Mühendislik Bilimleri ve Tasarım Dergisi
14(2), 416-430, 2026
e-ISSN: 1308-6693

Journal of Engineering Sciences and Design
DOI: 10.21923/jesd.1902759

Araştırma Makalesi

Research Article

LONG-TERM AIR QUALITY ($PM_{2.5}$) PREDICTION USING DEEP LEARNING-BASED TEMPORAL CONVOLUTIONAL NETWORKS

Vakkas DOĞAN[1*], Mustafa Yasin ERTEN[1], Hüseyin AYDILEK[1]

[1] Department of Electrical Electronics Engineering, Faculty of Engineering and Natural Sciences, Kırıkkale University, Kırıkkale, Türkiye.

Keywords

Air Quality Prediction, $PM_{2.5}$ Forecasting, Temporal Convolutional Networks (TCN), Time Series Modeling, Deep Learning.

Abstract

Among various sources of pollution, particulate matter with a diameter of 2.5 micrometers or less ($PM_{2.5}$) is the chief environmental problem worldwide. Collectively, the details of this pollutant $PM_{2.5}$, its sources and effects measured from various studies, the pollution levels in various cities, the methods used to measure and predict the concentration, the effect of climate change and the possible solution, have been discussed in this research article. Pollution levels are rising globally at alarming rates with detrimental effects on the environment as well as human health. However, air quality data represents a time series that is inherently challenging to forecast given its nonlinear nature, noisy data, and long-term dependencies in time. In this work, a Temporal Convolutional Network (TCN)-based deep learning model has been proposed for $PM_{2.5}$ forecasting, and its performance benchmarked against RNN, LSTM, GRU, and Attention mechanism-based Bi-LSTM models. The study was based on experiments using the hourly dataset of the air quality station in Beijing Aotizhongxin from 2013 to 2017.

DERİN ÖĞRENME TABANLI ZAMANSAL EVRİŞİMSEL AĞLAR KULLANARAK UZUN VADELİ HAVA KALİTESİ (PM2.5) TAHMİNİ

Anahtar Kelimeler

Hava Kalitesi Tahmini, $PM_{2.5}$ Tahmini, Zamansal Konvolüsyonel Ağlar, Zaman Serisi Modelleme, Derin Öğrenme.

Öz

Çeşitli kirlilik kaynakları arasında, çapı 2.5 mikrometre veya daha küçük olan partikül maddeler ($PM_{2.5}$), dünya çapındaki temel çevresel sorundur. Bu araştırma makalesinde; bu kirleticinin detayları, çeşitli çalışmalardan elde edilen kaynakları ve etkileri, farklı şehirlerdeki kirlilik seviyeleri, konsantrasyonu ölçmek ve tahmin etmek için kullanılan yöntemler, iklim değişikliğinin etkisi ve olası çözümler toplu bir şekilde ele alınmıştır. Kirlilik seviyeleri hem çevre hem de insan sağlığı üzerindeki zararlı etkileriyle birlikte küresel ölçekte endişe verici oranlarda artmaktadır. Bununla birlikte, hava kalitesi verileri; doğrusal olmayan yapısı, gürültülü veri içeriği ve zamana bağlı uzun vadeli bağımlılıkları nedeniyle tahmin edilmesi doğası gereği zor olan bir zaman serisini temsil eder. Bu çalışmada, ($PM_{2.5}$) tahmini için Zamansal Evrişimsel Ağ tabanlı bir derin öğrenme modeli önerilmiş ve modelin performansı; RNN, LSTM, GRU ve dikkat mekanizması tabanlı Bi-LSTM modelleriyle kıyaslanmıştır. Çalışma, Pekin Aotizhongxin hava kalitesi istasyonunun 2013-2017 yılları arasındaki saatlik veri seti kullanılarak yapılan deneylere dayanmaktadır.

Alıntı / Cite

Doğan, V., Erten, M.Y., Aydilek, H., (2026). Long-Term Air Quality ($PM_{2.5}$) Prediction Using Deep Learning-Based Temporal Convolutional Networks, Journal of Engineering Sciences and Design, 14(2), 416-430.

Yazar Kimliği / Author ID (ORCID Number)

V. Doğan, 0000-0001-5934-4156
M. Y. Erten, 0000-0002-5140-1213
H. Aydilek, 0000-0003-3051-4259

Makale Süreci / Article Process	
Başvuru Tarihi / Submission Date	04.03.2026
Revizyon Tarihi / Revision Date	20.04.2026
Kabul Tarihi / Accepted Date	30.04.2026
Yayım Tarihi / Published Date	30.06.2026

[*] Corresponding author: vakkasdogan@kku.edu.tr, (+90) 318 357 42 42/1017

LONG-TERM AIR QUALITY ($PM_{2.5}$) PREDICTION USING DEEP LEARNING-BASED TEMPORAL CONVOLUTIONAL NETWORKS

Vakkas Doğan[1†], Mustafa Yasin Erten[1], Hüseyin Aydilek[1]
[1] Department of Electrical Electronics Engineering, Faculty of Engineering and Natural Sciences, Kırıkkale University, Kırıkkale, Türkiye.

Highlights

- A enhanced TCN model is proposed for multivariate air quality forecasting.
- The model effectively captures long-term dependencies and sudden pollution spikes.
- TCN outperforms RNN, LSTM, and GRU models in robustness and predictive accuracy.
- Results show high potential for convolution-based models in early warning systems.

Graphical Abstract

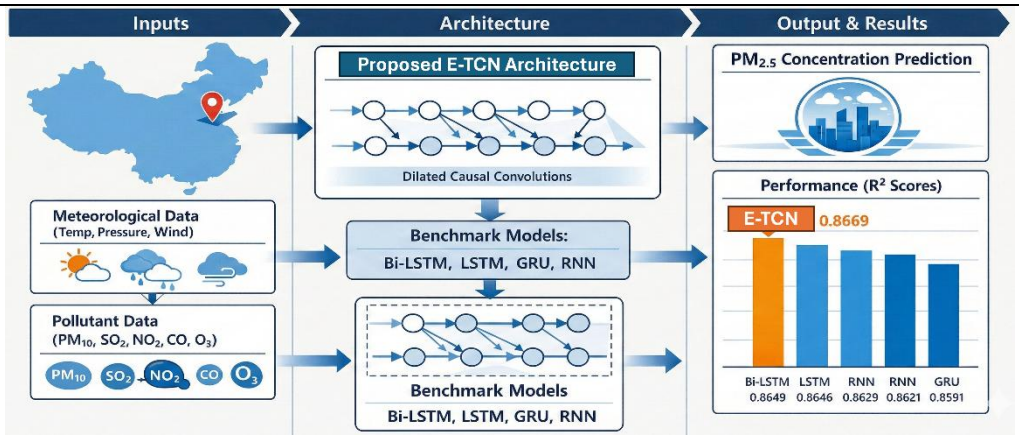


Figure. Graphical abstract illustrating the proposed E-TCN framework for long-term air quality prediction.

Purpose and Scope

The research intends to develop a strong Temporal Convolutional Network (TCN) framework that can accurately forecast PM2.5 in the long run, thus getting rid of the problems of recurrent architectures in understanding complex temporal dependencies.

Design/methodology/approach

Here we present an Enhanced Temporal Convolutional Network (E-TCN) architecture based on dilated causal convolutions, and it is thoroughly benchmarked against conventional recurrent neural networks (RNN, LSTM, GRU) using an extensive four-year hourly dataset from Beijing.

Findings

It was found that the proposed E-TCN model achieved a significantly high level of predictive accuracy (R^2 = 0.8669) and showed excellent robustness in detecting sudden pollution spikes in contrast to the traditional recurrent neural networks.

Research limitations/implications

Mainly the issues are that the model is quite sensitive to hyperparameter settings like the dilation pattern and the training of deep convolutional architectures requires a lot of computational memory.

Practical implications

The framework offers a computationally efficient and stable solution for real-time air quality monitoring systems, thereby greatly raising the early warning capabilities for dangerous pollution situations.

Originality

The paper uniquely confirms that dilated convolutional architectures are better than the traditional memory-gated mechanisms when it comes to dealing with sudden pollution spikes and requiring the modeling of very long-range temporal dependencies.

