Zero-shot learning for predicting transportation traffic volume

Roshan Kotian*[1], Feixiong Liao [2]

[1] Assistant Professor, Urban Planning and Transportation Group, TU/e, The Netherlands

[2] Assistant Professor, Urban Planning and Transportation Group, TU/e, The Netherlands

SHORT SUMMARY

Traffic prediction is essential for the success of Intelligent Transport Systems (ITS), as accurate and reliable traffic information directly impacts stakeholders' ability to make informed decisions regarding route selection. A range of statistical and deep learning methods have been employed for these predictions; however, these approaches are often time-consuming due to the extensive training, validation, and testing required on historical datasets. Zero-shot learning—a category of machine learning algorithms—has emerged to overcome these challenges. Unlike conventional models, zero-shot techniques are pre-trained on extensive historical time series data from diverse domains, so they do not require manual training. This study evaluates the prediction accuracy of various zero-shot learning models with statistical and deep learning models. Through comprehensive experiments involving seven types of machine learning models and multiple frequencies of time series data, our findings reveal that zero-shot learning models excel in predicting traffic volume.

Keywords: Zero-shot learning, Autoencoders, Neural Networks, Transformer, Transportation

1. INTRODUCTION

Background

Traffic congestion is a significant issue with increasingly negative impacts such as unreliable travel times and increased fuel consumption and emissions. Therefore, predicting traffic conditions accurately greatly enhances network capacity utilization and alleviates congestion by enabling traffic management personnel to manage traffic more efficiently.

Most publicly available traffic volume and congestion data are time series data, which consist of a sequence of data points measured and recorded at specific time intervals. The digital revolution across various domains, such as transportation and e-commerce, has led to an explosion of different types of data, particularly time series data. This data often exhibits complex patterns that classical models, like Autoregressive Integrated Moving Average(ARIMA), may struggle to analyze effectively, especially when detecting drift, abrupt changes, and outliers (Akhtar & Moridpour, 2021).

Machine learning models have demonstrated remarkable capabilities in understanding the relationships between successive data points to address traditional models' limitations and better capture the data's non-linearities.

Literature Review

Numerous machine learning models with different architectures, including Graph Neural Networks (Liu et al., 2024), LSTMs (Gopali et al., 2024), Transformers (Wen et al., 2022), and encoder-decoder models (Das et al., 2023) have been proposed. However, many of these models require lengthy and computationally intensive training and validation processes, as illustrated in Figure 1. This complexity can make it daunting for non-experts to adopt machine-learning techniques.

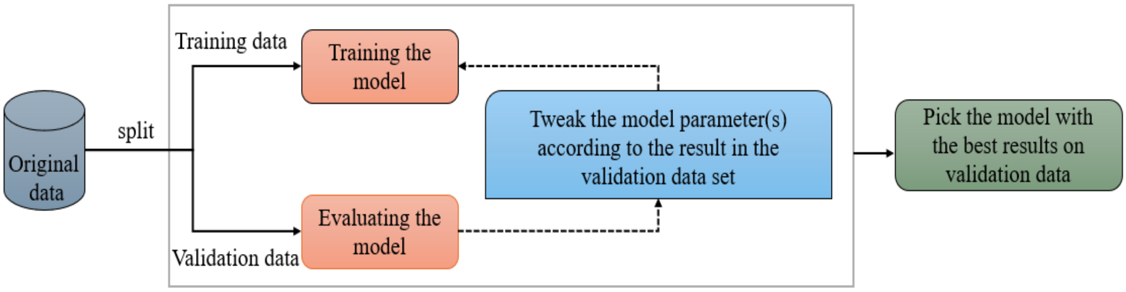

Figure 1: General Training and Validation of deep learning models

Zero-shot learning has emerged as a method to expedite the prediction process (Ye et al., 2024). These models have been pre-trained on large datasets that are publicly available and exhibit a variety of data distributions and statistical properties. The term "zero-shot" refers to the capability of these models to predict future data points without the need for lengthy and computationally intensive training and validation phases.

This innovation empowers non-AI professionals to deploy models quickly and easily without the burden of lengthy training and validation processes (Radford et al., 2021).We believe that insufficient research has been conducted comparing the performance of zero-shot models in predicting traffic flow within the transportation domain.

