ENSEMBLE DEEP LEARNING FOR CLASSIFICATION OF POLLUTION PEAKS

PHUONG N. CHAU[1], RASA ZALAKEVICIUTE[2] & YVES RYBARCZYK[1]
[1]School of Information and Engineering, Dalarna University, Sweden
[2]Grupo de Biodiversidad Medio Ambiente y Salud (BIOMAS), Universidad de Las Américas, Ecuador

ABSTRACT
The concentration peaks of atmospheric pollutants are the most challenging and important phenomena in air quality forecasting. The fact that these elevated levels of pollution do not seem to follow any specific pattern explains why current models still struggle to provide an accurate prediction of these harmful events for human health. The present study tackles this issue by testing several supervised learning methods to discriminate between peak and no peak of concentrations of five contaminants: NO_2, CO, SO_2, $PM_{2.5}$, and O_3. The classification performance of ensemble decision tree (gradient boosting machine (GBM)) models and ensemble deep learning (EDL) models are compared. The results reveal that the EDL outperforms the GBM model. An analysis of the variable importance (SHapley additive exPlanations (SHAP)) shows that both temporal and meteorological features have an impact on the proposed models. In particular, time of day and wind speed are the most important features to explain the performance of the ensemble DL models.
Keywords: machine learning, deep learning, air pollution forecasting, data-driven modelling.

1 INTRODUCTION

Traditionally, air quality modelling has been done using atmospheric chemical transport models (CTMs), which provide a powerful framework for describing emission patterns, meteorology, and chemical transformation processes [1]. More recently, machine learning (ML) modelling has demonstrated its efficiency and reliability in predicting pollutant concentrations in the atmosphere [2]–[4]. Moreover, Grange et al. used ML and meteorological parameters to develop weather normalized models (WNMs) for developing air contaminant prediction models [5]. Rybarczyk and Zalakeviciute [4], Barré et al. [2] and Ceballos-Santos et al. [3] developed WNMs using gradient boosting machine (GBM) to simulate and quantify the effects of human activities on the environment in the context of COVID-19 lockdowns.

The past years have seen the rapid developments of deep learning (DL) models that have been applied to both timeseries data and air quality modelling. Particularly, long-short term memory (LSTM) model was developed in 1997 by Hochreiter and Schmidhuber [6] to capture both long- and short-term memory from the sequence data. Afterwards, LSTM models have been more adapted for time series prediction [7], [8] and have also been applied for predicting CO, NO_2, O_3, PM_{10}, SO_2 and pollen concentrations in Madrid [9]. In 2019, Krishan et al. used LSTM to develop air quality models for India [10]. After recognizing the limitations of LSTM, Schuster and Paliwal combine two hidden LSTM layers in the opposite direction to create the bidirectional recurrent neural network (BiRNN) [11]. It has been demonstrated to be an effective DL architecture for sequence data with two hidden LSTM layers in opposite directions. In 2019, BiRNN was used by Li et al. [12] to capture deeper characteristics of timeseries data. Besides, Tong et al. used BiRNN for modelling the $PM_{2.5}$ concentrations and exploring the correlations of spatiotemporal properties [13]. Also, Zhang et al. developed a hybrid model based on BiRNN for $PM_{2.5}$ forecasting, which outperformed the traditional models [14].

WIT Transactions on Ecology and the Environment, Vol 259, © 2022 WIT Press
www.witpress.com, ISSN 1743-3541 (on-line)
doi:10.2495/AWP220031

Although the performance of air quality forecasting is improving, the concentration peaks seem to be extremely difficult to handle or predict [15]–[17]. The reason is that elevated levels of pollutants do not seem to follow any specific pattern. Consequently, the current models continue to struggle to provide an accurate prediction of harmful air quality. In addition, when we can determine the concentration peak, the other concentrations are equal or lower than the peak value. Therefore, the air pollutant trend is not going up after the highest concentration, and we can implement many strategies to improve the predicted models such as external features. A model able to detect peaks can provide early warnings of air pollutant concentrations exceeding the WHO guidelines [18].

This study addresses the daily peaks classification problem of five air pollutants (NO_2, SO_2, CO, O_3 and $PM_{2.5}$) by proposing advanced techniques, based on ensemble decision tree GBM and ensemble deep learning (EDL), to automatically classify peak vs no peak of pollutant concentrations. Additionally, we leveraged SHapley Additive exPlanations (SHAP) to explore the variable correlations in the EDL models. This method enables us to discover the principal factors responsible for the daily peak of concentrations.

The main goal of this study is to create supervised learning models that use data-driven techniques to classify the daily air pollutants according to the peaks of concentration. The remainder of this paper includes four sections. Section 2 describes the study site, data collection, and data processing. Section 3 depicts the supervised methods. The results and discussion are presented in Section 4. Finally, this research is summarized and concluded in the last section.

