

Article

Applying Machine Learning Techniques in Air Quality Prediction—A Bucharest City Case Study

Grigore Cican ^{1,2,*} , Adrian-Nicolae Buturache ³ and Radu Mirea ² 

¹ Faculty of Aerospace Engineering, Polytechnic University of Bucharest, 1-7 Polizu Street, 1, 011061 Bucharest, Romania

² National Research and Development Institute for Gas Turbines COMOTI, 220D Iuliu Maniu, 061126 Bucharest, Romania

³ FasterEdu.com, 075100 Otopeni, Romania; office@fasteredu.com

* Correspondence: grigore.cican@upb.ro

Abstract: Air quality forecasting is very difficult to achieve in metropolitan areas due to: pollutants emission dynamics, high population density and uncertainty in defining meteorological conditions. The use of data, which contain insufficient information within the model training, and the poor selection of the model to be used limits the air quality prediction accuracy. In this study, the prediction of NO₂ concentration is made for the year 2022 using a long short-term memory network (LSTM) and a gated recurrent unit (GRU). this is an improvement in terms of performance compared to traditional methods. Data used for predictive modeling are obtained from the National Air Quality Monitoring Network. The KPIs(key performance indicator) are computed based on the testing data subset when the NO₂ predicted values are compared to the real known values. Further, two additional predictions were performed for two days outside the modeling dataset. The quality of the data is not as expected, and so, before building the models, the missing data had to be imputed. LSTM and GRU performance in predicting NO₂ levels is similar and reasonable with respect to the case study. In terms of pure generalization capabilities, both LSTM and GRU have the maximum R² value below 0.8. LSTM and GRU represent powerful architectures for time-series prediction. Both are highly configurable, so the probability of identifying the best suited solution for the studied problem is consequently high.

Keywords: machine learning; air quality; LSTM; GRU; NO₂; Bucharest



Citation: Cican, G.; Buturache, A.-N.; Mirea, R. Applying Machine Learning Techniques in Air Quality Prediction—A Bucharest City Case Study. *Sustainability* **2023**, *15*, 8445. <https://doi.org/10.3390/su15118445>

Academic Editors: Norazian Mohamed Noor, Ahmad Zia Ul-Saufie Mohamad Japeri and Mohd Remy Rozainy Mohd Arif Zainol

Received: 27 March 2023

Revised: 7 May 2023

Accepted: 10 May 2023

Published: 23 May 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Air pollution is affecting the global climate, ecosystems and human health [1,2]. It is responsible for millions of deaths all over the world [3]. Atmospheric pollution impacts human health, particularly in urban environments [4]. The concentration of pollutants corresponds to the population distribution among areas due to human activities [5,6]. One of the most important atmospheric components that has a direct relationship with pollution is nitrogen dioxide (NO₂), which is released mainly from diesel and petrol engines as reported in [7], with road transportation contributing approximately 40% of the land-based NO_x emissions in European countries. Nitrogen dioxide (NO₂) is one of the most active gaseous pollutants emitted in the industrial era and is highly correlated with human industrial activities.

Two main meteorological components, wind speed and wind direction, are directly influencing the dispersion of highly concentrated pollutants, which are emitted within the atmosphere. Thus, low and uniform wind speeds are favorable conditions for gaseous pollutant accumulation near their source [8], while high and turbulent winds are responsible for gaseous pollutant dispersion.

Nitrogen oxide (NO_x) is the generic name of a gaseous mixture normally containing various amounts of nitrogen and oxygen and represents a “family” of compounds (N₂O, NO₂, NO, N₂O₃, N₂O₂, N₂O₄, N₂O₅) [9] with NO₂ and NO as the main components. The

main NO_x quantity found in the atmosphere is produced by human activity, and their main action path is by actively participating in acid rain formation through interacting with atmospheric gasses and forming HNO₃. Moreover, these rains actively contribute to the accumulation of nitrates in the soil [10]. Another action path of the NO_x is its contribution to smog formation and in decreasing water quality.

The main components of NO_x are usually monitored separately, since both of them (NO and NO₂) cause specific issues. Thus, NO is a precursor for O₃ depletion within the upper layers of the atmosphere, and the main sources of NO emissions are jet engines. Moreover, even though NO does not directly affect the environment, it participates in the formation of nitric acid and particulate nitrate [11]. On the other hand, NO₂ acts as a “catalyst” between NO and O₃ and the main precursor for nitric acid formation. NO₂ is mainly produced by NO oxidation within the atmospheric layers. There are some sources of NO₂ formation, such as the burning of fossil fuels and biomass and diesel engines. Within urban areas, NO₂ concentrations vary between 0.1 and 0.25 ppm [12]. It is to be mentioned that NO₂ is four times more toxic than NO and children are primarily affected by it. NO₂ is very toxic for living creatures [13]. During exposure to NO₂, the lungs are drastically affected and high concentrations may be fatal. People exposed to low NO₂ concentrations but for long periods of time may suffer from respiratory issues [14]; therefore, NO₂ levels within the atmospheric environment are regulated and drastically monitored [15,16].

Tracking NO₂ emissions and predicting their concentrations represent important steps toward controlling pollution and setting rules to protect people’s health indoors, such as in factories and in outdoor environments.

Air pollution forecasting techniques include numerical models and statistical models [17]. The numerical models achieve the simulation of the transformation and diffusion of air pollutants and reflect their change law. However, they are based on a large amount of meteorological information, air pollutant discharge source data and atmospheric monitoring data. The models need to master the mechanism of pollution change, and the calculation time is long [18].

NO₂ concentration prediction is a nonlinear, multivariable problem with strong coupling between predictors, so NO₂ numerical forecasting will be an extraordinarily complex systems engineering problem. Statistical models are widely used in operational prediction, due to the advantages of their easy calculation, low data requirements and high precision. Nevertheless, most statistical models align with linear regression theory; assuming that there is a nonlinear relationship between pollutant concentration and weather conditions, linear regression is difficult to be applied to nonlinear, strongly coupled systems [19]. For air quality prediction, LSTM-based models can produce better performance than statistical models such as the ordinary least-squares regression and the Bayesian ridge regression, or machine learning models such as support vector regression, multilayer perceptron and random forest regression [20]. The same superiority of the LSTM is also with respect to ARIMA [21]. When comparing the most important recurrent neural network architectures (standard, LSTM and GRU) for air quality prediction, it is found that GRU exceeds the performance of LSTM and standard networks [22]. Moreover, a survey on machine learning algorithms used for air quality prediction showed that LSTM and multilayer perceptron are the most used models for such tasks [23].

Artificial intelligence (AI) techniques have been extensively applied in a variety of research areas [24,25].

Regarding the use of machine learning (ML) for air quality prediction, there are many studies on how related techniques are used. The studies discuss various types of pollutants such as PM₁₀, PM_{2.5}, NO, NO₂, etc. Another approach is to build hybrid models consisting of core statistics and machine learning-based models such as WANN, whereby the wavelet transform is applied prior to feeding the data into an artificial neural network [26].

In [27], a model using artificial neural networks (ANNs) was developed to forecast the pollutant concentration of PM₁₀, PM_{2.5}, NO₂, and O₃ for the current day and subsequent 4 days in a highly polluted region (32 different locations in Delhi). The model was trained

using meteorological parameters and hourly pollution concentration data for the year 2018, and then used for generating air quality forecasts in real-time. In [28], the authors developed new machine learning models, namely random forest (RF) and support vector regression (SVR), to estimate PM_{2.5} concentrations across Malaysia for the first time, covering the years 2018 and 2019.

For gaseous pollutants such as NO or NO₂, modeling should accommodate high levels of variability and nonlinearity. Therefore, the concept of artificial neural networks (ANNs) is used. The authors of [29] had developed the so-called multiplayer perception (MLP) artificial network in order to model NO and NO₂ pollution in London, and the main conclusion was that the pollutant's variations can be modeled by using the time of day and the day of the week as input variables. Moreover, the effectiveness of ANN and MLP was assessed by performing a sensitivity analysis following three predefined scenarios [30]. The main conclusion of the study was that the calculated values within all three scenarios were similar to the values measured onsite.

MLP was also used to model NO [31], NO₂ [32] and O₃ [33] within a port area (Shanghai port) and in a city area (Zagreb). The obtained results were similar with the measured concentrations of the above mentioned gaseous pollutants.

Another useful characteristic of MLP is that it can allow forecasts concerning gaseous pollution to be drafted, as shown in [34], where a three-day forecast regarding NO₂ and O₃ pollution was drafted for the city of Athens. A study that compared the performances of MLP and linear regression [35] was drafted for NO₂ and O₃. The obtained results proved to be very good in terms of predicting pollutions, using the vector regression as support, as demonstrated within [36,37]. Other studies also searched for the best model for series forecasting. They utilized various types of tools from support vector regression (SVR), time series fuzzy inference system (TSFIS) and MLP for the prediction of NO_x and O₃. Other researchers used generalized regression neural networks (GRNN), SVR, MLP and radial basis function (RBF) neural networks for predicting the NO₂ in urban areas [38–40].

Complex studies such as that of [26] used a mixture of three methods and one test. The method consists of interquartile range (IQR), isolation forest and local outlier factor (LOF), while the test is the generalized extreme standardized deviate (GESD) test. The models built within paper [26] are autoregressive integrated moving average (ARIMA), generalized regression neural networks (GRNN) and hybrid ARIMAGRNN, and their processing was possible after the removal of aberrant values. The main results of the study emphasized that the first approach produced the best performance results in terms of statistic modeling, but, nevertheless, the best model was obtained by the hybrid ARIMAGRNN.

The scope of this paper is to conceptualize and build a machine learning-based model to predict the hourly levels of NO₂ in one selected location in Bucharest where historical data are available. Among the existing machine learning techniques, artificial neural networks represent one of the best solutions for this type of predictive task. LSTM and GRU are architectures designed for time series data, having in place the right mechanisms for capturing long-term and short-term dependencies in data. To ensure that the approach and the results are meaningful in terms of performance, but also for the professionals and scientific community, the theoretic fundamentals and model evaluation are conducted in such a way that they can be replicated or compared with other similar research.

