Environmental Science and Pollution Research

Forecasting Carbon Footprint: Temporal Fusion Transformer Modeling for Predicting CO2 Emissions in Bangladesh

--Manuscript Draft--

Manuscript Number:	ESPR-D-24-14674
Full Title:	Forecasting Carbon Footprint: Temporal Fusion Transformer Modeling for Predicting CO2 Emissions in Bangladesh
Article Type:	Research Article
Keywords:	Carbon emissions; Carbon Dioxide Forecasting; Time series; Temporal Fusion Transformer; Bangladesh
Corresponding Author:	Mohammad Sadman Tahsin BRAC University BANGLADESH
Corresponding Author Secondary Information:	
Corresponding Author's Institution:	BRAC University
Corresponding Author's Secondary Institution:	
First Author:	Shuvro Ahmed
First Author Secondary Information:	
Order of Authors:	Shuvro Ahmed
	Neda Firoz
	Mohammad Sadman Tahsin
Order of Authors Secondary Information:	
Funding Information:	
Abstract:	Climate change has emerged as a pressing concern in recent years, with carbon emissions playing a significant role in its progression. The study of carbon dioxide (CO2) emissions has garnered considerable attention, particularly in the context of sustainable development. China, as the world's largest energy consumer and greenhouse gas emitter, has pledged to peak carbon emissions before 2030 and achieve carbon neutrality before 2060. Various modeling approaches have been proposed to forecast carbon emissions, including support vector machines (SVM), random forest, artificial neural networks (ANN), and deep learning (DL). In this paper, the Temporal Fusion Transformer is employed to forecast carbon emissions in Bangladesh. This model, based on the transformer architecture, is adept at handling complex datasets, including static, dynamic, and categorical data. Our results show significant performance improvements over SMAPE of 0.17 over existing benchmarks for forecasting carbon emissions in Bangladesh.
Suggested Reviewers:	Faiza Tafannum UT Arlington: The University of Texas at Arlington fxt8308@mavs.uta.edu
	Muhammad Iqbal Hossain BRAC University iqbal.hossain@bracu.ac.bd
	Musaddiq Karim AIUB: American International University Bangladesh musaddiqmk19@gmail.com
Opposed Reviewers:	
Additional Information:	

Question	Response
§Are you submitting to a Special Issue?	No

Mohammad Sadman Tahsin (Corresponding author)
Email: sadman.tahsin3@gmail.com
Department of Computer Science and Engineering,
BRAC University, Dhaka, Bangladesh
Editors-in-Chief,
Environmental Science and Pollution Research

Dear Sir,

I am writing to submit our manuscript titled " **Forecasting Carbon Footprint: Temporal Fusion Transformer Modeling for Predicting CO2 Emissions in Bangladesh.**" for consideration and publication in **Environmental Science and Pollution Research**. I believe that this research significantly contributes to the field of Carbon Emission and has the potential to impact the detection and prevention of Carbon. **I have not submitted my manuscript to a preprint server before submitting it to Environmental Science and Pollution Research**.

Thank you for considering our submission. We hope that our research on Carbon Emission prediction will be of interest to the readership of **Environmental Science and Pollution Research**. We look forward to the opportunity to contribute to the journal's mission of advancing knowledge in Carbon Emission prediction and making a positive impact on the Carbon Emission.

Sincerely,
Mohammad Sadman Tahsin

Forecasting Carbon Footprint: Temporal Fusion Transformer Modeling for Predicting CO2 Emissions in Bangladesh

1. Shuvro Ahmed, Computer Science and Engineering, BRAC University, Dhaka. ahmedshuvro01@gmail.com
2. Neda Firoz, Institute of Applied Mathematics and Computer Science, National Research Tomsk State University, Tomsk Russia. nedafiroz1910@gmail.com
3. *Mohammad Sadman Tahsin, Computer Science and Engineering, BRAC University, Dhaka. Sadman.tahsin3@gmail.com

***Corresponding Author Email id: sadman.tahsin3@gmail.com**

Abstract

Climate change has emerged as a pressing concern in recent years, with carbon emissions playing a significant role in its progression. The study of carbon dioxide (CO2) emissions has garnered considerable attention, particularly in the context of sustainable development. China, as the world's largest energy consumer and greenhouse gas emitter, has pledged to peak carbon emissions before 2030 and achieve carbon neutrality before 2060. Various modeling approaches have been proposed to forecast carbon emissions, including support vector machines (SVM), random forest, artificial neural networks (ANN), and deep learning (DL). In this paper, the Temporal Fusion Transformer is employed to forecast carbon emissions in Bangladesh. This model, based on the transformer architecture, is adept at handling complex datasets, including static, dynamic, and categorical data. Our results show significant performance improvements over SMAPE of 0.17 over existing benchmarks for forecasting carbon emissions in Bangladesh.

Keywords

Carbon emissions, Carbon Dioxide Forecasting, Time series, Temporal Fusion Transformer, Bangladesh.

1. Introduction

Climate change has emerged as a prominent focal point within environmental concerns (Althor et al., 2016). Due to their significant impact on climate change, carbon emissions—which are primarily the result of human activity such as industries, transportation, etc.—have become a major global concern(Deutch, 2017). Nations across the globe have come to acknowledge the significance of mitigating carbon emissions while simultaneously sustaining economic development (H. Liu & Lin, 2017). To put it simply, one of the main goals of sustainable development is to separate the effects of global economic growth (EG) and carbon emissions (CEs). Global carbon dioxide emissions have experienced a significant and rapid escalation (Z. Liu et al., 2022). The reduction of carbon dioxide (CO2) emissions has emerged as a prominent topic within the scope of research on global climate change (Sun & Liu, 2016). The World Resources Institute reports that there has been a significant increase in CO2 emissions, approximately 150-fold, from the year 1850 to 2011. Furthermore, this upward trajectory has persisted up until the year 2018 (Emami Javanmard & Ghaderi, 2022). Currently, China holds the distinction of being the foremost global user of energy as well as the primary producer of greenhouse gases. During the 75th Session of the UN General Assembly on September 22, 2020, President Xi Jinping articulated China's commitment to achieving the peaking of carbon emissions before 2030 and attaining carbon neutrality before 2060 (Y. Wu & Xu, 2022). This objective will be pursued through the amplification of China's intended nationally determined contributions and the implementation of more robust policies and measures. During the Climate Ambition Summit 2020, held on December 12 of the same year, President Xi Jinping declared that China intends to reduce its carbon dioxide (CO2) emissions per unit of gross domestic product (GDP) by over 65% compared to the 2005 level. Additionally, China aims to raise the proportion of non-fossil fuels in primary energy consumption to approximately 25% by the year 2030 (F. Yang et al., 2023).

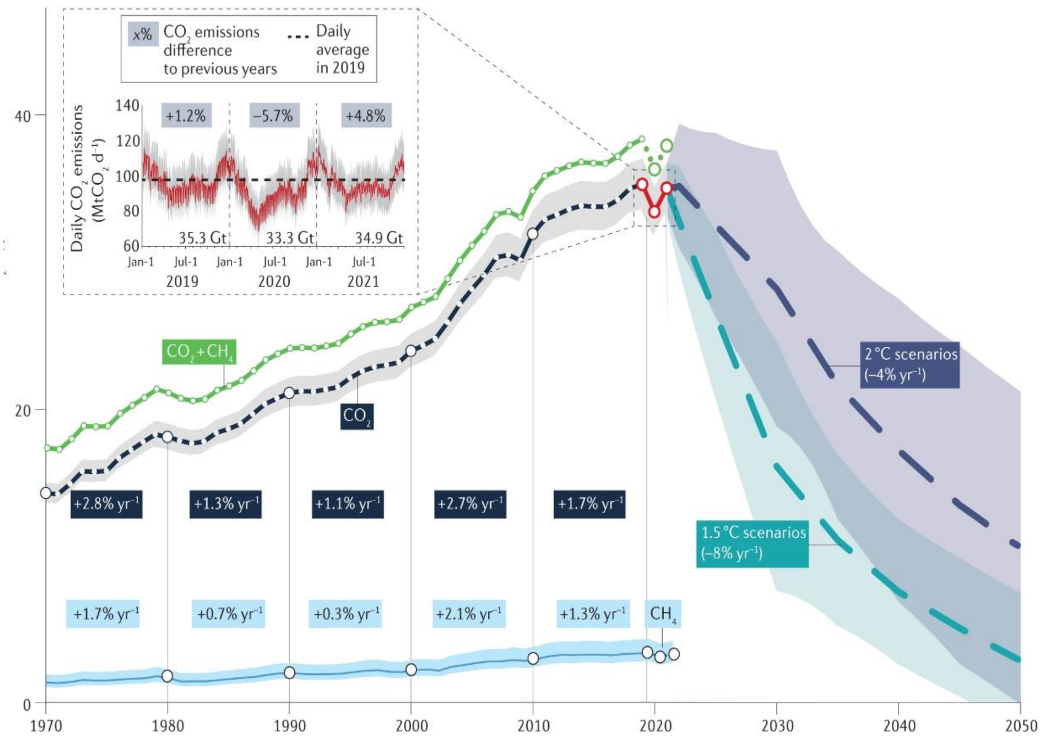

Fig 1. Global Carbon Emission (Z. Liu et al., 2022).

Various models have been suggested to address the task of projecting carbon dioxide (CO2) emissions (Sun & Liu, 2016). Several models that are commonly used in machine learning include support vector machine (SVM) (González Costa et al., 2017), (D. Liu et al., 2017), (Pham et al., 2019), random forest (Dou et al., 2019), (Hong et al., 2019), (Rodriguez-Galiano et al., 2014), artificial neural networks (ANN) (Ganzenmüller et al., 2019), (Pradhan et al., 2020), (Schmidt et al., 2018), (M. Yang et al., 2020) and deep learning (DL) (Dou et al., 2020), (Z. Wu et al., 2020). In artificial neural networks (ANN), the algorithm must possess a high degree of adaptability to successfully tackle a wide range of jobs. A learning mechanism is employed to facilitate the training of individual components inside the network, enabling them to acquire the necessary skills to accomplish the intended task. The field of deep learning has made significant advancements in recent years. Time series algorithms like Arima (H. Yang & O'Connell, 2020), (Sen et al., 2016), (Q. Wang et al., 2020) Sarima (Cui et al., 2023), (Lau et al., 2014), Auto Regression (Ahmed et al., 2022), (Xu & Lin, 2016), (Danish & Ulucak, 2021), Prophet (Primandari et al., 2022), (Meng & Dou, 2023), (Shaw et al., 2022) etc. are used for forecasting and have achieved excellent results.

In this paper, the Temporal Fusion Transformer has been used to forecast the carbon emissions of Bangladesh. The model is part of the transformer architecture (Lim et al., 2021) and is known for handling complex sets of inputs, which can be static and dynamic as well as categorical data.

2. Literature Review

In examining the carbon emissions landscape, the Chinese pulp and paper industry (CPPI) emerges as a noteworthy contributor (H. Wang, 2023), prompting an in-depth investigation into its intricacies. From 2005 to 2019, a lot of research was done using the LMDI method and the Tapio decoupling model to try to figure out the complicated link between CPPI's carbon emissions and economic growth. This research includes using the STIRPAT model to make predictions, which shows a changing path and focuses on important factors such as the value of industrial output per person. The study underscores the formidable challenge of achieving carbon peaking by 2030, emphasizing the imperative for effective low-carbon policies to ensure sustainable development within the CPPI.

