https://math-sci.ui.ac.ir

Mathematics and Society
ISSN (print): 2345-6493, ISSN (on-line): 2345-6507
Vol. 11 No. 2 (2026), pp. 97-128.

https://www.ui.ac.ir

TRANSFORMERS IN MULTIVARIATE TIME SERIES FORECASTING: A REVIEW

ESMAEIL CHITGAR, MAHDIEH HASHEMINEZHAD* AND SEPEHR EBRAHIMI MOOD

Abstract. Long-term forecasting of multivariate time series is a fundamental challenge in the field of machine learning, with critical applications in numerous domains such as energy, transportation, and financial markets. The inherent complexity of this data, stemming from seasonal patterns, non-stationary trends, and interdependencies among variables, has constrained traditional forecasting methods. The importance of addressing this problem is evident in strategic decision-making, as the accuracy and efficiency of predictions directly impact energy resource management, traffic control, and financial risk analysis.

This review article provides a comprehensive framework for analyzing state-of-the-art Transformer-based architectures in the domain of multivariate time series forecasting. The proposed framework is structured around three main aspects: (1) quantitative performance evaluation of models on standard datasets, (2) analysis of computational considerations including time and memory complexity, and (3) examination of the models' capability in handling high-frequency time series. Accordingly, prominent models such as Autoformer, iTransformer, GTformer, FEDformer, ETSformer, and Pathformer are reviewed and comparatively analyzed.

The findings of this review indicate that recent innovations, including the application of frequency-domain processing, auto-correlation mechanisms, graph-based learning, and multi-scale architectures, have simultaneously led to improvements in forecasting accuracy and computational efficiency. These advancements outline a clear research path for the development of future architectures with a focus on greater interpretability, better scalability, and broader generalization capabilities.

Keywords: Transformer, Multivariate, Encoder, Decoder, Time Series.

Article Type: Review Paper.

Communicated by Khalil Shafie.

Received: 14-04-2025, Accepted: 17-11-2025, Published Online: 28-01-2026.

Cite this article: E. Chitgar, M. Hasheminezhad and S. Ebrahimi Mood, Transformers in multivariate time series forecasting: a review, *Mathematics and Society*, **11** no. 2 (2026) 97–128. https://dx.doi.org/10.22108/msci.2025.144922.1740 .

1. Introduction

Multivariate time series forecasting represents a fundamental challenge in machine learning, with essential applications in fields such as energy systems, transportation networks, and financial markets. The complexity of such data arises from seasonal patterns, non-stationary behaviors, and the intricate relationships among variables, which often limit the performance of traditional forecasting methods. Understanding and effectively solving this challenge is crucial for strategic decision-making because improvements in prediction accuracy and computational efficiency directly influence energy management, traffic optimization, and financial risk assessment.

The transformer architecture has emerged as one of the most important deep learning models specifically designed for sequence processing. Initially developed for machine translation, transformers demonstrate high capability in learning long-term dependencies due to their use of attention mechanisms, and have been rapidly employed in various domains including natural language processing, computer vision, and more recently in time series forecasting. Key applications of transformers include machine translation, text summarization, computer vision, and time series forecasting, where they show high efficiency due to their ability to understand long-term and cross-variable dependencies.

However, the application of transformers to multivariate time series forecasting faces significant challenges such as sensitivity to noise and outliers, high computational costs for long sequences, limited interpretability, and the difficulty of modeling complex nonlinear relationships. To overcome these obstacles and enhance prediction accuracy and efficiency, numerous research studies have been conducted in recent years, leading to the emergence and development of diverse and innovative transformer architectures.

This review article provides a comprehensive framework for analyzing advanced transformer-based architectures in multivariate time series forecasting. The proposed framework is structured around three main axes: quantitative performance evaluation of models based on standard datasets, analysis of computational considerations including time complexity, and examination of model capabilities in handling high-frequency time series. Accordingly, prominent models such as Autoformer, iTransformer, GTformer, FEDformer, ETSformer, and Pathformer are examined and compared.

