A deep learning model for air quality prediction in smart cities

İbrahim KÖK
Department of Computer Science
Gazi University
Ankara, Turkey
ikok@gazi.edu.tr

Mehmet Ulvi ŞİMŞEK
Department of Computer Engineering
Gazi University
Ankara, Turkey
mehmet.ulvi.simsek@gazi.edu.tr

Suat ÖZDEMİR
Department of Computer Engineering
Gazi University
Ankara, Turkey
suatozdemir@gazi.edu.tr

Abstract—**In recent years, Internet of Things (IoT) concept has become a promising research topic in many areas including industry, commerce and education. Smart cities employ IoT based services and applications to create a sustainable urban life. By using information and communication technologies, IoT enables smart cities to make city stakeholders more aware, interactive and efficient. With the increase in number of IoT based smart city applications, the amount of data produced by these applications is increased tremendously. Governments and city stakeholders take early precautions to process these data and predict future effects to ensure sustainable development. In prediction context, deep learning techniques have been used for several forecasting problems in big data. This inspires us to use deep learning methods for prediction of IoT data. Hence, in this paper, a novel deep learning model is proposed for analyzing IoT smart city data. We propose a novel model based on Long Short Term Memory (LSTM) networks to predict future values of air quality in a smart city. The evaluation results of the proposed model are found to be promising and they show that the model can be used in other smart city prediction problems as well.**

Keywords—*Deep learning; Internet of things(IoT); Smart city; Long short term memory(LSTM); IoT data analytics; Air quality prediction*

I. INTRODUCTION

Internet of Things (IoT) is known as an interconnection of everyday objects [1]. It improves the quality of human life by connecting an emergent number of objects. It is estimated that there will be 50 billion connected devices in IoT by 2020 [2]. Devices from different domains are expected to be interconnected to provide valuable information to consumers, manufacturers and utility providers [3]. Enrichment of these devices emerges smart applications that facilitate the quality of human life. IoT applications generate vast amount of data by monitoring the physical phenomenon surrounding them. Generated data are heterogeneous, distributed and continuously changing and increasing day by day, hence maximizing the data analysis efficiency in such data is important [4].

Smart Cities (SCs) make the city services and monitoring of cities more aware, interactive and efficient [5]. Citizens and stakeholders get better interaction with benefit of improving quality of urban services and city life [6]. Data, which is essential for smart solutions, is the most important aspect of SCs. Despite the opportunities of SC data, creating smart solutions and applications for the needs of city users is a

challenging task. This is simply due to volume, variety and dynamics of the generated data. Hence, efficient techniques must be developed to process and obtain valuable information from SC data.

Due to environmental and climate changes, today, many cities are facing air pollution problem [7]. According to a recent report exposure to particulate matter which is a type of air pollution caused about 4.2 million deaths and it ranks 5th worldwide among all risks [8]. This dramatic expansion motivates us to focus on air pollution problem and propose a solution based on IoT data analysis. Air pollution causes various impacts on human health, including respiratory problems, hospitalization for heart or lung diseases, and premature death [9]. Especially, ozone and nitrogen dioxide gases effect people who suffer from lung diseases such as asthma, chronic bronchitis and emphysema that worsen their diseases. Children and elderly adults are also effected by their gases concentration. It depends on air quality that effects how you live and breathe. Air Quality Index (AQI) is used for understanding local air quality that tell us how clean or unhealthy your air is [10]. Therefore, forecasting air pollution parameters such as ozone and nitrogen dioxide is important to take precautionary measures. These parameters have also significant effect on air pollution degree and are measured by sensing devices deployed in many SCs. As stated earlier, the integration of sensing devices into cities produces a great amount of data that are timely annotated. Processing and forecasting a great amount of data bring out the Deep Learning (DL) concept. Increased chip processing abilities, significantly lowered cost of computing hardware and recent advances in machine learning and information processing studies are the main reasons for popularity of DL [11]. Thus, DL has a great attention in academic and industrial areas of interest and it has been applied with successfully in many domain such as classification tasks, natural language processing, dimensionality reduction, object detection, motion modeling, and so on [12]. Therefore, monitored gases produce temporal data that can be evaluated in early manner to predict future value.

The main contribution of this paper is two-fold: First, we present a DL model that can be applied to SC IoT data. Second, we design and implement a novel LSTM based prediction model that is helpful to solve future air quality problems in SCs.

