Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 198 (2022) 530–535

Procedia Computer Science

www.elsevier.com/locate/procedia

International Workshop of Innovation and Technologies (IWIT 2021)
November 1-4, 2021, Leuven, Belgium

Machine learning in the service of a clean city

Essamlali Ismail[a]*, Bahnasse Ayoub[a,b], Khiat Azeddine[a], Ouajji Hassan[a]

[a] LAB SSDIA, ENSET Mohammedia, University Hassan II of Casablanca, Morocco
[b] ENSAM Casablanca, University Hassan II of Casablanca, Morocco

Abstract

Clean city, green city and sustainable city are slogans describing our hope to live in areas free of pollution. Pollution is an issue that is becoming more and more worrying to the world's population, climate change is affecting our globe at all levels, and cities are undergoing the worst consequences. Combining the useful with the pleasant, the combat against pollution becomes intelligent and efficient in a smart city where artificial intelligence is widely used as an implement. In this paper we discuss the use of artificial intelligence, especially machine learning methods to predict and alert about pollution in urban areas. Supervised learning as well as unsupervised learning is utilized. We focus in this study on four fields: Air pollution; Water quality; urban noise and traffic and waste management.

© 2021 The Authors. Published by Elsevier B.V.
This is an open access article under the CC BY-NC-ND license (https://creativecommons.org/licenses/by-nc-nd/4.0)
Peer-review under responsibility of the Conference Program Chairs

Keywords: Artificial intelligence, smart city, pollution, machine learning:

1. Introduction

In Smart cities we use technologies and connectivity as infrastructures in order to make all vital domains connected which allow citizens to interact comfortably and efficiently with their environment .It also helps to notify alert and raise awareness between people of how precious our environment is. However, in polluted areas this interaction is being jeopardized which can badly affect our welfare.

* Corresponding author. Tel.:+212 60 55 79 39;
E-mail address: essamlali.ismail01@gmail.com

1877-0509 © 2021 The Authors. Published by Elsevier B.V.
This is an open access article under the CC BY-NC-ND license (https://creativecommons.org/licenses/by-nc-nd/4.0)
Peer-review under responsibility of the Conference Program Chairs
10.1016/j.procs.2021.12.281

Inhaling fresh air, drinking good water, living in an area free of noise and where garbage is managed and valorized is an aspiration shared by scientists, policy makers and citizens who are mobilized to find rapid and smart solutions.
Pollution can be defined as the introduction of contaminants into the natural environment that cause adverse change [1]. In urban areas, pollution is mainly of human origin and it's up to humans to solve it. Air, water, waste and urban traffic and noise constitute real issues where pollution prevail the most.

We estimate that 4.2 million deaths every year occur as a result of exposure to ambient (outdoor) air pollution and 91% of the world's population live in places where air quality exceeds guideline limits.[2]
Concerning water, globally, at least 2 billion people use a drinking water source contaminated with faeces. Contaminated water can transmit diseases such as Diarrhoea, Cholera, Dysentery, Typhoid, and Polio. Drinking contaminated water is estimated to cause 485 000 diarrhoeal deaths each year. [3]

As for waste, the world generates 2.01 billion tonnes of municipal solid waste annually, with at least 33 percent of that—extremely conservatively—not managed in an environmentally safe manner. Worldwide, waste generated per person per day averages 0.74 kilogram but ranges widely, from 0.11 to 4.54 kilograms [4].

According to a European Union (EU) publication about Traffic noise[5] :

- about 40% of the population in EU countries is exposed to road traffic noise at levels exceeding 55 db(A);
- 20% is exposed to levels exceeding 65 dB(A) during the daytime;
- more than 30% is exposed to levels exceeding 55 dB(A) at night [5]

Reading these numbers, it seems that it is urgent to take actions in order to reduce the impacts of pollution and avoid provoking it in urban areas, that's why scientists must intervene upstream. In the last decade we notice the affluence of articles treating issues of pollution using artificial intelligence, a rich field permitting dealing with pollution by predicting it.

The rest of this paper is organized as follows. Concerning Section 2, we present some applications of machine learning in city pollution. As for Section 3 we discuss briefly the current work. Finally, we conclude this paper with the fourth section.

2. Machine learning and city pollution

This section has been divided into two parts. The first part deals with supervised machine learning techniques and pollution while the second part treats the unsupervised ones, each part gives a brief overview in four chosen fields of pollution: air, water, urban noise and garbage.

