reshape the data to the 3D format that is expected by LSTM model. Next step is to build the LSTM model then decide if it is not optimized, the optimization will be made by using Adam optimizer and three layers will be used in the model. If the model was optimized, then it will give the predicted values of PM10. All these steps were done using multiple python packages like keras, pandas, sklearn, and matplotlib.
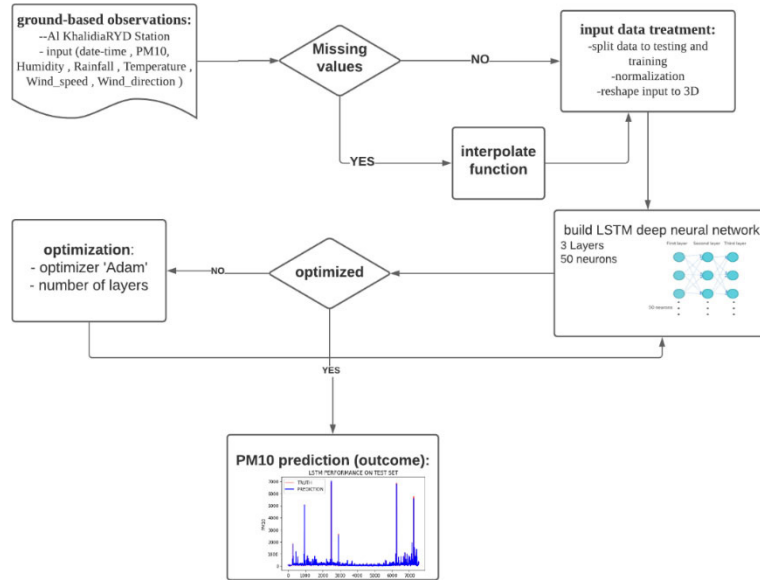
Fig. 4. The proposed algorithm

5   Results

In this section, the results of the model were presented. Figure 5 shows how the actual data correlate with the predicted data. From the plot, it can be observed that predicted values of this model perfectly match the actual values in most of the time.
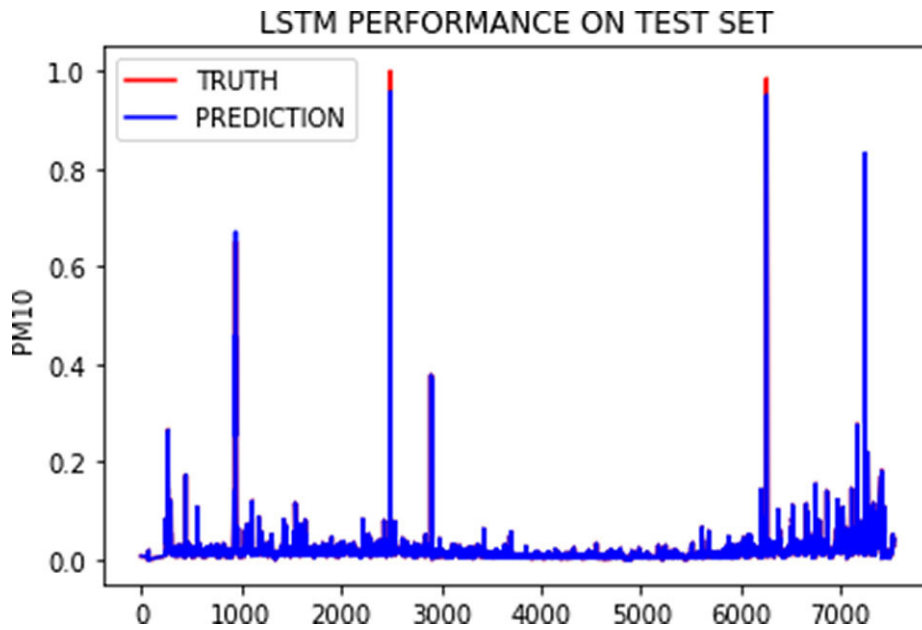

