

Article

Mode Decomposition Bi-Directional Long Short-Term Memory (BiLSTM) Attention Mechanism and Transformer (AMT) Model for Ozone (O₃) Prediction in Johannesburg, South Africa

Israel Edem Agbehadji ^{1,*} and Ibidun Christiana Obagbuwa ^{2,*} 

¹ Centre for Global Change, Faculty of Natural and Applied Sciences, Sol Plaatje University, Kimberley 8301, South Africa

² Department of Computer Science and Information Technology, Faculty of Natural and Applied Sciences, Sol Plaatje University, Kimberley 8301, South Africa

* Correspondence: israel.agbehadji@spu.ac.za (I.E.A.); ibidun.obagbuwa@spu.ac.za (I.C.O.)

Abstract: This paper presents a model that combines mode decomposition approaches with a bi-directional long short-term memory (BiLSTM) attention mechanism and a transformer (AMT) to predict the concentration level of ozone (O₃) in Johannesburg, South Africa. Johannesburg is a densely populated city and the industrial and economic hub of South Africa. Being the industrial hub, air pollution is a major concern as it affects human health. Using air pollutants and meteorological datasets, a model was proposed that uses a mode decomposition approach to address the nonlinear nature of O₃ concentration. This nonlinearity is one of the most challenging issues in air quality prediction, and this study proposed a model to decompose input data and identify the most relevant features and leverage attention mechanisms to produce weighted parameters that can enhance the model's performance. The model's performance enhancement approach was aimed at ensuring an effective model that easily adapts to frequently changing pollutant data in air quality prediction. The performance was evaluated statistically with root mean squared error (RMSE), mean absolute error (MAE), and mean square error (MSE). The proposed EEMD-CEEMDAN-BiLSTM-AMT model produced the most optimal result with MSE (4.80×10^{-6}), RMSE (0.002), and MAE (0.001). When compared with the other similar models, the proposed model was best in terms of MSE value. Future work seeks to enhance the proposed model to fine-tune its performance on different air pollutant concentrations in South Africa.

Keywords: empirical mode decomposition (EMD); enhanced empirical mode decomposition (EEMD); complete ensemble empirical mode decomposition with adaptive noise (CEEMDAN); BiLSTM; attention mechanism and transformer; ozone prediction; South Africa



Academic Editor: Jun A. Zhang

Received: 25 January 2025

Revised: 27 March 2025

Accepted: 31 March 2025

Published: 2 April 2025

Citation: Agbehadji, I.E.; Obagbuwa, I.C. Mode Decomposition Bi-Directional Long Short-Term Memory (BiLSTM) Attention Mechanism and Transformer (AMT) Model for Ozone (O₃) Prediction in Johannesburg, South Africa. *Forecasting* **2025**, *7*, 15. <https://doi.org/10.3390/forecast7020015>

Copyright: © 2025 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 a major problem in most countries, and in recent times, countries with high levels of pollution are putting up mitigation measures to forestall any health impact. Several air pollutants can be identified to include particulate matter, nitrogen, oxygen, sulfur dioxide (SO₂), ozone (O₃), carbon dioxide (CO₂), and nitrogen dioxide (NO₂). One of the challenges with air quality prediction is that the concentration of these pollutants varies at different locations with time, which may be influenced by meteorological conditions [1]. The ground-level ozone (O₃) is very harmful to human health and climate, which can

also impact negatively on plants if not monitored [2]. It has been noted that ground-level O_3 has several detrimental health effects on major cities [3], including Johannesburg in South Africa.

South Africa is one of the countries that emit larger amounts of SO_2 and CO_2 , which is a concern that needs to be addressed. SO_2 is mostly emitted from coal, oil, and gas industries, while CO_2 is a result of burning fossil fuels from vehicles and power plants [4]. Mostly, power stations use low-grade coal, which emits CO, NO_2 , and SO_2 that negatively affect the environment [5]. Additionally, NO_2 is emitted from vehicles, factories, and plants. Thus, anthropogenic activities are causing serious environmental or ecological challenges that invariably have health implications [6]. These gaseous pollutants are a concern to most research scientists, leading them to find an effective approach to predict gaseous pollutant concentrations. The health implications of O_3 include coughing and chest pain that can even worsen the health condition of people with asthmatic conditions [7].

Globally, air pollution accounted for 8.1 million deaths in 2021, and 3365 children died due to air pollution in South Africa in the same year (www.unicef.org/southafrica accessed on 20 January 2025). Gaseous pollutants such as O_3 are formed in the presence of sunlight, which could affect humans either in the short or long term. Wu and He [8] attempted to predict hourly the short-time exposure of O_3 and NO_2 using a combination of residual neural network (ResNet), BiLSTM, and graph convolutional neural network (CNN) at three air monitoring stations in which their model performed well in predicting O_3 concentration at some monitoring stations. Thus, a suggestion that hybrid models can predict air pollutant concentration at different locations in a country.

Yafouz and Al Dahoul [9] indicated that the formation of O_3 is complex due to the influence of its precursors and meteorological conditions. Two types of such precursors are the VOC and nitrogen oxides (that is, NO_x , which is a combination of NO and NO_2) [10].

The Center for Global Change in Kimberley, South Africa, recognizes the importance of air pollution prediction models in the developmental agenda of the country. Though countries have policies that regulate the harmful effects of air pollutants, analyzing the recent air pollutant data could help strengthen policy interventions at air pollution monitoring stations [11]. Among the models to analyze air pollutant data are traditional statistical models, machine learning, and deep learning models. As technology for capturing air pollutants advances, air monitoring stations need to enhance their traditional or machine learning-based air pollutant predictive models as more data become available. Generally, deep learning models are well suited for large volumes of data because of the number of layers in the architecture. These layers can adapt to dynamically changing features in temporal air pollutants and meteorological datasets.

Donzelli and Suarez-Varela [12] suggested that ozone is one of the air pollutants that has a complex dynamic formation and distribution. In recent times, there has been a growing interest in adopting deep learning and machine learning models in designing air quality prediction systems. On one hand, machine learning models' computational power decreases when processing large volumes of air pollutant and meteorological datasets. On the other hand, deep learning models are more efficient in processing high volumes of datasets. This notwithstanding, machine learning models have been combined with deep learning models to improve the predictive performance of air quality prediction models. Unfortunately, air monitoring stations may find it challenging to change their machine learning-based legacy air quality prediction systems or models. Thus, necessitating the need to combine machine learning with deep learning models to capture the dynamic and volatile characteristics of air pollutants.

One of the concerns raised about O_3 is its volatility and ease of reaction with the weather and temperature. For instance, Xiao [13] suggested a machine learning model

that combined the long short-term memory (LSTM) and GARCH models to capture the nonlinearity and volatility of O₃ concentrations in time-series analysis. In this regard, the machine learning algorithms like support vector machine (SVM), artificial neural network (ANN), and random forests (RF) models were identified to have performed well in predicting O₃ concentration [13].