Transitioning to the broader context of global warming, a study (F. Yang et al., 2023) scrutinizes Chengdu's carbon dioxide (CO_2) emissions dynamics from 2009 to 2020. Using the STIRPAT and Tapio decoupling models, the study finds that Chengdu's rising energy use is the main cause of CO_2 emissions. Notably, the city appears on track to fulfill its carbon peak commitment by 2025, with a "business as usual" scenario proving effective for low-carbon development. This promising trajectory underscores Chengdu's shift from weak to strong decoupling between CO_2 emissions and economic growth, offering valuable insights for global carbon peaking endeavors.

Turning to the global concern of air pollution and greenhouse gas (GHG) emissions, another work (Emami Javanmard & Ghaderi, 2022) introduces a novel hybrid approach integrating machine learning and mathematical programming. Focused on predicting GHG emissions in the Iranian energy sector from 1990 to 2018, the study employs various forecasting algorithms and metaheuristic algorithms (PSO and GWO). The results show a big boost in the accuracy of predictions, highlighting how useful this method is for improving GHG emissions forecasting and management. By 2028, it is expected that emissions will be more than 1096 Mt/year.

Shifting the narrative to the profound implications of global warming, a study (Fang et al., 2018) addresses its impact on immigration, agriculture, and human conflicts. The introduction of an advanced Gaussian processes regression (GPR) method, enhanced by a modified Particle Swarm Optimization (PSO) algorithm, facilitates carbon dioxide emission forecasting from historical data (1980-2012) in the United States, China, and Japan. The study suggests varying emission control scenarios, providing policy insights for emission reduction efforts.

A study (Sen et al., 2016) stresses how important it is to predict energy use and greenhouse gas (GHG) emissions in India's iron and steel sector to promote environmentally conscious manufacturing (ECM). Using the autoregressive integrated moving average (ARIMA) method, the study finds the best models for both indicators. These models provide accurate predictions that are needed to make smart decisions about environmental policy and improve environmental management.

A critical examination of the escalating concern surrounding China's CO_2 emissions takes center stage in another work (Sun & Liu, 2016). In line with the 12th Five-Year Plan's goals to reduce CO_2 emissions, the study sorts total CO_2 emissions into groups based on industry and uses cointegration and Granger causality tests to find factors that affect each group. Using the Least Squares Support Vector Machine (LSSVM) to guess different CO_2 emissions, the study shows that this method is much more accurate than other models during predictions.

A study (Niu et al., 2020) uses an IFWA-GRNN model and an RF method to figure out how well China can meet its 2030 carbon emission intensity reduction goals. Results suggest potential challenges under the Basic as Usual scenario but favorable outcomes with policy tightening and market allocation scenarios. Addressing remaining hurdles requires policy adjustments, market incentives, and advancements in energy practices, emphasizing the importance of cleaner energy sources and markets.

Lastly, a study addresses (Qiao et al., 2020) the critical concern of global warming by introducing a novel hybrid algorithm that combines the Lion Swarm Optimizer and Genetic Algorithm with the Least Squares Support Vector Machine model. When compared to eight other algorithms, this one show better global optimization, faster convergence, and higher accuracy. This makes the model a useful tool for fighting global warming problems because it predicts emissions from 2018 to 2025.

3. Methodology

3.1.1. Data Description

The dataset was collected from Global Carbon Atlas website (*Global Carbon Atlas*, n.d.). The dataset is in a csv format and has information about total co2 emissions from 1960-2021 in Bangladesh. It is a time series dataset and the dataset (features and target combined) have 62 rows and 2 columns. The following table 1 is a general overview of the dataset:

Index	Unique	Missing	Duplicated	Types

Date	62	0	0	int64
Bangladesh	62	0	0	float64

Table 1. Overview of the dataset

3.1.2. Data Preprocessing:

To begin with, we checked whether the dataset contained any null values, and as seen in Table 1, the dataset is compact and no missing values is detected. Then, we saved the date values (e.g. year) in a separate column named as "Year" and then converted the "Date" column to date time yearly format. Next, we set the "Date" column as our index column for ease of visualization. Besides, as the dataset contained only one feature, our priority was to generate appropriate features from the given information. As the dataset is time series based, we tried to find the trend of it and see whether there are any seasonal patterns to be found using statistical methods (Lim et al., 2021). To achieve that, we first visualized the overall trend of yearly co2 emission in Bangladesh as seen in Fig 2. At a glance, we can see that the co2 emission has been increasing over the years and had a rapid growth since 2000.

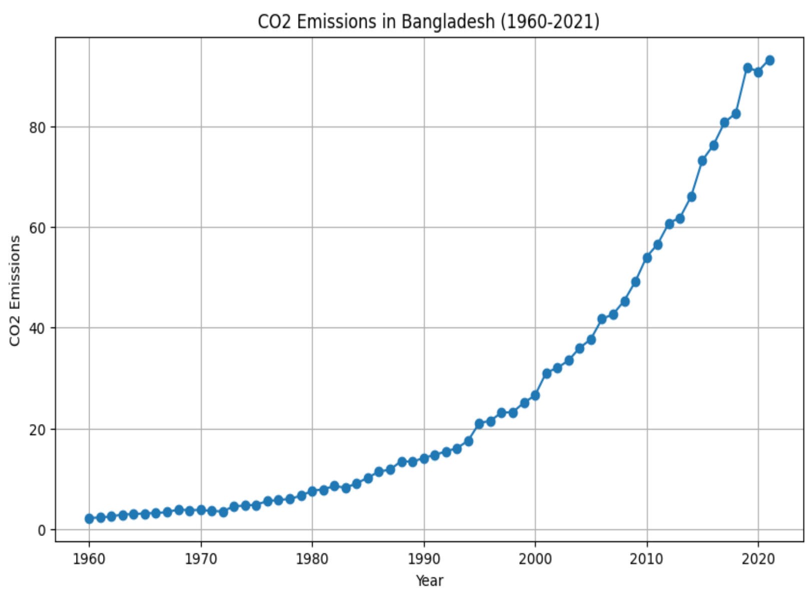

Fig 2. Time Series plot of Yearly Co2 emission.

Similar observation can be inferred from the histogram of Fig 3. as we can clearly see the frequency of high value emissions are in the lower bound, which tracks with higher level of Co2 emission being a recent phenomenon.

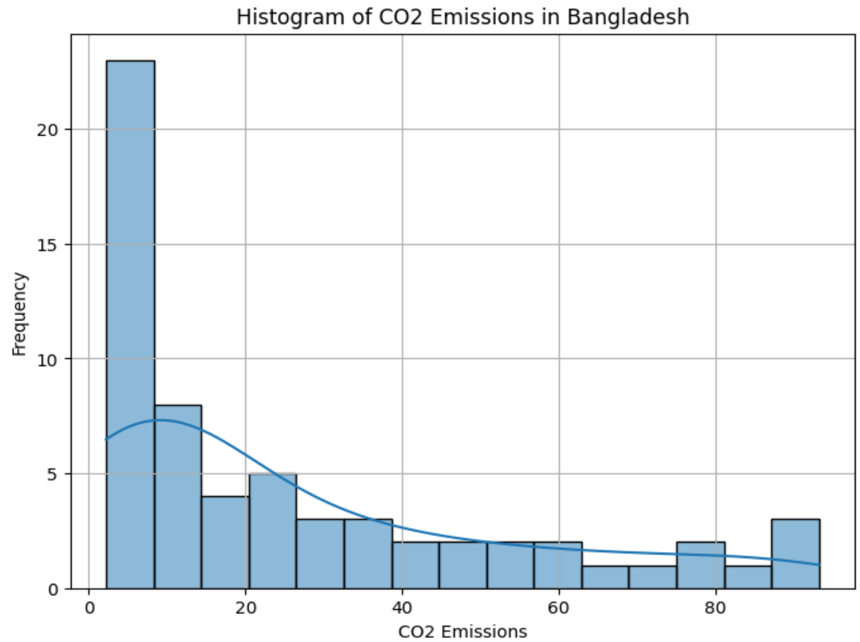

Fig 3. Histogram of CO2 emissions.

Next, using Seasonal Decomposition techniques (Holt, 2004), we observed whether the dataset has any seasonality and residuals as well as trends. As shown in Fig 4., the dataset only has a trend but no residuals or seasonality detected.

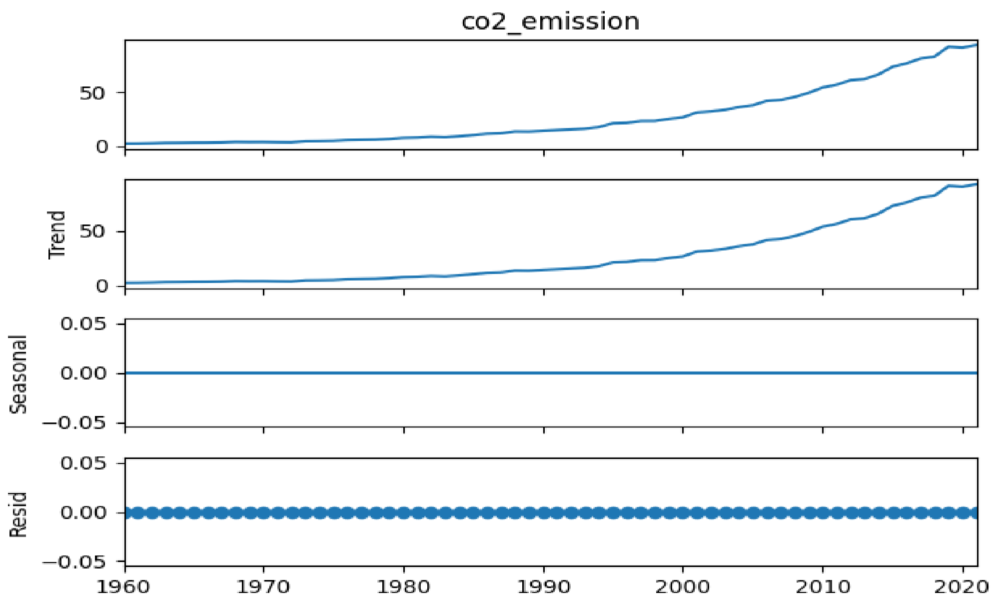

Fig 4. Seasonal Decomposition of Yearly Co2 emission.

Now, the task was to find out the type of trend and for that, we first used to move average method (He et al., 2022) with a window size of 62. However, it seems hard to define the trend only from the moving average graph as shown in Fig 5.