Fig. 5. The plot of the predicted and real PM10 values

Figure 6 shows a sample of predicted and actual data. The numbers prove that the model is effective and efficient in forecasting PM10 values since the difference between these two variables is very small.
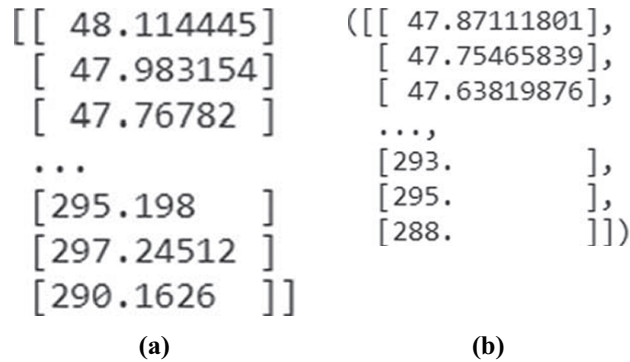
```
[[ 48.114445]        ([[ 47.87111801],
 [ 47.983154]          [ 47.75465839],
 [ 47.76782 ]          [ 47.63819876],
 ...                   ...,
 [295.198   ]          [293.        ],
 [297.24512 ]          [295.        ],
 [290.1626  ]]         [288.        ]])
```
        (a)   **(b)**

Fig. 6. A comparison between the predicted and real PM10 values. (a) PM10 predicted values. (b) PM10 real values

5.1   Major findings

After testing the LSTM model that was used in this study, some error rates measurements have been applied. These measurements can be applied to a time-series prediction problem, and the model has achieved the following:

- MAE measurement:

  LSTM model has achieved the value of MAE as 0.98.

- RMSE measurement:

  LSTM model has achieved the value of RMSE as 8.68.

- R-squared measurement:

  LSTM model has achieved the value of R-Squared as 0.999.

The results have been compared with other researches that have been made in the field of air pollution prediction. After the comparison, the proposed LSTM model showed a lower error rate compared with other error rates in the researches, which mean that the LSTM model has high accuracy across the 24 hours due to the low error rate. The next section will show more details about the comparison of these error rates.

5.2   Comparison of error rates with other studies

Table 1 shows the comparison of RMSE values between this study and the other studies that were made to predict the value of air pollution across the next 24 hours.

Table 1. A comparison between this study and other studies

The Study Name	The used Technique	RMSE Value	Year	Number of Hours
Air Pollution Forecasting Using Deep Learning (this study).	LSTM.	8.68	2020	Next 24 hours.
Real Time Attention based Bidirectional Long Short-Term Memory Networks for Air Pollution Forecasting [21].	LSTM.	96.58	2019	Next 24 hours.
Development of daily PM10 and PM2.5 prediction system using a deep long short-term memory neural network model [30].	LSTM.	15.19	2019	Next 24 hours.

This comparison can prove that this study has a less error rate than the other studies. The proposed methodology achieved more accurate results than other studies due to the improvements that have been made in this paper. These improvements included some steps such as normalization step and adding more layers to increase the learning process. This study has shown that the LSTM method that was used is considered as a reliable method in predicting the value of air pollution across the next 24 hours since it has the lowest error rate value.

6 Discussion

The methodology in this study has been applied to one region in Saudi Arabia that is Riyadh, and it can be applied to any other regions in Saudi Arabia.

The proposed method is shown to be highly accurate at predicting the concentrations of PM10. The predicted PM10 values that were obtained after training the LSTM model were very close to the real values of PM10, which means that the LSTM model is very reliable, and as a result, it can be applied in smart cities in the future. Due to the high accuracy of this methodology, it can overcome all the problems of air pollution.

In future, this study can be extended to be embedded with the Internet of Things (IoT) devices where it can be used in dangerous situations. When having a very high pollution value, the window will automatically close, and the application will alert the people in the house of the PM10 level. After integrating this project with IoT devices, it can be used efficiently in smart cities.