Again, aside from the influence of sunlight on O₃, the high concentration of O₃ experienced in certain tropical areas is influenced by anthropogenic activities [3]. Monitoring these activities is imperative; however, Yar and Henna [14] indicated that systems for monitoring O₃ concentration could be faulty, thus creating a gap in continuous air pollution monitoring. Cao and Bhatti [15] suggested a method that can retain important information about the air quality for a longer prediction such that the model's accuracy is maintained. This method fused the enhanced empirical mode decomposition (EEMD) with Complete Ensemble Empirical Mode Decomposition with Adaptive Noise (CEEMDAN) to decompose input data and further reduce the pollutant complexity; then, LSTM was applied to predict the air quality index. The EEMD algorithm provided an accurate reconstruction of the original input data. Therefore, EEMD and CEEMDAN are very useful tools in time series analysis and analysis of complex nonlinear data. While the EEMD avoids mixing of modes, the CEEMDAN uses the adaptive noise approach to remove averaging of signals and produce accurate and reliable intrinsic mode functions (IMFs). Du and Yuan [16] proposed CEEMDAN-VMD-GRU to predict the daily O₃ concentration. Initially, the CEEMDAN decomposed features; the VMD was applied as the second decomposition layer; and the Gate Recurrent Unit (GRU) was applied as the predictive layer. Pei and Huang [17] assessed Adaptive Variational Mode Decomposition (AVMD) combined with Multivariate Temporal Graph Neural Network (MtemGNN) to predict PM_{2.5} concentration. In their approach, AVMD was applied to select the optimal parameters for the predictive task. Sun and Huang [18] predicted NO₂ and SO₂ concentrations using a mode decomposing method where the pollutants were decomposed into sub-sequences with the CEEMDAN model. Wang and Zhang [19] proposed a model that used deep learning, a method (that is, BiLSTM) combined with the VMD and a graph attention network for PM_{2.5} concentration prediction. By using this model, the possible noise in the pollutant was addressed with the VDM model by decomposing the PM_{2.5} concentration into sub-sequences. The review highlights the robustness of mode decomposition models like VMD, CEEMDAN, EEMD, and other deep learning models, including BiLSTM, at revealing nonlinear and non-stationary characteristics of air pollutants. Thus, the novelty of this research lies in the hybridization of mode decomposition models, attention mechanisms and transformers, and deep learning BiLSTM models to predict O₃ concentration. Transformer and attention mechanism models support a two-phase training, where the model can be pre-trained on a dataset and fine-tuned for a predictive task. In this regard, our study aims to explore the current air pollutants and meteorological datasets to propose a predictive model for air quality prediction, focusing on Johannesburg, South Africa. Our approach benefits air monitoring stations that use historical O₃ concentration to effectively carry out predictions. Thus, this study contributes to the development of an enhanced predictive model that could benefit air monitoring stations in South Africa and worldwide. The sections of this paper are organized as follows: Section 2 (Materials and Methods), Section 3 (Results), Section 4 (Discussions), and Section 5 (Conclusion and Future Directions).

2. Materials and Methods

This section presents an overview of the study area, and the methodological stages for model development are dataset pre-processing, empirical mode decomposition, model

creation, model training, hyper-parameter tuning, and performance evaluation. The O₃ concentrations are measured in µg/m³.

2.1. Study Area

In South Africa, Johannesburg is the industrial and economic hub with an estimated population of 10 million. While cities considered to be the industrial hub are faced with many challenges, including air pollution, the level of O₃ concentration is increasing daily in the Delta Park area and also in the area of Newtown and Buccleuch [20]. Studies reveal that SO₂ and NO₂ are among the gaseous pollutants causing serious problems in Johannesburg, and these pollutants peak during winter, that is, July to August, with concentration levels of 38.6 µg/m³ and 12.6 µg/m³ [21]. Unfortunately, these industrial cities have recorded high concentrations of O₃, which peak during the Austral Spring [20]. In these regards, this study used Johannesburg as a case to develop a suitable model for O₃ concentration prediction. The weather and air pollution data were collected from 2023 to 2024. The weather data features are temperature, relative humidity, and wind speed, which cover the four seasons in Johannesburg: spring (September, October, and November), summer (December, January, and February), autumn (March, April, and May) and winter (June, July, August).

2.2. Model Development

Our study developed a hybrid model that consists of the mode decomposition layer, BiLSTM layer, attention mechanism, and transformer encoding layer and it outputs the result of O₃ prediction.

2.2.1. Pre-Processing Layer

Data were loaded on the model for data pre-processing. Data pre-processing helped to remove missing values from the dataset. Normally, missing values can be a result of device failure or other technical challenges that may affect data collection. The challenge of not removing these missing values is that it could reduce the model's predictive performance, and the result is poor prediction. The data pre-processing is significant in generating clean data and removing any other outliers that may affect the predictive performance of the model. Missing values were handled using interpolation, and having achieved this, the clean data are then normalized and reshaped to create the overlapping sequence input structure for the model's layers. Such normalization was achieved using the "MinMaxScaler" method. Furthermore, the likelihood of having a data attribute that recorded datetime was resolved through feature engineering, which means extracting further attributes from a DateTime feature as data and time.

2.2.2. Mode Decomposition Approach

In this study, the Ensemble Empirical Mode Decomposition (EEMD) and Complete Ensemble Empirical Mode Decomposition (CEEMDAN) were utilized to analyze complex nonlinear ozone concentration levels. It also considered the meteorological variables such as temperature, humidity, and wind speed in analyzing the O₃ concentration levels. After data pre-processing, the EEMD-CEEMDAN decomposed data into a sequence of modes as Intrinsic Mode Function (IMF) and its residual values. By using the IMF, our model handles the variability in O₃ concentration. The EEMD uses random noise addition, while CEEMDAN improves the noise by adding adaptive noise, which is more suitable for nonlinear time-series analysis. The BiLSTM, which is a time-series model, was applied to capture the temporal dependencies in the features at each time step to understand the sub-sequence in the data and use the attention mechanism to focus on the most relevant time step for each of the IMFs using simple additive attention. The output of the attention

mechanism was later fed into the Transformer encoder layer. The output was the predicted O_3 concentration.

To find the most relevant IMF, this study computed the correlation coefficient using the Pearson correlation coefficient (r), which measures the linear relationship between the two variables: IMF and O_3 levels. Equation (1) is expressed as follows:

$$r = \frac{cov(IMF, O_3)}{\sigma_{IMF}\sigma_{O_3}} \quad (1)$$

where $cov(IMF, O_3)$ is the covariance between IMF and original O_3 concentration, σ_{IMF} and σ_{O_3} are the standard deviations of IMF and O_3 , respectively, and the coefficient (r) ranges from -1 to 1 , showing the strength and direction of IMF and O_3 concentration level. We also analyzed the statistical properties of each IMF in terms of the mean, standard deviation, skewness, power spectral density (PSD), and kurtosis. The PSD was computed using continuous-time signal $x(t)$, where PSD $P_x(f)$ is expressed in Equation (2).

$$P_x(f) = \lim_{T \rightarrow \infty} \frac{1}{T} \left| \int_{-\frac{T}{2}}^{\frac{T}{2}} x(t) e^{-j2\pi ft} dt \right|^2 \quad (2)$$

where $P_x(f)$ refers to the power spectral density at frequency f , $x(t)$ is the time domain signal, j is the imaginary unit, and the integral computes the Fourier Transform of $x(t)$ in the limit as T approaches infinity for average power over an infinite set of periods. Analyzing these statistical properties helped to understand the dynamic nature of the underlying data to inform the modeling and forecasting task. The adaptive noise computation is expressed by (3).

$$n_i(t) = \alpha_i \cdot \eta(t) \quad (3)$$

where $n_i(t)$ presents the noise added to the data during each time (t), α_i is a scaling factor to control changing noise, and $\eta(t)$ is the random white noise series with a mean of zero and unit variance (between 0 and 1).

The most relevant IMF was inputted into the BiLSTM model for training, testing, and validation to show the model's performance curves. The validation loss monitors the model's performance during training and helps detect any overfitting. The test loss gives the final indication of how the model has performed after tuning all the hyper-parameters, thus, helping in model generalization. Mean ($\bar{x}$) is computed in Equation (4):

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i \quad (4)$$

Standard deviation s is computed in Equation (5):

$$s = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2} \quad (5)$$

Skewness U_1 is computed in Equation (6):

$$U_1 = \frac{n}{(n-1)(n-2)} \sum_{i=1}^n \left(\frac{x_i - \bar{x}}{s} \right)^3 \quad (6)$$

where n represents the number of observations, x_i is the individual data points, $\bar{x}$ is the sample mean, and s is the sample standard deviation.

