

RESEARCH ARTICLE

Comparative Analysis of Deep Learning and Statistical Models for Air Pollutants Prediction in Urban Areas

FAREENA NAZ¹, CONOR MCCANN¹, MUHAMMAD FAHIM¹, TUAN-VU CAO², RUTH HUNTER³, NGUYEN TRUNG VIET⁴, LONG D. NGUYEN⁵, (Member, IEEE), AND TRUNG Q. DUONG¹, (Fellow, IEEE)

¹School of Electronics, Electrical Engineering and Computer Science, Queen's University Belfast, BT7 1NN Belfast, U.K.

²Norwegian Institute for Air Research, 2007 Oslo, Norway

³Centre for Public Health, School of Medicine, Dentistry and Biomedical Sciences, Queen's University Belfast, BT7 1NN Belfast, U.K.

⁴Thuyloi University, Hanoi 116705, Vietnam

⁵Duy Tan University, Da Nang 50000, Vietnam

Corresponding author: Trung Q. Duong (trung.q.duong@qub.ac.uk)

The work of Muhammad Fahim, Tuan-Vu Cao, and Trung Q. Duong was supported in part by UKRI and European Commission under MISO Project, "Autonomous Multi-Format In-Situ Observation Platform for Atmospheric Carbon Dioxide and Methane Monitoring in Permafrost and Wetlands." The work of Ruth Hunter and Trung Q. Duong was supported by the SPACE (Supportive Environments for Physical and Social Activity for Cognitive Health) Project (<https://www.qub.ac.uk/sites/space/>) funded by the ESRC Healthy Ageing Challenge under Grant ES/V016075/1.

ABSTRACT Rapid growth in urbanization and industrialization leads to an increase in air pollution and poor air quality. Because of its adverse effects on the natural environment and human health, it's been declared a "silent public health emergency". To deal with this global challenge, accurate prediction of air pollution is important for stakeholders to take required actions. In recent years, deep learning-based forecasting models show promise for more effective and efficient forecasting of air quality than other approaches. In this study, we made a comparative analysis of various deep learning-based single-step forecasting models such as long short term memory (LSTM), gated recurrent unit (GRU), and a statistical model to predict five air pollutants namely Nitrogen Dioxide (NO₂), Ozone (O₃), Sulphur Dioxide (SO₂), and Particulate Matter (PM_{2.5}, and PM₁₀). For empirical evaluation, we used a publicly available dataset collected in Northern Ireland, using an air quality monitoring station situated in Belfast city centre. It measures the concentration of air pollutants. The performance of forecasting models is evaluated based on three performance metrics: (a) root mean square error (RMSE), (b) mean absolute error (MAE) and (c) R-squared (R²). The result shows that deep learning models consistently achieved the least RMSE compared to the statistical models with a value of 0.59. In addition, the deep learning model is also found to have the highest R² score of 0.856.

INDEX TERMS Air quality, machine learning, deep learning, predictive models, statistical methods.

I. INTRODUCTION

Over the past few years, air pollution has become a major global challenge. Air pollution has a direct impact not only on the environment but also on human health and well-being. It has been observed that air pollution leads to increase mortality and morbidity such as respiratory diseases, impaired cognitive function, cardiovascular diseases, and cancer [1], [2]. Each year, over 3 million deaths are recorded due to air

pollution especially in low and middle income countries [3]. In addition, the United Nations (UN) has defined sustainable development goals (SDG) such as 3, 7 and 11 where targets are set for 2030 to reduce deaths, illness and the adverse environmental effect in cities by improving air quality and other factors [4]. Similarly, in the United Kingdom (UK), the government has set a target to reduce 35% of air pollution by 2040 [5].

There are multiple factors involved in deteriorating air quality such as manufacturing, industrial emissions, transportation (in land, air and sea) emissions, dust, and coal

The associate editor coordinating the review of this manuscript and approving it for publication was Kathiravan Srinivasan¹.

consumption [6]. Air pollution is the introduction of harmful materials and gases into the environment which is of great concern to humans and other living organisms. These harmful materials (solids, liquids or gases) are called pollutants. When these pollutants like PM_{2.5} (Particulate Matter) are produced in higher concentrations than usual, it reduces the quality of our environment and causes serious harmful health effects [7], [8], [9].

Education and raising public awareness relating to this issue requires interdisciplinary approaches with professionals and other stakeholders. Local councils are playing their part and have set up many air quality monitoring stations throughout the country to monitor the concentration of air pollutants. Data collected from such monitoring stations can be used in the prediction of pollutants. Prediction of air quality is important to control air pollution and to identify areas which require solutions to overcome air pollution and related impacts. However, how to model air quality accurately is a challenge on its own and depends on available data and modelling approaches. The main contributions of this study include: (1) We propose a combination of meteorological parameters such as temperature, wind speed and wind direction with lagged air quality feature which is based on the concentration value of the previous hour for the pollutant being predicted for the next hour. We also make use of the datetime index by splitting it into hour, day and month for additional features to improve prediction accuracy and reduction of error. (2) We propose a comprehensive study to predict the five major air pollutants and a comparison of deep learning base models with statistical model. (3) We provide detail architectural and parameters information for both deep learning and statistical models, to predict each pollutant, which can be useful in various applications in the scope of smart cities or can provide benchmarking for more better and accurate models.

The remainder of the paper is organized as follows: related work is described in Section II, the considered statistical and deep learning models with architectural details are provided in Section III. Section IV describe the dataset. Model training and testing is discussed in Section V. Results and discussion are provided in Section VI and finally, we concluded the paper in Section VII.

II. RELATED WORK

In recent years, machine learning (ML), particularly deep learning (DL), models for regression problems have received a great amount of attention to predict air pollution [10], [11]. In [12] a light gradient boosting machine model is proposed to process high dimensional large-scale data, collected from 35 air quality monitoring stations in Beijing. The PM_{2.5} concentration levels are predicted for the next 24 hours. They compared model performance with Adaptive boosting (Adaboost), gradient boosting decision tree (GBDT), extreme gradient boosting (XGboost), and deep neural network (DNN). Based on symmetric mean absolute percentage error (SMAPE), mean square error (MSE) and

mean absolute error (MAE), the results show that their model outperformed. Furthermore, the integration of historical data significantly improves the performance of the model. In a similar study [13] to predict PM_{2.5}, the dataset was collected by a Taiwan air quality monitoring station from 2012 to 2017. The data was gathered from 5 major polluted cities of Taiwan and contains meteorological data as well as air pollution data. In pre-processing, methods like Fourier arrangement and spline multinomial approaches are used to impute missing values in the dataset. Although data was collected on an hourly basis, it was converted into a daily and monthly basis to make the monthly prediction. The proposed model, gradient boosting regression model, based on the coefficient of determination (R^2), RMSE, MSE and MAE are compared with models like linear regression, lasso regression, random forest regression, K-nearest neighbours regression, decision tree regression and many others. In [14], data were collected from various air quality monitoring stations including local and nearby industrial areas. The PM_{2.5} pollutant is predicted using a modified LSTM model and found to be better for prediction up to 8 hours by comparing with models such as LSTM, support vector machine-based regression (SVR) and gradient boosted tree regression (GBTR).