The limitations were rarely found because this study was done based on a scheduled plan. However, one limitation was about the data cleaning process, since it is a time-consuming step.

7 Conclusion

In Saudi Arabia, dust frequently occurs during specific periods, and there is the fact that no system has been found in it to predict the percentage of air pollution. Consequently, this paper applied an air pollution prediction system on the most affected area in Saudi Arabia, the Riyadh city.

This research applied the LSTM technique on one of the stations that exist in Riyadh city. The achieved results were exceptional due to some improvements that have been made on the proposed methodology. One of these improvements was the normalization step that has been made on the data set before entering it to the model. According to the promising testing results that were obtained by this research, the LSTM model achieved an RMSE value of 8.68, MAE value of 0.98, and R-Squared value of 0.999. With the reliable results that were achieved, the model showed a lower error rate than the other researches that have been made in the field of air pollution prediction.

This study showed high accuracy in predicting the percentage of PM10 values compared with the other studies. Hence, this study can be applied to any region in Saudi Arabia. To enhance the performance of the proposed algorithm some improvements have been made to it to achieve a lower error rate than the other researches. The improvements were like increasing the number of layers to increase the deep learning technique in order to get high accuracy and normalize the data set that aided in a huge improvement on the accuracy.

8   References

[1] Alghamdi, A., & Akbar, M. K. (2018). Implementation of an Implicit Solver in ADCIRC Storm Surge Model. Journal of Marine Science and Engineering, 6(2), 62. https://doi.org/10.3390/jmse6020062

[2] General authority of meteorology and environmental protection Saudi. Accessed: 20.8.2021. [Online]. Available: https://ncm.gov.sa/Ar/Environment/AirQuality/Pages/AQ-Dashboard.aspx

[3] Alla, S., and Adari, S. K. (2019). Long Short-Term Memory Models. In Beginning Anomaly Detection Using Python-Based Deep Learning, Apress, Berkeley, CA., pp. 213–256. https://doi.org/10.1007/978-1-4842-5177-5_6

[4] Vlachogianni, A., Kassomenos, P., Karppinen, A., Karakitsios, S., and Kukkonen, J. (2011). Evaluation of a multiple regression model for the forecasting of the concentrations of NOx and PM10 in Athens and Helsinki. Science of the total environment, 409(8): 1559–1571. https://doi.org/10.1016/j.scitotenv.2010.12.040

[5] Sahu, S. K. Bayesian Spatio-Temporal Modelling to Deliver More Accurate and Instantaneous Air Pollution Forecasts. In UK Success Stories in Industrial Mathematics, 2016, Springer, Cham, pp. 67–74. https://doi.org/10.1007/978-3-319-25454-8_9

[6] Athanasiadis, I. N., Kaburlasos, V. G., Mitkas, P. A., and Petridis, V. Applying machine learning techniques on air quality data for real-time decision support. In First international NAISO symposium on information technologies in environmental engineering (ITEE'2003), June 2003, Gdansk, Poland.

[7] Yao, L., Lu, N., and Jiang, S. Artificial neural network (ANN) for multi-source PM2.5 estimation using surface, MODIS, and meteorological data. International Conference on Biomedical Engineering and Biotechnology, May 2012, IEEE, pp. 1228–1231. https://doi.org/10.1109/iCBEB.2012.81

[8] Ip, W. F., Vong, C. M., Yang, J. Y., and Wong, P. K. Forecasting daily ambient air pollution based on least squares support vector machines. In The 2010 IEEE International Conference on Information and Automation, June 2010, IEEE, pp. 571–575. https://doi.org/10.1109/ICINFA.2010.5512401

[9] Telagam, N., Kandasamy, N., & Nanjundan, M. (2017). Smart sensor network based high quality air pollution monitoring system using labview. International Journal of Online Engineering (iJOE), 13(08), 79–87. https://doi.org/10.3991/ijoe.v13i08.7161

