



Air Quality Monitor and Forecast in Norway Using NB-IoT and Machine Learning

Andreas Lepperød¹, Hai Thanh Nguyen^{1,2}(✉), Sigmund Akselsen²,
Leendert Wienhofen², Pinar Øzturk¹, and Weiqing Zhang²

¹ Norwegian University of Science and Technology, Trondheim, Norway
hai.nguyen@ntnu.no

² Telenor ASA, Fornebu, Norway

Abstract. In recent years, air quality has become a significant environmental and health related issue due to rapid urbanization and industrialization. As a consequence, real-time monitoring and precise prediction of air quality gained increased importance. In this paper, we present a complete solution to this problem by using NB-IoT (Narrowband-Internet-of-Things) sensors and machine learning techniques. This solution includes our own compiled cheap micro-sensor devices that are planned to be deployed at stationary locations as well as on the moving vehicles to provide a comprehensive overview of air quality in the city. We developed our own IoT data and analysis platform to support the gathering of air quality data as well as weather and traffic data from external sources. We applied seven machine learning methods to predict air quality in the next 48-h, which showed promising results. Finally, we developed a mobile application named Lufta, which is now available in Google play for testing purposes.

Keywords: Air quality · Internet of Things · Machine learning

1 Introduction

Air quality has increasingly attracted attention from environment managers and citizens all over the world. New tools continue to emerge to raise air quality awareness worldwide. Continuous improvements in air quality monitoring are occurring along with the advancements of smart cities and with the rapidly increasing deployment of internet-of-things sensor devices. As a natural consequence, air pollution forecasting has become a hot topic, aiming the prediction of the atmospheric composition of pollutants at a given time and location. With an accurate air quality forecast, individuals can take action to reduce the possible adverse effects of air pollution on their health, such as choosing the cleanest routes for the commute or the best time for outdoor activities. From the

policy-making perspective, accurate forecasting contributes to better planning and establishment of procedures to reduce the severity of local pollution levels.

Much effort has already been made by researchers to create accurate forecasting models capable of fitting the underlying time series, which is challenging for various reasons. Often, air quality prediction involves a noisy and limited amount of historical data, mainly due to the poor quality of sensors used. Furthermore, the prediction of a single observation usually depends on many factors, such as weather conditions, traffic flow, time of the week, and so on. Besides, the air changes rapidly in short time frames, with hourly data being more uncertain compared with monthly and yearly trends and seasonality. These problems make it hard to generalize a created model to be transferable to other locations.

We studied air quality monitoring and forecasting in Trondheim, one of the largest cities in Norway. Typically, and similar to many other cities in Scandinavia, the air quality in Trondheim is on average at a healthy level but has periods with severe pollution, especially in the winter months. This is mainly because people drive a lot and use wood-burners during wintertime. Also, municipalities often put sand on the roads to make them less slippery under conditions with snow and ice.

We developed a complete solution for air quality monitoring and forecasting using Narrowband IoT (NB-IoT) and machine learning. Our holistic IoT solution contains self-compiled micro-sensor devices, IoT data platform, and analytics tools [1]. The solution aggregates different data sources and performs air quality prediction by using machine learning methods. We also developed a mobile application named Lufta [2] to visualize the air quality data as well as to give users forecast information.

Our study demonstrates the benefits of machine learning for predicting general patterns of air pollutants and foresees sudden spikes of a high pollution level. The study has tested seven different machine learning algorithms for modeling and forecasting the pollution of $PM_{2.5}$, PM_{10} , and NO_2 . The data of pollutants, and meteorological and traffic data with statistical temporal-spatial feature engineering were taken into account to build models for multi-step-ahead air quality forecasts for 24 and 48-h. Results express that ensemble techniques could significantly improve the stability and accuracy of the prediction of the general trends of air quality. Among the ensemble techniques, gradient boosting with dropouts resulted in prediction errors with the lowest deviation. For the prediction of sudden changes in air pollution, using a recurrent neural network with a memory unit provided the highest accuracy of classified spikes. Lastly, the machine learning results were compared with that of the national air quality services, which uses a knowledge-driven model. The predictions of general patterns and anomalies are shown to be superior for 24-h, and more comparable results for the 48-h forecasts. The data-driven approach is considered to be an excellent complement for the knowledge-driven model.