The rest of the paper is structured as follows. We first introduce the related works in Section II. Then, we give essential information about IoT SC and DL in Section III and IV, respectively. We present the LSTM based air quality prediction model in Section V. The experiments and experimental results is given in Section VI. Finally, we conclude the paper in Section VII.

II. RELATED WORK

The development of SC technologies has favored researchers to focus their efforts on improving to manage resources of SC. These efforts are made for enhancing quality of life in cities via monitoring sensor device deployed in SC. So, it is important to predict situations in SCs to take precautions before an unpleasant situation occurs. In this context, traffic management systems are studied for helping drivers and citizens to improve quality of city life [13]. A neural network model of parking availability prediction is also presented that give information about parking occupancy prediction in selected regions [14]. Concentration of air pollutants in SC is another problem that affect people's health. Therefore, an air quality monitoring and alarm system could hamper the degree of air pollutions when taking precautionary measures. There are many studies proposing solutions for the prediction problem of AQI in the literature. These methods are physical deterministic model, statistical, empirical and neural network approach [15] that can be categorized in deterministic and statistical methods. Hybrid model uses neural network that get higher prediction performance [16-18]. Hybrid approaches including fuzzy models [19, 20] are also used to predict AQI. In another study, an LSTM based approach is used for predicting pollutant gas to warn citizens [21].

DL models are regarded as a powerful models for showing excellent performance on difficult learning tasks [22]. There are many studies proposing solutions for the prediction problem of DL in the literature. LSTM which is a special type of DL is the state of the art recurrent neural network (RNN) for supervised temporal sequence learning [23]. LSTM has a structure of loops that memorize previous events to make better use of its input [24]. LSTM is used to solve a wide range of problems such as speech recognition [36], machine translation, traffic flow prediction [25, 26], translation [22], text recognition in videos [27] and acoustic modeling [28] etc. It has a great performance in all this wide area that model sequential data with mapping input sequence to target sequence [22].

III. IOT AND SMART CITY

City populations are growing much faster than available resources. So, cities need sustainable ways to manage resources for growing populations. We have to enhance quality of life with implementing a SC concept [29]. SC is a framework that implements information services to monitor public areas and infrastructures [30]. SC has gained an importance by using information and communication technologies to make city stakeholders more aware, interactive and efficient. SC utilizes information and communication technologies to improve quality of life [31]. SC produces rich sources of information with deployed various sensory devices and IoT infrastructures. This information is derived from the environments in which enormous amount of data can be collected to be harnessed by many smart city applications [32]. These applications consist of several SC models but they have common characteristic such as smart economy, smart mobility, smart governance, smart environment, smart living and smart people [6].

A large number of research projects are conducted that collect data with benefit of IoT. These collected data coming from heterogeneous sources are various types such as traffic, weather, pollution, noise data [32]. The data are low level raw sensor data that can be analyzed by analytics for detecting important city events. They have been described as a valuable source for the deployment of big data businesses [25]. It is beneficial for triggering actions to keep from bad situations for city citizens. Most SC applications have to deal with huge volumes of data, with high velocity, dynamicity and variety [33]. Heterogeneity of data is another problem that solved by semantic technologies provides interpret and interrelation among data [34].

IV. DEEP LEARNING

Natural data that has a raw form that can be processed limitedly by conventional machine learning methods. The term of Deep Learning (DL) which is a type of machine learning [12] has emerged with the advance of representation learning. DL methods can automatically discover the representations for classification and detection in natural data. DL has contributed major advances in solving problems in artificial intelligence community for many years with using machine learning algorithms [35]. It also provides better representation and classification on time series problems compared to shallow approaches when configured and trained properly.

In machine learning, artificial neural networks (ANN) consist of neurons which are activated through weighted connections from previously active neurons [36]. DL named Deep Neural Network (DNN) [37] uses the structure of ANN that processes information and signals more efficiently. It uses hidden layers which are contained between input and output layer that determines weighted layers from input layer to last layer [38]. Therefore, it provides a non-linear transformation of the deep learning data and can model complex relationships with multi-layered structures instead of shallow ones.

Analysis of time series data has been studied by researchers for decades [39]. Traditional shallow models do not have the capacity for modeling sequential data with high accuracy rate. As time series data are more complex, high-dimensional and noisy, DL is often associated with complex real-world data and have been applied to the problem of time series prediction [40]. In the next section, we propose a prediction model based on LSTM to predict air quality in a SC.