This review systematically examines transformer-based models for multivariate time series forecasting, covering architectural fundamentals, key innovations, and comprehensive performance evaluation. The analysis concludes with future research directions in this rapidly evolving field.

2. Main Results

Our analysis reveals significant advancements in transformer-based approaches for multivariate time series forecasting. This section presents a multi-dimensional evaluation framework designed

https://dx.doi.org/10.22108/msci.2025.144922.1740

to rigorously assess these models beyond conventional accuracy metrics, addressing critical gaps in existing literature.

2.1. **Comprehensive Evaluation Framework.** Our evaluation framework systematically assesses transformer models across three critical dimensions: predictive accuracy on standard benchmarks, computational efficiency, and performance on high-frequency time series data. This multi-faceted approach provides a more holistic understanding of model capabilities and limitations than traditional evaluation methods that focus solely on prediction accuracy. In most research studies, transformer-based models are investigated and compared mainly by using error metrics on standard datasets and models are ranked based on these results. Despite the value of such comparisons, relying solely on these criteria is insufficient for practical model selection. Numerous factors including computational considerations, architectural features of models, and also their performance in specific scenarios such as processing high-frequency time series or different sampling patterns play a determining role in the final selection.

Three main criteria are considered for evaluating models: numerical performance on standard datasets, computational efficiency, and performance evaluation on high-frequency time series. Finally, by synthesizing these results, we provide a roadmap for selecting models appropriate to specific application conditions.

2.2. **Quantitative Performance Analysis.** Comparative evaluation of time series forecasting models based on standard and well-known datasets enables fair and systematic comparison between different methods. Benchmark datasets that have been widely used in previous studies including the Electricity Transformer Temperature (ETT) dataset (hourly and minute-level), the Electricity Consumption (ECL) dataset, the Traffic dataset, the Weather dataset, the Influenza-like Illness (ILI) dataset, the Exchange Rate dataset, and the Solar Energy dataset.

For evaluating model performance, two common metrics, Mean Squared Error (MSE) and Mean Absolute Error (MAE) are generally used. Lower values in both metrics indicate higher model accuracy. The performance of selected transformer models that have been trained and evaluated on a subset of standard datasets is summarized in the comparison table. These data have been extracted from reliable sources and used under multivariate forecasting scenarios with different time horizons. For each dataset, values of both metrics have been calculated and their averages have been reported as final performance indices.

Evaluations have been conducted for four different forecasting horizons. The best performance for each horizon is indicated by underlining. Based on the overall average errors reported in the table, GTformer and Pathformer models show the best performance among the examined models with the lowest squared error and absolute error values, respectively. In contrast, the Autoformer

https://dx.doi.org/10.22108/msci.2025.144922.1740

model shows the weakest performance with the highest error values across both measurement criteria. Other models including iTransformer, FEDformer, and ETSformer are ranked in the middle positions. However, this overall ranking tells only part of the story and for a complete understanding of each model's capabilities, more precise examination based on different criteria is necessary.

A more complete analysis of the obtained results includes: Overall prediction accuracy, Long-term forecasting, Computational efficiency, Interpretability, High-dimensional data and interdependencies, and Learning multi-scale and seasonal patterns. GTformer and Pathformer have the lowest average error in total and show that they are capable of capturing temporal and cross-series correlations. iTransformer has good performance in short horizons, but has slight accuracy drop in longer horizons. FEDformer and two decomposition-based models, i.e., ETSformer and Autoformer, usually yield higher errors, although they remain stable on some datasets.

GTformer and Pathformer models still provide satisfactory results in long-term horizons. FEDformer is also suitable for very long forecasting horizons due to computational efficiency and frequency architecture. In contrast, Autoformer and ETSformer usually have more error in longer horizons. In terms of computational efficiency, FEDformer is the most efficient model with near-linear complexity, suitable for large data. iTransformer is designed for high-dimensional data and consumes less memory. GTformer has higher computational cost but provides very good accuracy if sufficient resources are available. Pathformer establishes a good balance between accuracy and speed.

