Article

A Comparative Study of Deep Learning Models on Tropospheric Ozone Forecasting Using Feature Engineering Approach

Reza Rezaei [1,*], Behzad Naderalvojoud [2] and Gülen Güllü [1]

1   Department of Environmental Engineering, Hacettepe University, 06800 Ankara, Turkey
2   Department of Computer Engineering, Hacettepe University, 06800 Ankara, Turkey
*   Correspondence: rezarezaei2008@gmail.com

Abstract: This paper investigates the effect of the architectural design of deep learning models in combination with a feature engineering approach considering the temporal variation in the features in the case of tropospheric ozone forecasting. Although deep neural network models have shown successful results by extracting features automatically from raw data, their performance in the domain of air quality forecasting is influenced by different feature analysis approaches and model architectures. This paper proposes a simple but effective analysis of tropospheric ozone time series data that can reveal temporal phases of the ozone evolution process and assist neural network models to reflect these temporal variations. We demonstrate that addressing the ozone evolution phases when developing the model architecture improves the performance of deep neural network models. As a result, we evaluated our approach on the CNN model and showed that not only does it improve the performance of the CNN model, but also that the CNN model in combination with our approach boosts the performance of the other deep neural network models such as LSTM. The development of the CNN, LSTM-CNN, and CNN-LSTM models using the proposed approach improved the prediction performance of the models by 3.58%, 1.68%, and 3.37%, respectively.

Keywords: air pollution; feature engineering; ozone forecasting; deep neural network

1. Introduction

Tropospheric ozone is a major air quality issue that constitutes 10% of the total atmospheric ozone. Unlike stratospheric ozone, which protects life from ultraviolet radiation, tropospheric ozone has damaging impacts on human health, crop growth, and climatic conditions. An increased risk of cardiovascular and respiratory diseases, premature mortality, and some eye problems due to increased ground-level ozone exposure has been demonstrated in epidemiological studies [1–4]. The study conducted by Silva et al. (2013) shows that ozone exposure is associated with 470K annual premature deaths, globally [5]. A high ozone concentration has a detrimental effect on crop growth by reducing the carboxylation efficiency and carbon dioxide assimilation rate [6]. In a study on the ozone-induced economic losses of China, the annual loss of the agricultural sector was estimated at billions of USD [2].

Photochemical production, horizontal transport, and stratosphere-to-troposphere exchanges (STE) are the main sources of tropospheric ozone [6–8]. However, tropospheric ozone formation is dominated by in situ photochemical production. As tropospheric ozone is a secondary pollutant, the presence of precursors is important, alongside favored meteorological conditions. Nitrogen oxides (NO_x = NO_2 + NO), VOCs, and CO are the main precursors of tropospheric ozone production [6,9]. There is a direct linear relationship between the VOCs and ozone concentration, whereas the relationship between ozone and NO_x is inverse [10]. In addition to the direct role of NO_x in the photochemical chain reaction, it impacts the ozone level by reacting with hydrocarbons [11] and contributing to

the generation of hydroxyl radicals [12]. Therefore, the dependency of the ozone level on the fluctuations in NO_x concentration is higher than the other precursors.

The tropospheric ozone level is also sensitive to meteorological variables such as temperature, solar radiation, wind speed and direction, precipitation, and humidity [13–15]. For instance, a rise in temperature or solar radiation contributes to higher levels of ozone, whereas an increase in humidity causes a decrease in ozone by increasing the destruction of ozone. Moreover, as precipitation causes the ozone precursors to be washed down, the ozone concentration decreases [16,17]. Furthermore, wind-induced precursor transport affects the ozone level in both positive and negative ways.

Analysis of historical ozone data and future projections under climate change show an increasing trend in ozone concentration, especially in the urban area [6,9,18]. Hence, it is necessary to make accurate forecasts of ozone concentration to warn vulnerable groups and conduct reduction scenarios. However, (1) the multiple influential factors, (2) the complexity of the interactions between the factors, (3) the occurrence of sudden changes (extreme events), (4) dispersion of pollution, and (5) having sequential properties are the main challenges in air quality forecasting processes [19,20]. The importance of the challenges noted above can be varied under the geographical and environmental conditions and affect the quality of ozone forecasting.

Air quality forecasting models are divided into deterministic, statistical, and hybrid approaches [21,22]. In order to predict an atmospheric pollutant concentration, the deterministic approach uses physical and chemical mechanisms, whereas the statistical approach uses historical data as inputs to learn the pattern of changes in ozone concentration by relying on changes in predictor variables over time. Due to the linearity of the early statistical models (such as linear-regression-based models), improved machine learning and deep learning models have been utilized to tackle the non-linearity issue of such data. Statistical approaches have received much attention in recent years. The multi-layer architecture in deep learning algorithms provides the possibility of extracting air quality features with complex characteristics. In addition, sequential deep models, such as RNN and LSTM, are capable of learning temporal dynamic patterns from a sequence of input parameters. Some studies have shown that deep learning models can even outperform sophisticated deterministic models [23–25]. For instance, Feng et al. (2019) [24] compared the accuracy of a traditional chemical transport model, WRF-CMAQ, with machine learning and deep learning methods (i.e., random forest (RF), MLP, and the extreme learning machine (ELM)) in the tropospheric ozone forecasts. Their results demonstrate that the machine learning methods, excluding ELM, have a better prediction than the traditional WRF-CMAQ model. The worst prediction belonged to ELM, which is a linear model and could not capture the non-linear and non-smooth characteristics of the ozone formation process.

Several examples of successful applications of less complicated machine learning models in air quality prediction have been reported in the literature. Among these less complicated models, the MLP model has been widely used in air quality forecasting studies. In a study conducted by Mekparyup and Saithanu (2014) [26], the MLP and radial basis function (RBF) models with feed-forward and principal component feed-forward input nodes were applied to forecast the tropospheric ozone. Regardless of the simplicity of MLP, it shows a better performance than the advanced RBF model. The satisfactory performance of the MLP model has also been reported in other studies [27–33]. Song et al. (2019) [34] applied the principal component regression (PCR) method in the prediction of the air quality index (AQI) in Wuhan, China. According to their findings, meteorological variables significantly affected the AQI, and the model was able to predict the AQI effectively at the station level. The results of a study conducted by Su et al. (2020) [35] demonstrated the high performance of support vector regression (SVR) and the kernel extreme learning machine (KELM) in the prediction of tropospheric ozone. Their results indicate that kernel-based models are more robust than the ELM, step-wise regression (SR), and back propagation neural network (BPNN) models. The SVR model has been successfully applied to predict air quality in several studies [36–40]. Jumin et al. (2020) [41] applied a boosted decision

tree (BDTR) model for ozone prediction and compared the model's performance with that of the neural network (ANN) and linear regression (LR) models. The results indicate that the BDTR and LR models made the best and worst predictions, respectively. Plocoste and Laventure (2023) [42] used SVR, kNN, random forest regression (RFR), gradient boosting regression (GBR), Tweedie regression (TR), and Bayesian ridge regression (BRR) models to forecast PM_{10} concentration. They used the preprocessed daily averages of PM_{10}, daily average of temperature, and the daily sum of rainfall (RR) as input data. The results show a higher prediction performance of the GBR model; however, all models underestimated the PM_{10} concentration.

With the recent advances in deep learning models, the forecasting performance has even improved. Pak et al. (2018) [43] proposed a deep model for ozone forecasting in Beijing City using a CNN-LSTM hybrid network. They evaluated the proposed hybrid approach using four different architectures, including CNN-LSTM, CNN-Pooling-LSTM, CNN-LSTM-LSTM, and CNN-Pooling-LSTM-LSTM. According to the reported results, the fourth and third models, respectively, showed the best and worst performances. Overall, the evaluation metrics, i.e., RMSE, MAE, and MAPE, indicate that their proposed hybrid model, CNN-Pooling-LSTM-LSTM, outperforms the LSTM and MLP models in the prediction of the ozone level. In research conducted by Freeman et al. (2018) [44], the LSTM model was used to predict the 8 h average ozone concentration. They used a decision tree technique to reduce the number of input features, which improved the precision of the model. Their results show that the proposed LSTM model significantly outperforms the feed-forward neural network (FFNN) and ARIMA models. Eslami et al. (2019) [45] used the CNN model for the real-time prediction of the hourly tropospheric ozone concentration over Seoul, South Korea, and compared its performance with the benchmark models, including the ANN, MLP, LSTM, and a stacked autoencoder (SAE). According to their results, the CNN model shows the highest index of agreement (IOA) and the lowest mean absolute error (MAE) of the benchmark models. Similarly, the better performance of the CNN model compared to the other benchmark models was reported by Sayeed et al. (2020) [46]. They developed a five-layer deep CNN model to predict the future 24 h of ozone concentration. The performance of the model was compared with ridge regression, lasso regression, deep neural networks (DNNs), MLP, and GRU. The best index of agreement based on the hourly concentration and daily maximum concentration belonged to the CNN, with an IOA > 0.85 at the majority of the stations. All of the models underpredict the daily maximum ozone concentration.

Since the LSTM model was developed to solve the vanishing gradient problems, it has a good performance in forecasting time series and unsegmented data, such as atmospheric pollution data [47]. Liu et al. (2020) [48] developed an attention-based LSTM model to predict the $PM_{2.5}$ concentration in the next 24 h. They proposed a wind-sensitive attention mechanism with the LSTM model (WALSTM) and a wind-sensitive attention LSTM neural network with the XGBoost ensemble and weather forecast model (WALSTME-W). The prediction results from the proposed models were compared with the results from the MLP, SVR, XGBoost, and conventional LSTM models. The comparison results show an improvement in $PM_{2.5}$ forecasting using the proposed models. An aggregated LSTM-based model (ALSTM) was used by Chang et al. (2020) [49] to forecast the $PM_{2.5}$ concentration. With this approach, the data from local, external, and industrial air quality monitoring stations were combined. Moreover, the SVR, GBTR, and conventional LSTM models were developed as benchmark models. The higher performance of the proposed model compared to the benchmark models shows that the ALSTM model is capable of learning the weights of different air quality monitoring stations. The LSTM model was applied in a study conducted by Navares and Aznarte (2020) [50] to investigate the possibility of using a comprehensive model instead of multiple single models for station-scale air quality forecasting. Their objective was to impose the spatial effect of the pollutant concentrations on each other using a single comprehensive LSTM model. To this end, they proposed four different architectures using LSTM layers to capture the temporal dependencies as

well as the spatial relations between pollutant concentrations. Their results show that a single comprehensive model performs better than multiple single models in air quality forecasting. Zhang et al. (2021) [51] successfully conducted a semi-supervised bidirectional LSTM model to forecast the $PM_{2.5}$ concentration. To improve the model's prediction accuracy, an empirical mode decomposition was used in this study. Compared with the conventional LSTM model, the proposed BiLSTM model had a better performance in terms of daily and hourly error rates (RMSE, MAE, MAPE, and R^2).