A multivariant lag-FLSTM (lag-LSTM-fully connected network) model is proposed with Bayesian optimisation to predict PM_{2.5} concentration [15]. The dataset is collected from four monitoring stations situated in Wayne county, United States. It contains hourly air pollutants data, meteorological and temporal features over 2 years. To impute missing values, values were replaced by values of the previous day having same time index. The proposed model was compared with other models including autoregressive integrated moving average (ARIMA), LASSO regression, ridge regression, SVR, artificial neural network (ANN), recurrent neural network (RNN), LSTM, Lag-FLSTM and evaluated using RMSE, MAE and mean absolute percentage error (MAPE). It was emphasized that combining meteorological and other air pollutant data in the training of the model can yield better prediction of PM_{2.5} concentration. In [16], authors developed forecasting models based on DL methods such as convolutional neural network (CNN), LSTM, CNN-LSTM, spatiotemporal clustering and their combination. The dataset includes hourly meteorological, air pollutant, spatial and temporal data collected from 12 monitoring stations at Beijing based on 2 years. The models performance is evaluated using RMSE and the index of agreement for prediction over 1-6 hours. The results indicated that LSTM is found to be the best model for multi-hour forecasting. However, there is a relatively small difference in the performance of the LSTM and CNN-LSTM.

In [17], a hybrid model is proposed by combining a one dimensional (1D) CNN with bi-directional LSTM for PM_{2.5} single-step and multi-step forecasting for next 48 hours. The local trends and spatial features based on hourly collected data over 4 years are extracted using 1D-CNN and proposed model is compared with SVR, variants of LSTM,

CNN and RNN to achieve lowest RMSE. An ensemble empirical mode decomposition (EEMD) method is combined with a LSTM model to improve PM2.5 forecasting [18]. The EEMD method improved the performance of forecasting by introducing white noise in the target data and decomposing the time series data into several sub-series. To predict PM2.5, data from 74 cities of China were used and predicted by combining LSTM and GRU models [19]. The proposed model produced better prediction in terms of RMSE by comparing with a LSTM model only. A bi-directional GRU model, a simplified version of LSTM without any memory gate, is used to learn long term dependencies by combining with 1D-CNN [20]. Different combinations of features were tested and found that model perform better based on historical data of pollutant and meteorological parameters like temperature, dew point, wind speed and direction. The proposed model achieved significant improvement in accuracy by comparing with traditional ML models and variants of RNN.

In [21], various ML approaches were implemented for the hourly PM2.5 concentration only and prediction is performed in Beijing using meteorological, temporal and PM2.5 concentrations data. The dataset was based on nearby stations and proposed a gradient-boosted regressor model. In a similar study using the publicly available dataset for PM2.5 prediction, a hybrid CNN-LSTM model is proposed and compared with other deep learning approaches [22]. A large-scale study is conducted by incorporating a dataset from 1,085 stations in China and PM2.5 is predicted for the next 72 hours using a state-of-the-art model architecture based on a transformer model. The performance of the proposed architecture is enhanced using a two-stage approach in which along with spatial and temporal dependencies and a stochastic approach is used to capture uncertainty of the air quality data [23]. A combination of graph neural networks (GNN) and GRU model is used to predict PM2.5 for next 72 hours by incorporating both spatial and temporal domain knowledge [24]. In the construction of graphs, the node and edges are defined by meteorological information for each city while geographical information is embedded into graph structures. Although aforementioned studies mostly investigated PM2.5, in this paper we have covered broad range of pollutants such as NO₂, O₃, SO₂, PM2.5, and PM10 and provided prediction model for each pollutant.

III. TIME SERIES FORECASTING MODELS

The details about the considered forecasting models are provided in this section as follows:

A. STATISTICAL MODEL

An autoregressive integrated moving average (ARIMA) is a classic statistical modelling approach also referred as Box-Jenkin's method for time series forecasting [25]. For non-stationary time series, generally the case for air pollution data, it requires auto regressive (AR) and moving average (MA) terms. To make data stationary, common approaches

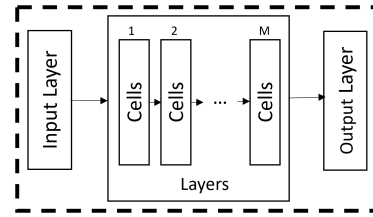


FIGURE 1. A simplified architecture of a RNN model.

such as differencing technique can be used on data $x(t)$ and order of difference (also known as lag) is defined by parameter d as given in (1). Here, B is a backward shift operator. Moreover, the model to forecast $x_d(t)$ also requires identification of parameters like p and q , which defines number of AR and MA terms, respectively, as given in (2). Here, w are weights relating AR and MA terms and $n(t)$ is Gaussian noise with zero mean. The optimum values of ARIMA model parameters such as p , d and q , can be found using autocorrelation function (ACF) and partial autocorrelation function (PACF) based on the data and its associated differences.

$$x_d(t) = x(t) - B^d x(t) \tag{1}$$

$$x_d(t) = \sum_{i=1}^p w_x^i x_d(t-i) + \sum_{j=0}^q w_n^j n(t-j) \tag{2}$$

B. DEEP LEARNING MODELS

The deep learning models contains recurrent neural network (RNN), which is based on sequence or time series data and found to provide improved performance in applications like natural language processing and speech recognition [26], [27]. The RNN model has memory units to capture and learn dependencies between input and output over short or long term. However, with the increase in the network layers and iterations, RNN tends to forget the dependencies and suffer with vanishing gradient problem. Fig 1 shows a simplified architecture of RNN model with M layers and each layer is compose of cells to process the data.

The RNN has vanishing gradients or long-term dependence issue.

1) LONG SHORT TERM MEMORY

A variant of RNN known as long short-term memory (LSTM) solves this issues [28]. A LSTM cell is shown in Fig. 2.

In Fig. 2, each cell has its state and three gates such as forget gate, an input gate and an output gate. Output of forget gate $f(t)$ decides the contribution of previous cell state $c(t-1)$ to produce current cell state $c(t)$. In addition, input gate output $i(t)$ decides amount of new information, which is required to produce current cell state. Whereas, output gate generates output $h(t)$ based on current cell state, current input (also known as new information) $x(t)$ and previous hidden state (also known as previous output or past information) of cell $h(t-1)$. These gates are represented mathematically in (3)-(8).

$$f(t) = \sigma(w_x^f x(t) + w_h^f h(t-1) + b_f) \tag{3}$$

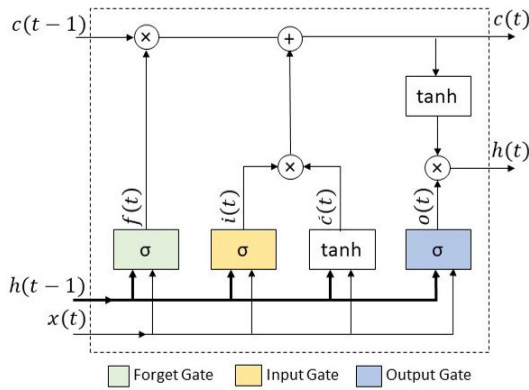


FIGURE 2. LSTM Cell.

