ISSN 3049 - 3692 (Print)

A DUAL-STAGE DEEP LEARNING FRAMEWORK FOR ROBUST TIME-SERIES FORECASTING UNDER NON-STATIONARY CONDITIONS

M. Revathi, S.Nithya
Computer Science & Engineering, Coimbatore Institute of Technology, Coimbatore, Tamilnadu

Abstract

Time-series forecasting plays a critical role in applications such as smart grids, financial markets, weather prediction, and industrial monitoring. However, most deep learning models struggle under non-stationary conditions involving concept drift, abrupt regime changes, and noise-induced disturbances. This paper proposes a Dual-Stage Deep Learning Forecasting Framework (DS-DLFF) combining (1) a Variational Mode Decomposition (VMD)-based preprocessing module for decomposing non-stationary time-series into intrinsic mode components, and (2) a Transformer–LSTM hybrid architecture that captures long-range dependencies and local temporal patterns within each decomposed component. A drift-adaptive calibration layer is introduced to detect distribution shifts using Maximum Mean Discrepancy (MMD) and dynamically update model parameters. Experiments conducted on four real-world datasets—electricity load, financial stock indices, traffic speed, and environmental pollution—demonstrate that DS-DLFF achieves significant improvements in RMSE, MAE, and MAPE compared to state-of-the-art baselines (Tables 2–4). This framework provides a robust forecasting solution for non-stationary environments, outperforming both classical ML models and advanced DL architectures

Keywords: Time-series forecasting; Deep Learning; Transformer Networks; LSTM; Concept Drift; Non-Stationary Data; Variational Mode Decomposition (VMD).

1. Introduction

Deep learning (DL) models such as LSTM, GRU, and Transformers have achieved remarkable performance in predicting temporal data patterns across domains including finance, weather systems, and industrial automation [1, 2]. However, real-world time-series often exhibit non-stationarity, meaning their statistical properties vary over time due to external factors such as seasonal variation, market volatility, or sensor drift [3]. Traditional ML models like ARIMA, SVR, and XGBoost handle low-variance environments well but fail under non-linear and highly dynamic systems [4, 5].

Recent studies have explored decomposition-based forecasting frameworks, including Wavelet Transforms, Empirical Mode Decomposition (EMD), and Variational Mode Decomposition (VMD), for reducing noise and extracting intrinsic components [6, 7]. Others utilize hybrid deep architectures such as CNN-LSTM, Seq2Seq, and attention-based models to capture multi-scale dependencies [8,9].

However, two challenges remain:

(1) Handling abrupt concept drift, where model performance deteriorates due to sudden distribution changes.

(2) Combining decomposition and attention, while maintaining computational feasibility.

To address these gaps, this paper proposes DS-DLFF, a dual-stage framework integrating VMD with a Transformer–LSTM hybrid network and a drift-detection calibration layer. An overview of DS-DLFF components is provided in Table 1, and full evaluation results in Tables 2–4.

2 Related Work

2.1 Classical Time-Series Forecasting Models

ARIMA, SARIMA, and Holt-Winters are widely used but rely heavily on stationarity assumptions [10, 11]. Kernel-based models like SVR and machine learning methods such as Random Forest
(RF) show improvements but lack temporal sequence awareness [12, 13].

2.2 Deep Learning for Forecasting

LSTM, GRU, TCN, and Transformer variants have become the state-of-the-art for sequence modeling [14, 15]. Transformers, in particular, excel at capturing long-range dependencies but often require large datasets and may overfit under noisy conditions [16].

2.3 Decomposition-Driven Forecasting

Techniques such as EMD, Wavelet Transform, and VMD enhance model stability by decomposing data into intrinsic components [17, 18]. VMD is superior in separating oscillatory modes with minimal mode mixing [19].

2.4 Drift Detection and Adaptive Learning

Approaches such as ADWIN, DDM, and MMD-based detection are used widely in streaming data analysis [20, 21]. However, integrating drift detection with deep forecasting models remains relatively unexplored.

3 Proposed Methodology

The architecture consists of:

Variational Mode Decomposition (VMD) for preprocessing

Transformer–LSTM Hybrid Forecaster

MMD-Based Drift Detection and Calibration Layer

Detailed descriptions appear below.

3.1 Variational Mode Decomposition (Stage 1)

The input time-series x(t)x(t)x(t) is decomposed into
K intrinsic mode functions (IMFs):

$$x(t) = \sum_{k=1}^{k} u_k(t)$$

VMD Optimizes:

$$\min_{u_k, w_k} \sum \| \partial_t [(\delta(t) + j / \pi t) * u_k(t)] e^{-j w_k t} \|_2^2$$

This eliminates noise and reveals hidden patterns.