2. Methodology

Bucharest [41,42], the capital of Romania, is the largest Romanian city and is the country's main political, administrative, economic, financial, educational, scientific and cultural center. It is located in the SE part of the country, on the banks of the Dâmbovița River, less than 60 km (37.3 mi) North of the Danube River and the Bulgarian border. Bucharest has either a continental or a humid subtropical climate, with hot, humid summers and cold, snowy winters. Due to its position on the Romanian Plain, the city's winters can be windy, although some of the winds are mitigated due to urbanization. Winter temperatures often drop below 0 °C, sometimes even to −20 °C. During summer, the

average temperature is 23 °C (the average for July and August). Temperatures frequently reach 35 to 40 °C in midsummer in the city center. Although the average precipitation and humidity during the summer are low, occasional heavy storms occur. During spring and autumn, daytime temperatures vary between 17 and 22 °C, and precipitation during spring tends to be higher than in summer, with more frequent yet milder periods of rain. Bucharest has a relatively developed industrial area at its suburbs, and household heating is still dependent on large thermo-energetic plants even though a percentage of households have their own heating system. Traffic is the most important source of air quality degradation in the capital, according to the Research Report on the State of the Environment in Bucharest. More specifically, 80% of air pollution in the metropolis comes from traffic. The contribution made by road traffic is 90% of carbon monoxide emissions, 59% of nitrogen oxide emissions, 45% of volatile organic compounds and 95% of lead emissions, according to the report recently released by the Environment Platform for Bucharest. It is not surprising, given that there are approximately 1.84 million vehicles in the city, of which 1.5 million are registered in Bucharest. A total of 80% of these are cars, of which more than half are more than 12 years old. Less than a quarter of personal cars can be considered new—less than four years old—and 43% are diesel, according to data from the same study [43].

Overall, all these factors contribute to a significant increase in pollution, especially during the cold season.

2.1. Air Quality Data

There are 41 centers where the National Air Quality Monitoring Network from Romania collects data which is then transmitted and validated at the Air Quality Assessment Centre of National Agency for Environmental Protection. Specific laws regulate the gaseous pollutants' concentrations and allows the classification of agglomerations within 3 different classes (A, B or C) based on pollution measurements and assessment. The measured concentrations obtained from the measuring stations of the above mentioned network are mathematically modeled in order to assess the dispersion of the gaseous pollutants.

Law 104/2001 [35,44] sets the limits for various pollutants as follows: NO_x/NO₂, alert threshold—400 µg/m³; hourly limit for human health protection: 200 µg/m³; annual average limit for human health protection: 40 µg/m³; annual average limit for vegetation protection: 30 µg/m³.

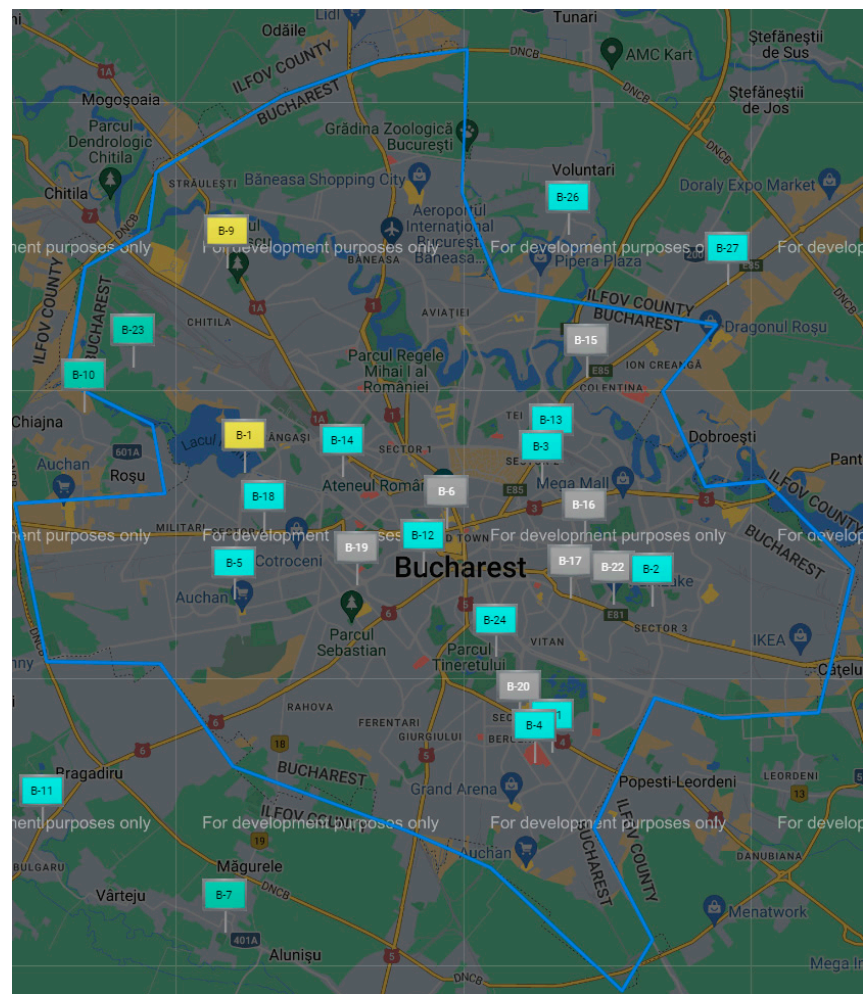
Following the worldwide trend stated in [45], the air quality of the largest Romanian cities, i.e., Bucharest, Cluj-Napoca, Timisoara, etc., has been decreasing each year [46–50].

Figure 1 shows the air quality monitoring stations around Bucharest. It is to be mentioned that only stations B3 and B6 are traffic-type stations.

The monitoring stations within Bucharest's administrative area, which measure NO_x, are the following: B-1 urban background stations /urban, B-2 industrial/urban, B-3 traffic/urban, B-4 industrial/urban, B-5 industrial/urban, B-6 traffic/urban and B-9 urban background stations/urban; only B-3 and B-6 are traffic measuring stations. Measuring station B-9 has not recorded any NO_x values within the last 5 years, so station B-9 was not taken into account. Only the stations B-1–B-6 and B-9 monitor the levels of NO_x within the Bucharest area.

Table 1 shows the average measured values of the last 5 years.

As it can be observed in Table 1, traffic measuring stations have recorded average values above the enforced limit of 40 µg/m³ except for the year 2022. This may be associated with the post-pandemic period and the enforced regulations regarding the vehicles' movements and the improvement of Bucharest's air quality. It is well known that several laws and regulations have been adopted by the municipality in order to improve air quality within the city [44,51]. Other aspects that may have influenced the low value registered for 2022 can be the higher winter temperatures, which led to a decrease in residential fossil-fuel use, an increased usage of public transport, etc.



(a)



(b)

Figure 1. Map showing the location of the sampling site (a) www.calitateair.ro, accessed on 15 March 2023; (b) <https://www.nationsonline.org/oneworld/map/romania-political-map.htm> accessed on 15 March 2023.

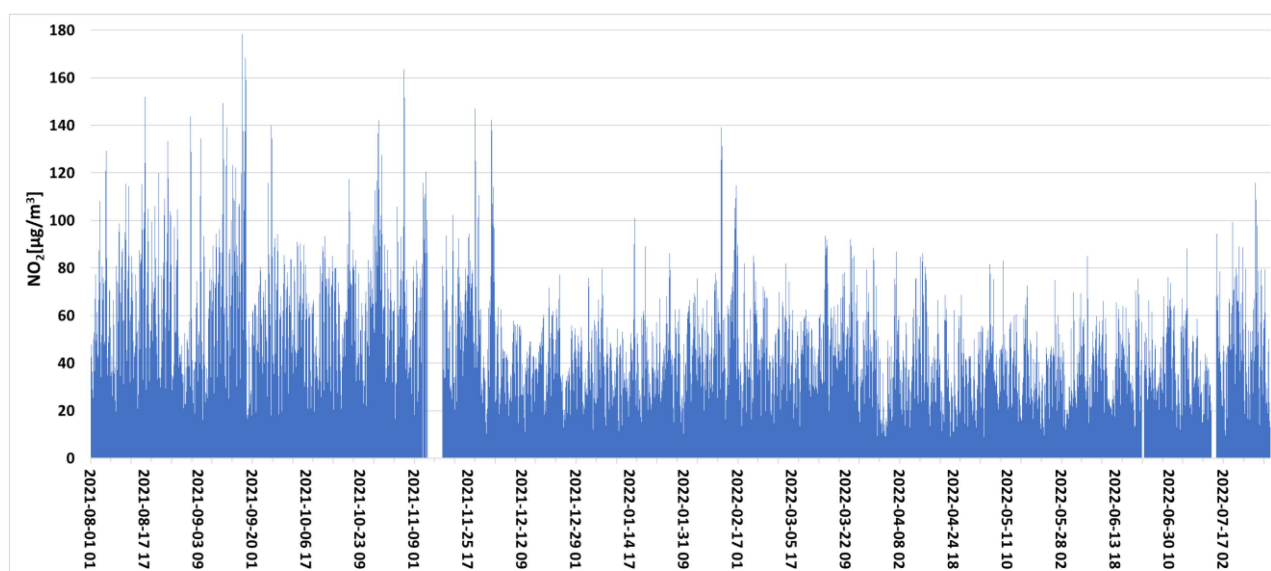
Table 1. Mean concentrations of NO₂ observed in Bucharest city in the last 5 years.

Pollutant	Year	Station/Concentration					
		B-1 µg/m ³	B-2 µg/m ³	B-3 µg/m ³	B-4 µg/m ³	B-5 µg/m ³	B-6 µg/m ³
NO ₂	2018	27.73	31.62	59.33	27.57	35.5	62.79
	2019	30.4	31.35	51.92	29.52	39.14	57.44
	2020	26.78	28.1	40.2	24.35	29.74	41.63
	2021	29.44	29.74	44.81	25.47	32.26	49.39
	2022	21.95	29.78	39.3	25.29	30.07	38.65

Note: where red = exceeds NO₂ enforced limits of 40 µg/m³ annual avg. limit for human health protection.

Since the highest values were recorded at station B-6, this station was used within this paper. The dataset used consists of 8760 records representing one year of hourly data between 1 August 2021 to 31 July 2022. The dependent variable B-6 is available online on www.calitateaer.ro, accessed on 12 September 2022.