[10] Oprea, M., Mihalache, S. F., and Popescu, M. (2017). Computational intelligence-based PM2. 5 air pollution forecasting. International Journal of Computers Communications and Control, 12(3): 365–380. https://doi.org/10.15837/ijccc.2017.3.2907

[11] Abimannan, S., Chang, Y. S., and Lin, C. Y. Air Pollution Forecasting Using LSTM-Multivariate Regression Model. In International Conference on Internet of Vehicles, Springer, Cham, November 2019, pp. 318–326. https://doi.org/10.1007/978-3-030-38651-1_25

[12] Saraubon, K., Wiriyanuruknakon, N., & Tangthirasunun, N. (2021). Flashover Prevention System Using IoT and Machine Learning for Transmission and Distribution Lines. International Journal of Interactive Mobile Technologies, 15(11). https://doi.org/10.3991/ijim.v15i11.20753

[13] Fan, J., Li, Q., Hou, J., Feng, X., Karimian, H., and Lin, S. (2017). A spatiotemporal prediction framework for air pollution based on deep RNN. ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences, 4, 15. https://doi.org/10.5194/isprs-annals-IV-4-W2-15-2017

[14] Song, W., Han, J., Xie, J., Gao, Y., and Song, L. System for detecting and forecasting pm2. 5 concentration levels using long short-Term memory and lora. International Conference on Internet of Things (iThings) and IEEE Green Computing and Communications (GreenCom) and IEEE Cyber, Physical and Social Computing (CPSCom) and IEEE Smart Data (SmartData), July 2019, IEEE, pp. 834–841. https://doi.org/10.1109/iThings/GreenCom/CPSCom/SmartData.2019.00151

[15] Liu, X., Liu, Q., Zou, Y., and Wang, G. A self-organizing LSTM-based approach to PM2.5 forecast. In International Conference on Cloud Computing and Security, Springer, Cham, June 2018, pp. 683–693. https://doi.org/10.1007/978-3-030-00015-8_59

[16] Yan, L., Zhou, M., Wu, Y., and Yan, L. Long short-term memory model for analysis and forecast of PM2.5. International Conference on Cloud Computing and Security, June 2018, Springer, Cham, pp. 623–634. https://doi.org/10.1007/978-3-030-00006-6_57

[17] Tsai, Y. T., Zeng, Y. R., and Chang, Y. S. Air pollution forecasting using RNN with LSTM. 16th Intl Conf on Dependable, Autonomic and Secure Computing, 16th Intl Conf on Pervasive Intelligence and Computing, 4th Intl Conf on Big Data Intelligence and Computing and Cyber Science and Technology Congress (DASC/PiCom/DataCom/CyberSciTech), August 2018, IEEE, pp. 1074–1079. https://doi.org/10.1109/DASC/PiCom/DataCom/CyberSciTec.2018.00178

[18] Kang, S., Kim, N., and Lee, B. D. Fine Dust Forecast Based on Recurrent Neural Networks. 21st International Conference on Advanced Communication Technology (ICACT), February 2019, IEEE, pp. 456–459. https://doi.org/10.23919/ICACT.2019.8701978

[19] Hassan, S., & ul Haq, I. (2019). Pervasive Pollution Problems Caused by Plastics and its Degradation. International Journal of Online & Biomedical Engineering, 15(10). https://doi.org/10.3991/ijoe.v15i10.10873

[20] Kim, S., Lee, J. M., Lee, J., and Seo, J. (2019). Deep-dust: Predicting concentrations of fine dust in Seoul using LSTM. arXiv preprint arXiv:1901.10106.