2 State of the Art on Air Quality Prediction

Air quality prediction methods can be split into two main categories: classical deterministic models and data-driven models [3]. The traditional dispersion models consist of heavy domain knowledge of air quality behavior with expertise from multiple areas among other on chemical, emission, and climatological processes. These factors help to create complex numerical models to predict the future. However, these dispersion models are computationally heavy and expensive in maintenance. The second category refers to data-driven models, where various machine learning methods have been applied to predict air pollution.

2.1 Influential Variables

Due to the complexity of air quality behavior, it is crucial to include multiple influential variables. In recent studies, several pollutants and meteorological variables have been included. The different pollutants are often PM_x , NO_x , SO_x , CO_x , Ozone, and VOC. Meteorological variables are those which describe the weather and the atmospheric composition. The most common meteorological variables are temperature, pressure, humidity, and surface wind with speed and direction. The meteorological variables vary from location to location and affect air pollution differently. Various air pollutants and meteorological variables have been extensively studied in the literature [4, 6–8, 12, 13, 15]. Other variables such as traffic [5] and weather forecast [4, 6, 11, 16] have been investigated to find their relations with air quality changes.

2.2 Air Quality Prediction Methods

Multiple research studies apply variations of Recurrent Neural Networks (RNN) to capture temporal dependencies. [6] includes an LSTM model to learn short-term and long-term temporal dependencies by using the weather forecast. [9] adopts an LSTM solution on IoT sensor data to perform short-term prediction. [12] provides a performance overview of different RNN cells and concludes that GRU cell has a slightly higher accuracy of learning PM_{10} concentration. [13] consists of an LSTM model that considers spatio-temporal relations for predicting air quality concentrations. From their results comparing an extended LSTM to SVR, deep learning-based models exhibit better prediction.

Multiple specialized multilayer perceptron (MLP) networks [11] were implemented for each weather class, determined by clustering. They further learn the relation between a high concentration of air pollutants and different weather classes to improve the classification of sudden spikes. In [17], they show how a deep learning regression model can learn patterns of pollutants and weather data collected from 449 sensors all around Aarhus city in Denmark. Their DNN model can outperform SVM in predictions of the next hour.

In [10], they use fuzzy inference of the results from an ensemble of random forest (RF) and feed-forward neural network. They combine the power of a non-linear relationship in a neural network and the averaging strategies of an

ensemble approach to generalize the results. [14] predicts daily NO₂ exposure and compares an RF model with an LR model at a national scale. [15] also applied an RF model to predict PM_{2.5} with features including other pollutants and meteorological variables. Their RF model shows better performance than their implementation of a generalized additive model.

2.3 Norwegian Air Quality Service

A new nationwide air quality information service was launched on December 18th, 2018 in Norway, by the Norwegian Environment Agency, which will be referenced in this paper as MET [18]. Their urban EMEP (uEMEP) model is a downscaling model of EMEP, a knowledge-driven model that calculates the transboundary transport of air pollutants [19]. uEMEP initiates with low spatial data (10 km-2.5 km resolution) from the EMEP model, which is scaled down to an approximately 50 m grid resolution based on proxy data from each grid. The proxy data consist of meteorological forecasts, historical emissions and traffic volume, and geographic variables. Each grid calculates its local contribution of emissions and with a Gaussian model to find non-local concentrations.

Notable strengths of uEMEP are its consideration of all primary sources of air quality pollution with a direct connection to weather forecasts and geographical terrain. Although adding weather forecasts into the model is a strength, it can also be a weakness if the forecasts deviate from the real values and thereby induce warnings of too high or low air pollution levels.

3 Complete Solution

This section introduces the complete solution we developed for air quality monitoring and forecasting using NB-IoT and machine learning. The overall architecture is presented in Fig. 1.

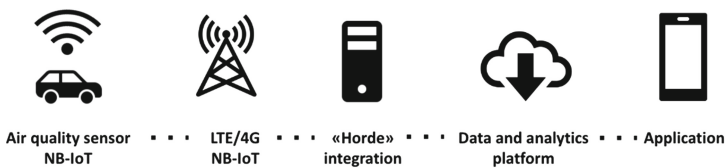


Fig. 1. IoT pipeline.