For domains that require explainability, ETSformer and Autoformer are the best options. These two models provide the possibility of more precise analysis with decomposition of trend and seasonal components. FEDformer is also somewhat explainable from the frequency perspective. GTformer clarifies cross-series relationships with graphs, though its explanation is more difficult for the end user. For data with a high number of variables or complex relationships, iTransformer and GTformer are superior choices. Other models have relative weakness in these conditions. Due to multi-scale architecture, Pathformer is the best option for data with complex seasonal patterns. FEDformer also has appropriate performance in this domain with reliance on frequency analysis. Autoformer and ETSformer focus more on trend/seasonal separation.

2.3. **Computational Efficiency Analysis.** One of the main challenges in employing transformer models for time series forecasting is their computational load and infrastructure requirements. These considerations are particularly important in industrial applications or when dealing with large-scale data. A review of the literature shows that each architecture has adopted specific strategies to reduce time complexity and optimize resource consumption:

https://dx.doi.org/10.22108/msci.2025.144922.1740

Autoformer reduces time complexity from $\mathcal{O}(L^2)$ to $\mathcal{O}(L \log L)$ by replacing the standard self-attention with an auto-correlation mechanism. This model requires moderate to high computational resources.

iTransformer by inverting the standard architecture, considers individual time series as tokens. This design transfers attention calculation from temporal dimension to variable dimension and transforms the dominant complexity of the model to $\mathcal{O}(N^2 \cdot D)$. Since in many practical problems $N \ll L$, this model significantly reduces memory consumption and training time especially when dealing with high-dimensional data.

FEDformer achieves linear complexity $O(L)$ using frequency domain analysis and random selection of Fourier modes. This characteristic makes it highly efficient for long-term forecasting with large-scale data.

GTformer has a computational complexity of $\mathcal{O}(M \cdot (L/P)^2 + M^2)$ and uses graph-based representation and sequence division into patches, thus providing acceptable efficiency for datasets with a large number of variables.

ETSformer combines exponential-smoothing attention mechanisms with frequency attention, achieving $\mathcal{O}(L \log L)$ time complexity. In terms of resource consumption, this model falls in the intermediate level, faster than standard transformers but heavier than linear models. This design enables the model to capture level, growth, and seasonality of data more accurately.

Pathformer by introducing adaptive pathways, gives the model the ability to dynamically select and compute only K scales from available M scales for each input sample. This mechanism prevents computational waste on irrelevant scales and leads to significant reduction in computational load and training time compared to fixed multi-scale architectures.

In general, it can be said that Autoformer and ETSformer balance between accuracy and computational cost, FEDformer and iTransformer have the most improvement in scalability and memory efficiency, GTformer is designed for heavy multivariate data, and Pathformer is a practical option for environments with resource constraints.

2.4. **Architectural Innovations and Design Principles.** Each of the evaluated models introduces unique architectural innovations that address specific challenges in time series forecasting:

Autoformer introduces an innovative architecture in the domain of long-term time series forecasting that is designed based on encoder-decoder structure. The main innovation of this model is that instead of standard attention, it employs series decomposition blocks and an auto-correlation mechanism. In each encoder layer, first the signal is decomposed into two components: trend and seasonal. The smoothed trend is forwarded and the seasonal part is analyzed by the auto-correlation mechanism to identify periodic patterns in historical sub-sequences. Then the feedforward network and residual connections cause stability and improvement of representation. In the decoder, the trend part is gradually reconstructed and finally combined with the seasonal part to produce the final prediction.

iTransformer provides an innovative approach to employing transformer architecture for multivariate time series forecasting by introducing the concept of inverted transformer. The core of this innovation is changing the common approach in data representation. Unlike conventional methods where values of all variables at one time step are considered as one token, iTransformer considers each time series as an independent token. This fundamental change involves swapping the temporal and variable dimensions. In this architecture, the attention mechanism is used to understand correlations between different variables instead of modeling temporal dependencies, and directly records dependencies between variables.