V. Air Quality Prediction Model

In this section, air pollution, air quality index, LSTM, SVR and proposed prediction model are explained in detail.

A. Air Pollution and Quality Index (AQI)

Air pollution is an environmental and social problem that must be struggled in many ways. The most important gases causing air pollution are ozone (O_3), carbon monoxide (CO), sulfur dioxide (SO_2), nitrogen dioxide (NO_2) and particulate matter [41].

These gases have important effects on climate change, ecosystems, human health, cultural heritage and economy [42]. Therefore, analysis of pollutant gas concentrations is important for the protection of the inhabitants.

AQI is an index for indicating daily air quality. It is calculated from major pollutant gas concentrations and it is divided into six levels of health concern [10].

Air Quality Index (AQI) Values	Levels of Health Concern	Colors
0 to 50	Good	Green
51 to 100	Moderate	Yellow
101 to 150	Unhealthy for Sensitive Groups	Orange
151 to 200	Unhealthy	Red
201 to 300	Very Unhealthy	Purple
301 to 500	Hazardous	Maroon

Figure 1. Air Quality Index Levels

Environmental Protection Agency (EPA) defines the value of 100 AQI as a critical level for the protection of public health [10]. Air quality is considered satisfactory when AQI values below 101, but values higher than 101 is considered unhealthy. All AQI levels determined by EPA are shown in Fig 1. In this study we follow EPA's AQI index to develop the proposed system.

B. Long Short Term Memory (LSTM)

LSTM is a special kind of recurrent neural network (RNN) and capable of learning long-term dependencies. It was introduced by Hochreiter and Schmidhuber in order to overcome vanishing gradient problem in 1997 [43]. In this neural network model, a memory block takes the place of each ordinary neuron in the hidden layer of standard recurrent neural network [44].

The LSTM block shown in Fig. 2 has an input gate, a forget gate and an output gate which regulate the flow of information into and out of the cell. These gates, block input and block output as follows:
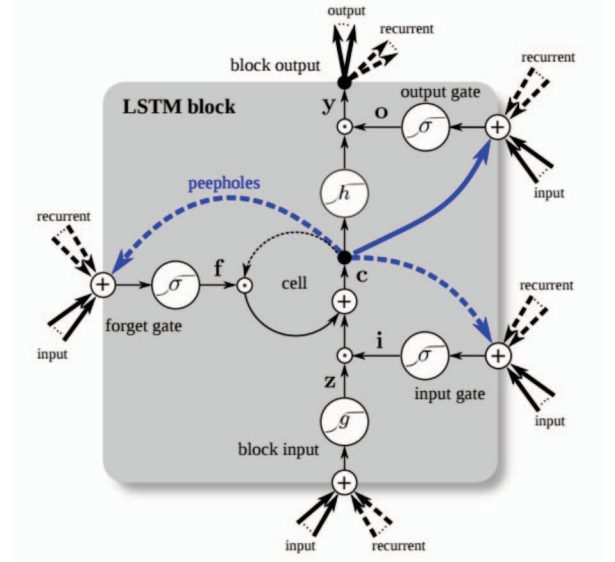

Figure 2. A Long Short-Term Memory Block [45]

block input:
$$z^t = g(W_z x^t + R_z y^{t-1} + b_z) \qquad (1)$$

input gate:
$$i^t = \sigma(W_i x^t + R_i y^{t-1} + p_i \odot c^{t-1} + b_i) \qquad (2)$$

forget gate:
$$f^t = \sigma(W_f x^t + R_f y^{t-1} + p_f \odot c^{t-1} + b_f) \qquad (3)$$

cell state:
$$c^t = i^t \odot z^t + f^t \odot c^{t-1} \qquad (4)$$

output gate:
$$o^t = \sigma(W_o x^t + R_o y^{t-1} + p_o \odot c^t + b_o) \qquad (5)$$

block output:
$$y^t = o^t \odot h(c^t) \qquad (6)$$

where x^t is input vector at time t, W_z, W_i, W_f, W_o are the weights matrices connecting x^t to the three gates and block input, R_z, R_i, R_f, R_o are recurrent weight matrices connecting y^{t-1} to the three gates and block input, b_z, b_i, b_f, b_o are the bias vectors. σ represents the logistic sigmoid function and h represents hyperbolic tangent function. σ is used for as activation of the gates and g is used as the block input and output activation function.