3.1 Air Quality NB-IoT Sensor Device

Recent progress in sensory and communication technologies has made the development of portable air-quality micro-sensing devices feasible. For our project,

a device consisting of a board with sensors and a communication modem was compiled¹. The initial price of this device was around EUR 100, excluding costs tied to writing software for the integration of parts. The sensors report levels of particle dust (PM_{2.5} and PM₁₀), temperature, humidity, CO₂ equivalents, and VOC equivalents. The communication modem includes a GPS module, and it supports both LTE-M and NB-IoT connectivity². So far, only off-the-shelf low-cost micro-sensors have been used in these designs. In the first version, a particle sensor made by Honeywell was used³.

The quality of data from micro-sensors has been questioned, and there is a need for assessment of the sensors' performance in varied applications and environments. This need has been addressed by [20]. An initial test of the data quality from the device compiled in this project (compared to an industrial sensor of particle dust in the same location) was made, indicating that the measurements were influenced by variations in temperature, humidity, and pollutant levels. A thorough and systematic testing of the differences in performance between our micro-sensor devices and the standardized industrial equipment over time remains to be done. Based on the initial test, we decided to use data only from the standardized industrial equipment for training the machine learning models (see Chap. 4). The plan for this project includes systematic testing of micro-sensor devices with more expensive components and various designs, e.g., positioning and structure of the inlet which tends to have a big impact on the measurement quality of PM devices. Depending on the results of these tests a number of micro-sensor devices will be deployed throughout the city and on vehicles. The initial plan is in the range of 25–50 devices.

3.2 IoT Data Platform and Analytics

The Lambda Architecture [21] is used in the design of the air quality data platform. The architecture is a data-processing architecture designed to handle massive quantities of data by taking advantage of both batch-processing and stream-processing methods. It attempts to balance latency, throughput, and fault-tolerance by using batch processing to provide comprehensive and accurate views of batch data, while simultaneously using real-time stream processing to provide views of online data. The IoT data platform adopts the Lambda Architecture in its design to build a robust system that is fault-tolerant and able to serve a wide range of workloads and low-latency reads and updates.

¹ <https://blog.exploratory.engineering/post/where-the-air-is-crisp/>.

² The modem is based on Nordic Semiconductor technology, see <https://www.nordicsemi.com/Products/Low-power-cellular-IoT/nRF9160> and had an indicative price of USD 45, see <https://shop.exploratory.engineering/>.

³ Honeywell HPM Series, Particle Sensor, 32322550, Issue E, using a light scattering method. The price of this sensor measuring both PM_{2.5} and PM₁₀ was less than 40 USD as of November 2018.



Fig. 2. Air quality micro-sensor devices mounted to buses and stationary locations

In this project, we use the Horde IoT data platform [1] (beta version) from Telenor Exploratory Engineering team to support running NB-IoT devices experiments quickly and efficiently. Horde handles data encryption and provides back-end services to manage IoT devices, inspect payloads, share devices between team members, and send data from the devices. Through Horde, users are able to get data online quickly either as a WebSocket (for a simple single-page web apps) or WebHooks (for a quick demo service), and into IFTTT (known as IF This Then That [23], a free web-based service to create chains of simple conditional statements, for quick prototypes and hackathons) or MQTT (Message Queuing Telemetry Transport [22], for flexible and reliable delivery) (Fig. 2).

3.3 Mobile Application for Air Quality Monitoring and Forecasting

The mobile application we developed, named Lufta, is now available in Google Play for testing purpose [2]. The app analyses the aggregated data provided by both sensors in stationary locations as well as sensors mounted to moving vehicles. That gives users the possibility to have a better and real-time monitoring of air quality in the areas where they live. In the app, users can also see recommendations or set up to get notifications about the level of air quality, whether it is good, moderate, or bad. We plan to provide the forecast information to users in the app as well through APIs to third parties.

4 Experiment for Exploring Air Quality Prediction

This section describes the experiments performed in the project. First, in Subsect. 4.1, we introduce the experimental setting, including the description of datasets, extracted features, machine learning methods used, and the evaluation metrics. Next, the experiment results are presented in Subsect. 4.2.