[21] Dua, R. D., Madaan, D. M., Mukherjee, P. M., and Lall, B. L. Real time attention based bidirectional long short-term memory networks for air pollution forecasting. 5th International conference on Big Data computing service and applications (BigDataService), April 2019, IEEE, pp. 151–158. https://doi.org/10.1109/BigDataService.2019.00027

[22] González-Enrique, J., Ruiz-Aguilar, J. J., Moscoso-López, J. A., Urda, D., Deka, L., and Turias, I. J. (2021). Artificial Neural Networks, Sequence-to-Sequence LSTMs, and Exogenous Variables as Analytical Tools for NO2 (Air Pollution) Forecasting: A Case Study in the Bay of Algeciras (Spain). Sensors, 21(5): 1770. https://doi.org/10.3390/s21051770

[23] Choi, S. W., and Kim, B. H. (2021). Applying PCA to Deep Learning Forecasting Models for Predicting PM2. 5. Sustainability, 13(7): 3726. https://doi.org/10.3390/su13073726

[24] Huang, C. J., and Kuo, P. H. (2018). A deep cnn-lstm model for particulate matter (PM2. 5) forecasting in smart cities. Sensors, 18(7): 2220. https://doi.org/10.3390/s18072220

[25] Xayasouk, T., Lee, H., and Lee, G. (2020). Air pollution prediction using long short-term memory (LSTM) and deep autoencoder (DAE) models. Sustainability, 12(6): 2570. https://doi.org/10.3390/su12062570

[26] Park, J. H., Yoo, S. J., Kim, K. J., Gu, Y. H., Lee, K. H., and Son, U. H. PM10 density forecast model using long short-term memory. 9th International Conference on Ubiquitous and Future Networks (ICUFN), July 2017, IEEE, pp. 576–581. https://doi.org/10.1109/ICUFN.2017.7993855

[27] Bansal, M., Aggarwal, A., Verma, T., and Sood, A. (2019). Air quality index prediction of Delhi using LSTM. Int. J. Emerg. Trends Technol. Comput. Sci, 8: 59–68.

[28] Noor, M. N., Yahaya, A. S., Ramli, N. A., and Al Bakri, A. M. M. (2014). Filling missing data using interpolation methods: Study on the effect of fitting distribution, Trans Tech Publications Ltd., 594: 889–895. https://doi.org/10.4028/www.scientific.net/KEM.594-595.889

[29] Memon, F. A., Khan, U. A., Shaikh, A., Alghamdi, A., Kumar, P., & Alrizq, M. (2021). Predicting Actions in Videos and Action-Based Segmentation Using Deep Learning. IEEE Access, 9, 106918–106932. https://doi.org/10.1109/ACCESS.2021.3101175

[30] Kim, H. S., Park, I., Song, C. H., Lee, K., Yun, J. W., Kim, H. K., and Han, K. M. (2019). Development of a daily PM 10 and PM 2.5 prediction system using a deep long short-term memory neural network model. Atmospheric Chemistry and Physics, 19(20): 12935–12951. https://doi.org/10.5194/acp-19-12935-2019

9   Authors

Manal Alghieth is an assistant professor at computer faculty with PhD in Artificial intelligence and a vice dean of IT deanship, Qassim University, Saudi Arabia (mgietha@qu.edu.sa). Manal Alghieth received the BSc in computer science from Qassim university with two excellence awards. She received the MSc and PhD with distinction from De Montfort University, UK. She has received several research grants from the scientific research deanship at Qassim University. She published many research papers in many refereed journals and conferences and her research interest: Recommendation Systems, Machine Learning, Deep Learning.

Raghad Alawaji is an IT graduated student in College of Computer, Qassim University, Buraydah, Saudi Arabia (E-mail: 361204693@qu.edu.sa).

Safaa Husam Saleh is a graduated Information Technology student in College of Computer, Qassim University, Buraydah, Saudi Arabia (E-mail: 362218946@qu.edu.sa).

Seham Alharbi is an Information Technology graduated student in College of Computer, Qassim University, Buraydah, Saudi Arabia (E-mail: 362206373@qu.edu.sa).