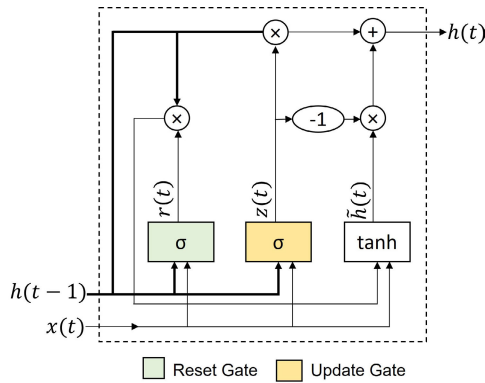


FIGURE 3. GRU Cell.

$$i(t) = \sigma(w_x^i x(t) + w_h^i h(t-1) + b_i) \quad (4)$$

$$\tilde{c}(t) = \tanh(w_x^{\tilde{c}} x(t) + w_h^{\tilde{c}} h(t-1) + b_{\tilde{c}}) \quad (5)$$

$$o(t) = \sigma(w_x^o x(t) + w_h^o h(t-1) + b_o) \quad (6)$$

$$c(t) = f(t)c(t-1) + i(t)\tilde{c}(t) \quad (7)$$

$$h(t) = o(t)\tanh(c(t)) \quad (8)$$

where σ and \tanh are the activation functions and b is the bias for respective gates in the LSTM cell. The output of each gate is generated by applying respective activation function on weighted sum, for e.g. w_x^f is the weight for input in forget gate, and bias.

2) GATED RECURRENT UNIT

Gated Recurrent Unit (GRU) is another variant of RNN and a simplified version of LSTM cell with fewer parameters to achieve faster training compare to LSTM [29]. The GRU cell is based on two gates such as a reset gate and a update gate as shown in Fig. 3.

The reset gate output $r(t)$ control the amount of previous output of cell, $h(t-1)$, information to forget while output of the update gate $z(t)$ control the amount of previous output of cell to remember. The output $h(t)$ is produce by combining gate outputs and $\tilde{h}(t)$. These gates are represented mathematically as given in (9)-(12).

$$r(t) = \sigma(w_x^r x(t) + w_h^r h(t-1) + b_r) \quad (9)$$

$$z(t) = \sigma(w_x^z x(t) + w_h^z h(t-1) + b_z) \quad (10)$$

$$\tilde{h}(t) = \tanh(w_x^{\tilde{h}} x(t) + w_h^{\tilde{h}} h(t-1) + b_{\tilde{h}}) \quad (11)$$

$$h(t) = h(t-1)z(t) + (1-z(t))\tilde{h}(t) \quad (12)$$

where σ and \tanh are the activation functions and b is the bias for respective gates in the GRU cell. The output of each gate is generated by applying the respective activation function on weighted sum, for e.g. w_x^r is the weight for input in reset gate, and bias.

IV. DATASET

The dataset used in this study is collected from air quality in Northern Ireland, which is publicly available [30]. This dataset includes hourly concentration level of the air quality parameters such as Nitrogen Dioxide (NO₂), Ozone (O₃), Sulphur Dioxide (SO₂), and Particulate Matter (PM_{2.5} and PM_{2.10}). In addition, it contains meteorological data such as temperature, wind speed and wind direction. This data is measured at the Belfast city centre using air quality monitoring station between 2015 and 2020. This dataset contains over 50,000 samples.

Statistical information of meteorological data (i.e., temperature, wind speed, wind direction) is provided in Table. 1. It includes statistical information such as the total number of samples, mean, standard deviation, minimum and maximum value of each data. Total number of samples is found to be above 50,000. Mean and standard deviation of all parameters range from 5.63 to 213.19 and 2.77 to 84.87, respectively. Minimum value of all parameters is 0 and maximum value ranges between 24 to 360. Table. 2 provides a summary of the statistical information of all air quality parameters. The NO₂ concentration data ranges from 1 to 203 with a mean and standard deviation of 26.11 and 17.87, respectively. While, lowest mean is found to be of SO₂ with standard deviation of 1.6.

V. MODEL TRAINING AND TESTING

This section provides details about the data preparation, model training, hyperparameters optimization, and testing of the single-step forecasting models.

1) DATA PRE-PROCESSING

Generally, any dataset may contain outliers, and invalid values, and data may need to be normalized as per forecasting model requirements. Outliers are extreme or odd values that are unlike other dataset values and their presence may affect the overall distribution of data. However, outliers must be removed to improve the forecasting model's performance. Likewise, the dataset may have a missing or periodic sequence of values known as invalid values that needs to be removed or replaced by some estimated values before modeling. Finally, normalization is performed on the dataset by re-scaling data to fall in the predefined range. Normalisation can help the forecasting models to perform better on the data with a smaller scale and improve the convergence speed of the models. The dataset used in this paper is pre-processed for outliers using interquartile range method (IQR) and invalid

TABLE 1. Statistics of meteorological data.

	Temperature (°C)	Wind Speed (m/s)	Wind Direction (°)
Count	52567	52567	52567
Mean	8.27	5.63	213.19
Std	4.43	2.77	84.87
Min	0	0	0
25%	4.9	3.5	156.4
50%	8	5.2	236.64
75%	11.5	7.3	273.5
Max	24	20.3	360

TABLE 2. Statistics of air pollutants in $\mu\text{g}/\text{m}^3$.

	NO ₂	O ₃	SO ₂	PM2.5	PM10
Count	52567	52608	52562	52591	52557
Mean	26.11	43.24	1.55	9	14.14
Std	17.87	20.65	1.6	7.94	10.6
Min	1	0	0	0	0
25%	13	29	1	4	8
50%	22	44	1	7	12
75%	35	58	2	11	17
Max	203	150	20	104	143

values by removing them. In pre-processing to replace the missing values, we are grouping data into month, day and hour. Then, the missing values are replaced by taking an average of the available concentration values on same month, day and hour in all years of the dataset. This approach allows for a greater spread of values for the missing data.

The meteorological data is considered an addition input features for the forecasting models. Other than the features available in the dataset, we have also created lagged feature by taking the concentration value of the previous hour of the pollutant being predicted and the datetime index is split into hour, day, and month to create additional features. Therefore, for the prediction of each pollutant, we have considered a combination of features such as previous hour concentration, meteorological (temperature, wind speed and wind direction) and temporal (day, month, hour) information for next hour prediction. The workflow for pre-processing of the dataset is shown in Fig. 4. The input features are normalised using Min-Max normalization as defined by (13), where x_{min} and x_{max} are the minimum and maximum values of data x . Fig. 5 shows a sample of NO₂ before and after pre-processing data.