Upon analyzing the variation of NO₂ levels at station B-6 within the entire dataset (Figure 2), it can be seen that there was a decrease in pollution beginning from December of 2021.

**Figure 2.** Hourly mean value concentrations of NO₂ during the entire period at B-6 monitoring station.

2.2. Meteorological Data

The independent variables representing hourly weather data are available via the visual crossing weather application programming interface (API) [52]. The meteorological station is the Filaret station and it is located within Bucharest city, a few hundred meters from the air-quality-monitoring station B-6.

As can be seen in Table 2, there are missing values for both dependent and some of the independent variables.

The missing data can be explained in two ways. The first one, in relation to variables such as snow depth or solar energy and based on the logic implemented in the data extract, the absence of the transform and load flow is sensible since it is not expected to snow throughout the entire year or for solar energy to be used every day for 24 h. Second, in the case of B-6, the missing data cannot be explained by any phenomenon other than miscommunication that likely led to data loss. For this second scenario, the missing data are filled using polynomial interpolation.

Table 2. Dataset statistics.

	Count	Mean	SD	Min	25%	50%	75%	Max
B-6	8218	41.5	22.0	6.8	25.2	37.2	53.1	178.3
temp	8760	12.9	9.7	−7.0	4.8	11.7	20.3	39.6
feelslike	8760	12.5	10.1	−10.0	4.1	11.7	20.3	39.3
dew	8760	5.9	7.4	−14.3	0.4	6.2	11.7	22.1
humidity	8760	67.1	22.1	12.6	49.3	67.7	86.2	100
precip	8760	0.1	0.7	0	0	0	0	30.0
precipprob	8760	1.7	12.8	0	0	0	0	100
snow	4845	0.0	0.0	0	0	0	0	0.6
snowdepth	5139	0.0	0.2	0	0	0	0	5
windgust	5113	17.8	11.1	0	10.8	14.4	22	139.9
windspeed	8760	4.8	3.1	0	3.6	3.6	7.2	26.6
winddir	8760	150.7	106.9	1	50	129	240	360
sealevelpressure	8760	1017.7	7.6	993	1013	1017	1023	1043
cloudcover	8760	51.1	40.4	0	0	50	90	100
visibility	8760	9.7	1.9	0	10	10	10	64.6
solarradiation	8734	108.3	226.7	0	0	1	38	934
solarenergy	4941	0.7	1.0	0	0	0.1	1.2	3.4
uvindex	8734	1.0	2.3	0	0	0	0	9
severerisk	4857	9.8	1.6	3	10	10	10	30

The variables used and their abbreviations are as follows: temp = ambient temperature [°C], feelslike = real feel (temperature) [°C], dew = dew point [°C], humidity = relative humidity [%], precip = precipitation [mm], precipprob = precipitation chance [%], snow = snow [mm], snowdepth = snow depth [mm], windgust = wind gust [km/h], windspeed = wind speed [km/h], winddir = wind direction [degrees], sealevelpressure = sea level pressure [mb], cloudcover = cloud cover [%], visibility = visibility [km], solarradiation = solar radiation [W/m²], solarenergy = solar energy [MJ/m²], uvindex = UV index, severerisk = severe risk.

As part of the data preprocessing step out of the timestamp, there are extracted features representing the month, day of the month, day of week, hour and year. From the perspective of traffic peaks, their impact on the air quality of the day of the week as well as the hour are important for the feature list. For the temperature and wind speed, new features are generated as the rolling average over the previous 6, 12 and 24 h. Apart from the numerical data described in Table 1, there are two additional categorical variables representing precipitation type and cloud cover. Even these two might be redundant as the data preprocessing step are encoded and used further. To avoid the bias induced by multicollinearity, all the features with a Pearson correlation coefficient higher than 0.9 were removed. At the end of the preprocessing step, 42 features were kept. The features removed due to high correlation are: “feels like”, “solar energy”, “UV-index” and “year”. The number of features is important in selecting the number of neurons on the hidden layers. The three selected candidates are calculated as $n/2$, n , $2n + 1$, where n is the number of input features.

2.3. Machine Learning Recurrent Neural Network (RNN) Models

Artificial neural networks (ANNs) represent an area in which concepts derived from other major knowledge domains such as biology, mathematics, programming, engineering, statistics, or informatics are merged with the aim to mimic the way human neurons function. The neuron is the main structural element of the brain. The human brain is composed of neurons varying between 1×10^{11} and 2×10^{11} .

By default, ANNs are considered capable to generalize very well in very specific, well-defined use cases. Moreover, they are expected to be capable of modeling nonlinear data (no direct relationship between independent and dependent variables), scalability, rational and contextualized outcome.

In a pollution-prediction problem where historical data are known, it can be considered as falling under the scope of supervised learning. By definition, supervised learning means that a model is trained with both independent and dependent variables available in the training dataset. During the training, the prediction made by the trained model during the intermediate steps is compared to the known actual values. Based on the error between the prediction and actual values, the model parameters are adjusted as part of the training. When the error is reasonable in relation to the studied problem, then the training is stopped. The last configuration of the parameters is kept as the final one.

Therefore, the main known biological neuron data processing and propagation mechanisms are implemented in ANNs. Artificial neurons are for ANNs what biological neurons are for the human brain (Figure 3). It is common for neurons to have multiple inputs and one output. Neuron inputs are signals coming from the outside environment or other neurons of the network, while the output is the signal the neuron that is propagating back to the environment or to another neuron of the network. Each connection between neurons has its own synaptic weight attached, where the information is stored. The synaptic weight represents, roughly, how important an input is for the neuron. These weights are adjusted during the training until the error is minimized according to the defined criteria.

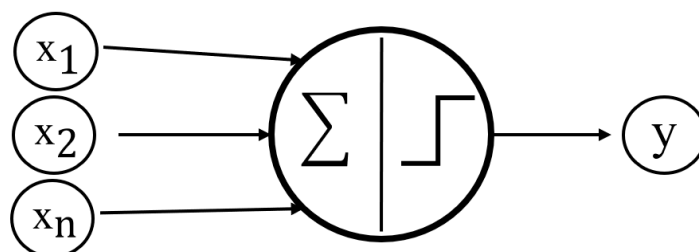


Figure 3. Artificial neuron mathematical representation.

For the use cases where the data available consist of a time series, then the type of neural networks used must be suited for this type of data. Of course, the use of feedforward neural networks can provide reasonable performance, but other aspects, mostly time-dependency-related, must be taken into consideration.

Recurrent neural networks (RNNs) represent a popular choice among professionals for time series-based problems. There are multiple types of RNNs. The standard RNNs [53] have the simplest mechanisms for processing the input data and delivering predictions while trying to minimize errors [54].

The mechanisms are simple and straightforward, but the main two issues of the standard RNNs are gradient exploding or disappearing and data morphing. However, it is not mandatory for any of them to occur [55,56].

2.3.1. Long Short-Term Memory Networks (LSTM)

A more complex RNN architecture has been proposed in order to overcome the weakness of the standard RNN (Figure 4). This new architecture is called the long short-term memory (LSTM).

Additionally, to the standard RNN, a new mechanism for keeping and taking into consideration the short-term and long-term dependencies within the data has been implemented. This new system consists of three logic gates that govern the way information flows through the network. Concretely, the relevant information is kept and the irrelevant information is discarded. The term cell is coined and incorporates the new mechanisms. At each time step, there will be three inputs (input data at time step t , hidden state at time step $t - 1$, cell state at time step $t - 1$) and two outputs (hidden state at time step t , cell state at time step t). This new architecture is called the long short-term memory (LSTM) [57]. The logic gate system consists of an input gate, an output gate and a forget gate. The input gate takes into consideration the information coming from the current time step's assigned input vector x_t and the previous step's hidden state vector h_{t-1} . Both have assigned their

own synaptic weights vector, U^i for the current step input and W^i for the previous state. After each dot, the product is computed then the results are summed. At the end of the bias, b^i , is also added. On top of this, a sigmoid function is applied. The sigmoid function retains the values between 0 and 1 (Equation (1)):

$$i_t = \sigma(x_t U^i + h_{t-1} W^i + b^i) \quad (1)$$

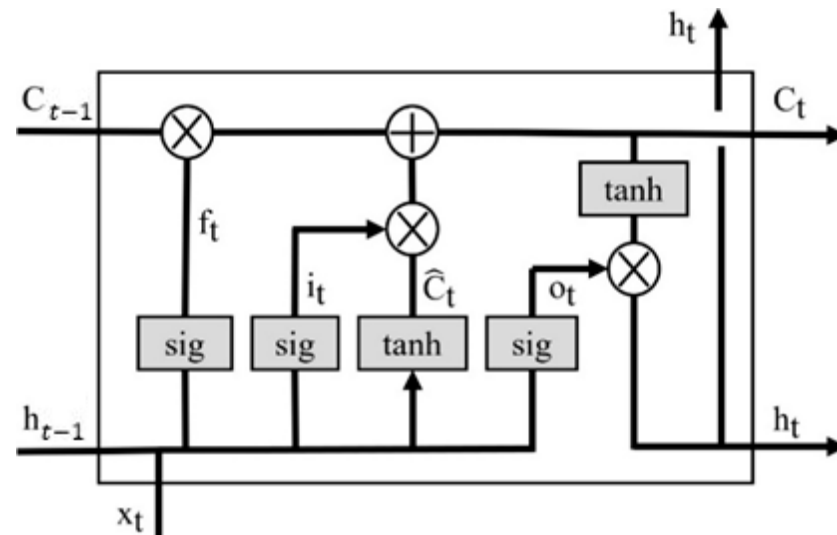


Figure 4. LSTM design overview.

