Short-Term Forecasting of Urban Carbon Monoxide Concentrations Using Deep Learning and Time Series Models: A Case Study of Mashhad

Rahil Ahmadi[1], Toktam Zoughi[2*]

[1] *Department of Computer Engineering, Technical and Vocational University (TVU), Tehran, Iran, ORCiD:, Email: rahil.ahmadi2595@gmail.com*

[2] *Department of Computer Engineering, Technical and Vocational University (TVU), Tehran, Iran, ORCiD: 0000-0002-1797-6910, Email: t.zoughi@shariaty.ac.ir*

1. Abstract

Accurate short-term forecasting of urban air pollutants is essential for effective air quality management and the development of early warning systems. This study proposes a data-driven forecasting framework, termed UniTCN-CO (Univariate Temporal Convolutional Network for CO Forecasting), for predicting daily carbon monoxide (CO) concentrations in the city of Mashhad, Iran. The model is developed using real monitoring data collected at the Sajad air quality station over the period 2017–2021. Prior to modeling, the CO time series was carefully preprocessed by removing unrealistic observations, filling missing values through linear interpolation, and applying Min–Max normalization to ensure stable model training. The processed series was then transformed into supervised learning samples using a sliding time window of seven consecutive days, enabling short-term prediction of daily CO concentrations. Within a unified univariate input framework, several forecasting models were implemented and comparatively evaluated, including Linear Regression, Prophet, Long Short-Term Memory (LSTM), Bidirectional LSTM (Bi-LSTM), a Temporal Convolutional Network (TCN), and a hybrid Prophet–LSTM model. Model performance was assessed on an independent test dataset using three widely adopted error metrics: mean absolute error (MAE), root mean square error (RMSE), and symmetric mean absolute percentage error (SMAPE). The empirical results indicate that deep learning–based approaches outperform classical statistical models in capturing the temporal dynamics and variability of urban CO concentrations. Among the evaluated models, the proposed UniTCN-CO framework—based on a Temporal Convolutional Network architecture—demonstrates the most stable and reliable performance across all evaluation metrics. Further analysis through visual comparison of observed and predicted time series, together with residual diagnostics, reveals that the selected model produces prediction errors largely centered around zero without evident systematic bias, while effectively capturing both the dominant temporal trends and a substantial portion of short-term fluctuations in CO levels. Overall, the findings highlight the effectiveness of the proposed UniTCN-CO framework as a practical and computationally efficient solution for short-term urban air quality forecasting. Such a framework can support environmental monitoring agencies in developing reliable early warning systems and improving pollution management strategies. Future research may extend this approach by incorporating multi-station observations, additional pollutants, and higher temporal resolution data to further enhance predictive capability and spatial representativeness in urban air quality modeling.

2. Keywords:

Carbon monoxide (CO); Air quality forecasting; Time series prediction; Deep learning; Temporal Convolutional Network (TCN); Urban air pollution; Sliding window approach; Early warning systems.

3. introduction

Air pollution has emerged in recent decades as one of the most critical global challenges to human health. Despite substantial advances in industry and technology, millions of people worldwide remain exposed to polluted air that poses serious direct and indirect risks to human well-being. This growing concern has led to the recognition of access to clean air as a fundamental public health necessity. Numerous epidemiological studies have documented strong associations between air pollution exposure and increased mortality, as well as the prevalence of chronic respiratory and cardiovascular diseases. Recent investigations further indicate that air pollution is not confined to outdoor environments; interacting factors such as climate change, indoor air conditions, and rapid developments in transportation systems have significantly increased the complexity of this problem. Consequently, continuous assessment of recent advances, challenges, and research directions in air pollution studies is essential for improving health impact assessment and supporting evidence-based environmental policy-making [1].

Among atmospheric pollutants, carbon monoxide (CO) is recognized as one of the most hazardous toxic gases generated by the incomplete combustion of fossil fuels. CO emissions originate primarily from mobile sources, particularly motor vehicles, and to a lesser extent from stationary urban sources. Its colorless and odorless nature, combined with its ability to bind with hemoglobin and disrupt oxygen transport in the human body, makes CO especially dangerous in urban environments. The temporal variability of CO concentrations is strongly influenced by traffic flow patterns, meteorological conditions, and urban morphology, collectively rendering accurate prediction of this pollutant a complex and challenging task [2]. These characteristics highlight the necessity of developing precise and efficient predictive models to support urban air pollution control and management strategies.

While air quality monitoring systems play a fundamental role in assessing current pollution levels, monitoring alone is insufficient for effective urban environmental management and decision-making. Preventive and mitigation measures—such as traffic regulation policies, public health warnings, and emergency responses during pollution episodes—require timely and reliable forecasts of air pollutant concentrations. In this context, data-driven prediction of air pollution using advanced machine learning techniques has emerged as a crucial complement to traditional monitoring approaches. Over the past decade, this predictive paradigm has attracted increasing attention in environmental science and urban planning due to its potential to reduce the adverse public health impacts of air pollution [3].

Conventional air quality forecasting approaches, including statistical regression methods, classical time-series models, and physicochemical dispersion simulations, have been widely applied. However, these methods often rely on simplified assumptions and require extensive physical parameterization, limiting their ability to capture nonlinear relationships, abrupt variations, and long-term temporal dependencies in pollutant concentrations. Such limitations are particularly evident in complex urban environments characterized by heterogeneous traffic patterns and incomplete data infrastructures. In contrast, machine learning approaches offer greater flexibility by learning complex nonlinear patterns directly from data and by effectively handling high-dimensional inputs, leading to improved forecasting performance in many applications [4].

Building upon these developments, machine learning and deep learning techniques have opened new perspectives in air pollution forecasting. By exploiting historical observations and uncovering latent temporal patterns, models such as Random Forest and Recurrent Neural Networks (RNNs), including Long Short-Term Memory (LSTM) and Bidirectional LSTM (Bi-LSTM), have demonstrated promising results in forecasting urban air pollutants such as particulate matter, nitrogen oxides, and carbon monoxide [5]. Nevertheless, the predictive performance of these models remains highly dependent on architectural design choices, feature selection strategies, and their adaptability to local environmental and urban characteristics.

Mashhad, one of the most populous metropolitan areas and a major religious center in Iran, faces distinct air pollution challenges. The city experiences an annual influx of millions of pilgrims, resulting in substantial increases in traffic volume during specific periods of the year. When combined with unfavorable meteorological conditions and atmospheric stability, these factors lead to pronounced temporal fluctuations in pollutant concentrations, particularly carbon monoxide. Such conditions make accurate temporal prediction of CO concentrations in Mashhad both scientifically significant and operationally challenging [6].

Despite the importance of carbon monoxide as a critical urban air pollutant, much of the existing air quality research in Iran has predominantly focused on particulate matter. Studies specifically addressing CO concentration forecasting—especially those employing advanced machine learning and time-series-based neural network approaches—remain relatively limited. Moreover, existing research has rarely provided systematic comparisons between classical statistical models, recurrent neural network architectures, and convolution-based temporal models within a unified, data-driven framework.

To address these gaps, the present study proposes a comprehensive one-step-ahead forecasting framework based on a sliding window approach ($L = 7$ days) for predicting daily CO concentrations in Mashhad, Iran. The framework evaluates and compares multiple modeling paradigms, including linear regression, Prophet, LSTM, Bi-LSTM, Temporal Convolutional Networks (TCNs), and a hybrid Prophet–LSTM model. In particular, TCNs, with their causal convolutional structure and flexible receptive fields, are investigated as a promising yet underexplored alternative for short-term urban CO forecasting. The results are intended to provide both quantitative and qualitative insights into model performance and to support the development of reliable, data-driven tools for urban air quality management.

The remainder of this paper is organized as follows. Section 2 reviews related literature and describes the air quality data obtained from Mashhad's monitoring stations. Section 3 presents the proposed UniTCN-CO methodology and model implementation details. Section 4 reports and discusses the experimental results, and Section 5 concludes the study and outlines directions for future research.

4. Literature Review

Reviewing previous studies plays a fundamental role in identifying prevailing scientific trends, commonly employed methodologies, and existing research gaps in the field of air pollution prediction. A comprehensive examination of earlier research provides a clearer understanding of the achievements and limitations of different approaches and facilitates the appropriate positioning of the present study within the existing body of knowledge. Accordingly, this section first reviews studies related to air pollution in general and carbon monoxide as a critical pollutant. Subsequently, classical air quality prediction methods and approaches based on machine learning and deep learning are discussed. Finally, particular emphasis is placed on studies conducted in Iran—especially in the city of Mashhad—to clearly identify the research gaps that motivate the present work.

2.1. Studies Related to Air Pollution and the Carbon Monoxide Pollutant

Air pollution, particularly in urban environments, represents one of the most serious environmental and public health challenges worldwide. Understanding the temporal and spatial behavior of air pollutants and identifying their major sources are essential steps toward effective air quality management and the mitigation of adverse health effects. In recent years, advances in satellite remote sensing and spatiotemporal analysis techniques have enabled more comprehensive and accurate assessments of atmospheric pollutants.

In a study conducted by Mathew et al. (2024), the spatiotemporal distribution of air pollutants in India was analyzed using Sentinel-5P TROPOMI satellite imagery. The study examined pollutants including carbon monoxide (CO), nitrogen dioxide (NO_2), sulfur dioxide (SO_2), and ozone over the period from 2019 to 2022, encompassing pre-pandemic, pandemic, and post-pandemic phases. The results indicated higher pollutant concentrations during winter months, with NO_2 hotspots predominantly observed in metropolitan regions and SO_2 hotspots located near industrial areas. Furthermore, a notable reduction in NO_2 concentrations during the COVID-19 lockdown period in 2020 was reported. This study demonstrates the effectiveness of satellite-based data in large-scale air quality assessment and in identifying major pollution sources, providing valuable insights for air pollution management and public health protection strategies [7].