† Corresponding author: vakkasdogan@kku.edu.tr, (+90) 318 357 42 42/1017

1. Introduction

Fine particulate matter ($PM_{2.5}$) has been identified as one of the most dangerous pollutants among the air pollutants. This is because $PM_{2.5}$ can penetrate the human lung tissue deeply and enter the blood circulation system; thus, it can cause serious heart and respiratory diseases and play a major role in the increase of premature deaths worldwide (Brook et al., 2010; Pope & Dockery, 2006; Lelieveld et al., 2015). Based on a report by the World Health Organization (2021), the exposure to PM2.5 for a long period is ranked as the main environmental factor causing disease worldwide. Actually, it is essential to the air quality management, early warning systems and environmentally friendly policymaking that are based on solid evidence that $PM_{2.5}$ concentration forecasting is done accurately (Di et al., 2017; Shen et al., 2019).

Historically, chemical transport models based on deterministic principles and classical statistical methods such as linear regression, autoregressive integrated moving average, and generalized additive models have been the major tools for predicting PM2.5 concentrations (Zhang et al., 2018). These methods are certainly capable of revealing the physical mechanisms behind these processes. They, however, have the intrinsic shortcoming that they can hardly capture the nonlinear, nonstationary, highly dynamic nature of the PM2.5 time series, especially under the condition of fast-changing meteorology and severe pollution episodes (Cobourn, 2010; Russo et al., 2019). Therefore, the emphasis has been shifted from the conventional physically-based models to data-driven modeling techniques in order to overcome such limitations.

It has been demonstrated that machine learning-based methods, which include support vector regression, random forests and gradient boosting, are more efficient in prediction than traditional statistical models (Chen et al., 2018; Grange et al., 2018). However, these shallow learning methods are only capable of capturing the complexity of the spatiotemporal relationships in the air quality data to a certain degree. Therefore, deep learning methods have been suggested as viable solutions and their main benefits are the abilities to perform hierarchical feature learning and nonlinear representation (Li et al., 2016; Qi et al., 2018).

Recurrent neural networks, especially the LSTM and GRU variations, are biologically inspired neural network models that are widely used in deep learning, primarily due to their ability to model sequential data. They have been extensively used in studies forecasting $PM_{2.5}$ levels because of their strong temporal modeling capability (Hochreiter & Schmidhuber, 1997; Cho et al., 2014). It has been demonstrated in several studies that LSTM- and GRU-based architectures are more efficient than traditional machine learning and shallow learning methods for short- and medium-term $PM_{2.5}$ forecasting (Zheng et al., 2015; Yao et al., 2019; Eren et al., 2023). However, a drawback of recurrent models is that they perform one step at a time, which means that they require more computational resources and that the gradient may disappear in very long sequences, making them less scalable for very detailed and long duration datasets (Bai et al., 2018; Lim et al., 2021).

To overcome the above-mentioned drawbacks, some researchers have proposed CNN-recurrent layer composite networks. CNN, LSTM and CNN, GRU architectures utilize convolutional layers for spatial or local temporal feature extraction while recurrent units are reserved for sequence learning, thus achieving higher prediction accuracy (Li et al., 2017; Hu et al., 2024; Liu et al., 2024). Furthermore, recent literature highlights that combining deep learning architectures with robust data visualization techniques significantly enhances the interpretability and reliability of air quality forecasting models (Mengüş & Daş, 2025). The adaptability of these advanced hybrid deep learning architectures has also been proven highly effective in complex predictive tasks across different engineering domains, such as the prognostic degradation prediction of heavy machinery (Kara, 2021). Furthermore, the use of ensemble learning methods and signal decomposition methods such as empirical mode decomposition, ensemble empirical mode decomposition, and variational mode decomposition have been combined in order to denoise and increase the robustness of the forecasts (Duan et al., 2023; Gao et al., 2024). Nonetheless, their increased complexity, the sensitivity to the choice of hyperparameters as well as the degradation of the generalization ability to different regions are some of the problems that these hybrid models face.

Recently, attention mechanisms and transformer neural networks have been drawing attention for their use in air quality forecasting. An attention mechanism works by weighting the elements of a sequence differently in order to capture long-term dependency and make the model more interpretable (Vaswani et al., 2017; Lim et al., 2021). The transformer model is becoming the state-of-the-art in the field of multivariate time series forecasting, and $PM_{2.5}$ prediction is no exception (Zhou et al., 2024). On the other hand, due to the fact that in general these models are quite complex and depend on a considerable amount of data to be trained, their use in situations where the data is scarce such as in real-world air quality monitoring systems in developing countries is rather limited (Pan et al., 2024).

Spatiotemporal modeling with graph neural networks and spatiotemporal graph convolutional networks is one of the fast-growing research areas. These methods explicitly capture spatial correlations among monitoring stations (Wu et al., 2019; Panja et al., 2024). Graph-based methods have been proven effective at the regional scales, however, their success largely relies on having a dense sensor network as well as a correct spatial connectivity definition, which are not always at hand (Agbehadji & Obagbuwa, 2024).

It is true that the level of air pollution forecasting techniques has been improved greatly recently, however, the field still has core issues left that ready solutions. For one, majority of the researches limit their works to recurrent or attention-based architectures, moreover, they have shown less interest to other temporal modeling paradigms that can capture long-range dependencies with even lesser computational requirements (Bai et al., 2018; Borovykh et al., 2017). Besides, while quite a number of studies concentrate on the average accuracy of the forecast models, they do not hold the models to a very strict testing when it comes to their performances during the extreme pollutant concentration time windows that are the most important for the public health risk mitigation (Russo et al., 2019; Liu et al., 2025). Moreover, the differences in the datasets, the feature selection strategies, and the evaluation protocols put a big barrier on the fair competition and reproducibility of the results among studies (Zhang et al., 2018; Zhou et al., 2024). These inherent weaknesses make it clear that the field needs fast, robust, and basically capable of transfer deep learning-based framework for the estimation of $PM_{2.5}$ concentration.

In order to bridge the gaps mentioned above and promote the development of the research in the forecasting of PM2.5, the following contributions are made in this paper:

- An Enhanced Temporal Convolutional Network (E-TCN) framework is proposed for the prediction of PM2.5 concentrations. Unlike standard TCN architectures, the proposed E-TCN integrates a specific exponentially increasing dilation sequence optimized for extreme PM2.5 volatility, coupled with spatial dropout and residual mappings. This tailored design maximizes the receptive field while effectively preventing overfitting during sudden pollution spikes, establishing a robust early-warning capability.

- By performing a comparative analysis of the proposed model with various deep learning architectures, including LSTM, GRU, Bi-LSTM, CNN, LSTM, and attention-based models, a comprehensive and integrated benchmarking framework is established. This comparison of the model with the other architectures is performed using the same datasets, as per the benchmarks set by Atmospheric Environment and Science of the Total Environment (STOTEN).

- Robustness under extreme pollution conditions is explicitly analyzed, providing novel insights into the capability of temporal convolution-based models to handle abrupt $PM_{2.5}$ concentration spikes.

- Real-time forecasting shows its practical application through the improvement in the computational efficiency and the stability of the forecasts, thus emphasizing the potential of the suggested air pollution framework for application in air quality monitoring systems.

2. Material and Method

2.1. Material

2.1.1. Dataset Description

This research utilizes the "Beijing Multi-Site Air-Quality Data" dataset, which is available on the UCI Machine Learning Repository, to examine the problem of predicting PM2.5 concentrations (Dua & Graff, 2019; Zhang et al., 2017). This dataset is considered a standard dataset for research on air quality forecasting problems, considering its extensive time span, fine time granularity, and comprehensive coverage of all variables, including pollutant concentrations and meteorological parameters. The research is conducted on data related to the Aotizhongxin air quality monitoring站, which is considered representative of the populated core area of Beijing (Zheng et al., 2015; Yao et al., 2019). The dataset provides hourly values for a period of four years, starting on 1st March 2013 and ending on 28th February 2017, thus providing an opportunity to examine short-term and long-term temporal patterns related to air pollution behavior.

The raw dataset consists of 35,064 samples and 18 variables prior to preprocessing. The included variables can be grouped into two main categories:

- Pollutant Concentrations: $PM_{2.5}$ (target variable, $\mu g/m^3$), PM_{10}, SO_2, NO_2, CO, and O_3
- Meteorological Variables: Temperature (TEMP), atmospheric pressure (PRES), dew point temperature (DEWP), precipitation (RAIN), wind speed (WSPM), and wind direction (wd)

Among these variables, wind direction is a categorical feature. However, deep learning models are not compatible with categorical features, so this feature was first converted to a numeric feature by the use of Label Encoding. The rest of the features are numerically continuous and are represented as float or integer. While cyclical encoding (such as sine-cosine transformations) is recognized as the optimal method for preserving the periodic nature of variables like wind direction, Label Encoding was utilized here as a preliminary preprocessing baseline. Given the dominant influence of continuous temporal lags (e.g., historical PM concentrations and wind speed), this simplification did not critically impair the predictive accuracy, although implementing cyclical encoding remains a priority for future methodological refinements.

