1D Convolutional Neural Network for Air Quality Forecasting

1. Introduction and Problem Statement

The objective of this study is to develop a deep learning model for high-precision **Air Quality forecasting** using multivariate time series data. A **1D Convolutional Neural Network (1D-CNN)** architecture was selected to automatically extract relevant temporal features from sequential sensor readings, enabling accurate prediction of a target air quality metric (a univariate regression task). The final model performance was evaluated using standard regression metrics.

---

2. Methodology: Data Preparation and Preprocessing

2.1. Dataset Characteristics and Preprocessing

The model was trained on a time-series dataset that was loaded from a cleaned CSV file.

Parameter	Detail
Data Type	Multivariate Time Series
Feature Count (Nfeatures)	18 independent input variables (features).
Input Sequence Length (Lseq)	24 time steps (e.g., the previous 24 hours of data).
Output Type	Univariate Regression (prediction of a single future air quality value).
Total Training Sequences	34,425 samples.

he raw data was preprocessed by applying a **scaling or normalization technique** (details on the specific method are not explicit in the code, but normalization is confirmed by the use of `inverse_scale` for final evaluation). The data was then transformed into 3D sequences of shape **(Samples, Lseq, Nfeatures)** using a custom `create_sequences` function, preparing it for the 1D-CNN input.

2.2. Hardware and Environment

Model training was conducted on an accelerated computing environment utilizing an **NVIDIA T4 GPU**.

---

3. Model Architecture

The prediction model utilizes a sequential 1D-CNN architecture designed for time-series feature extraction and subsequent regression. The total number of trainable parameters in the model is **9,985**.

Layer Type	Filters/Units	Kernel/Pool Size	Output Shape	Parameters	Activation (Assumed)
Input Layer	N/A	N/A	(None, 24, 18)	N/A	N/A
Conv1D	**64**	Presumed **3**	(None, 22, 64)	3,520	ReLU (Standard for Conv layers)
MaxPooling1D	N/A	Presumed **2**	(None, 11, 64)	0	N/A
Dropout	N/A	N/A	(None, 11, 64)	0	N/A
Conv1D_1	**32**	Presumed **3**	(None, 9, 32)	6,176	ReLU (Standard for Conv layers)
Flatten	N/A	N/A	(None, 288)	0	N/A
Dense (Output)	**1**	N/A	(None, 1)	289	Linear (Standard for Regression)
Total				**9,985**	

*Note: The **Kernel Size** for Conv1D layers and **Pool Size/Strides** for MaxPooling1D are inferred as 3 and 2, respectively, based on the input sequence length (24) and the resulting output sequence lengths (22 and 11), assuming a default `padding='valid'` and `strides=1` (except for MaxPooling, where `strides` is typically equal to `pool_size` for downsampling).*

---

4. Training Methodology

4.1. Compilation and Training Parameters

The model was compiled and trained for a regression task with the following settings:

Parameter	Detail
Loss Function	**Mean Squared Error (MSE)** (Inferred as the standard loss for deep learning regression).
Optimizer	**Adam** (Inferred as the common default optimizer for deep learning).
Metrics	**Mean Absolute Error (MAE)**.

Parameter	Detail
Max Epochs	**50**.
Steps per Epoch	1,076 (Reflects the total batch size/training set size ratio).
Validation	A hold-out validation set was used to monitor performance during training.

4.2. Training Log Excerpt

The training log shows significant convergence across the initial epochs. The process stopped at Epoch 24 (out of 50 maximum set epochs), suggesting that an **Early Stopping** callback was likely implemented to prevent overfitting.

Epoch	Loss	MAE	Val. Loss	Val. MAE
1	0.0203	0.0952	0.0081	0.0635
10	0.0039	0.0381	0.0028	0.0336
20	0.0031	0.0332	0.0023	0.0307
24	0.0029	0.0318	0.0021	0.0299

5. Evaluation and Results

5.1. Evaluation Metrics

Model performance was rigorously assessed on both the validation and test sets using three primary metrics:

1. **Mean Absolute Error (MAE):** Measures the average magnitude of the errors.
2. **Root Mean Squared Error (RMSE):** Measures the standard deviation of the residuals (prediction errors).
3. **Coefficient of Determination (R^2 Score):** Indicates the proportion of the variance in the dependent variable that is predictable from the independent variables (a measure of model fit).

5.2. Final Performance Results

The final results were computed on the inverse-scaled predictions and true values, representing the performance in the original data unit scale.

```
📌 Validation Evaluation
MAE  = 13.9862
RMSE = 20.9046
R²   = 0.9079
--------------------------------------
```

```
📌 Test Evaluation
MAE  = 12.4748
RMSE = 16.9847
R²   = 0.8809
--------------------------------------
```

The R2 values on both the validation and test sets (0.9079 and 0.8809, respectively) indicate that the 1D-CNN model achieved a **good fit** and **generalization ability** for the air quality prediction task. The MAE values are minimal, demonstrating high predictive accuracy.