Carbon monoxide is recognized as one of the most critical air pollutants due to its direct and severe impacts on human health, making its accurate monitoring and assessment essential. In a study by Salthammer (2024), CO was investigated as a priority pollutant in indoor environments. The findings revealed that elevated CO concentrations often occur abruptly as a result of incomplete combustion processes and may lead to severe or even fatal health outcomes. From a technological perspective, recent advances in CO sensing technologies—particularly infrared spectroscopy combined with machine learning algorithms—have enabled accurate, real-time monitoring of this pollutant. The study also emphasized that CO can serve as an effective indicator of indoor air quality and that integrating sensor data with environmental variables such as temperature, humidity, PM2.5, and CO_2 can enhance air quality control and energy management in indoor settings [8].

2.2. Air Quality Prediction Methods

2.2.1. Classical Air Quality Prediction Methods

With the growing availability of urban environmental datasets and the increasing demand for reliable air quality forecasts, researchers have extensively employed classical statistical models and time-series approaches to analyze and predict air pollutants such as carbon monoxide and particulate matter.

In a recent study, Bunnag (2023) investigated the prediction of PM10 concentrations in Bangkok using classical time-series models, including SARIMA and SARIMA-GARCH. Monthly air pollution data from 2008 to 2023 were analyzed, and model performance was evaluated using standard error metrics such as MAE and RMSE. The results demonstrated that the SARIMA(3,1,3)(1,1,2)12 model outperformed the hybrid SARIMA-GARCH approach, achieving lower prediction errors. These findings indicate that classical statistical time-series models can still provide satisfactory performance in air pollution forecasting and remain useful tools for environmental analysis and decision-making [9].

One of the seminal studies in air pollution forecasting was conducted by Gocheva-Ilieva et al. (2013), who examined the air quality status of the city of Blagoevgrad in Bulgaria. Using hourly data collected over a one-year period, the study focused on six major pollutants, including NO, NO_2, NOx, PM10, SO_2, and O_3. Factor analysis was first applied to identify interrelationships among pollutants and to group them into three dominant factors. Subsequently, short-term pollutant forecasting was performed using SARIMA models based on the Box–Jenkins methodology, with the Yeo–Johnson transformation employed to stabilize variance. The results demonstrated that SARIMA models could accurately predict pollutant concentrations—particularly PM10 and ozone—up to 72 hours in advance. This study provides a solid foundation for the application of classical statistical models in air quality monitoring and forecasting, particularly in small urban environments [10].

In another study, Ansari and Alam (2023) proposed an intelligent air quality prediction framework based on the integration of IoT technologies and cloud computing. Using univariate time-series data collected from IoT sensors across different states in India, the authors developed a hybrid BO-HyTS model that combines SARIMA with a Long Short-Term Memory (LSTM) network. This hybrid approach successfully captured both linear and nonlinear characteristics of the data and achieved superior predictive accuracy compared with traditional time-series models and standalone neural networks. The proposed model achieved error values of MSE = 632.20, RMSE = 25.14, and MAE = 20.49, highlighting the effectiveness of integrating advanced computational technologies with hybrid prediction models for air quality management and decision-making [11].

2.2.2. Machine Learning–Based Methods

With the expansion of urban environmental datasets, machine learning techniques have been increasingly adopted to model complex relationships among pollutants and to enhance prediction accuracy.

In a study by Almubaidin et al. (2024), carbon monoxide concentration prediction in urban environments was examined using a variety of machine learning models. Data from Petaling Jaya, Malaysia, were analyzed using six approaches, including linear regression, decision trees, Gaussian process models, ensemble tree methods, support vector machines, and artificial neural networks. The results showed that the Gaussian process model with a Matérn 5/2 kernel achieved the best performance, with an R^2value of 0.97 and RMSE values ranging from 0.084 to 0.088. Despite the high accuracy achieved, the authors noted that reliance on data from a single urban area limits model generalizability and emphasized the need for incorporating meteorological and traffic-related variables in future studies [12].

In another study, Alrasheedi et al. (2025) conducted a systematic review of machine learning applications for air pollution prediction, analyzing studies published between 2016 and 2023. A wide range of methods—including classical algorithms, deep learning models, ensemble techniques, and hybrid approaches—were evaluated. The findings indicated that deep learning models are particularly effective in capturing complex temporal patterns, while ensemble and hybrid methods offer greater robustness in heterogeneous environments. However, challenges such as high computational costs, limited cross-regional generalizability, and data quality constraints were also identified. The authors recommended that future research focus on developing scalable, interpretable, and decision-oriented predictive models [13].

2.2.3. Deep Learning–Based Methods (with Emphasis on LSTM)

As air pollution datasets become increasingly complex due to pronounced temporal and spatial variability, deep learning and hybrid modeling approaches have gained considerable attention for air quality prediction.

In a study by Guo et al. (2024), air quality index prediction and major pollutant identification were performed for nine representative cities in China using a hybrid VMD-CSA-CNN-LSTM framework. Data spanning from January 2015 to November 2021 were used, and the model combined convolutional neural networks and LSTM networks, with hyperparameter optimization carried out using the Chameleon Swarm Algorithm. The model demonstrated high predictive accuracy (RMSE < 7.5, MAPE < 0.1, and adjusted R^2above 96%) and strong generalization capability across different cities. Random Forest analysis further identified PM2.5, PM10, and ozone as dominant contributors to AQI variations in several cities, underscoring the effectiveness of advanced LSTM-based hybrid models for air quality management [14].

In a related long-term study conducted in Pakistan, Mujtaba et al. (2025) evaluated the performance of SARIMA, NAR, and LSTM models for predicting air quality and urban pollutants in Lahore using data from 2003 to 2022. The results indicated that PM2.5 and PM10 were the primary contributors to AQI levels, with the NAR model achieving the lowest RMSE value of 23.52. The study also projected an increase in AQI of up to 16% by 2030, emphasizing the importance of long-term datasets and model integration for improving prediction accuracy and supporting urban planning initiatives [15].

Wu et al. (2024) investigated the application of geographic artificial intelligence and neural networks for predicting air quality and PM2.5 concentrations through a bibliometric analysis of 607 published articles. The study identified key research trends, leading countries, and commonly employed models, concluding that RNN-based and hybrid approaches incorporating spatiotemporal features are among the most effective techniques for air quality prediction. The authors also highlighted challenges such as sensitivity to pollution peaks and the need for multi-source data integration, suggesting directions for future research aimed at enhancing model generalization [16].

Overall, international studies consistently indicate that machine learning and deep learning approaches—particularly hybrid and LSTM-based models—offer significant improvements in air quality prediction accuracy compared with classical methods, especially when dealing with complex and high-dimensional datasets.

2.3. Studies Conducted in Iran and the City of Mashhad

In a study conducted in Iran, Moezan et al. (2024) examined the relationship between air pollution and respiratory conditions among children admitted to the emergency department of Akbar Hospital in Mashhad. This descriptive cross-sectional study included 932 pediatric patients with respiratory problems during the period from December 2017 to December 2019. Air pollutant data—including SO_2, NO_2, CO, PM2.5, PM10, and AQI—were obtained from the Mashhad Environmental Pollution Monitoring Center, while clinical and demographic data were extracted from hospital records. Spearman correlation analysis revealed no statistically significant association between pollutant concentrations and hospital admissions. Nevertheless, the authors emphasized the importance of investigating long-term health effects using larger datasets to inform environmental health policies [17].

In another study, the relationship between urban green spaces, air pollution, and respiratory mortality in Mashhad was analyzed using linear regression and spatial analysis techniques. Pollutant concentration data from 23 monitoring stations were interpolated using GIS software to generate spatial distribution maps, while respiratory mortality data were obtained from the Ferdows Organization of Mashhad. The results indicated that increased urban green space was associated with reduced concentrations of nitrogen dioxide and carbon monoxide, as well as lower respiratory-related mortality rates, although the observed correlations were relatively weak. This study highlights the role of urban planning and green infrastructure in improving air quality and public health outcomes[18].

A recent study conducted in Mashhad examined the combined effects of meteorological factors and air pollution on respiratory mortality during the period from 2017 to 2021. Daily data on major pollutants, including PM2.5, PM10, NO_2, SO_2, and CO, along with meteorological variables such as atmospheric pressure, temperature, humidity, and solar radiation, were analyzed using a Random Forest model. The findings indicated that PM10 and NO_2 exerted the greatest influence on respiratory mortality, while atmospheric pressure also played a significant role. These results underscore the critical interaction between climate variability and air pollution and highlight the need for comprehensive air quality management strategies in densely populated urban areas such as Mashhad [19].

2.4. Summary of the Literature Review and Identification of the Research Gap

In summary, existing studies at both international and national levels demonstrate that classical time-series models, machine learning techniques, and deep learning approaches are capable of predicting air pollution with varying degrees of accuracy. They also confirm that meteorological conditions, pollutant interactions, and urban characteristics can significantly influence air quality and public health outcomes. However, in Iran—and particularly in the city of Mashhad—there remains a lack of long-term, integrated studies that simultaneously incorporate pollutant concentrations, meteorological variables, and localized urban characteristics within a robust predictive framework. This gap highlights the need for developing accurate, data-driven, and location-specific forecasting models to support effective air pollution management and policy formulation.