2 MATERIALS

2.1 Study site

Bellisario (2,835 meters above sea level (m.a.s.l), coord. 78°29'24" W, 0°10'48" S) study site is one of the monitoring stations in Quito located on a school and in a heavily populated area of the capital city of Ecuador. The weather in the study site is mild and consistent throughout the year. These characteristics are caused by the city's two distinct seasons. The wet season is from September to May, and the dry season is between June and August [19]. Furthermore, this city is a high elevation city established at 2,850 m.a.s.l. with a population of about 2.2 million people in 2011 [20], [21]. Due to 30% reduction of oxygen concentrations at this altitude, traffic emissions are the principal causes of long-term pollution problems for this urban center [19], [22].

2.2 Data

The data is provided by the Secretariat of the Environment of the Municipality of the Metropolitan District of Quito. All devices follow the Environmental Protection Agency of the United States (US-EPA) standards. The data is from a central urban study site Bellisario with five pollutants, namely NO_2, CO, SO_2, $PM_{2.5}$, O_3 and seven meteorological features such as solar radiation (SR), wind direction (WD), wind speed (WS), atmospheric pressure (p), precipitation (Prec), temperature (T) and relative humidity (RH). Additionally, we created four temporal features such as "Julian day" (or day of the year), "week day" (day of the week), hours (the time of the day), and index (i.e., starting from 1 January 2014 and increasing by one at each instance). According to research by Grange et al. [5] and Henneman et al. [23], the temporal variables are independent features in meteorological normalization models. "Julian day", "week day", and "hours" are indicators for emission pattern cycles

rather than direct influences on levels of pollutants in the atmosphere. “hours”, for example, is a term used to explain cyclic emissions such as traffic-related rush hour emissions. “week day” represents weekdays or weekends in a week. Meanwhile, “Julian day” is a seasonal term that strongly presents seasonal emissions or pollutant concentrations.

2.3 Pre-process data

We selected the data for a total of five years from 1 January 2014 to 31 December 2018. The instance is removed from the dataset if the missing value is pollutant concentrations. The missing values in meteorological features are replaced by their averages. Since the filled values in this phase are less than 1%, it allows us to use the average method for imputation [24], [25]. Next, the output features will be labelled with two labels: 0 and 1. The daily peak concentration is labelled “1”, and the label “0” is randomly selected for another concentration on the same day. We chose the label “0” randomly, but it must differ from the highest daily concentration. It allows us to generate balanced datasets with two classes.

3 METHODS

Fig. 1 illustrates an overview of our process. First, we split the data into training and testing set with 80% (four years from 1 January 2014 to 31 December 2017) for training and 20% (one year from 1 January 2018 to 31 December 2018) for testing. Second, we develop EDL models based on LSTM and BiRNN architectures for classifying the peak concentrations with five air pollutants (NO_2, SO_2, CO, O_3 and $PM_{2.5}$) in Section 3.2. Additionally, we develop the GBM model for this problem in Section 3.1. Next, the GBM and EDL models are compared, which is based on the assessment metrics (F1-score and AUC score) to evaluate the classification performance (Section 3.3). The models with the highest F1 and AUC scores are selected as the best models for classification problems. According to the best models, we can use it to identify the daily peak values of air pollutants. Besides, it allows us to explore the variable correlations and discover the significant factors in the pollutant concentration peaks with SHAP in Section 3.4.

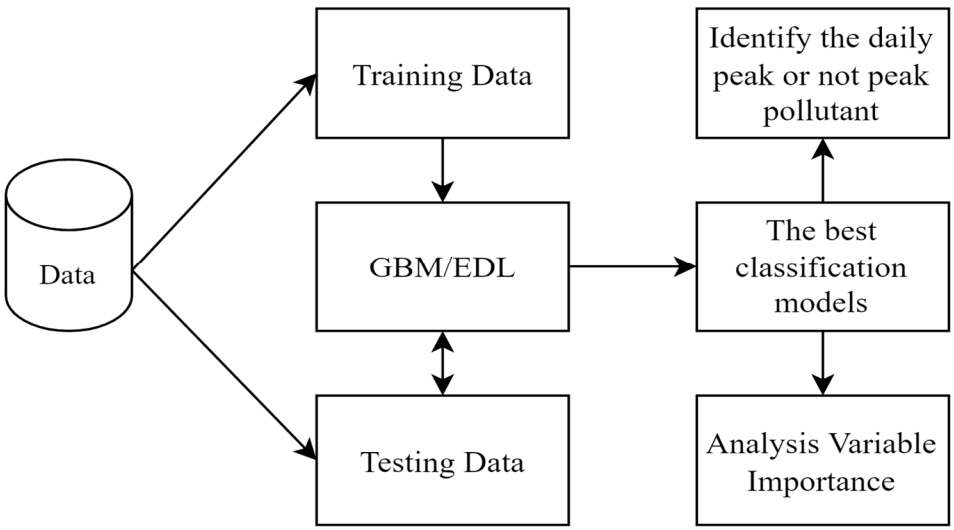

Figure 1: Workflow of our process.