- **air pollution**:

The authors An Wang et al attempt to show the limits of LUR approaches compared with machine learning models that offer a great potential. Artificial Neural Networks (ANN) and gradient boost are developed and applied with the help of mobile data, PM2.5 and BC values are gathered between March and June 2019 in downtown Toronto, Canada, the two machine learning models used showed a better performance than LUR approaches [6].The same thing happened in [7] where ANN provided the best results (R 2 ¼ 0.58,RMSE ¼ 20.8, MAE ¼ 14.4) in predicting PM10 concentrations from a gathered data from 7 stations in Ankara province in Turkey , the data trained-in addition to ANN- with other machine learning algorithms LASSO, SVR, RF, kNN, xGBoost.

Considering the unsupervised techniques, the authors in [14] combine combine the Deep Belief Networks (DBN) model and a one-class support vector machine (OCSVM) in order to get unlabelled ozone measurements, for the nonlinear variations in the ground -level ozone concentrations we use DBN while OCSVM IS USED to detect unusual ozone measurement, the DBN-OCSVM model shows its efficiency in detecting ozone measurement compared to RBM and DSA-based K-Means, EM and Birch algorithms.

Unsupervised model combined with supervised model gives birth to semi-supervised model, that's what propose the authors in [15], the aim is to detect PM2.5 concentrations. As an unsupervised feature learning method we use EMD to decompose the data and extract the frequency and amplitude features while BiLSTM is used as a supervised learning method ,the China National Environmental Monitoring Center provide the dataset , the results show that the model surpass the standard LSTM- based model using four indicator values hourly (RMSE:6.86 $\mu g/m^3$, MAE:4.92 $\mu g/m^3$, MAPE:10.66%, R^2 : 0.989) and daily (RMSE: 22.58 $\mu g/m^3$, MAE: 16.67 $\mu g/m^3$, MAPE: 60.87%, R^2: 0.742) scales.

- **water quality**:

The Two Classification Algorithms Decision Tree and Support Vector Machine were considered in many articles to assess water quality, Decision tree performed better than SVM and naive bayes with an accuracy of 98.50% [8]. They implemented the algorithms based on, pH, DO, BOD and electrical conductivity used as parameters. While SVM had a superior performance in mapping the salinity of groundwater in [9] compared to other models, namely, flexible discriminant analysis (FDA), mixture discriminant analysis (MAD), boosted regression tree (BRT), multivariate adaptive regression spline (MARS) and random forest (RF) in mapping the salinity of groundwater.

In [16] a system based on K-means clustering and Gated Recurrent Unit (GRU) neural network is proposed to predict dissolved oxygen. K-means produces DO times series while GRU is used for the DO prediction. and we use PCA to select the elements impacting the fluctuation of DO. The average absolute error of the 30-min interval model was 0.264, and the mean absolute percentage error was 3.5 %, therefore the results show that the model can attain a better prediction accuracy and flexibility than the conventional approach.

Concerning groundwater Omolbani Mohammadrezapour et al propose in [17] a model of two clustering algorithms fuzzy c-means (FCM) and K-means. The data is constituted of 14 hydrochemical parameters from 108 wells sampled in 2016, Golestan province, northeast of Iran, different clusters are obtained such as optimal clusters, clusters in unfavorable condition in terms of TDS and CL and clusters for agricultural use based on the electrical conductivity (EC). Based on uncertainty conditions in determining the class boundary, the FCM algorithm outperforms K-means.

- **Garbage and waste management:**

In [10] KNN presents an accuracy of 93.3%, in generating an alert message from various combinations of biodegradable , non-biodegradable and concentration of poisonous gas, the authors build the simulation using pandas library to import csv file and sklearn or Scikit-Learn to use machine learning library for technique like KNN classifier using python. KNN also used in [11] with SVM, NB and RF to predict the accuracy of sending alert messages about dustbin level, metal level and poisonous gas level .RF has the best forecasting accuracy with 85.29%

For the purpose of monitoring and optimizing garbage collection, authors use in [18] K-means that determine the times where to proceed to the clean up and locate places where to put new dustbins for optimization. The dustbin's level is detected with a sensor and the servers are informed if the threshold is exceeded by a developed online application programming interface (API), which also stores other parameters like fill time, cleanup time and location.

Categorizing waste into organic and recycled types is proposed in [19] . They associate autoencoders and CNNs. Autoencoder reconstructs the dataset used for the classification and CNNs extracts the feature sets, and SVM was used as a classifier. authors have been able to achieve an accuracy of 99.95%.

- **Urban noise-traffic** :

The question of the existence of a link between air pollution and noise pollution is always approached, for this purpose a spatio-temporal relationship has been studied in [12] in four different locations. The concentration of CO, CO2 NO2, PM2.5, humidity, temperature and intensity of noise are used as key parameters and Pearson correlation confirms a strong association between the two types of pollution. In order to examine the effect of noise pollution on predicting air quality, three machine learning models are used: Decision Tree, Random Forest and K-Nearest Neighbors, a prediction accuracy of up to 95% was obtained and random forest performed the best.