Nabavi et al. (2021) [52] proposed a hybrid model to predict the ozone concentration. The model consists of the multiple linear regression-based eXtreme Gradient Boosting Machines (MLR-XGBM) and reanalyzed air quality models outputs (CAMS-EU). The model was trained and tested using 14 atmospheric pollutants' concentration and meteorological data as well as surface ozone concentration data from the satellite and the station category. In the hybrid model, the ozone estimates from the CAMS-EU model were also used as an input. The results show that the proposed hybrid model had a better performance than the baseline machine learning algorithms and the deterministic regional dynamic model.

Jia et al. (2021) [25] used a sequence-to-sequence model to make a real-time prediction system for ozone concentration. The model was trained and tested using atmospheric pollutant concentration data from 123 monitoring stations' hourly CO, O_3, SO_2, NO_2, $PM_{2.5}$, and PM_{10} measurements, as well as meteorology station data, including the temperature, relative humidity, wind speed, wind direction, precipitation, daylight hours, and air pressure. In the proposed approach, the input data encode to the GRU layer using an encoding-forecasting network. The attention mechanism concurrently applies weights to the input time-series data. The model uses the past 24 h input data to predict the ozone concentration of the future 6 h. Moreover, the WRF-Chem model was used to evaluate the results of the proposed prediction model. The performance analysis results show an accurate and stable prediction and good performance of the proposed model.

Machine learning methods are also used to analyze the air pollution issues other than tropospheric ozone pollution. Dai et al. (2022) [53] employed the PCA-MEE-ISPO-LightGBM algorithm to predict the risk of haze before the major haze events. In this study, the annual mean concentrations of NO_2, SO_2, PM_{10}, $PM_{2.5}$, and VOCs were used as indicators of the haze component vulnerability. Moreover, some other factors, including economics, population density, health services, and urban factors, were used as indicators of haze. After the calculation of each indicator's weight, five models were developed to evaluate the health risk of the population, transportation damage risk, crop damage risk, economic loss risk, and integrated risk. The findings demonstrate that the suggested model was successful in predicting the haze risk for the aforementioned categories.

Dai et al. (2022) [54] used different combinations of GARCH models with XGBoost and MLP (XGBoost-GARCH-MLP) to predict $PM_{2.5}$ volatility. From the GARCH models group, the authors used the GARCH, TARCH, EGARCH, and PARCH models. The proposed method focuses on the prediction of the fluctuation in $PM_{2.5}$ concentration to capture the change trajectory of the pollutant. To this end, hourly concentrations of $PM_{2.5}$, PM_{10}, NO_2, SO_2, O_3, and CO, along with 26 meteorological parameters, were used as the model input. Sayeed et al. (2022) [55] introduced a new technique for the post-processing of the deterministic chemical transport models' output using a deep learning algorithm. They fed the CNN model with outputs from the WRF-CMAQ modeling system to forecast the tropospheric ozone concentration. The results show an improvement in the prediction accuracy and reduction in the bias rate. Kim et al. (2022) [56] proposed a hybrid CNN-LSTM model to forecast the daily $PM_{2.5}$ concentration. The observational data and GFS (Global Forecast System) forecasts were used for training the model. Moreover, the data from ground stations and CMAQ outputs were used to evaluate the model prediction. The results show a higher prediction accuracy of the proposed model compared to the CMAQ model. In a recently published study by Zhao et al. (2023) [57], a genetic algorithm-optimized backpropagation (GA-BP) model was used to forecast the ozone concentration. The proposed model was only trained using meteorological parameters, including the

temperature, relative humidity, wind speed and direction, and visibility. The model results were compared with the results of a multiple linear regression (MLR) model, a BP neural network model, a model based on the RF algorithm, and the LSTM model. According to the results, the average relative error of the GA-BP model was much lower than the baseline models.

Nevertheless, the majority of approaches proposed in the literature have focused on the architecture of forecasting models without considering the nature of the data. For example, the CNN and LSTM models have been widely used in complex neural network architectures to extract features representing temporal dependencies between input variables [58]. However, more complex deep models are resource- and time-intensive, limiting their utility in real-world ozone forecasting applications, especially when real-time forecasts are required. On the other hand, the huge number of training parameters in deep models increases the overfitting problem and may reduce the generalizability of the models. As a result, less complex models with a comparable performance will have a greater chance of being deployed in real-world applications. In this study, we propose a simple but effective approach that improves the performance of basic deep learning models in ozone forecasting. In contrast to previous deep learning models, which do not account for the ozone evolution cycles across successive time steps, we propose incorporating these cycles into the model development. This is important because a change in the evolution of a pollutant may not be well-detected when variables are analyzed in short (hourly) or long (daily) time steps. In light of this, we propose ozone forecasting models that adhere to the cycles of ozone evolution. We also propose and make available pre-processed datasets to address the aforementioned ozone forecasting challenges. As different types of variables are considered in air quality forecasting models, they are obtained from various sources and necessitate pre-processing for handling outliers, missing values, and so on. This makes it challenging to compare models under comparable conditions. As a result, this paper contributes further by proposing a series of benchmarks for air quality forecasting. All of the datasets and source codes related to this paper are released at https://github.com/nbehzad/air-quality2, accessed on 17 January 2023.

We summarize our contribution as follows: (1) incorporating the role of ozone evolution phases into basic deep learning models, which improves the performance of deep learning models; (2) proposing a publicly available dataset for ozone forecasting.

2. Material and Methods

2.1. Study Area

Five years (2015–2019) of Istanbul air quality data, including hourly NO_2, NO_x, and O_3 concentrations, were obtained from the Turkish Ministry of Environment and Urbanization. Out of the existing 35 monitoring stations, only 8 stations contained five years of measured ozone values. Hence, the data associated with these stations were used in this study. From these stations, Alibeyköy, Esenyurt, Kadıköy, Kağıthane, and Sultanbeyli stations are urban air quality monitoring stations, and Başakşehir, Beşiktaş, and Sultangazi stations are industrial, traffic, and urban-traffic stations, respectively. Furthermore, the hourly meteorological records of five ground stations, including pressure, relative humidity, temperature, precipitation, wind speed and direction, and solar radiation were obtained from the Turkish State Meteorological Service. The location of the monitoring stations is presented in Figure 1.

2.2. Data Preprocessing

All datasets were analyzed in regard to invalid data and missing values. Tables 1 and 2 show the statistical summary of air quality and meteorological stations' data. To deal with the missing values, the use of mean values, the elimination of data lines containing missing points, and interpolation are more common methods. Using mean values for data gap filling assigns a constant value for missing points, while, due to the high temporal resolution of the input data, a gradual change in the variables' magnitude is important in the model

prediction accuracy. Therefore, this method is not a good choice for filling missing points of the time series data. Moreover, the application of the interpolation method for treating the missing values at the beginning or end of the dataset is problematic, as it leads to extrapolation. Furthermore, the elimination of data rows is not reasonable for solving the missing values when considering the sequential characteristics of meteorological and air quality data.
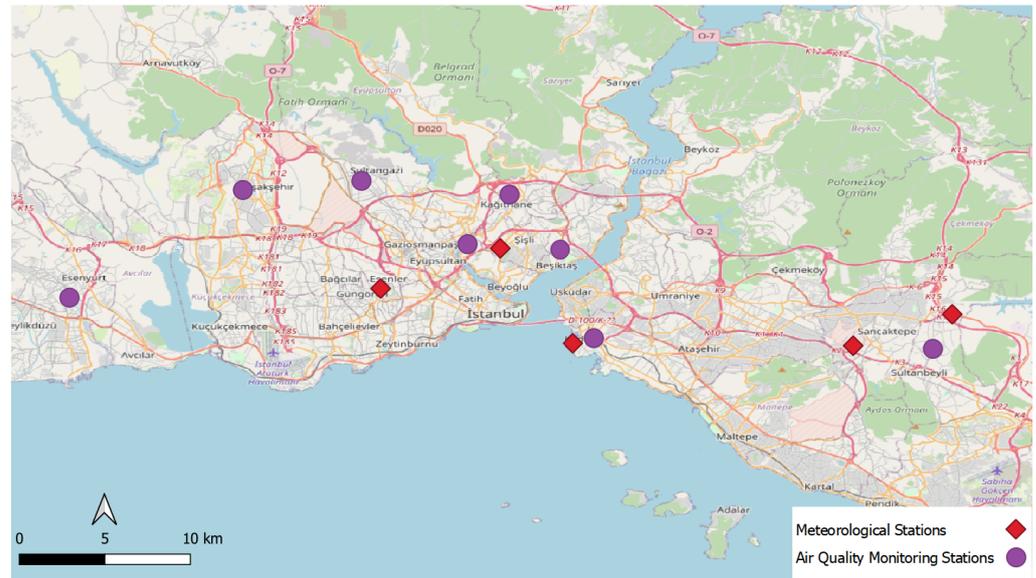

Figure 1. Location of meteorological and air quality monitoring stations.

Table 1. Statistical summary of the air quality monitoring stations' data.

Station		NO$_2$	NO$_x$	O$_3$	Station		NO$_2$	NO$_x$	O$_3$
Alibeyköy	Mean	45.583	100.627	22.802	Başakşehir	Mean	31.681	56.149	55.013
	Std.	30.603	134.910	23.088		Std.	25.632	70.125	28.306
	Q1 *	25.900	30.400	4.000		Q1	14.376	19.265	34.245
	Q2	39.000	56.400	15.400		Q2	23.740	33.124	57.900
	Q3	57.290	109.900	34.200		Q3	40.734	60.993	75.680
	Miss. (%) **	14.13				Miss. (%)	5.85		
Beşiktaş	Mean	73.898	182.003	27.330	Esenyurt	Mean	25.762	88.345	35.030
	Std.	35.057	127.797	17.759		Std.	17.922	116.479	26.203
	Q1	48.600	88.900	12.900		Q1	12.850	30.631	13.743
	Q2	68.200	145.869	24.000		Q2	21.250	53.107	32.100
	Q3	92.588	241.900	39.000		Q3	34.335	93.976	51.130
	Miss. (%)	7.67				Miss. (%)	5.45		
Kadıköy	Mean	56.129	153.261	20.158	Kağıthane	Mean	36.530	101.051	44.949
	Std.	31.364	224.346	14.927		Std.	28.642	120.980	30.859
	Q1	35.800	49.500	9.600		Q1	16.670	37.239	20.650
	Q2	49.434	85.900	16.200		Q2	28.901	62.552	42.500
	Q3	68.800	156.100	28.600		Q3	48.630	115.374	65.872
	Miss. (%)	7.36				Miss. (%)	4.27		
Sultanbeyli	Mean	19.497	45.148	58.245	Sultangazi	Mean	35.068	75.146	35.329
	Std.	20.942	75.336	33.931		Std.	22.142	81.717	23.783
	Q1	4.802	8.059	30.800		Q1	20.610	33.527	14.390
	Q2	10.819	17.473	61.600		Q2	30.925	55.171	34.050
	Q3	27.755	47.033	83.700		Q3	44.690	88.654	52.941
	Miss. (%)	3.15				Miss. (%)	3.69		

* Q1: First quartile. ** Miss. (%): The percentage of missing values in the station data.