5. Research Methodology

3.1. Study Area and Data Description

The dataset used in this study comprises daily observations of carbon monoxide (CO) concentrations along with corresponding meteorological variables collected in the city of Mashhad. All data were obtained from the Sajjad urban air quality monitoring station for the period spanning 2017 to 2021. This temporal coverage allows for an in-depth investigation of the behavior of CO concentrations under real urban conditions, capturing both short-term fluctuations and medium-term temporal trends. The data were recorded at a daily temporal resolution and synchronized based on the Gregorian calendar to ensure temporal consistency across all variables. In total, the dataset includes 1,319 daily records.

The primary target variable of this study is the daily average concentration of carbon monoxide measured at the Sajjad monitoring station. In addition to CO, the dataset contains a set of meteorological variables that are known to influence air pollutant dispersion and accumulation. These variables include minimum, maximum, and mean air temperature; relative humidity; precipitation; evaporation; sunshine duration; maximum and mean wind speed; wind direction corresponding to the maximum wind speed; and solar radiation.

In the baseline forecasting framework, model inputs are defined exclusively based on the historical time series of CO concentrations, enabling an assessment of the predictive capability of univariate models. To further investigate the contribution of meteorological factors, a multivariate forecasting configuration was implemented using a Long Short-Term Memory (LSTM) network, in which selected meteorological variables were incorporated as auxiliary input features alongside past CO observations. The complete list of variables used in this study, together with their corresponding field names in the dataset and measurement units, is provided in Table 1.

Table 1. Variables used in the CO concentration prediction model

Variable Name	Dataset Field Name	Suggested Symbol	Unit	Description
Station Name	Station_Name	–	–	Name of the air quality monitoring station
Station Code	Station_Code	–	–	Identification code of the monitoring station
Date	Date	t	–	Date of data recording (Gregorian, daily)
Minimum Temperature	Temp_Min	(T_{min})	°C	Daily minimum air temperature
Maximum Temperature	Temp_Max	(T_{max})	°C	Daily maximum air temperature
Average Temperature	Temp_Avg	(T_{avg})	°C	Daily average air temperature
Relative Humidity	Humidity	H	%	Daily average relative humidity
Precipitation	Precipitation	P	mm	Total daily precipitation
Evaporation	Evaporation	E	mm	Daily evaporation amount
Sunshine Duration	Sunshine	S	h	Duration of sunshine hours
Maximum Wind Speed	WindSpeed_Max	(WS_{max})	m/s	Maximum daily wind speed
Wind Direction	WindDir_Max	WD	°	Wind direction at time of maximum wind speed
Average Wind Speed	WindSpeed_Avg	(WS_{avg})	m/s	Average daily wind speed
Solar Radiation	Radiation	R	W/m^2	Solar radiation intensity
Carbon Monoxide (CO)	CO	CO	mg/m^3	Daily CO concentration (target variable)

3.2. Data Preprocessing

To enhance data quality and reduce the influence of measurement noise and sensor-related errors on the model learning process, a systematic data preprocessing procedure was applied prior to model development. First, unrealistic observations of daily carbon monoxide (CO) concentrations were identified and removed from the dataset. Specifically, CO values lower than 0.3 were considered invalid and excluded from further analysis. This threshold was determined based on an exploratory analysis of the data distribution in conjunction with the technical detection limits of the monitoring sensors, ensuring that near-background or noise-driven measurements did not adversely affect model training.

In the subsequent step, missing values in the CO time series were addressed using linear interpolation based on adjacent temporal observations. This approach preserves the inherent temporal continuity of the dataset and enables effective utilization of available information without introducing artificial trends. Linear interpolation is commonly adopted in environmental time-series studies with daily resolution due to its simplicity and reliability. Following imputation, all observations were temporally aligned according to their corresponding dates and organized into a continuous daily time series.

For models sensitive to differences in variable scales, CO concentration values were normalized to the range [0, 1] using the Min–Max normalization technique prior to training. The processed dataset was then restructured using a sliding window approach to generate supervised input–output pairs for time-series forecasting. The window length was set to $L = 7$days in order to capture weekly cyclic patterns commonly observed in urban air pollution while maintaining a balance between temporal context and model complexity.

3.3 Problem statement

From a problem formulation perspective, short-term forecasting of urban carbon monoxide (CO) concentration is addressed as a one-step-ahead time series prediction task based on historical daily observations. Given the strong temporal dependence of CO concentrations in urban environments—primarily driven by recent pollutant levels and short-term emission dynamics—a univariate input structure is adopted in this study to focus exclusively on the temporal behavior of the CO time series. This design choice allows the models to learn intrinsic temporal patterns while avoiding additional complexity associated with exogenous variables, thereby enabling a fair and consistent comparison across different forecasting approaches.

The forecasting task is formulated using a sliding window strategy with a fixed window length of $L = 7$ days. This window size is selected to capture short-term temporal dependencies and weekly cyclic patterns commonly observed in urban air pollution, while maintaining a sufficient number of effective training samples. The objective is to learn a common nonlinear mapping from the reconstructed input space to the CO concentration of the subsequent day, ensuring methodological consistency across all evaluated models.

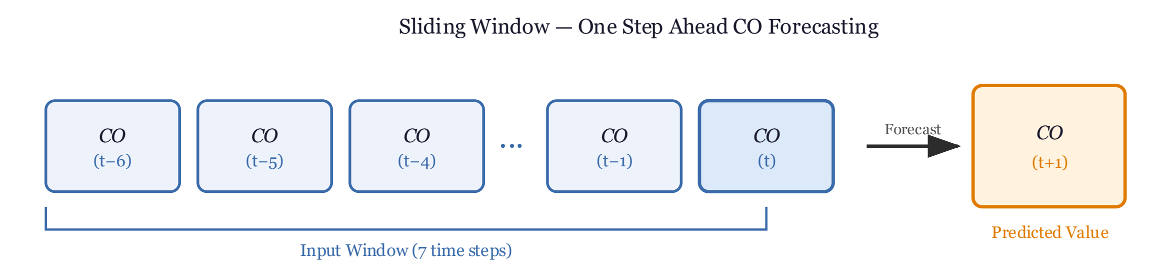

Figure 1. Schematic representation of the sliding window approach used for one-step-ahead CO concentration forecasting, where historical observations from $t - 6$to $t$are used to predict the concentration at time $t + 1$.

Within this framework, the daily CO concentration at time step $t$is denoted as CO_t. The goal of the forecasting models is to estimate the CO concentration at the next time step based solely on historical information. To this end, model inputs are constructed using a sliding time window of length $L = 7$, such that the input vector at time $t$is defined as

$$\mathbf{X}_t = [\mathrm{CO}_{t-L}, \dots, \mathrm{CO}_{t-1}]$$

The output of the forecasting model corresponds to the predicted CO concentration at time step t, denoted as $\widehat{\mathrm{CO}}_t$. Accordingly, the forecasting process can be expressed as a nonlinear mapping between the input feature vector and the model output:

$$\widehat{\mathrm{CO}}_t = f(\mathbf{X}_t)$$

where $f(\cdot)$represents the prediction function implemented by the forecasting models, including classical time series approaches as well as machine learning and deep learning architectures evaluated in this study.

This formulation ensures that, during both the training and evaluation phases, only information available up to the prediction time is utilized. As a result, any potential leakage of future information into the modeling process is effectively prevented, thereby guaranteeing the validity and reliability of the forecasting results.

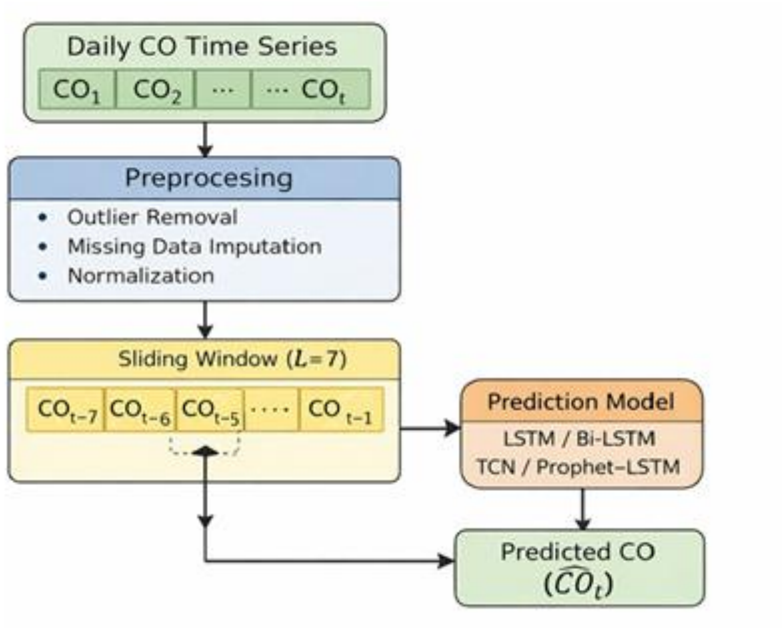

Figure 2. Overview of the short-term forecasting framework for daily urban CO concentration. The framework includes data preprocessing, sliding window reconstruction with L = 7days, model training, and prediction of the next-day CO concentration.