Traffic not only plays a big part in air pollution by the vehicles emissions but contribute in urban noise also, in [13] the authors aim to predict the annoyance caused by traffic-noise using machine learning technique ANN provide the best results obtaining 42% and 35% error reduction in training subsets compared to the MRL and SVM models, respectively

In [20] K-means and K- medoids are implemented to classify the driving-style of conductors by exploiting Basic Safety Messages (BSMs) generated by connected vehicles, a groups of aggressive, normal and calm driving mode are determined , the authors remark that the threshold aggressive and calm driving fluctuate due to the disparity of roadway types and changes in environment conditions and in [21] density based clustering methods is associated with fuzzy C-means to hotspots, the DBSCAN algorithm is used for data preprocessing while fuzzy C-means divides the urban area into four clusters. Even if FCM is unsupervised machine learning, here the authors determine the number and characteristics of clusters which the model should a priori know.

Table 1. Machine learning techniques and pollution.

		Article	Author	Date	Publisher	algorithm
supervised	Air	[6]	An Wang et al	2020	ELSEVIER	ANN/GBOOST
		[7]	Aslı Bozdağ et al	2020	ELSEVIER	LASSO,SVR,RF, KNN,xGBoost,ANN
	Water	[8]	Neha Radhakrishnan, Anju S Pillai	2020	IEEE	SVM,DT,Naïve Bayes
		[9]	Amir Hossein Mosavi , et al	2020	SPRINGER	(FDA),(MAD),(BRT),(MARS),(RF), (SVM)
	Garbage	[10]	Sonali Dubey et al	2020	ELSEVIER	KNN
		[11]	Sahana Parvin Miquit et al	2018	IEEE	KNN
	Urban traffic and noise	[12]	Arindam Ghosh et al	2018	IEEE	DT, RF, KNN.
		[13]	Luis Bravo-Moncayo et al	2019	ELSEVIER	(ANN), (SVM) and (MLR)
	Air	[14]	Fouzi Harrou et al	2018	IEEE	DBN,OCSVM

unsupervised		[15]	Luo Zhang et al	2020	ELSEVIER	(EMD,BiLSTM)
	Water	[16]	Xinkai Cao et al	2020	ELSEVIER	K-means ,GRU
		[17]	Omolbani Mohammadrezapour et al	2018	SPRINGER	Fuzzy c-means , K-means
	Garbage	[18]	Shinjini Ray, et al	2018	IEEE	K-means
		[19]	Mesut Toğaçar,et al	2019	ELSEVIER	Autoencoder ,CNN,SVM
	Urban noise and traffic	[20]	Amin Mohammad Nazar, et al	2020	ELSEVIER	K-means and K- medoids
		[21]	Dailiang Jin, et al	2018	IEEE	fuzzy C-means

3. Discussion

Thanks to Artificial intelligence scientists have been able to predict several forms of pollution; it constitutes a great tackle preventing the expansion of pollution. However AI can become a burden itself if not controlled environmentally, since it consumes a lot of electricity in processing systems, carbon emissions grow when the use of AI grows so the ecological footprint of AI should be evaluated in order to assess the real compact of AI on environment, in a new paper, researchers at the University of Massachusetts, Amherst, performed a life cycle assessment for training several common large AI models. They found that the process can emit more than 626,000 pounds of carbon dioxide equivalent—nearly five times the lifetime emissions of the average American car [22], which makes us question the sustainability of artificial intelligence and how green it is.

4. Conclusion

Since the level of pollution is growing, the need for the use of artificial intelligence is growing also. The above paper discusses some of the applications of machine learning to combat pollution in cities. Four fields are studied: air pollution, water quality, garbage, urban noise and traffic. A diversity of machine learning techniques are applied in outdoor and indoor air pollution detection, water of surface and groundwater, collecting and valorizing waste and in urban noise and traffic which demonstrates that artificial intelligence can be truly trusted in fighting against pollution. The techniques have been evaluated and compared using metrics i.e. MAE, RMSE, MAPE, MSE etc,and the accurate model of each field of pollution is determined. In the future we will extend this research to include the majority of known machine learning algorithms applied in aforementioned fields, a comparison will be effectuated to determine the best and appropriate technique for the considered four fields.

References

[1] "Pollution – Definition from the Merriam-Webster Online Dictionary". Merriam-Webster. 2010-08-13. Retrieved 2010-08-26

[2] Who.int. 2021. Air pollution.

[3] Who.int. 2021. Drinking-water.

[4] Datatopics.worldbank.org. 2021. Trends in Solid Waste Management.