3.1 Gradient boosting machine

Friedman proposed GBM in 2001 as a decision tree-based ensemble learning algorithm [26]. Typically, an ensemble model is developed to obtain a better generalization model from weak learning models. GBM algorithm builds regression trees for all features from the dataset sequentially, which means that the trees are built independently. When all trees are generated, the final output from this algorithm is obtained by eqn (1). $\hat{f}^n(x)$ is the output from regression tree n. Considering its high predictive power, more and more authors have been using GBM to predict air quality [3], [4].

$$\hat{f}(x) = \sum_{n=1}^{N} \hat{f}^n(x). \quad (1)$$

All experiments are implemented with the Scikit-learn library (version 0.23.2). The tuning parameters are presented in Table 1. For this Scikit-learn version, we defined the learning rate, max_depth and random_state for all GBM models. The best results from GBM models are reported in the results section. The learning_rate was tuned to 0.05 to satisfy the convergence criterion quickly. The other parameters were used as the default values of the Scikit-learn library.

Table 1: Parameters for GBM in Scikit-learn library.

Parameters	Values
learning_rate	0.05
max_depth	5,10,15,30
random_state	1234

3.2 Ensemble deep learning

The EDL model is composed of given sub-DL models. Sub-model 1 and 2 are based on LSTM and BiRNN, respectively. While sub-model 3 is a combination of two LSTM layers, sub-model 4 is a combination of two BiRNN layers. Finally, sub-model 5 is a hybridization of BiRNN and LSTM layers.

In Fig. 2, the input data include the meteorological and temporal features. Next, 20 models (five sub-models × four "the number of nodes" in Table 2) were created for each pollutant. Afterwards, we select the top three best sub-models in the "meta learner" phase. The final output is computed by eqn (2), and this is the average output from these sub-models. Note that $\hat{f}^n(x)$ is the output from the three best sub-models. Therefore, N has a value equal to three.

$$\hat{f}(x) = \frac{1}{N}\sum_{n=1}^{N} \hat{f}^n(x). \quad (2)$$

The parameters for all the DL models are shown in Table 2. Tensorflow library, version 2.3.0, was used for all the DL sub-models. "The number of nodes" are changed to find the best model for each pollutant. "Epochs" parameter is a condition to stop the model. Specifically, the DL models will stop after 300 iterations if the model cannot obtain global optimization or improve the mean square error (MSE). By contrast, the model will cease the training phase by the "patience" parameter. Therefore, the training model finishes if the model accuracy cannot improve after 50 iterations (patience = 50). The "drop out" (20%) eliminates many connections between two layers to reduce the overfitting. The learning rate

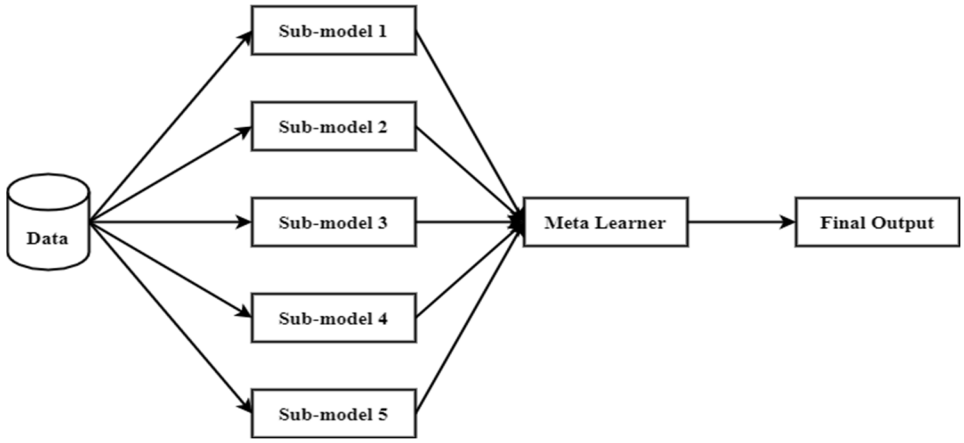

Figure 2: EDL model with its five sub-models.

Table 2: Parameters for all the DL models.

Parameters	Values
The number of nodes	16, 32, 64, 128
Patience	50
Drop out	0.2
Loss Function	mse
Learning_rate	0.05
Batch size	500
Epochs	300

is 0.05, which improves the convergent speed of the DL models. "Batch size" controls the number of training samples to update the weights in DL models.

3.2.1 Sub-model 1 and sub-model 2

In these models, the first layer is the input layer, which includes the meteorological and temporal features (Fig. 3). This layer transforms the raw format to the DL layer formats. In the sub-model 1, the next layer is the LSTM layer, which consists of many LSTM cells. The number of the LSTM cells are defined as "the number of nodes" (see Table 2). A "drop out" layer is introduced to reduce the overfitting of the model. Next, a dense layer with 20 nodes connects outputs from the "drop out" layer to the output layer. Since this is a binary classification, the Output Layer is a node with the sigmoid activation function to determine if the output belongs to peak ("1") or not peak ("0") levels of pollutant concentrations. The sub-model 2 (BiRNN) has a similar design to the LSTM sub-model. However, the LSTM cells and LSTM layer are replaced by the BiRNN cells and BiRNN layer.