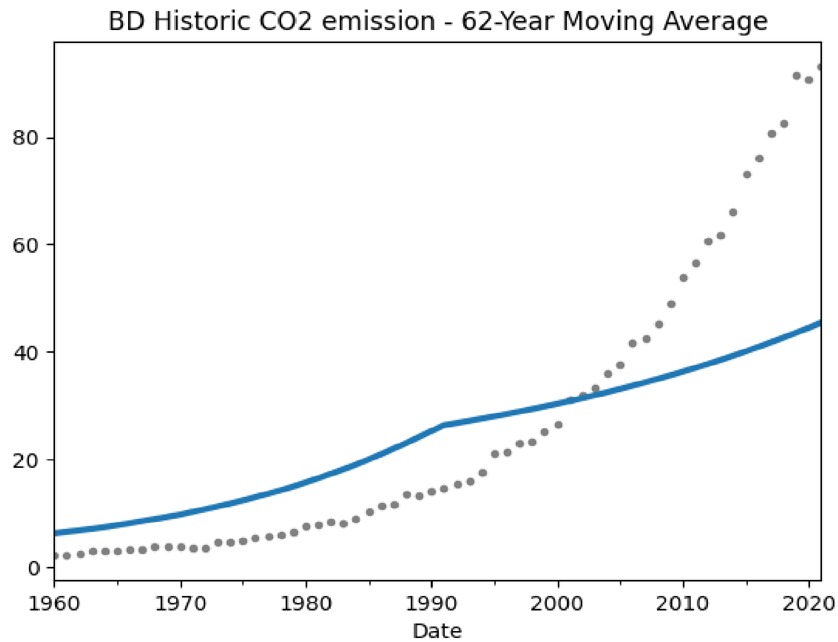

Fig 5. Moving Average of CO2 Emission.

Next, we used deterministic process (Bazionis & Georgilakis, 2021) to generate new features and found out that at order = 2, the linear regression fitted graph as seen in the Fig 6. gives us a very similar trend shown in Fig 2.

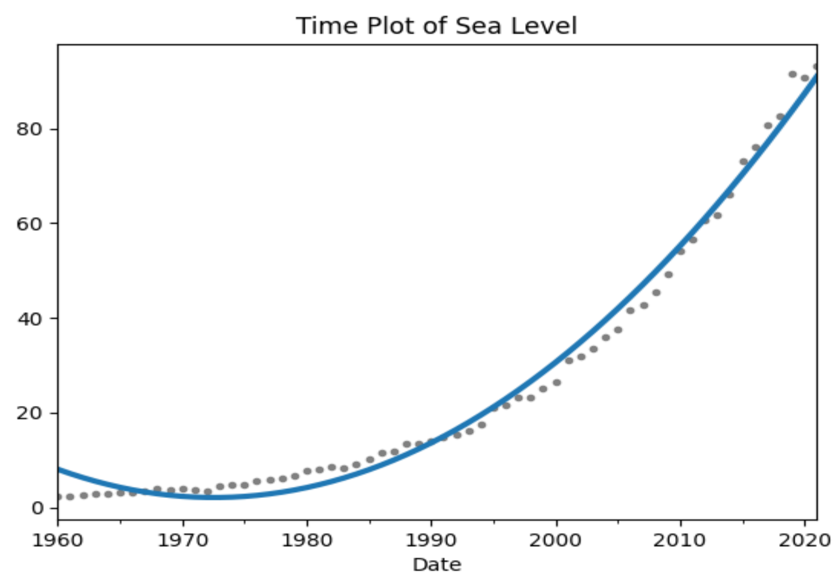

Fig 6. Linear Regression Plot of deterministic process (order = 2).

So, we can confidently say that the co2 emission represents a order 2 polynomial trend or a quadratic trend. In Fig 7., we can clearly see that the newly generated feature correlates the highest with our target value which is labled as "co2_emssion".

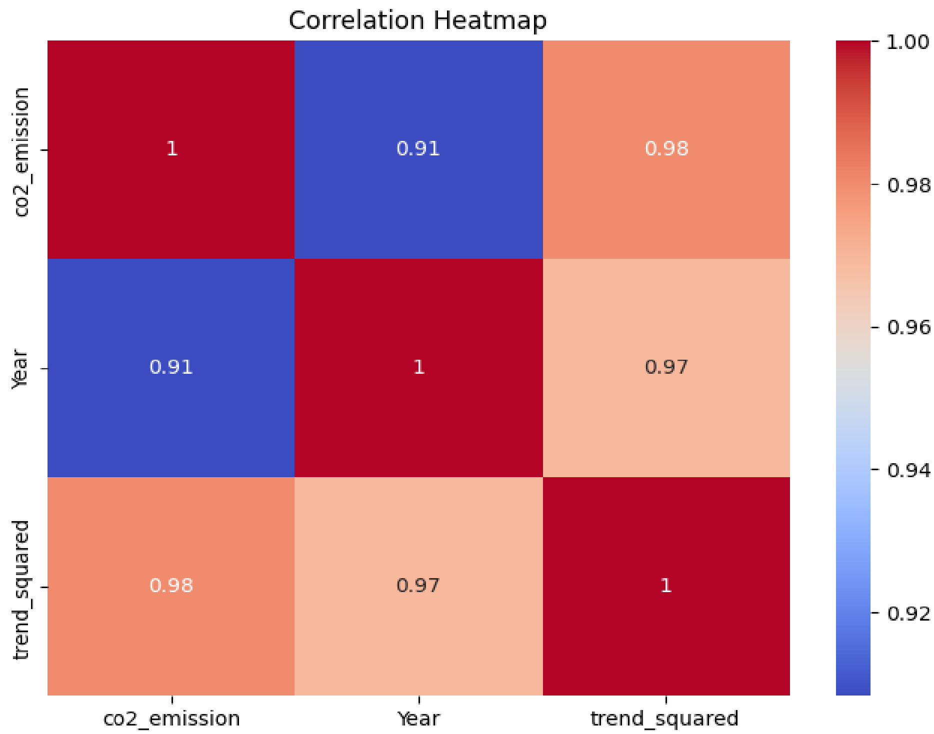

Fig 7. Correlation heatmap of the dataset.

So, from here, we generated an important feature known as trend_squared to complement our model.
As stated before, we have chosen Temporal Fusion Transformer as our model. The Temporal Fusion Transformer (TFT) is a cutting-edge Deep Neural Network designed for superior performance and interpretability. TFT handles three types of features: temporal data with known future inputs, temporal data up to the present, and time-invariant features. It supports multi-series training from diverse distributions, dividing processing into local and global components for event-specific and collective characteristics capture. TFT enables multi-step predictions and prediction intervals using the quantile loss function. Built on a transformer-based architecture, it features a unique multi-head attention mechanism for enhanced feature importance insights.

3.1.3. Data Loader:

Now, we start preparing our data for the Temporal fusion transformer model. To begin with, a column named id was created and assigned the value of 0. This is due to the fact that our dataset is a univariate time series and hence would require to be grouped as one entity. Moreover, we generate another column known as “time_idx” and set the starting value to 0 and increment for each observation. The primary attribute within our dataset of significance is the "time_idx" column, as it dictates the chronological order of the samples. In the absence of any missing data points, the values in this column should exhibit a consistent increment of +1 for each time series. Subsequently, we encapsulate our dataframe within an instance of the TimeSeriesDataset. First of all, we set our max_prediction_length to 12 and max_encoder_length to 50. The maximum encoder length, denoted as max_encoder_length, establishes the retrospective period, while the max_prediction_length parameter determines the number of datapoints to be forecasted. In our specific scenario, we examine the preceding 50-time steps to generate 12 predictions. This is also done to keep a balance of 80-20 split of train-test and the training cutoff is based on that.
Then, we assign “co2_emission” as our target value, “time_idx” to time_idx and “id” to group_ids. Furthermore, 'Year', 'trend_squared', and 'time_idx'x column are also time-varying known features as their change of values in time are known while “co2_emission” column get assigned as time varying unknown feature for the opposite reason. Finally, we used Group Normalizer to normalize the entire dataset.

3.2. Workflow:

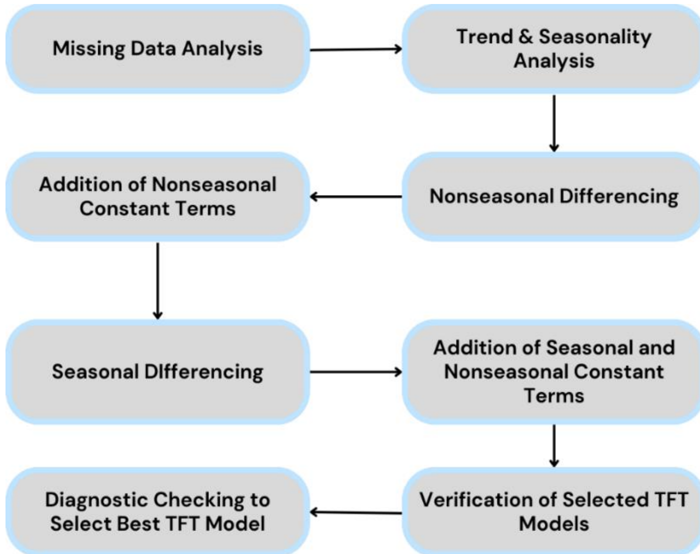

Fig 8. Workflow diagram of TFT model.

Here (Fig 8), TFT is chosen due to its ability to capture temporal patterns and handle both linear and nonlinear relationships in time series data. First of all, after completing the data loader, we use a naïve baseline model to predict co2 emission which gave us a Mean absolute error of 24.79.

Then, we focused on finding the optimal learning rate as the dataset is very small. Firstly, trainer was set up by using gpu as the accelerator & setting gradient_val_cap to 0.1 which is a hyperparameter and important to prevent divergance. Next, we used Tuner from pytorch and found the following result as shown in Fig 7. Which is 0.33 as seen in Fig 9.

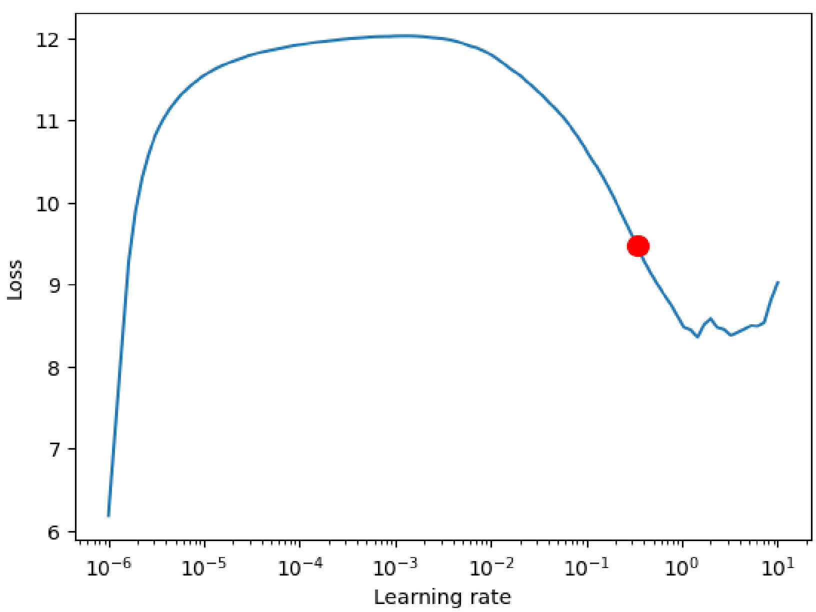

Fig 9. Optimal lr suggested by PyTorch Lightning learning rate finder.

However, as lr 0.33 is an unusual number, thus we used 3 types of lr along with other hyperparameters as shown below to find the best result:

- Dropout rate – 0.1, 0.2, 0.3
- Learning rate – 0.1, 0.33, 0.5
- Num. heads – 1, 4

Finally, we use early callbacks with patience of 20 to avoid overfitting. Next, we run the model and save the best one. Our optimal model took around 3 minutes to finalize the output.