Figure 2 illustrates a schematic overview of the proposed UniTCN-CO methodological framework for short-term forecasting of daily CO concentrations. In this framework, the raw daily CO time series is first subjected to a series of preprocessing steps, including outlier removal, missing data imputation, and normalization. Subsequently, a sliding window reconstruction with a window length of $L = 7$ is applied to generate model input sequences from historical observations, spanning from $\mathrm{CO}_{t-7}$to CO_{t-1}. These input sequences are then provided to the forecasting models, and the corresponding model outputs represent the predicted CO concentration at the next time step, $\widehat{\mathrm{CO}}_t$. The framework is explicitly designed to ensure that both model training and evaluation rely exclusively on past information, thereby preventing any leakage of future data into the prediction process.

3.4. Structure of the Prediction Models

Based on the problem formulation described in the preceding section, the forecasting function $f(\cdot)$ is approximated using a set of classical time-series models alongside advanced deep learning–based approaches. To ensure a fair and consistent comparison of predictive performance, all models examined in this study adopt an identical input structure derived from the sliding time-window methodology.

In total, six forecasting models are evaluated, namely Linear Regression, Prophet, LSTM, Bi-LSTM, TCN, and a hybrid Prophet–LSTM model. Furthermore, to assess the impact of incorporating additional meteorological information on prediction accuracy, the LSTM model is also implemented in a multivariate configuration, in which selected meteorological variables are included as auxiliary input features alongside historical CO concentrations.

The general structure and key configuration aspects of each forecasting model are described in the following subsections, while the comparative performance assessment based on the selected evaluation metrics is presented in the Results section.

3.4.1. LSTM and Bi-LSTM Models

Recurrent Neural Networks (RNNs) are specifically designed for modeling sequential data and time-series processes, as they are capable of capturing temporal dependencies among consecutive observations during the learning process. However, conventional RNN architectures often encounter the vanishing gradient problem when attempting to learn long-term dependencies. To overcome this limitation, Long Short-Term Memory (LSTM) networks and their bidirectional variant (Bi-LSTM) are employed in this study for predicting carbon monoxide concentration. These architectures are well suited for learning both short-term fluctuations and long-term temporal patterns in CO time-series data.

In general, the hidden-state update mechanism in LSTM networks can be expressed as:

$$h_t = \phi(x_t, h_{t-1}), \widehat{CO}_t = g(h_t)$$

where x_t denotes the input vector at time step t, and h_t represents the hidden state of the network. The nonlinear function $\phi(\cdot)$ corresponds to the internal state transition governed by the LSTM gating mechanisms, while $g(\cdot)$ maps the hidden state to the output space associated with the predicted CO concentration.

In the Bi-LSTM architecture, the learning process is performed simultaneously in both forward and backward temporal directions, enabling the model to extract more complex and informative temporal relationships from the input sequence.

In this study, the LSTM and Bi-LSTM architectures consist of one or two recurrent layers, followed by a fully connected output layer to generate the final CO concentration prediction. To mitigate the risk of overfitting, Dropout layers are incorporated into the network structure. Model training is conducted using the Mean Squared Error (MSE) loss function in conjunction with the Adam optimization algorithm. The primary hyperparameters include the number of recurrent layers, the number of memory units, the time-window length, the number of training epochs, and the batch size. The detailed configuration of these hyperparameters is reported in the Results section.

3.4.2. Temporal Convolutional Network (TCN)

Temporal Convolutional Networks (TCNs) are convolution-based architectures specifically designed for modeling sequential and time-series data. Unlike recurrent neural networks, TCNs are capable of capturing long-term temporal dependencies through stacked convolutional layers, without relying on recurrent connections. A key feature of TCNs is the use of causal convolutions, which ensure that the output at each time step depends exclusively on current and past inputs, thereby preventing any leakage of future information into the prediction process.

In general, the output of a temporal convolutional layer can be expressed as:

$$h_t = \sum_{k=0}^{K-1} w_k \; x_{t-d\cdot k}$$

where Kdenotes the convolutional filter size, $d$represents the dilation rate, and $w_k$corresponds to the trainable convolutional filter coefficients. In the implementation adopted in this study, the dilation rate is fixed at $d = 1$, and causal padding is applied to preserve the temporal ordering of the input sequence while maintaining a consistent output length.

The TCN architecture employed in this research consists of two one-dimensional (1D) causal convolutional layers with a filter size of $K = 2$and filter counts of 64 and 32, respectively. To reduce the risk of overfitting and improve model generalization, a Spatial Dropout layer with a dropout rate of 0.2 is incorporated after the first convolutional layer. Subsequently, a Flatten layer is used to map the extracted temporal features into a vector representation, which is then passed to a Dense layer to generate the final prediction of the CO concentration.

Model training is conducted using the Mean Squared Error (MSE) loss function in combination with the Adam optimizer, with a learning rate set to 10^{-3}. In addition, Early Stopping with a patience value of 10 epochs, along with learning-rate reduction strategies, is employed to enhance training stability and convergence efficiency.

3.4.3. Final Mapping and Prediction Process

Across all forecasting models considered in this study, the prediction task is ultimately formulated as a nonlinear mapping from the input feature space to the output space representing the predicted CO concentration. In general, this prediction process can be summarized as follows:

$$\widehat{CO}_t = f(X_t \,|\, \Theta)$$

where $\mathbf{X}_t$denotes the input vector constructed using the sliding time-window approach described in Section 3.3, $f(\cdot)$ represents the nonlinear mapping associated with the forecasting model employed, and Θdenotes the set of trainable model parameters. During the training phase, the parameters $\Theta$are optimized using historical observations. After the convergence of the learning process, the optimized parameters are subsequently used to generate the final predictions of carbon monoxide concentration.

3.4.4. Prophet Model

The Prophet model is an additive time series forecasting approach that explicitly decomposes a time series into trend, seasonality, and residual components. In this study, Prophet is configured using a linear trend component combined with both annual and weekly seasonal patterns to model the temporal behavior of CO concentration. These components are particularly suitable for air pollution data, which often exhibit strong periodic characteristics at multiple temporal scales.

The main configurable parameters of the Prophet model include the number of potential trend changepoints and the forecasting horizon. These parameters are calibrated using the training dataset to ensure an optimal balance between model flexibility and generalization capability. Due to its simplicity, interpretability, and robustness in handling noisy and irregular time series, Prophet is selected as one of the baseline forecasting models in this study.

3.4.5. Prophet–LSTM Hybrid Model

The Prophet–LSTM hybrid model is designed to combine the strengths of statistical time series decomposition and deep learning–based nonlinear modeling. In this approach, the long-term structural components of the CO time series, including trend and seasonal patterns, are first extracted using the Prophet model. The residual series obtained after removing these components represents short-term variations and nonlinear fluctuations that are not captured by Prophet.

These residuals are subsequently used as input to an LSTM network, which is trained to model the remaining complex temporal dependencies. The final forecast is generated by combining the Prophet prediction with the LSTM-based residual forecast. This hybrid strategy enables an effective separation of long-term deterministic structures from short-term stochastic dynamics, thereby improving overall forecasting accuracy and robustness.

3.4.6. Linear Regression Model

Linear regression is employed as a classical baseline model for comparison with more advanced machine learning and deep learning approaches. In this model, a linear relationship is assumed between the historical values of CO concentration and its future value. The forecasting process is formulated as a linear mapping from the input feature space, constructed using the sliding time window, to the output variable.

Although linear regression has limited capability in capturing nonlinear relationships and complex temporal dependencies, its transparency, simplicity, and interpretability make it a suitable benchmark. The performance of this model provides a reference point for evaluating the degree of improvement achieved by more sophisticated architectures.

3.4.7. Architecture Design and Model Justification

The architectural design of the forecasting models in this study is guided by a unified, data-driven framework aimed at ensuring a fair, interpretable, and reliable comparison of predictive performance. To isolate the impact of model architecture from other influencing factors, all deep learning models are trained using an identical univariate input structure and the same sliding window configuration with a fixed length of $L = 7$ days. This consistent experimental setup enables a direct assessment of how different modeling paradigms capture the temporal dynamics of urban CO concentration.

Recurrent neural network architectures, including Long Short-Term Memory (LSTM) and Bidirectional LSTM (Bi-LSTM), are employed to model sequential dependencies through gated memory mechanisms that regulate information flow across time steps. While these models are well suited for learning temporal patterns, their inherently sequential processing may limit training efficiency and constrain the effective modeling of long-range temporal dependencies, particularly in environmental time series with relatively limited sample sizes.

In contrast, the Temporal Convolutional Network (TCN) adopts a fundamentally different architectural strategy based on causal and dilated convolutions. This design enables the construction of hierarchical temporal representations that capture multi-scale temporal dependencies without relying on recurrent connections. By expanding the effective receptive field through dilation, TCNs can efficiently incorporate information from multiple recent time steps while maintaining stable gradient propagation during training. These properties make TCNs particularly well suited for short-term air pollution forecasting, where influential temporal patterns may be distributed across several preceding days rather than confined to immediate past observations.

The adoption of a univariate modeling strategy further emphasizes the intrinsic temporal behavior of CO concentration and reduces the potential influence of noise introduced by auxiliary variables that may suffer from measurement uncertainty or inconsistent availability. Within this unified framework, a diverse set of forecasting approaches—including linear regression, Prophet, recurrent neural networks, convolution-based temporal models, and a hybrid Prophet–LSTM architecture—are systematically evaluated under identical conditions. This architecture-driven comparative analysis provides clear insights into the relative strengths and limitations of each modeling paradigm and establishes a robust justification for the selection of the most suitable model for short-term forecasting of urban CO concentration.