2) MODEL PARAMETERS AND TUNING

The training of the ARIMA model and associated parameters are optimised and tested to forecast air pollutants. We trained the model for predicting all five air pollutants in the dataset. However, for analysis purpose, only NO₂ is explained as a use case. To find optimum values of model parameters such as p, d and q, investigation of ACF and partial autocorrelation (PACF) is required, based on the data and its associated differences. As ARIMA works on the stationary data, we observe ACF of actual data to find if the dataset is stationary or non-stationary. Fig. 6 shows that the ACF of NO₂ has all positive values and gradually dropping which indicates that

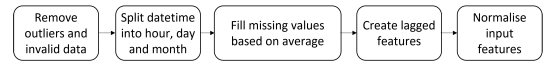


FIGURE 4. Pre-processing of dataset.

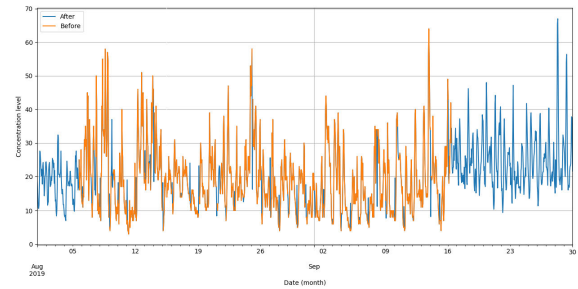


FIGURE 5. NO₂ data sample (over 2 months) representing addition of missing values.

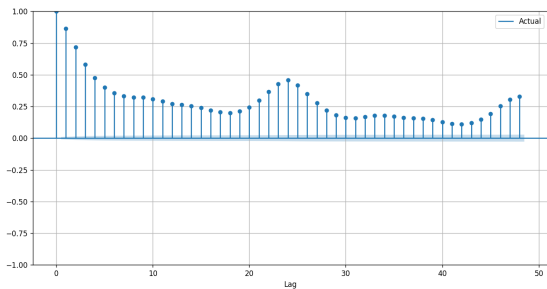


FIGURE 6. ACF of NO₂ data.

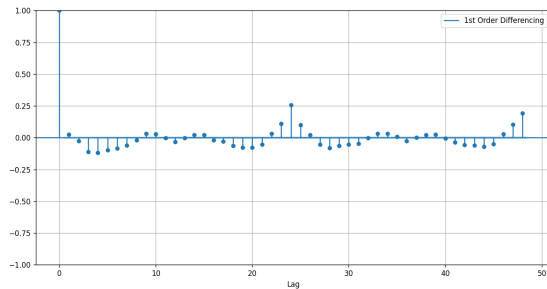


FIGURE 7. ACF of 1st order difference of NO₂ data.

the data is non-stationary. By comparing ACF of 1st order differencing (i.e., Fig.7) with the 2nd order differencing (i.e., Fig.8), we can observe that the lag at 1 of 2nd order differencing is negative, which indicates that the data may get over differenced. Whereas, this is not the case in the ACF of 1st order differencing and is sufficient to make the data stationary. Thus, we found that the value of d is 1 which defines order of differencing (d).

$$x_{norm} = \frac{x - x_{min}}{x_{max} - x_{min}} \tag{13}$$

Similarly, the order of MA terms can be found based on over differenced data by looking at ACF cuts off point. In our case, 2nd order differencing indicates over differenced due to the lag at 1 being negative and lag at 1 also shows the cut off point, which leads to q equals to 1.

The order of AR terms is found based on the first cut off point, where PACF at lag 1 is positive. Fig. 9 shows the PACF of NO₂ data. In our case as shown in Fig.10, lag at 1 of PACF of 1st order differencing is positive and shows the

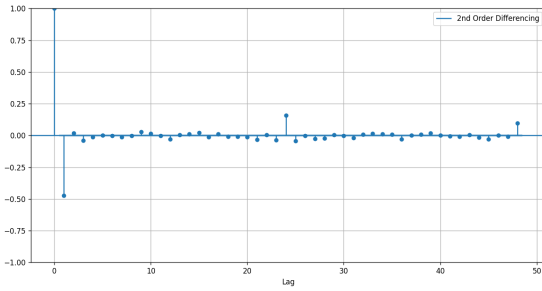


FIGURE 8. ACF of 2nd order difference of NO₂ data.

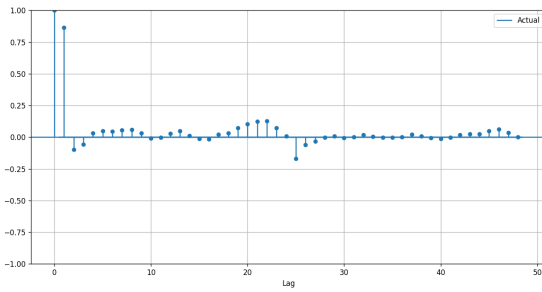


FIGURE 9. PACF of NO₂ data.

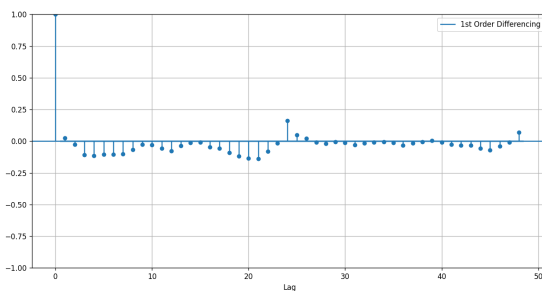


FIGURE 10. PACF of 1st order difference of NO₂ data.

cuts off point, which leads to p equals to 1. Hence, we found (1, 1, 1) as an optimum set of parameters (p, d, q) for NO₂. The detail of the optimum set of parameters for the rest of the pollutants is given Table. 4. In summary, the optimum parameters such as d and q are both found to be 1 for all the pollutants. However, only the value of p is either 0 or 1 for the respective pollutant.

The dataset is split into training, validation, and testing sets with ratio of 70%, 20% and 10%, respectively. In each split, the indices kept higher than previous set, which will avoid the shuffling (i.e., inappropriate in time-series). Fig. 11 shows the workflow of deep learning training and testing with key components. Fig. 12 shows the bounds and step size of parameters for model hyper-parameters optimisation in terms of cell size per layer, total number of layers and dropout rate for both LSTM and GRU models. In DL models, input layer pass features to model where we have considered a maximum of 5 layers and each layer can have minimum of 5 and maximum of 30 cells. In DL models, we are optimising number of layers in the range from 1 to 5 with a step size of 1. In case of number of cells, we are considering minimum of 5 to maximum of 30 cells by increasing with a step size of 5.

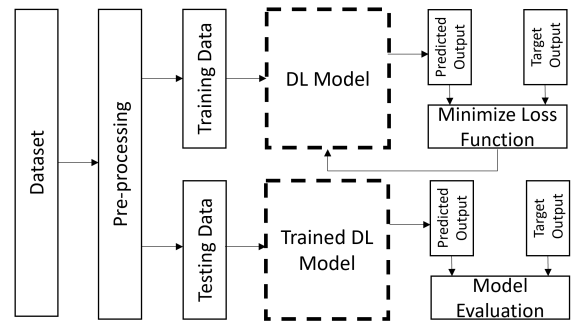


FIGURE 11. Workflow of deep learning model training and testing.