4.1 Experimental Settings

Dataset. Because of the unstable data provided by our micro-sensors during the testing phase, we decided to do our experiments with the data from expensive sensors given by the Norwegian Institute for Air Research through open APIs [24]. Four air quality stations are used in this research. Air quality in Trondheim has improved due to initiatives taken by the municipality, such as road cleaning and dust suppression. Thus the data analysis and machine learning models utilized data from January 2014 to April 2019 to avoid learning on too old data with unrelated distributions.

The weather dataset consists mainly of the hourly data recorded at a station at Voll in Trondheim, which includes features like temperature, precipitation, humidity, pressure, wind speed, and wind direction. The traffic data consists of traffic information on the road network in Trondheim. The traffic information was included by taking the mean of the closest three traffic stations for each of the locations predicted. The recorded variables are hourly vehicle count in both driving directions. This paper uses the sum of the numbers of passing vehicles in both driving directions and assumes that this sum of recordings is sufficient for analyzing the relationship between traffic and air pollutants (Fig. 3).

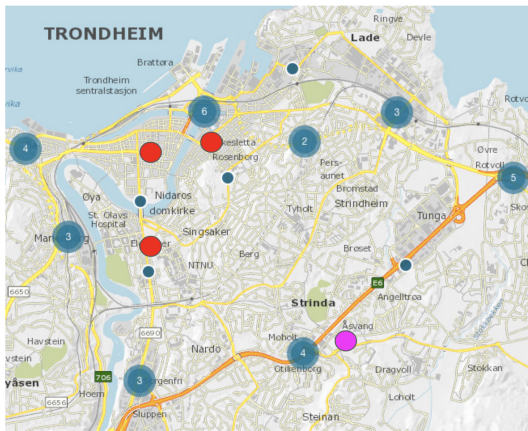


Fig. 3. Map of the data stations in Trondheim where red marks show air quality stations, pink is a weather station, and blue (small and large) is traffic stations. The numbers within the circles are indicators of the total number of stations in that area. (Color figure online)

Extracted Features. The extracted features are divided into different categories. See Fig. 4 for an overview of all with their shorthand ID, type, feature, and a short description. In this paper, we deal with three kinds of features that measure some qualities in nature (meteorological, air quality, and traffic). We also identify three types of features (temporal, statistical, and spatial). In total, we have used 655 high-level extracted features.

ID	Type	Feature	Description
M	Meteorological	humidity	Hourly average relative humidity (%)
		pressure	Hourly average surface pressure (Pa)
		temperature	Hourly average air temperature ($^{\circ}C$)
		wind direction	Hourly average wind direction (degrees)
		wind speed	Hourly average wind speed (m/s)
O	Air Quality	precipitation	Hourly measured sum of precipitation (mm)
		PM2.5	Hourly measured particular dust below $2.5 \mu g/m$
		PM10	Hourly measured particular dust above $2.5 \mu g/m$
		NO2	Hourly measured ($\mu g/m$)
		NO	Hourly measured ($\mu g/m$)
V	Traffic	NOx	Hourly measured ($\mu g/m$)
		Traffic volume	Hourly traffic count
T	Temporal	hour	Hour of the timestamp (0-23)
		month	Month of the timestamp (1-12)
		day of week	Day of week of the timestamp (0-6)
		day of month	Day of month of the timestamp (0-30)
		holiday	Is the day of the timestamp a Norwegian holiday
		season	The timestamp season (1-4)
		N lagged of X	X parameter during the past N hours
C	Statistical	Moving Average	Moving average, $n = [3, 6, 12]$
		Difference	Difference between previous values, $n = [3, 6, 12]$
		Deviation	Deviation from previous values, $n = [3, 12, 24]$
		Minimum	Minimum of previous values, $n = [24, 48]$
		Maximum	Maximum of previous values, $n = [24, 48]$
S	Spatial	PM2.5	Mean value of neighbouring stations
		PM10	Mean value of neighbouring stations
		NO2	Mean value of neighbouring stations
		NO	Mean value of neighbouring stations
		NOx	Mean value of neighbouring stations

Fig. 4. The final set of extracted feature

The temporal features are mainly generated by the use of the timestamp of the measurement. The timestamp includes information of the hour of the day, the day of the week, the day of the month, the month, and the season of the year. The Norwegian holiday calendar is matched against the date to see if it is a day off. The last temporal feature is created out of the historical values of the parameter. The spatial features contain properties from neighboring stations. These are calculated from the mean of the nearby stations. These features are included to help the models capture spatial relations of the air quality.