As part of the input gate, a new candidate for the cell state, \hat{C}_t , is calculated based on the current time step's input, x_t , and the previous step's hidden state, h_{t-1} . This layer has its own synaptic weight and bias. On top of the computed values, a tanh activation function is applied, as in Equation (2):

$$\hat{C}_t = \tanh(x_t U^c + h_{t-1} W^c + b^c) \quad (2)$$

In a similar way to the input gate, the forget gate decides if information coming from the previous hidden state and current state's input should be forgotten (Equation (3)). As expected, the weights U^f , W^f and the bias, b^f , belong to this gate.

$$f_t = \sigma(x_t + h_{t-1} W^f + b^f) \quad (3)$$

The relevant information is passed through the input and forget gates, and then, taking into consideration the previous cell state, C_{t-1} , the new cell state for the time step t is calculated using Equation (4):

$$C_t = \sigma(i_t \cdot \hat{C}_t + f_t \cdot C_{t-1}) \quad (4)$$

The two outputs are computed using Equations (5) and (6):

$$o_t = \sigma(x_t U^o + h_{t-1} W^o + b^o) \quad (5)$$

$$h_t = \tanh(C_t) \cdot o_t \quad (6)$$

where o_t represents the output value for the current time step and h_t represents the current time step's hidden state. The time dependencies are kept in the cell state, designed for long-term memory, and the hidden state, which is designed for short-term memory. With

the gating system in place, the network predicts at timestep t using relevant information gained upstream starting from step $t - 1$.

2.3.2. Gated Recurrent Unit (GRU)

Another type of recurrent neural network inspired by the standard one is the gated recurrent unit (GRU). This architecture (Figure 5) rapidly became popular in 2014 when it was presented for the first time [58].

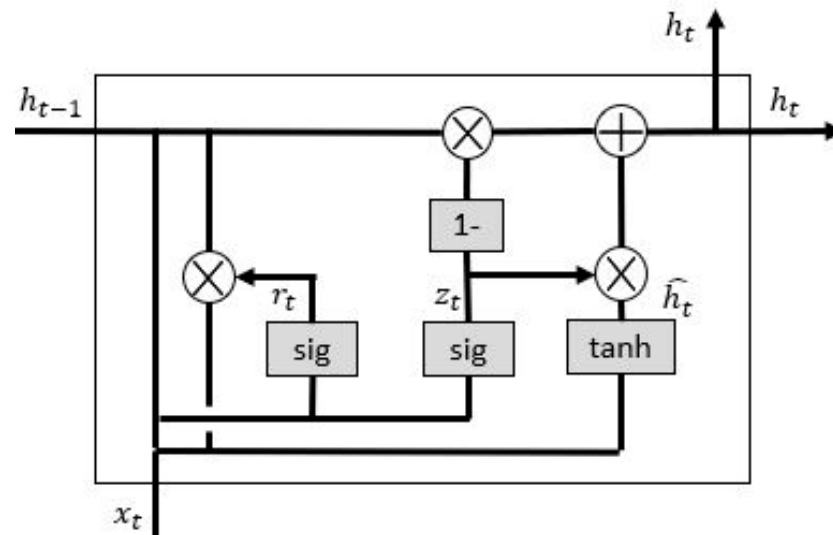


Figure 5. GRU design overview.

Similar to the LSTM, the information flow within the GRU is governed by a gate system, but with two gates instead of three. The notion of the hidden state is kept, while the notion of cell state is discarded from the design of GRU compared to LSTM. These decisions lead to a shorter training time due to the reduced computational load.

The reset gate is processing data from a short-term perspective. The functionality of this gate is similar to the LSTM's forget gate and is governed by Equation (7):

$$r_t = \sigma(x_t U^r + h_{t-1} W^r + b^r) \quad (7)$$

The update gate is used for the purpose of long-term memory and is implemented by Equation (8):

$$z_t = \sigma(x_t U^z + h_{t-1} W^z + b^z) \quad (8)$$

The same activation function, sigmoid, is used for both the forget and update gates. The difference is in the weight matrices and bias. As expected, the closer the values in the weight matrices are to 1, the more relevant are the data.

The hidden state, h_t , is calculated in two steps. The first step, Equation (9), is used to calculate a new candidate hidden state for the time step t , \hat{h}_t . The key in understanding GRU's mechanism used for the information moving upstream is the way the previous hidden state, h_{t-1} , is multiplied by reset gate vector, r_t . All precious acquired information are discarded if the values equal to 0, and are kept if the values equal to 1.

$$\hat{h}_t = \tanh(x_t U^{\hat{h}} + (h_{t-1} \cdot r_t) W^{\hat{h}} + b^{\hat{h}}) \quad (9)$$

The second step is to calculate the hidden state for the time step t as in Equation (10). The information passing through the update gate, z_t , the hidden state candidate, \hat{h}_t , and the previous hidden state, h_{t-1} , are used to modulate the output of the hidden state at the time step t .

$$h_t = (1 - z_t) \cdot h_{t-1} + z_t \cdot \hat{h}_t \quad (10)$$

For both LSTM and GRU, while the new time steps are added, the equations above are recomputed. LSTM and GRU are sophisticated designs suited for time series data. The gating systems implemented in both provide the much-needed mechanisms to capture time dependencies to avoid data morphing and overall information loss.

2.3.3. Model Performance

Model performance is analyzed from two perspectives. The first perspective is where the model generalization capability is assessed using KPIs based on error calculation. The second is represented by the learning curves that provide learning while the model converges. A rigorous approach to KPI setting must provide the opportunity to understand the magnitude of the error relative to the data used, but must also be independent to the dataset to provide the opportunity for other researchers to compare the results of their work with the current ones. The mean absolute error (MAE) is a measure of error between actual and predicted values as the average of absolute errors (Equation (11)). R^2 is a measure of how well the independent variables can explain the variance of the dependent variable (Equation (12)). MAE is relative to the dataset, while R^2 is independent.

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (11)$$

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (12)$$

where y_i and \hat{y}_i are the actual and predicted values. Additionally, the training time will also be added to the performance-related KPIs.

Being in a supervised learning paradigm, the initial dataset is split into two distinct parts. The training subset representing 70% of the initial dataset is used for training purposes only, and the remaining 30% representing the testing subset is only used for testing. None of the records in the testing subset were part of the training. The initial dataset is split into training and testing sets while keeping the temporal dependencies. All the entries are ordered by date, which means that in the testing subset, there are no data that occurred earlier than the latest in the training subset.

The selected metrics are computed for the testing subset only. The only relevant performance metrics for this kind of modeling is on the datasets that are not used during the training.

To identify the best model for the use-case, various configurable parameters had to be tested before deciding on the final model. Neural network-based models are highly configurable, so the best model selection can become time and computationally expensive. For both LSTM and GRU, the parameters in Table 3 are tested. The model configurations are built as unique configurations of each of the parameters listed below. A total number of 5832 models were trained and tested.

Table 3. LSTM and GRU parameters.

Parameter Name	Values
Optimization algorithm	Adagrad, Adam, RMSProp
Activation function	Relu, Sigmoid, Tanh
Weight initialization	LeCun normal, LeCun uniform, Xavier normal, Xavier uniform
Number of epochs	10, 30, 50
Batch size	64, 128, 256
Number of hidden layers	2
Number of neurons on each hidden layer	21, 42, 85

3. Results and Discussion

In terms of generalization capabilities, both LSTM and GRU have maximum R^2 values below 0.8. As can be seen in Figures 6 and 7, most of the tested models achieve peak performance. This leads to the conclusion that used neural-based models show potential in providing generalization capabilities in various configurations of parameters. Having such a large number of models performing at maximum is a confirmation of the versatility of LSTM and GRU, but also builds the confidence in the final selected model.

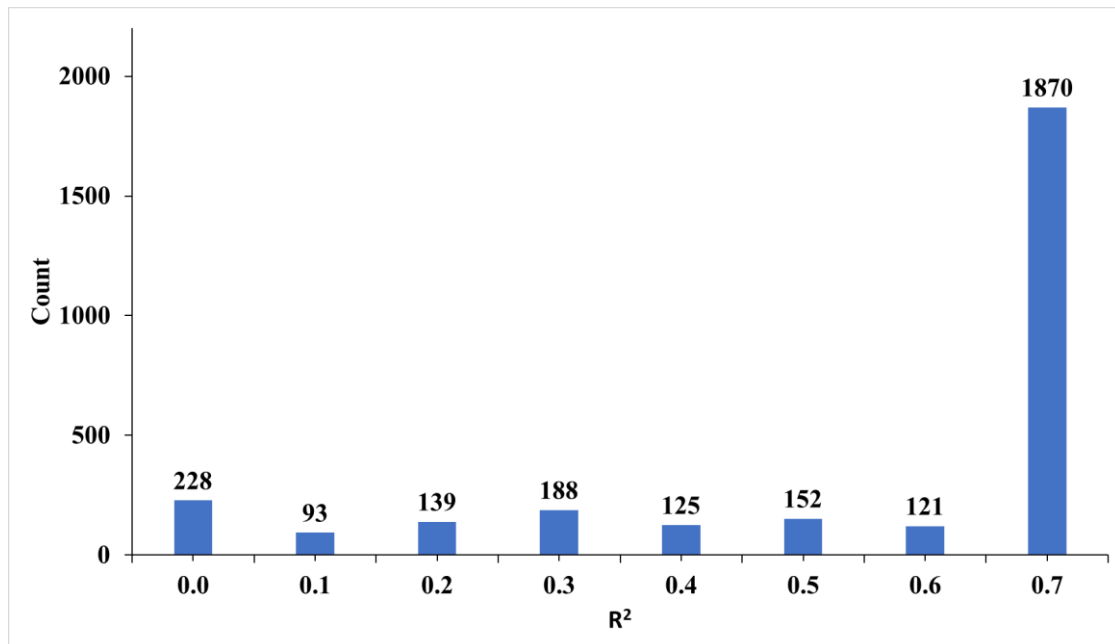


Figure 6. LSTM models' R^2 grouping by generalization capability.