To this end, our paper addresses the main research question: "Do zero-shot learning models provide better predictive accuracy for traffic flow than traditional statistical and deep learning models with default hyperparameters? If so, to what extent are they better or worse than their counterparts?"

The remainder of this paper is organized as follows: Section 2 describes the dataset, and the models used in the experiments. It also includes information on the training and testing procedures, as well as the hardware configuration utilized for running the models. Section 3 presents a detailed analysis of the results for various models across different datasets. Finally, Section 4 summarizes the overall performance of the zero-shot learning models and discusses potential strategies for improving model performance in cases of high kurtosis and skewness in the data.

2. METHODOLOGY

This section provides an overview of the dataset, the training and testing of the machine learning algorithms, and the hardware configuration used in the experiments.

Dataset

We utilized two popular datasets for our analysis:

Swiss Tunnel Data: This dataset contains 747 data points, each representing the daily traffic volume recorded in Baregg, Switzerland (Galit Shmueli & Kenneth C. Lichtendahl Jr et al., 2016). The data were collected over two years, from 2003 to 2004.

Kaggle Data: This dataset includes approximately 48,120 data points across four road junctions. The traffic volume at each intersection was recorded hourly for three years, from 2015 to 2017. It is worth noting that all years except 2017 feature data from three junctions. Statistical descriptions of both datasets are provided in Tables 1 and 2, respectively (Kaggle, 2020).

Table 1: Swiss Tunnel Data Description

Seasonality	Std. Dev	Variance	Kurtosis	Skew	Stationarity
TRUE	12456.35	155160667	0.38	-0.57	FALSE

Table 2: Kaggle Data Description

Kaggle Data 2015, total data points=4392						
Junction	**Seasonality**	**Std. Dev**	**Variance**	**Kurtosis**	**Skew**	**Stationarity**
1	True	7,78	60,60	-0,28	0,51	True
2	True	3,33	11,10	26,55	3,25	True
3	True	4,35	18,93	14,85	3,26	False
Kaggle Data 2016, total data points=26352						
1	True	16,82	283,06	-0,22	0,62	True
2	True	4,29	18,42	-0,40	0,31	True
3	True	9,71	94,47	24,57	3,55	True
Kaggle Data 2017, total data points=17376						
1	True	22,84	522,08	-0,61	0,36	True
2	True	8,25	68,18	-0,37	0,54	True
3	True	11,42	130,45	33,64	3,83	True
4	True	3,52	12,4	4,73	1,33	True

Models

To address the research question mentioned above in section 2.1 we have chosen the following three sets of models as shown in Table 3.

Table 3: Three types of models

Zero-shot	Statistical	Deep Learning
TimeGPT Chronos_tiny	ETS Seasonal Naïve	PatchTST Simple FF NN TiDE

TimeGPT is the first foundation model specifically designed for time series data and is based on the self-attention mechanism proposed by Vaswani (Vaswani et al., 2017). It features a multi-layer encoder-decoder architecture, each incorporating residual connections and layer normalization. TimeGPT has been trained on an extensive dataset containing 100 billion data points (Garza et al., 2023).

Chronos_tiny is one of the pre-trained time series forecasting models developed by Amazon. This model first transforms a time series into a sequence of tokens through normalization and quantization techniques. Following this transformation, a language model is trained on the normalized and quantized tokens. Chronos has been pre-trained on 55 publicly available datasets, including the popular Monash Time Series and the M-Competition datasets (Ansari et al., 2024).

ETS (Exponential Smoothing with Trend and Seasonality) can identify general patterns in the data and can be extended to account for trends and seasonal variations. Exponential smoothing assigns greater weight to recent observations while progressively reducing the influence of older data as the distance from the current data points increases (Ryan Tibshirani, 2023).

Seasonal Naïve is one of the simplest forecast models that employs the concept of random walk. In this model, the one-time-step ahead forecast value is equal to the most recent past value (Ivan Svetunkov, 2023).

Simple Feed Forward Neural Network (Simple FF NN) is a basic neural network consisting of three main layers: the input, the hidden, and the output. Each layer contains three key components: nodes (or units), weighted connections, and activation functions. In this type of network, information flows in a single direction, from the input layer to the output layer. The learning process involves backpropagation, which adjusts the initial weights in the network to minimize prediction errors (Shaygan et al., 2022).