GTformer, a graph-based temporal order-aware transformer, is designed for long-term forecasting of multivariate time series and specifically focuses on dependencies between series and strict temporal order of data. For learning relationships between time series, GTformer introduces an adaptive graph learning method that identifies unidirectional and bidirectional relationships between series. In this model, inputs are divided into patches and a temporal-order-aware unit produces positional mappings so that the order of moments in data is not lost. Following this, an adaptive graph learning layer records unidirectional and bidirectional relationships among variables.

FEDformer is an innovative architecture that overcomes limitations of conventional models in processing long-term time series by combining time series decomposition and frequency domain analysis. This model uses a mixture-of-experts architecture decomposition and employs Fourier transform or wavelet transform to transfer data from time domain to frequency domain. In FEDformer architecture, first the input signal is optimally decomposed into trend and seasonal parts using mixture of experts approach. Then the seasonal part is transferred to frequency domain and attention mechanism is applied only to a subset of dominant frequency components.

ETSformer is an interpretable model inspired by classical exponential smoothing methods, decomposes time series data into three interpretable components: level, growth, and seasonal. The core of this model is based on two innovative attention mechanisms: exponential smoothing attention that effectively models temporal dependencies with a gradual emphasis on more recent data, and frequency attention that automatically identifies repetitive seasonal patterns using Fourier transform. In this model's architecture, layer-wise decomposition blocks separately extract and reconstruct each of the three parts.

Pathformer is an advanced multi-scale transformer architecture designed for understanding temporal patterns at different levels. This model precisely analyzes temporal features at various scales and combines two key aspects: temporal resolution and temporal distance, to efficiently model complex patterns. This model intelligently divides the time series into patches of different sizes and uses

https://dx.doi.org/10.22108/msci.2025.144922.1740

two types of specialized attention: inter-patch attention that analyzes connections between different patches, and intra-patch attention that examines details within each patch.

2.5. **Performance on High-Frequency Time Series.** In the domain of time series forecasting, one of the less explored aspects is the impact of sampling patterns on the performance of transformer models. In many real applications including high-speed financial systems and IoT sensor networks, data is generated with very high frequency. This unique characteristic creates fundamental questions about the impact of sampling methods on the forecasting accuracy of models. Two main approaches exist in sampling high-frequency data: domain increase and infill.

This difference in sampling patterns significantly affects the statistical characteristics of data. For example, in some conditions high-frequency series suffer from induced negative correlations that can make forecasting more difficult. Studies have also shown that asymptotic characteristics of estimators can exhibit different behaviors in these two cases. It should be noted that most introduced standard datasets are at medium frequency level and do not reflect actual very high-frequency conditions. Therefore, the results of these studies cannot be fully generalized to high-frequency data and need for more theoretical and conceptual analyses exists.

Model performance in facing these challenges is as follows: Autoformer shows separation of trend and seasonal components and stable performance in datasets with strong autocorrelation, though higher error has been reported in some scenarios. iTransformer shows high scalability for data with large numbers of variables and suitable performance in short-term predictions. GTformer learns interdependencies with graph structure, strong performance in transformer and electricity datasets, and suitability for noisy conditions and complex cross-series dependencies. FEDformer shows high computational efficiency with near-linear complexity and suitability for large-scale data and long-term prediction lengths. ETSformer provides high interpretability through decomposition of level, growth, and seasonal components, suitable for domains such as finance and medicine that require explanation. Pathformer shows multi-scale modeling capability and complex seasonal patterns, with strong performance in weather datasets and some transformer datasets.

The comprehensive summary of capabilities and limitations of these models is presented in a comparison table. The results of this analysis show that optimal model selection should not only be based on intrinsic characteristics of data, but also should consider sampling structure and existence of negative correlations. Understanding these complex relationships can lead to more informed model selection and improvement of prediction accuracy in practical applications.