Data-Driven Methodological Framework

The proposed UniTCN-CO is not restricted to the application of a single forecasting model but is built upon a data-driven analytical framework that guides the selection of input structures, the design of network architectures, and the fair comparison of different forecasting approaches. The overarching objective is to develop a stable, reliable, and practically applicable framework for predicting carbon monoxide (CO) concentration in the urban environment of Mashhad.

Univariate versus Multivariate Analysis as a Methodological Decision Step

As an initial step, the LSTM network is implemented and evaluated in both univariate and multivariate configurations. The univariate model relies solely on the historical CO concentration, whereas the multivariate model incorporates additional meteorological variables. The purpose of this comparison is to assess the impact of meteorological information on forecasting accuracy and model stability.

The evaluation results indicate that, for the dataset considered in this study, the univariate LSTM model outperforms its multivariate counterpart based on the selected error metrics. This finding suggests that the temporal dependency structure of CO concentration at the selected monitoring station constitutes the dominant source of predictive information for short-term forecasting. In this specific setup, the inclusion of meteorological variables does not necessarily enhance model performance.

Based on this empirically supported observation, all subsequent forecasting models are implemented using a univariate input structure with a sliding window approach with length $L = 7$ days. This choice reflects a deliberate and data-driven methodological decision rather than an a priori assumption.

Neural Network Architecture Design and Rationale for Layer Choices

In recurrent neural network–based models, the LSTM architecture consists of two consecutive layers with 64 and 32 units, respectively. The first layer is responsible for extracting primary temporal patterns and short-term dependencies from the input sequence, while the second layer captures deeper and more complex nonlinear temporal relationships. This hierarchical design facilitates progressive learning of temporal features across different levels of abstraction.

To mitigate overfitting and improve generalization performance, a Dropout layer with a rate of 0.2 is applied between the recurrent layers. This regularization strategy is particularly effective for environmental time series data, which often contain noise and strong seasonal variability.

For all neural network models, a final Dense layer with a linear activation function is employed to produce the predicted CO concentration. This layer maps the learned feature representation to a continuous output value and effectively serves as the final regression stage. The use of a linear Dense output layer is a standard and well-established practice in time series forecasting tasks involving continuous variables.

The Role of Complementary Architectures in the Proposed Framework

In addition to the LSTM architecture, several complementary models are incorporated to capture different aspects of temporal dynamics. The Bi-LSTM network processes the input sequence in both forward and backward temporal directions, enabling the extraction of more complex temporal patterns. The TCN model, which relies on causal convolutions rather than recurrent connections, preserves temporal causality while offering improved computational efficiency and more stable training behavior.

The Prophet model is employed to explicitly model long-term trend and seasonal components, while the hybrid Prophet–LSTM model integrates statistical decomposition with deep learning–based residual modeling. This combination allows the framework to simultaneously leverage the interpretability of classical methods and the flexibility of neural networks.

Key Strengths of the Proposed Method Compared to Previous Studies

The main strengths of the proposed framework can be summarized as follows:

1. **Data-driven decision-making:** explicit comparison between univariate and multivariate LSTM models, with the final input structure selected based on empirical results.
2. **Unified input framework:** consistent use of the same sliding-window input structure across all models, enabling fair and transparent performance comparison.
3. **Architectural diversity:** integration of classical statistical models, recurrent neural networks, convolutional architectures, and hybrid approaches within a cohesive framework.
4. **Transparent architecture design:** clear specification of network depth, number of neurons, and the functional role of each layer.
5. **Interpretability and practical applicability:** maintaining a balance between predictive accuracy and usability in real-world urban air quality monitoring systems.

To provide a clear visual illustration of the neural network architectures employed in this study, Figure 3 presents the structure of the LSTM model as a representative example. Since the remaining deep learning models share a similar input configuration based on the sliding window and a common mapping to a single predicted CO value, the LSTM architecture sufficiently represents the general modeling structure. A quantitative evaluation of model performance and a detailed numerical comparison based on error metrics are presented in the Results section.

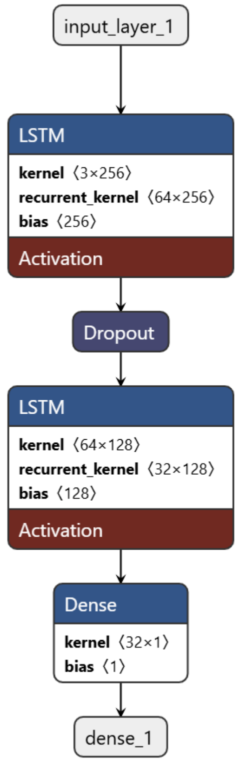

Figure 3. Layer-by-layer architecture of the LSTM network used for daily CO concentration prediction, illustrating the sequential processing of the sliding-window input and the generation of the short-term forecasting.

3.5. Data Splitting and Validation

Given the time-series nature of the dataset, the data were split chronologically to prevent information leakage from future observations into the training process. Accordingly, the first 80% of the observations were used for model training, while the remaining 20% were reserved for testing. Specifically, data collected between 2017 and 2019 were employed for training the models, whereas observations from 2020 to 2021 were used to evaluate their predictive performance.

For all models, the temporal order of the data was strictly preserved, and any form of random shuffling was intentionally avoided. This strategy ensures that the evaluation procedure accurately reflects realistic forecasting conditions in which future values are predicted solely based on past information.

3.6. Performance Evaluation Metrics

To evaluate and compare the forecasting performance of the proposed models, three widely used error metrics in regression and time-series forecasting were employed. Each metric captures a different aspect of prediction accuracy and model behavior.

The Mean Absolute Error (MAE) quantifies the average magnitude of the prediction errors and is relatively insensitive to outliers. The Root Mean Squared Error (RMSE) assigns greater weight to larger errors and is therefore particularly effective for assessing model performance under high or extreme CO concentration levels. In addition, the Symmetric Mean Absolute Percentage Error (SMAPE) is used as a normalized metric, enabling model performance to be compared on a percentage basis.

The combined use of MAE, RMSE, and SMAPE provides a comprehensive and reliable evaluation framework for assessing both accuracy and robustness across different forecasting models.

6. Results

In this section, the performance of the forecasting framework described in the methodology is evaluated using the daily carbon monoxide (CO) concentration dataset collected from the Sajjad monitoring station in Mashhad. The primary objective is to assess the capability of the proposed models to perform short-term forecasting of CO concentration and to conduct a fair comparison under a unified input structure based on the sliding time-window approach.

First, the characteristics of the dataset and the data splitting strategy are described. Subsequently, the quantitative results and visual comparisons of the models are presented and analyzed using the MAE, RMSE, and SMAPE evaluation metrics.

4.1. Dataset and Its Characteristics

The dataset used in this study comprises daily measurements of carbon monoxide (CO) concentration recorded at the Sajjad air quality monitoring station in Mashhad, together with the corresponding meteorological variables, over the period from 2017 to 2021. Prior to model implementation, the data were prepared through a series of preprocessing steps, including temporal synchronization, removal of unrealistic values, completion of missing observations using linear interpolation, and normalization for models sensitive to input scale.

The target variable is the daily CO concentration measured in mg/m^3. For the primary comparison of forecasting models, the univariate formulation—based exclusively on the historical CO time series—is adopted.

To generate samples suitable for time-series forecasting models, the data were reorganized using a sliding time window with a length of $L = 7$. Accordingly, to predict the CO concentration at time t, denoted as CO_t, the model input is defined as:

$$\mathbf{X}_t = [CO_{t-7}, \dots, CO_{t-1}]$$

The selection of the window length $L = 7$is motivated by the intention to capture weekly patterns associated with urban traffic activities, which represent one of the dominant contributors to daily variations in carbon monoxide concentration. Preliminary experiments showed that increasing the window length beyond this value did not result in a significant improvement in forecasting accuracy and, in some cases, led to increased model complexity and overfitting.

Given the total number of observations $N = 1319$, the number of usable samples after applying the sliding window is calculated as:

$$N' = N - L = 1319 - 7 = 1312$$

The data splitting procedure was performed in a strictly sequential manner, without any random shuffling, to prevent future information leakage. Following standard practice in time-series forecasting, 80% of the earliest samples were allocated to the training and validation phase, while the remaining 20% were reserved for the test phase. The validation set was extracted from the end of the training portion, again in chronological order, to ensure that all evaluation stages remain consistent with real-world forecasting conditions.

Table 2. Dataset characteristics and data splitting configuration

Feature	Value
Dataset name	Daily Carbon Monoxide (CO) Monitoring Dataset of Mashhad – Sajad Station
Time period	2017–2021
Temporal resolution	Daily
Total number of raw observations	1,319 samples
Problem type	Time series forecasting (regression)
Forecasting horizon	Short-term
Time window length	L = 7 days
Data splitting method	Sequential time-based split (no shuffling)
Training / testing ratio	80% training – 20% testing
Number of training data (raw)	1,055 samples
Number of testing data (raw)	264 samples
Number of training samples after windowing	1,048 samples
Number of testing samples after windowing	257 samples
Validation set	No independent validation dataset
Validation strategy	Use of internal validation_split from the training data in deep learning models
Data leakage considerations	Preservation of temporal order; scaler fitted only on training data

4.2. Quantitative Results

In this section, the quantitative forecasting performance of the proposed prediction models for daily carbon monoxide (CO) concentration is evaluated using the Mashhad dataset. To ensure a fair and consistent comparison, all models were trained and tested using an identical input formulation based on a sliding time window of length $L = 7$ days. Model performance was assessed on the test set using the Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and Symmetric Mean Absolute Percentage Error (SMAPE) metrics.