Therefore, we adopted a local regression method to reduce the error rates caused by the missing data filling process and to construct a reliable benchmark dataset, which is one of the contributions of the study. In order to fill in the missing values of an air quality variable, we have used either the correlation between air quality variables of the

same dataset (same station) or the correlation between the same variables of the nearest neighboring station. The former approach is used where the other air quality variables of the target dataset exist; otherwise, the latter approach is used. Since a strong relationship ($R^2 > 0.9$) exists between different variables of a dataset (station), such as O_3 and NO_x, this is used to fill in missing values. Moreover, the regression coefficients demonstrate a good association between the same variables of the neighbor stations (mostly $R^2 > 0.7$). A similar regression method was also conducted to fill in the missing meteorological values using data from neighboring stations.

Table 2. Statistical summary of the meteorological stations' data.

Station	Parameter	Mean	Std.	Q1	Q2	Q3	Miss. (%)
Güngören D.	Pressure (hPa)	1008.30	6.55	1003.90	1007.80	1012.50	3.27
	R. humidity (%)	72.72	15.68	62.00	74.00	85.00	
	Temperature (°C)	15.89	7.78	9.60	15.90	22.40	
	Precipitation (mm)	0.12	1.06	0.00	0.00	0.00	
	Wind speed (ms^{-1})	3.21	1.64	1.90	3.00	4.20	
	Solar rad. (Wm^{-2})	9806.21	15,195.13	0.00	0.00	15,600.00	
Kadıköy R.	Pressure (hPa)	1014.69	6.67	1010.10	1014.10	1018.90	3.27
	R. humidity (%)	73.02	13.59	64.00	74.00	83.00	
	Temperature(°C)	16.40	7.59	10.20	16.30	22.60	
	Precipitation(mm)	0.08	0.59	0.00	0.00	0.00	
	Wind speed (ms^{-1})	3.27	1.83	1.80	2.90	4.40	
Şişli	R. humidity (%)	73.00	17.60	61.00	74.00	87.00	3.71
	Temperature(°C)	16.09	7.67	9.80	16.20	22.50	
	Precipitation (mm)	0.09	0.61	0.00	0.00	0.00	
	Wind speed (ms^{-1})	1.96	0.96	1.30	1.90	2.60	
Sancaktepe	Temperature(°C)	14.77	8.05	8.20	14.90	21.00	2.92
	Precipitation(mm)	0.10	0.70	0.00	0.00	0.00	
	Wind speed (ms^{-1})	2.52	1.72	1.10	2.20	3.60	
Samandıra H.	Pressure (hPa)	1002.11	6.77	997.50	1001.50	1006.40	4.13
	R. humidity (%)	77.45	17.27	65.00	81.00	92.00	

It should be noted that the study area has a variety of topographic features in terms of altitude, land use, etc., and that the monitoring stations are located far away from each other. Given the size and topographic complexity of the study area, the presence of a correlation between all stations is implausible. The correlation used in the data preprocessing is limited to neighboring stations, and we used data from the nearest neighboring station to fill the data gap of the modeled station. However, the neighboring stations that were used to impute missing values were not used for forecasting because they lacked 5-year air quality data. These monitoring stations are also not presented in Figure 1.

Finally, all data sets were normalized using the MinMax scaler (Equation (1)):

$$n(x) = \frac{x - \min(y)}{\max(y) - \min(y)} \tag{1}$$

where y denotes all values of the given metric, x is the metric value belonging to y, and $n(x)$ is the normalized value of x.

2.3. Dataset Analysis and Feature Engineering

Tropospheric ozone formation and decomposition are under the control of some chemical mechanisms, meteorological conditions, and spatio-temporal parameters. In most air quality forecasting studies, the specific characteristics of the study site may not be considered, and they only focus on the model architecture to improve the performance. On the contrary, we propose a simple CNN model based on the ozone evolution cycles

observed in the study area and adopted it in the other deep neural network models. Considering the spatio-temporal differentiation of the ozone evolution process, understanding these influential parameters helps to make a better prediction and contributes to a better discussion on the ozone level.

As tropospheric ozone formation is a light-dependent reaction chain, temperature and sunlight intensity are the most influential environmental factors in ozone concentration. Depending on the temperature and light intensity, the daily tropospheric ozone cycle includes the formation, stationary, and decomposition phases. Figure 2 demonstrates the temporal characteristics of tropospheric ozone formation. To apply these characteristics to the forecasting model, we analyzed the diurnal ozone pattern in the input data. The findings indicate that each daylight evolution phase lasts roughly four hours, and the evening and overnight minimums last approximately twelve hours (a multiple of four). Here, our objective was to boost the performance of deep neural network models, such as CNN, by providing a setting to better learn the cyclic ozone evolution pattern. The important point here is that we set our dataset entries to the beginning of the cycle, with no missing entries throughout the dataset. For example, station entries begin at 00:00 a.m. (1 January 2015). Therefore, a certain filter size can overlap the ozone evolution phases.
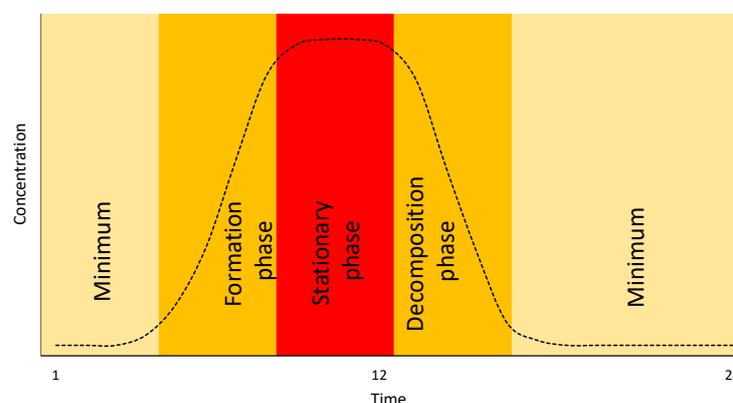

Figure 2. Theoretical daily ozone evolution cycle.

Our hypothesis is that using a filter size that corresponds to ozone evolution phases improves the model performance by allowing it to capture ozone change patterns over successive evolution phases. However, the hourly ozone concentration changes gradually because the ozone formation and decomposition are slow reactions [10,59]. Moreover, feeding the model with hourly strides makes tracking the ozone change tougher by overlapping ozone phases. To demonstrate the impact of overlapping in evolution phases, we propose multi-stride models with no overlapping (stride size equals filter size) compared to single-stride models (stride size equals 1) with maximum overlapping. Using the multi-stride setting allows the model to separate the ozone evolution phases and prepares more linear input samples for the model, whereas, in the single-stride setting overlapping, the phases increase the non-linearity of the input samples.

2.4. Proposed Neural Networks Models

In all proposed models, the ozone concentration at time $t + k$ is predicted based on parameters observed between $t - h$ and t, where $k = 1, 2, 3, \ldots, 48$ and h is 240 h (past 10 days). At any time step, the input data are represented by 10 key parameters, including 7 meteorological parameters and 3 air-quality parameters. Moreover, 2 additional parameters related to air quality are added to the input parameters: the average and maximum of NO_x for the past 8 h. Overall, the input of all models is a matrix with a shape of 240×12, and the output is a vector with the size of 48.

2.4.1. MLP Model

Multilayer perceptrons (MLPs) have been widely used in ozone forecasting, where the ozone concentration at time $t + k$ is predicted based on air-quality and meteorological data at time t. Different techniques are used to train MLPs using time-series data. The simplest method is to flatten the time series data. As the input is a matrix 240×12, this approach generates an input layer with a size of 2880. The input layer is given to the hidden layer with a size of 100 units. Finally, the output layer with a size of 48 generates the outputs of the model. This model was used as a baseline in this study to evaluate the performance of the proposed approach.

2.4.2. CNN Model

To extract features representing the behavior of the ozone concentration over time, we propose a convolutional neural network based on the ozone evolution phase intervals, with a *m-hour* filter size and *m-hour* striding without pooling operations. Based on our hypothesis, large filter sizes allow the ozone phases to be mixed together so that changes between phases may be neutralized. On the other hand, an hourly stride highlights hourly changes in the ozone concentration that are not more noticeable due to the gradual evolution of the ozone concentration. Therefore, the stride size is considered to be the same as the filter size. As a result of this convolution, we put together the effect of all input parameters in each phase of ozone evolution. Under this condition, the pooling operation may eliminate the sequence of changes during ozone phases. Figure 3 shows the details of the CNN model based on the proposed feature engineering approach.

We set m to 2, 4, 8, 16, and 24 in the experiments, and the obtained models are denoted by CNN2, ..., CNN24. In order to evaluate the effectiveness of the proposed approach, following Sayeed et al. (2020) [46], we also adopted a CNN model with a filter size of 2 and stride of 1. We also considered a MaxPooling layer with pool and stride size of 2. This model is called CNN-base in the experiments.
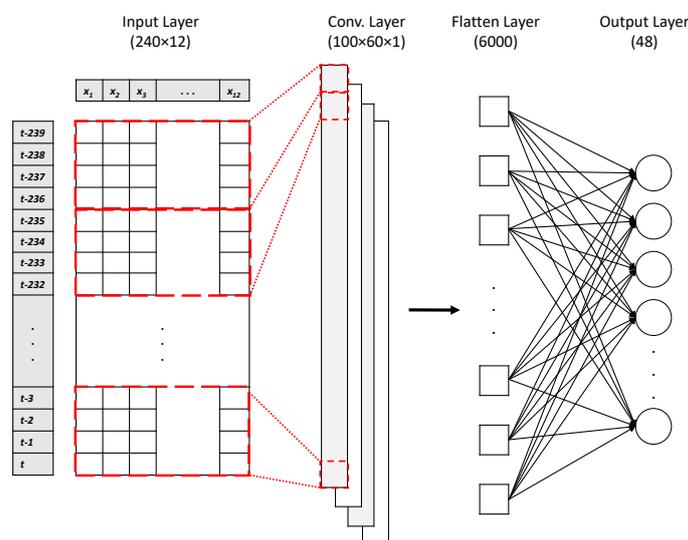

Figure 3. CNN model architecture with filter and stride size of 4.