2.6. **Interpretability and Practical Applications.** Interpretability remains a critical consideration, particularly in sensitive domains such as healthcare, finance, and energy management where understanding model decisions is as important as prediction accuracy. The evaluated models offer varying levels of interpretability through different architectural approaches.

ETSformer and Autoformer provide the highest level of interpretability through explicit decomposition of time series into meaningful components. This decomposition allows domain experts to understand which components are driving the forecasts and provides valuable insights for decision-making processes. The clear separation of trend and seasonal components in Autoformer, and level, growth, and seasonal components in ETSformer, enables straightforward analysis of temporal patterns and their contributions to future predictions.

FEDformer offers interpretability from a frequency perspective, enabling analysis of which frequency components contribute most significantly to the predictions. This frequency-based interpretability is particularly valuable for applications with strong seasonal characteristics and periodic patterns, allowing domain experts to identify dominant frequencies and their impact on forecasting outcomes.

GTformer provides insights through its learned graph structure, revealing relationships between different variables in multivariate time series. The adaptive graph learning mechanism identifies meaningful connections between variables, supporting domain knowledge validation and discovery of previously unknown relationships in complex systems.

Pathformer's multi-scale analysis reveals which temporal scales are most informative for different forecasting tasks. The adaptive pathway selection reveals the relevance of different time scales for specific prediction scenarios, providing insights into temporal pattern hierarchies and their relative importance.

iTransformer adds value by focusing on cross-variable relationships, which can be particularly useful in applications where understanding variable interactions is important. The attention mechanisms highlight dependencies between different time series.

The practical applications of these transformer models span numerous critical domains. In energy management, they enable accurate load forecasting, grid optimization, and resource allocation. In financial markets, they support risk assessment, trading strategies, and market analysis. In healthcare, they facilitate patient monitoring, disease progression prediction, and medical decision support. In transportation systems, they improve traffic flow optimization, demand forecasting, and infrastructure planning. In industrial applications, they enable predictive maintenance, process optimization, and quality control.

3. **Conclusions**

This comprehensive review demonstrates that transformer-based architectures have significantly advanced the state-of-the-art in multivariate time series forecasting, offering substantial improvements over traditional statistical methods and earlier deep learning approaches. The architectural innovations

https://dx.doi.org/10.22108/msci.2025.144922.1740

introduced in models such as Autoformer, iTransformer, GTformer, FEDformer, ETSformer, and Pathformer have addressed key limitations of standard transformers while leveraging their strengths in capturing long-range dependencies.

The synthesis of this study indicates that recent innovations, including the application of frequency domain processing, auto-correlation mechanisms, graph-based learning, and multi-scale architectures, have simultaneously led to improvements in both prediction accuracy and computational efficiency. These advancements chart a clear research path for developing future architectures with focus on higher interpretability, better scalability, and broader generalization capabilities.

The key conclusions from our analysis can be summarized as follows: First, no single architecture emerges as universally superior across all evaluation dimensions. The optimal model choice depends on specific application requirements, including forecasting horizon, data characteristics, computational constraints, and interpretability needs. This highlights the importance of context-aware model selection and the value of having diverse architectural options for different scenarios.

Second, the most significant advances have come from incorporating domain-specific inductive biases into the transformer architecture. Models that explicitly model temporal patterns such as seasonality, trends, and periodicity generally outperform generic transformer architectures. These domain adaptations enable more efficient learning of temporal patterns and improve forecasting accuracy across various applications.

Third, computational efficiency remains a critical consideration, particularly for long sequences and real-time applications. The evolution from quadratic to linear or log-linear complexity represents a major advancement enabling practical deployment in resource-constrained environments. The various efficiency optimization strategies have made transformer-based forecasting feasible for large-scale applications.