LSTM has been applied on a large variety of real world problems such as machine translation [46], speech recognition [47], neural language model [48] and image recognition [49]. In this study, a novel prediction model based on LSTM is proposed on IoT SC data.

C. Proposed Prediction Model

The proposed air quality prediction model consists of three parts. In the first part, a deep learning model consisting of LSTM neural networks is realized. Numerous experiments have been conducted with varying hyper parameters to build best LSTM structure. After the experiments, the network structure that gives the best result is constructed.

This network structure has an input layer, a hidden layer with 24 LSTM units, and an output layer. Other hyper parameters of the network are denoted in Table I.

TABLE I. PROPOSED MODEL HYPER PARAMETERS

Hyperparameters	Values
Input Sequence	8
Hidden Layer	1
LSTM Units	24
Output Layer	1
Batch Size	50
Number of Epoch	100

Another prediction method SVR is trained in model for evaluating the LSTM success.

In the second part, a labeling unit is created that labels data according to the daily AQI values. In the last part, a decision unit is developed which maps according to the observed and predicted alarm situations. The complete structure of the proposed model is shown in Fig.3.
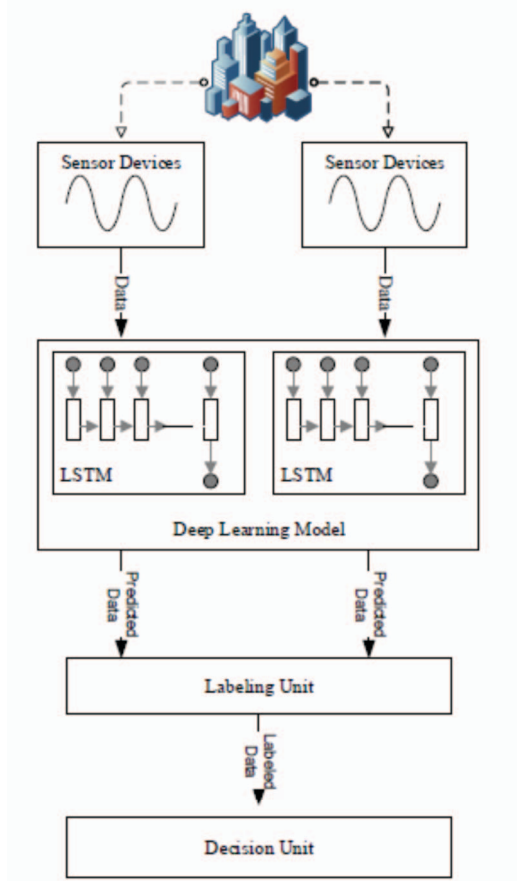

Figure 3. Proposed Prediction Model

In the proposed model, a certain number of past sequential values are taken into account and predictions are made for the next step, denoted by Eq. (7).

$$a(t-n), \ldots, a(t-2), a(t-1), a(t) = \acute{a}(t+1) \qquad (7)$$

where $a(t-n), \ldots, a(t-2), a(t-1), a(t)$ are the observed values (input sequence) and $\acute{a}(t+1)$ is the predicted value.

Then, the network loss is calculated by comparing the observed values with the predicted values at time t + 1. The overall network performance was evaluated using performance metrics given in Eq. (11) and (12).

D. Support Vector Regressor (SVR)

Support Vector Machine (SVM) is a method used for classifying linear and non-linear data [50]. It can easily solve problems such as small sample size, high dimension. It can be used for numeric prediction as well as classification. For this purpose, a model called Support Vector Regression (SVR) is used to solve the regression problems. SVR can also achieve both linear and nonlinear regression case [51]. For nonlinear regression cases, a kernel function is used for mapping, whereby the equation is defined as follows.

$$f(x) = \sum_{i=1}^{n} (\alpha_i - \acute{\alpha}_i) K(x_i, x) + b \qquad (8)$$

where $K(x_i, x)$ is the kernel function, x_i is input vector and n is the number of samples.

VI. EXPERIMENTS

A. Dataset:

CityPulse EU FP7 Project [52] involves several SC applications based on IoT. Within the scope of the project, road traffic, pollution, weather, cultural, library and parking data were collected from the cities of Aarhus and Brasov in Denmark and Romania respectively between 2013 and 2015.