2.4.3. LSTM-CNN Model

Long short-term memory (LSTM) is a variant of the recurrent neural network (RNN) that can deal with the vanishing gradient problem and can remember the values over arbitrary intervals. Unlike the RNN, LSTM is well-suited to the processing of time-series data, where time lags are of unknown duration [58,60]. However, as noted above, the concentration of ozone follows almost a known pattern depending on the seasonal and environmental conditions of the study area. Here, our objective was to model 10-day changes in the ozone concentration at different seasonal times using LSTM, and to use it

for predicting the next 48 h of ozone concentration. To improve the performance of the LSTM model, as shown in Figure 4, an additional CNN layer was added to the model that leverages the applied feature engineering approach. The use of the CNN layer after the LSTM enables the model to better understand the phase changes in the ozone concentration over successive days, as well as temporal dependencies between ozone phases.
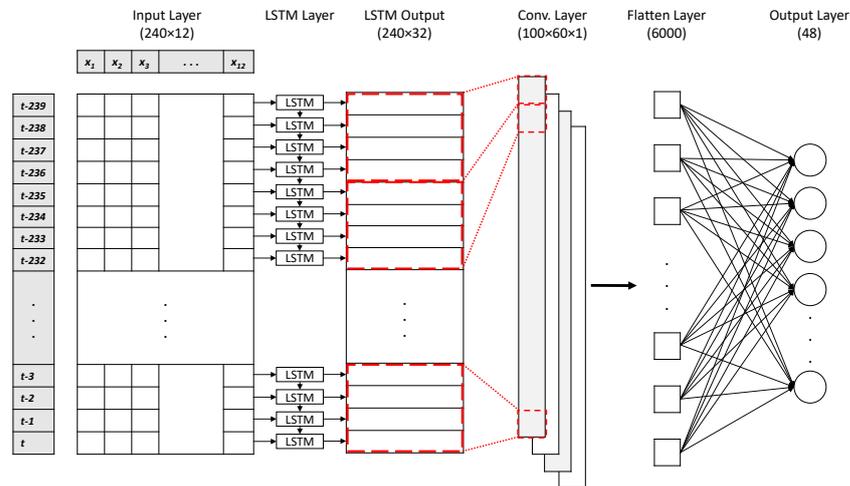

Figure 4. LSTM-CNN model architecture with filter and stride size of 4.

In order to observe the effect of the CNN layer in the performance of the proposed LSTM-CNN model, we adopted a baseline LSTM model, which is shown in Figure 5.
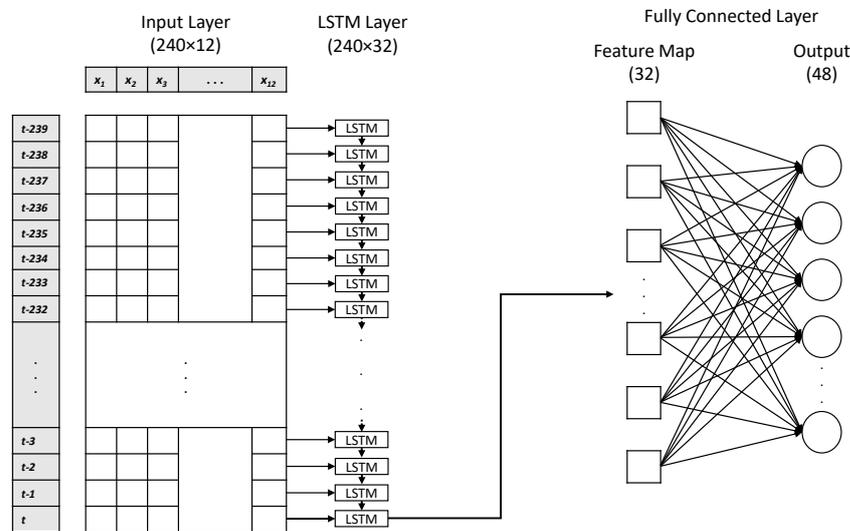

Figure 5. Baseline LSTM model.

2.4.4. CNN-LSTM Model

In the LSTM-CNN model, the CNN layer learns the ozone phases and cyclic patterns from the outputs of the LSTM layer (hidden state vectors), reflecting the hourly changes in the ozone concentration at each time step. In fact, a hidden state vector at each time step reflects all of the changes until that time step. Therefore, adding our proposed CNN layer to the LSTM outputs allows the model to better learn ozone concentration phases. To evaluate the reverse case, the CNN-LSTM model is also proposed. In this model, the LSTM layer learns the temporal dependencies from the features extracted by the CNN model. The details of the architecture of this model is shown in Figure 6. The impact of the applied feature engineering approach on the performance of the CNN-LSTM model was also assessed in the experiments.
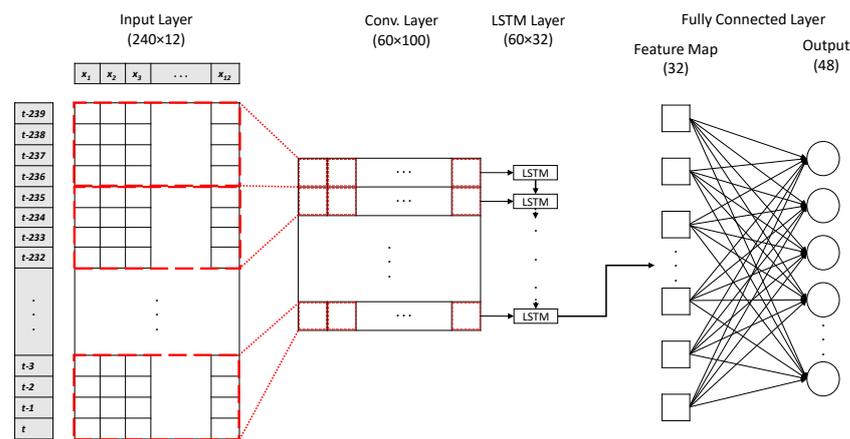
Figure 6. CNN-LSTM model architecture with filter and stride size of 4.

2.5. Model Evaluation Metrics

The mean absolute error (MAE), mean square error (MSE), and root mean square error (RMSE) are the most commonly used model evaluation metrics for ozone forecasting. The differences between the metric formulations bring some advantages or disadvantages. Therefore, to balance the disadvantages of the metrics, it is recommended to use various distance metrics. For example, while the RMSE is sensitive to extreme values (outliers) [61], the MAE metric is a good choice when the error distribution is not Gaussian [62]. On the contrary, the use of absolute values in MAE calculation is considered to be a disadvantage against RMSE, as, in many mathematical calculations, the absolute value is not favored [62]. In this study, the above-mentioned distance metrics were used to evaluate the performance of the models.

$$MAE = \frac{\Sigma_{i=1}^{n} \mid y_i - \hat{y}_i \mid}{n} \tag{2}$$

$$MSE = \frac{\Sigma_{i=1}^{n} (y_i - \hat{y}_i)^2}{n} \tag{3}$$

$$RMSE = \sqrt{\frac{\Sigma_{i=1}^{n} (y_i - \hat{y}_i)^2}{n}} \tag{4}$$

where n is the number of samples in the test dataset, and y_i and $\hat{y}_i$ are the predicted and observed values, respectively.

2.6. Implementation Details

All of the models were implemented by Keras (https://keras.io/ accessed on 17 January 2023) with TensorFlow as the backend. The MAE metric was used for the loss function, and the Adam optimization algorithm was adopted to train models through 20 epochs. The default settings were used in the definition of layers. To avoid over-fitting, a dropout approach [63] with a rate of 0.2 was used in the output of the CNN and LSTM layers.

3. Experiments

In order to show the effectiveness of involving the ozone evolution phases in the deep learning models on the prediction performance, we conducted two types of experiments. Firstly, we evaluated the effectiveness of the feature engineering approach, including filter and stride sizes based on the ozone evolution phases, on the performance of three groups of models: CNN, LSTM-CNN, and CNN-LSTM. Then, the best-performing multi-stride models and their corresponding single-stride models, described in Section 2.3, were chosen for a more detailed analysis. Finally, a monthly evaluation was conducted on the baseline

and the best-performing models. To assess the significance of feature engineering on the performance of the proposed models, the paired *t*-test was conducted on monthly performance results. Figure 7 shows the conceptual framework of this study.

In all experiments, the following convention was used to denote models and settings: $< model\ name > < filter\ size > - < stride\ size >$. In cases where the stride size was missing, it equalled the filter size. For example, while CNN4-1 indicates a CNN model with a filter size of 4 and a stride size of 1, CNN4 is equivalent to CNN4-4.
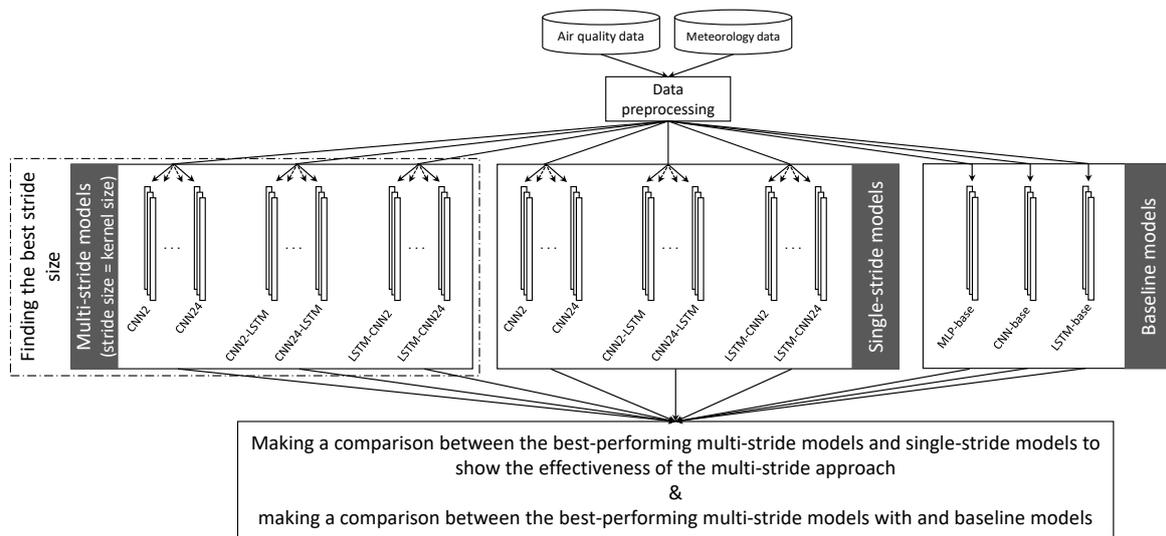

Figure 7. Conceptual framework of the study.