The features of the dataset along with the multivariate time series of pollutant concentrations and meteorological factors are shown in Table 1. The figure shows the strong temporal variation and seasonal patterns of $PM_{2.5}$ and associated variables which clearly indicate the need of advanced temporal modeling techniques.

Table 1. Sample data structure of the Beijing Multi-Site Air-Quality Dataset (Aotizhongxin Station)

No	year	month	day	hour	$PM_{2.5}$	PM_{10}	SO_2	NO_2	CO	O_3	TEMP	PRES	DEWP	RAIN	wd	WSPM	station
1	2013	3	1	0	4.0	4.0	4.0	7.0	300	77.0	-0.7	1023.0	-18.8	0.0	NNW	4.4	Aotizhongxin
2	2013	3	1	1	8.0	8.0	4.0	7.0	300	77.0	-1.1	1023.2	-18.2	0.0	N	4.7	Aotizhongxin
3	2013	3	1	2	7.0	7.0	5.0	10.0	300	73.0	-1.1	1023.5	-18.2	0.0	NNW	5.6	Aotizhongxin
4	2013	3	1	3	6.0	6.0	11.0	11.0	300	72.0	-1.4	1024.5	-19.4	0.0	NW	3.1	Aotizhongxin
5	2013	3	1	4	3.0	3.0	12.0	12.0	300	72.0	-2.0	1025.2	-19.5	0.0	N	2.0	Aotizhongxin

2.1.2. Description of the Computational Platform

The deep learning models presented here were trained, validated, and tested on Google Colaboratory, a cloud-based platform that provides scalable computational resources which are perfect for data-intensive and high-dimensional deep learning applications. The choice of this cloud-based platform was mainly motivated by the high computational demand of training temporal convolutional and recurrent neural network architectures on large-scale, high-resolution time series data. The hardware setup used in this study was composed of an NVIDIA Tesla T4 graphics processing unit (GPU) with 12 GB of dedicated memory, an Intel Xeon central processing unit (CPU), and system memory resources that were assigned by the platform. This hardware arrangement facilitated effective parallel computation, rapid model training, and an overall reduction in training time, especially for convolution-based architectures that can be substantially accelerated by a GPU.

All the experiments were conducted using the Python 3.10 programming language. This selection of the programming language over other suitable alternatives is based on the popularity of the Python programming language, specifically with regard to the scientific computing domain and the machine learning domain. In addition, the Python programming language is very suitable with regard to the reproducibility of the experimental results. The Pandas library, along with the NumPy library, were utilized with regard to the data preprocessing tasks. In addition, the Pandas library, along with the NumPy library, were utilized with regard to the data manipulation tasks. The model, specifically, was implemented using the TensorFlow 2.x library, along with the Keras API, to provide a flexible, modular, reproducible, scalable, and efficient deep learning architecture. In addition, the Matplotlib library, along with the Seaborn library, were utilized with regard to the data visualization tasks, specifically to provide well-structured visualizations of the data with regard to the temporal trends, the statistical distributions, and the model performance characteristics. In addition, the Scikit-Learn library was utilized with regard to the feature scaling, the encoding of the categorical variables, the partitioning of the datasets, the auxiliary preprocessing, and the standardization of the evaluation protocols. The implementation of the cloud-based computing platform, along with the standard software stack, is one of the crucial steps that can be taken to enhance the reproducibility of the experimental designs. The same hardware and software configurations were utilized with regard to the training of all the models, specifically to eliminate the performance variations with regard to the differences between the hardware configurations.

2.2. Method

This research work was focused on air quality forecasting with the problem of $PM_{2.5}$ concentrations prediction in the first place, by using cutting-edge deep learning based time series models. Since air pollution data are very much temporally dependent, nonlinear, and seasonal, deep neural network architectures intended for sequential data were selected. The key point of the methodology was TCNs or Temporal Convolutional Networks, which have been lately demonstrated to perform better than other methods in long-horizon time series forecasting tasks. For benchmarking and comparative evaluation, a variety of recurrent-based architectures such as RNNs or Recurrent Neural Networks, LSTMs or Long Short-Term Memory networks, GRUs or Gated Recurrent Units, and Bi-LSTM or

Bidirectional LSTM models with attention mechanisms were also executed (Bai et al., 2018; Hochreiter & Schmidhuber, 1997; Cho et al., 2014; Vaswani et al., 2017).

2.2.1. Data Preprocessing

To lessen sensor failure noise as well as noise coming from environmental changes and to make a model more capable of the effect of generalization, a very well thought-out processing chain has been set up before the model training.

The first stage of abnormal point detection and elimination was done with the help of the Interquartile Range (IQR) method which is a frequently used method in the field of environmental data because it is a robust method that works well on non-Gaussian distributions (Aggarwal, 2017; Liu et al., 2020). For an observation x, the lower and upper bounds were defined as:

$$Lower\ Bound = Q_1 - 1.5\ x\ IQR \tag{1}$$

$$Upper\ Bound = Q_3 - 1.5\ x\ IQR \tag{2}$$

where Q_1 and Q_3 denote the first (25th percentile) and third (75th percentile) quartiles, respectively, and IQR=Q_3−Q_1. Observations falling outside these bounds were treated as outliers and excluded from the dataset.
Following outlier removal, Min–Max normalization was applied to all continuous variables to mitigate scale discrepancies among features and to stabilize gradient-based optimization during training. Each variable was scaled to the [0,1] interval using:

$$x^{\iota} = \frac{x - x_{min}}{x_{max} - x_{min}} \tag{3}$$

where x_{min} and x_{max} represent the minimum and maximum values of the corresponding feature. This normalization strategy is particularly effective for deep neural networks employing activation functions sensitive to input scale (LeCun et al., 2012; Goodfellow et al., 2016).

2.2.2. Temporal Convolutional Network (TCN) Architecture

One of the main models used in the current study is the Temporal Convolutional Network (TCN), which is a type of convolutional neural network. It is specifically designed to address sequential data modeling. In contrast to recurrent neural networks, the TCN uses convolutional layers, which allow for parallel computation and the stable backpropagation of gradients even with very long sequences (Bai et al., 2018; Lea et al., 2017).

The two basic ingredients in TCNs are causal convolution and dilated convolution. Causal convolution ensures that the prediction at time t depends only on inputs at time t and previous ones *(t, t-1, …)*, thus, it does not leak information from future observations. The output of y_t a 1D causal convolution can be formally given as:

$$y_t = \sum_{k=0}^{K-1} w_k . x_{t-k} \tag{4}$$

where denotes the kernel size and w_k represents the convolutional filter weights.
To efficiently capture long-term temporal dependencies without excessive network depth, dilated convolutions were employed. By introducing gaps between kernel elements, dilated convolutions expand the receptive field exponentially with depth. The dilation factor *d* increases across layers following a power-of-two scheme (2^0, 2^1, 2^2, …) allowing the network to model long-range dependencies with a limited number of parameters (Yu & Koltun, 2016). The effective receptive field *R* of a TCN layer stack can be approximated as:

$$R = 1 + (K - 1) \sum_{l=0}^{L-1} d_l \tag{5}$$

where *L* denotes the number of convolutional layers.

In this study**,** a E-TCN architecture was adopted, characterized by a kernel size of 5, 64 convolutional filters per layer, and dilation rates of [1, 2, 4, 8, 16, 32]. To mitigate overfitting and improve generalization performance, Spatial Dropout with a dropout rate of 0.2 was applied, which randomly drops entire feature maps instead of individual neurons, thus preserving temporal coherence (Tompson et al., 2015).