2.2.3. BiLSTM Layer

The BiLSTM layer received the most relevant IMF as an input sequence into its input layer. The bidirectional model allows backward and forward processing of the input sequence. This layer ensures the processing of O₃ concentration using the past days or time steps. The LSTM cells are memory cells, and their gated mechanisms act as filters and mitigate vanishing gradients. The mathematical formula underpinning the BiLSTM layer can be expressed in Equation (7):

$$\begin{aligned} f_t &= \sigma(W_f \cdot [h_{t-1}, x_t] + b_f): \text{Forget gate} \\ i_t &= \sigma(W_i \cdot [h_{t-1}, x_t] + b_i): \text{Input gate} \\ o_t &= \sigma(W_o \cdot [h_{t-1}, x_t] + b_o): \text{Cell state update} \\ h_t &= o_t * \tanh(S_t): \text{Hidden state} \end{aligned} \quad (7)$$

where σ denotes the sigmoid function at time t , whereas S is activation vectors, which can be expressed in Equation (8):

$$\begin{aligned} \sigma(x) &= \frac{1}{1 + e^{-x}} \\ S_t &= f_t * S_{t-1} + i_t * \hat{S}_t \\ \hat{S}_t &= \tanh(W_S \cdot [h_{t-1}, x_t] + b_S) \end{aligned} \quad (8)$$

where $\hat{S}$ is the candidate value computed in each cell.

2.2.4. Attention Mechanism Transformer Layer

The attention mechanism enhances the model's performance by weighing the inputs dynamically to identify the best weight. Furthermore, the transformer uses multi-head attention to capture multiple correlations in the data while maintaining efficient parallelization. Thus, the attention mechanism provides a dynamic approach to generating weight for performance improvement. The attention weight focuses on the current context c_t at timestamp t and input sequence of hidden states h_t and corresponding weight a_t , which can be generalized by Equations (9) and (10):

$$c_t = \sum_{j=1}^T a_t * h_t \quad (9)$$

$$a_t = \text{softmax}\left(\frac{A_{weight}}{\sqrt{D}}\right) \quad (10)$$

where z_t represents the final context that shows the attention weight of the original query with all other inputs. A_{weight} is the unnormalized attention weight, D is the length of weight, and a_t is the normalized attention weight using the softmax function. T is the total number of hidden states. In this study, the context was O₃ concentration levels, which were the input sequences. The set of weights was generated by the attention mechanism. Additionally, the multi-head Transformer networks were utilized to train the entire sequence simultaneously and normalize the layer to avoid gradient vanishing. The multiple heads were set with 32 key dimensions and a dropout rate of 0.1.

2.2.5. Model Evaluation

Evaluating the model's performance provided a surety that the model adapted well to the observed and predicted values of O₃ concentration. Though there are several methods of model evaluation, statistical methods such as mean squared error and root mean square error were used due to their simplicity. Thus, when a lower error rate was recorded, it suggested that the prediction was closer to the observed value on the test dataset [22]. To

visualize the performance, the predicted and actual O₃ concentration level was plotted to assess the model's accuracy. The statistical computation was computed using the following methods: mean square error (MSE), root mean squared error (RMSE), and mean absolute error (MAE). These methods are expressed in Equations (11)–(13), respectively.

$$MSE = \frac{1}{N} \sum_{i=1}^N (y_i - y)^2 \quad (11)$$

$$RMSE = \sqrt{MSE} \quad (12)$$

$$MAE = \frac{1}{N} \sum_{i=1}^N |(y_i - y)| \quad (13)$$

where y_i represents the actual value, y represents the predicted value, and N is the number of samples.

The trained model was used to predict the O₃ concentration level, and the final results were displayed in graphs.

2.3. Flowchart of the Proposed Model

The flowchart of the proposed model is expressed in Figure 1. Furthermore, Appendix A presents the algorithm to implement the proposed model (EEMD-CEEMDAN-BiLSTM-AMT).

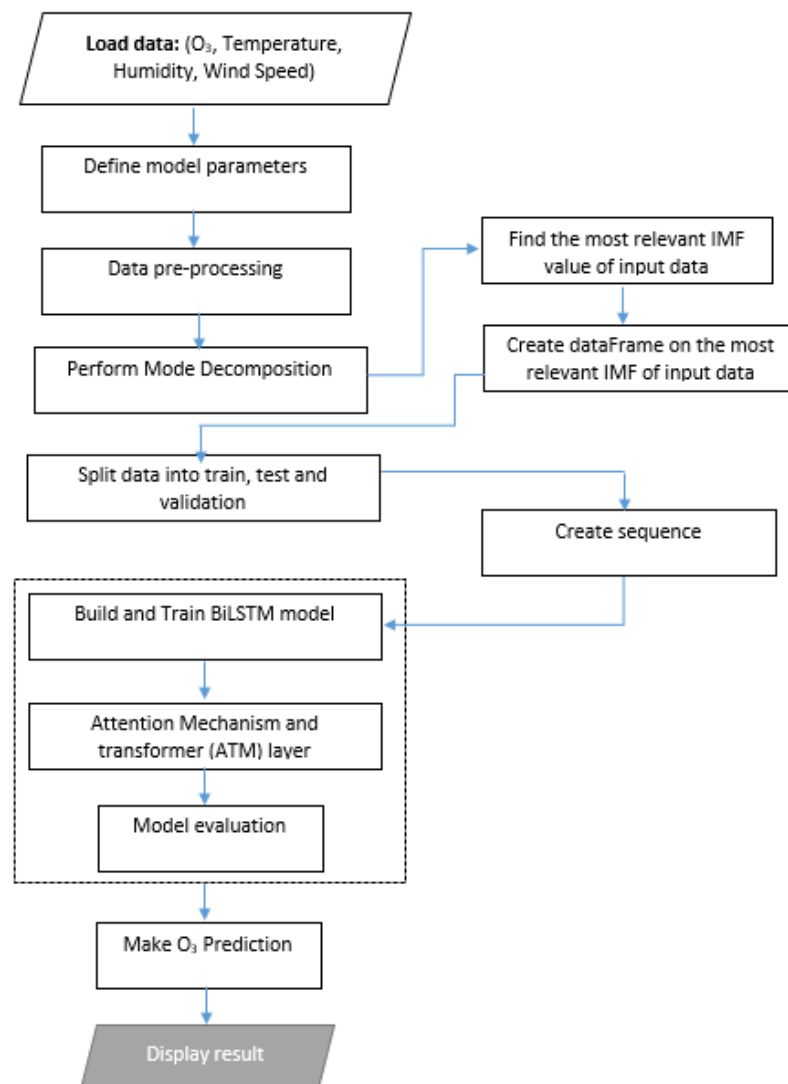


Figure 1. Flowchart of the proposed model.

2.4. Model Description and Parameters

During the experiment, the Google Colab platform with Python 3.10 was used, in which Deep Learning libraries were utilized along with other libraries for data manipulation, pre-processing, model development, and training. The experiment was carried out on a personal computer with a Core i3 16 gigabyte CPU. The hyper-parameter settings for the BiLSTM are presented in Table 1. In setting these hyper-parameters, the Bayesian Optimization approach was used for model parameter tuning.

Table 1. Hyper-parameter settings of the BiLSTM model.

Hyper-Parameter	Value
Batch size	32
Maximum epoch	100
Dropout	0.1
Epoch	100
Learning rate	0.01
Units	50

3. Results

During the data cleansing stage, 521 data points and 12 unique features (including city, country, and date) were recorded, and no missing values were recorded in the original dataset (Appendix B). The values recorded in the dataset cover the four seasons in Johannesburg, as mentioned earlier. The features on each row represent a unique measurement taken on a specific date at different time scales, with information on various pollutants and meteorological conditions. Figure 2 shows the results of the missing values in the dataset.

```

City          0
Country       0
Date          0
PM2.5         0
PM10          0
NO2           0
SO2           0
CO            0
O3            0
Temperature   0
Humidity      0
Wind Speed    0

```

Figure 2. Missing data values.

Table 2 shows the statistical analyses in terms of the minimum (min), maximum (max), mean, and variance of the air pollutants and the meteorological data. This analysis information shows the constituents of the data used for this research, in which this study focused on only O₃ concentration.