Empirical Results

Figure 8 shows the effect of the filter (kernel) size on the performance of the three types of ozone forecasting models, namely CNN, CNN-LSTM, and LSTM-CNN, as well as the impact of the stride size.

The impact of the filter and stride size on the performance of the CNN model can be observed in Figure 8a. This figure shows the MAE over different filter sizes while the stride is equal to the filter size (called multi-stride) or equal to 1 (called single-stride). The results indicate that the MAE values in the single-stride models are higher than those in the multi-stride models. In addition, it is observed that the CNN4 achieves the best performance (the lowest MAE) among the other CNN models. As can be seen, the CNN model does not perform well with both lower and higher filter and stride sizes, and the best performance is achieved by CNN4, which is in compliance with the ozone evolution process.

The results of the CNN-LSTM models are shown in Figure 8b. It is observed that, in the single stride models, the growth in the filter size has a negative impact on the performance of the CNN-LSTM model, so the best MAE is achieved by the CNN4-1-LSTM model. On the contrary, in the multi-stride models, where the stride size is equal to the filter size, the CNN-LSTM models with higher filter sizes perform better than lower ones, so the best model is achieved by CNN24-LSTM among all single and multi-stride models. This demonstrates that the CNN-LSTM model requires a different setting for better training. In this model, the LSTM layer is applied to the output of the CNN layer, and the final features are obtained from the LSTM model. Here, the LSTM model captures the temporal/sequential dependencies between the feature vectors obtained from the CNN layer. Therefore, the large filter sizes—for instance, 24 h—provide a better view of the ozone evolution process for the LSTM model so that the difference between two consecutive vectors is more tangible and more informative.
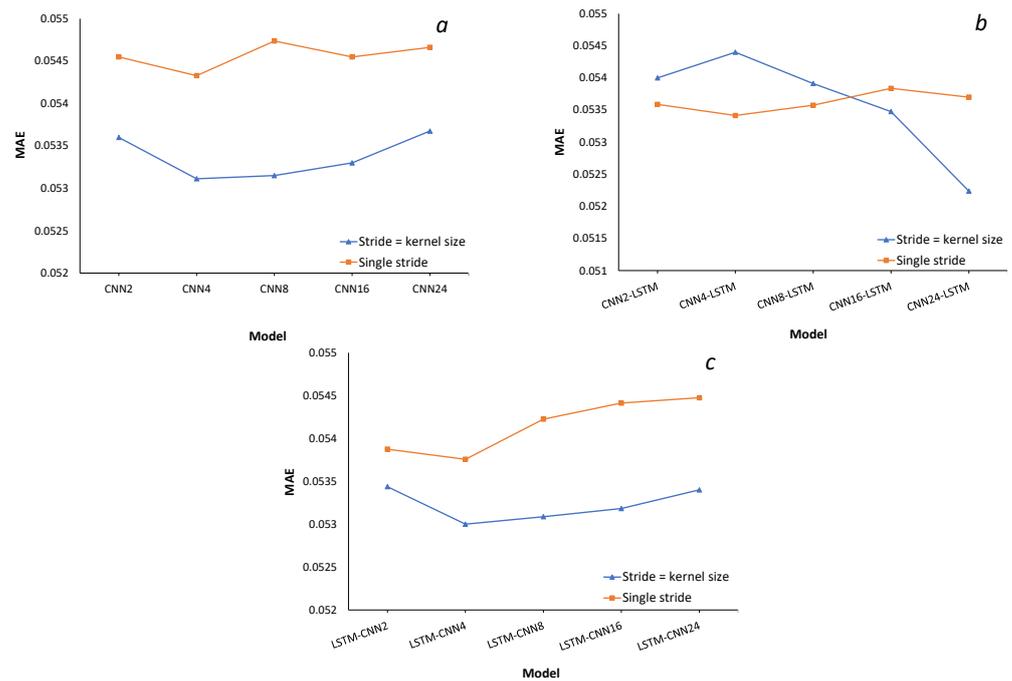
Figure 8. Mean absolute error of single and multi-stride: (**a**) CNN model, (**b**) CNN-LSTM model, and (**c**) LSTM-CNN model over different kernel sizes

The performance of the LSTM-CNN model is shown in Figure 8c. Unlike CNN-LSTM, the LSTM-CNN model supports our hypothesis that the best models are achieved by a filter size of 4 in both single and multi-stride settings, which is in compliance with the ozone evolution process. Since the final feature vectors are produced from the CNN layer, the performance of the LSTM-CNN and CNN models almost follows the same patterns as our hypothesis. According to our hypothesis, the single-stride feature extraction approach cannot provide better features for ozone forecasting models. In addition, filter sizes that are more consistent with ozone evolution phases are more effective than the others.

In the second experiment, the yearly and monthly performances of different models, indicated by MAE, MSE, and RMSE metrics, are compared to each other in order to determine the best ozone forecasting model. Figure 9 shows the yearly performance of four groups of models, including baseline models, LSTM-CNN, CNN-LSTM, and CNN, evaluated by three distance metrics. Each group includes the best-performing model identified from the first experiment and its equivalent single-stride model. Moreover, a baseline group, including MLP, LSTM, and CNN models described in Section 2.4, was considered in our experiments.

From Figure 9, it is observed that involving the ozone evolution phases in the model has a substantial impact on the performance of the three groups of models, especially on the CNN group. The performance evaluation metrics show that the CNN24-LSTM model achieves the best performance among all models.

As the seasonal variables are determinant factors in ozone formation patterns, the models' prediction accuracy may differ between seasons. Therefore, this should be considered in the evaluation of models in order to make a more detailed judgment on their performance. Hence, the statistical analysis hereafter was conducted on the monthly error values of the models. Table 3 indicates the monthly performance of the proposed and baseline models using three evaluation metrics. To show the significance of the difference between the performance of the proposed multi-stride and single-stride models and between the multi-stride and baseline models, the paired *t*-test was adopted. The paired *t*-test was performed on the monthly errors, shown in Table 3, to evaluate the statistical significance of the improvements associated with the proposed approach.
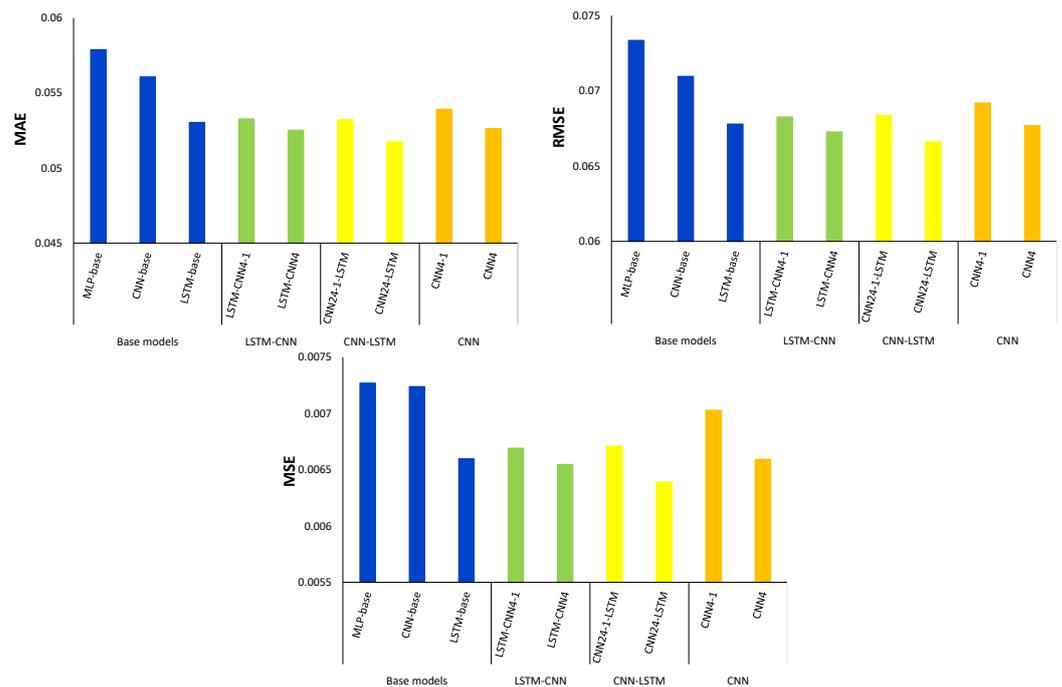
Figure 9. Yearly performance of the proposed and baseline models.

Table 3. Monthly performance of the proposed and baseline models using three evaluation metrics.