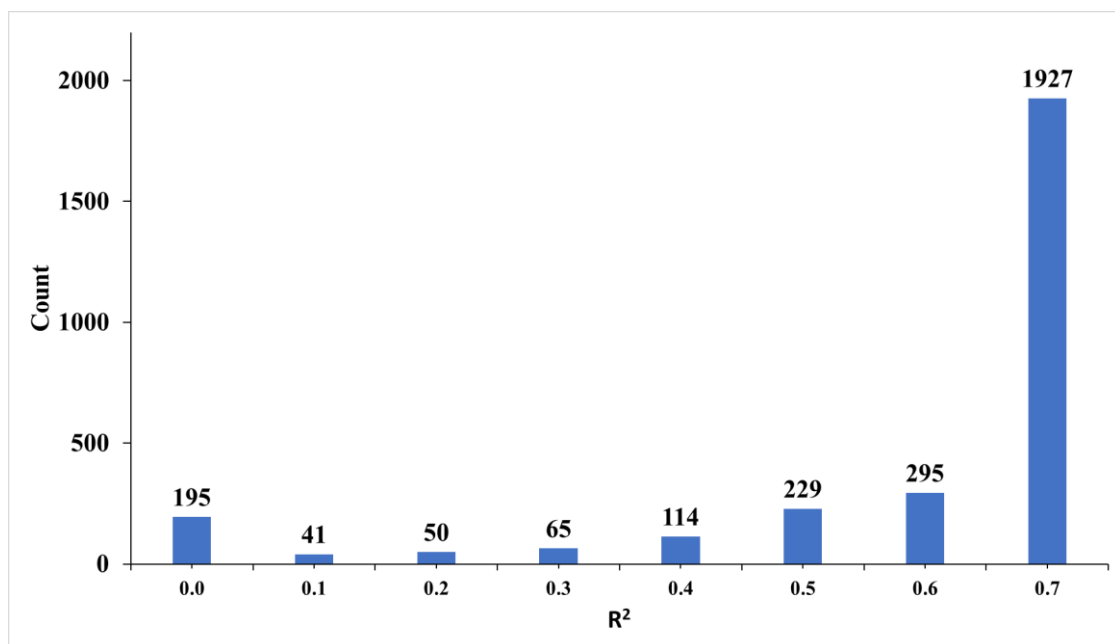


Figure 7. GRU models' R^2 grouping by generalization capability.

To narrow the list, the three best LSTM and the three best GRU models are selected and plotted in Figure 8. The performances of all selected models are very close, result-

ing in the decision of the final model to be dependent on the other criteria rather than generalization capabilities.

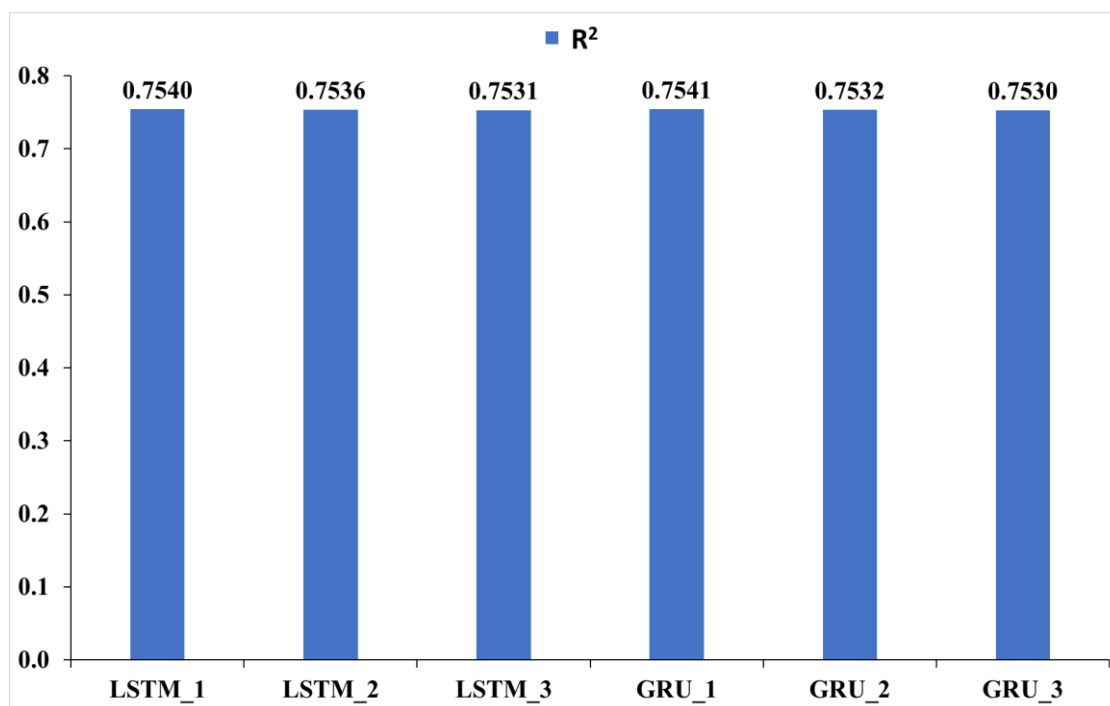


Figure 8. Top three best performing models by model type.

The six models compared in Figure 8 have the parameters presented in Table 4.

Table 4. Top LSTM and GRU models.

Parameter	LSTM_1	LSTM_2	LSTM_3	GRU_1	GRU_2	GRU_3
Optimizer	Adam	Adam	RMSProp	RMSProp	RMSProp	RMSProp
Activation function	Relu	Relu	Relu	Tanh	Tanh	Tanh
Initialization	Xavier Normal	LeCun Normal	LeCun Uniform	LeCun Uniform	LeCun Normal	LeCun Normal
Epochs	50	50	30	50	50	50
Batch	64	64	64	128	64	256
Hidden neurons on the first layer	85	42	85	85	42	85
Hidden neurons on the second layer	85	85	42	84	21	85

In terms of training time, as expected, there are important differences between the considered models (Figure 9). Training time variation within the model type can be explained by batch size and epochs. The smaller the batch size, the higher the training time. The larger the number of epochs, the higher the training time. Of course, other parameters have an impact as well, but can be negligible in the overall context.

Another aspect mentioned in the theoretical fundamentals is that when assessing GRU and LSTM, there will be a difference due to the smaller number of equations that govern GRU compared to LSTM.

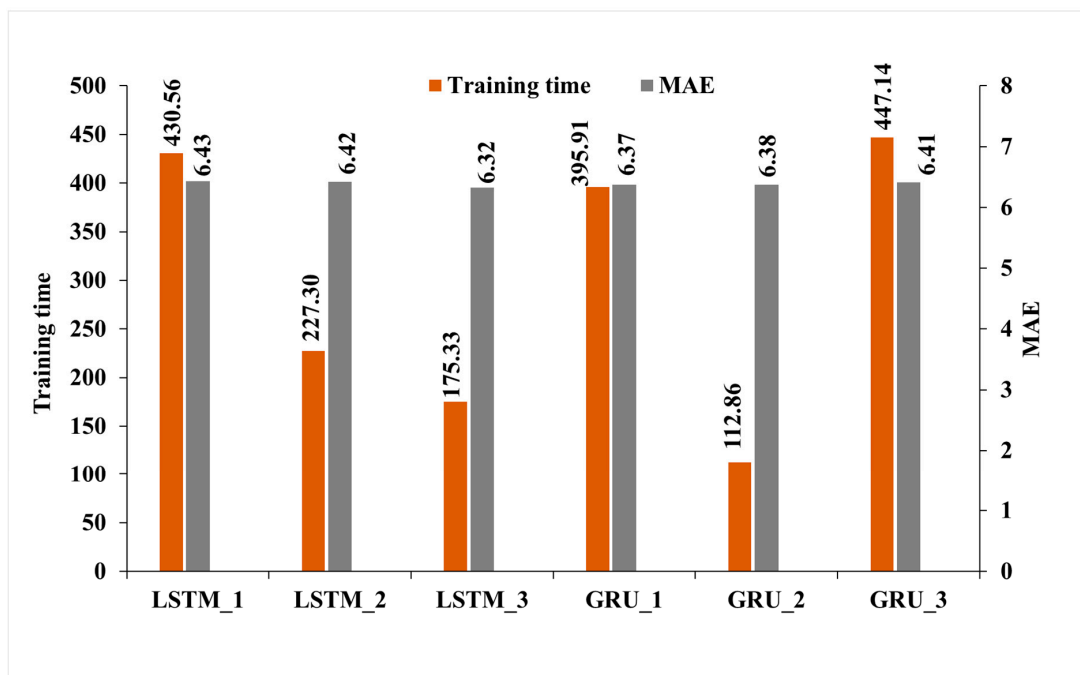


Figure 9. Training time and MAE comparison of top models.

An approach to assess how the models converge is by plotting learning curves. This type of visualization shows how the loss evolves epoch by epoch during training. The visualization is made based on the epochs since this hyperparameter is the only one in the list controlling the degree with which the models are trained. The learning curves, where the blue line is for learning on the training subset and the green line is for testing the model on its subset, can provide information if over-fitting or under-fitting should occur. For the selected LSTM (Figure 10) and GRU models (Figure 11), the solution converged correctly without spikes or multiple intersections of the graphs. Thus, the training of the convergence of the GRU model is smoother compared to LSTM.

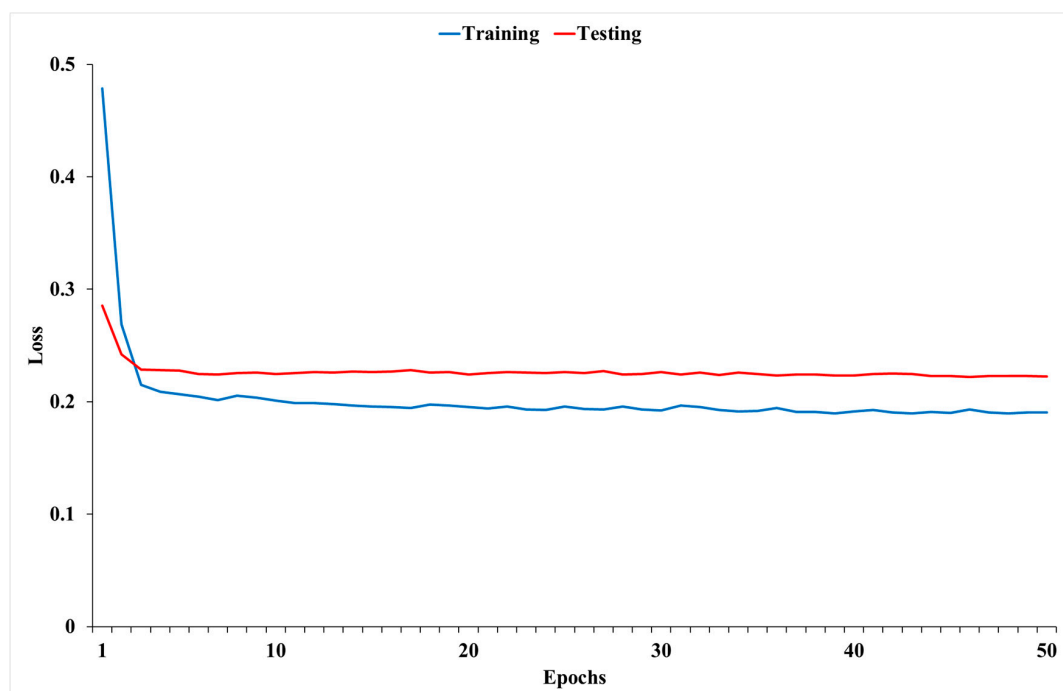


Figure 10. LSTM_2 learning curves.