3.2.2 Sub-model 3, sub-model 4 and sub-model 5

Fig. 4 represents the architecture of sub-model 3 (LSTM–LSTM) with its six layers. The input, "LSTM layer 1", "drop out layer", and the output layers are alike sub-model 1. We also tune the number of nodes for "LSTM layer 1". However, we added one more "LSTM

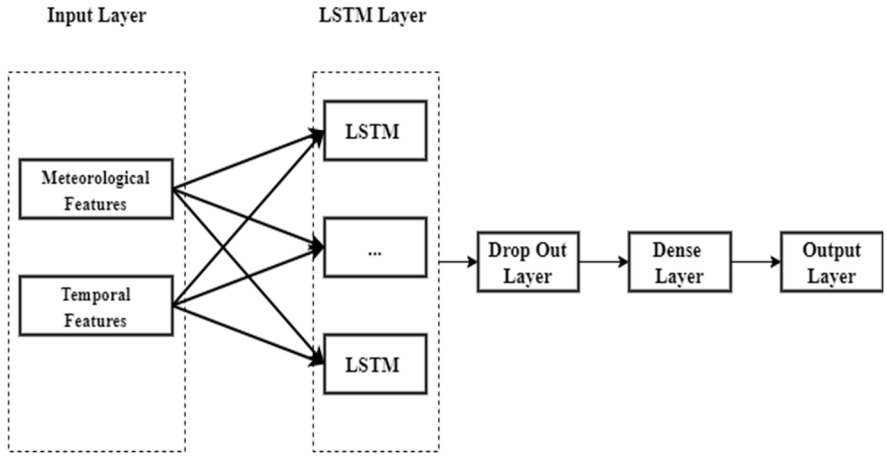

Figure 3: Sub-model 1 (LSTM).

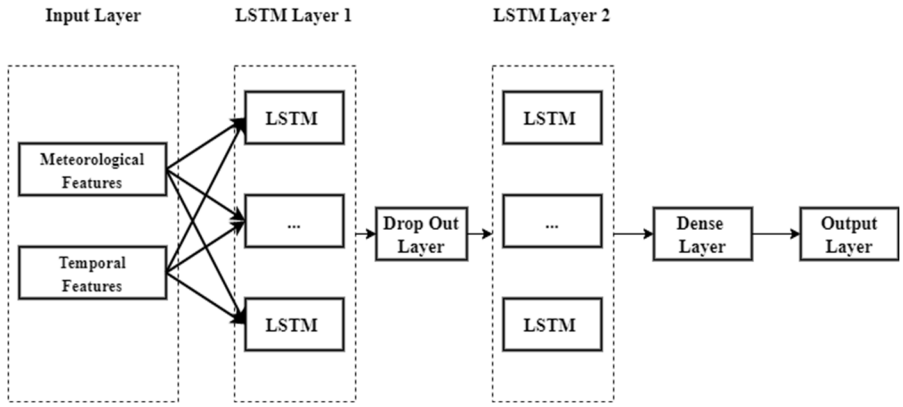

Figure 4: Sub-model 3 (LSTM–LSTM).

Layer 2" with 50 LSTM cells to obtain deeper information from the data. This layer is between the "drop out layer" and "dense layer". Additionally, dense layer has ten nodes and do not change when the parameters are tuned. In sub-model 4, the main difference is that the BiRNN cells in both BiRNN layers of sub-model 4 take the place of the LSTM cells in sub-model 3. Sub-model 5 is similar to sub-model 3, except for the fact that "LSTM layer 1" is BiRNN layer. The number of nodes in sub-model 4 and 5 is tuned similar to sub-model 3.

3.3 Metrics

Two metrics are used to compare the performance of each model: area under the ROC curve (AUC) and F1 score (or F-score). ROC is defined as a curve plotted from two parameters, which are the true positive rate (TPR) and false positive rate (FPR) resulting from the classification. The area under the ROC curve is the AUC value, and ranges from 0.5 to 1. *TPR* and *FPR* are computed by true positive (TP), false negative (RN), false positive (FP)

and true negative (TN) as described in eqns (3) and (4). F1 score (eqn (7)) is based on the calculation of the *Precision* (eqn (5)) and the *Recall* (eqn (6)). The range values of the F1 score are from 0 to 1. Both metrics (AUC and F1) are calculated by the Scikit-learn library (version 0.23.2). The closer to 1, the better model is.

$$TPR = \frac{TP}{TP+FN}, \tag{3}$$

$$FPR = \frac{FP}{FP+TN}, \tag{4}$$

$$Precision = \frac{TP}{TP+FP}, \tag{5}$$

$$Recall = \frac{TP}{TP+FN}, \tag{6}$$

$$F1\ score = \frac{2 \times precision \times recall}{precision+recall}. \tag{7}$$

3.4 SHapley additive exPlanations

Although DL has high accuracy, it is a kind of black-box which limits its interpretation and the identification of the weight of each feature in the prediction. Currently, two popular methods were introduced to disclose this opacity. The first one, SHAP [27], can provide the variable importance of input features in the DL models. These values are computed in terms of the game theory knowledge. The second, local interpretable model-agnostic explanations (LIME) uses coefficients among features to explain the model performance. However, it is tricky to choose the correct parameters from the LIME method, which can lead to miss significant variables [28]. Hence, we use SHAP for variable importance in this study.