Fourth, the handling of high-frequency time series presents unique challenges that require specialized approaches. The differences between sampling strategies and phenomena such as induced negative correlations necessitate careful model selection and potentially domain-specific adaptations. The varying model performance across different sampling scenarios underscores the need for continued research in high-frequency time series forecasting.

Several important challenges and research directions remain open, providing opportunities for future investigations: The integration of time-domain and frequency-domain approaches represents a promising direction for future research, potentially enabling more robust handling of both short-term patterns from dense sampling and long-term patterns from extended observation windows. Developing models with greater robustness to the statistical peculiarities of high-frequency data would significantly enhance practical applicability.

The creation of more adaptive architectures that can automatically adjust their processing strategies based on data characteristics, sampling patterns, and application requirements would reduce the need for manual model selection and hyperparameter tuning. Improving interpretability while maintaining competitive performance remains an important challenge, particularly for applications in regulated domains or scenarios requiring human oversight.

The development of more comprehensive benchmark datasets that better reflect real-world high-frequency conditions would enable more meaningful evaluation and comparison of forecasting models. Standardized evaluation protocols and diverse application scenarios would facilitate rigorous assessment of model capabilities and support reproducible research in this rapidly evolving field.

In conclusion, transformer-based architectures have fundamentally transformed the landscape of multivariate time series forecasting. The continued innovation in this field promises even more powerful, efficient, and interpretable forecasting models that will enable better decision-making across numerous application domains. The diversity of architectural approaches and the ongoing methodological advancements ensure that transformer-based models will remain at the forefront of time series forecasting research and applications, driving progress in energy management, financial forecasting, healthcare monitoring, intelligent transportation systems, and beyond.

References

[1] D. Cao and S. Zhang, AD-autoformer: Decomposition transformers with attention distilling for long sequence time-series forecasting, *The Journal of Supercomputing*, **2024** 1–21.

[2] P. Chen, Y. Zhang, Y. Cheng, Y. Shu, Y. Wang, Q. Wen, B. Yang and C. Guo, Pathformer: Multi-scale transformers with adaptive pathways for time series forecasting, *arXiv preprint arXiv:2402.05956*, 2024.

[3] W. Chen, W. Wang, B. Peng, Q. Wen, T. Zhou and L. Sun, Learning to rotate: Quaternion transformer for complicated periodical time series forecasting, *Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, (2022) 146–156.

[4] Y. Chen, S. Liu, J. Yang, H. Jing, W. Zhao and G. Yang, A Joint Time-Frequency Domain Transformer for multivariate time series forecasting, *Neural Networks*, **176** (2024) 106334.

[5] Z. Chen, M. Ma, T. Li, H. Wang and C. Li, Long sequence time-series forecasting with deep learning: A survey, *Information Fusion*, **97** (2023) 101819.

[6] R.-G. Cirstea, C. Guo, B. Yang, T. Kieu, X. Dong and S. Pan, Triformer: Triangular, Variable-Specific Attentions for Long Sequence Multivariate Time Series Forecasting–Full Version, *arXiv preprint arXiv:2204.13767*, 2022.

[7] W. Guan, I. Smetannikov and T. Man, Survey on automatic text summarization and transformer models applicability, *Proceedings of the 2020 1st International Conference on Control, Robotics and Intelligent System*, (2020) 176–184.

 https://dx.doi.org/10.22108/msci.2025.144922.1740

[8] K. Han, Y. Wang, H. Chen, X. Chen, J. Guo, Z. Liu, Y. Tang, A. Xiao, C. Xu, Y. Xu, et al., A survey on vision transformer, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **45** no. 1 (2022) 87–110.

[9] N. Kitaev, Ł. Kaiser and A. Levskaya, Reformer: The efficient transformer, *arXiv preprint arXiv:2001.04451*, 2020.

[10] S. Li, X. Jin, Y. Xuan, X. Zhou, W. Chen, Y.-X. Wang and X. Yan, Enhancing the locality and breaking the memory bottleneck of transformer on time series forecasting, *Advances in Neural Information Processing Systems*, **32** (2019) 5243–5253.