3.3. Model architecture:

TFT optimizes forecasting performance with canonical components: gating mechanisms for adaptive complexity, variable selection networks, static covariate encoders, temporal processing with sequence-to-sequence and multi-head attention, and prediction intervals via quantile forecasts. Furthermore, TFT forecasts future outcomes using past target values, time-dependent exogenous inputs (unknown and known), and static covariates. Instead of a single value, TFT provides prediction intervals through quantiles for each τ-step-ahead forecast at time t. This can be seen in equation (i) below:

$$\hat{y}i(q,t,\tau) = fqi(\tau, yi, t-k{:}\, t; zi, t-k{:}\, t; xi, t-k{:}\, t+\tau; si) \ldots\ldots\ldots\ldots\ldots (1)$$

Here:

- $\hat{y}i(q,t,\tau)$ is the predicted (or forecasted) value for the i-th time series at time t for the quantile q at the forecast horizon τ.
- q: The quantile for which the prediction is made. In probabilistic forecasting, instead of predicting a single value, predictions are made for several quantiles of the conditional distribution of the future value.
- t: The current time index.
- τ: The forecast horizon, which is the amount of time into the future for which the prediction is being made.
- f q i : The forecasting function for the i-th time series, parameterized by quantile q.
- y i,t−k:t : The historical target values for the i-th time series from time t−k to t. This is the window of past observations used to make the prediction.
- z i,t−k:t : The unknown (future) inputs for the i-th time series from time t−k to t. These could be things like future planned promotions or events which are known at time t, but whose effects are unknown.
- x i,t−k:t+τ : The known inputs or observed covariates for the i-th time series from time t−k up to t+τ. These include things like past values of temperature or other factors that could affect the prediction.
- s i : The static covariates for the i-th time series, which do not change over time. These could be attributes like the location of a store or the type of product being forecasted.

The TFT uses historical data, known future inputs, and static covariates to forecast future values of a time series across different quantiles, which allows for an estimation of the uncertainty of the predictions.
Besides, Temporal Fusion Transformer (TFT) is a model built on the Transformer architecture, utilizing self-attention to comprehend intricate temporal dynamics across multiple time sequences.

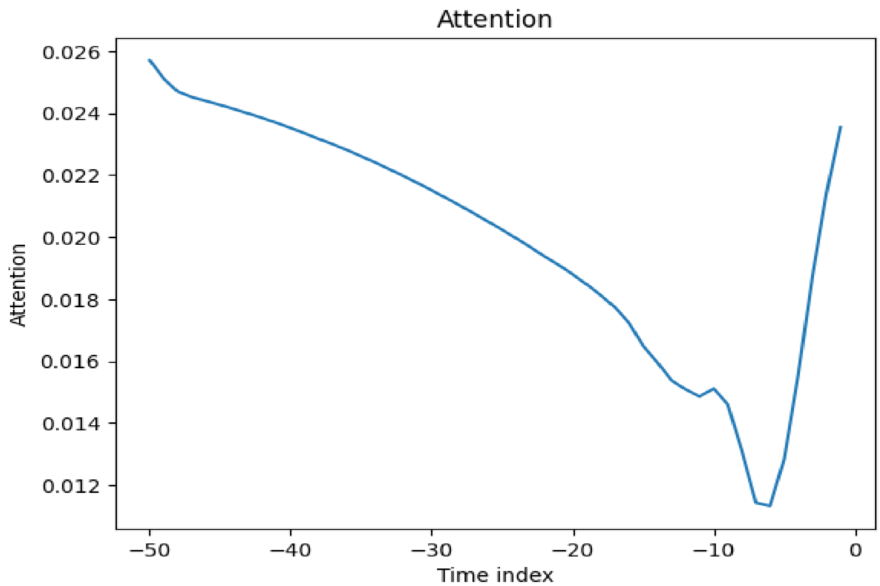

Fig 10. Attention of the optimal TFT model.

Moreover, as shown in Fig 10., the figure depicted represents the attention weights assigned by a Temporal Fusion Transformer model to various elements across different time indices. The following interpretations can be drawn from the graph:

Attention Mechanism Insight: The graph exhibits the model's utilization of an attention mechanism, which is pivotal in determining how it prioritizes different segments of the input sequence. The 'Attention' label on the y-axis likely denotes the weights, or significance, the model places on various inputs at specific time indices.

Time Index Representation: On the x-axis, the time indices are portrayed, typically representing sequential time steps in the dataset. The graph suggests a backward-looking perspective, with negative indices pointing to past observations.

Attention Weight Trend: There is a discernible declining trend in the attention weights from time index -50 to approximately -10, indicating the model's decreasing interest in historical data as it moves towards more recent points in time.

Notable Attention Shift: Around the -10-time index, there is a marked fluctuation—a dip followed by a spike in attention weights. This implies that the model has identified a particular event or data point around this time as being especially salient for the prediction task.

Recent Data Emphasis: The stark escalation in attention weights near the time index 0 illustrates that the model ascribes substantial importance to the latest data points, aligning with the common practice in time series forecasting where recent events are more heavily weighted for near-future predictions.

TFT leverages this Interpretable Multi-head attention technique by using the following equation:

$$\text{IMH (Q, K, V)} = \frac{1}{h}\sum_{i=1}^{h} head_i \; W_H \ldots\ldots\ldots\ldots\ldots.(2)$$

$$\text{Where } head_i = \text{Attention } (QW_Q^{(i)}, KW_K^{(i)}, VW_V)$$

Here, Q, K, and V represent queries, keys, and values respectively while W_H, W_Q, W_K, and W_V represent weight matrices for projection, with h denoting the number of attention heads.

4. Result:

We evaluated our model using mean absolute error, Quantile Loss, mean absolute percentage error, Root Mean Squared Error, and Symmetric mean absolute percentage error.

The Mean Absolute Error (MAE) constitutes a performance metric frequently utilized within the domain of regression analysis to assess the precision and accuracy of a predictive model. Its principal function involves quantifying the average absolute disparity that exists between the predicted values generated by the model and the actual observed values present within the dataset. MAE is computed by summing the absolute disparities between each predicted and actual value, then dividing this sum by the total number of data points. The MAE formula is expressed as:

$$MAE = \left(\frac{1}{n}\right) \Sigma(\mathrm{i} = 1 \text{ to } \mathrm{n}) |y_{\mathrm{i}} - \hat{\mathrm{y}}_{\mathrm{i}}| \ldots \ldots \ldots \ldots \ldots (3)$$

In this formula:

1) MAE represents the Mean Absolute Error.
2) n represents the total number of data points.
3) y_i signifies the actual observed value for the ith data point.
4) ŷ_i represents the predicted value generated by the regression model for the ith data point.

Quantile loss, also known as quantile regression loss, is a crucial metric in machine learning. It assesses models that predict different quantiles (e.g., 25th, 50th, 75th percentiles) of a target variable's distribution, unlike traditional regression that predicts the mean. This is particularly valuable for non-normally distributed data, offering a more comprehensive understanding of prediction uncertainty. However, TFT employs the pinball loss function to predict quantiles in the target ŷ distribution, emphasizing realistic under- and overestimations for various quantiles. The optimization process, akin to maximizing MAE loss for the median prediction, ensures nuanced model performance. Training TFT involves minimizing cumulative quantile losses across all outputs, enhancing predictive accuracy across the quantile spectrum.

For a given quantile level q, the quantile loss is defined as:

$$Quantile\ Loss = q \max(0, y - \hat{y}) + (1 - \mathrm{q}) \max(0, \hat{y} - \mathrm{y}) \ldots \ldots \ldots \ldots (4)$$

Here:
q represents the quantile level.
y is the true target value.
ŷ is the predicted value.

The Mean Absolute Percentage Error (MAPE) is a percentage-based metric that gauges the average percentage difference between the predicted and actual values in a dataset. It is particularly valuable for understanding the relative magnitude of errors in the context of the actual values. The MAPE formula is given by:

$$MAPE = \frac{1}{n \sum_{t=1}^{n} \left|\frac{At - Ft}{At}\right|} \ldots \ldots \ldots \ldots (5)$$

Here:
n represents the total number of data points.
At represents actual value
Ft represents forecast value

The Root Mean Squared Error (RMSE) serves as another critical metric in the evaluation of regression models, offering a comprehensive assessment of the model's predictive accuracy. Unlike the Mean Absolute Error (MAE), RMSE not only accounts for the absolute disparities between predicted and actual values but also incorporates a squared term for each disparity. This characteristic makes RMSE more sensitive to larger errors, emphasizing the significance of outliers in the dataset. The RMSE formula is defined as follows:

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n} (yi - \hat{y}i)^2 \dots\dots\dots\dots (6)}$$

In this formula:
n represents the total number of data points.
yi denotes the actual observed value for the ith data point.
ŷi represents the predicted value generated by the regression model for the ith data point.

Moving on to the Symmetric Mean Absolute Percentage Error (SMAPE), it is a metric that assesses the percentage disparity between predicted and actual values. SMAPE is particularly useful when dealing with data exhibiting different scales. It is expressed as:

$$SMAPE = \frac{100}{n}\sum_{t=1}^{n} \frac{|Ft - At|}{(|At| + |Ft|)/2} \dots\dots\dots\dots (7)$$

Here:
n is the total number of data points.
At is the actual value
Ft represents the forecast value
In table 2. We can see the values of mean absolute error and Quantile loss using the model.

Model	**Mean Absolute Error**	**Quantile Loss**	**Mean absolute percentage error**	**Root Mean Squared Error**	**Symmetric mean absolute percentage error**
Temporal Fusion Transformer	13.05	11.20	0.16	15.69	0.17

Table 2. Performance score of the model.

Here (Fig 11), we can observe that the mean absolute error is around 13 which is decent for such a small dataset with very few features.The final output of the best performing model can be seen in the following figure:

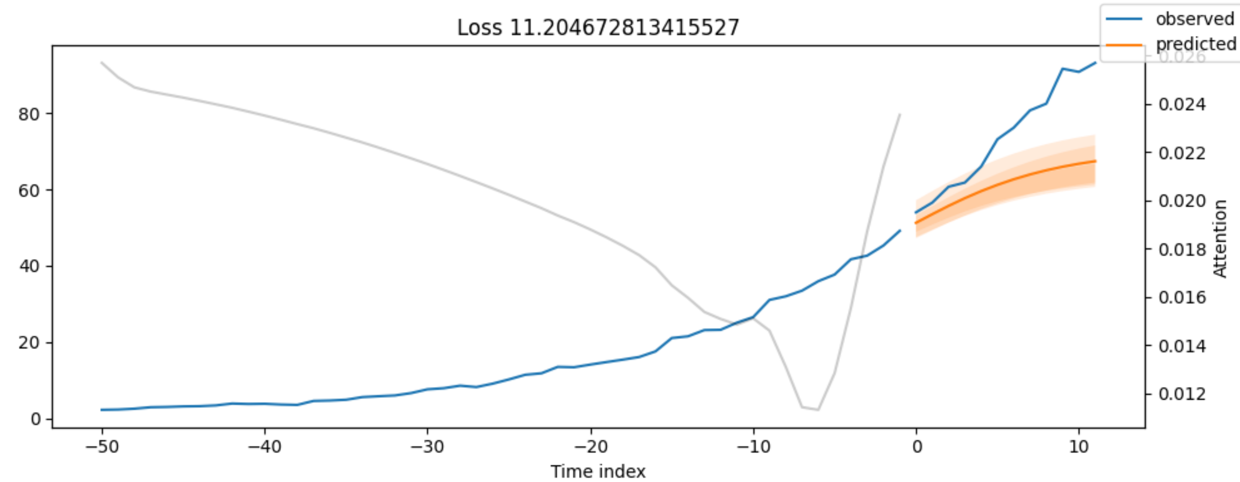

Fig 11. TFT Quantile Loss.