The statistical features are produced by applying a set of mathematical functions to the time series to derive unique properties, such as lagged value difference, moving average, moving standard deviation, moving minimum, and moving maximum, as shown in Fig. 4. The goal of statistical features is to add a more general and broader temporal dependency by including historical values. The statistical functions will consist of a smarter relation of the past, that the models will easier learn. The statistical features will provide reliable and more straightforward ties between the past and the forecasts. Statistical feature engineering can help smooth the raw values of the time series to decipher the complexity. The functions minimum, maximum, and moving average are used to capture trends in the series. The difference and deviation can help to detect sudden changes by learning what happened just before the change.

Machine Learning Models. It was decided to implement seven forecasting techniques, each with its unique trait, and identified as potentially advantageous approaches for air quality prediction. Autoregressive Integrated Moving Average

(ARIMA) and Ridge Regression (Ridge) had been applied to time series problems with reliable results in the past. Multilayer Perceptron (MLP) and Random Forest (RF) had been used in the recent literature with reliable results within air quality prediction problems. A version of Recurrent Neural Network with Gated Recurrent Unit cells (GRU) was included due to its predicting powers of time series problems. Finally, because of the ability of gradient boosting to minimize error in complex problem domains, and because it is less used in the literature, two unique variations of gradient boosting were implemented, Gradient Boosting Decision Tree (GBM) and Dropouts meet Multiple Additive Regression Trees (DART).

Gradient Boosting implemented with Microsoft’s version of LightGBM [25]. It is an optimized version of gradient boosting and is faster with the same accuracy than its competitors XGBoost and Scikit Learns version. In this paper, we use the implementation of the traditional GBM, and DART. PyTorch is used to implement the Recurrent Neural Network (RNN) model. The implementation can utilize either GRU or LSTM cells. Several model hyperparameters were optimized using randomized search; the RNN cell (LSTM or GRU), number of layers, number of RNN cells, learning rate, sequence length, dropout rate, and batch size.

Evaluation Metrics. In the literature of air quality, there is no single superior evaluation method. For our experiments, we have used the set of multiple performance metrics, including Mean Absolute Error (MAE), Relative Absolute Error (RAE), and Root Mean Squared Error (RMSE). However, because of the limitation on the number of pages in this paper, we chose to present the RAE results only. Other evaluation metrics provided consistent results.

In addition to regular air quality patterns, it frequently occurs sudden changes in the pollution concentration, which are essential to predict for real-time monitoring as they can have more impact on the daily life of most people. While the evaluation metrics defined in the previous section cover the total error and how good the model fit the actual values, they are not suitable metrics for anomaly prediction. Instead, we used the F1-Score as the evaluation metric for anomaly prediction. The predicted anomalies are matched against the real observed time series and are counted as hits if they were in the span of 1 h into the past and 1 h into the future. The smoothing and interval calculation will then account for a range of 4 h that needs to overlap. The interval of 4 h is fine since a typical sudden change lasts for about 4–6 h, and there are few partial overlaps of lengthy anomalies in the time series dataset. This straightforward approach for anomaly prediction ignores the residuals of the predicted spikes, but it related well to classifying the specific warning levels. These warning levels (good, OK, or bad) are a simple indicator for the city populations to grasp the air quality at their location.

4.2 Experimental Results

The models are trained on data from 1. January 2014, to 30. November 2018, and tested on data until 30. April 2019. The results are split up into two evaluations where the first concerned with the model's regression error for general air quality pattern, while the second for its classification accuracy toward anomaly prediction of sudden changes and spikes. All results are shown in Fig. 5. As we can see, the DART model performed the best in terms of predicting the ordinary situations of air quality whereas the GRU model provided the best results in predicting spikes in air quality pollution.