	Models	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
MSE	MLP-base	0.0063	0.0085	0.0102	0.0113	0.0111	0.0078	0.0064	0.0058	0.0070	0.0043	0.0050	0.0036
	CNN-base	0.0061	0.0089	0.0108	0.0113	0.0110	0.0078	0.0062	0.0063	0.0066	0.0041	0.0045	0.0033
	LSTM-base	0.0059	0.0073	0.0095	0.0103	0.0099	0.0069	0.0052	0.0053	0.0074	0.0043	0.0042	0.0031
	LSTM-CNN4-1	0.0060	0.0078	0.0100	0.0107	0.0104	0.0071	0.0054	0.0052	0.0064	0.0037	0.0044	0.0032
	LSTM-CNN4	0.0060	0.0073	0.0100	0.0106	0.0102	0.0069	0.0052	0.0050	0.0064	0.0036	0.0043	0.0031
	CNN24-1-LSTM	0.0061	0.0073	0.0094	0.0108	0.0103	0.0070	0.0056	0.0054	0.0069	0.0041	0.0045	0.0032
	CNN24-LSTM	0.0058	0.0070	0.0092	0.0102	0.0098	0.0068	0.0054	0.0052	0.0066	0.0039	0.0040	0.0029
	CNN4-1	0.0065	0.0084	0.0097	0.0115	0.0108	0.0069	0.0056	0.0053	0.0071	0.0041	0.0050	0.0035
	CNN4	0.0059	0.0077	0.0095	0.0106	0.0104	0.0070	0.0055	0.0052	0.0065	0.0038	0.0044	0.0028
RMSE	MLP-base	0.0673	0.0799	0.0882	0.0938	0.0910	0.0773	0.0721	0.0682	0.0705	0.0581	0.0609	0.0530
	CNN-base	0.0646	0.0801	0.0890	0.0921	0.0881	0.0750	0.0682	0.0689	0.0666	0.0549	0.0562	0.0482
	LSTM-base	0.0635	0.0728	0.0835	0.0879	0.0844	0.0711	0.0629	0.0648	0.0692	0.0556	0.0535	0.0445
	LSTM-CNN4-1	0.0634	0.0751	0.0862	0.0895	0.0860	0.0718	0.0638	0.0639	0.0653	0.0522	0.0552	0.0469
	LSTM-CNN4	0.0632	0.0730	0.0856	0.0890	0.0845	0.0706	0.0626	0.0627	0.0648	0.0516	0.0543	0.0455
	CNN24-1-LSTM	0.0642	0.0729	0.0829	0.0898	0.0856	0.0713	0.0652	0.0650	0.0681	0.0545	0.0551	0.0459
	CNN24-LSTM	0.0625	0.0718	0.0822	0.0874	0.0831	0.0703	0.0633	0.0638	0.0659	0.0535	0.0520	0.0438
	CNN4-1	0.0650	0.0764	0.0828	0.0924	0.0856	0.0686	0.0639	0.0636	0.0685	0.0551	0.0588	0.0497
	CNN4	0.0632	0.0747	0.0835	0.0892	0.0855	0.0712	0.0643	0.0633	0.0657	0.0529	0.0548	0.0441
MAE	MLP-base	0.0539	0.0648	0.0713	0.0737	0.0727	0.0596	0.0561	0.0522	0.0561	0.0451	0.0477	0.0414
	CNN-base	0.0513	0.0649	0.0731	0.0718	0.0705	0.0585	0.0534	0.0531	0.0531	0.0425	0.0439	0.0371
	LSTM-base	0.0494	0.0577	0.0670	0.0686	0.0672	0.0542	0.0485	0.0493	0.0543	0.0431	0.0420	0.0351
	LSTM-CNN4-1	0.0502	0.0601	0.0689	0.0695	0.0686	0.0551	0.0489	0.0480	0.0514	0.0397	0.0432	0.0357
	LSTM-CNN4	0.0498	0.0585	0.0688	0.0691	0.0677	0.0539	0.0482	0.0470	0.0512	0.0393	0.0424	0.0345
	CNN24-1-LSTM	0.0502	0.0577	0.0663	0.0698	0.0678	0.0541	0.0500	0.0491	0.0530	0.0414	0.0429	0.0365
	CNN24-LSTM	0.0494	0.0572	0.0649	0.0678	0.0658	0.0533	0.0482	0.0478	0.0515	0.0406	0.0407	0.0340
	CNN4-1	0.0519	0.0615	0.0661	0.0716	0.0685	0.0523	0.0485	0.0470	0.0536	0.0419	0.0457	0.0385
	CNN4	0.0498	0.0596	0.0667	0.0694	0.0681	0.0539	0.0490	0.0472	0.0518	0.0400	0.0425	0.0337

The results of the paired *t*-test are reported in Table 4. The test results indicate that the multi-stride models, where the stride size equals the filter size, outperform the corresponding single-stride models significantly with respect to all evaluation metrics ($p < 0.05$).

According to the *t*-test result, the CNN4 model, using our proposed approach, significantly outperforms the CNN-base, CNN4-1, and MLP-base models, as well as LSTM-base and some single-stride hybrid models, such as LSTM-CNN4-1 and CNN24-1-LSTM, although the improvements are not statistically significant in some cases. Overall, the CNN4, LSTM-CNN4, and CNN24-LSTM models, by relying on the proposed feature engineering approach, perform much better than the base and single-stride models for ozone forecasting. From Table 4, we also observe that, while the performance of the CNN24-LSTM is significantly higher than the LSTM-base model, there is no statistically significant difference between the performance of the LSTM-base model and that of the the proposed CNN4 and LSTM-CNN4 models.

Table 4. The results of paired *t*-test analysis (*p*-values) performed on the monthly calculated performance metrics.

	Models	MLP-Base	CNN-Base	LSTM-Base	LSTM-CNN4-1	LSTM-CNN4	CNN24-1-LSTM	CNN24-LSTM	CNN4-1
MSE	CNN-base	0.7692							
	LSTM-base	0.0006	0.0078						
	LSTM-CNN4-1	0.0000	0.0003	0.4847					
	LSTM-CNN4	0.0000	0.0003	0.6821	0.0030				
	CNN24-1-LSTM	0.0001	0.0103	0.2080	0.8447	0.0987			
	CNN24-LSTM	0.0000	0.0003	0.0083	0.0099	0.1265	0.0000		
	CNN4-1	0.0482	0.2517	0.0141	0.0072	0.0012	0.0118	0.0006	
	CNN4	0.0000	0.0002	0.9655	0.0703	0.5359	0.1233	0.0284	0.0006
RMSE	CNN-base	0.0017							
	LSTM-base	0.0000	0.0021						
	LSTM-CNN4-1	0.0000	0.0001	0.4757					
	LSTM-CNN4	0.0000	0.0000	0.3834	0.0000				
	CNN24-1-LSTM	0.0000	0.0037	0.1055	0.8204	0.0362			
	CNN24-LSTM	0.0000	0.0000	0.0008	0.0067	0.1792	0.0000		
	CNN4-1	0.0001	0.0915	0.0973	0.2032	0.0169	0.2170	0.0055	
	CNN4	0.0000	0.0000	0.8485	0.1020	0.2233	0.0697	0.0103	0.0454
MAE	CNN-base	0.0068							
	LSTM-base	0.0000	0.0013						
	LSTM-CNN4-1	0.0000	0.0001	0.6583					
	LSTM-CNN4	0.0000	0.0000	0.3091	0.0001				
	CNN24-1-LSTM	0.0000	0.0015	0.5357	0.9095	0.1048			
	CNN24-LSTM	0.0000	0.0000	0.0002	0.0035	0.0827	0.0000		
	CNN4-1	0.0000	0.0408	0.2149	0.2961	0.0384	0.2233	0.0036	
	CNN4	0.0000	0.0000	0.3876	0.0229	0.6471	0.1203	0.0174	0.0349

As a qualitative analysis, a 48 h forecasting instance of the proposed and baseline models associated with the Esenyurt air quality monitoring station is shown in Figure 10. Among the baseline models, MLP-base shows the worst prediction. According to this 48 h example forecast, as shown in Figure 10, CNN4 and LSTM-CNN4, which fully support our hypothesis, predict the peak ozone concentration better than the CNN24-LSTM, although CNN24-LSTM achieves the best yearly performance.

Figure 11 illustrates the observed and predicted ozone concentrations over the test set belonging to 2019 using the proposed models. The graphs indicate a better performance of the CNN24-LSTM model during the winter and early spring period (days 0 to 100). As seen, the peak ozone levels are better predicted by the CNN24-LSTM model and a smoother forecast is produced, although these points are underpredicted by all of the models. On the other hand, it is observed that the minimum ozone levels are overpredicted by the models.
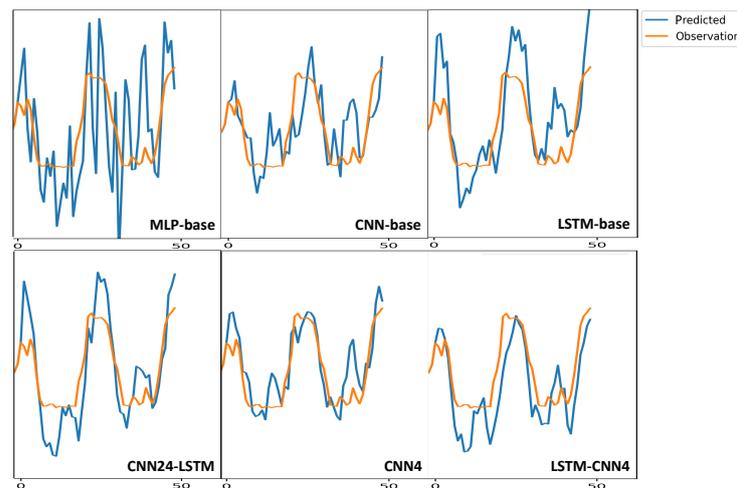
Figure 10. A 48 h ozone forecasting by the proposed and baseline models for Esenyurt station (predicted period: 16 April 2019 8:00 a.m.–18 April 2019 8:00 a.m.).
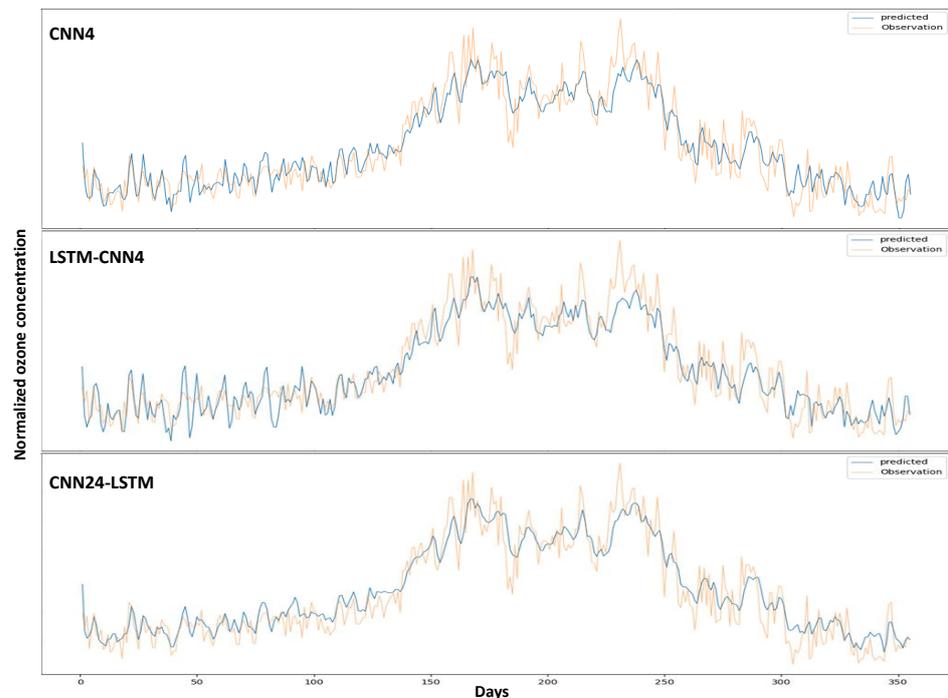

Figure 11. Yearly ozone forecasting by the proposed three models over Esenyurt station.