4.1. Feature-wise Interpretability:

TFT leverages its Variable Selection Network module to calculate the importance of every feature (Lim et al., 2021). Year feature can be seen as the top variable in both encoder and decoder importance as seen in Fig 13.,14. This is expected as it is the most relevant date present in each row. On the other hand, it can be seen that co2_emission_scale is the most important in terms of static variables importance as shown in Fig 12.

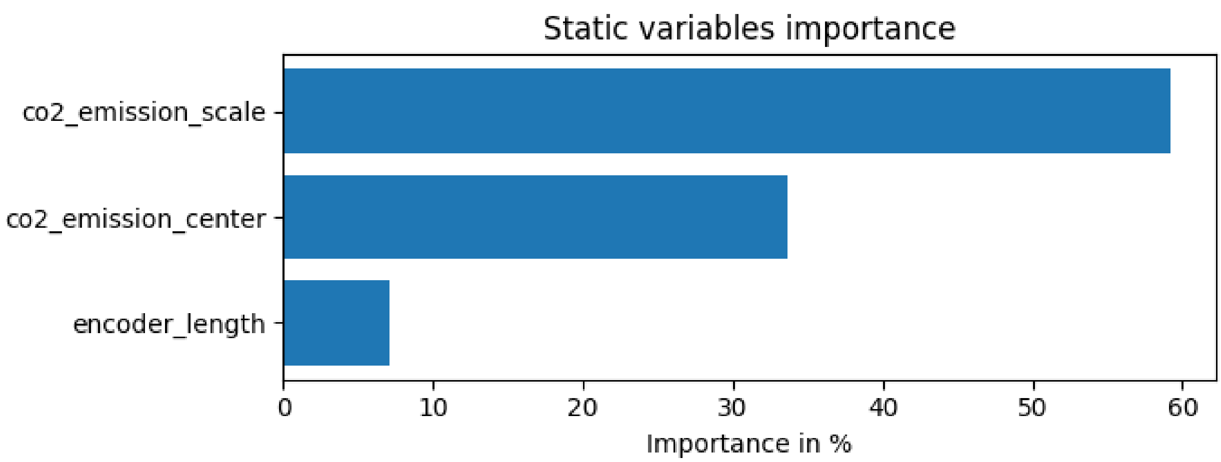

Fig 12. Static Variable Importance.

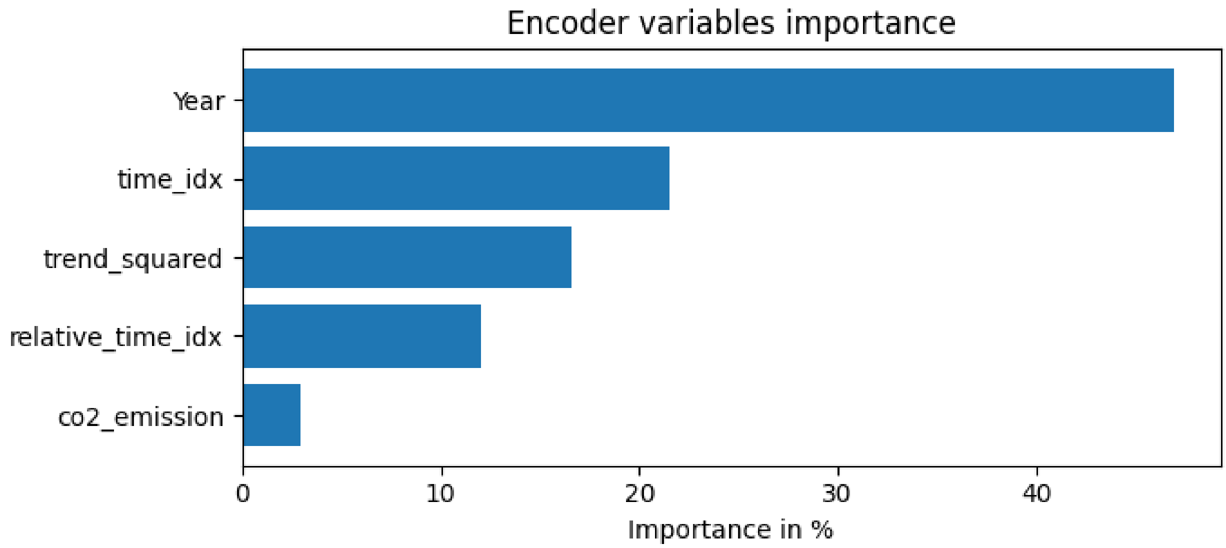

Fig 13. Encoder Variables Importance.

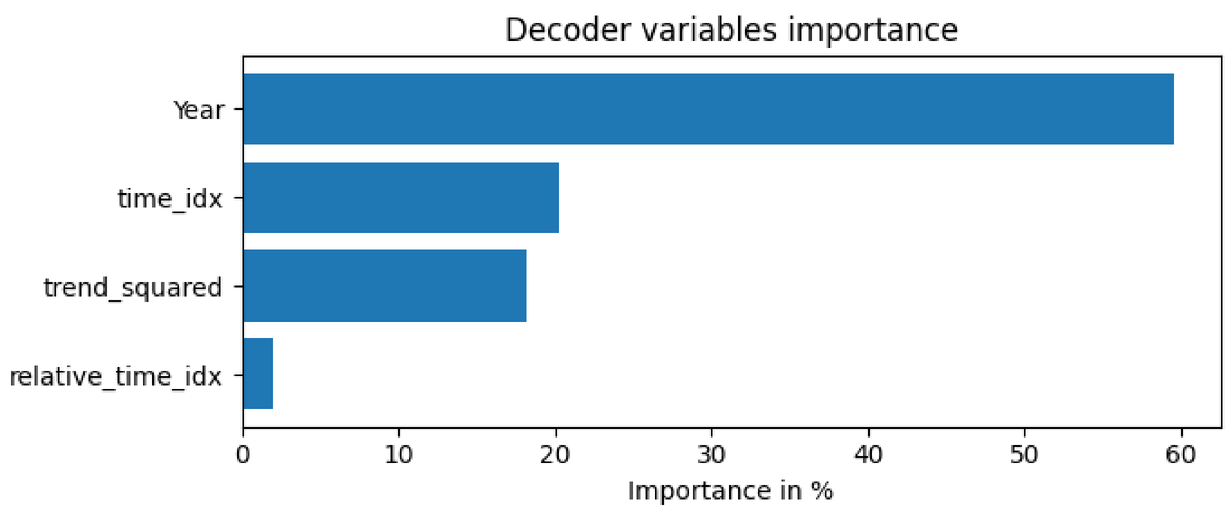

Fig 14. Decoder Variable Importance.

5. Discussion

Reference	Works	Model	MAE	MAPE	RMSE	SMAPE
(Niu et al., 2020)	Niu, D et. al.	improved fireworks algorithm (IFWA) - General regression neural network (GRNN)	0.003	0.017	0.003	-
(Emami Javanmard & Ghaderi, 2022)	Javanmard, M. et. al.	Particle Swarm Optimization (PSO)	0.0364	0.0009	0.0729	-

(Ahmed et al., 2022)	Ahmed, M. U., et. al.	Autoregression	0.00028	0.97	-	-
Our Proposed Work		Temporal Fusion Transformer	13.05	0.16	15.69	0.17

Table 3. Comparison with previous work.

According to Table 3, In (Niu et al., 2020), the IFWA-GRNN model shines with remarkable performance, showcasing exceptionally low MAE and RMSE values. This suggests a high degree of accuracy in predicting carbon emissions. Similarly, (Emami Javanmard & Ghaderi, 2022) present a robust performance in their PSO-based model, featuring relatively low MAE and MAPE values, indicating commendable predictive capabilities. In (Ahmed et al., 2022) the autoregression model demonstrates an extraordinarily low MAE but raises concerns with a notably high MAPE value. The absence of RMSE and SMAPE values further complicates a comprehensive evaluation of its accuracy.

Contrastingly, our initial implementation of the proposed Temporal Fusion Transformer model reveals noticeably higher MAE and RMSE values when compared to the cited works. However, direct comparisons are challenging as the referenced studies did not employ our model and omitted mentioning SMAPE values, preventing a holistic assessment. Further refinement and analysis are warranted to enhance the predictive accuracy of the Temporal Fusion Transformer and elucidate its performance relative to existing models.

5.1. Model Performance Across Variables:

Evaluating the model's output over various segments of the dataset helps identify areas of underperformance. Displayed subsequently are the average predicted values versus the actual values for each variable, segmented into 100 categories. With this segmentation, predictions can be made directly on the resultant data utilizing methods such as calculate_prediction_actual_by_variable() and plot_prediction_actual_by_variable(). The gray bars represent the distribution of data for each variable within these categories, functioning as a histogram.
CO2 Emission: The model's predictions closely track the actual values for most of the range, but there are discrepancies at certain points where the prediction either underestimates or overestimates the actual values (Fig 15.). The frequency bars suggest that the model may struggle with less frequent higher CO2 emission values.

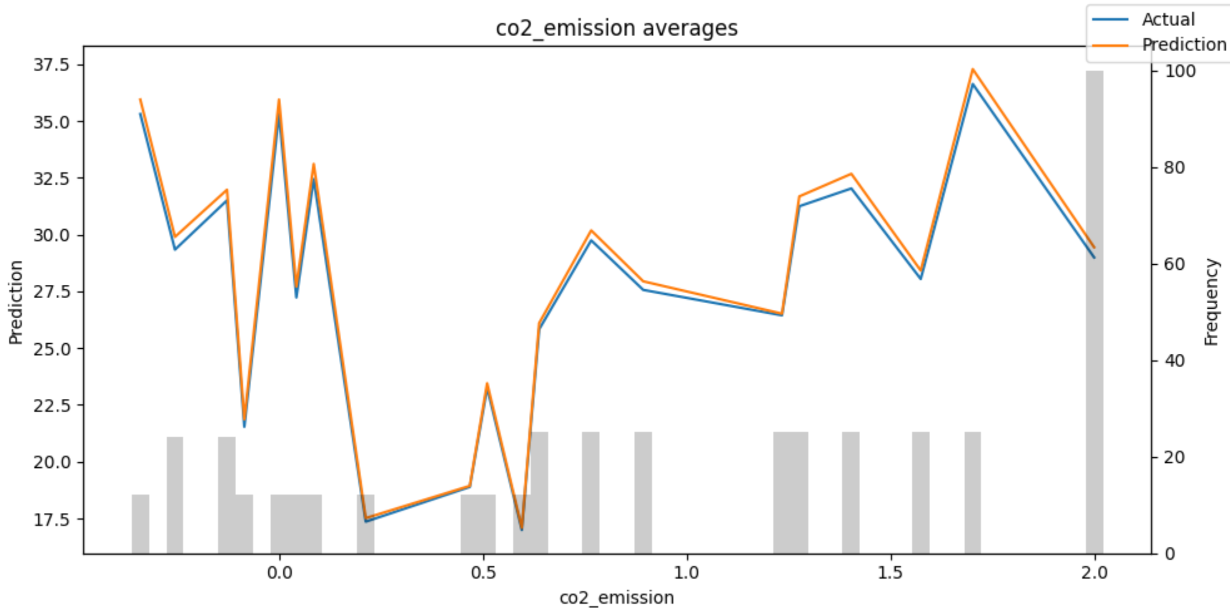

Fig 15. Robustness of co2_emission averages.