[5] Euro.who.int. 2021. Data and statistics.

[6] Wang, A., Xu, J., Tu, R., Saleh, M., & Hatzopoulou, M. (2020). "Potential of machine learning for prediction of traffic related air pollution". *Transportation Research Part D: Transport and Environment*, 88, 102599

[7] Bozdağ, A., Dokuz, Y., & Gökçek, Ö. B. (2020). "Spatial prediction of PM10 concentration using machine learning algorithms in Ankara, Turkey." *Environmental Pollution*, 263, 114635.

[8] Radhakrishnan, N., & Pillai, A. S. (2020, June). "Comparison of Water Quality Classification Models using Machine Learning." *In 2020 5th International Conference on Communication and Electronics Systems (ICCES)* (pp. 1183-1188). IEEE.

[9] Mosavi, A., Hosseini, F. S., Choubin, B., Taromideh, F., Ghodsi, M., Nazari, B., & Dineva, A. A. (2021). "Susceptibility mapping of groundwater salinity using machine learning models." *Environmental Science and Pollution Research*, *28*(9), 10804-10817.

[10] Dubey, S., Singh, P., Yadav, P., & Singh, K. K. (2020). "Household waste management system using iot and machine learning." *Procedia Computer Science, 167*, 1950-1959.

[11] Muquit, S. P., Yadav, D., Bhaskar, L., & Ahmed, W. F. (2018, February). "IoT based Smart Trash Bin for Waste Management System with Data Analytics." *In 2018 International Conference on Communication, Computing and Internet of Things (IC3IoT)* (pp. 137-142). *IEEE.*

[12] Ghosh, A., Pramanik, P., Banerjee, K. D., Roy, A., Nandi, S., & Saha, S. (2018, November). "Analyzing Correlation Between Air and Noise Pollution with Influence on Air Quality Prediction." In *2018 IEEE International Conference on Data Mining Workshops (ICDMW)* (pp. 913-918). *IEEE.*

[13] Bravo-Moncayo, L., Lucio-Naranjo, J., Chávez, M., Pavón-García, I., & Garzón, C. (2019) "A machine learning approach for traffic-noise annoyance assessment." *Applied Acoustics, 156*, 262-270.

[14]Harrou, F., Dairi, A., Sun, Y., & Kadri, F. (2018) "Detecting abnormal ozone measurements with a deep learning-based strategy." *IEEE Sensors Journal*, *18*(17), 7222-7232.

[15] Zhang, L., Liu, P., Zhao, L., Wang, G., Zhang, W., & Liu, J. (2021) "Air quality predictions with a semi-supervised bidirectional LSTM neural network." *Atmospheric Pollution Research, 12*(1), 328-339.

[16] Cao, X., Liu, Y., Wang, J., Liu, C., & Duan, Q. (2020) "Prediction of dissolved oxygen in pond culture water based on K-means clustering and gated recurrent unit neural network." *Aquacultural Engineering*, *91*, 102122.

[17] Mohammadrezapour, O., Kisi, O., & Pourahmad, F. (2020) "Fuzzy c-means and K-means clustering with genetic algorithm for identification of homogeneous regions of groundwater quality." *Neural Computing and Applications, 32*(8), 3763-3775.

[18] Ray, S., Tapadar, S., Chatterjee, S. K., Karlose, R., Saha, S., & Saha, H. N. (2018, January) "Optimizing routine collection efficiency in IoT based garbage collection monitoring systems." *In 2018 IEEE 8th Annual Computing and Communication Workshop and Conference* (CCWC) (pp. 84-90). *IEEE.*

[19] Toğaçar, M., Ergen, B., & Cömert, Z. (2020) "Waste classification using AutoEncoder network with integrated feature selection method in convolutional neural network models." *Measurement*, *153*, 107459.

[20] Mohammadnazar, A., Arvin, R., & Khattak, A. J. (2021) "Classifying travelers' driving style using basic safety messages generated by connected vehicles: Application of unsupervised machine learning." *Transportation research part C: emerging technologies, 122*, 102917.

[21] Jin, D., Zhao, X., & Pang, L. (2018, June) "Track mining based on density clustering and fuzzy C-means." *In 2018 IEEE 20th International Conference on High Performance Computing and Communications; IEEE 16th International Conference on Smart City; IEEE 4th International Conference on Data Science and Systems (HPCC/SmartCity/DSS) (pp. 1429-1434). IEEE.*

[22]Training a single AI model can emit as much carbon as five cars in their lifetimes. (2021). *Retrieved 21 June 2021, from https://www.technologyreview.com/2019/06/06/239031/training-a-single-ai-model-can-emit-as-much-carbon-as-five-cars-in-their-lifetimes/*