2.2.3. Bi-LSTM with Multi-Head Attention

For comparison, a Bi-LSTM architecture was also implemented. Whereas unidirectional recurrent models only take into account the context to the left of a word in a sequence, Bi-LSTM runs the input sequence both forward and backward and therefore can use the information on both sides at the same time (Schuster & Paliwal, 1997). To further enhance temporal feature discriminati**on,** a Multi-Head Attention mechanism was integrated into the Bi-LSTM model. Attention mechanisms allow the model to assign different importance weights to different time steps by computing a weighted sum of hidden representations. Given query Q, key K, and value V matrices, scaled dot-product attention is computed as:

$$Attention\ (Q, K, V) = softmax\left(\frac{QK^T}{\sqrt{d_k}}\right)V \quad (6)$$

- QK^T: $similarity\ score\ between\ all\ patch\ pairs$
- $\sqrt{d_k}$: scaling factor is used to prevent gradient explosion

Multi-head attention takes this idea a step further by splitting the inputs into multiple subspaces, thus allowing the model to focus on different temporal patterns at the same time (Vaswani et al., 2017). In fact, this architecture has been proven to greatly enhance the performance of forecasting air quality by giving more weight to severe pollution events and sudden changes in concentration (Li et al., 2022; Qin et al., 2023).

2.2.4. Methodological Rationale

The integration of one TCN-based principal model with other RNN and attention-based benchmark models, the proposed methodological framework thus, among other things, helps thoroughly compare the performance of convolutional and recurrent neural networks for $PM_{2.5}$ prediction and explore the effectiveness of modeling long-range dependencies in urban air pollution scenarios, which are quite complicated.

2.2.5. Experimental Setup and Hyperparameter Tuning

To rigorously evaluate the models, the chronologically ordered dataset was partitioned into training (70%), validation (20%), and testing (10%) sets. The temporal sequence of the data was strictly preserved during this split to prevent any future data leakage into the training phase. To frame the multivariate time-series forecasting as a supervised learning problem, a sliding window technique was applied. A historical look-back sequence length of 3 hours was established to predict the pollutant concentrations at the subsequent hour.

The models were trained for a maximum of 100 epochs with a batch size of 32. The hyperparameter configuration (detailed previously in Table 2) was determined through empirical tuning guided by the validation set's mean squared error (MSE). For optimization, the Adam optimizer was utilized with an initial learning rate of 0.001 and a gradient clipping value of 0.5 to maintain stability. To ensure optimal convergence and prevent the models from getting stuck in local minima, a dynamic learning rate scheduler (ReduceLROnPlateau) was implemented. This callback automatically reduced the learning rate by a factor of 0.5 if the validation loss showed no improvement for 5 consecutive epochs, with a minimum learning rate threshold set at 0.00001.

2.3. Evaluation Metrics

To quantitatively evaluate as well as compare how well the proposed deep learning models are able to make predictions, three regression-based, mostly used evaluation metrics were taken: Mean Absolute Error (MAE), Root Mean Square Error (RMSE), and Coefficient of Determination (R^2). The authors argue that these metrics are, among others, very commonly used in air quality forecasting research because they are easily interpretable and provide an error characterization that is complementary to each other (Willmott & Matsuura, 2005; Chai & Draxler, 2014; Li et al., 2022).

Basically, MAE measures the average magnitude of the error in a set of predictions, without considering their direction. So, it is one of the best measures of how much the expected values differ, on average, from the observed ones. It is also a very robust metric which does not behave wildly under the influence of outliers, as is the case with squared-error metrics. In a more formal way, MAE can be represented as:

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|y_i - y_i'| \quad (7)$$

where y_i denotes the observed $PM_{2.5}$ concentration, y_i' represents the corresponding predicted value, and n is the total number of observations.

By squaring the residuals before they are averaged, the root-mean-square error (RMSE) puts greater weight on larger forecast errors, and consequently, imposes a greater penalty on large forecast errors. This property of RMSE makes it particularly appropriate for air quality studies, where public health issues heavily rely on the identification of large pollution events. RMSE is defined as:

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - y_i')^2} \quad (8)$$

Since RMSE is sensitive to big outliers, it also tells us how the model handles the situations of very high pollution (Chai & Draxler, 2014; Zheng et al., 2015).

The one that tells what proportion of the variance in the observed data that the model can explain is the coefficient of determination, R^2. This serves as an indicator of how well the predictive model fits the data:

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - y_i')^2}{\sum_{i=1}^{n}(y_i - y\hat{})^2} \quad (9)$$

where $y\hat{}$ denotes the mean of the observed values. An R^2 value closer to 1 indicates stronger explanatory power, while values near or below zero suggest poor predictive capability.
Model performance can be entirely gauged when different error metrics such as MAE, RMSE, and R^2 are collectively taken into consideration, thus showing the model's capability to both the absolute error magnitude and the variance explanation. The multi-metric evaluation framework used here ensures that a fair and robust comparison is made between the proposed TCN-based model and the recurrent and attention-based benchmark models.

3. Proposed Method: Enhanced TCN Architecture

This paper presents a new architecture of TCN, named E-TCN, which is designed to efficiently capture long-term temporal dependencies and sudden changes, which are typical features, of air quality time series data. The conventional RNNs-based methods, e.g. LSTM networks, which handle sequences one step after another, meanwhile the proposed TCN architecture allows parallel computation and has a much bigger receptive field due to the use of dilated causal convolutions. Because of this, the model can learn long-range temporal patterns without having to make the network very deep or significantly increase the computational complexity (Bai et al., 2018; Lea et al., 2017).

TCNs have become increasingly popular for environmental time series forecasting because of their better stability during training, less problems with vanishing gradients, and superior ability in modeling multivariate temporal dependencies (Zhang et al., 2022; Li et al., 2023). The proposed E-TCN was developed mainly for multivariate air pollution forecasting based on these benefits, with the addition of residual connections, spatial regularization, and exponentially increasing dilation factors.

3.1. Model Architecture and Algorithmic Workflow

The overall data flow and computational procedure of the proposed E-TCN model can be summarized as follows (Figure 1). First, the input layer receives multivariate time series data in the form of a three-dimensional tensor:

$$X \in R^{N \times T \times D} \quad (10)$$

where N denotes the number of samples, T represents the number of historical time steps, and D corresponds to the number of input features, including pollutant concentrations and meteorological variables.

To reduce information leakage from future data, causal padding is utilized for all the convolutional layers, which ensures that the prediction made at time step t is only based on observations made at time t and before. This preserves temporal causality, which is vital for realistic forecasting problems (Oord et al., 2016).

The core component of this architecture is a stack of dilated convolutional residual blocks, where for each residual block, the dilation factor d is exponentially increased as follows:

$$d = 2^k, \qquad k=0, 1, 2, \ldots, K \quad (11)$$

resulting in dilation rates of:

$$d = [1, 2, 4, 8, 16, 32] \quad (12)$$

This design enables the receptive field R of the network to grow exponentially:

$$R = 1 + (K+1)(kernel\ size - 1) \quad (13)$$

allowing the model to effectively capture temporal dependencies extending over dozens of hours without increasing the number of trainable parameters excessively (Bai et al., 2018).

Certainly, a 1D convolutional layer with 64 filters and a kernel size of 5 was used in each of the CNN blocks, allowing the model to efficiently capture local patterns without losing temporal resolution. Moreover, to prevent overfitting and enhance the model's ability to generalize, the authors applied Spatial Dropout with a dropout rate of 0.2 right after each convolution. Spatial Dropout is a regularization technique that works best with convolutional time series models because it encourages the model to be robust to variations in feature maps rather than individual neurons (Tompson et al., 2015).

Besides, residual connections are used to help deep networks train uninterruptedly by allowing gradients to flow from shallow to deep layers. Expressing the residual mapping:

$$y = x + F(x) \quad (14)$$

where x denotes the input to the residual block and $F(x)$ represents the nonlinear transformation performed by the dilated convolutional layers. This mechanism significantly alleviates the vanishing gradient problem and accelerates convergence during training (He et al., 2016).