PatchTST (Patch Time Series Transformer) is a transformer-based forecast model that segments time series data into different patches and employs channel independence. It is recognized for its superior accuracy in long-horizon forecasting while reducing computational and memory demands (Nie et al., 2023).

TiDE (Time Series Dense Encoder) is a multilayer perceptron encoder-decoder model for long-horizon time series forecasting. It combines the simplicity and speed of linear models, achieving optimal error rates for linear dynamical systems (Das et al., 2023).

For the zero-shot model, we chose TimeGPT and Chronos as they are open-sourced, easy to use, and pre-trained on 100 billion data points while others such as Lag-Llama (Rasul et al., 2023) and GTT (Feng et al., 2024) are pre-trained on 0.3 and 2.4 billion data points.

Most research papers on traffic prediction concentrate on Recurrent Neural Networks (RNNs) and Graph Neural Networks (GNNs). However, only a few studies examine the performance of transformer-based models. This work addresses this gap by providing empirical evidence on how well transformer-based models perform.

ML Training and Testing

98% of all datasets are utilized for training. The model's performance in forecasting the remaining 2% of data points in the time series is evaluated using Mean Absolute Error (MAE) and Root Mean Square Error (RMSE).

Since our goal is to predict time series with large values, particularly in the Swiss Tunnel data, we employ MAE and RMSE to assess the accuracy of the machine learning models. Additionally, because the datasets exhibit abrupt changes and outliers, MAE and RMSE are preferred metrics since they are less sensitive to outliers and perform well when future values vary (Rob J Hyndman & George Athanasopoulos, 2021).

All models were executed with the default hyperparameter settings provided in the AutoGluon framework on an Intel i7 2.6 GHz CPU with 64 GB of RAM. To ensure reproducibility, all the code for our experiments has been made available on GitHub along with the hyperparameters used in each of the models https://github.com/rkoti/Zero-shot_traffic_Volume.

3. RESULTS AND DISCUSSION

In this section, we analyze the prediction accuracy of the models in three different combinations as shown in Table 4 below.

Table 4: Evaluation schema of various models

Models	**Models**
Zero-shot	Zero-shot Statistical Deep Learning

We present the RMSE and MAE for all models across all datasets and compare the percentage difference between these metrics.

Swiss Tunnel Data

Figure 2 provides the overall metric values for all the models. The x-axis provides the MAE and RMSE scores for all the models in the y-axis.

Figure 2: RMSE AND MAE scores for all the models

Zero-shot vs. Zero-shot

As shown in Figure 3, the Chronos_tiny model has RMSE and MAE scores of 23% and 7% higher than the TimeGPT model (indicated by the red arrow and red background). Since lower scores for these metrics indicate better performance, it is evident that the TimeGPT model outperforms the others.

	RMSE	MAE
Model	TimeGPT	TimeGPT
Chronos_tiny	23%	7%

Figure 3: Comparison of zero-shot models by their percentage difference between the two metrics

Zero-shot vs. Statistical and other Deep Learning models

When comparing zero-shot models to statistical models in Figure 4, we observe that zero-shot models significantly outperform the latter, as indicated by the lower RMSE and MAE scores highlighted by the green arrows and green background. Additionally, zero-shot models excel beyond other deep-learning models, as demonstrated in Figure 5.

RMSE		
Model	ETS	Seasonal Naïve
Chronos_tiny	7%	15%
TimeGPT	30%	38%

MAE		
Model	ETS	Seasonal Naïve
Chronos_tiny	9%	12%
TimeGPT	16%	19%

Figure 4: Comparison of zero-shot and statistical models by their percentage difference of the two metrics

RMSE

Model	PatchTST	Neural Network	TiDE
Chronos_tiny	13%	1%	16%
TimeGPT	34%	24%	39%

MAE

Model	PatchTST	Neural Network	TiDE
Chronos_tiny	16%	8%	26%
TimeGPT	23%	15%	34%

Figure 5: Comparison of zero-shot and deep learning models by their percentage difference of the two metrics