[11] A. Liang, X. Chai, Y. Sun and M. Guizani, GTformer: Graph-Based Temporal-Order-Aware Transformer for Long-Term Series Forecasting, *IEEE Internet of Things Journal*, (2024) 31467–31478.

[12] B. Lim, S. Ö. Arık, N. Loeff and T. Pfister, Temporal fusion transformers for interpretable multi-horizon time series forecasting, *International Journal of Forecasting*, **37** no. 4 (2021) 1748–1764.

[13] Y. Lin, I. Koprinska and M. Rana, SSDNet: State space decomposition neural network for time series forecasting, *2021 IEEE International Conference on Data Mining (ICDM)*, (2021) 370–378.

[14] S. Liu, H. Yu, C. Liao, J. Li, W. Lin, A. X. Liu and S. Dustdar, Pyraformer: Low-complexity pyramidal attention for long-range time series modeling and forecasting, *The Tenth International Conference on Learning Representations (ICLR)*, (2022).

[15] X. Liu and W. Wang, Deep Time Series Forecasting Models: A Comprehensive Survey, *Mathematics*, **12** no. 10 (2024) 1504.

[16] Y. Liu, T. Hu, H. Zhang, H. Wu, S. Wang, L. Ma and M. Long, iTransformer: Inverted Transformers Are Effective for Time Series Forecasting, *The Twelfth International Conference on Learning Representations*, (2024).

[17] Y. Liu, H. Wu, J. Wang and M. Long, Non-stationary transformers: Exploring the stationarity in time series forecasting, *Advances in Neural Information Processing Systems*, **35** (2022) 9881–9893.

[18] Y. Liu, H. Zhang, C. Li, X. Huang, J. Wang and M. Long, Timer: Generative Pre-trained Transformers Are Large Time Series Models, *Forty-first International Conference on Machine Learning*, (2024).

[19] Q. Ma, Z. Liu, Z. Zheng, Z. Huang, S. Zhu, Z. Yu and J. T. Kwok, A survey on time-series pre-trained models, *IEEE Transactions on Knowledge and Data Engineering*, (2024) 7536–7555.

[20] J. A. Miller, M. Aldosari, F. Saeed, N. H. Barna, S. Rana, I. B. Arpinar and N. Liu, A survey of deep learning and foundation models for time series forecasting. arXiv 2024, *arXiv preprint arXiv:2401.13912*, (2024).

[21] Y. Nie, N. H. Nguyen, P. Sinthong and J. Kalagnanam, A time series is worth 64 words: Long-term forecasting with transformers, *arXiv preprint arXiv:2211.14730*, (2022).

[22] K. Olorunnimbe and H. Viktor, Ensemble of temporal Transformers for financial time series, *Journal of Intelligent Information Systems*, (2024) 1–25.

[23] X. Piao, Z. Chen, T. Murayama, Y. Matsubara and Y. Sakurai, Fredformer: Frequency debiased transformer for time series forecasting, *Proceedings of the 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, (2024) 2400–2410.

[24] A. Shabani, A. Abdi, L. Meng and T. Sylvain, Scaleformer: Iterative multi-scale refining transformers for time series forecasting, *arXiv preprint arXiv:2206.04038*, (2022).

[25] D. V. da Silva, V. Estevam and D. Menotti, Towards a realistic Libras to Portuguese translation, *2023 36th SIBGRAPI Conference on Graphics, Patterns and Images (SIBGRAPI)*, (2023) 1–6.

[26] J. Tong, L. Xie, W. Yang, K. Zhang and J. Zhao, Enhancing time series forecasting: A hierarchical transformer with probabilistic decomposition representation, *Information Sciences*, **647** (2023) 119410.

[27] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. U. Kaiser and I. Polosukhin, Attention is all you need, *Advances in Neural Information Processing Systems*, **30** (2017) 1–15.