Table 3 summarizes the quantitative results obtained from implementing Linear Regression, Prophet, LSTM, Bi-LSTM, TCN, and the hybrid Prophet–LSTM model. The results indicate that the deep learning–based models exhibit comparable levels of accuracy, with relatively small differences observed in their MAE and RMSE values. This suggests that all neural architectures are capable of capturing the dominant temporal patterns in the CO time series.

Among these models, the TCN does not exhibit a pronounced numerical superiority over the other deep learning approaches; however, it consistently delivers stable and competitive performance across all three-evaluation metrics. In particular, its uniform behavior in terms of SMAPE, together with the residual analysis presented in Section 4.3, suggests a balanced predictive behavior and the absence of systematic bias in its forecasts.

Accordingly, the selection of the TCN as the preferred model in this study is motivated not solely by absolute error minimization, but by a combination of acceptable predictive accuracy, robustness across different metrics, behavioral stability, and structural simplicity. These characteristics make the TCN a reliable and practical choice for short-term forecasting of urban CO concentration.

Table 3. Comparison of the Quantitative Performance of Different Models in Forecasting CO Concentration (Test Set)

Model	MAE (mg/m³)	RMSE (mg/m³)	SMAPE (٪)
Linear Regression	0.0675	0.0981	5.81%
Prophet	0.0794	0.1065	6.84%
LSTM	0.0676	0.0990	5.81%
Bi-LSTM	0.0663	0.0981	5.70%
TCN	0.0675	0.0986	5.81%
Prophet–LSTM	0.0699	0.0976	6.02%

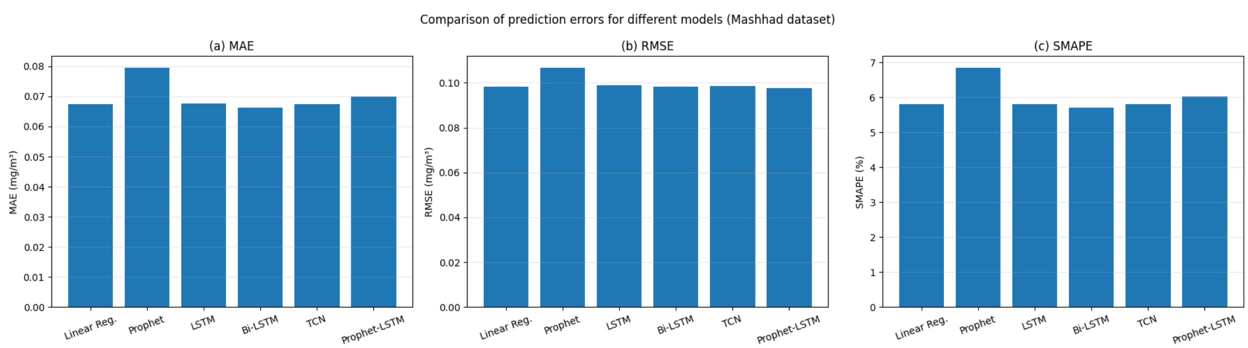

Figure 4. Comparative visualization of error metrics (MAE, RMSE, and SMAPE) for different prediction models applied to the Mashhad daily CO dataset.

To facilitate a visual comparison of the forecasting performance of the different prediction models, Figure 4 illustrates the MAE, RMSE, and SMAPE values reported in Table 3 for each model. As shown, deep learning–based models generally achieve lower error values compared to classical approaches, indicating their superior capability in capturing the nonlinear temporal dynamics of the CO concentration time series.

Among these models, the TCN demonstrates stable and competitive performance across all three-evaluation metrics, which is consistent with the numerical results presented in Table 3. This visual representation enables a more intuitive and rapid comparison of model performance and further corroborates the quantitative findings discussed in the previous subsection. It should be emphasized that Figure 4 serves solely as a graphical representation of the numerical values reported in Table 3 and does not introduce any additional evaluation criteria.

4.3. Visual Analysis of the Selected Model's Performance (TCN)

In addition to the comparative analysis of quantitative error metrics, a visual assessment was conducted to further examine the predictive behavior, stability, and potential bias of the selected model. To this end, Figure 5 presents a comparison between the observed and predicted daily carbon monoxide (CO) concentrations over the entire test period. In this figure, the horizontal axis represents time (in days), while the vertical axis denotes the CO concentration measured in mg/m^3.

As illustrated in Figure 5, the TCN model successfully captures the overall trend of the CO time series and accurately reproduces a substantial portion of the short-term fluctuations. The close overlap between the observed and predicted curves indicates that the model has learned the underlying temporal structure of the data effectively and demonstrates stable generalization performance when applied to unseen observations.

Nevertheless, during periods characterized by abrupt and pronounced spikes in CO concentration, the model tends to produce slightly conservative estimates, predicting peak values that are marginally lower than the observed maxima. This behavior, which is commonly observed in deep learning–based time series forecasting models, reflects an inherent trade-off between sensitivity to extreme events and overall prediction stability. By smoothing highly irregular variations, the model reduces the influence of random noise while maintaining robust performance across the full range of observations.

Overall, the visual analysis presented in this subsection is fully consistent with the low MAE, RMSE, and SMAPE values reported in the previous section. These findings confirm the suitability and reliability of the TCN model for short-term forecasting of urban CO concentration under real-world conditions.

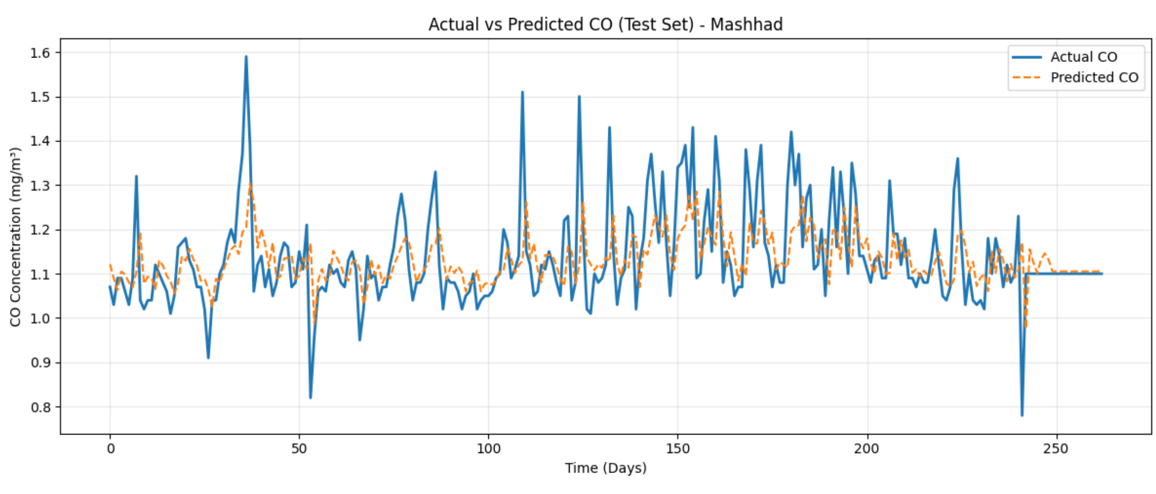

Figure 5. Comparison between observed and predicted daily CO concentrations on the test dataset, illustrating the temporal forecasting performance of the evaluated models.

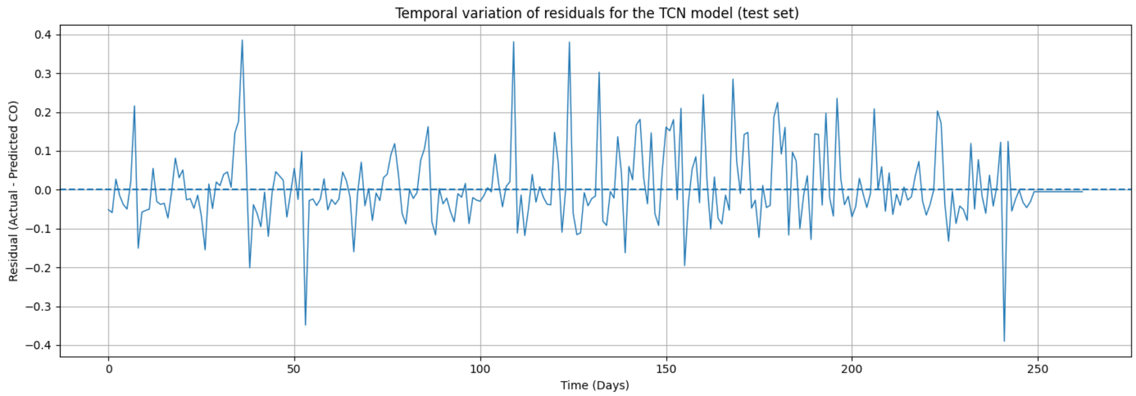

Figure 6. Temporal distribution of short-term forecasting residuals (Actual − Predicted CO) obtained by the TCN model on the independent test dataset.