Table 2. Statistical analyses of pollutants and meteorological data.

Features	Min	Max	Mean	Variance
PM _{2.5}	5.02	149.78	77.6660	1663.56
PM ₁₀	10.00	199.99	105.1351	3162.87
NO ₂	5.83	99.98	53.1469	752.70
SO ₂	1.02	49.93	24.6742	207.35
CO	0.10	9.97	5.09339	7.55
O ₃	10.05	199.76	106.1828	3095.26
Temperature	−9.93	39.97	14.45266	206.79
Relative Humidity	10.05	99.86	55.44952	647.58
Wind Speed	0.50	20.00	10.08606	31.25

The unit of measurement for O_3 concentration is $\mu g/m^3$, while temperature ($^{\circ}C$), relative humidity (time of measurement (%)), and wind speed (time of measurement (m/s)) were also measured. Figure 3 shows their monthly average recordings.

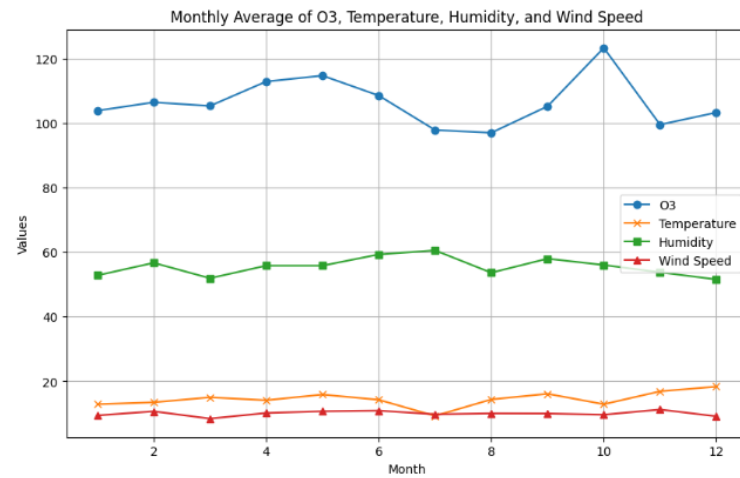


Figure 3. Monthly average of O_3 , temperature, humidity, and wind speed.

Figure 4 shows the empirical mode decomposition of O_3 . Also, the original O_3 data and IMFs are shown. These IMFs represent different groups of O_3 concentrations, of which the most relevant is suitable for further processing.

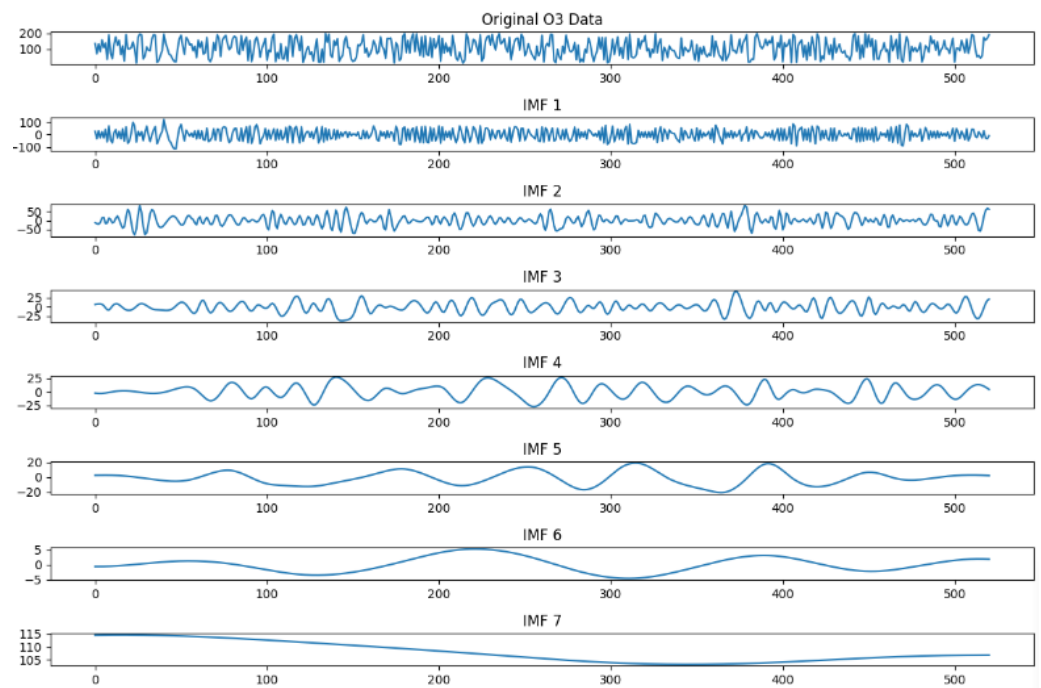


Figure 4. Mode decomposition.

Figure 5 shows all the most relevant IMF values for O_3 , temperature, humidity, and wind speed for different time scales with all the most relevant IMF values scaled between -100 and 100 amplitude. This amplitude scale subjects the IMF values of O_3 concentration to a range. In so doing, the most relevant IMF values of O_3 concentration fall within the defined range.

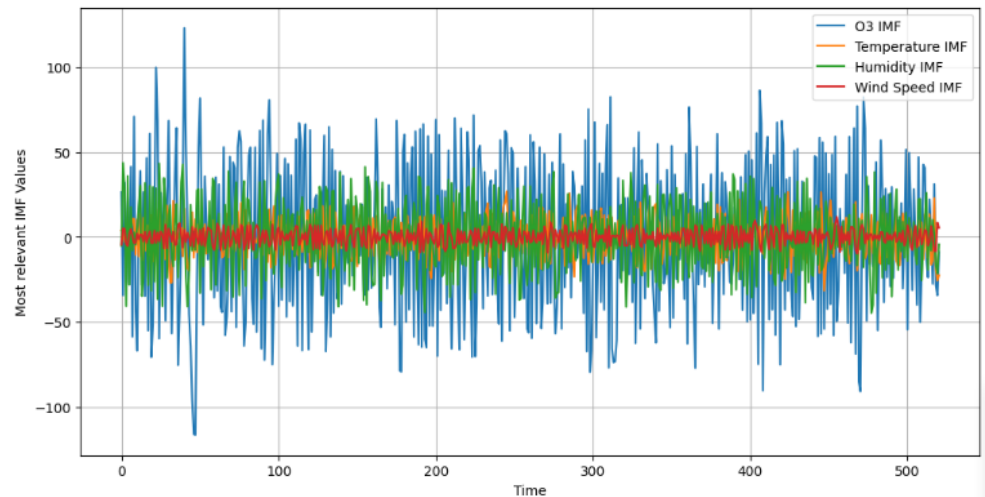


Figure 5. The most relevant IMFs values for O₃, temperature, humidity, and wind speed.

Table 3 shows the statistical analyses of the IMFs in terms of mean, standard deviation, skewness, and kurtosis.

Table 3. Statistical analyses of the IMFs values of O₃.

IMF	Mean	Standard Deviation	Skewness	Kurtosis
1	−0.42202	44.9020	−0.02123	−1.0677
2	−0.1517	27.5651	−0.00772	0.26913
3	−0.50037	13.6808	−0.09326	0.2250
4	0.35778	12.0271	−0.00269	−0.4866
5	−1.0607	9.0833	0.0859	−0.59361
6	0.02320	2.5492	0.14982	−0.7002
7	107.9368	3.85897	0.49060	−1.1875

This study computed the correlation of each O₃ IMF with the original O₃ concentration, and the result is presented in Table 4. It also shows IMF1 as the most relevant feature that correlated with the original O₃ concentration, with a value of 0.7797.

Table 4. Correlation of O₃ IMFs with original O₃ concentration.