[28] Y. Wang, H. Wu, J. Dong, Y. Liu, M. Long and J. Wang, Deep time series models: A comprehensive survey and benchmark, *arXiv preprint arXiv:2407.13278*, (2024).

[29] Q. Wen, T. Zhou, C. Zhang, W. Chen, Z. Ma, J. Yan and L. Sun, Transformers in time series: A survey, *arXiv preprint arXiv:2202.07125*, (2022).

[30] G. Woo, C. Liu, D. Sahoo, A. Kumar and S. Hoi, CoST: Contrastive Learning of Disentangled Seasonal-Trend Representations for Time Series Forecasting, *International Conference on Learning Representations*, (2022).

[31] G. Woo, C. Liu, D. Sahoo, A. Kumar and S. Hoi, Etsformer: Exponential smoothing transformers for time-series forecasting, *arXiv preprint arXiv:2202.01381*, (2022).

[32] H. Wu, J. Xu, J. Wang and M. Long, Autoformer: Decomposition transformers with auto-correlation for long-term series forecasting, *Advances in Neural Information Processing Systems*, **34** (2021) 22419–22430.

[33] Y. Yu, R. Ma and Z. Ma, Robformer: A robust decomposition transformer for long-term time series forecasting, *Pattern Recognition*, **153** (2024) 110552.

[34] A. Zeng, M. Chen, L. Zhang and Q. Xu, Are transformers effective for time series forecasting?, *Proceedings of the AAAI Conference on Artificial Intelligence*, **37** no. 9 (2023) 11121–11128.

[35] J. Zhang, H. Luan, M. Sun, F. Zhai, J. Xu, M. Zhang and Y. Liu, Improving the Transformer Translation Model with Document-Level Context, *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, (2018) 533–542.

[36] Y. Zhang, L. Ma, S. Pal, Y. Zhang and M. Coates, Multi-resolution Time-Series Transformer for Long-term Forecasting, *International Conference on Artificial Intelligence and Statistics*, (2024) 4222–4230.

[37] Y. Zhang and J. Yan, Crossformer: Transformer utilizing cross-dimension dependency for multivariate time series forecasting, *The Eleventh International Conference on Learning Representations*, (2023).

[38] H. Zhou, S. Zhang, J. Peng, S. Zhang, J. Li, H. Xiong and W. Zhang, Informer: Beyond efficient transformer for long sequence time-series forecasting, *Proceedings of the AAAI Conference on Artificial Intelligence*, **35** no. 12 (2021) 11106–11115.

[39] T. Zhou, Z. Ma, Q. Wen, X. Wang, L. Sun and R. Jin, Fedformer: Frequency enhanced decomposed transformer for long-term series forecasting, *International Conference on Machine Learning*, (2022) 27268–27286.

[40] S. Bandyopadhyay and S. N. Lahiri, Asymptotic properties of discrete Fourier transforms for spatial data, *Sankhyā: The Indian Journal of Statistics, Series A*, **71** no. 2 (2009) 221–259.

https://dx.doi.org/10.22108/msci.2025.144922.1740

[41] R. S. Tsay, *Analysis of Financial Time Series*, 2nd ed., Hoboken, NJ: Wiley-Interscience [John Wiley & Sons], 2005, ISBN 978-0-471-69074-0, doi: 10.1002/0471746193.

Esmaeil Chitgar

Department of Computer Science, Faculty of Mathematical Sciences, Yazd University, Yazd, Iran

`Email: ec.chitgar@gmail.com`

Mahdieh Hasheminezhad

Department of Computer Science, Faculty of Mathematical Sciences, Yazd University, Yazd, Iran

`Email: hasheminezhad@yazd.ac.ir`

Sepehr Ebrahimi Mood

Department of Computer Science, Faculty of Mathematical Sciences, Yazd University, Yazd, Iran

`Email: s.ebrahimi@yazd.ac.ir`