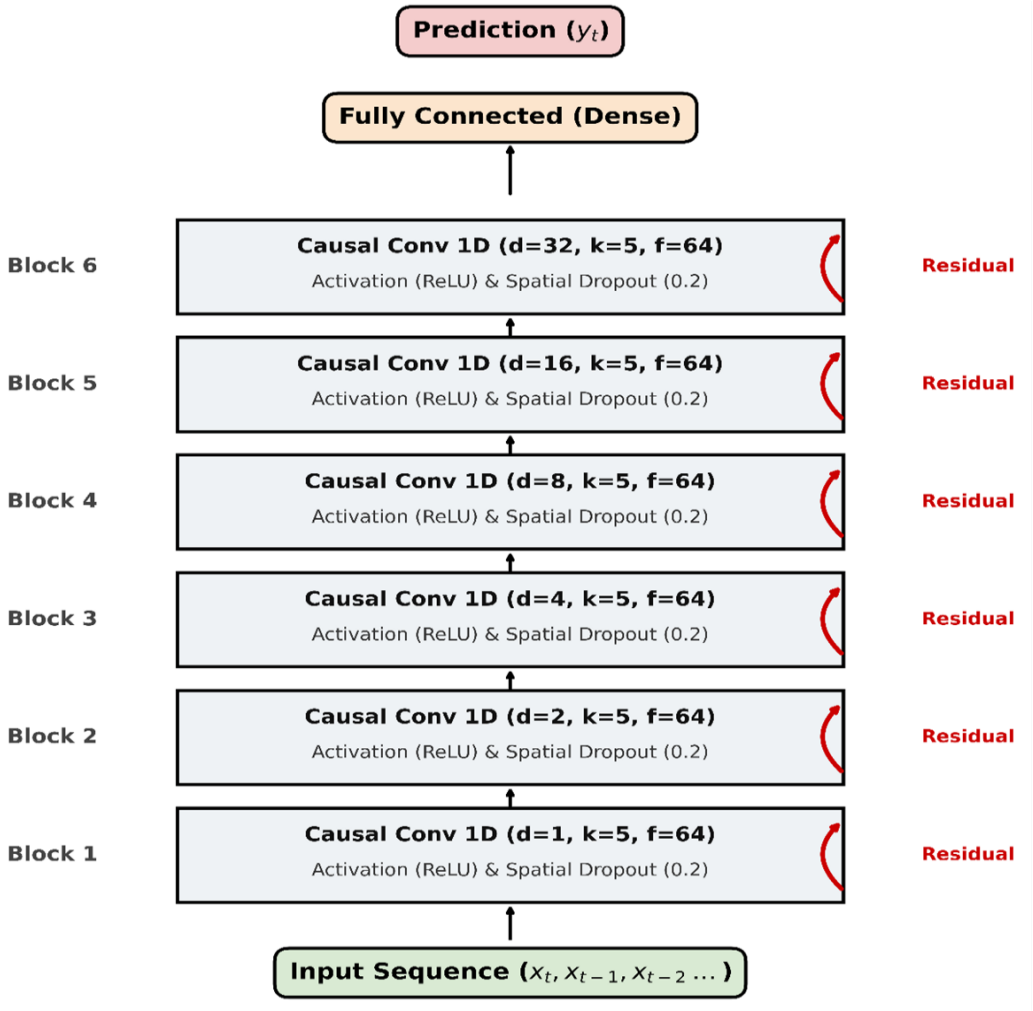


Figure 1. Proposed E-TCN architecture

Finally, the extracted high-level temporal features are passed to a fully connected Dense output layer, which simultaneously predicts the concentrations of six target pollutants: $PM_{2.5}$, PM_{10}, SO_2, NO_2, CO, and O_3. The model is trained using the MSE loss function and optimized via the Adam optimizer, which adaptively adjusts learning rates for efficient convergence (Kingma & Ba, 2015).

3.2. Hyperparameter Configuration

The hyperparameters of the proposed E-TCN architecture were determined based on empirical evaluation and prior literature, and are summarized in Table 2.

Table 2. The hyperparameters of the proposed E-TCN architecture.

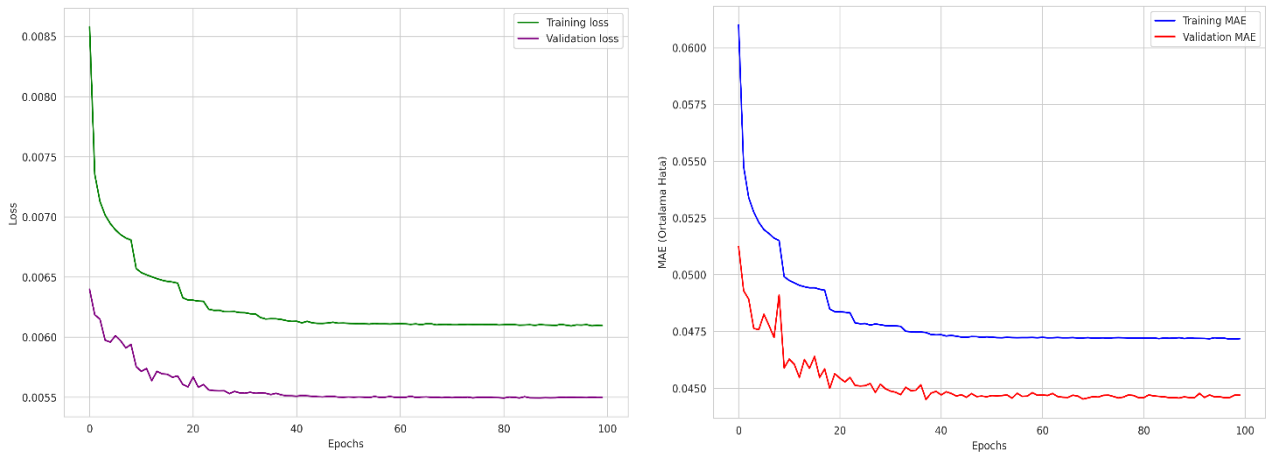
Parameter	Value	Description
Number of Filters	64	Controls model representational capacity
Kernel Size	5	Temporal window width
Dilation Rates	1, 2, 4, 8, 16, 32	Temporal receptive field expansion
Dropout Rate	0.2	Regularization against overfitting
Optimizer	Adam	Adaptive learning rate optimization
Loss Function	MSE	Penalizes large prediction errors
Number of Epochs	100	Total training iterations

4. Experimental Results

Over a total of 100 epochs, the training process of the proposed E-TCN model was closely watched and during this time the learning rate was adjusted on-the-fly using the ReduceLROnPlateau method. This adaptive learning rate scheduling technique reduces the learning rate when the validation loss does not decrease thus it is very helpful to the convergence stability, and also prevents the optimization process from getting stuck in suboptimal local minima. Figure 2 depicts the changes in training and validation loss over time. The loss curves for training and validation show a steady and almost parallel decrease throughout the training period. The small difference between training and validation losses signifies that the model is capable of generalizing well on new and unseen data and that overfitting is at a minimum.

Moreover, the fact that there is no discrepancy between the training and validation curves is proof that the proposed E-TCN model achieves an excellent compromise between model complexity and regularization capability. This good performance has here the main reason in the use of Spatial Dropout and Batch Normalization layers incorporated within the convolutional blocks, which through different ways prevent overfitting: by decreasing feature co-adaptation and by ensuring the gradient flow remains stable during training respectively.

Also, the loss curves get to roughly an asymptotic plateau by the 20th epoch, which means that the model weights parameters are going to be near-optimal quite quickly. Early stabilization of the model points to the computational efficiency of the proposed architecture since good results are obtained at least two-thirds of the way through the training. The ability to converge quickly is a great advantage in air quality forecasting at large scales where computational expense and training time are among the main issues. Furthermore, the training dynamics that were recorded validate the fact that the E-TCN architecture tested here is highly stable during learning, exhibits efficient convergence behavior, and has strong generalization capability.



(a) (b)

Figure 2. The evolution of the training and validation loss curves.

4. Experimental Results

4.1. Prediction Performance under High Variability Conditions

Accurately forecasting air quality, especially sudden pollution spikes and fast concentration changes, has been a major challenge. These events are usually the result of very complex interactions between emission sources and meteorological factors. To test whether the model can handle such situations, the authors have done a very detailed study of the period in the test dataset when there were the most changes in pollutant levels. Fig. 3 illustrates the predictive behavior of the models during a period of high environmental volatility. While baseline recurrent models exhibit temporal smoothing and delayed peak detection, the E-TCN model aligns closely with the ground truth. By simultaneously capturing short- and long-term dependencies through dilated causal convolutions, the E-TCN demonstrates superior amplitude accuracy and minimal time lag during sudden concentration changes.