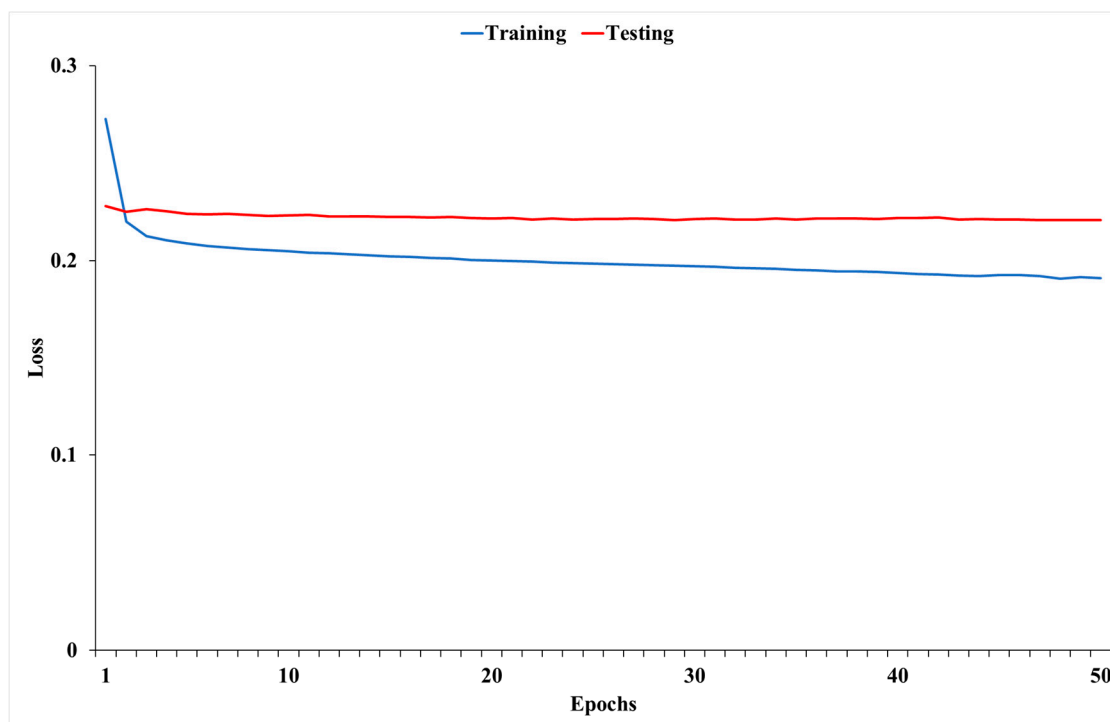


Figure 11. GRU_2 learning curves.

Once the predictive capability is revealed by computing the metrics, another way to assess the predictions with respect to the actual values is to visualize them. Three time frames were selected for visualization purposes, one extracted from the testing subset and two from outside the modeling dataset. The three time frames which are not overlapping are also for different days of the week. The first interval extracted for the testing subset contains predictions for 7–8 July 2022, Thursday and Wednesday (Figure 12). The second interval contains data for 10–11 October 2022, Monday and Tuesday (Figure 13). The third interval for 12–13 November address the predictions for a weekend (Figure 14). The graphs were produced with GRU_2.

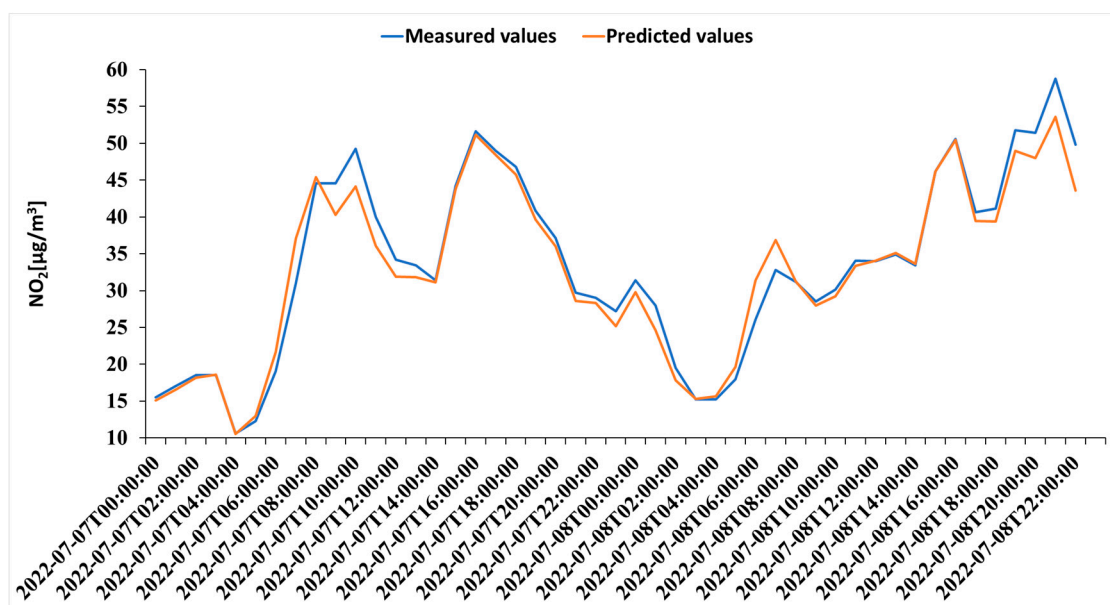


Figure 12. Predictions for 7–8 July 2022, Thursday and Wednesday.

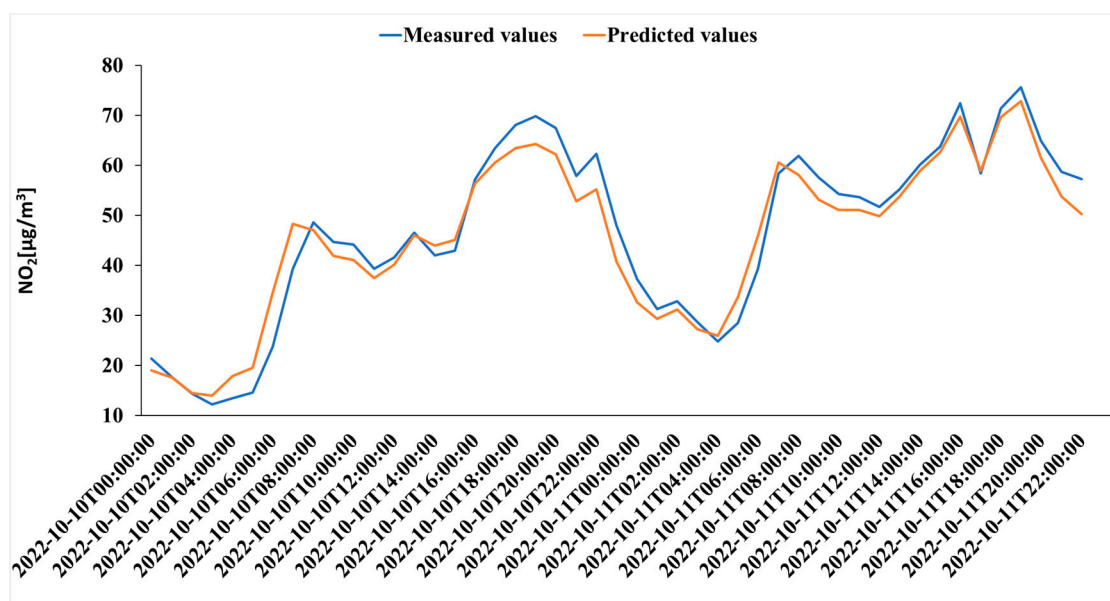


Figure 13. Predictions for 10–11 October 2022, Monday and Tuesday.

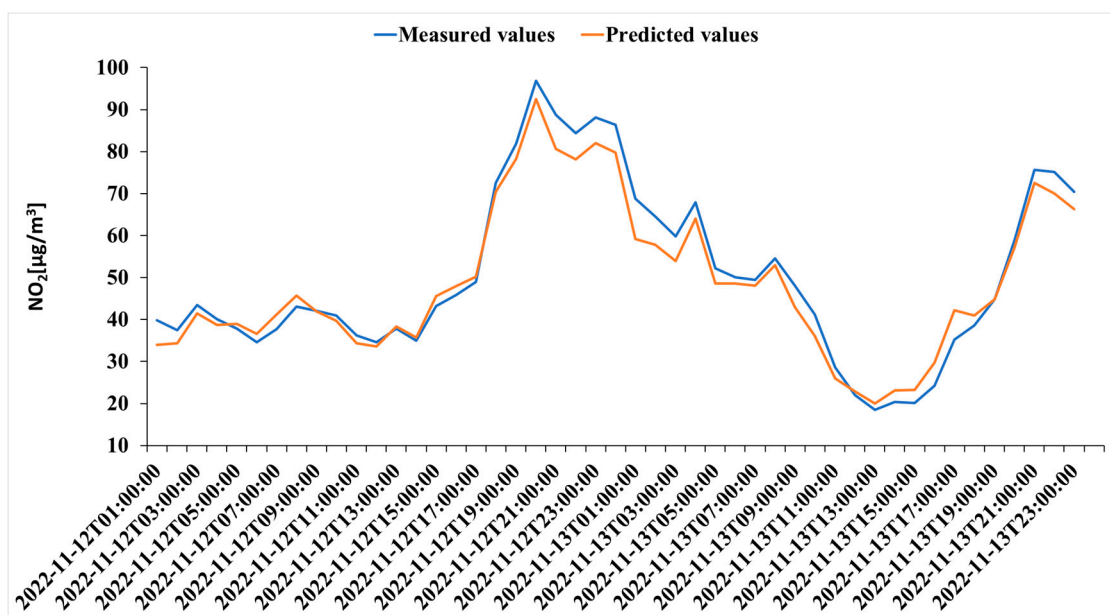


Figure 14. Predictions for 12–13 November.