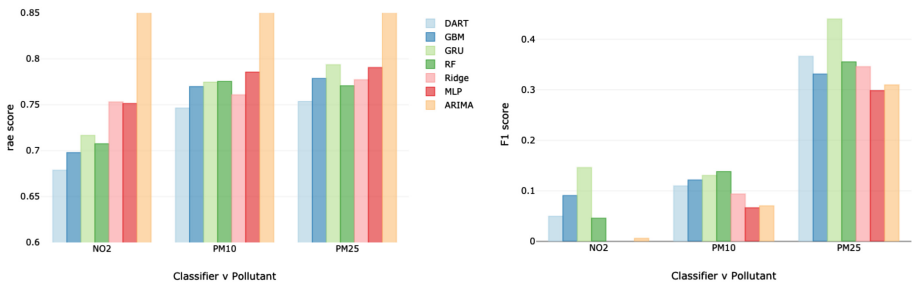


Fig. 5. Models performance RAE (left chart) and anomaly prediction F1-score (right chart) with different pollutants.

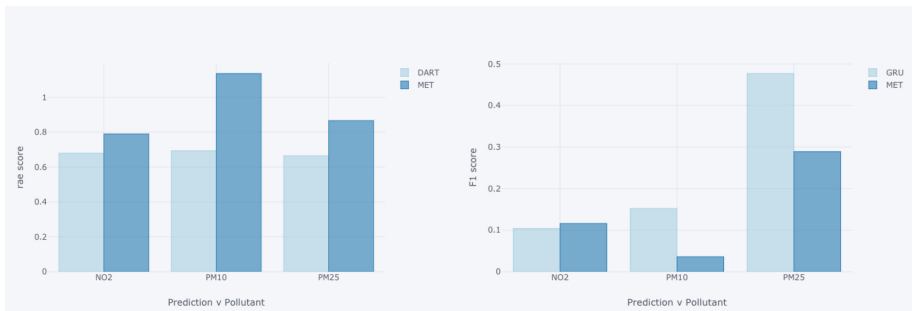


Fig. 6. Results of regression error RAE (left chart) and anomaly prediction F1-score (right chart) grouped by pollutant type.

We compared the results of machine learning predictions with the Norwegian national air quality service, a knowledge-driven model described in Sect. 2. The evaluation of the results is presented in two parts in Fig. 6: The first includes the RAE regression error of 24-h predictions. The second evaluation shows the results of the accuracy of classifying anomalies found. Our tested machine learning models DART and GRU outperformed the MET expert-based forecast model in both ordinary situations and in case of sudden changes in the air quality.

5 Conclusion

We have developed a complete solution for air quality monitoring and forecasting, which contains the holistic IoT pipeline with our own developed micro-sensors, IoT data platform, data analytics, machine learning for prediction, and mobile app for visualization. The goal of this study is to evaluate the performance of machine learning methods for air quality prediction in Trondheim. We started with an analysis of datasets of Trondheim, including air pollutants, historical weather observations, traffic volume count, and wood burners. Further, we created more features with statistical feature engineering and tested multiple state-of-the-art machine learning techniques. Seven machine learning models were implemented, optimized, trained, and tested to determine the strengths and weaknesses of air quality prediction. The results showed that DART has the best performance of predicting the overall air quality for all the pollutants studied (PM_{2.5}, PM₁₀, NO₂). We found that GRU can classify sudden changes better than the other methods. Lastly, the machine learning results were compared with the national air quality service, a knowledge-driven model, to evaluate real-world practice. The predictions of general pattern and anomalies of this study are shown to be superior for 24-h, and more comparable for the 48-h forecast.

Acknowledgements. This work is part of the AI4EU project (<https://www.ai4eu.eu/>) which has received funding from the European Union's Horizon 2020 research and innovation program, under the Grant Agreement No 825619.