IMF Index	IMF	Correlation of Each IMF with O ₃
0	IMF1	0.7797
1	IMF2	0.4603
2	IMF3	0.2915
3	IMF4	0.1715
4	IMF5	0.1471
5	IMF6	0.1061
6	IMF7	0.0575

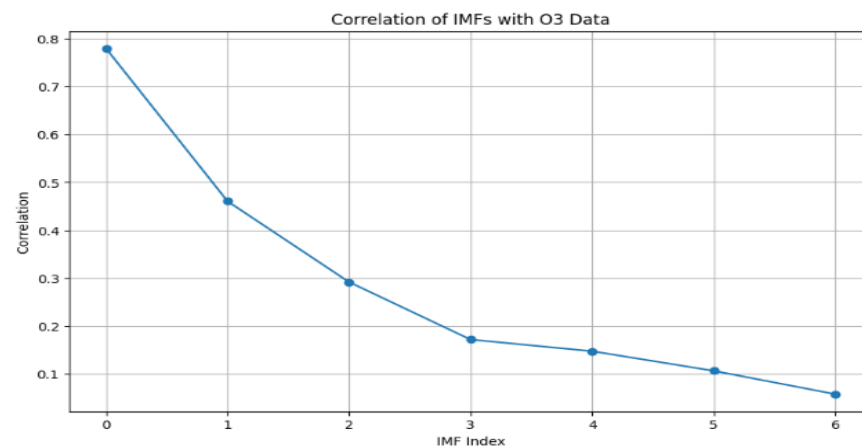
Note: Bold row represents the most relevant O₃ IMF value that correlates with the original O₃.

Furthermore, Table 5 shows the correlation summary of all the most relevant IMF features (temperature IMF, humidity IMF, and wind speed IMF). This correlation shows how each most relevant IMF value correlated with their respective feature in the original dataset.

Table 5. Correlation summary of all the most relevant IMFs with O₃.

IMF Features	Correlation
Temperature IMF	0.8206
Humidity IMF	0.8632
Wind speed IMF	0.8025

Figure 6 shows all the IMF correlations with the original O₃ concentration (Figure 3).

**Figure 6.** Plot of IMF correlations with the original O₃ concentration.

The most relevant IMF (IMF1) values for temperature, humidity, and wind speed were used to train the BiLSTM model and the comparative models. During the model development, 70% of the most relevant IMF was used for the training model, 15% for testing, and 15% for validation. The model's training and validation losses are shown in Figure 7 for 100 epochs.

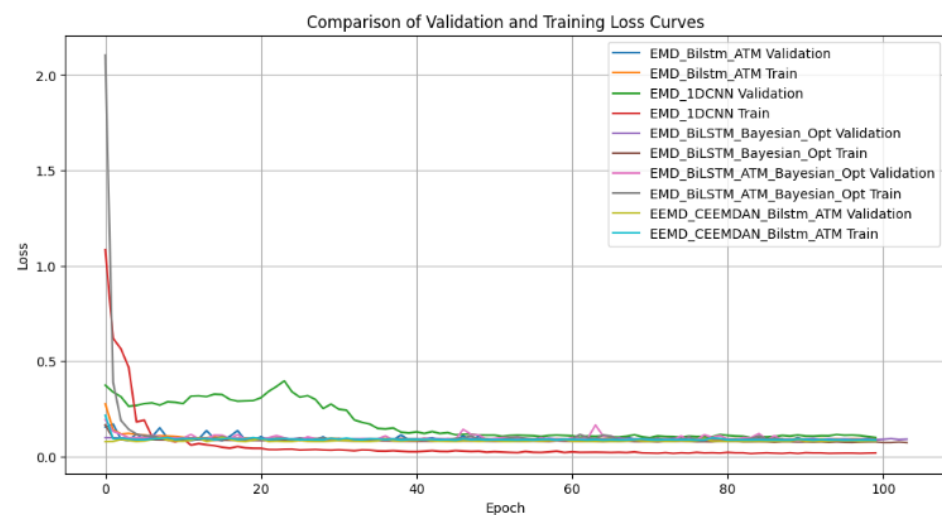
**Figure 7.** Comparison of validation and training loss curve.

Figure 7 shows the comparison of validation and training loss, where the curves converge as the number of epochs increases up until the 100th epoch. Table 6 shows the validation and test losses of the models. While a low test loss is ideal in model evaluation, it can be observed that EEMD-CEEMDAN-BiLSTM-AMT guaranteed a test loss of 6.35×10^{-4} , which is smaller compared to the other models.

Table 6. Validation and test losses of comparative models.

Models	Validation Loss	Test Loss
EMD-BiLSTM-AMT	0.088	0.065
EMD-1DCNN	0.099	0.081
EMD-BiLSTM-Bayesian Opt.	0.090	0.068
EMD-BiLSTM-AMT-Bayesian Opt.	0.103	0.083
EEMD-CEEMDAN-BiLSTM-AMT	6.32×10^{-6}	6.35×10^{-6}

Table 7 shows further performance evaluation results of the comparative models in terms of MSE, RMSE, and MAE.

Table 7. Performance evaluation metrics.

Models	MSE	RMSE	MAE
EMD-BiLSTM-AMT	0.065	0.254	0.211
EMD-1DCNN	13,277.24	115.22	102.915
EMD-BiLSTM-Bayesian Opt.	0.068	0.239	0.194
EMD-BiLSTM-AMT-Bayesian Opt.	3001.06	54.78	47.056
EEMD-CEEMDAN-BiLSTM-AMT	4.80×10^{-6}	0.002	0.0019

Note: Bayesian Opt. refers to Bayesian Optimization.

Figure 8 shows the comparison of O₃ prediction with actual O₃ concentration. In making the predictions, data on the past 20 days (20 time steps) was used to make daily future predictions of O₃ concentration.

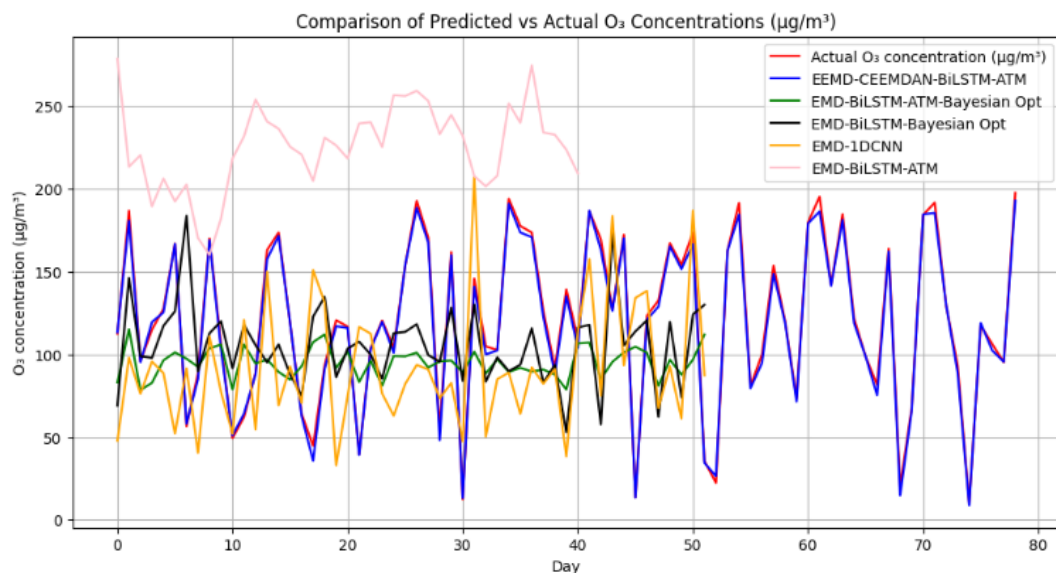
**Figure 8.** Comparison of predicted vs. actual O₃ concentration.

Figure 9 shows the prediction accuracy (using MSE) over lead time (in seconds using 10 lead times). It can be observed that, at the function of lead time, the prediction accuracy increases initially to the 4th lead time, then decreases steadily. The lead time analysis shows how the model makes short- and long-term predictions over time. In this context, the lead time refers to how far into the future the model is predicting and how the prediction accuracy changes as the lead time increases or decreases. In this context, as the lead time

increases to a point, the error in terms of MSE becomes smaller because the model smooths out or averages the error over long-term patterns.