As can be seen in all three visualizations, the predicted values are close to the actual measured values. The model selected can predict when the NO_2 level will increase or decrease, but in some cases, the inaccuracy is given by the magnitude of it. In this way, the tendency of the model shifts from overestimation to underestimation and vice versa. However, for all three evaluated time intervals, there are also subintervals in which the predicted values match the exact measured values.

Finally, a better understanding of the prediction results can be achieved by assessing the feature importance. As can be seen in Figure 15, the most important features are wind speed, temperature, dew point, humidity and cloud cover. Feature importance is calculated by permuting the features. Based on the impact which the permutation makes, the importance of the feature can be assessed. When the feature is not important, the model performance is then not much altered. When the feature is important, the model performance is then altered in a perceptible way.

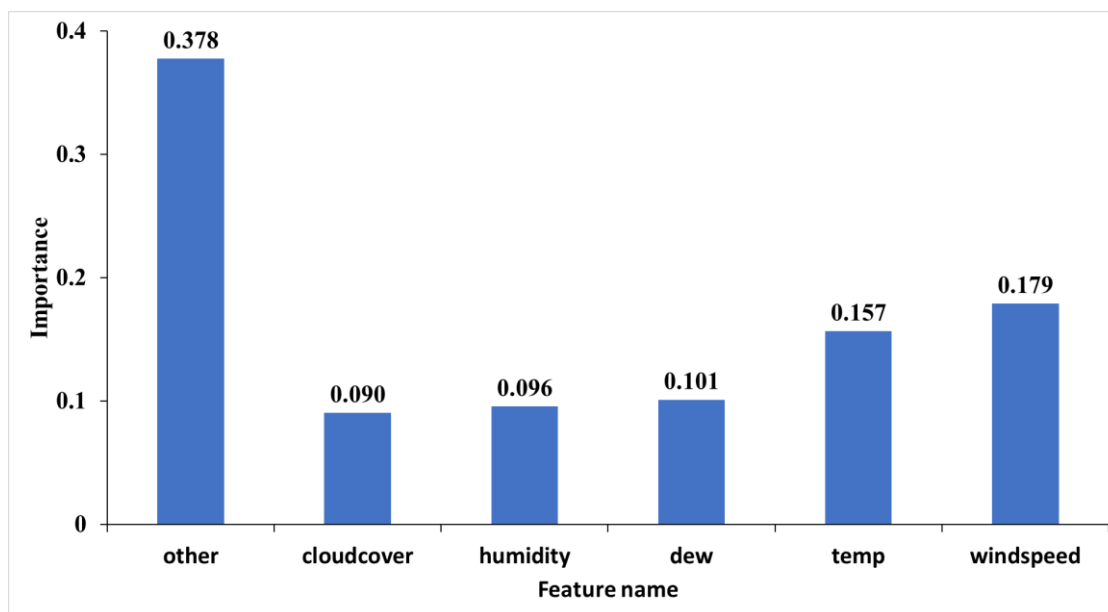


Figure 15. The most important features.

The remaining features are summed up in the last category: other. It is expected to have wind speed and temperature as top contributors, being in line with what other researchers found. Wind speed is important in dispersing the pollution; the higher the wind speed, the faster the dispersion. In this way, the polluting particles are moved away from the source. On the other hand, the impact of temperature has an explanation that can be provided by basic physics. Throughout the specified area, convection transports pollutants from ground level to higher altitudes, thus reducing the measured pollution. However, the impacts of wind speed and temperature must be studied in more detail due to their complex interactions with the environment.

4. Conclusions

LSTM and GRU represent powerful architectures that produce good performance when training data are appropriate for the studied phenomenon, and the hyperparameters are chosen accordingly. Both are highly configurable, so the probability to identify the best suited solution for the studied problem is also high. This can become problematic due to the need for building a large number of models with various configurations in order to identify the best hyperparameter configuration. Training and testing all these models lead to a process that can be characterized as computationally expensive.

As expected, pollution prediction is quite a complex task due to the need of taking into account multiple variables that impact the studied phenomenon. Not all of these variables which impact pollution levels are available. Some data such as traffic, road maintenance or accidents are not available in a big data fashion, so their usefulness is close to none. Nevertheless, the performance of the final models are good, with respect to the available data and performance obtained by other researchers on similar applications. Both LSTM and GRU are capable of providing even better performances, but the lack of independent variables needed to fully model the phenomenon makes it impossible to obtain exceptional results.

Author Contributions: G.C. and R.M., conceptualization; A.-N.B., R.M. and G.C. carried out simulations. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The datasets used and analyzed during the current study are available from the corresponding author upon request.

Acknowledgments: The research was carried out with INCDT COMOTI's support with respect to its interest in environmental sciences within project "Nucleu" as part of National Research, Development and Innovation Plan, under Romanian Ministry of Research and Digitalization, project no.: PN 23.12.02.02.

Conflicts of Interest: The authors declare no conflict of interest.

References

- Balogun, A.-L.; Tella, A.; Baloo, L.; Adebisi, N. A review of the inter-correlation of climate change, air pollution and urban sustainability using novel machine learning algorithms and spatial information science. *Urban Clim.* **2021**, *40*, 100989. [CrossRef]
- Ding, A.; Nie, W.; Huang, X.; Chi, X.; Sun, J.; Kerminen, V.M.; Xu, Z.; Guo, W.; Petäjä, T.; Yang, X.; et al. Long-term observation of air pollution-weather/climate interactions at the SORPES station: A review and outlook. *Front. Environ. Sci. Eng.* **2016**, *10*, 15. [CrossRef]
- Lelieveld, J.; Evans, J.S.; Fnais, M.; Giannadaki, D.; Pozzer, A. The contribution of outdoor air pollution sources to premature mortality on a global scale. *Nature* **2015**, *525*, 367–371. [CrossRef]
- Yang, X.; Wang, Y.; Zhao, C.; Fan, H.; Yang, Y.; Chi, Y.; Shen, L.; Yan, X. Health risk and disease burden attributable to long-term global fine-mode particles. *Chemosphere* **2022**, *287*, 132435. [CrossRef]
- Santiago, J.L.; Rivas, E.; Gamarra, A.R.; Vivanco, M.G.; Buccolieri, R.; Martilli, A.; Lechón, Y.; Martín, F. Estimates of population exposure to atmospheric pollution and health-related externalities in a real city: The impact of spatial resolution on the accuracy of results. *Sci. Total Environ.* **2020**, *819*, 152062. [CrossRef]
- Fan, H.; Zhao, C.; Yang, Y. A comprehensive analysis of the spatio-temporal variation of urban air pollution in China during 2014–2018. *Atmos. Environ.* **2020**, *220*, 117066. [CrossRef]
- Jonson, J.E.; Borken-Kleefeld, J.; Simpson, D.; Nyíri, A.; Posch, M.; Heyes, C. Impact of excess NO_x emissions from diesel cars on air quality, public health and eutrophication in Europe. *Environ. Res. Lett.* **2017**, *12*, 094017. [CrossRef]
- Leelosy, Á.; Molnár, F.; Izsák, F.; Havasi, Á.; Lagzi, I.; Mészáros, R. Dispersion modeling of air pollutants in the atmosphere: A review. *Cent. Eur. J. Geosci.* **2014**, *6*, 257–278. [CrossRef]
- EPA Nitrogen Oxides (NO_x), Why and How They Are Controlled. Technical Bulletin. 1999. Available online: <https://www3.epa.gov/ttn/catc/dir1/fnoxdoc.pdf> (accessed on 25 March 2023).
- Thurston, G.D. Outdoor Air Pollution: Sources, Atmospheric Transport, and Human Health Effects. In *International Encyclopedia of Public Health*; Heggenhougen, H.K., Ed.; Academic Press: Cambridge, MA, USA, 2008; pp. 700–712.
- Wood, E.C.; Herndon, S.C.; Onasch, T.B.; Kroll, J.H.; Canagaratna, M.R.; Kolb, C.E.; Worsnop, D.R.; Neuman, J.A.; Seila, R.; Zavala, M.; et al. A case study of ozone production, nitrogen oxides, and the radical budget in Mexico City. *Atmos. Chem. Phys.* **2009**, *9*, 2499–2517. Available online: www.atmos-chem-phys.net/9/2499/2009/ (accessed on 20 March 2023). [CrossRef]
- Bărbulescu, A.; Dumitriu, C.S.; Ilie, I.; Barbeș, S.-B. Influence of Anomalies on the Models for Nitrogen Oxides and Ozone Series. *Atmosphere* **2022**, *13*, 558. [CrossRef]
- Addison, C.C. *Nitrogen Oxides*, AccessScience; McGraw-Hill Education: New York, NY, USA, 2018.
- Manisalidis, I.; Stavropoulou, E.; Stavropoulos, A.; Bezirtzoglou, E. Environmental and health impacts of air pollution: A review. *Front. Public Health* **2020**, *8*, 14. [CrossRef] [PubMed]
- EEA. Assessing the Risks to Health from Air Pollution. 2021. Available online: <https://www.eea.europa.eu/publications/assessing-the-risks-to-health> (accessed on 15 April 2021).
- Available online: <https://www.calitateaer.ro/public/description-page/general-info-page/?locale=en> (accessed on 15 January 2023).
- Jiang, X.; Wei, P.; Luo, Y.; Li, Y. Air Pollutant Concentration Prediction Based on a CEEMDAN-FE-BiLSTM Model. *Atmosphere* **2021**, *12*, 1452. [CrossRef]
- Su, X.; An, J.; Zhang, Y.; Zhu, P.; Zhu, B. Prediction of ozone hourly concentrations by support vector machine and kernel extreme learning machine using wavelet transformation and partial least squares methods. *Atmos. Pollut. Res.* **2020**, *11*, 51–60. [CrossRef]
- Kshirsagar, A.; Shah, M. Anatomization of air quality prediction using neural networks, regression and hybrid models. *J. Clean. Prod.* **2022**, *369*, 133383. [CrossRef]
- Chen, Y.; Cui, S.; Chen, P.; Yuan, Q.; Kang, P.; Zhu, L. An LSTM-based neural network method of particulate pollution forecast in China. *Environ. Res. Lett.* **2021**, *16*, 044006. [CrossRef]
- Mani, G.; Volety, R. A comparative analysis of LSTM and ARIMA for enhanced real-time air pollutant levels forecasting using sensor fusion with ground station data. *Cogent Eng.* **2021**, *8*, 1936886. [CrossRef]
- Athira, V.; Geetha, P.; Vinayakumar, R.; Soman, K.P. DeepAirNet: Applying Recurrent Networks for Air Quality Prediction, International Conference on Computational Intelligence and Data Science (ICCIDS 2018). *Procedia Comput. Sci.* **2018**, *132*, 1394–1403.