Relative Time Index: The predictions are relatively close to the actuals across the normalized relative time index (Fig 16.). However, there is a tendency to underestimate at the very beginning and overestimate towards the peak. This could indicate that the model may not capture the initial and peak trends accurately.

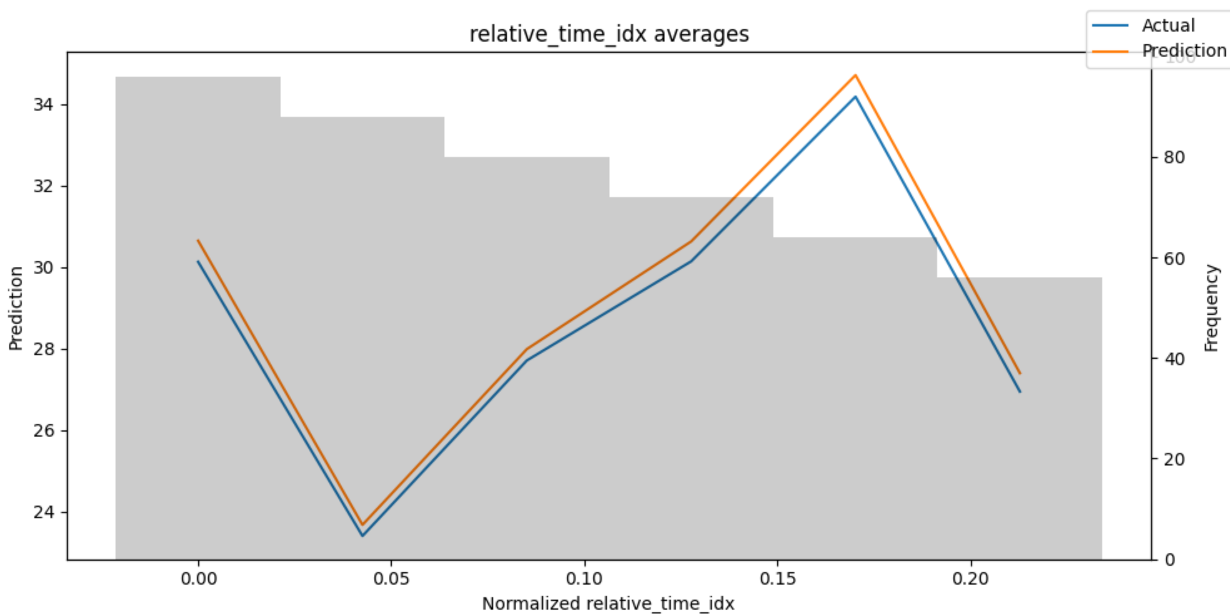

Fig 16. Robustness of relative_time_idx averages.

Time Index: The model seems to perform well across different time indices, but there are visible gaps between predictions and actual at certain intervals, particularly where there is a sudden change in trend or at the edges of the time index range as seein in Fig 17. This suggests a possible weakness in adapting to abrupt changes in the time series.

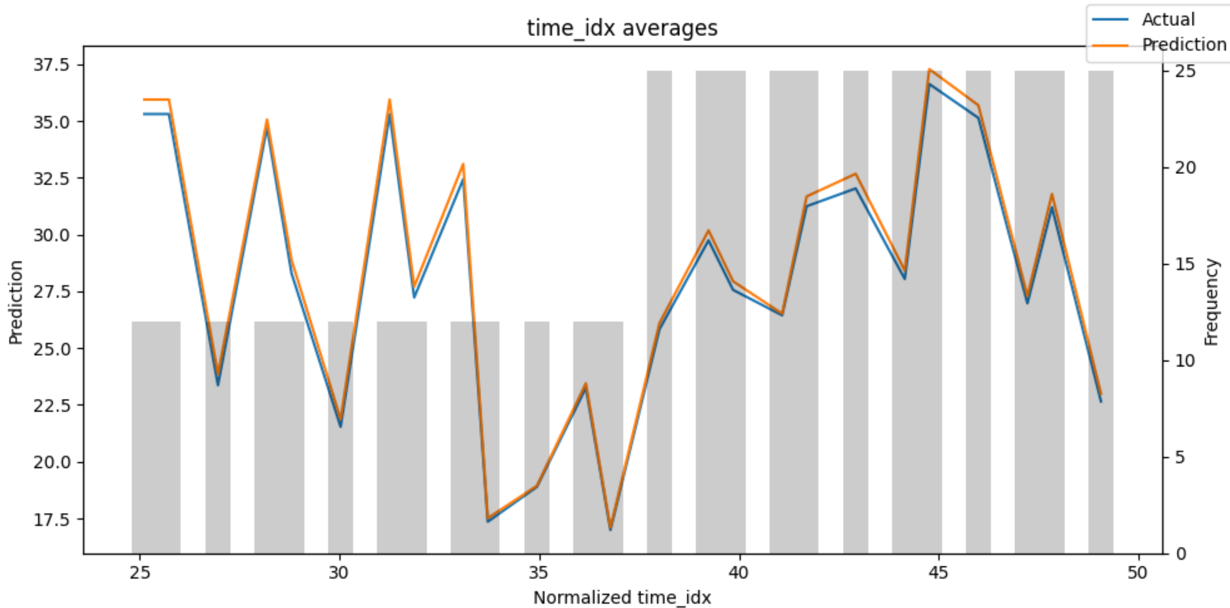

Fig 17. Robustness of time_idx averages.

Trend Squared: Here, we see a significant divergence in the predictions at higher values of the normalized trend squared (Fig 18.). The model's accuracy decreases as the trend squared value increases, indicating a potential difficulty in capturing nonlinear relationships.

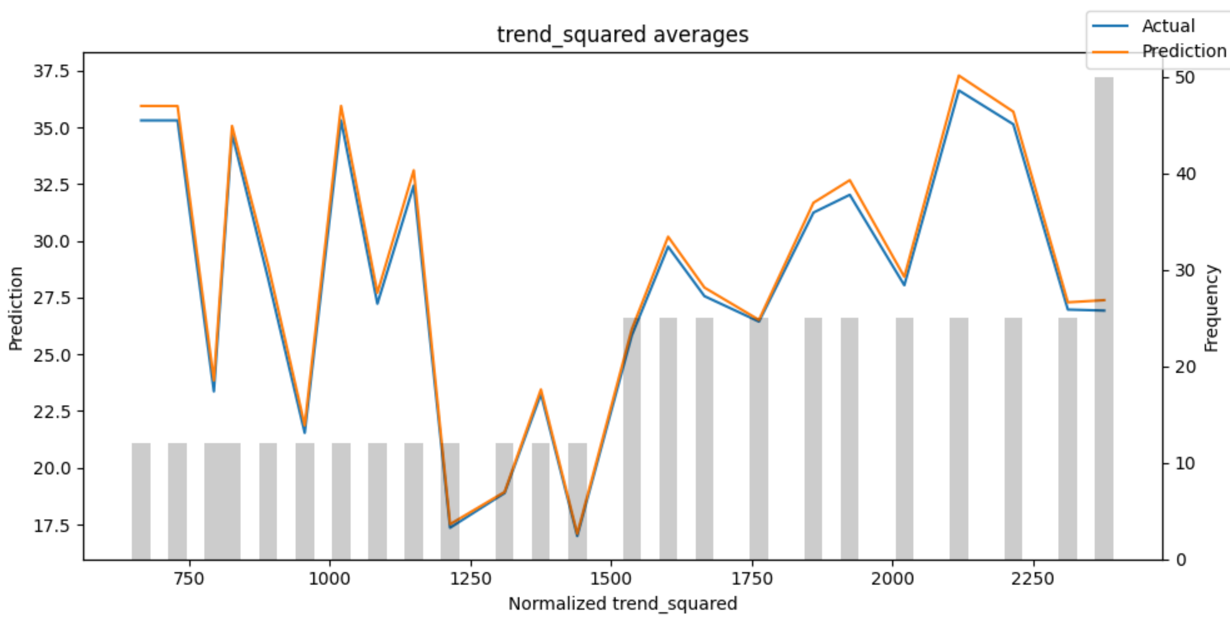

Fig 18. Robustness of trend_squared averages.

Year: The predictions follow the actuals closely over the years but with occasional overestimations or underestimations (Fig 19.). This could be due to external factors that the model does not account for or from complex patterns that arise over longer time scales.

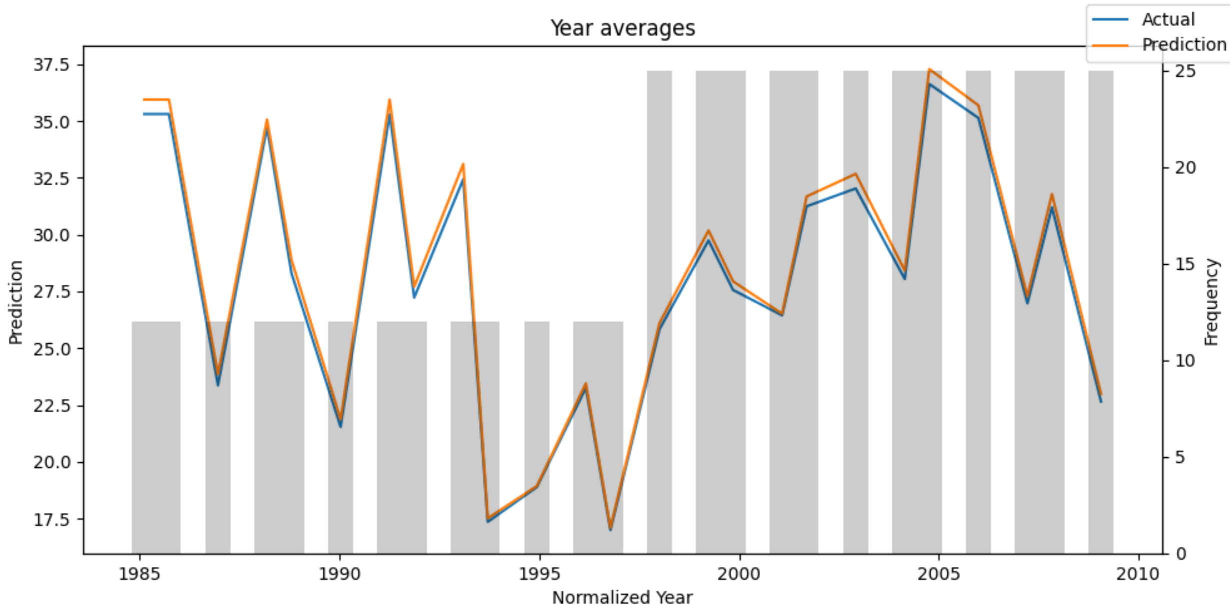

Fig 19. Robustness of Year averages.