Kaggle Data

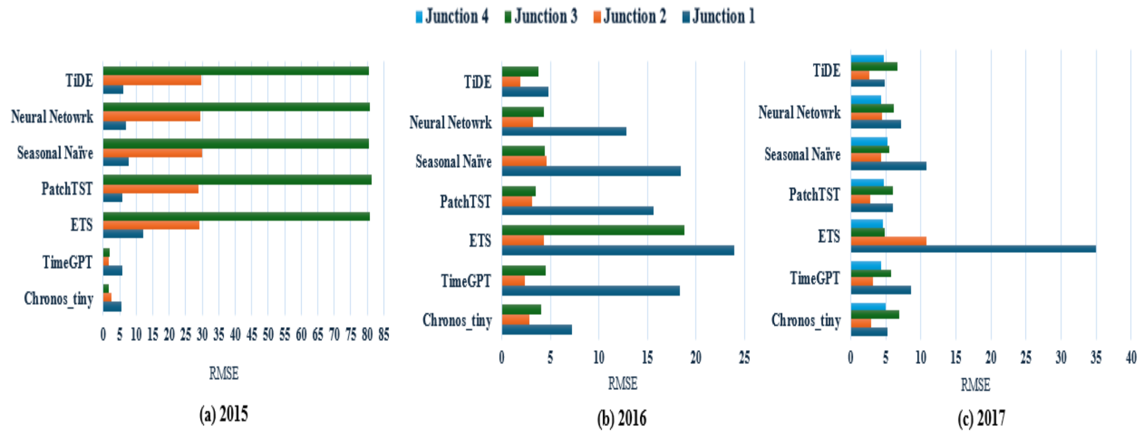

Figure 6: RMSE score for various junctions

Figures 6 and 7 show the overall RMSE and MAE scores along the x-axis for all models and junctions. To evaluate the Kaggle dataset, we calculated the average RMSE and MAE for each junction by year.

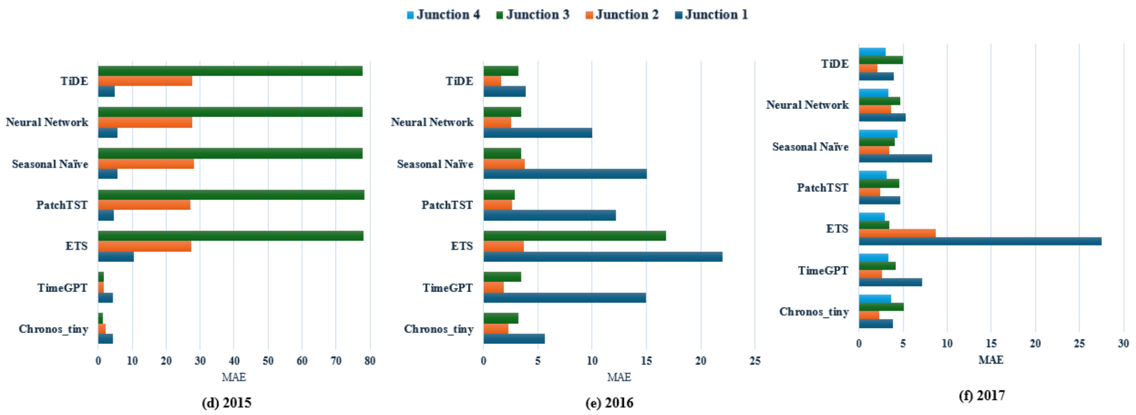

Figure 7: MAE score for various junctions

Zero-shot vs. Zero-shot

2015, RMSE	
Model	TimeGPT
Chronos_tiny	↑ 1%

2015, MAE	
Model	TimeGPT
Chronos_tiny	↑ 2,30%

2016, RMSE	
Model	TimeGPT
Chronos_tiny	↓ 57%

2016, MAE	
Model	TimeGPT
Chronos_tiny	↓ 58%

2017, RMSE	
Model	TimeGPT
Chronos_tiny	↓ 8%

2017, MAE	
Model	TimeGPT
Chronos_tiny	↓ 15%

Figure 8: Comparison of zero-shot models by their percentage difference between the two metrics

From Figure 8, we can infer that although Chronos_tiny has a marginally higher error rate for the year 2015, it outshines the TimeGPT model for the two consecutive years of 2016 and 2017.

Zero-shot vs. Statistical models

Figure 9 shows that zero-shot models have a lower average error score than statistical models. Chronos_tiny demonstrates relatively better performance.