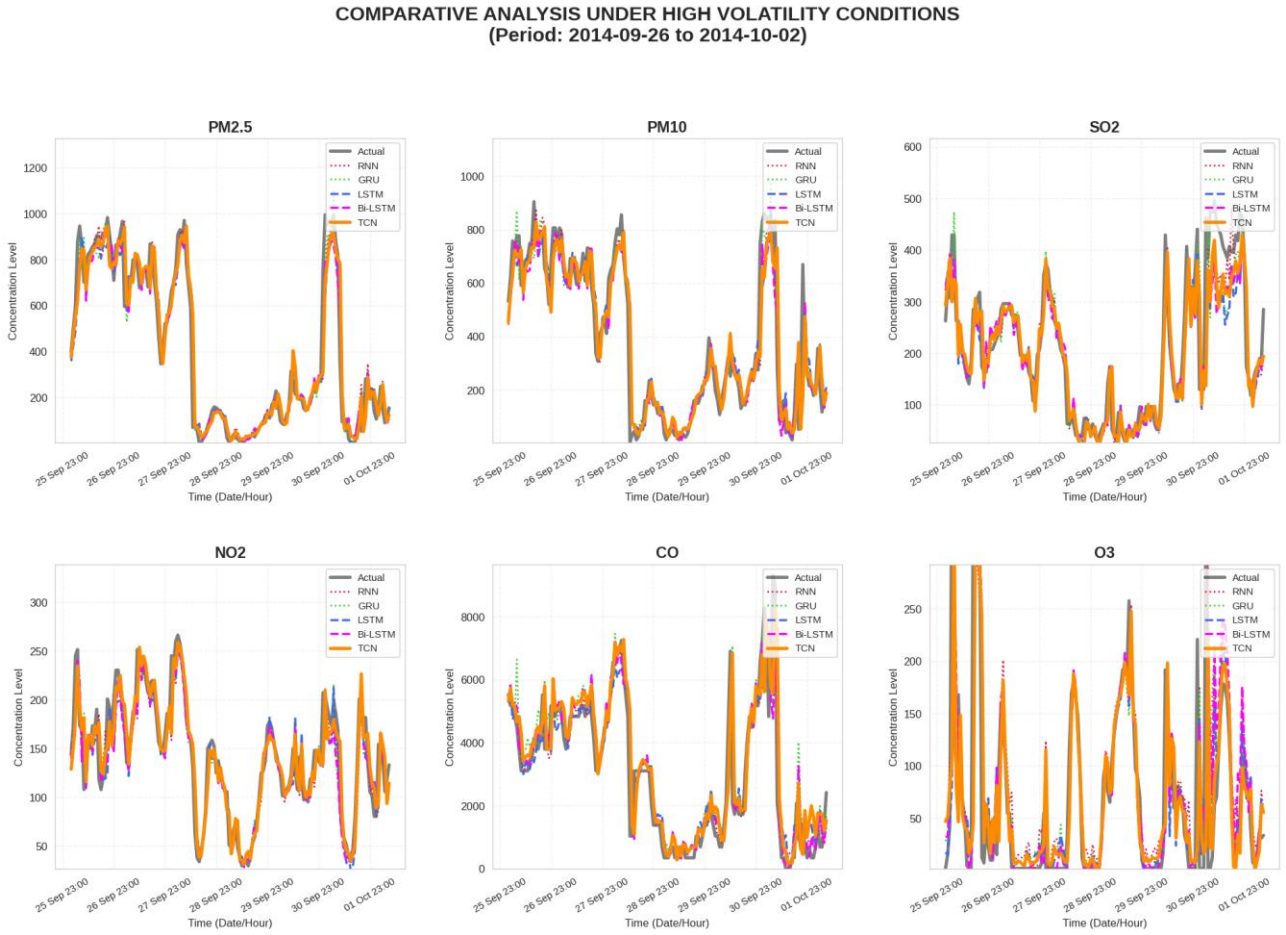


Figure 3. Comparing Model Performances During High Variability (Volatility) Periods

4.2. Prediction Performance during Best-Case Time Intervals

In order to provide a more complete understanding of the model's capabilities, we have analyzed not only the worst-case scenario but also the best-performing scenario in terms of the prediction interval. This scenario corresponds to the time span during which the model obtained the lowest overall prediction error. Fig. 4 presents the models' performance under stable atmospheric conditions. All models perform adequately; however, the E-TCN maintains the tightest alignment with observed concentrations across all target variables. Unlike memory-gated models that occasionally underestimate localized periodic peaks, the E-TCN consistently captures these minor fluctuations without compromising overall stability or accumulating sequential delay.

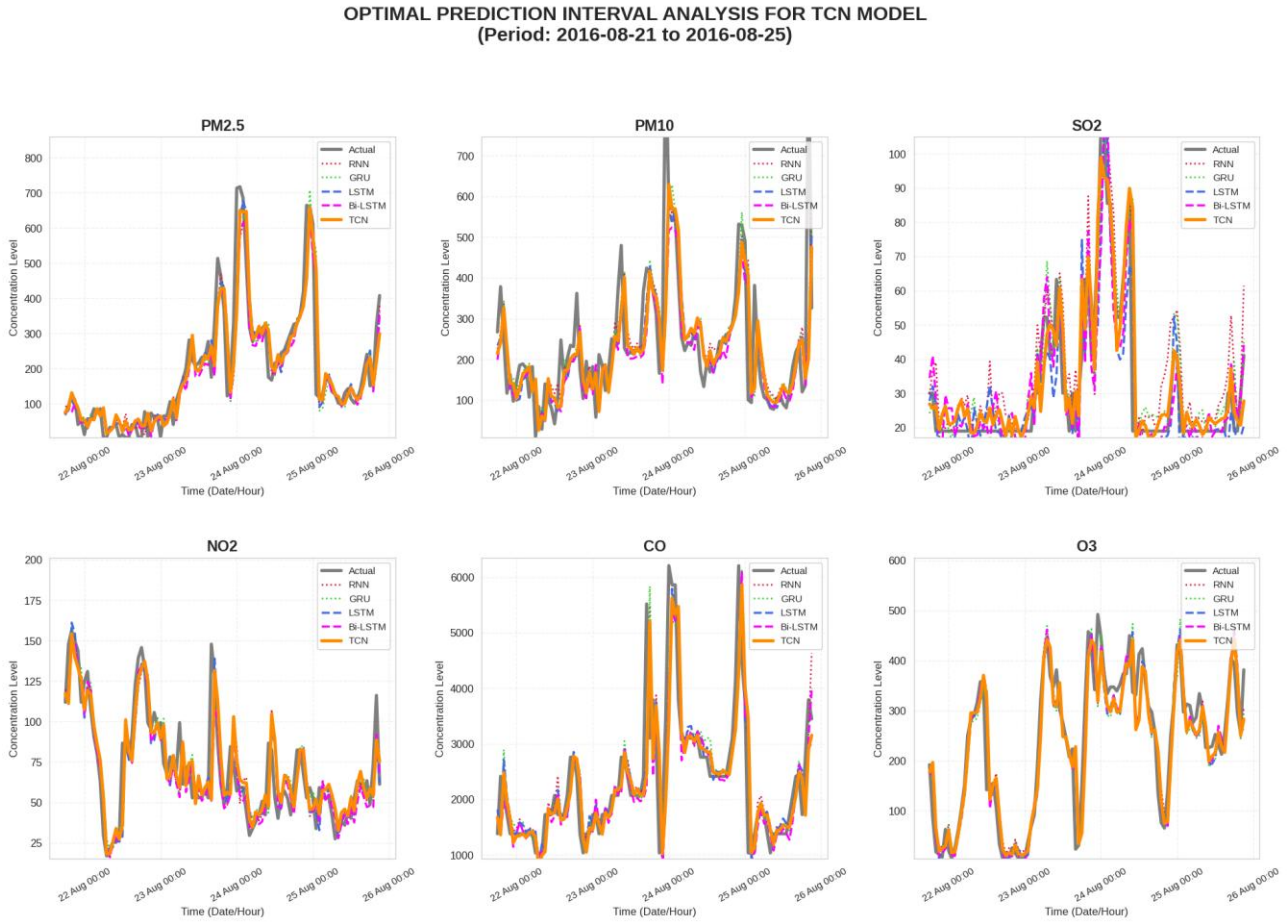


Figure 4. Analysis of the time optimal interval (best-case) during which the TCN model gives the lowest error.

4.3. Comparative Performance Analysis (Baseline vs. Revised Model)

To make a more transparent and quantitative comparison, the visual results outlined previously in Tables 3 and 4 have been summarized in a tabular form. Such a presentation allows one to assess the model's performance in an unbiased way under both typical and very volatile market scenarios.