CO2 Emission Scale and Center: The last two variables (Fig 20.,21) show very narrow distributions for actuals, which the model predictions don't seem to capture, as they appear to be constant over the range. This lack of variability in predictions may indicate the model's insensitivity to these specific normalized features.

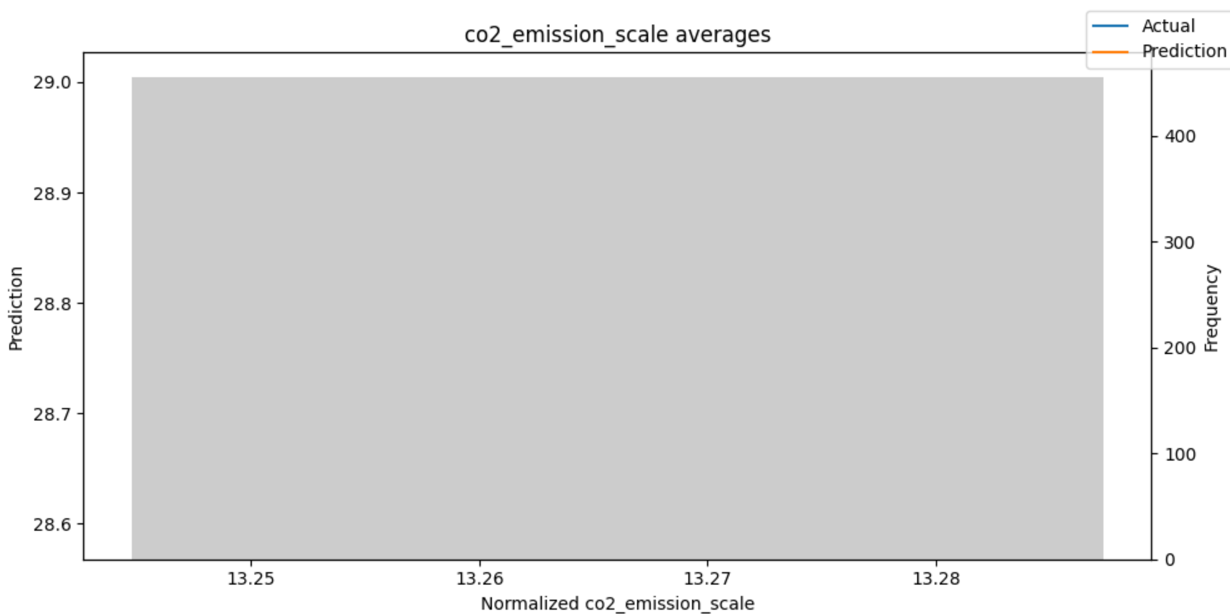

Fig 20. Robustness of co2_emission_scale averages.

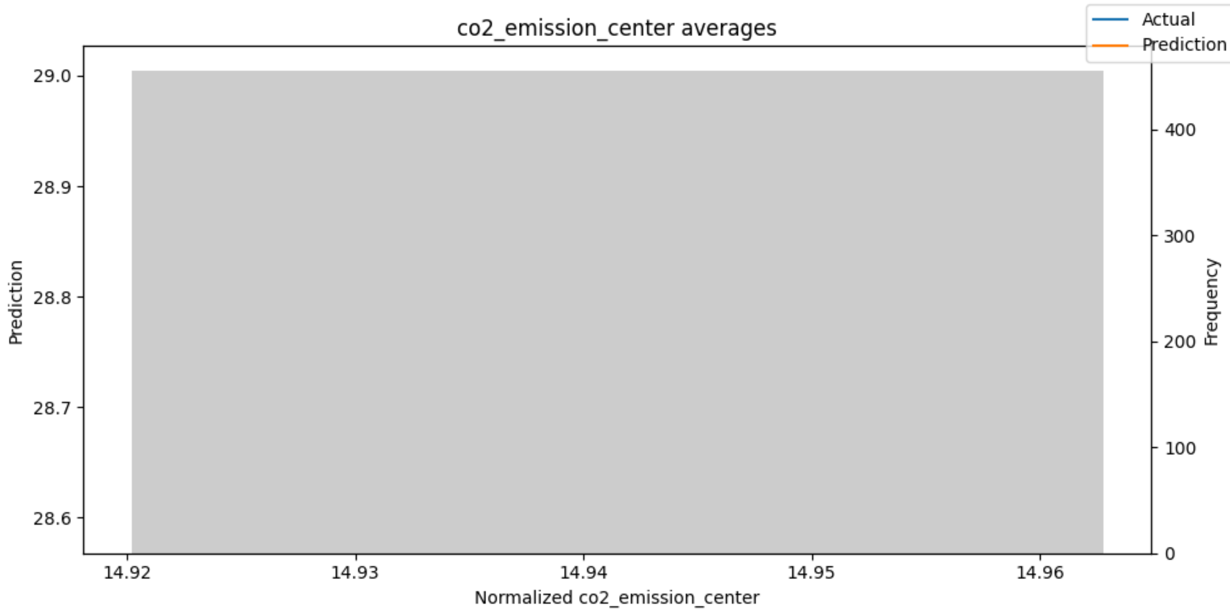

Fig 21. Robustness of co2_emission_center averages.

Encoder Length: There is a significant variance between the predictions and the actuals across different values of the normalized encoder length, with the model tending to underpredict for lower values and overpredict for higher values as seen in Fig 22. This pattern suggests a potential misalignment in the model's understanding of how the length of input sequences affects the target variable.

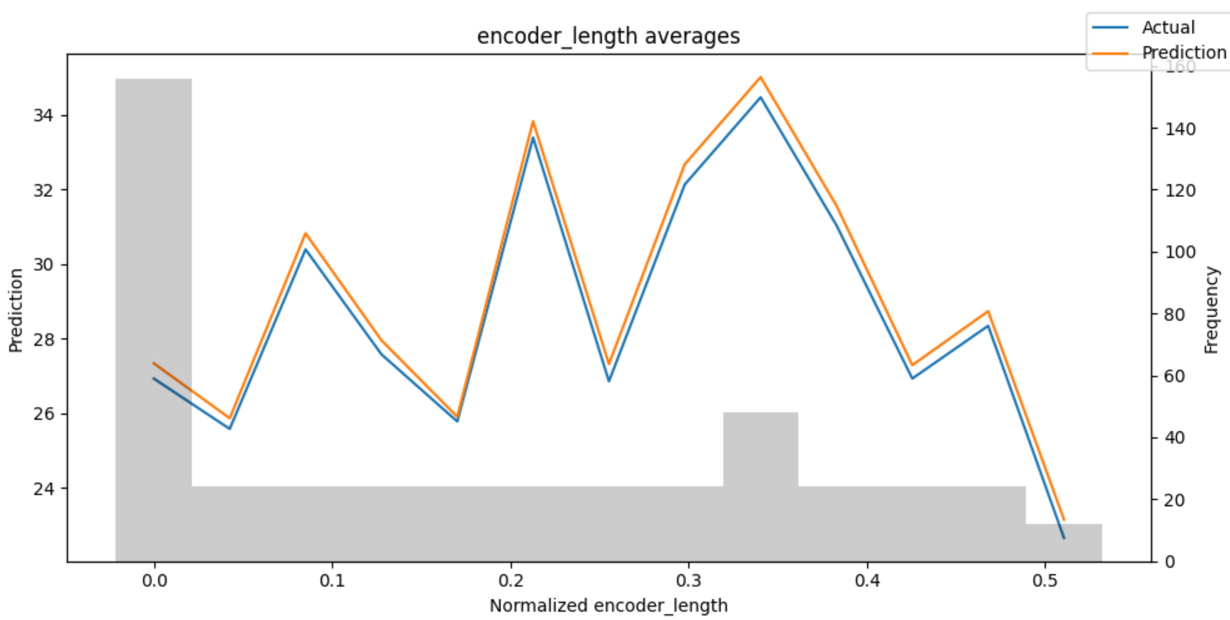

Fig 22. Robustness of encoder_length averages.

5.2. Overall Weaknesses:

The model may have difficulty with rare or extreme values across the various variables, as indicated by less frequent but more significant discrepancies.

Nonlinear relationships and abrupt trend changes seem to challenge the model's predictive capabilities.
The model may not be incorporating long-term dependencies or external factors affecting the data, which could be critical for accurate predictions.
A potential overfitting to the majority data distribution might be present, as indicated by a lack of sensitivity to less frequent data points.

6. Conclusion and future work

The successful application of the Temporal Fusion Transformer (TFT) to forecast carbon emissions in Bangladesh highlights its potential as a robust and versatile forecasting tool. By effectively capturing complex relationships and trends in the data, the TFT outperforms traditional time series forecasting methods and demonstrates its ability to provide accurate insights into future carbon emissions.
This breakthrough is particularly significant in the context of Bangladesh, where rapid economic growth and urbanization have led to significant increases in carbon emissions. The TFT's ability to handle diverse economic and social factors, along with its inherent interpretability, makes it an invaluable tool for policymakers and stakeholders seeking to mitigate climate change and promote sustainable development in Bangladesh.
As the TFT continues to evolve and gain widespread adoption, its impact on carbon emission forecasting is poised to extend beyond Bangladesh to a global scale. By providing accurate and timely projections, the TFT can empower governments and industries to make informed decisions that promote environmental sustainability and combat climate change effectively.
The demonstration of the TFT's capabilities in Bangladesh serves as a clear testament to the transformative potential of artificial intelligence in addressing pressing environmental challenges. As we strive to build a more sustainable future, the TFT and other AI-powered technologies will undoubtedly play a critical role in shaping our path towards a greener planet.

6.1. Future research directions include:

- Incorporate additional features or external data that might explain the discrepancies observed in the rare or extreme values.
- Enhance the model's capacity to handle nonlinear relationships, potentially by increasing model complexity or by feature engineering.
- Introduce regularization techniques or expand the training dataset to address overfitting and improve the model's generalizability.
- Investigate further into the data preprocessing steps, particularly normalization, to ensure that they are aligning well with the model's architecture and assumptions.
- Investigating the applicability of the Temporal Fusion Transformer to other countries and regions.
- Exploring the use of different data preprocessing techniques and feature engineering methods to further improve the model's performance.
- Developing ensemble models that combine the Temporal Fusion Transformer with other forecasting models to enhance predictive accuracy.