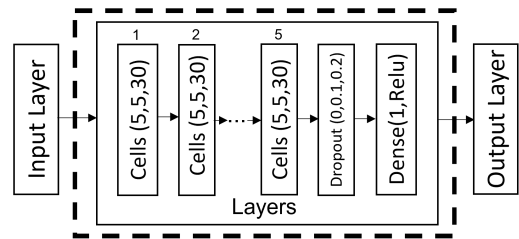


FIGURE 12. Architecture of deep learning model.

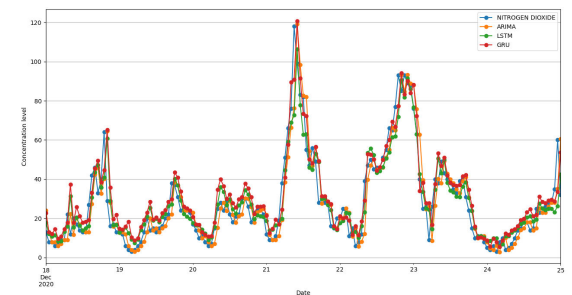


FIGURE 13. Comparison between actual and predicted data of NO₂ over a week.

TABLE 3. Summary of deep learning model parameters.

Parameters	Value
No. of layer	1, 1, 5
No. of cells in each layer	5, 5, 30
Dropout layer	0, 0.1, 0.2
Dense layer	1, Relu
Optimiser	Adam
Tuning algorithm	Hyperband

In the dropout layer, which randomly drops out number of cells to handle over fitting, Dropout rate is also considered as an optimisation parameter. At last, a fully connected dense layer with Relu activation function is used. We are using Adam optimiser during training and for the tuning of hyperparameters, the Hyperband algorithm [31] is used which tunes the number of units in a layer, the amount of layers used and the learning rate of the model. Here, the Hyperband algorithm optimise the hyperparameters by minimising the validation loss during model training to reduce the training time. The summary of the parameters with architectural details of deep learning models is given in Table. 3. After hyperparameter optimisation, the LSTM model used the least number of layers that is 2 for SO₂ and the maximum of 4 layers for NO₂. Whereas, the GRU model used the mini-

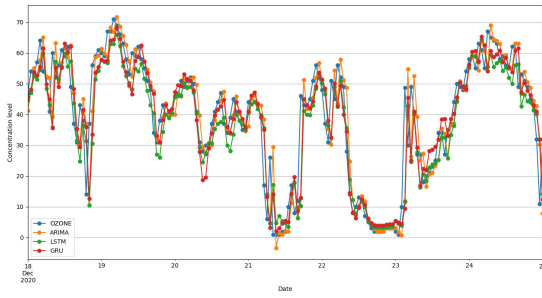


FIGURE 14. Comparison between actual and predicted data of O₃ over a week.

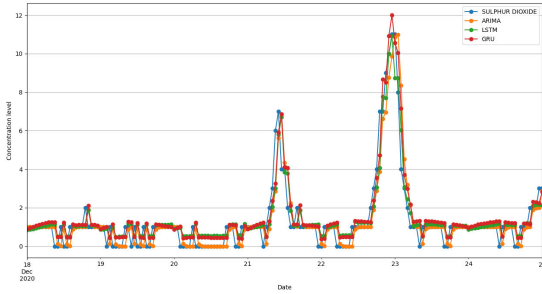


FIGURE 15. Comparison between actual and predicted data of SO₂ over a week.

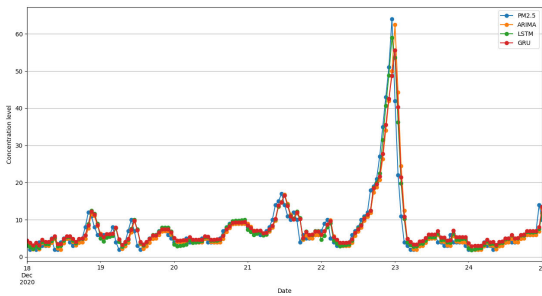


FIGURE 16. Comparison between actual and predicted data of PM_{2.5} over a week.

num number of layers that is 1 for PM_{2.5} and a maximum of 4 layers for PM₁₀. Moreover, LSTM and GRU models used an optimised number of cells in a range of [35 70] and [20 70] for considered pollutants, respectively. In addition, both DL models used optimised dropout rates with the value of 0 or 0.1 for the respective pollutant. The detail of the optimum set of hyperparameters for all the pollutants is given in Table. 4 and these parameters can be further used to evaluate the complexity and computational effort.

3) PERFORMANCE METRICS

In this work, the performance metrics used to evaluate the effectiveness of the forecasting model are RMSE, R² and MAE as defined in (14), (15) and (16) respectively.

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2} \tag{14}$$

$$R^2 = 1 - \frac{\sum_{i=1}^N (y_i - \hat{y}_i)^2}{\sum_{i=1}^N (y_i - \bar{y})^2} \tag{15}$$

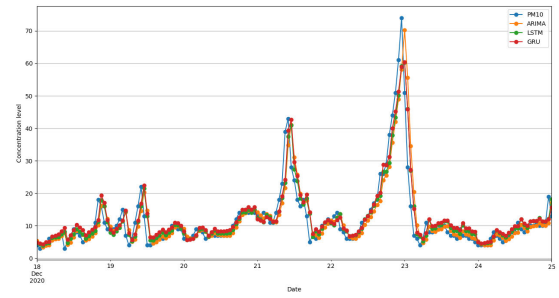


FIGURE 17. Comparison between actual and predicted data of PM₁₀ over a week.

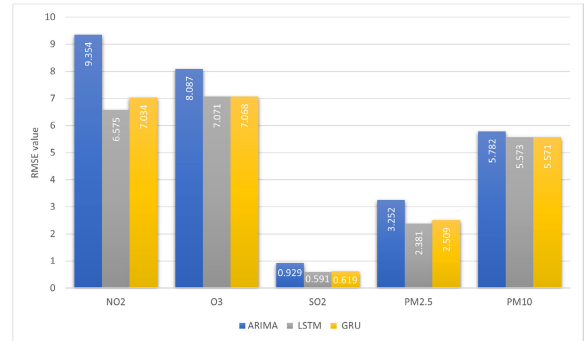


FIGURE 18. Comparison of forecasting models in terms of RMSE.

$$MAE = \frac{1}{N} \sum_{i=1}^N |y_i - \hat{y}_i| \tag{16}$$

where y_i and \hat{y}_i are target output and predicted output from forecasting model at i^{th} sample, respectively. N is total samples available to compute metrics in testing and \bar{y} is mean based on samples of target output. RMSE is a standard metric used in supervised learning applications for measuring the quality of predictions. It explains how far predicted value falls from the targeted value. The lower value of RMSE represents smaller error in the forecasting. R² also known as coefficient of determination is a statistical measure, that shows how well the regression model fits the target data. In general, the model fits the data well if R² value is closer to 1 (means the difference between the target and predicted values is small).