In this study, pollution dataset of CityPulse EU FP7 Project is used to realize the proposed system. The dataset contains 8 features including ozone, particulate matter, carbon monoxide, sulfur dioxide, nitrogen dioxide, longitude, latitude and timestamp was used for experiment. The dataset has 17568 samples that are collected at five-minute intervals. Each sample value is given in the form of EPA's AQI standard. In this study, ozone and nitrogen dioxide pollutants are selected for air quality prediction.

B. Model Training

The proposed model is trained on the Python platform by using TensorFlow and Keras DL framework. Primarily, ozone and nitrogen dioxide data are divided into two parts as follows: training set % 69.5 and test set % 30.5. Then, the network is configured according to the hyper parameters shown in Table I. Training are carried out by SVR and using separate LSTM models for each type of gas. In the process of LSTM training, Adaptive Moment Estimation (*Adam*) optimizer which computes individual adaptive learning rate for different parameters, is used to minimize the loss function [53].

C. Thresholds

In this section, AQI levels are divided into three threshold values by considering AQI critical level (100). These threshold values are defined as alarm color and shown in Table II.

TABLE II. THRESHOLD AIR QUALITY INDEX

AQI Values	Level of Health Concern	Alarm Color
0 – 50	Good	Green
51 – 100	Moderate	Yellow
101 – 500	Unhealthy– Hazardous	Red

Ozone and nitrogen dioxide test datasets are separately labeled as Good, Modarate and Unhealty-Hazardous according to these threshold values as listed in Table III.

TABLE III. ALARM DECISION TABLE

Ozone	Nitrogen Dioxide	Alarm Color
Good	Good	Green
Good	Moderate	Green
Good	Unhealthy-Hazardous	Yellow
Moderate	Good	Green
Moderate	Moderate	Yellow
Moderate	Unhealthy-Hazardous	Red
Unhealthy-Hazardous	Good	Yellow
Unhealthy-Hazardous	Moderate	Red
Unhealthy-Hazardous	Unhealthy-Hazardous	Red

According to Table III, the final alarm colors are labeled by evaluating observed values and predicted values of both pollutant concentrations. Eq. 9 and 10 were applied separately for labeling. Then, the correct matching performance of the prediction model is calculated by comparing the observed $a_{alarm}(t+1)$ and predicted $ά_{alarm}(t+1)$ alarm color state Eq. (11). The classification performance is calculated by using final alarm situations via the equations below.

$$a_{ozone}(t+1) - a_{nitrogen\ dioxide}(t+1) \xrightarrow{decision} a_{alarm}(t+1) \quad (9)$$

$$ά_{ozone}(t+1) - ά_{nitrogen\ dioxide}(t+1) \xrightarrow{decision} ά_{alarm}(t+1) \quad (10)$$

$$a_{alarm}(t+1) \xrightarrow{matching\ ?} ά_{alarm}(t+1) \quad (11)$$

D. Model Performance Metrics

The proposed model's performance is assessed by MAE (Mean Absolute Error) and RMSE (Root Mean Squared Error) error criteria. These are defined by Eq. (12) and (13);

$$MAE = \frac{1}{n} \sum_{i=1}^{n} |\alpha_i - ά_i| \quad (12)$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (|\alpha_i - ά_i|)^2} \quad (13)$$

where α represents the actual observed data and ά represents the predicted data and n is the number of prediction.

The alarm color matching performance of the proposed prediction model is measured by well known Precision, Recall, F1-Score and Accuracy metrics.

E. Experimental Results

In this section, we introduced the results of evaluation of our air quality prediction model and the forecasting results of the two air pollutant concentrations. We compared the proposed model with Support Vector Regressor (SVR) technique to verify the prediction ability. To evaluate the prediction performance from multiple perspectives, we used two error criteria (RMSE, MAE) and Confusion Matrix (CM), Precision, Recall and F1-Score metrics.

TABLE IV. PREDICTION RESULTS

	Train		Test	
	RMSE	MAE	RMSE	MAE
Our model-O_3	**3,22**	**2,77**	**3,26**	**2,81**
SVR-O_3	6,48	4,97	6,79	5,15
Our model-NO_2	**5,49**	**4,49**	**3,76**	**3,11**
SVR-NO_2	6,91	4,94	4,91	3,79

For both prediction methods, we obtained the prediction results which are shown in Table IV. It seems that the proposed model provides a considerable improvement in terms of error criteria. According to these results, we were obtained following Confusion Matrix (CM), Precision, Recall and F1-Score values from the classification which made by using the threshold values in Table II and the decision table in Table III.