Reference

Ahmed, M. U., Karim, M. Al, Tahsin, M. S., Rahman, Y., & Tafannum, F. (2022). Analyzing Co2 Emission in Developing Countries Using Auto Regression and Auto Regression Walk Forward: A Time Series Approach. *2022 IEEE Delhi Section Conference (DELCON)*, 1–5. https://doi.org/10.1109/DELCON54057.2022.9752807

Althor, G., Watson, J. E. M., & Fuller, R. A. (2016). Global mismatch between greenhouse gas emissions and the burden of climate change. *Scientific Reports*, *6*(1), 20281. https://doi.org/10.1038/srep20281

Bazionis, I. K., & Georgilakis, P. S. (2021). Review of Deterministic and Probabilistic Wind Power Forecasting: Models, Methods, and Future Research. *Electricity*, *2*(1), 13–47. https://doi.org/10.3390/electricity2010002

Cui, T., Shi, Y., Lv, B., Ding, R., & Li, X. (2023). Federated learning with SARIMA-based clustering for carbon emission prediction. *Journal of Cleaner Production*, *426*, 139069. https://doi.org/10.1016/j.jclepro.2023.139069

Danish, & Ulucak, R. (2021). Renewable energy, technological innovation and the environment: A novel dynamic auto-regressive distributive lag simulation. *Renewable and Sustainable Energy Reviews*, *150*, 111433. https://doi.org/10.1016/j.rser.2021.111433

Deutch, J. (2017). Decoupling Economic Growth and Carbon Emissions. *Joule*, *1*(1), 3–5. https://doi.org/10.1016/j.joule.2017.08.011

Dou, J., Yunus, A. P., Merghadi, A., Shirzadi, A., Nguyen, H., Hussain, Y., Avtar, R., Chen, Y., Pham, B. T., & Yamagishi, H. (2020). Different sampling strategies for predicting landslide susceptibilities are deemed less consequential with deep learning. *Science of The Total Environment*, *720*, 137320. https://doi.org/10.1016/j.scitotenv.2020.137320

Dou, J., Yunus, A. P., Tien Bui, D., Merghadi, A., Sahana, M., Zhu, Z., Chen, C.-W., Khosravi, K., Yang, Y., & Pham, B. T. (2019). Assessment of advanced random forest and decision tree algorithms for modeling rainfall-induced landslide susceptibility in the Izu-Oshima Volcanic Island, Japan. *Science of The Total Environment*, *662*, 332–346. https://doi.org/10.1016/j.scitotenv.2019.01.221

Emami Javanmard, M., & Ghaderi, S. F. (2022). A Hybrid Model with Applying Machine Learning Algorithms and Optimization Model to Forecast Greenhouse Gas Emissions with Energy Market Data. *Sustainable Cities and Society*, *82*, 103886. https://doi.org/10.1016/j.scs.2022.103886

Fang, D., Zhang, X., Yu, Q., Jin, T. C., & Tian, L. (2018). A novel method for carbon dioxide emission forecasting based on improved Gaussian processes regression. *Journal of Cleaner Production*, *173*, 143–150. https://doi.org/10.1016/j.jclepro.2017.05.102

Ganzenmüller, R., Pradhan, P., & Kropp, J. P. (2019). Sectoral performance analysis of national greenhouse gas emission inventories by means of neural networks. *Science of The Total Environment*, *656*, 80–89. https://doi.org/10.1016/j.scitotenv.2018.11.311

Global Carbon Atlas. (n.d.). FOUNDATION BNP PARIBAS.

González Costa, J. J., Reigosa, M. J., Matías, J. M., & Covelo, E. F. (2017). Soil Cd, Cr, Cu, Ni, Pb and Zn sorption and retention models using SVM: Variable selection and competitive model. *Science of The Total Environment*, *593–594*, 508–522. https://doi.org/10.1016/j.scitotenv.2017.03.195

He, R., Zhang, L., & Chew, A. W. Z. (2022). Modeling and predicting rainfall time series using seasonal-trend decomposition and machine learning. *Knowledge-Based Systems*, *251*, 109125. https://doi.org/10.1016/j.knosys.2022.109125

Holt, C. C. (2004). Forecasting seasonals and trends by exponentially weighted moving averages. *International Journal of Forecasting*, *20*(1), 5–10. https://doi.org/10.1016/j.ijforecast.2003.09.015

Hong, Y., Shen, R., Cheng, H., Chen, Y., Zhang, Y., Liu, Y., Zhou, M., Yu, L., Liu, Y., & Liu, Y. (2019). Estimating lead and zinc concentrations in peri-urban agricultural soils through reflectance spectroscopy: Effects of fractional-order derivative and random forest. *Science of The Total Environment*, *651*, 1969–1982. https://doi.org/10.1016/j.scitotenv.2018.09.391

Lau, E. T., Yang, Q., Forbes, A. B., Wright, P., & Livina, V. N. (2014). Modelling carbon emissions in electric systems. *Energy Conversion and Management*, *80*, 573–581. https://doi.org/10.1016/j.enconman.2014.01.045

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. https://doi.org/10.1016/j.ijforecast.2021.03.012

Liu, D., Mishra, A. K., Yu, Z., Yang, C., Konapala, G., & Vu, T. (2017). Performance of SMAP, AMSR-E and LAI for weekly agricultural drought forecasting over continental United States. *Journal of Hydrology*, *553*, 88–104. https://doi.org/10.1016/j.jhydrol.2017.07.049

Liu, H., & Lin, B. (2017). Energy substitution, efficiency, and the effects of carbon taxation: Evidence from China's building construction industry. *Journal of Cleaner Production*, *141*, 1134–1144. https://doi.org/10.1016/j.jclepro.2016.09.119

Liu, Z., Deng, Z., Davis, S. J., Giron, C., & Ciais, P. (2022). Monitoring global carbon emissions in 2021. *Nature Reviews Earth & Environment*, *3*(4), 217–219. https://doi.org/10.1038/s43017-022-00285-w

Meng, F., & Dou, R. (2023). Prophet-LSTM-BP Ensemble Carbon Trading Price Prediction Model. *Computational Economics*. https://doi.org/10.1007/s10614-023-10384-5

Niu, D., Wang, K., Wu, J., Sun, L., Liang, Y., Xu, X., & Yang, X. (2020). Can China achieve its 2030 carbon emissions commitment? Scenario analysis based on an improved general regression neural network. *Journal of Cleaner Production*, *243*, 118558. https://doi.org/10.1016/j.jclepro.2019.118558

Pham, B. T., Nguyen, M. D., Dao, D. Van, Prakash, I., Ly, H.-B., Le, T.-T., Ho, L. S., Nguyen, K. T., Ngo, T. Q., Hoang, V., Son, L. H., Ngo, H. T. T., Tran, H. T., Do, N. M., Van Le, H., Ho, H. L., & Tien Bui, D. (2019). Development of artificial intelligence models for the prediction of Compression Coefficient of soil: An application of Monte Carlo sensitivity analysis. *Science of The Total Environment*, *679*, 172–184. https://doi.org/10.1016/j.scitotenv.2019.05.061

Pradhan, P., Tingsanchali, T., & Shrestha, S. (2020). Evaluation of Soil and Water Assessment Tool and Artificial Neural Network models for hydrologic simulation in different climatic regions of Asia. *Science of The Total Environment*, *701*, 134308. https://doi.org/10.1016/j.scitotenv.2019.134308

Primandari, A. H., Thalib, A. K., & Kesumawati, A. (2022). Analysis of Changes in Atmospheric CO2 Emissions Using Prophet Facebook. *Enthusiastic : International Journal of Applied Statistics and Data Science*, 1–9. https://doi.org/10.20885/enthusiastic.vol2.iss1.art1

Qiao, W., Lu, H., Zhou, G., Azimi, M., Yang, Q., & Tian, W. (2020). A hybrid algorithm for carbon dioxide emissions forecasting based on improved lion swarm optimizer. *Journal of Cleaner Production*, *244*, 118612. https://doi.org/10.1016/j.jclepro.2019.118612

Rodriguez-Galiano, V., Mendes, M. P., Garcia-Soldado, M. J., Chica-Olmo, M., & Ribeiro, L. (2014). Predictive modeling of groundwater nitrate pollution using Random Forest and multisource variables related to intrinsic and specific vulnerability: A case study in an agricultural setting (Southern Spain). *Science of The Total Environment*, *476–477*, 189–206. https://doi.org/10.1016/j.scitotenv.2014.01.001

Schmidt, A., Creason, W., & Law, B. E. (2018). Estimating regional effects of climate change and altered land use on biosphere carbon fluxes using distributed time delay neural networks with Bayesian regularized learning. *Neural Networks*, *108*, 97–113. https://doi.org/10.1016/j.neunet.2018.08.004

Sen, P., Roy, M., & Pal, P. (2016). Application of ARIMA for forecasting energy consumption and GHG emission: A case study of an Indian pig iron manufacturing organization. *Energy*, *116*, 1031–1038. https://doi.org/10.1016/j.energy.2016.10.068

Shaw, J. T., Allen, G., Topping, D., Grange, S. K., Barker, P., Pitt, J., & Ward, R. S. (2022). A case study application of machine-learning for the detection of greenhouse gas emission sources. *Atmospheric Pollution Research*, *13*(10), 101563. https://doi.org/10.1016/j.apr.2022.101563

Sun, W., & Liu, M. (2016). Prediction and analysis of the three major industries and residential consumption CO2 emissions based on least squares support vector machine in China. *Journal of Cleaner Production*, *122*, 144–153. https://doi.org/10.1016/j.jclepro.2016.02.053

Wang, H. (2023). Analysis on influencing factors of carbon emissions from China's pulp and paper industry and carbon peaking prediction. *Environmental Science and Pollution Research*, *30*(37), 86790–86803. https://doi.org/10.1007/s11356-023-28483-z

Wang, Q., Li, S., & Pisarenko, Z. (2020). Modeling carbon emission trajectory of China, US and India. *Journal of Cleaner Production*, *258*, 120723. https://doi.org/10.1016/j.jclepro.2020.120723

Wu, Y., & Xu, B. (2022). When will China's carbon emissions peak? Evidence from judgment criteria and emissions reduction paths. *Energy Reports*, *8*, 8722–8735. https://doi.org/10.1016/j.egyr.2022.06.069

Wu, Z., Zhou, Y., Wang, H., & Jiang, Z. (2020). Depth prediction of urban flood under different rainfall return periods based on deep learning and data warehouse. *Science of The Total Environment*, *716*, 137077. https://doi.org/10.1016/j.scitotenv.2020.137077

Xu, B., & Lin, B. (2016). Reducing carbon dioxide emissions in China's manufacturing industry: a dynamic vector autoregression approach. *Journal of Cleaner Production*, *131*, 594–606. https://doi.org/10.1016/j.jclepro.2016.04.129

Yang, F., Shi, L., & Gao, L. (2023). Probing CO2 emission in Chengdu based on STRIPAT model and Tapio decoupling. *Sustainable Cities and Society*, *89*, 104309. https://doi.org/10.1016/j.scs.2022.104309

Yang, H., & O'Connell, J. F. (2020). Short-term carbon emissions forecast for aviation industry in Shanghai. *Journal of Cleaner Production*, *275*, 122734. https://doi.org/10.1016/j.jclepro.2020.122734

Yang, M., Mou, Y., Meng, Y., Liu, S., Peng, C., & Zhou, X. (2020). Modeling the effects of precipitation and temperature patterns on agricultural drought in China from 1949 to 2015. *Science of The Total Environment*, *711*, 135139. https://doi.org/10.1016/j.scitotenv.2019.135139

Ethical approval

Not applicable.

Consent to participate

Not applicable.

Consent for publication

Not applicable.

Data Availability Statement

Data will be made public upon request.

Author contribution

Shuvro Ahmed: Methodology, Model Architecture, Result, Model Performance Across Variables. **Neda Firoz:** Abstract, Literature Review, Work Flow, Discussion, Overall Weakness. **Mohammad Sadman Tahsin:** Introduction, Data Collection, Conclusion and Future Work.

Funding

Not applicable.

Competing Interests

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.