SHAP values will allow us to get a more in-depth understanding of the contributions of the meteorological and temporal features to classify the level of pollution. In the EDL models, we applied SHAP for each sub-model. The SHAP values of EDL are the averages of SHAP values from the entire sub-models.

4 RESULTS AND DISCUSSION

4.1 Performance

Table 3 represents the average performance of all models. Overall, the EDL models outperform the other models. It is to note that EDL models are better than both GBM and sub-models in CO (F1 score = 0.8094; AUC = 0.8025), O_3 (F1 score = 0.8775; AUC = 0.8712), SO_2 (F1 score = 0.6969; AUC = 0.7126) and $PM_{2.5}$ (F1 score = 0.7079; AUC = 0.7195). However, even if the accuracy of the EDL model is always better than GBM, its AUC is slightly lower than sub-model 1 for NO_2. The reason is that the average of the five sub-models contributes to the final outputs of the EDL, as the top three sub-models are used for the calculation. Therefore, each selected sub-model contributes one third to the final classification.

It is to highlight the fact that the performance of EDL is higher than 0.7 for both metrics (only the F1 score of SO_2 is approximately 0.7 (0.6969) [29]. On the contrary, GBM scores are lower than 0.7 for SO_2 (F1 score = 0.7065; AUC = 0.6690) and $PM_{2.5}$ (F1 score = 0.6749; AUC = 0.6731). Hence, it can be concluded the EDL models are more reliable than both GBM and simple deep learning models to discriminate between daily peak or not daily peak levels of pollution concentration.

Table 3: Performance of GBM, EDL models with sub-models on testing set.

			Sub-model					
Pollutant	Metrics	GBM	1	2	3	4	5	EDL
NO_2	F1 score	0.7596	0.7736	0.7686	0.7611	0.7730	0.7641	0.7737
	AUC	0.7521	0.7638	0.7529	0.7485	0.7562	0.7512	0.7622
CO	F1 score	0.7769	0.7945	0.7894	0.7937	0.7979	0.7971	0.8094
	AUC	0.7726	0.7871	0.7858	0.7830	0.7855	0.7860	0.8025
O_3	F1 score	0.8729	0.8725	0.8562	0.8717	0.8714	0.8698	0.8775
	AUC	0.8671	0.8658	0.8501	0.8658	0.8644	0.8627	0.8712
$PM_{2.5}$	F1 score	0.7065	0.6911	0.6882	0.6888	0.7018	0.7008	0.7079
	AUC	0.6690	0.7082	0.7003	0.7038	0.7176	0.7179	0.7195
SO_2	F1 score	0.6749	0.6915	0.6622	0.6980	0.6842	0.6793	0.6969
	AUC	0.6731	0.6984	0.6934	0.7019	0.6967	0.6926	0.7126

To summarize, the best model performance in classifying daily peaks was found for O_3 and CO. The O_3 contaminant is one of the easiest pollutants to predict due to its high dependence on solar radiation activity, which always peaks at noon anywhere on the planet [30]. Some variations may be common due to increased cloudiness; however, those are also easy to predict using trends of relative humidity and atmospheric pressure. Similarly, CO may be affected by photochemical oxidation [31]. However, a week to months lifetime of CO (e.g., Holloway et al.), makes it an easier pollutant to predict, due to its longer persistence in the urban airshed [32].

On the contrary, NO_2, $PM_{2.5}$ and SO_2 are a bit less predictable, as they highly depend on anthropogenic emissions, but may also come from other sources [33]. Moreover, specific environmental conditions might help transport the emissions, or, on the other hand, might help the accumulation of air pollutants in an urban canopy. Variations in sources and the impact of meteorological parameters must be considered in addition to the complex trends of urban mobile circulation. While SO_2 can be emitted from high sulfur content fossil fuel burning, it can also be emitted by the active Andean volcanos. This latter factor is a completely unpredictable phenomenon given the set of parameters used in this study. This might help explain the poorest performance for this air pollutant.

4.2 Variable importance

SHAP is an advanced library for analyzing the inter-correlations among features of the GBM and DL models. Therefore, we used SHAP to identify the main factors that affect the concentration peaks of pollution. Fig. 5 represents the mean of the SHAP values for the five best EDL models (one for each pollutant). The higher the SHAP value is, the more important is the feature in the predicting output.

The results show that both the meteorological and temporal variables played a significant role in the best EDL models. Especially, “hours” is the most important variable in all five pollutant peaks’ models, whereas the second and third crucial features are slightly different from one model to another. Since Bellisario is an urban area with heavy traffics, the pollution is highly affected by the time of the day. A previous study has identified two obvious pollution peaks at the rush hours (around 8 am and 6 pm) in the capital city of Ecuador [17].