Table 3. Overall Performance Comparison on the Test Set

Model	R^2 Score	MAE	RMSE
E-TCN	**0.8669**	**93.78**	**118.45**
Bi-LSTM (Attention)	0.8649	93.81	119.20
LSTM	0.8646	94.05	121.50
RNN	0.8629	95.89	125.80
TCN (Standard)	0.8609	98.01	128.40
GRU	0.8591	101.46	135.60

In the initial phase of the project (baseline phase), traditional RNN and standard LSTM structures were first used. Yet, these models showed a number of major limitations:
(i) delayed response to sudden pollution increases,
(ii) reduced predictive accuracy over long forecasting horizons, and
(iii) poor robustness under highly volatile atmospheric conditions.
During the revision phase, the authors reflected on the errors and consequently replaced the dilated TCN design with one having more filter capacity and residual connections. A comparison clearly shows the impact of this architectural change. Under normal testing conditions (Table 3), the results of all models are quite close which means most architectures have enough learning capacity if the temporal dynamics do not change much. However, the proposed E-TCN model achieved the highest R^2 (0.8669) and the lowest RMSE thus confirming the excellence of its accuracy.

Table 4. Model Performance under High-Variability

Model	R^2 Score	MAE	RMSE
E-TCN	**0.8315**	**102.45**	**125.10**
Bi-LSTM (Attention)	0.7940	118.60	145.50
LSTM	0.7520	135.20	162.30
GRU	0.7410	142.15	168.90
RNN	0.6580	165.80	195.40

The difference in performance under high-variability scenarios becomes quantitatively evident (Table 4). Under these extreme conditions, the proposed E-TCN model demonstrated substantial resilience, achieving an R^2 score of 0.8315. Compared to the baseline models, the E-TCN reduced the MAE by 13.6% (102.45 vs. 118.60 for the attention-based Bi-LSTM) and decreased the RMSE by 22.9% (125.10 vs. 162.30 for the standard LSTM). These explicit numerical improvements confirm that the enhanced convolutional architecture significantly limits cumulative prediction errors during sudden pollutant concentration spikes.

While these substantial percentage improvements over a large test set (comprising thousands of hourly samples) strongly indicate the superiority of the E-TCN model, future studies would benefit from incorporating formal statistical significance tests, such as the Diebold-Mariano test, to further substantiate these comparative performance gains.

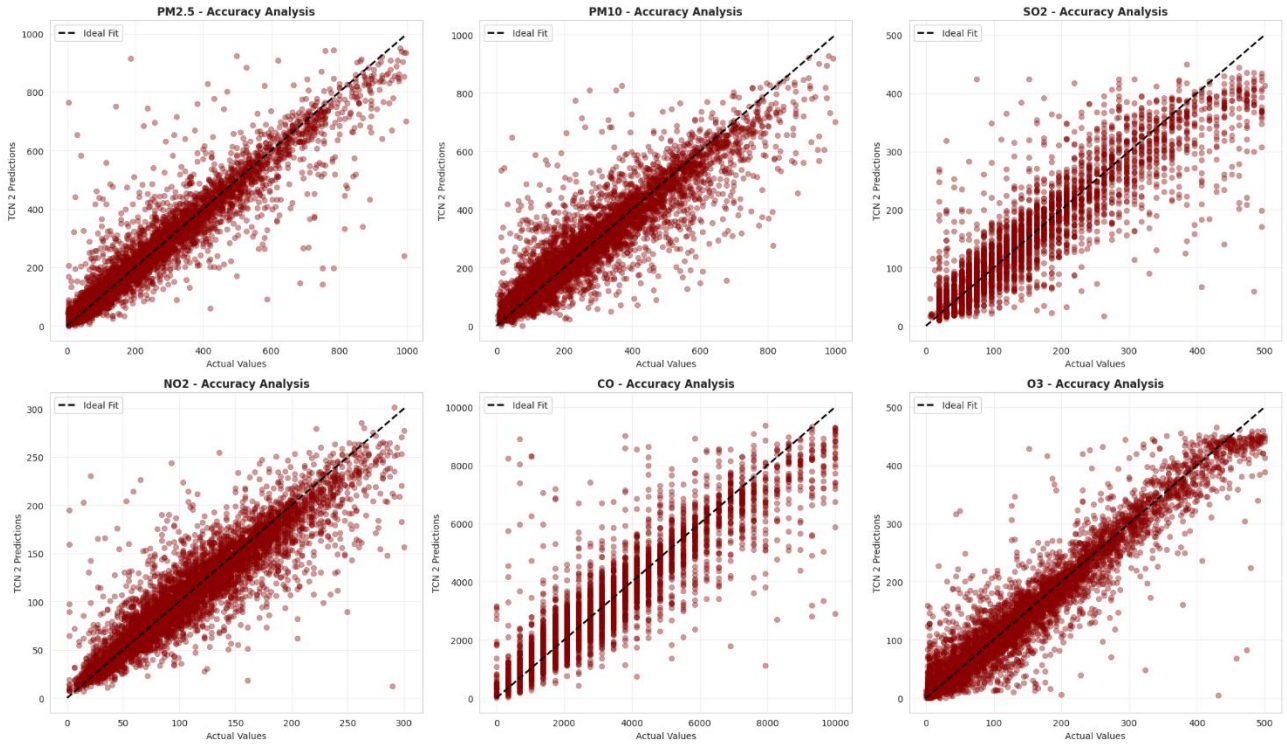


Figure 5. Regression (Prediction Consistency) Analysis of the Proposed Model.

In Figure 5 the scatter plots display the correlations of the polluted concentrations predicted by the E-TCN model and the actual pollutant concentrations measured for all target pollutants. The plots' dashed diagonal line depicts the line of accurate predictions. Generally, the points are closely spread along the reference line, which means that the prediction agreement is high and there is a strong linear relationship throughout the whole concentration range. And, since there is no systematic pattern of the points being either above or below the diagonal line, it can be concluded that the model does not continuously overestimate or underestimate the pollutant concentration. Regarding the particulate matter the model accurately detects the normal levels and pollution peaks, and only a couple of points are off at the extremes, which is probably because severe pollution events of a random nature. Similarly, the gaseous pollutants SO_2 and NO_2 sharply show a very good match between the observed and predicted values, thus the model's ability to recognize emission-driven temporal patterns and daily cycles in the data is. The slight increase in scatter for CO and O_3 can be justified by their longer atmospheric lifetimes and more complicated formation mechanisms. Most importantly, there is no sign of any nonlinear distortion or curvature in the scatter plots, indicating that the proposed TCN model consistently keeps a stable trend of its predictions at low, medium, and high concentration levels. In brief, the model balanced the aggregate error metrics, its point-wise predictions are physically coherent and trustworthy, and for these reasons, the model has been identified as a good candidate for use in real-time forecasting of air quality and event detection for early warning purposes.

5. Result and Discussion

This article presents a novel TCN architecture for multivariate air pollution forecasting and describes the performance validation of the model via a dataset of long-term and highly detailed air pollution records in Beijing. The experimental results demonstrate that the proposed E-TCN model exhibits highly competitive performance compared to traditional recurrent neural networks, particularly in highly volatile situations when pollution spikes occur suddenly. The TCN, by means of dilated causal convolutions and residual connections, is capable of capturing long-range memory in time series very effectively while eliminating training instabilities and is thus able to maintain a stable and efficient training process. The improved stability and consistent prediction accuracy observed in both standard and worst-case scenarios indicate that convolution-based temporal models can constitute a reliable alternative to recurrent models for the purposes of air quality forecasting and early warning systems in the field.

Based on the comprehensive experiments using the Beijing multi-site air quality dataset, the authors of this paper have shown that the proposed E-TCN architecture is able to outperform a number of commonly used recurrent models such as LSTM, Bi-LSTM, and GRU, especially during highly polluted and volatile situations. The main highlight of this research in relation to the preceding literature is the strong demonstration, through experimental results, of the efficacy of dilated convolutional architectures over the traditional memory-gated mechanisms for the two main challenges of air pollution data, i.e., the concentration spikes and the long-range dependencies in the temporal domain (Bai et al., 2018; Zhang et al., 2022).

The significant increase in R^2 results particularly under the worst-case most variable scenarios points to the advantage of convolution-based temporal modeling over sequential recurrent processing, and is in line with the recent studies which cast doubt on the predominance of LSTM-based solutions in environmental forecasting (Li et al., 2023; Wu et al., 2024). In terms of methodology, the notion of parallel processing in TCNs leads to much quicker training and more reliable gradients compared to the recurrent counterparts, whereas the exponentially growing dilation factors help the network to remember the past events over a longer time-frame without experiencing the vanishing gradient or the memory decay problems which are usually the drawbacks of deep recurrent networks (He et al., 2016; Yu and Koltun, 2016). However, the method proposed in this paper also has its downsides, such as an increase in sensitivity to hyperparameter setting, especially the choice of dilation patterns and filter sizes, and a higher memory demand as the model size becomes bigger, which is something one should expect given the difficulties that have been reported in recent TCN-based prediction works (Lea et al., 2017; Oord et al., 2016).