23. Méndez, M.; Merayo, M.G.; Núñez, M. Machine learning algorithms to forecast air quality: A survey. *Artif. Intell. Rev.* **2023**. [CrossRef]
24. Masood, A.; Ahmad, K. A review on emerging artificial intelligence (AI) techniques for air pollution forecasting: Fundamentals, application and performance. *J. Clean. Prod.* **2021**, *322*, 129072. [CrossRef]
25. Ma, S.; Wu, T.; Chen, X.; Wang, Y.; Tang, H.; Yao, Y.; Wang, Y.; Zhu, Z.; Deng, J.; Wan, J.; et al. An artificial neural network chip based on two-dimensional semiconductor. *Sci. Bull.* **2022**, *67*, 270. [CrossRef]
26. Guo, Q.; He, Z.; Wang, Z. Predicting of Daily PM_{2.5} Concentration Employing Wavelet Artificial Neural Networks Based on Meteorological Elements in Shanghai, China. *Toxics* **2023**, *11*, 51. [CrossRef]
27. Agarwal, S.; Sharma, S.; Suresh, R.; Rahman, M.H.; Vranckx, S.; Maiheu, B.; Blyth, L.; Janssen, S.; Gargava, P.; Shukla, V.K.; et al. Air quality forecasting using artificial neural networks with real time dynamic error correction in highly polluted regions. *Sci. Total Environ.* **2020**, *735*, 139454. [CrossRef] [PubMed]
28. Zaman, N.A.F.K.; Kanniah, K.D.; Kaskaoutis, D.G.; Latif, M.T. Evaluation of Machine Learning Models for Estimating PM_{2.5} Concentrations across Malaysia. *Appl. Sci.* **2021**, *11*, 7326. [CrossRef]
29. Gardner, M.W.; Dorling, S.R. Neural network modeling and prediction of hourly NO_x and NO₂ concentrations in urban air in London. *Atmos. Environ.* **1999**, *33*, 709–719. [CrossRef]
30. Dragomir, C.M.; Voiculescu, M.; Constantin, D.-E.; Georgescu, L.P. Prediction of the NO₂ concentration data in an urban area using multiple regression and neuronal networks. *AIP Conf. Proc.* **2015**, *1694*, 040003.
31. Baawain, M.S.; Al-Serhi, A.S. Systematic Approach for the Prediction of Ground-Level Air Pollution (around an Industrial Port) Using an Artificial Neural Network. *Aerosol Air Qual. Res.* **2014**, *14*, 124–134. [CrossRef]
32. Jiang, D.; Zhang, Y.; Hu, X.; Zeng, Y.; Tan, J.; Shao, D. Progress in Developing an ANN Model for Air Pollution Index Forecast. *Atmos. Environ.* **2004**, *38*, 7055–7064. [CrossRef]
33. Hrust, L.; Klaić, Z.B.; Križan, J.; Antonić, O.; Hercog, P. Neural Network Forecasting of Air Pollutants Hourly Concentrations Using Optimised Temporal Averages of Meteorological Variables and Pollutant Concentrations. *Atmos. Environ.* **2009**, *43*, 5588–5596. [CrossRef]
34. Moustiris, K.P.; Ziomas, I.C.; Paliatsos, A.G. 3-Day-ahead Forecasting of Regional Pollution Index for the Pollutants NO₂, CO, SO₂, and O₃ Using Artificial Neural Networks in Athens, Greece. *Water Air Soil Pollut.* **2010**, *209*, 29–43. [CrossRef]
35. Agirre-Basurko, E.; Ibarra-Berastegi, G.; Madariaga, I. Regression and Multilayer Perceptron-based Models to Forecast Hourly O₃ and NO₂ Levels in the Bilbao Area. *Environ. Model. Softw.* **2006**, *21*, 430–446. [CrossRef]
36. Wang, W.; Men, C.; Lu, W. Online Prediction Model Based on Support Vector Machine. *Neurocomputing* **2008**, *71*, 550–558. [CrossRef]
37. Osowski, S.; Garanty, K. Forecasting of the Daily Meteorological Pollution using Wavelets and Support Vector Machine. *Eng. Appl. Artif. Intell.* **2007**, *20*, 745–755. [CrossRef]
38. Hajek, P.; Olej, V. Ozone Prediction on the Basis of Neural Networks, Support Vector Regression and Methods with Uncertainty. *Ecol. Inf.* **2012**, *12*, 31–42. [CrossRef]
39. Lin, K.P.; Pai, P.F.; Yang, S.L. Forecasting Concentrations of Air Pollutants by Logarithm Support Vector Regression with Immune Algorithms. *Appl. Math. Comput.* **2011**, *217*, 5318–5327. [CrossRef]
40. Singh, K.P.; Gupta, S.; Kumar, A.; Shukla, S.P. Linear and Nonlinear Modeling Approaches for Urban Air Quality Prediction. *Sci. Total Environ.* **2012**, *426*, 244–255. [CrossRef]
41. Nae, M.; Turnock, D. The new Bucharest: Two decades of restructuring. *Cities* **2011**, *28*, 206–219. [CrossRef]
42. Maria, A.Z.; Roxana, S.S.; Savastu, D.M.; Tautan, M.N. Impacts of exposure to air pollution, radon and climate drivers on the COVID-19 pandemic in Bucharest, Romania: A time series study. *Environ. Res.* **2022**, *212*, 113437. [CrossRef]
43. Available online: <https://www.wall-street.ro/special/romaniaverde/284397/traficul-din-bucuresti-produce-80-din-poluarea-din-aer-cate-masini-sunt-in-oras-in-acest-moment.html> (accessed on 25 April 2023).
44. Quarmby, S.; Santos, G.; Mathias, M. Air Quality Strategies and Technologies: A Rapid Review of the International Evidence. *Sustainability* **2019**, *11*, 2757. [CrossRef]
45. Law 24/15 June 2011 on Ambient Air Quality. (In Romanian). Available online: https://www.calitateaer.ro/export/sites/default/galleries/Legislation/national/Lege-nr-104_2011-calitatea-aerului-inconjurator.pdf_2063068895.pdf (accessed on 15 March 2022).
46. Iorga, G. Air pollution monitoring: A case study from Romania. In *Air Quality—Measurement and Modeling*; Sallis, P., Ed.; InTech: London, UK, 2016.
47. Levei, L.; Hoaghia, M.A.; Roman, M.; Marmureanu, L.; Moisa, C.; Levei, E.A.; Ozunu, A.; Cadar, O. Temporal trend of PM₁₀ and associated human health risk over the past decade in Cluj-Napoca city, Romania. *Appl. Sci.* **2020**, *10*, 5331. [CrossRef]
48. Bărbulescu, A.; Barbeș, L. Statistical assessment and modeling of benzene level in atmosphere in Timiș County, Romania. *Int. J. Environ. Sci. Technol.* **2022**, *19*, 817–828. [CrossRef]
49. Bărbulescu, A.; Barbeș, L.; Nazzal, Y. New model for inorganic pollutants dissipation on the northern part of the Romanian Black Sea coast. *Rom. J. Phys.* **2018**, *63*, 806.
50. Ichim, P.; Sfîcă, L. The Influence of Urban Climate on Bioclimatic Conditions in the City of Iași, Romania. *Sustainability* **2020**, *12*, 9652. [CrossRef]

51. Available online: https://calitateaer.ro/public/legislation-page/national-legislation-page/?__locale=en (accessed on 25 April 2023).
52. Visual Crossing Weather Data. 2022. Available online: <https://www.visualcrossing.com/resources/documentation/weather-data/weather-data-documentation/> (accessed on 23 November 2022).
53. Pascanu, R.; Gulcehre, C.; Cho, K.; Bengio, Y. How to Construct Deep Recurrent Neural Networks. *arXiv* **2014**, arXiv:1312.6026.
54. Wu, L.; Noels, L. Recurrent neural networks (RNNs) with dimensionality reduction and break down in computational mechanics; application to multi-scale localization step. *Comput. Methods Appl. Mech. Eng.* **2022**, *390*, 114476. [[CrossRef](#)]
55. Wei, X.; Zhang, L.; Yang, H.; Zhang, L.; Yao, Y. Machine learning for pore-water pressure time-series prediction: Application of recurrent neural networks. *Geosci. Front.* **2021**, *12*, 453–467. [[CrossRef](#)]
56. Wang, J.; Li, X.; Li, J.; Sun, Q.; Wang, H. NGCU: A new RNN model for time-series data prediction. *Big Data Res.* **2022**, *27*, 100296. [[CrossRef](#)]
57. Hochreiter, S.; Schmidhuber, J. Long Short-Term Memory. *Neural Comput.* **1997**, *9*, 1735–1780. [[CrossRef](#)]
58. Cho, K.; van Merriënboer, B.; Gulcehre, C.; Bahdanau, D.; Bougares, F.; Schwenk, H.; Bengio, Y. Learning Phrase Representations using RNN Encoder-Decoder for Statistical Machine Translation. In Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP), Doha, Qatar, 26–28 October 2014.

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.