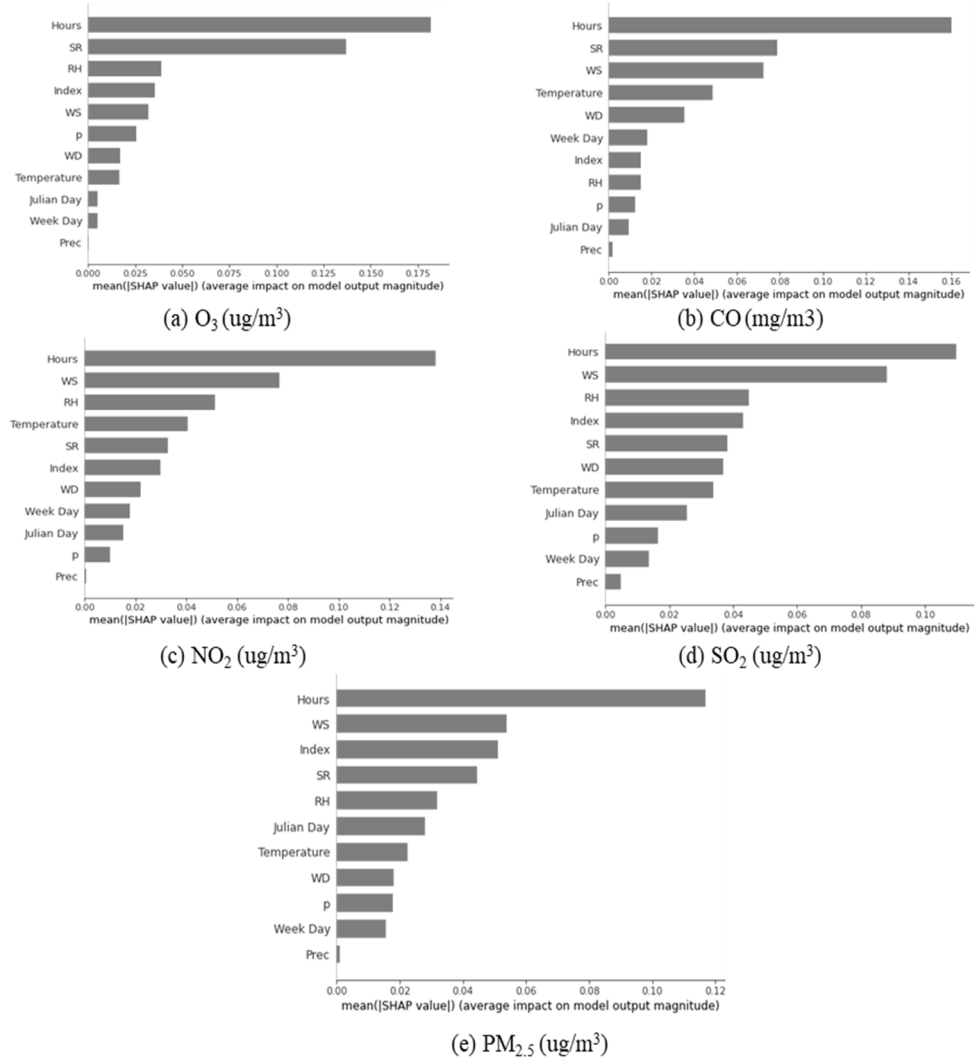

Figure 5: Variable importance with mean (SHAP value) for the best models of five pollutants.

As discussed above, the most important predictor in all the models is the "hours". Daily trends of human activities carry the most important role in predicting air quality levels in this problematic high elevation city. While most of the pollutants might show the increased levels during the rush hours (low atmospheric mixing and increased traffic), O_3 shows a midday peak and thus also highly depends on SR (mean SHAP values ≈ 0.135) and next on RH (mean SHAP values ≈ 0.04) that might indicate cloud cover.

The second and third most important variables for CO are SR (mean SHAP values ≈ 0.08) and WS (mean SHAP values ≈ 0.07). As discussed above, CO is affected by photochemical oxidation, explaining the importance of SR. Now, WS is extremely important for all

pollutants but O_3. WS is the above-mentioned ventilation parameter. An increase in wind speed can help ventilate the air pollutants, and transport them away from the emission sources, while the lack of it might create stagnant atmospheric conditions and generate an accumulation of anthropogenic urban emissions.

5 CONCLUSIONS

The prediction of atmospheric pollutant concentrations is always challenging to tackle, even if it is fundamental to anticipate on harmful effects of bad air quality on human health. It is even more difficult to forecast pollution peaks because they do not seem to follow any specific pattern. This study aims to determine the concentration peaks to get a better prediction of air pollution. Here, we propose an EDL approach to identify the daily peaks of concentration. The performance of EDL models ranges from 0.6969 to 0.8775, which demonstrates that the proposed method is promising for improving the prediction of pollution peaks. The highest accuracy is obtained for O_3, CO and NO_2 (above 0.75). As already reported in other studies, SO_2 and $PM_{2.5}$ are slightly more difficult to predict [34], [35]. According to the F1 and AUC metrics, we can conclude that the EDL outperforms both the traditional machine learning algorithm (GBM) and simple deep learning methods (LSTM and BiRNN) whatever the pollutants are considered.