VI. RESULTS AND DISCUSSION

The results and related experimental discussion over the performance of all single-step forecasting models for all the pollutants is provided in this section. From the test data, we are showing comparative graphs over one week for better understanding and clarity. Fig. 13 shows one-week prediction performance of all forecasting models over test data for NO₂. The results show that for NO₂, the DL models performed better in achieving the highest R² score and least error in terms of both RMSE and MAE values than the statistical model. In both DL models, the LSTM model clearly indicates better performance than GRU. Similarly, in the case of O₃, DL models scored better than the ARIMA model on all grounds. Although, both DL models achieved similar accuracy but GRU is having a slight edge over LSTM model in attaining the least RMSE and MAE values. Fig. 14, 15, 16

TABLE 4. Optimised hyperparameters of single-step forecasting models for all the pollutants.

Models	Parameters	NO ₂	O ₃	SO ₂	PM2.5	PM10
ARIMA	(p, d, q)	(1,1,1)	(1,1,1)	(0,1,1)	(0,1,1)	(0,1,1)
LSTM	Layers	4	3	2	3	3
	Cells	25,10,15,20	20,10,10	30,5	15,15,25	25,20,15
	Dropout rate	0.0	0.0	0.1	0.1	0.0
GRU	Layers	3	3	2	1	4
	Cells	25,30,15	15,10,10	25,25	20	15,20,15,5
	Dropout rate	0.1	0.0	0.0	0.1	0.1

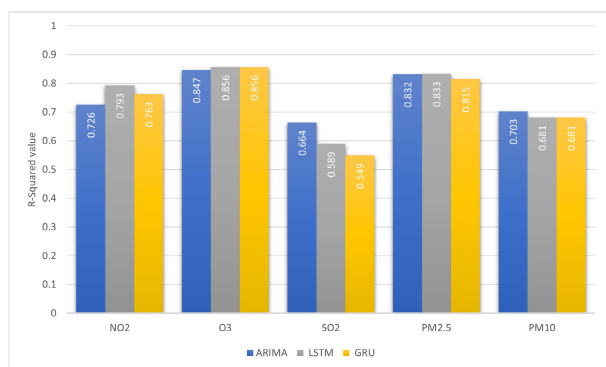


FIGURE 19. Comparison of forecasting models in terms of R².

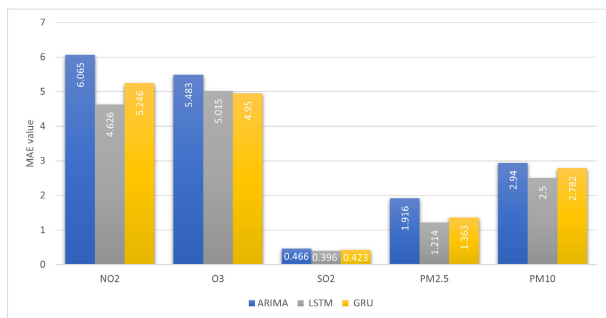


FIGURE 20. Comparison of forecasting models in terms of MAE.

and 17 shows the prediction of all forecasting models over test data for O₃, SO₂, PM_{2.5} and PM₁₀ respectively. For SO₂ and PM₁₀, ARIMA secured the highest accuracy than DL models with respect to R² but on the other hand, attained the highest error in terms of RMSE and MAE. In comparison to ARIMA model, both DL models achieved similar performance for PM₁₀ but LSTM performed better than GRU in all respects for SO₂. In case of PM_{2.5}, both LSTM and ARIMA achieved over 83% accuracy in terms of R². However, LSTM attained least error with respect to RMSE and MAE. From the above discussion, we can easily conclude that the DL models outperformed the statistical model in attaining the lowest error values in terms of both RMSE and MAE consistently for all of the air pollutants. In addition to this, DL models also achieved highest R² score for most of the air pollutants except for the PM₁₀ and SO₂, where ARIMA outperformed DL models. The performance comparison of all forecasting models in

terms of RMSE, R² and MAE is presented in Fig. 18,19 and 20 respectively.

VII. CONCLUSION

Air pollution is a global health challenge and its accurate prediction is vital to reduce health risks and environmental concerns. This work aims at single-step air pollution prediction for most of the pollutants (e.g. NO₂, O₃, SO₂, PM_{2.5}, PM₁₀) using various forecasting approaches based on DL and statistical models. The performance of the forecasting models is tested using evaluation metrics such as RMSE, MAE and R². At the broader level, among all the forecasting models and pollutants, LSTM achieved the lowest RMSE and MAE of 0.591 and 0.396 respectively, in predicting SO₂ time series data, whereas the highest RMSE and MAE is found to be 9.354 and 6.065 respectively, for NO₂ by ARIMA model. In terms of R², among all forecasting models, both DL models performed similarly in achieving the highest score of around 86% while predicting O₃. On the other end, GRU model is the one found to be having least predictive accuracy of around 55% for SO₂. Overall findings through results revealed that among all considered forecasting models, DL models outperform statistical models consistently in achieving the least error in terms of RMSE and MAE for all the pollutants and attained better predictive accuracy in terms of R² for most of the pollutants. While ARIMA model could only perform better in predicting two pollutants (i.e. SO₂ and PM₁₀) in terms of R² score only, however with the highest error value of RMSE and MAE for all of the pollutants. In future, we aim to target multi-step prediction and improve the performance of the DL models using new feature engineering approaches and relating optimisation of hyper-parameters of the models.

REFERENCES

- [1] R. D. Brook, "Cardiovascular effects of air pollution," *Clin. Sci.*, vol. 115, no. 6, pp. 175–187, Sep. 2008.
- [2] M. Stafoggia and T. Bellander, "Short-term effects of air pollutants on daily mortality in the Stockholm county—A spatiotemporal analysis," *Environ. Res.*, vol. 188, Sep. 2020, Art. no. 109854.
- [3] WHO. *Household Air Pollution and Health*. Accessed: Dec. 29, 2022. [Online]. Available: <https://www.who.int/news-room/fact-sheets/detail/household-air-pollution-and-health>
- [4] WHO. *Air Quality and Health*. Accessed: Jan. 10, 2023. [Online]. Available: <https://www.who.int/teams/environment-climate-change-and-health/air-quality-and-health/policy-progress/sustainable-development-goals-air-pollution>