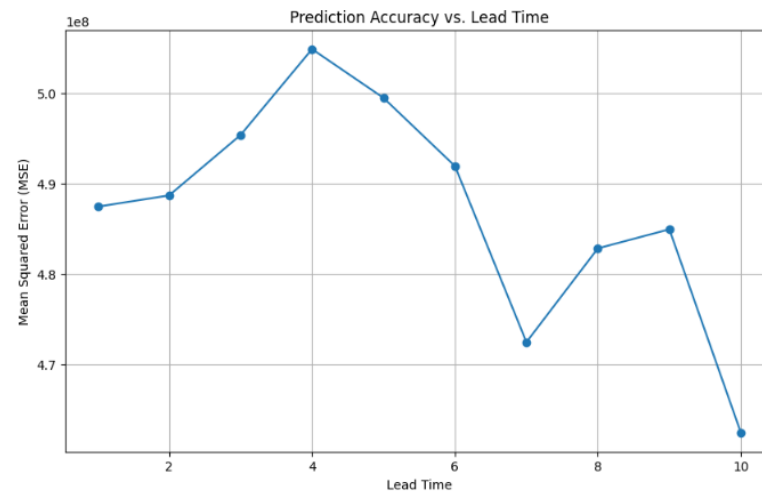


Figure 9. Prediction accuracy over lead time.

Figure 10 provides insight into how climate-relevant factors such as wind speed, temperature, and relative humidity correlate with O₃ concentration levels in the original dataset.

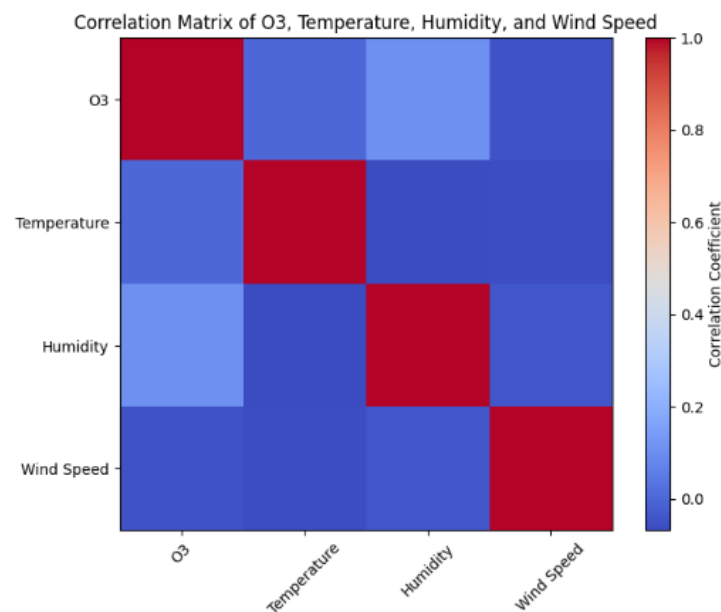


Figure 10. Correlation of O₃ with climate-related variables.

Having analyzed the correlation of O₃ with climate-related variables (Figure 10), it is noted that (see Table 8) the correlation coefficients for temperature, relative humidity, and wind speed are approximately -0.001 , 0.103 , and -0.050 , respectively. Thus, the magnitude of the correlation coefficient ($r = -0.001$) suggests an extremely weak negative correlation between temperature and O₃ concentration in the original dataset. Again, relative humidity with a correlation coefficient value of $r = 0.103$ indicates a weak positive correlation between humidity and O₃ concentration. This suggests that humidity has a weak positive influence on O₃ concentration. Finally, wind speed ($r = -0.05$) indicates an extremely weak negative correlation between wind speed and O₃ concentration.

Table 8. Pearson correlation matrix.

	Temperature	Humidity	Wind Speed	O ₃
Temperature	1.000000	−0.067947	−0.062734	−0.001001
Humidity	−0.067947	1.000000	−0.040013	0.103426
Wind Speed	−0.062734	−0.040013	1.000000	−0.050442
O ₃	−0.001001	0.103426	−0.050442	1.000000

Thus, concluding that these variables have low correlation (Figure 10 and Table 8), the IMFs are adopted as factors due to their relatively high correlation (Table 9). Finally, the IMFs were used as factors to predict O₃ concentration.

Table 9. Correlation of IMF values.

	O ₃ IMF	Temperature IMF	Humidity IMF	Wind Speed IMF
O ₃ IMF	1.0000	0.0226	0.0854	−0.0648
Temperature IMF	0.0224	1.0000	−0.0153	−0.0942
Humidity IMF	0.0854	−0.0153	1.0000	−0.0534
Wind Speed IMF	−0.0648	−0.0942	−0.0534	1.0000

4. Discussions

The experiment results provide an insight into models' performances. While this study highlights some drawbacks of deep learning models, model hybridization was employed in this research. During the experiment, the mode was decomposed into seven IMFs, and the most relevant IMF was used to train the BiLSTM model. Having set the hyper-parameters of the BiLSTM model, the attention mechanism and transformer layers were added to ensure adequate learning on the nonlinear nature of the O₃ concentrations. Initially, the most relevant IMF was split into training (70%), testing (15%), and validation (15%). Afterwards, the most relevant IMF was reshaped to fit the input structure of the BiLSTM model.

Correlation analysis helped in understanding of the relations among each IMF. This was achieved with the Pearson correlation coefficient, where the IMFs showed a value of 0.7797, where the most relevant IMF was IMF1 (Table 4). This suggests that the BiLSTM model should be trained with the most relevant IMF. Further statistical analyses were performed on each IMF to understand the distribution of each IMF in terms of mean, standard deviation, skewness, and kurtosis (Table 3). While our approach used the most relevant IMF, which represents the O₃ feature, Jiang and Wei [23] instead combined all the features or IMFs and reconstructed the features for training by the BiLSTM model. The reconstruction was aimed at reducing the noise in the sequence. The power spectral density is constructed on all the IMFs to obtain the exact frequencies of each peak value. Having achieved this, each IMF is correlated with original O₃ concentration, and the strong correlation is produced as IMF1 (0.7797), and the least correlated is IMF5 (0.0575).

This study employed mode decomposition models (EMD, EEMD, CEEMDAN) combined with BiLSTM, an attention mechanism, a transformer (that is, AMT), and Bayesian Optimization. The performance evaluation metrics are MSE, MAE, and RMSE. The comparative models were EMD-BiLSTM-AMT, EMD-1DCNN, EMD-BiLSTM-Bayesian Optimization, EMD-BiLSTM-AMT-Bayesian Optimization, and EEMD-CEEMDAN-BiLSTM-AMT. The experiment results show EMD-BiLSTM-AMT guaranteeing MSE, RMSE, and MAE values as 0.065, 0.254, and 0.211, respectively. The EMD-1DCNN model guaranteed MSE (13,277.24), RMSE (115.22), and MAE (102.915). Also, EMD-BiLSTM with Bayesian Optimization produced MSE (0.068), RMSE (0.239), and MAE (0.194). The EMD-BiLSTM-AMT

with Bayesian Optimization produced MSE (3001.06), RMSE (54.78), and MAE (47.056). The EEMD-CEEMDAN-BiLSTM-AMT guaranteed MSE (4.80×10^{-6}), RMSE (0.002), and MAE (0.0019). Based on the MSE values, the EEMD-CEEMDAN-BiLSTM-AMT model guaranteed a minimal value of 4.80×10^{-6} . This implies that the proposed EEMD-CEEMDAN-BiLSTM-AMT shows promising results for O₃ concentration prediction. Chen and Li [24] attested that data limitation could affect a model's performance, leading to overfitting in models used for air quality prediction tasks. For instance, overfitting was observed in the EMD-1DCNN model. Bayesian Optimization was used to tune the hyper-parameters (such as the number of LSTM units, learning rate, and epochs) for the BiLSTM model. Thus, the goal is to find the best hyper-parameter combination that minimizes validation loss to produce the best model. Based on validation and test losses (see Table 6), EMD-BiLSTM-AMT-Bayesian Optimization had a validation loss value of 0.103, and the EMD-BiLSTM-Bayesian Optimization model had a validation loss value of 0.090. Despite including AMT and Bayesian Optimization in the hybrid EMD-BiLSTM models in this study to fine-tune the model's hyper-parameters, optimal performance was a limitation. Again, data limitation is one of the main challenges in developing the models.