The SHAP library allowed us to capture the interrelations among input features and pollutant concentrations in the EDL models. The time of the day (hours) – a marker of human activity tendencies – has been identified as the most significant feature in classifying the concentration peaks in all five contaminants. It can be explained by the fact that the capital city of Ecuador has two obvious daily peaks at the morning (8 am) and the evening rush hour (6 pm) [36]. Additionally, RH and WS are relevant factors in the NO_2 and SO_2 models. The second and third significant features depend on the nature of the pollutant. They are SR and RH for O_3, SR and WS for CO, WS and Index for $PM_{2.5}$. Further work will consist in using this classification to apply optimized models, specifically designed to predict high vs low concentrations.

ACKNOWLEDGEMENT

We thank the Secretariat of the Environment of the Municipality of the Metropolitan District of Quito for providing us with the air quality and meteorological data.

REFERENCES

[1] Steinfeld, J.I., Atmospheric chemistry and physics: from air pollution to climate change. *Environment: Science and Policy for Sustainable Development*, **40**(7), p. 26, 1998. DOI: 10.1080/00139157.1999.10544295.

[2] Barré, J., Petetin, H., Colette, A., Guevara, M., Peuch, V.H., Rouil, L., Engelen, R., Inness, A., Flemming, J., Pérez García-Pando, C. & Bowdalo, D., Estimating lockdown-induced European NO_2 changes using satellite and surface observations and air quality models. *Atmospheric Chemistry and Physics*, **21**(9), pp. 7373–7394, 2021.

[3] Ceballos-Santos, S., González-Pardo, J., Carslaw, D.C., Santurtún, A., Santibáñez, M. & Fernández-Olmo, I., Meteorological normalisation using boosted regression trees to estimate the impact of COVID-19 restrictions on air quality levels. *International Journal of Environmental Research and Public Health*, **18**(24), p. 13347, 2021. DOI: 10.3390/ijerph182413347.

[4] Rybarczyk, Y. & Zalakeviciute, R., Assessing the COVID-19 impact on air quality: A machine learning approach. *Geophysical Research Letters*, **48**(4), e2020GL091202, 2021. DOI: 10.1029/2020GL091202.

[5] Grange, S.K., Carslaw, D.C., Lewis, A.C., Boleti, E. & Hueglin, C., Random forest meteorological normalisation models for Swiss PM_{10} trend analysis. *Atmospheric Chemistry and Physics*, **18**(9), pp. 6223–6239, 2018.
[6] Hochreiter, S. & Schmidhuber, J., Long short-term memory. *Neural Computation*, **9**(8), pp. 1735–1780, 1997.
[7] Kuremoto, T., Kimura, S., Kobayashi, K. & Obayashi, M., Time series forecasting using a deep belief network with restricted Boltzmann machines. *Neurocomputing*, **137**, pp. 47–56, 2014.
[8] Ong, B.T., Sugiura, K. & Zettsu, K., Dynamically pre-trained deep recurrent neural networks using environmental monitoring data for predicting $PM_{2.5}$. *Neural Computing and Applications*, **27**(6), pp. 1553–1566, 2016.
[9] Navares, R. & Aznarte, J.L., Predicting air quality with deep learning LSTM: Towards comprehensive models. *Ecological Informatics*, **55**, 101019, 2020.
[10] Krishan, M., Jha, S., Das, J., Singh, A., Goyal, M.K. & Sekar, C., Air quality modelling using long short-term memory (LSTM) over NCT-Delhi, India. *Air Quality, Atmosphere and Health*, **12**(8), pp. 899–908, 2019.
[11] Schuster, M. & Paliwal, K.K., Bidirectional recurrent neural networks. *IEEE Transactions on Signal Processing*, **45**(11), pp. 2673–2681, 1997. DOI: 10.1109/78.650093.
[12] Li, Q., Ness, P.M., Ragni, A. & Gales, M.J., Bi-directional lattice recurrent neural networks for confidence estimation. *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 6755–6759, 2019.
[13] Tong, W., Li, L., Zhou, X., Hamilton, A. & Zhang, K., Deep learning $PM_{2.5}$ concentrations with bidirectional LSTM RNN. *Air Quality, Atmosphere and Health*, **12**(4), pp. 411–423, 2019. DOI: 10.1007/s11869-018-0647-4.
[14] Zhang, Z., Zeng, Y. & Yan, K., A hybrid deep learning technology for $PM_{2.5}$ air quality forecasting. *Environmental Science and Pollution Research*, **29**, pp. 39409–39422, 2021. DOI: 10.1007/s11356-021-12657-8.
[15] Catalano, M., Galatioto, F., Bell, M., Namdeo, A. & Bergantino, A.S., Improving the prediction of air pollution peak episodes generated by urban transport networks. *Environmental Science and Policy*, **60**, pp. 69–83, 2016.
[16] Dutot, A.L., Rynkiewicz, J., Steiner, F.E. & Rude, J., A 24-h forecast of ozone peaks and exceedance levels using neural classifiers and weather predictions. *Environmental Modelling and Software*, **22**(9), pp. 1261–1269, 2007.
[17] Zalakeviciute, R., Bastidas, M., Buenaño, A., Rybarczyk, Y., A traffic-based method to predict and map urban air quality. *Applied Sciences*, **10**(6), p. 2035, 2020.
[18] WHO, *WHO Global Air Quality Guidelines*, 2021. https://apps.who.int/iris/bitstream/handle/10665/345329/9789240034228-eng.pdf?sequence=1&isAllowed=y. Accessed on: 6 Jan. 2022.
[19] Zalakeviciute, R., López-Villada, J. & Rybarczyk, Y., Contrasted effects of relative humidity and precipitation on urban $PM_{2.5}$ pollution in high elevation urban areas. *Sustainability*, **10**(6), p. 2064, 2018. DOI: 10.3390/su10062064.
[20] EMASEO, De Quito, Municipio del Distrito Metropolitano. Plan de desarrollo 2012–2022, 2011. *Quito: Municipio de Quito*.
[21] INEC Q, Poblacion, superficie (km^2), densidad poblacional a nivel parroquial.