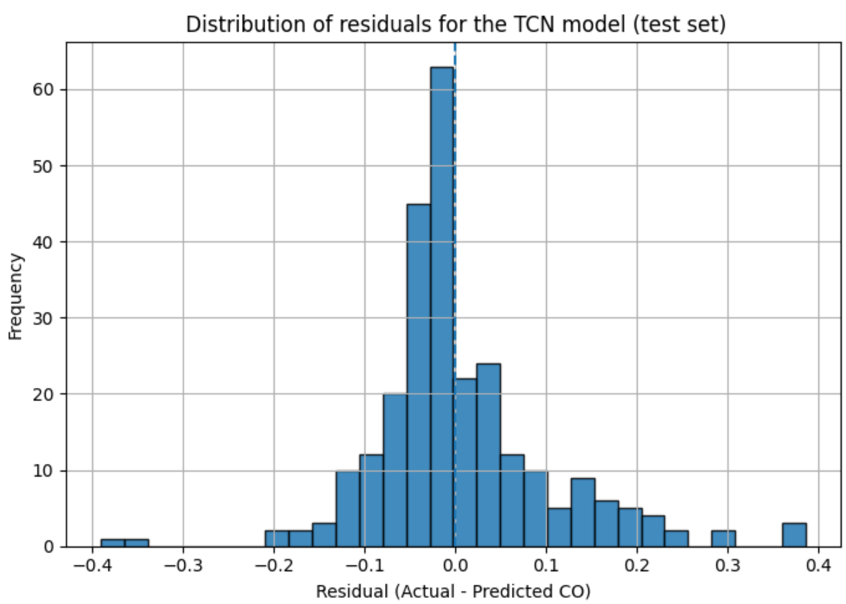

Figure 7. Histogram of residuals obtained from the UniTCN-CO model on the test dataset, illustrating the distribution of prediction errors around zero and indicating the absence of systematic bias.

To further examine the behavior of prediction errors and to verify the absence of systematic bias in the selected model, a residual analysis of the TCN was conducted on the test set. As illustrated in Figure 6, the residuals fluctuate predominantly around zero and do not exhibit any evident temporal structure, periodicity, or persistent increasing or decreasing trend over time. This behavior indicates that the TCN model does not suffer from overfitting or structural bias and has effectively captured the underlying dynamics of the CO concentration time series.

In addition, the statistical distribution of the residuals presented in Figure 7 shows that the prediction errors are approximately symmetrically distributed around zero, with the majority of residuals concentrated within a relatively narrow range. Although a limited number of larger errors can be observed during specific periods—primarily associated with abrupt peaks in CO concentration—the overall residual distribution remains well-behaved. Such deviations are expected in real-world air quality time series and do not undermine the general stability of the model.

Collectively, these findings confirm the robustness and generalization capability of the TCN model under realistic urban conditions. While the numerical differences in performance metrics among the evaluated models are relatively small, the residual analysis demonstrates that the TCN provides consistent and reliable predictive behavior from both temporal and statistical perspectives.

4.4. Model Comparison

To contextualize the results of the present study within the existing body of literature and to assess the relative position of the proposed approach, this subsection compares the performance of the selected model with findings reported in several related studies. Table 4 provides a summary of the datasets used, the forecasting methods applied, the evaluation metrics considered, and the reported performance outcomes in these studies.

Table 4. Comparison of the proposed study with selected previous works on CO concentration forecasting.

Study	**Year**	**Location**	**Models**	**Best Metric**
[20]	2025	response signals of electronic nose (E-nose)	VMD-LSTNet-Attention	R^2
[21]	2025	Fonollosa et al. E-nose dataset	PMH-TCN	RMSE, MAE , R^2
[22]	2024	Central Pollution Control Board (CPCB) of India	CNN–ARIMA (TBS)	MAE, RMSE, R^2, MAPE
UniTCN CO	2026	Mashhad, Iran (Sajjad air quality monitoring station)	UniTCN CO	RMSE, MAE , R^2

Beyond the quantitative error metrics, a qualitative comparison of the evaluated models reveals distinct behavioral differences. Linear Regression, as expected, exhibits limited capability in capturing nonlinear temporal dependencies inherent in urban air pollution time series. Prophet demonstrates acceptable performance in modeling long-term trends and seasonal components; however, its accuracy deteriorates during abrupt concentration changes, indicating limited adaptability to short-term fluctuations. LSTM and Bi-LSTM models improve predictive accuracy by learning temporal dependencies, yet their performance shows higher variability across the test period, suggesting sensitivity to local fluctuations and potential overfitting. In contrast, the TCN model delivers more stable predictions with consistently balanced error values. This robustness can be attributed to its causal convolution structure, which effectively captures short-term temporal patterns while maintaining stable gradient propagation, making it particularly suitable for daily CO concentration forecasting.

It is important to note that due to inherent differences in dataset characteristics, temporal coverage, input variables, data preprocessing procedures, and evaluation protocols, this comparison is intended to be qualitative rather than strictly quantitative. Consequently, the reported results should be interpreted as a comparative overview rather than a direct numerical benchmark.

Table 5. Comparison of the proposed method's performance with selected previous studies in the field of air pollution forecasting

Reference	Region/Dataset	Pollutant	Method / Model	Evaluation Metrics	Reported Result
Gan et al. (2025) [20]	Laboratory E-nose dataset (530 samples)	CO	VMD–LSTNet–Attention	MAE, RMSE, R^2	R^2 = 0.993, MAE = 1.330, RMSE = 2.414
Zhuo et al. (2025) [21]	Fonollosa et al. E-nose dataset (C_2H_4–CO gas mixtures, laboratory experimental data)	CO	PMH-TCN	R^2, MAE, RMSE	R^2 = 0.9850, MAE = 0.0448, RMSE = 0.0651
Jayaraman et al. (2025) [22]	CPCB India air quality dataset (22 major cities in India)	CO, PM2.5, PM10, NO_2, SO_2, O_3, NH_3	CNN–ARIMA (TBS)	MAE, RMSE, R^2, MAPE	R^2 = 0.81, MAE = 31.73, RMSE = 44.60
Kırelli (2025) [23]	Beşiktaş, Istanbul air quality monitoring data (2018–2023)	CO	GRU	MAE, RMSE, MAPE, R^2	RMSE = 123.12, MAE = 95.33
UniTCN-CO	Mashhad, Iran (Sajjad air quality monitoring station, daily CO data)	CO	TCN (proposed model)	MAE, RMSE, SMAPE	MAE = 0.0675, RMSE = 0.0986, SMAPE = 5.81%

As shown in Table 5, the reviewed studies exhibit substantial differences in terms of dataset characteristics, spatial scale, data acquisition conditions, and modeling strategies. Some investigations, such as those conducted by Gan et al. [20] and Zhuo et al. [21], relied on laboratory-based gas sensor (E-nose) data collected under controlled environmental conditions, where pollutant variability is relatively limited and measurement noise is minimized. In contrast, studies such as Jayaraman et al. [22] and Kırelli [23] employed real-world urban air quality monitoring data obtained from multiple stations, which inherently involve higher levels of complexity and uncertainty due to the combined effects of traffic emissions, meteorological variability, and urban morphology.

Consequently, direct numerical comparisons of reported error values across these studies should be interpreted with caution, as differences in dataset composition and experimental design significantly influence model performance. Within this context, the present study—based on daily urban air quality measurements collected at the Sajad monitoring station in Mashhad and employing the Temporal Convolutional Network (TCN) as the final selected model—demonstrates competitive and practically acceptable performance under realistic urban conditions.

The results indicate that the TCN model, despite its non-recurrent architecture and relatively simple structural design, is capable of effectively learning temporal dependencies in carbon monoxide concentration data and achieves a limited prediction error on the test set. These findings suggest that the proposed framework not only attains adequate predictive accuracy but also offers strong potential for deployment in urban air quality forecasting and management systems, owing to its stability, computational efficiency, and applicability to real-world monitoring data. As such, it may serve as a supportive tool for environmental assessment and decision-making processes.

4.5. Discussion

Despite the overall satisfactory performance of the proposed models, several important limitations must be acknowledged. First, the analysis relies on data from a single air quality monitoring station, which restricts the spatial representativeness of the findings and may not fully reflect pollutant dynamics across the broader metropolitan area of Mashhad. Second, all evaluated models exhibit reduced accuracy during sudden concentration spikes, indicating ongoing challenges in capturing extreme short-term variability in CO levels. Moreover, the use of daily aggregated data may smooth intra-day fluctuations, potentially leading to the loss of valuable fine-scale temporal information. These considerations highlight the need for caution when generalizing the results and suggest that future research should incorporate multi-station measurements and higher temporal resolution data to enhance model robustness and predictive reliability.

In this study, the proposed forecasting framework was assessed using real-world daily carbon monoxide observations collected in Mashhad, and the performance of several prediction models was evaluated under a unified input structure. The quantitative results show that deep learning models generally achieve lower MAE, RMSE, and SMAPE values compared to classical time series methods, demonstrating their stronger capability in learning the temporal dependencies and variability inherent in CO concentration dynamics.

Although the numerical differences among the deep learning models are relatively modest, the Temporal Convolutional Network (TCN) consistently demonstrates stable and reliable behavior across all three evaluation metrics on the test dataset. Visual examination of the observed and predicted time series, accompanied by residual analysis, confirms that the selected model effectively captures the dominant temporal trends and a substantial portion of short-term fluctuations, while maintaining prediction errors predominantly centered around zero.

Furthermore, comparison with selected previous studies indicates that the proposed framework—despite being applied to real, noisy, and highly variable urban air quality data—achieves performance levels comparable to state-of-the-art approaches reported in the literature. Overall, the findings of this chapter suggest that combining a sliding-window input structure with deep learning architectures, particularly the TCN, provides a robust and reliable foundation for developing short-term CO concentration forecasting systems in urban environments.