The evaluation of lead time assesses the model's accuracy for predicting future O₃ concentration levels. Figure 9 shows how prediction accuracy degrades as time increases. Thus, a longer lead time generally indicates a decrease in prediction accuracy of the models. This notwithstanding, the EEMD-CEEMDAN model guaranteed optimal performance due to its adaptive noise reduction approach, thus, making it suitable for the long-term prediction of O₃ concentrations.

These results demonstrate the error-canceling effect when models are hybridized [22,25] so that both can leverage their strengths and limitations for effective performance. In this research, it has been demonstrated that a mode decomposition approach combined with an attention mechanism and transformer, BiLSTM can predict O₃ concentration. Thus, the major contribution of this research is the integration of AMT with mode decomposition approaches such as EMD, EEMD, and CEEMDAN.

In this research, it can be identified that relative humidity influences O₃ concentration level in the study area; however, there was a weak correlation of relative humidity with O₃ concentration. Similarly, temperature and wind speed also exhibited a weak correlation with O₃ concentration level (see Table 8). In spite of the correlations, the EEMD-CEEMDAN-BiLSTM-AMT model guaranteed smaller MSE than the comparative mode decomposition hybrid models. As indicated earlier, O₃ depletion influences climate change [26], and using predictive models such as the EEMD-CEEMDAN-BiLSTM-AMT experimented with in this research could help pre-empt the pollutant distribution over the long term when given 20 days of data on pollutant concentrations and meteorological readings. In this regard, the EEMD-CEEMDAN-BiLSTM-AMT model predicted the O₃ concentration for the next 80 days.

5. Conclusions and Future Directions

In this study, several models for air pollutant concentration predictions were highlighted. Model hybridization is very topical in the current research dispensation, which could be attributed to the drawbacks of single models in dealing with the complex and nonlinear characteristics of air pollutants and their meteorological variables, which are very dynamic. Such dynamism requires the use of dynamic and nonlinear models that can pre-process air pollutants and provide insight into future predictions. Though our study proposed the EEMD-CEEMDAN-BiLSTM-AMT model in the context of South Africa, the model can be applied to countries seeking to leverage this model in their air quality monitoring stations.

Having demonstrated the practicality of the proposed model for O₃ concentration level prediction using historical datasets, future work could apply these models to different datasets from other provinces in South Africa. Again, transfer learning models are encouraged to explore the effectiveness of the proposed EEMD-CEEMDAN-BiLSTM-AMT in predicting air pollutant concentrations from resource-constrained air monitoring stations in South Africa. Transfer learning is emphasized because of data limitations in some locations in South Africa, which make predictions across the entire country very challenging. While acknowledging that the meteorological/weather data could be considered as variables in developing a mode decomposition-based air quality prediction model, this research provides the baseline model for future research. In view of these findings, the Center of Global Change is recommending the use of this promising EEMD-CEEMDAN-BiLSTM-AMT model for real-time air quality prediction tasks. Furthermore, the use of explainable artificial intelligence is recommended to help with the model’s interpretability and transparency for any air pollution predictive task.

Author Contributions: Conceptualization, I.E.A. and I.C.O.; methodology, I.E.A.; data curation, I.E.A. and I.C.O.; writing—original draft preparation, I.E.A. and I.C.O.; writing—review and editing, I.C.O.; supervision, I.C.O. All authors have read and agreed to the published version of the manuscript.

Funding: This research is funded by the Center for Global Change, Sol Plaatje University, Kimberley, South Africa with the National Research Foundation (NRF) (Number: 136097).

Data Availability Statement: The dataset used for this research is available at: <https://www.kaggle.com/datasets/waqi786/global-air-quality-dataset> (accessed on 20 September 2024).

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

Appendix A. Implementation Algorithm

Algorithms for implementation in the proposed flowchart are expressed by Algorithms A1 to A4.

Algorithm A1: Data Pre-Processing Layer
Input: Load original unprocessed data in CSV file format
Output: Clean data for Johannesburg in CSV file format
Process 1.1: Import libraries
Process 1.2: Load CSV file
Process 1.3: Extract country and city
Process 1.4: Rename the columns
Process 1.5: Re-engineer timestamp into day, hour, and year
Process 1.6: Remove empty entries
Process 1.7: Print (Output the clean data)
Algorithm A2: Mode Decomposition Layer
Input: Load the clean_data.csv
Output: Most relevant IMF
Process 2.1: Load data (e.g., O ₃ concentration)
Process 2.2: Perform EEMD
Process 2.3: Calculate the correlation between IMF and original data
Process 2.4: Find the IMF with the highest correlation
Process 2.5: Output the most relevant IMF and correlation with O ₃

Algorithm A3: Analysis of IMFs**Input:** Most relevant IMF**Output:** Statistical properties of IMF

Process 3.1: Analyze the frequency of IMF
 Process 3.2: Perform statistical analysis of each IMF
 Process 3.3: Calculate the PSD
 Process 3.4: Find the peaks in PDS
 Process 3.5: Print frequencies
 Process 3.6: Plot the frequencies and PDS with peak marks
 Process 3.7: Output the statistical value of the IMF

Algorithm A4: BiLSTM Layer for Training**Input:** Most relevant IMF**Output:** Model for prediction of O₃

Process 4.1: Normalize and reshape data
 Process 4.2: Split data into test, training, and validation
 Process 4.3: Create BiLSTM layer with input sequence as the most relevant IMF
 Process 4.4: Hyper-parameter settings on BiLSTM
 Process 4.5: Create an Attention mechanism layer
 Process 4.6: Create a transformer layer with dropout and layer normalization
 Process 4.7: Dense layer final output
 Process 4.8: Create the hybrid model for training
 Process 4.9: Evaluate the model
 Process 4.10: Output evaluation performance
 Process 4.11: Inverse transform the prediction to the original scale
 Process 4.12: Output the graph on training, test, and validation loss

Appendix B. Dataset

	City	Country	Date	PM2.5	PM10	NO2	SO2	\
47	Johannesburg	South Africa	2023-10-09	35.41	134.39	6.99	5.40	
54	Johannesburg	South Africa	2023-04-06	135.47	105.67	10.03	6.66	
132	Johannesburg	South Africa	2023-05-26	81.57	125.21	27.25	10.14	
168	Johannesburg	South Africa	2023-08-17	108.32	153.08	14.25	34.62	
180	Johannesburg	South Africa	2023-08-13	95.78	120.41	13.53	2.44	
...	...	...	...	...	...	...	...	
9916	Johannesburg	South Africa	2023-07-10	148.64	181.29	65.58	29.19	
9920	Johannesburg	South Africa	2023-05-04	95.24	135.30	21.84	7.25	
9922	Johannesburg	South Africa	2023-08-19	134.86	63.57	33.59	44.31	
9938	Johannesburg	South Africa	2023-02-14	80.16	159.88	58.62	36.49	
9995	Johannesburg	South Africa	2023-09-16	147.85	184.34	90.33	34.93	

	CO	O3	Temperature	Humidity	Wind Speed
47	0.77	134.92	28.65	49.11	7.49
54	5.58	70.22	27.81	95.91	17.88
132	5.23	133.79	19.49	84.25	18.29
168	6.10	83.46	18.41	20.11	9.27
180	4.76	158.63	8.73	99.86	7.65
...	...	...	...	...	...
9916	8.28	55.62	-9.90	99.70	11.52
9920	4.58	173.65	32.74	87.99	2.65
9922	5.82	157.63	-2.10	59.32	2.01
9938	4.33	169.56	-3.89	46.74	18.58
9995	2.81	191.45	-1.92	65.22	15.48