[22] Zalakeviciute, R., Rybarczyk, Y., López-Villada, J. & Suarez, M.V., Quantifying decade-long effects of fuel and traffic regulations on urban ambient $PM_{2.5}$ pollution in a mid-size South American city. *Atmospheric Pollution Research*, **9**(1), pp. 66–75, 2018. DOI: 10.1016/j.apr.2017.07.001.
[23] Henneman, L.R., Holmes, H.A., Mulholland, J.A. & Russell, A.G., Meteorological detrending of primary and secondary pollutant concentrations: Method application and evaluation using long-term (2000–2012) data in Atlanta. *Atmospheric Environment*, **119**, pp. 201–210, 2015.
[24] Enders, C.K., *Applied Missing Data Analysis*, Guilford Press, 2010.
[25] Lynch, S.M., *Introduction to Applied Bayesian Statistics and Estimation for Social Scientists*, Springer Science and Business Media, 2007.
[26] Friedman, J.H., Greedy function approximation: A gradient boosting machine. *Annals of Statistics*, **1**, pp. 1189–1232, 2001.
[27] Lundberg, S.M. & Lee, S.I., A unified approach to interpreting model predictions. *Advances in Neural Information Processing Systems,* **30**, 2017.
[28] Garreau, D. & Luxburg, U., Explaining the explainer: A first theoretical analysis of LIME. *International Conference on Artificial Intelligence and Statistics,* pp. 1287–1296, 2020.
[29] Luque, A., Carrasco, A., Martín, A. & de Las Heras, A., The impact of class imbalance in classification performance metrics based on the binary confusion matrix. *Pattern Recognition,* **91**, pp. 216–231, 2019.
[30] Juarez, E.K. & Petersen, M.R., A comparison of machine learning methods to forecast tropospheric ozone levels in Delhi. *Atmosphere*, **13**(1), p. 46, 2022. DOI: 10.3390/atmos13010046.
[31] Buchholz, R.R., Worden, H.M., Park, M., Francis, G., Deeter, M.N., Edwards, D.P., Emmons, L.K., Gaubert, B., Gille, J., Martínez-Alonso, S. & Tang, W., Air pollution trends measured from Terra: CO and AOD over industrial, fire-prone, and background regions. *Remote Sensing of Environment*, **256**, 112275, 2021. DOI: 10.1016/j.rse.2020.112275.
[32] Holloway, T., Levy, H. & Kasibhatla, P., Global distribution of carbon monoxide. *Journal of Geophysical Research: Atmospheres*, **105**(D10), pp. 12123–12147, 2020. DOI: 10.1029/1999JD901173.
[33] US EPA, Criteria air pollutants, 2022. https://www.epa.gov/criteria-air-pollutants. Accessed on: 13 Jan. 2022.
[34] Abdul-Wahab, S.A. & Al-Alawi, S.M., Prediction of sulfur dioxide (SO_2) concentration levels from the Mina Al-Fahal refinery in Oman using artificial neural networks. *American Journal of Environmental Sciences*, **4**(5), pp. 473–481, 2008. DOI: 10.3844/ajessp.2008.473.481.
[35] Li, J., Li, X., Wang, K. & Cui, G., Atmospheric $PM_{2.5}$ concentration prediction and noise estimation based on adaptive unscented Kalman filtering. *Measurement and Control*., **54**(3–4), pp. 292–302, 2021. DOI: 10.1177%2F0020294021997491.
[36] Rybarczyk, Y. & Zalakeviciute, R., Regression models to predict air pollution from affordable data collections. *Machine Learning: Advanced Techniques and Emerging Applications*, pp. 15–48, 2018. DOI: 10.5772/intechopen.71848.