References

1. Horde IoT Backend Service. <https://blog.exploratory.engineering/post/horde-iot-backend/>. Accessed 13 Aug 2019
2. Lufta app. <https://play.google.com/store/apps/details?id=com.luftaapp.lufta>. Accessed 13 Aug 2019
3. Zhang, Y., Bocquet, M., Mallet, V., Seigneur, C., Baklanov, A.: Real-time air quality forecasting, part I: history, techniques, and current status. *Atmos. Environ.* **60**, 632–655 (2012a)
4. Yi, X., et al.: Deep distributed fusion network for air quality prediction. In: Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining (KDD 2018), pp. 965–973. ACM, New York. <https://doi.org/10.1145/3219819.3219822>
5. Chen, L., et al.: Spatially fine-grained urban air quality estimation using ensemble semi-supervised learning and pruning. In: International Joint Conference on Pervasive and Ubiquitous Computing, pp. 1076–1087. ACM (2016)
6. Wang, J., Song, G.: A deep spatial-temporal ensemble model for air quality prediction. *Neuro Comput.* **314**, 198–206 (2018)
7. Qi, Z., et al.: Deep air learning: interpolation, prediction, and feature analysis of fine-grained air quality. *IEEE Trans. Knowl. Data Eng.* **30**(12), 2285–2297 (2018)
8. Lin, Y., et al.: Exploiting spatiotemporal patterns for accurate air quality forecasting using deep learning. In: Proceedings of the 26th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems, pp. 359–368 (2018)

9. K ok,  . I., et al.: A deep learning model for air quality prediction in smart cities. In: IEEE International Conference on Big Data, pp. 1983–1990 (2017)
10. Bougoudis, I., et al.: HISYCOL a hybrid computational intelligence system for combined machine learning: the case of air pollution modeling in athens. *Neural Comput. Appl.* **27**(5), 1191–1206 (2016)
11. Tamas, W., et al.: Hybridization of air quality forecasting models using machine learning and clustering: an original approach to detect pollutant peaks. *Aerosol Air Qual. Res.* **16**, 405–416 (2016)
12. Athira, V., et al.: DeepairNet: applying recurrent networks for air quality prediction. *Procedia Comput. Sci.* **132**, 1394–1403 (2018)
13. Li, X., et al.: Long short-term memory neural network for air pollutant concentration predictions: method development and evaluation. *Environ. Pollut.* **231**, 997–1004 (2017)
14. Zhan, Y., et al.: Satellite-based estimates of daily NO₂ exposure in China using hybrid random forest and spatiotemporal Kriging model. *Environ. Sci. Technol.* **52**(7), 4180–4189 (2018)
15. Chen, G., et al.: A machine learning method to estimate PM_{2.5} concentrations across China with remote sensing, meteorological and land use information. *Sci. Total Environ.* **636**, 52–60 (2018)
16. Zheng, Y., et al.: Forecasting fine-grained air quality based on big data. In: Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 2267–2276. ACM (2015)
17. Ghoneim, O.A., Manjunatha, B., et al.: Forecasting of ozone concentration in smart city using deep learning. In: 2017 International Conference on Advances in Computing, Communications and Informatics (ICACCI). IEEE, pp. 1320–1326 (2017)
18. Denby, B.R., et al.: The norwegian air quality service: model forecasting (2018). <https://www.met.no/>
19. T rseth, K., et al.: Introduction to the European monitoring and evaluation programme (EMEP) and observed atmospheric composition change during 1972–2009. *Atmos. Chem. Phys.* **12**(12), 5447–5481 (2012)
20. Fishbain, B., et al.: An evaluation tool kit of air quality micro-sensing units. *Sci. Total Environ.* **575**(1), 639–648 (2017). <https://doi.org/10.1016/j.scitotenv.2016.09.061>
21. Kiran, M., Murphy, P., Monga, I., Dugan, J., Baveja, S.S.: Lambda architecture for cost-effective batch and speed big data processing. In: 2015 IEEE International Conference on Big Data (Big Data), Santa Clara, CA, pp. 2785–2792 (2015)
22. Singh, M., Rajan, M.A., Shivraj, V.L., Balamuralidhar, P.: Secure MQTT for Internet of Things (IoT). In: 2015 Fifth International Conference on Communication Systems and Network Technologies, Gwalior, pp. 746–751 (2015)
23. Kauling, D., Mahmoud, Q.H.: Sensorian Hub: an IFTTT-based platform for collecting and processing sensor data. In: 2017 14th IEEE Annual Consumer Communications & Networking Conference (CCNC), Las Vegas, NV 2017, pp. 504–509 (2017)
24. Norwegian Institute for Air Research. <https://www.nilu.no/en/>
25. Ke, G., et al.: LightGBM: a highly efficient gradient boosting decision tree. In: Advances in Neural Information Processing Systems, vol. 30, pp. 3146–3154. Curran Associates Inc. (2017)