TABLE V. CONFUSION MATRIX FOR OUR MODEL

Alarm Color	Red	Yellow	Green
Red	1821	73	3
Yellow	42	1639	74
Green	0	81	1617

TABLE VI. CONFUSION MATRIX FOR SVR

Alarm Color	Red	Yellow	Green
Red	1851	53	0
Yellow	101	1630	24
Green	3	198	1497

From the comparison of two CMs in Table V and VI, we can see that our proposed model have the lowest error rates in Yellow and Green alarms. In red alarms, it has a bit higher error rates than SVR. According to CMs, we calculated the precision, recall and F1-Score values are presented in Table VII.

TABLE VII. PRECISION, RECALL AND F1-MEASURE

	Precision		Recall		F1-Score	
	LSTM	SVR	LSTM	SVR	LSTM	SVR
Red	0.98	0.95	0.96	0.97	0.97	0.96
Yellow	0.91	0.87	0.93	0.93	0.92	0.90
Green	0.95	0.98	0.95	0.88	0.95	0.93

Please note that if the model falsely predicts green and yellow alarm colors as red, it is tolerable, but the opposite is not acceptable. This means that the red alarm prediction values is more important than the others. If we evaluate in this context, the best Precision (% 98) and F1-Score (% 97) results were achieved on red alarms with our model. Regarding F1-Score, it performed better performance in all alarm colors than SVR. Accordingly, F1-scores were obtained as % 97, % 92 and % 95 per alarm color. In SVR these score were seen as %96, % 90 and % 93.

On the other hand, the prediction accuracy of the LSTM and SVR based model were calculated as % 95 and % 92.9, respectively. These scores have shown that LSTM based model outputs has a very good performance. The overall results indicate that LSTM based prediction model takes its advantages of memorizing of long historical data and achieve higher prediction accuracy even if it has a simple network structure.

VII. CONCLUSION

IoT paradigm brings new heterogeneous data particularly collected from sensor devices. Extracting and retrieving the hidden knowledge from this type of IoT data is a challenging task. It also needs robust models that perform these tasks fast and efficiently.

In this paper, we propose a DL model to overcome air pollution problems in SC. We firstly configured the network with the best hyper parameters according to the results obtained from the experiments. Then, the proposed model is trained, and evaluated with widely used RMSE and MAE metrics.

Later, labeling and decision units are designed and the prediction performance of model is calculated with the Precision, Recall, F1-measure and accuracy metrics. The overall results verify that our LSTM based model has a better performance than SVR based model.

Consequently, the obtained results show that the employment of the LSTM based prediction model to the IoT data is effective and promising. In the future, we focus on more advanced models that include different DL methods for IOT data analytics.

REFERENCES

[1] F. Xia, L. T. Yang, L. Wang, and A. Vinel, "Internet of Things," *International Journal of Communication Systems,* vol. 25, pp. 1101-1102, 2012.

[2] J. Chase, "The evolution of the internet of things," Texas Instruments2013.

[3] O. Monnier, "A smarter grid with the Internet of Things," Texas Instruments2013.

[4] M. I. Ali, N. Ono, M. Kaysar, Z. U. Shamszaman, T.-L. Pham, F. Gao*, et al.*, "Real-time data analytics and event detection for IoT-enabled communication systems," *Web Semantics: Science, Services and Agents on the World Wide Web,* vol. 42, pp. 19-37, 1// 2017.

[5] J. Bélissent, "Getting Clever About Smart Cities New Opportunities Require New Business Models," Forrester Research2010.

[6] H. B. Sta, "Quality and the efficiency of data in "Smart-Cities"," *Future Generation Computer Systems,* 2016.

[7] D. Campbell-Lendrum and C. Corvalan, "Climate change and developing-country cities: implications for environmental health and equity," *J Urban Health,* vol. 84, pp. i109-17, May 2007.

[8] IHME, "State of Global Air 2017," Institute for Health Metrics and Evaluation (IHME)2017.

[9] H. K. Cigizoglu, K. Alp, and M. Kömürcü, "Estimation of Air Pollution Parameters Using Artificial Neural Networks," presented at the Advances in Air Pollution Modeling for Environmental Security, 2005.

[10] U. S. E. P. Agency, "Air Quality Index: A Guide to Air Quality and Your Health," 2014.