7. Conclusions and Recommendations

This study proposed and evaluated a data-driven framework for forecasting daily carbon monoxide (CO) concentrations in the city of Mashhad using real-world air quality monitoring data collected at the Sajad station. By relying on a univariate input structure based solely on short-term historical CO observations and applying an identical input definition across all models, the framework enabled a fair, transparent, and systematic comparison of classical time series methods and deep learning–based approaches under realistic urban conditions. The results demonstrate that short-term temporal information contains sufficient predictive value for short-term CO forecasting, even in the presence of noise and complex urban dynamics.

A central finding of this research is the demonstrated effectiveness of deep learning models in modeling daily CO concentration dynamics in real urban environments. Despite the inherent variability and uncertainty of urban air quality data, these models were able to capture a substantial portion of the temporal behavior of the pollutant using relatively simple inputs. Among the evaluated methods, the Temporal Convolutional Network (TCN) emerged as the most stable model, exhibiting consistent performance across multiple evaluation metrics as well as favorable behavioral characteristics in visual and residual analyses. The strong performance of the TCN highlights the potential of non-recurrent convolutional architectures as efficient alternatives to traditional recurrent models for short-term air pollution forecasting tasks. From a practical standpoint, the simpler structure of TCN reduces computational overhead and implementation complexity, making it particularly attractive for operational applications.

The outcomes of this study also carry important practical implications. The predictions generated by the selected model can support early warning systems for urban air pollution, enabling timely communication of potential risk conditions to vulnerable populations and assisting authorities in planning short-term mitigation actions. Moreover, the proposed framework could be integrated with traffic management and urban planning systems to assess the short-term impacts of traffic control strategies on CO concentration levels. At a broader level, reliable short-term forecasts of key air pollutants can serve as an effective decision-support tool for urban environmental management and policy formulation.

Despite its contributions, this study has several limitations that should be acknowledged. The analysis was based on data from a single monitoring station and focused exclusively on carbon monoxide, which restricts the direct generalization of the results to other locations within the city and to other pollutants. In addition, the use of daily data limits the ability to capture intraday variability and short-duration pollution peaks, while the univariate modeling structure does not explicitly account for the influence of meteorological conditions, traffic patterns, or other external drivers. These factors may affect the model's capacity to reproduce extreme or anomalous pollution events.

Based on these considerations, several directions for future research are recommended. Extending the framework to a multi-station and multi-pollutant setting would allow for a more comprehensive assessment of urban air quality patterns and facilitate the analysis of interactions among different pollutants. Employing higher temporal resolution data, such as hourly measurements, and systematically incorporating meteorological and traffic-related variables into multivariate models could improve the prediction of short-term peaks and acute pollution episodes. Furthermore, exploring more advanced hybrid and spatiotemporal architectures—such as combining TCN with attention mechanisms or spatial dependency modeling—may lead to forecasting systems that not only achieve higher accuracy but also provide enhanced interpretability and practical applicability for urban air quality management.

8. References

[1] D. Vilcins *et al.*, "Updates in Air Pollution: Current Research and Future Challenges.," *Ann. Glob. Heal.*, vol. 90, no. 1, p. 9, 2024, doi: 10.5334/aogh.4363.

[2] S. Bedi, K. Tiwari, P. A. P., S. H. Kota, and N. M. A. Krishnan, "A neural operator for forecasting carbon monoxide evolution in cities," *npj Clean Air*, vol. 1, no. 1, p. 2, 2025, doi: 10.1038/s44407-024-00002-5.

[3] K. Chatterjee *et al.*, "Ecosense: a revolution in urban air quality forecasting for smart cities," *BMC Res. Notes*, vol. 18, no. 1, pp. 1–17, 2025, doi: 10.1186/s13104-025-07099-1.

[4] Y. Wang, F. Xia, L. Yao, S. Zhao, Y. Li, and Y. Cai, "An enhanced machine learning model for urban air quality forecasting under intense human activities," *Urban Clim.*, vol. 60, p. 102359, 2025, doi: https://doi.org/10.1016/j.uclim.2025.102359.

[5] I. Essamlali, H. Nhaila, and M. El Khaili, "Supervised Machine Learning Approaches for Predicting Key Pollutants and for the Sustainable Enhancement of Urban Air Quality: A Systematic Review," *Sustain.*, vol. 16, no. 3, 2024, doi: 10.3390/su16030976.

[6] F. Ahmadian, S. Rajabi, S. Maleky, and M. A. Baghapour, "Spatiotemporal analysis of airborne pollutants and health risks in Mashhad metropolis: enhanced insights through sensitivity analysis and machine learning," *Environ. Geochem. Health*, vol. 47, no. 2, p. 34, 2024, doi: 10.1007/s10653-024-02332-5.

[7] A. Mathew *et al.*, "Unveiling urban air quality dynamics during COVID-19: a Sentinel-5P TROPOMI hotspot analysis," *Sci. Rep.*, vol. 14, no. 1, pp. 1–20, 2024, doi: 10.1038/s41598-024-72276-4.

[8] T. Salthammer, "Carbon monoxide as an indicator of indoor air quality," *Environ. Sci. Atmos.*, vol. 4, no. 3, pp. 291–305, 2024, doi: 10.1039/d4ea00006d.

[9] T. Bunnag, "Forecasting PM10 Caused by Bangkok's Leading Greenhouse Gas Emission Using the SARIMA and SARIMA-GARCH Model," *Int. J. Energy Econ. Policy*, vol. 14, no. 1, pp. 418–426, 2024, doi: 10.32479/ijeep.15275.

[10] S. G. Gocheva-Ilieva, A. V. Ivanov, D. S. Voynikova, and D. T. Boyadzhiev, "Time series analysis and forecasting for air pollution in small urban area: An SARIMA and factor analysis approach," *Stoch. Environ. Res. Risk Assess.*, vol. 28, no. 4, pp. 1045–1060, 2014, doi: 10.1007/s00477-013-0800-4.

[11] M. Ansari and M. Alam, "An Intelligent IoT-Cloud-Based Air Pollution Forecasting Model Using Univariate Time-Series Analysis," *Arab. J. Sci. Eng.*, vol. 49, no. 3, pp. 3135–3162, 2024, doi: 10.1007/s13369-023-07876-9.

[12] M. A. Almubaidin *et al.*, "Machine learning predictions for carbon monoxide levels in urban environments," *Results Eng.*, vol. 22, p. 102114, 2024, doi: https://doi.org/10.1016/j.rineng.2024.102114.

[13] N. D. N. Alrasheedi, N. Masseran, and R. R. M. Tajuddin, "A Systematic Literature Review on the Estimation of High Air Pollution Periods Using Machine Learning Approaches," *Contemp. Math.*, vol. 6, no. 3, pp. 3184–3208, 2025, doi: 10.37256/cm.6320255760.

[14] Z. Guo *et al.*, "Optimized air quality management based on air quality index prediction and air pollutants identification in representative cities in China," *Sci. Rep.*, vol. 14, no. 1, pp. 1–9, 2024, doi: 10.1038/s41598-024-68972-w.

[15] M. A. Mujtaba *et al.*, “Using machine learning for air quality prediction and sustainable urban planning,” *Sustain. Futur.*, vol. 10, p. 100981, 2025, doi: https://doi.org/10.1016/j.sftr.2025.100981.

[16] C. Wu, S. Lu, J. Tian, L. Yin, L. Wang, and W. Zheng, “Current Situation and Prospect of Geospatial AI in Air Pollution Prediction,” *Atmosphere (Basel).*, vol. 15, no. 12, 2024, doi: 10.3390/atmos15121411.

[17] N. Moazzen, A. Memari, and N. Todarbary, “The temporal-spatial association of respiratory manifestations and air pollution in children referred to the Emergency Department of Akbar Children’s Hospital, Mashhad, Iran,” *Egypt. Pediatr. Assoc. Gaz.*, vol. 72, no. 1, 2024, doi: 10.1186/s43054-024-00301-2.

[18] M. R. Rahnama and marzieh sabouri, “Space Analysis of the Effect of Urban Green Area on Air Pollution and Respiratory Death Using Linear Regression in City of Mashhad,” *J. Arid Reg. Geogr. Stud.*, vol. 14, no. 54, pp. 75–88, 2024, doi: 10.22034/jargs.2023.392295.1020.

[19] Y. Dowlatabadi, S. abadi, M. Sarkhosh, M. Mohammadi, and S. M. M. Moezzi, “Assessing the impact of meteorological factors and air pollution on respiratory disease mortality rates: a random forest model analysis (2017–2021),” *Sci. Rep.*, vol. 14, no. 1, pp. 1–12, 2024, doi: 10.1038/s41598-024-74440-2.

[20] W. Gan, R. Ma, W. Zhao, X. Peng, H. Cui, and J. Yan, “Sensors and Actuators : B . Chemical A VMD-LSTNet-Attention model for concentration prediction of mixed gases,” vol. 422, no. April 2024, 2025.

[21] J. Zhuo *et al.*, “Rapid and high-accuracy concentration prediction of gas mixtures based on,” *Measurement*, vol. 242, no. PC, p. 116003, 2025, doi: 10.1016/j.measurement.2024.116003.

[22] S. Jayaraman, T. Nathezhtha, S. Abirami, and G. Sakthivel, “Enhancing urban air quality prediction using time-based-spatial forecasting framework,” pp. 1–15, 2025.

[23] Y. Kı, “Air Quality Forecasting in Urban Environments : A Deep Learning Approach,” vol. 13, pp. 1445–1454, 2025, doi: 10.29130/dubited.1591784.