2015, RMSE		
Model	ETS	Seasonal Naïve
Chronos_tiny	↓ 170%	↓ 169%
TimeGPT	↓ 171%	↓ 170%

2015, MAE		
Model	ETS	Seasonal Naïve
Chronos_tiny	↓ 175%	↓ 174%
TimeGPT	↓ 175%	↓ 174%

2016, RMSE		
Model	ETS	Seasonal Naïve
Chronos_tiny	↓ 107%	↓ 64%
TimeGPT	↓ 60%	↓ 9%

2016, MAE		
Model	ETS	Seasonal Naïve
Chronos_tiny	↓ 117%	↓ 66%
TimeGPT	↓ 69%	↓ 9%

2017, RMSE		
Model	ETS	Seasonal Naïve
Chronos_tiny	↓ 93%	↓ 25%
TimeGPT	↓ 86%	↓ 17%

2017, MAE		
Model	ETS	Seasonal Naïve
Chronos_tiny	↓ 97%	↓ 30%
TimeGPT	↓ 84%	↓ 15%

Figure 9: Comparison of zero-shot and statistical models by their percentage difference of the two metrics

Zero-shot vs. other Deep Learning models

From Figure 10, we can infer that in 2015, the average error rate of the zero-shot models across all junctions was lower than that of their counterparts. However, in 2016 and 2017, these models experienced a higher error rate compared to other models, including both statistical and deep learning approaches. Specifically, TimeGPT showed a notably higher error rate than other models, particularly in 2017.

2015, RMSE

Model	PatchTST	Neural Network	TiDE
Chronos_tiny	169%	169%	169%
TimeGPT	169%	169%	169%

2015, MAE

Model	PatchTST	Neural Network	TiDE
Chronos_tiny	173%	174%	173%
TimeGPT	174%	174%	174%

2016, RMSE

Model	PatchTST	Neural Network	TiDE
Chronos_tiny	45%	36%	28%
TimeGPT	12%	21%	82%

2016, MAE

Model	PatchTST	Neural Network	TiDE
Chronos_tiny	45%	36%	25%
TimeGPT	14%	23%	80%

2017, RMSE

Model	PatchTST	Neural Network	TiDE
Chronos_tiny	3%	9%	7%
TimeGPT	11%	1%	15%

2017, MAE

Model	PatchTST	Neural Network	TiDE
Chronos_tiny	0,27%	14%	4%
TimeGPT	16%	2%	20%

Figure 10: Comparison of zero-shot and deep learning models by their percentage difference of the two metrics

This can be attributed to the data presented in Table 2 above, where the combination of higher variance, kurtosis, and skewness values diminishes the predictive power of the zero-shot models under consideration.

4. CONCLUSIONS

The cost of road congestion in Europe is estimated to be over €110 billion a year (Christidis, 2012). Therefore, it is crucial to utilize the transportation infrastructure efficiently to reduce congestion. One of the ways to solve the congestion problem is to predict the traffic flow accurately.

To this end, this paper examined the performance of zero-shot learning models compared to other statistical and deep learning models to predict traffic flow on hourly and daily data. These models enable non-domain experts to quickly leverage forecasting capabilities in their specific fields without the need for extensive computational resources.

Our experiments demonstrated that these models could generate fairly accurate traffic flow predictions compared to other established models. However, we observed that data characterized by high kurtosis and skewness sometimes performed slightly less accurately than the other deep-learning models. This suggests that these zero-shot models may require fine-tuning on fat-tailed datasets before they can be effectively deployed. With the availability of machine learning frameworks such as AutoGluon, the fine-tuning of the zero-shot models can be formed with ease even by non-AI transportation domain experts.

REFERENCES

Akhtar, M., & Moridpour, S. (2021). A Review of Traffic Congestion Prediction Using Artificial Intelligence. In *Journal of Advanced Transportation* (Vol. 2021). Hindawi Limited. https://doi.org/10.1155/2021/8878011

Ansari, A. F., Stella, L., Turkmen, C., Zhang, X., Mercado, P., Shen, H., Shchur, O., Rangapuram, S. S., Arango, S. P., Kapoor, S., Zschiegner, J., Maddix, D. C., Wang, H., Mahoney, M. W., Torkkola, K., Wilson, A. G., Bohlke-Schneider, M., & Wang, Y. (2024). *Chronos: Learning the Language of Time Series*.