- [5] B. Paul and S. Louise. (2022). *Air Quality: Policies, Proposals and Concerns—House of Commons Library*. Accessed: Jan. 10, 2023. [Online]. Available: <https://commonslibrary.parliament.uk/research-briefings/cbp-9600/>
- [6] P. J. Landrigan, “Air pollution and health,” *Lancet Public Health*, vol. 2, pp. e4–e5, Jan. 2017.
- [7] K. Abutalip, A. Al-Lahham, and A. El Saddik, “Digital twin of atmospheric environment: Sensory data fusion for high-resolution PM_{2.5} estimation and action policies recommendation,” *IEEE Access*, vol. 11, pp. 14448–14457, 2023.
- [8] J. Ma, Y. Ding, J. C. P. Cheng, F. Jiang, Y. Tan, V. J. L. Gan, and Z. Wan, “Identification of high impact factors of air quality on a national scale using big data and machine learning techniques,” *J. Cleaner Prod.*, vol. 244, Jan. 2020, Art. no. 118955.
- [9] I. Manisalidis, E. Stavropoulou, A. Stavropoulos, and E. Bezirtzoglou, “Environmental and health impacts of air pollution: A review,” *Frontiers Public Health*, vol. 8, p. 14, 2020.
- [10] S. Ameer, M. A. Shah, A. Khan, H. Song, C. Maple, S. U. Islam, and M. N. Asghar, “Comparative analysis of machine learning techniques for predicting air quality in smart cities,” *IEEE Access*, vol. 7, pp. 128325–128338, 2019.
- [11] Q. Chen, W. Wang, F. Wu, S. De, R. Wang, B. Zhang, and X. Huang, “A survey on an emerging area: Deep learning for smart city data,” *IEEE Trans. Emerg. Topics Comput. Intell.*, vol. 3, no. 5, pp. 392–410, Oct. 2019.
- [12] Y. Zhang, Y. Wang, M. Gao, Q. Ma, J. Zhao, R. Zhang, Q. Wang, and L. Huang, “A predictive data feature exploration-based air quality prediction approach,” *IEEE Access*, vol. 7, pp. 30732–30743, 2019.
- [13] Doreswamy, K. S. Harishkumar, Y. Km, and I. Gad, “Forecasting air pollution particulate matter (PM_{2.5}) using machine learning regression models,” *Proc. Comput. Sci.*, vol. 171, pp. 2057–2066, Jan. 2020.
- [14] Y.-S. Chang, H.-T. Chiao, S. Abimannan, Y.-P. Huang, Y.-T. Tsai, and K.-M. Lin, “An LSTM-based aggregated model for air pollution forecasting,” *Atmos. Pollut. Res.*, vol. 11, no. 8, pp. 1451–1463, Aug. 2020.
- [15] J. Ma, Y. Ding, J. C. P. Cheng, F. Jiang, V. J. L. Gan, and Z. Xu, “A lag-FLSTM deep learning network based on Bayesian optimization for multi-sequential-variant PM_{2.5} prediction,” *Sustain. Cities Soc.*, vol. 60, Sep. 2020, Art. no. 102237.
- [16] R. Yan, J. Liao, J. Yang, W. Sun, M. Nong, and F. Li, “Multi-hour and multi-site air quality index forecasting in Beijing using CNN, LSTM, CNN-LSTM, and spatiotemporal clustering,” *Expert Syst. Appl.*, vol. 169, May 2021, Art. no. 114513.
- [17] S. Du, T. Li, Y. Yang, and S. Hornig, “Deep air quality forecasting using hybrid deep learning framework,” *IEEE Trans. Knowl. Data Eng.*, vol. 33, no. 6, pp. 2412–2424, Jun. 2021.
- [18] N. Zaini, L. W. Ean, A. N. Ahmed, M. A. Malek, and M. F. Chow, “PM_{2.5} forecasting for an urban area based on deep learning and decomposition method,” *Sci. Rep.*, vol. 12, no. 1, p. 17565, Oct. 2022.
- [19] B. Wang, W. Kong, H. Guan, and N. N. Xiong, “Air quality forecasting based on gated recurrent long short term memory model in Internet of Things,” *IEEE Access*, vol. 7, pp. 69524–69534, 2019.
- [20] Q. Tao, F. Liu, Y. Li, and D. Sidorov, “Air pollution forecasting using a deep learning model based on 1D convnets and bidirectional GRU,” *IEEE Access*, vol. 7, pp. 76690–76698, 2019.
- [21] A. Bekkar, B. Hssina, S. Douzi, and K. Douzi, “Air quality forecasting using decision trees algorithms,” in *Proc. 2nd Int. Conf. Innov. Res. Appl. Sci., Eng. Technol. (IRASET)*, Mar. 2022, pp. 1–4.
- [22] A. Bekkar, B. Hssina, S. Douzi, and K. Douzi, “Air-pollution prediction in smart city, deep learning approach,” *J. Big Data*, vol. 8, no. 1, pp. 1–21, Dec. 2021.
- [23] Y. Liang, Y. Xia, S. Ke, Y. Wang, Q. Wen, J. Zhang, Y. Zheng, and R. Zimmermann, “Airformer: Predicting nationwide air quality in China with transformers,” Nov. 2022, *arXiv:2211.15979*.
- [24] S. Wang, Y. Li, J. Zhang, Q. Meng, L. Meng, and F. Gao, “PM_{2.5}-GNN: A domain knowledge enhanced graph neural network for PM_{2.5} forecasting,” in *Proc. 28th Int. Conf. Adv. Geographic Inf. Syst. (SIGSPATIAL)*, Nov. 2020, pp. 163–166.
- [25] G. Box, G. Jenkins, G. Reinsel, and G. Ljung, *Time Series Analysis: Forecasting and Control* (Wiley Series in Probability and Statistics). Hoboken, NJ, USA: Wiley, 2015. [Online]. Available: <https://books.google.co.uk/books?id=rNt5CgAAQBAJ>
- [26] Z. Yang, K. Wang, J. Li, Y. Huang, and Y. Zhang, “TS-RNN: Text steganalysis based on recurrent neural networks,” *IEEE Signal Process. Lett.*, vol. 26, no. 12, pp. 1743–1747, Dec. 2019.
- [27] G. Gelly and J. Gauvain, “Optimization of RNN-based speech activity detection,” *IEEE/ACM Trans. Audio, Speech, Language Process.*, vol. 26, no. 3, pp. 646–656, Mar. 2018.
- [28] S. Hochreiter and J. Schmidhuber, “Long short-term memory,” *Neural Comput.*, vol. 9, no. 8, pp. 1735–1780, Nov. 1997.
- [29] K. Cho, B. van Merriënboer, C. Gulcehre, D. Bahdanau, F. Bougares, H. Schwenk, and Y. Bengio, “Learning phrase representations using RNN encoder–decoder for statistical machine translation,” in *Proc. Conf. Empirical Methods Natural Lang. Process. (EMNLP)*. Doha, Qatar: Association for Computational Linguistics, Oct. 2014, pp. 1724–1734.
- [30] Northern Ireland Air. *Download Air Quality Data—Northern Ireland*. Accessed: Dec. 1, 2022. [Online]. Available: <https://www.airqualityni.co.uk/data>
- [31] L. Li, K. Jamieson, G. DeSalvo, A. Rostamizadeh, and A. Talwalkar, “Hyperband: A novel bandit-based approach to hyperparameter optimization,” *J. Mach. Learn. Res.*, vol. 18, no. 1, pp. 6765–6816, Jan. 2017.