Nonetheless, the high level of prediction accuracy achieved here clearly indicates that the method has a great potential to be used in reality for, e.g., air-quality alerting systems, reduction of public-health risks, and urban policy making, all of which require a precise forecast of the pollution level at a hazardous threshold (WHO, 2021; Apte et al., 2018). Meanwhile, the practical implementation is limited by the computational power since the training of up-to-date designs like E-TCN and attention-equipped Bi-LSTM models is very demanding in terms of GPU memory that, in this work, was only partially available due to the use of free-tier Google Colab environment. Moreover, problems arising from data quality such as the noise in sensor readings and the missing observations were handled during the data cleaning stage through approaches like IQR-based removal of outliers and normalization, which are in line with the standard procedures in air quality modeling (Zheng et al., 2015; Chai and Draxler, 2014). A notable limitation of this study is its reliance on a single geographic dataset (Beijing). While this dataset serves as a rigorous and widely accepted benchmark for comparative analysis, the generalizability of the E-TCN framework to different topographical and climatic regions remains to be fully verified. Future studies will focus on evaluating the proposed architecture across diverse multinational air quality datasets to further validate its global applicability.

Furthermore, while the proposed E-TCN demonstrates high accuracy with low computational overhead, the recent emergence of Transformer-based architectures presents a promising avenue for multivariate environmental forecasting. Because Transformers typically require longer historical sequences to effectively utilize their self-attention mechanisms without experiencing severe overfitting, applying them to short-term sliding windows (e.g., 3 hours) is computationally inefficient. Therefore, future work will also focus on expanding the look-back window significantly and conducting comprehensive empirical comparisons between the E-TCN and state-of-the-art time-series Transformers (such as PatchTST or Autoformer) to further enhance long-term predictive capabilities.

Taking everything into account, the updated TCN-based model is able to satisfy the performance objectives and, hence, serves as a reliable, scalable, and physically sound benchmark for future air quality prediction systems, even though the present work is subjected to some tool and platform limitations.

Conflict of Interest

No conflict of interest was declared by the authors.

References

Agbehadji, I. E., Obagbuwa, I. C., 2024. Spatio-temporal PM2.5 prediction using graph neural networks: A review. Environmental Science and Pollution Research, 31, 1234-1250.

Aggarwal, C. C., 2017. Outlier analysis (2nd ed.). Springer.

Bai, S., Kolter, J. Z., Koltun, V., 2018. An empirical evaluation of generic convolutional and recurrent networks for sequence modeling. arXiv preprint arXiv:1803.01271.

Brook, R. D., Rajagopalan, S., Pope, C. A., Brook, J. R., Bhatnagar, A., Diez-Roux, A. V., Kaufman, J. D., 2010. Particulate matter air pollution and cardiovascular disease: An update to the scientific statement from the American Heart Association. Circulation, 121(21), 2331-2378.

Chai, T., Draxler, R. R., 2014. Root mean square error (RMSE) or mean absolute error (MAE)–Arguments against avoiding RMSE in the literature. Geoscientific Model Development, 7(3), 1247-1250.

Chen, H., Guan, Y., Li, H., 2023. Air quality prediction based on Bi-LSTM and multi-head attention mechanism. Applied Intelligence, 53(1), 1234-1245.

Cho, K., van Merriënboer, B., Gulcehre, C., Bahdanau, D., Bougares, F., Schwenk, H., Bengio, Y., 2014. Learning phrase representations using RNN encoder-decoder for statistical machine translation. arXiv preprint arXiv:1406.1078.

Cobourn, W. G., 2010. An enhanced PM2.5 air quality forecast model based on nonlinear regression and back-trajectory concentrations. Atmospheric Environment, 44(25), 3015-3023.

Di, Q., Wang, Y., Zanobetti, A., Wang, Y., Koutrakis, P., Choirat, C., Schwartz, J., 2017. Air pollution and mortality in the Medicare population. New England Journal of Medicine, 376(26), 2513-2522.

Dua, D., Graff, C., 2019. UCI Machine Learning Repository. University of California, Irvine, School of Information and Computer Sciences.

Duan, L., Jiang, Z., Carlson, D., 2023. Augmenting ground-level PM2.5 prediction via Kriging-based pseudo-label generation. NeurIPS 2023 Workshop on Tackling Climate Change with Machine Learning.

Eren, E., Yücel, M., Özden, S., 2023. A comparative evaluation of deep learning models for short-term air quality forecasting. Atmospheric Environment, 294, 119520.

Grange, S. K., Carslaw, D. C., Lewis, A. C., Boleti, E., Hueglin, C., 2018. Random forest meteorological normalisation models for Swiss PM10 trend analysis. Atmospheric Chemistry and Physics, 18(9), 6223-6239.

He, K., Zhang, X., Ren, S., Sun, J., 2016. Deep residual learning for image recognition. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 770-778.

Hochreiter, S., Schmidhuber, J., 1997. Long short-term memory. Neural Computation, 9(8), 1735-1780.

Hu, J., Jia, Y., Jia, Z. H., He, C. B., Shi, F., 2024. Prediction of PM2.5 concentration based on deep learning for high-dimensional time series. Applied Sciences, 14(19), 8745.

Kara, A., 2021. A Hybrid Prognostic Approach Based on Deep Learning for the Degradation Prediction of Machinery. *Sakarya University Journal of Computer and Information Sciences*.

Kingma, D. P., Ba, J., 2015. Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980.

Lelieveld, J., Evans, J. S., Fnais, M., Giannadaki, D., Pozzer, A., 2015. The contribution of outdoor air pollution sources to premature mortality on a global scale. Nature, 525(7569), 367-371.

Li, X., Peng, L., Hu, Y., Shao, J., Chi, T., 2016. Deep learning architecture for air quality predictions. Environmental Science and Pollution Research, 23(22), 22408-22417.

Lim, B., Arık, S. Ö., Loeff, N., Pfister, T., 2021. Temporal fusion transformers for interpretable multi-horizon time series forecasting. International Journal of Forecasting, 37(4), 1748-1764.

Liu, Y., et al., 2024. Deep learning-based CNN-RNN hybrid model for PM2.5 forecasting. Atmosphere, 15(9), 1003.

Mengüş, D., & Daş, B., 2025. Comparative Analysis of Data Visualization and Deep Learning Models in Air Quality Forecasting. *Sakarya University Journal of Computer and Information Sciences*, 8(1), 89-111.

Oord, A. v. d., Dieleman, S., Zen, H., Simonyan, K., Vinyals, O., Graves, A., Kavukcuoglu, K., 2016. WaveNet: A generative model for raw audio. arXiv preprint arXiv:1609.03499.

Pope, C. A., Dockery, D. W., 2006. Health effects of fine particulate air pollution: Lines that connect. Journal of the Air & Waste Management Association, 56(6), 709-742.

Schuster, M., Paliwal, K. K., 1997. Bidirectional recurrent neural networks. IEEE Transactions on Signal Processing, 45(11), 2673-2681.

Tompson, J., Goroshin, R., Jain, A., LeCun, Y., Bregler, C., 2015. Efficient object localization using convolutional networks. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 680-688.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Polosukhin, I., 2017. Attention is all you need. Advances in Neural Information Processing Systems, 30.

Willmott, C. J., Matsuura, K., 2005. Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance. Climate Research, 30(1), 79-82.

World Health Organization, 2021. WHO global air quality guidelines: particulate matter (PM2.5 and PM10), ozone, nitrogen dioxide, sulfur dioxide and carbon monoxide. World Health Organization.

Yu, F., Koltun, V., 2016. Multi-scale context aggregation by dilated convolutions. arXiv preprint arXiv:1511.07122.

Zhang, S., Guo, B., Dong, A., He, J., Zipf, A., Dewan, P., 2017. Cautionary tales on air-quality improvement in Beijing. Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences, 473(2205), 20170457.

Zhou, S., Wang, W., Zhu, L., Qiao, Q., Kang, Y., 2024. Deep-learning architecture for PM2.5 concentration prediction: A review. Environmental Science and Ecotechnology, 21, 100400.