Christidis, P. (2012). *Measuring road congestion*.

Das, A., Kong, W., Leach, A., Mathur, S., Sen, R., & Yu, R. (2023). *Long-term Forecasting with TiDE: Time-series Dense Encoder*.

Feng, C., Huang, L., & Krompass, D. (2024). *Only the Curve Shape Matters: Training Foundation Models for Zero-Shot Multivariate Time Series Forecasting through Next Curve Shape Prediction*.

Galit Shmueli, & Kenneth C. Lichtendahl Jr. (2016). *Practical Time Series Forecasting With R* (2nd Edition). Axelrod Schnall publishers.

Garza, A., Challu, C., & Mergenthaler-Canseco, M. (2023). *TimeGPT-1*. http://arxiv.org/abs/2310.03589

Gopali, S., Siami-Namini, S., Abri, F., & Namin, A. S. (2024). *The Performance of the LSTM-based Code Generated by Large Language Models (LLMs) in Forecasting Time Series Data*. http://arxiv.org/abs/2411.18731

Ivan Svetunkov. (2023). *Forecasting and Analytics with the Augmented Dynamic Adaptive Model (ADAM)* (1st edition). Chapman and Hall/CRC.

Kaggle. (2020). *Traffic Prediction Dataset*.

Liu, Y., Rasouli, S., Wong, M., Feng, T., & Huang, T. (2024). RT-GCN: Gaussian-based spatiotemporal graph convolutional network for robust traffic prediction. *Information Fusion*, *102*. https://doi.org/10.1016/j.inffus.2023.102078

Nie, Y., Nguyen, N. H., Sinthong, P., & Kalagnanam, J. (2023). A TIME SERIES IS WORTH 64 WORDS: LONG-TERM FORECASTING WITH TRANSFORMERS. *11th International Conference on Learning Representations, ICLR 2023*.

Radford, A., Kim, J. W., Hallacy, C., Ramesh, A., Goh, G., Agarwal, S., Sastry, G., Askell, A., Mishkin, P., Clark, J., Krueger, G., & Sutskever, I. (2021). Learning Transferable Visual Models From Natural Language Supervision. *Proceedings of Machine Learning Research*, *139*, 8748–8763.

Rasul, K., Ashok, A., Williams, A. R., Ghonia, H., Bhagwatkar, R., Khorasani, A., Bayazi, M. J. D., Adamopoulos, G., Riachi, R., Hassen, N., Biloš, M., Garg, S., Schneider, A., Chapados, N., Drouin, A., Zantedeschi, V., Nevmyvaka, Y., & Rish, I. (2023). *Lag-Llama: Towards Foundation Models for Probabilistic Time Series Forecasting*.

Rob J Hyndman, & George Athanasopoulos. (2021). *Forecasting Principle and Practice* (3rd Edition). OTexts.

Ryan Tibshirani. (2023). *Lecture 7: Exponential Smoothing With Trend and Seasonality Introduction to Time Series, Fall 2023 Ryan Tibshirani*. https://www.stat.berkeley.edu/~ryantibs/timeseries-f23/lectures/ets.pdf

Shaygan, M., Meese, C., Li, W., Zhao, X. (George), & Nejad, M. (2022). Traffic prediction using artificial intelligence: Review of recent advances and emerging opportunities. In *Transportation Research Part C: Emerging Technologies* (Vol. 145). Elsevier Ltd. https://doi.org/10.1016/j.trc.2022.103921

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., & Polosukhin, I. (2017). Attention is all you need. *Advances in Neural Information Processing Systems*, *2017-December*, 5999–6009.

Wen, Q., Zhou, T., Zhang, C., Chen, W., Ma, Z., Yan, J., & Sun, L. (2022). *Transformers in Time Series: A Survey*. http://arxiv.org/abs/2202.07125

Ye, J., Zhang, W., Yi, K., Yu, Y., Li, Z., Li, J., & Tsung, F. (2024). *A Survey of Time Series Foundation Models: Generalizing Time Series Representation with Large Language Model*. http://arxiv.org/abs/2405.02358