FAREENA NAZ received the B.S. degree in computer science from the University of Agriculture, Pakistan, in 2008, and the M.S. degree in computer science from COMSATS University, Islamabad, Pakistan, in 2010. She is currently pursuing the Ph.D. degree with Queen's University Belfast, U.K. She is a part of EU SPACE Project. Her research interests include machine learning, data analytics, air pollution modeling, and natural language processing.



CONOR MCCANN received the B.Sc. degree in computer science from Queen's University Belfast, in 2022. He is currently a full-time software engineer. His research interest includes machine learning for air pollution modeling.



MUHAMMAD FAHIM received the B.S. degree (Hons.) from Gomal University, Pakistan, in 2007, the M.S. degree from the National University of Computer and Emerging Sciences (NUCES), Pakistan, in 2009, and the Ph.D. degree from Kyung Hee University, South Korea, in February 2014. He was a Postdoctoral Fellow with the Department of Computer Engineering, Kyung Hee University. He was an Assistant Professor with the Department of Computer and Software Engineering, Faculty of Engineering and Natural Sciences, Istanbul Sabahattin Zaim University, Istanbul, Turkey, for three years, where he also led the Machine Learning Research Laboratory. He was also an Assistant Professor with the Institute of Information Systems, Innopolis University, Innopolis, Russia, for four years. Currently, he is a Lecturer with the School of Electronics, Electrical Engineering and Computer Science, Queen's University Belfast, U.K. His research interests include machine learning models, wearable computing, digital signal processing, and behavior recognition in intelligent environments.



TUAN-VU CAO received the Ph.D. degree in electrical engineering from the University of Oslo, in 2012. He is a Senior Scientist with the Norwegian Institute for Air Research (NILU). He has a long experience in both academic and industry. From 2012 to 2018, he was a Postdoctoral Fellow with the Norwegian University of Science and Technology and a Senior Engineer with WINS Instrumentation AS and Prediktor AS, for various projects, such as monitoring nutrient food content,

wireless DST real-time instrumentation, and micropower sensor interface in nanometer CMOS technology. He joined NILU, in 2018. He is a Co-PI of EEA Grant Project–HAPADS (2020–2023), a novel air mobile monitoring system enables end-users to make information-driven decisions to mitigate air pollution exposure; and a Key Designer and a Technical Coordinator for NFR 273394–“Leopard-Wearable Particle Detector Enabling Safer Working Environments” (2018–2020). His research interests include enabling technologies and autonomous systems for environmental monitoring and management.



RUTH HUNTER is a Professor of public health and planetary health with the Centre for Public Health, Queen’s University Belfast; and the Director of the WHO collaborating center for research and training in systems thinking and complexity science for NCD prevention and control. Her work primarily involves investigating how we can improve where we live (i.e., our built and social and natural environments) for better population health and planetary health. She is particularly

interested in research at the intersection of public health and planetary health. Her work involves the application of systems-thinking and complexity science methods.



NGUYEN TRUNG VIET received the Ph.D. degree from Tohoku University, Japan, in 2007. He is currently a Professor with Thuyloi University, Vietnam, where he was promoted to the Vice President, in 2014. He has strong expertise in nearshore hydro-morphodynamics using both field measurements and numerical modeling and remote sensing video techniques. He was a Project Leader of numerous MOST Projects on Nha Trang Coast (2013–2019) and Cua Dai Beach (2015–2018) in

very close collaborations with IRD/EPOC/AFD-France, Japan, and U.K. He has published more than 100 articles in journals and international conferences. He has appointed as a Distinguished Member of IAHR-APD and the Executive Member of Asian and Pacific Coasts (APAC) Council. He has been a member of the State Council for Professorship, Vietnam, in the field of hydraulic engineering, since 2019.



LONG D. NGUYEN (Member, IEEE) was born in Dong Nai, Vietnam. He received the B.S. degree in electrical and electronics engineering and the M.S. degree in telecommunication engineering from the Ho Chi Minh City University of Technology (HCMUT), Vietnam, in 2013 and 2015, respectively, and the Ph.D. degree in electronics and electrical engineering from Queen’s University Belfast (QUB), U.K., in 2018. He was a Research Fellow with Queen’s University Belfast, for a part

of Newton Project, from 2018 to 2019. He is currently with Dong Nai University, Vietnam, as an Assistant Professor; and with Duy Tan University, as an Adjunct Assistant Professor. His research interests include convex optimization techniques for resource management in wireless communications, energy efficiency approaches, real-time embedded optimization for wireless networks, and the Internet of Things (IoT). He received the Best Paper Award at the IEEE Digital Signal Processing (DSP), in 2017; the IEEE International Conference on Recent Advances in Signal Processing, Telecommunication, and Computing (Sigtelcom), in 2018; the IEEE International Conference on Communications (ICC), in 2019; and the International Wireless Communications and Mobile Computing Conference (IWCMC), in 2019. He was also received the Exemplary Reviewer Award in IEEE COMMUNICATIONS LETTERS, in 2018. He is serving as a Reviewer for IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS, IEEE TRANSACTIONS ON COMMUNICATIONS, IEEE ACCESS, IEEE COMMUNICATIONS LETTERS, and *IET Communications*; and several international conferences.



TRUNG Q. DUONG (Fellow, IEEE) received the Ph.D. degree in telecommunications systems from the Blekinge Institute of Technology, Sweden, in 2012. In 2013, he joined Queen’s University Belfast, U.K., as an Academic Staff, where he is currently a Chair Professor of telecommunications. He is a Research Chair of the Royal Academy of Engineering. He was a Distinguished Advisory Professor with Inje University, South Korea, from 2017 to 2019. He has been an

Adjunct Professor and the Director of the Institute for AI and Big Data, Duy Tan University, Vietnam, since 2012. He was a Distinguished Professor with Thuyloi University, Vietnam, from 2023 to 2028; and a Visiting Professor (under Eminent Scholar Program) with Kyung Hee University, South Korea, from 2023 to 2024. His current research interests include quantum communications, wireless communications, signal processing, machine learning, and real-time optimization.

He is a fellow of the Asia–Pacific Artificial Intelligence Association (AAIA). He received the Best Paper Award at the IEEE VTC-Spring 2013; IEEE ICC 2014; IEEE GLOBECOM 2016, 2019, and 2022; IEEE DSP 2017; and IWCMC 2019 and 2023. He was a recipient of the prestigious Royal Academy of Engineering Research Fellowship, from 2015 to 2020, and has won the prestigious Newton Prize, in 2017. He has served as an Editor/a Guest Editor for the IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS, IEEE TRANSACTIONS ON COMMUNICATIONS, IEEE TRANSACTIONS ON VEHICULAR TECHNOLOGY, IEEE COMMUNICATIONS LETTERS, IEEE WIRELESS COMMUNICATIONS LETTERS, IEEE WIRELESS COMMUNICATIONS, *IEEE Communications Magazines*, and IEEE JOURNAL ON SELECTED AREAS IN COMMUNICATIONS. He is serving as an Executive Editor for IEEE COMMUNICATIONS LETTERS.

...