[11] L. Deng, "A tutorial survey of architectures, algorithms, and applications for deep learning," *APSIPA Transactions on Signal and Information Processing,* vol. 3, 2014.

[12] Y. Lv, Y. Duan, W. Kang, Z. Li, and F.-Y. Wang, "Traffic Flow Prediction With Big Data: A Deep Learning Approach," *IEEE Transactions on Intelligent Transportation Systems,* pp. 1-9, 2014.

[13] C. T. Barba, M. A. Mateos, P. R. Soto, A. M. Mezher, and M. A. Igartua, "Smart city for VANETs using warning messages,traffic statistics and intelligent traffic lights," presented at the IEEE Intelligent Vehicles Symposium (IV) 2012.

[14] E. I. Vlahogianni, K. Kepaptsoglou, V. Tsetsos, and M. G. Karlaftis, "A Real-Time Parking Prediction System for Smart Cities," *Journal of Intelligent Transportation Systems,* vol. 20, pp. 192-204, 2015.

[15] X. L. L. P. Y. H. J. S. T. Chi, "Deep learning architecture for air quality predictions," *Environ Sci Pollut Res,* vol. 23, pp. 22408–22417, 2016.

[16] R. S. P. Awkash Kumar, Anil Kumar Dikshit, Rakesh Kumar, "Application of AERMOD for short-term air quality prediction with forecasted meteorology using WRF model," *Clean Techn Environ Policy,* vol. 19, pp. 1955–1965, 2017.

[17] Z. Yang and J. Wang, "A new air quality monitoring and early warning system: Air quality assessment and air pollutant concentration prediction," *Environmental Research,* vol. 158, pp. 105-117, 2017.

[18] J. Wang, X. Zhang, Z. Guo, and H. Lu, "Developing an early-warning system for air quality prediction and assessment of cities in China," *Expert Systems with Applications,* vol. 84, pp. 102-116, 2017.

[19] S. Munawar, M. Hamid, M. S. Khan, A. Ahmed, and N. Hameed, "Health Monitoring Considering Air Quality Index Prediction Using Neuro Fuzzy Inference Model A Case Study of Lahore, Pakistan," *Journal of Basic & Applied Sciences,* 2017.

[20] L. Wang, X.-y. Xiao, and J.-y. Meng, "Prediction of air pollution based on FCM-HMM Multi-model," in *35th Chinese Control Conference*, 2016.

[21] E. Pardo and N. Malpica, "Air Quality Forecasting in Madrid Using Long Short-Term Memory Networks," *Biomedical Applications Based on Natural and Artificial Computing,* vol. 10338, 2017.

[22] I. Sutskever, O. Vinyals, and Q. V. Le, "Sequence to Sequence Learning with Neural Networks," *Computation and Language,* 2014.

[23] F. J. Gomez and J. Schmidhuber, "Co-Evolving Recurrent Neurons Learn Deep Memory POMDPs," in *7th annual conference on Genetic and evolutionary computation*, Washington, USA, 2005.

[24] Y. Tang, J. Xu, K. Matsumoto, and C. Ono, "Sequence-to-Sequence Model with Attention for Time Series Classification," presented at the IEEE 16th International Conference on Data Mining, 2016.

[25] A. Abella, M. Ortiz-de-Urbina-Criado, and C. De-Pablos-Heredero, "A model for the analysis of data-driven innovation and value generation in smart cities' ecosystems," *Cities,* vol. 64, pp. 47-53, 2017.

[26] N. G. Polson and V. O. Sokolov, "Deep learning for short-term traffic flow prediction," *Transportation Research Part C: Emerging Technologies,* vol. 79, pp. 1-17, 2017.

[27] S. Yousfi, S.-A. Berrani, and C. Garcia, "Contribution of recurrent connectionist language models in improving LSTM-based Arabic text recognition in videos," *Pattern Recognition,* vol. 64, pp. 245-254, 2017.

[28] H. Sak, A. Senior, and F. Beaufays, "Long short-term memory recurrent neural network architectures for large scale acoustic modeling," presented at the INTERSPEECH, 2014.

[29] R. Jain, "Internet of Things and Smart Cities: Challenges and Issues," 2016.

[30] L. Filipponi, A. Vitaletti, G. Landi, V. Memeo, G. Laura, and P. Pucci, "Smart City: An Event Driven Architecture for Monitoring Public Spaces with Heterogeneous Sensors," pp. 281-286, 2010.

