1. Introduction and Methodology

The project develops a **Time Series Regression Model** for single-step (one-hour ahead) forecasting of the **Air Quality Index (AQI)** using the attention-based **Transformer Encoder** architecture.

Data Source and Preprocessing

Component	Detail	Source
Dataset File	/content/final_cleaned_interpolated.csv	
Data Type	Hourly Multivariate Time Series	
Total Samples	43,062 entries	
Target Variable ($\mathbf{y}$)	**AQI** (Air Quality Index)	
Scaling Method	**MinMaxScaler** was applied separately to the input features (X) and the target variable (y).	

Input Feature Engineering (X)

The model utilizes **18 input features** derived from pollution and meteorological measurements, along with engineered temporal features. The feature dimension is D=18.

- **Air Pollutants & Meteorological Variables (12 features):**
    - Pollutants: pm2_5_ugm3, pm10_ugm3, co_ppm, no2_ppb, o3_ppm, so2_ppb
    - Meteorology: temperature_2m, relative_humidity_2m, surface_pressure, precipitation, cloudcover, windspeed_10m
- **Temporal Features (6 features):**
    - hour, month, dayofweek (although dayofweek is added, only hour and month are used in cyclic encoding)
    - **Cyclic Encoding:** hour_sin, hour_cos, month_sin, month_cos.

Data Splitting and Sequence Creation

The data was split chronologically.

Set	Proportion	Samples (Features)
Training Set	80%	(34449, 18)
Validation Set	10%	(4306, 18)
Test Set	10%	(4307, 18)

Sequence Creation:

- **Input Sequence Length (SEQ_LEN): 24 hours**. (The model uses the past 24 hourly data points to predict the next AQI value.)
- **Training Sequence Shape:** (34425, 24, 18) (Samples, Time Steps, Features).

---

2. Model Architecture

The model is a custom implementation of a **Transformer Encoder-based** network adapted for time series regression. The model does not explicitly use positional encoding, relying on the input sequence structure and engineered temporal features for time dependence.

Transformer Encoder Block Components

The core of the model consists of a single encoder block (although referred to as layers in the summary, they form one block) applied to the input sequence.

Layer Type	Purpose / Function	Units / Output Shape	Parameters
Input Layer	Accepts the sequence input.	(None, 24, 18)	0
MultiHeadAttention	Self-attention mechanism.	(None, 24, 18)	9,618
Dropout	Regularization with rate 0.1 (in code)	(None, 24, 18)	0
Add	Residual connection.	(None, 24, 18)	0
LayerNormalization	Normalizes the attention output.	(None, 24, 18)	36
Dense (FFN 1)	Feed-Forward Network (Expansion).	(None, 24, 64)	1,216
Dense (FFN 2)	Feed-Forward Network (Compression).	(None, 24, 18)	1,170

Final Output Head

The sequence-level output from the Transformer block is processed for the final single-value prediction.

1. **Flatten** (Converts (None, 24, 18) to (None, 432))
2. **Dense (64 units, ReLU)**
3. **Dense (1 unit, Linear)** (Final regression output)

- **Total Trainable Parameters: 14,583**.

---

3. Training Configuration

Parameter	Value	Detail
Optimization Algorithm	**Adam**	Adaptive Moment Estimation.
Loss Function	**Mean Squared Error (MSE)**	Standard for regression tasks.
Primary Metric	Mean Absolute Error (MAE)	Used alongside loss for monitoring.
Maximum Epochs	50	Predefined limit.
Batch Size	32	Deduced from 1076 steps per epoch.
Early Stopping Callback	Monitored: val_loss	Stops training if validation loss does not improve.
	Patience: **10 epochs**	Number of epochs with no improvement before stopping.
Best Epoch (Observed)	The minimum validation loss (**0.0023**) was observed at Epoch 18.	The model weights from this epoch were restored.

4. Evaluation Metrics and Results

Model performance was evaluated on the unseen test set, with predictions inverse-transformed back to the original AQI scale.

Evaluation Metrics

1. **Mean Absolute Error (MAE):** Average magnitude of errors.
2. **Root Mean Squared Error (RMSE):** Measures the standard deviation of the prediction errors.
3. **Coefficient of Determination (R^2 Score):** Proportion of the variance in the target variable that is predictable (perfect fit = 1.0).

Performance Summary

```
📌 Validation Evaluation
MAE  = 13.3044
RMSE = 20.0150
R²   = 0.9156
--------------------------------------
📌 Test Evaluation
MAE  = 11.3740
RMSE = 15.3013
R²   = 0.9034
--------------------------------------
```