4. Discussion

In order to examine the effect of the proposed approach on the model performance, firstly, the single-stride models were compared with the multi-stride models. The results, shown in Figure 8, illustrate that the multi-stride models have a better performance (smaller MAE value) than the single-stride models. For instance, the performance of the proposed multi-stride models (CNN4, CNN24-LSTM, and LSTM-CNN4) is higher than the performance of the MLP-base and CNN-base models ($p < 0.05$), as well as single-stride models. This supports our hypothesis that the models cannot be well-trained using feature maps extracted on an hourly basis. In addition, we observed that the CNN and LSTM-CNN models performed well with a 4 h filter size, which is consistent with the ozone evolution process observed in the study area. However, the performance of the CNN24-LSTM model showed that the CNN-based 24 h features are more suitable for the LSTM model. This means that the LSTM model can perform better on the sequence of feature maps obtained from the CNN model over daily ozone evolution (24 h convolving). The main difference

between LSTM-CNN and CNN-LSTM is that the output of the LSTM layer in each time step reflects the effect of all previous time steps, whereas the 4 h convolving process only considers the changes in 4 h. Therefore, applying the 4 h convolving process to the output of the LSTM layer, in LSTM-CNN4, can better reflect the overall ozone evolution pattern. Furthermore, the use of the LSTM model with feature maps obtained from daily convolving in the CNN24-LSTM model significantly improved the performance of the LSTM-base model ($p < 0.05$ in Table 4).

According to the results shown in Table 3, there were 3.58%, 1.68%, and 3.37% improvements in the mean performance (average of MAE, MSE, and RMSE) of the proposed CNN4, LSTM-CNN4, and CNN24-LSTM relative to their corresponding single-stride models, respectively. Moreover, the averaged performance of the proposed CNN4 model was 6.56% higher than the CNN-base model.

As a result of the monthly analysis, setting the kernel and stride sizes based on daily ozone evolution phases used in our feature engineering approach has a positive impact on the monthly prediction performances. The largest decline in the magnitude of the error was associated with March, February, July, and August, respectively. The forecasting power of all models is pushed to its limit when the maximum and minimum points are predicted. In air quality forecasting studies, an underprediction of the peak points is a general issue in machine learning and deep learning models [40,46,64]. Our findings showed that the peak points of the ozone concentration, especially in the late spring and summer months, are underpredicted by all of the models (Figure 11).

5. Conclusions

In this study, we investigated the performance of deep learning models in the prediction of ozone concentration. The temporal variation in ozone concentration was imposed on the models, considering the evolution phases in the setting of stride and filter sizes. The idea behind the proposed approach is that, as the daily evolution of ozone can be divided into separate phases, the time steps of the dataset can also be divided into phases to make more linear samples. The results indicate that the proposed approach has a significant impact on the performance of deep learning models, including CNNs.

Our findings show that, whereas the CNN4 and LSTM-CNN4 models perform well with a 4 h filter and stride size, CNN24-LSTM performs with a 24 h filter and stride size in Istanbul data. Our results obtained from Istanbul data also indicate that the filter and stride size used in the CNN model are related to the ozone evolution phases observed in the study area, although this conclusion requires more empirical evaluation.

In our future work, we will investigate the impact of imposing the ozone evolution phases on the performance of other deep learning architectures in combination with the impact of the neighboring stations (spatial forecasting) and boundary conditions.

Abbreviations

The following abbreviations are used in this manuscript:

CMAQ	Community Multiscale Air Quality modeling system
CNN	Convolutional Neural Network
LSTM	Long Short-Term Memory
MAE	Mean Absolute Error
MLP	Multilayer Perceptron
MSE	Mean Square Error
RMSE	Root Mean Square Error
RNN	Recurrent Neural Network
STE	Stratosphere-to-Troposphere Exchanges
VOCs	Volatile Organic Compounds
WRF	Weather Research and Forecasting model

References

1. Fleming, Z.L.; Doherty, R.M.; Von Schneidemesser, E.; Malley, C.; Cooper, O.R.; Pinto, J.P.; Colette, A.; Xu, X.; Simpson, D.; Schultz, M.G.; et al. Tropospheric Ozone Assessment Report: Present-day ozone distribution and trends relevant to human health. *Elem. Sci. Anthr.* **2018**, *6*. [CrossRef]
2. Feng, Z.; De Marco, A.; Anav, A.; Gualtieri, M.; Sicard, P.; Tian, H.; Fornasier, F.; Tao, F.; Guo, A.; Paoletti, E. Economic losses due to ozone impacts on human health, forest productivity and crop yield across China. *Environ. Int.* **2019**, *131*, 104966. [CrossRef] [PubMed]
3. Kim, Y.; Choi, Y.H.; Kim, M.K.; Paik, H.J.; Kim, D.H. Different adverse effects of air pollutants on dry eye disease: Ozone, PM2.5, and PM10. *Environ. Pollut.* **2020**, *265*, 115039. [CrossRef]
4. Maji, K.J.; Namdeo, A. Continuous increases of surface ozone and associated premature mortality growth in China during 2015–2019. *Environ. Pollut.* **2021**, *269*, 116183. [CrossRef] [PubMed]
5. Silva, R.A.; West, J.J.; Zhang, Y.; Anenberg, S.C.; Lamarque, J.F.; Shindell, D.T.; Collins, W.J.; Dalsoren, S.; Faluvegi, G.; Folberth, G.; et al. Global premature mortality due to anthropogenic outdoor air pollution and the contribution of past climate change. *Environ. Res. Lett.* **2013**, *8*, 034005. [CrossRef]
6. Tiwari, S.; Agrawal, M. *Tropospheric Ozone and Its Impacts on Crop Plants*; Springer: Delhi, India, 2018; p. 201.
7. Škerlak, B.; Sprenger, M.; Wernli, H. A global climatology of stratosphere-troposphere exchange using the ERA-Interim data set from 1979 to 2011. *Atmos. Chem. Phys.* **2014**, *14*, 913–937. [CrossRef]
8. Hofmann, C.; Kerkweg, A.; Hoor, P.; Jöckel, P. Stratosphere-troposphere exchange in the vicinity of a tropopause fold. *Atmos. Chem. Phys. Discuss* **2016**, 1–26. [CrossRef]
9. Revell, L.E.; Tummon, F.; Stenke, A.; Sukhodolov, T.; Coulon, A.; Rozanov, E.; Garny, H.; Grewe, V.; Peter, T. Drivers of the tropospheric ozone budget throughout the 21st century under the medium-high climate scenario RCP 6.0. *Atmos. Chem. Phys.* **2015**, *15*, 5887–5902. [CrossRef]
10. Jacob, D.J. *Introduction to Atmospheric Chemistry*, 11th ed.; Princeton University Press: Princeton, NJ, USA, 1999; p. 280.
11. Monks, P.S.; Archibald, A.; Colette, A.; Cooper, O.; Coyle, M.; Derwent, R.; Fowler, D.; Granier, C.; Law, K.S.; Mills, G.; et al. Tropospheric ozone and its precursors from the urban to the global scale from air quality to short-lived climate forcer. *Atmos. Chem. Phys.* **2015**, *15*, 8889–8973. [CrossRef]
12. Kim, S.; VandenBoer, T.C.; Young, C.J.; Riedel, T.P.; Thornton, J.A.; Swarthout, B.; Sive, B.; Lerner, B.; Gilman, J.B.; Warneke, C.; et al. The primary and recycling sources of OH during the NACHTT-2011 campaign: HONO as an important OH primary source in the wintertime. *J. Geophys. Res. Atmos.* **2014**, *119*, 6886–6896. [CrossRef]
13. Fiore, A.M.; Naik, V.; Leibensperger, E.M. Air quality and climate connections. *J. Air Waste Manag. Assoc.* **2015**, *65*, 645–685. [CrossRef]
14. Oswald, E.M.; Dupigny-Giroux, L.A.; Leibensperger, E.M.; Poirot, R.; Merrell, J. Climate controls on air quality in the Northeastern US: An examination of summertime ozone statistics during 1993–2012. *Atmos. Environ.* **2015**, *112*, 278–288. [CrossRef]
15. Lefohn, A.S.; Malley, C.S.; Smith, L.; Wells, B.; Hazucha, M.; Simon, H.; Naik, V.; Mills, G.; Schultz, M.G.; Paoletti, E.; et al. Tropospheric ozone assessment report: Global ozone metrics for climate change, human health, and crop/ecosystem research. *Elementa* **2018**, *1*, 1. [CrossRef] [PubMed]
16. Fowler, D.; Amann, M.; Anderson, R.; Ashmore, M.; Cox, P.; Depledge, M.; Derwent, D.; Grennfelt, P.; Hewitt, N.; Hov, O.; et al. *Ground-Level Ozone in the 21st Century: Future Trends, Impacts and Policy Implications*; The Royal Society: London, UK, 2008; Volume 15/08.
17. Jacob, D.J.; Winner, D.A. Effect of climate change on air quality. *Atmos. Environ.* **2009**, *43*, 51–63. [CrossRef]
18. Itahashi, S.; Hayami, H.; Uno, I. Comprehensive study of emission source contributions for tropospheric ozone formation over East Asia. *J. Geophys. Res. Atmos.* **2015**, *120*, 331–358. [CrossRef]
19. Yi, X.; Zhang, J.; Wang, Z.; Li, T.; Zheng, Y. Deep distributed fusion network for air quality prediction. In Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, London, UK, 19–23 August 2018; pp. 965–973.