[521 rows x 12 columns]

References

- Samad, A.; Garuda, S.; Vogt, U.; Yang, B. Air pollution prediction using machine learning techniques—An approach to replace existing monitoring stations with virtual monitoring stations. *Atmos. Environ.* **2023**, *310*, 119987. [CrossRef]
- Pan, Q.; Harrou, F.; Sun, Y. A comparison of machine learning methods for ozone pollution prediction. *J. Big Data* **2023**, *10*, 63. [CrossRef]
- Abdullah, S.; Nasir, N.H.A.; Ismail, M.; Ahmed, A.N.; Jarkoni, M.N.K. Development of Ozone Prediction Model in Urban Area. *Int. J. Innov. Technol. Explor. Eng.* **2019**, *8*, 2263–2267.
- Igamba, J. Air Pollution in South Africa: The Silent Killer That Demands Urgent Action. 2023. Available online: <https://www.greenpeace.org/africa/en/blog/54600/air-pollution-in-south-africa-the-silent-killer-that-demands-urgent-action/> (accessed on 7 September 2024).
- Morakinyo, O.M.; Mukhola, M.S.; Mokgobu, M.I. Ambient Gaseous Pollutants in an Urban Area in South Africa: Levels and Potential Human Health Risk. *Atmosphere* **2020**, *11*, 751. [CrossRef]
- Sharma, S.; Joshi, J.; Kataria, S.; Verma, S.K.; Chatterjee, S.; Jain, M.; Brestic, M. Chapter 27—Regulation of the Calvin cycle under abiotic stresses: An overview. In *Plant Life Under Changing Environment*; Tripathi, D.K., Chauhan, D.K., Sharma, S., Prasad, S.P., Dubey, N.K., Ramawat, K., Eds.; Academic Press: Cambridge, MA, USA, 2020; pp. 681–717.
- Agency, U.S.E.P. Ozone and Your Patients' Health. 2024. Available online: <https://www.epa.gov/ozone-pollution-and-your-patients-health> (accessed on 20 September 2024).
- Wu, C.-L.; He, H.-D.; Song, R.-F.; Zhu, X.-H.; Peng, Z.-R.; Fu, Q.-Y.; Pan, J. A hybrid deep learning model for regional O₃ and NO₂ concentrations prediction based on spatiotemporal dependencies in air quality monitoring network. *Environ. Pollut.* **2023**, *320*, 121075. [CrossRef] [PubMed]
- Yafouz, A.; AlDahoul, N.; Birima, A.H.; Ahmed, A.N.; Sherif, M.; Sefelnasr, A.; Allawi, M.F.; Elshafie, A. Comprehensive comparison of various machine learning algorithms for short-term ozone concentration prediction. *Alex. Eng. J.* **2022**, *61*, 4607–4622. [CrossRef]
- Yang, J.; Zhao, Y. Performance and application of air quality models on ozone simulation in China—A review. *Atmospheric Environ.* **2023**, *293*, 119446. [CrossRef]
- Wang, S.; Sun, Y.; Gu, H.; Cao, X.; Shi, Y.; He, Y. A deep learning model integrating a wind direction-based dynamic graph network for ozone prediction. *Sci. Total. Environ.* **2024**, *946*, 174229. [CrossRef] [PubMed]
- Donzelli, G.; Suarez-Varela, M.M. Tropospheric Ozone: A Critical Review of the Literature on Emissions, Exposure, and Health Effects. *Atmosphere* **2024**, *15*, 779. [CrossRef]
- Xiao, H. A Hybrid Model Integrating LSTM with GARCH Family Models for the Ozone Concentration Prediction. In Proceedings of the 2023 3rd International Conference on Computer Science, Electronic Information Engineering and Intelligent Control Technology (CEI), Wuhan, China, 15–17 December 2023; pp. 1–6.
- Yar, A.; Henna, S.; McAfee, M.; Gharbia, S.S. Air Pollution Monitoring Using Online Recurrent Extreme Learning Machine. In Proceedings of the 2023 31st Irish Conference on Artificial Intelligence and Cognitive Science (AICS), Letterkenny, Ireland, 7–8 December 2023; pp. 1–6.
- Cao, J.; Bhatti, U.A.; Feng, S.; Huang, M.; Hasnain, A. Air Quality Index Predictions with a Hybrid Forecasting Model: Combining Series Decomposition and Deep Learning Techniques. In Proceedings of the 2023 IEEE 6th International Conference on Pattern Recognition and Artificial Intelligence (PRAI), Haikou, China, 18–20 August 2023.
- Du, X.; Yuan, Z.; Huang, D.; Ma, W.; Yang, J.; Mo, J. Importance of secondary decomposition in the accurate prediction of daily-scale ozone pollution by machine learning. *Sci. Total. Environ.* **2023**, *904*, 166963. [CrossRef] [PubMed]
- Pei, Y.; Huang, C.-J.; Shen, Y.; Ma, Y. An Ensemble Model with Adaptive Variational Mode Decomposition and Multivariate Temporal Graph Neural Network for PM_{2.5} Concentration Forecasting. *Sustainability* **2022**, *14*, 13191. [CrossRef]
- Sun, W.; Huang, C. A hybrid air pollutant concentration prediction model combining secondary decomposition and sequence reconstruction. *Environ. Pollut.* **2020**, *266*, 115216. [CrossRef] [PubMed]
- Wang, X.; Zhang, S.; Chen, Y.; He, L.; Ren, Y.; Zhang, Z.; Li, J.; Zhang, S. Air quality forecasting using a spatiotemporal hybrid deep learning model based on VMD–GAT–BiLSTM. *Sci. Rep.* **2024**, *14*, 17841. [CrossRef] [PubMed]
- Borduas-Dedekind, N.; Naidoo, M.; Zhu, B.; Geddes, J.; Garland, R.M. Tropospheric ozone (O₃) pollution in Johannesburg, South Africa: Exceedances, diurnal cycles, seasonality, Ox chemistry and O₃ production rates. *Clean Air J.* **2023**, *33*, 1–16. [CrossRef]
- Matandirotya, N.R.; Dangare, T.; Matandirotya, E.; Mahed, G. Characterisation of ambient air quality over two urban sites on the South African Highveld. *Sci. Afr.* **2023**, *19*, e01530. [CrossRef]
- Wattal, K.; Singh, S.K. Multivariate Air Pollution Levels Forecasting. In Proceedings of the 2021 2nd International Conference on Advances in Computing, Communication, Embedded and Secure Systems (ACCESS), Ernakulam, India, 2–4 September 2021.
- Jiang, X.; Wei, P.; Luo, Y.; Li, Y. Air pollutant concentration prediction based on a CEEMDAN-FE-BiLSTM model. *Atmosphere* **2021**, *12*, 1452. [CrossRef]

24. Chen, X.; Li, Y.; Xu, X.; Shao, M. A Novel Interpretable Deep Learning Model for Ozone Prediction. *Appl. Sci.* **2023**, *13*, 11799. [\[CrossRef\]](#)
25. Zang, Z.; Guo, Y.; Jiang, Y.; Zuo, C.; Li, D.; Shi, W.; Yan, X. Tree-based ensemble deep learning model for spatiotemporal surface ozone (O₃) prediction and interpretation. *Int. J. Appl. Earth Obs. Geoinf.* **2021**, *103*, 102516. [\[CrossRef\]](#)
26. Barnes, P.W.; Williamson, C.E.; Lucas, R.M.; Robinson, S.A.; Madronich, S.; Paul, N.D.; Bornman, J.F.; Bais, A.F.; Sulzberger, B.; Wilson, S.R.; et al. Ozone depletion, ultraviolet radiation, climate change and prospects for a sustainable future. *Nat. Sustain.* **2019**, *2*, 569–579. [\[CrossRef\]](#)

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.