[31] J. Jin, J. Gubbi, S. Marusic, and M. Palaniswami, "An Information Framework for Creating a Smart City Through Internet of Things," *IEEE Internet of Things Journal,* vol. 1, pp. 112-121, 2014.

[32] D. Puiu, P. Barnaghi, R. Tönjes, D. Kümper, M. I. Ali, A. Mileo*, et al.*, "CityPulse Large Scale Data Analytics Framework for Smart Cities," *IEEE Access: Special Section on Smart Cities* 2016.

[33] P. Barnagni, M. Bermudez Edo, and R. Tönjes, "Challenges for Quality of Data in Smart Cities," *ACM Journal of Data and Information quality,* 2015.

[34] S. Consoli, M. Mongiovì, D. R. Recupero, S. Peroni, A. Gangemi, A. G. Nuzzolese*, et al.*, "Producing Linked Data for Smart Cities the case of Catania," *The Role of Semantics in Smart Cities,* 2014.

[35] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature,* vol. 521, pp. 436-444, 2015.

[36] J. Schmidhuber, "Deep Learning in Neural Networks An Overview," 2014.

[37] S. S. Tirumala and S. R. Shahamiri, "A review on Deep Learning approaches in Speaker Identification," pp. 142-147, 2016.

[38] N. D. Lane, S. Bhattacharya, P. Georgiev, C. Forlivesi, and F. Kawsar, "An Early Resource Characterization of Deep Learning on Wearables, Smartphones and Internet-of-Things Devices," pp. 7-12, 2015.

[39] M. Langkvist, L. Karlsson, and A. Loutfi, "A review of unsupervised feature learning and deep learning for time-series modeling," *Pattern Recognition Letters,* vol. 42, pp. 11-24, Jun 1 2014.

[40] I. M. Coelho, V. N. Coelho, E. J. d. S. Luz, L. S. Ochi, F. G. Guimarães, and E. Rios, "A GPU deep learning metaheuristic based model for time series forecasting," *Applied Energy,* 2017.

[41] M. M. Rathore, A. Ahmad, A. Paul, and S. Rho, "Urban planning and building smart cities based on the Internet of Things using Big Data analytics," *Computer Networks,* vol. 101, pp. 63-80, 6/4/ 2016.

[42] E. E. Agency, "Air quality in europe," 2016.

[43] B. Hochreiter and J. Schmidhuber, "Long Short Term Memory," *Neural Computation,* vol. 9, pp. 1735-1780, 1997.

[44] Z. C. Lipton, J. Berkowitz, and C. Elkan, "A Critical Review of Recurrent Neural Networks for Sequence Learning," *2015,* 2015.

[45] K. Greff, R. K. Srivastava, J. Koutn´ık, B. R. Steunebrink, and J. Schmidhuber, "Lstm:A Search Space Odyssey," presented at the IEEE Transactions on Neural Networks and Learning Systems, 2016.

[46] D. Bahdanau, K. Cho, and Y. Bengio, "Neural Machine Translation by Jointly Learning to Align and Translate," in *International Conference on Learning Representations*, 2014.

[47] A. Graves, A.-R. Mohamed, and G. Hinton, "Speech recognition with deep recurrent neural networks," presented at the 2013 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2013.

[48] O. Vinyals and O. V. Le, "A Neural Conversational Model," presented at the Proceedings of the 31 st International Conference on Machine Learning, Lille, France, 2015.

[49] J. Donahue, L. A. Hendricks, S. Guadarrama, M. Rohrbach, S. Venugopalan, K. Saenko*, et al.*, "Long Term Recurrent Convolutional Networks for Visual Recognition and Description," *Conference on Computer Vision and Pattern Recognition(CVPR),* 2015.

[50] J. Han, M. Kamber, and J. Pei, *Data Mining: Concepts and Techniques*: Morgan Kaufmann Publishers Inc., 2011.

[51] M. Niu, Y. Wang, S. Sun, and Y. Li, "A novel hybrid decomposition-and-ensemble model based on CEEMD and GWO for short-term PM2.5 concentration forecasting," *Atmospheric Environment,* vol. 134, pp. 168-180, 2016.

[52] T. C. Consortium, "CityPulse Annual Report," The CityPulse Consortium2016.

[53] D. P. Kingma and J. B. Ba, "Adam: A Method for Stochastic Optimization," presented at the International Conference on Learning Representations, 2015.