20. Wang, J.; Song, G. A deep spatial-temporal ensemble model for air quality prediction. *Neurocomputing* **2018**, *314*, 198–206. [CrossRef]

21. Li, X.; Peng, L.; Hu, Y.; Shao, J.; Chi, T. Deep learning architecture for air quality predictions. *Environ. Sci. Pollut. Res.* **2016**, *23*, 22408–22417. [CrossRef]

22. Wang, H.W.; Li, X.B.; Wang, D.; Zhao, J.; Peng, Z.R. Regional prediction of ground-level ozone using a hybrid sequence-to-sequence deep learning approach. *J. Clean. Prod.* **2020**, *253*, 119841. [CrossRef]

23. Gradišar, D.; Grašič, B.; Božnar, M.Z.; Mlakar, P.; Kocijan, J. Improving of local ozone forecasting by integrated models. *Environ. Sci. Pollut. Res.* **2016**, *23*, 18439–18450. [CrossRef]

24. Feng, R.; Zheng, H.J.; Zhang, A.R.; Huang, C.; Gao, H.; Ma, Y.C. Unveiling tropospheric ozone by the traditional atmospheric model and machine learning, and their comparison: A case study in Hangzhou, China. *Environ. Pollut.* **2019**, *252*, 366–378. [CrossRef]

25. Jia, P.; Cao, N.; Yang, S. Real-time hourly ozone prediction system for Yangtze River Delta area using attention based on a sequence to sequence model. *Atmos. Environ.* **2021**, *244*, 117917. [CrossRef]

26. Mekparyup, J.; Saithanu, K. Application of Artificial Neural Network Models to Predict the Ozone Concentration at the East of Thailand. *Int. J. Appl. Environ. Sci.* **2014**, *9*, 1291–1296.

27. Hoffman, S.; Jasiński, R. The Use of Multilayer Perceptrons to Model PM2.5 Concentrations at Air Monitoring Stations in Poland. *Atmosphere* **2023**, *14*, 96. [CrossRef]

28. Özbay, B.; Keskin, G.A.; Doğruparmak, Ş.Ç.; Ayberk, S. Predicting tropospheric ozone concentrations in different temporal scales by using multilayer perceptron models. *Ecol. Inform.* **2011**, *6*, 242–247.

29. ALves, L.; Nascimento, E.G.S.; Moreira, D.M. Hourly tropospheric ozone concentration forecasting using deep learning. *WIT Trans. Ecol. Environ.* **2019**, *236*, 129–138.

30. Chattopadhyay, G.; Midya, S.K.; Chattopadhyay, S. MLP based predictive model for surface ozone concentration over an urban area in the Gangetic West Bengal during pre-monsoon season. *J. Atmos. Sol.-Terr. Phys.* **2019**, *184*, 57–62. [CrossRef]

31. Hassan, M.A.; Dong, Z. Analysis of Tropospheric Ozone by Artificial Neural Network Approach in Beijing. *J. Geosci. Environ. Prot.* **2018**, *6*, 8–17. [CrossRef]

32. Hijjawi, M.A.M.S.M. Ground-level Ozone Prediction Using Machine Learning Techniques: A Case Study in Amman, Jordan. *Int. J. Autom. Comput.* **2020**, *17*, 667–677.

33. Deng, T.; Manders, A.; Jin, J.; Lin, H.X. Clustering-based spatial transfer learning for short-term ozone forecasting. *J. Hazard. Mater. Adv.* **2022**, *8*, 100168. [CrossRef]

34. Song, Z.; Deng, Q.; Ren, Z. Correlation and principal component regression analysis for studying air quality and meteorological elements in Wuhan, China. *Environ. Prog. Sustain. Energy* **2020**, *39*, 13278. [CrossRef]

35. 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]

36. Liu, B.C.; Binaykia, A.; Chang, P.C.; Tiwari, M.K.; Tsao, C.C. Urban air quality forecasting based on multi-dimensional collaborative Support Vector Regression (SVR): A case study of Beijing-Tianjin-Shijiazhuang. *PLoS ONE* **2017**, *12*, e0179763. [CrossRef] [PubMed]

37. Sanchez-Torres, G.; Bolaño, I.D. Support Vector Regression for PM10 Concentration Modeling in Santa Marta Urban Area. *Eng. Lett.* **2019**, *27*, 432–440.

38. Liu, H.; Li, Q.; Yu, D.; Gu, Y. Air quality index and air pollutant concentration prediction based on machine learning algorithms. *Appl. Sci.* **2019**, *9*, 4069. [CrossRef]

39. Murillo-Escobar, J.; Sepulveda-Suescun, J.; Correa, M.; Orrego-Metaute, D. Forecasting concentrations of air pollutants using support vector regression improved with particle swarm optimization: Case study in Aburrá Valley, Colombia. *Urban Clim.* **2019**, *29*, 100473. [CrossRef]

40. Castelli, M.; Clemente, F.M.; Popovič, A.; Silva, S.; Vanneschi, L. A Machine Learning Approach to Predict Air Quality in California. *Complexity* **2020**, *2020*, 8049504. [CrossRef]

41. Jumin, E.; Zaini, N.; Ahmed, A.N.; Abdullah, S.; Ismail, M.; Sherif, M.; Sefelnasr, A.; El-Shafie, A. Machine learning versus linear regression modelling approach for accurate ozone concentrations prediction. *Eng. Appl. Comput. Fluid Mech.* **2020**, *14*, 713–725.

42. Plocoste, T.; Laventure, S. Forecasting PM10 Concentrations in the Caribbean Area Using Machine Learning Models. *Atmosphere* **2023**, *14*, 134. [CrossRef]

43. Pak, U.; Kim, C.; Ryu, U.; Sok, K.; Pak, S. A hybrid model based on convolutional neural networks and long short-term memory for ozone concentration prediction. *Air Qual. Atmos. Health* **2018**, *11*, 883–895. [CrossRef]

44. Freeman, B.S.; Taylor, G.; Gharabaghi, B.; Thé, J. Forecasting air quality time series using deep learning. *J. Air Waste Manag. Assoc.* **2018**, *68*, 866–886. [CrossRef]

45. Eslami, E.; Choi, Y.; Lops, Y.; Sayeed, A. A real-time hourly ozone prediction system using deep convolutional neural network. *Neural Comput. Appl.* **2019**, *32*, 8783–8797. [CrossRef]

46. Sayeed, A.; Choi, Y.; Eslami, E.; Lops, Y.; Roy, A.; Jung, J. Using a deep convolutional neural network to predict 2017 ozone concentrations, 24 hours in advance. *Neural Netw.* **2020**, *121*, 396–408. [CrossRef]

47. Khamparia, A.; Singh, K.M. A systematic review on deep learning architectures and applications. *Expert Syst.* **2019**, *36*, e12400. [CrossRef]

48. Liu, D.R.; Lee, S.J.; Huang, Y.; Chiu, C.J. Air pollution forecasting based on attention-based LSTM neural network and ensemble learning. *Expert Syst.* **2020**, *37*, e12511. [CrossRef]

49. Chang, Y.S.; Chiao, H.T.; Abimannan, S.; Huang, Y.P.; Tsai, Y.T.; Lin, K.M. An LSTM-based aggregated model for air pollution forecasting. *Atmos. Pollut. Res.* **2020**, *11*, 1451–1463. [CrossRef]

50. Navares, R.; Aznarte, J.L. Predicting air quality with deep learning LSTM: Towards comprehensive models. *Ecol. Inform.* **2020**, *55*, 101019. [CrossRef]

51. Zhang, L.; Liu, P.; Zhao, L.; Wang, G.; Zhang, W.; Liu, J. Air quality predictions with a semi-supervised bidirectional LSTM neural network. *Atmos. Pollut. Res.* **2021**, *12*, 328–339. [CrossRef]

52. Nabavi, S.O.; Nölscher, A.C.; Samimi, C.; Thomas, C.; Haimberger, L.; Lüers, J.; Held, A. Site-scale modeling of surface ozone in Northern Bavaria using machine learning algorithms, regional dynamic models, and a hybrid model. *Environ. Pollut.* **2021**, *268*, 115736. [CrossRef]

53. Dai, H.; Huang, G.; Zeng, H.; Yu, R. Haze Risk Assessment Based on Improved PCA-MEE and ISPO-LightGBM Model. *Systems* **2022**, *10*, 263. [CrossRef]

54. Dai, H.; Huang, G.; Zeng, H.; Zhou, F. PM2.5 volatility prediction by XGBoost-MLP based on GARCH models. *J. Clean. Prod.* **2022**, *356*, 131898. [CrossRef]

55. Sayeed, A.; Eslami, E.; Lops, Y.; Choi, Y. CMAQ-CNN: A new-generation of post-processing techniques for chemical transport models using deep neural networks. *Atmos. Environ.* **2022**, *273*, 118961. [CrossRef]

56. Kim, H.S.; Han, K.M.; Yu, J.; Kim, J.; Kim, K.; Kim, H. Development of a CNN+ LSTM Hybrid Neural Network for Daily PM2.5 Prediction. *Atmosphere* **2022**, *13*, 2124. [CrossRef]

57. Zhao, Q.; Jiang, K.; Talifu, D.; Gao, B.; Wang, X.; Abulizi, A.; Zhang, X.; Liu, B. Simulation of the Ozone Concentration in Three Regions of Xinjiang, China, Using a Genetic Algorithm-Optimized BP Neural Network Model. *Atmosphere* **2023**, *14*, 160. [CrossRef]

58. Cirstea, R.G.; Micu, D.V.; Muresan, G.M.; Guo, C.; Yang, B. Correlated time series forecasting using multi-task deep neural networks. In Proceedings of the 27th ACM International Conference on Information and Knowledge Management, Torino, Italy, 22–26 October 2018; pp. 1527–1530.

59. National Research Council. *Rethinking the Ozone Problem in Urban and Regional Air Pollution*; National Academies Press: Washington, DC, USA, 1992.

60. Li, Y.; Yu, R.; Shahabi, C.; Liu, Y. Diffusion convolutional recurrent neural network: Data-driven traffic forecasting. *arXiv* **2017**, arXiv:1707.01926.

61. Ait-Amir, B.; Pougnet, P.; El Hami, A. *Embedded Mechatronic Systems 2*; Elsevier: Amsterdam, The Netherlands, 2015; pp. 151–179.

62. Chai, T.; Draxler, R.R. Root mean square error (RMSE) or mean absolute error (MAE)? *Geosci. Model Dev.* **2014**, *7*, 1525–1534. [CrossRef]

63. Srivastava, N.; Hinton, G.; Krizhevsky, A.; Sutskever, I.; Salakhutdinov, R. Dropout: A simple way to prevent neural networks from overfitting. *J. Mach. Learn. Res.* **2014**, *15*, 1929–1958.

64. Mehdipour, V.; Memarianfard, M. Application of support vector machine and gene expression programming on tropospheric ozone prognosticating for Tehran metropolitan. *Civ. Eng. J.* **2017**, *3*, 557–567. [CrossRef]