3.2 Transformer–LSTM Hybrid Network (Stage 2)

Each IMF is independently processed using:

- Transformer Encoder: global dependencies
- Bi-LSTM Layer: local trends + nonlinear tem poral memory

The final prediction is reconstructed as:

$$\hat{x}(t) = \sum_{k=1}^{k} \hat{u}_k(t)$$

3.3 Drift-Adaptive Calibration Using MMD

To detect drift, we compute MMD between:

- Current input window X
- Reference stable window Y

$$MMD(X,Y) = \left\| \frac{1}{n}\sum \phi(x_i) - \frac{1}{m}\sum \phi(y_i) \right\|^2$$

If MMD > threshold:

- Recalibrate Transformer and LSTM layers
- Update learning rates
- Increase attention dropout

4 Experimental Setup

4.1 Datasets

Four publicly available datasets:

- Electricity Consumption (UCI) [22]
- S&P 500 Index and NASDAQ Composite [23]
- PEMS Traffic Speed Dataset [24]
- Beijing Air Quality Dataset [25]

4.2 Baseline Models

- ARIMA
- SVR
- Random Forest
- LSTM
- GRU
- TCN
- Transformer
- VMD+LSTM
- VMD+Transformer

5 Results

5.1 Overall Performance Comparison

Table 2. RMSE Comparison Across Models

Model	Electricity	PEMS Traffic	Air Quality	Stock Index
ARIMA	0.241	8.34	7.92	18.51
SVR	0.221	7.89	7.02	15.46
LSTM	0.203	6.74	6.29	13.87
Transformer	0.196	6.18	6.01	12.92
VMD + LSTM	0.183	6.04	5.83	12.64
VMD + Transformer	0.178	5.89	5.62	12.28
DS-DLF-F (Proposed)	0.162	5.41	5.21	11.73

5.2 MAE Results

Table 3. MAE Performance

Model	Electricity	Traffic	Air Quality	Stock
LSTM	0.152	4.32	3.71	9.13
Transformer	0.147	4.11	3.56	8.74
VMD + Transformer	0.133	3.82	3.27	8.31
DS-DLF-F	0.119	3.41	2.96	7.89

5.3 MAPE Results

Table 4. MAPE (%) Comparison

Model	Electricity	Traffic	Air Quality	Stock
ARIMA	7.82%	14.1%	15.9%	21.7%
LSTM	6.34%	12.6%	13.2%	19.4%
Transformer	5.91%	11.9%	12.4%	18.3%
DS-DLFF	4.84%	10.7%	11.1%	16.9%

6 Discussion

Key findings:

- Decomposition (VMD) significantly reduces noise and improves model stability.
- The Transformer–LSTM hybrid effectively models long- and short-term patterns simultaneously.
- MMD-based drift detection improves robustness under volatile conditions.
- DS-DLFF consistently outperforms all baselines in RMSE, MAE, and MAPE.

7 Conclusion

This paper presents DS-DLFF, a dual-stage deep learning forecasting framework incorporating VMD decomposition, a Transformer–LSTM hybrid architecture, and an MMD-based drift calibration module. Experiments across four datasets demonstrate substantial improvements and robustness under non-stationary conditions. Future work will explore federated forecasting, reinforcement learning for drift management, and lightweight models for edge deployment.

References

[1] S. Hochreiter and J. Schmidhuber, “Long short-term memory,” Neural Computation, vol. 9, no. 8, pp. 1735–1780, 1997.

[2] A. Vaswani et al., “Attention is all you need,” in Proc. 31st Int. Conf. Neural Information Processing Systems (NeurIPS), Long Beach, CA, USA, pp. 5998–6008, 2017.

[3] R. J. Hyndman and G. Athanasopoulos, Forecasting: Principles and Practice, 2nd ed., Melbourne, Australia: OTexts, 2018.

[4] L. Breiman, “Random forests,” Machine Learning, vol. 45, no. 1, pp. 5–32, 2001.

[5] C. Cortes and V. Vapnik, “Support-vector networks,” Machine Learning, vol. 20, no. 3, pp. 273–297, 1995.

[6] N. E. Huang et al., “The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis,” Proc. Royal Society A, vol. 454, no. 1971, pp. 903–995, 1998.

[7] K. Dragomiretskiy and D. Zosso, “Variational mode decomposition,” IEEE Transactions on Signal Processing, vol. 62, no. 3, pp. 531–544, 2014.

[8] S. Bai, J. Z. Kolter, and V. Koltun, “An empirical evaluation of generic convolutional and recurrent networks for sequence modeling,” arXiv preprint arXiv:1803.01271, 2018.

[9] S. Shun, Y. Li, and H. Zhang, “Hybrid deep learning models for time-series forecasting,” Applied Soft Computing, vol. 105, Art. no. 107294, 2021.

[10] G. E. P. Box, G. M. Jenkins, G. C. Reinsel, and G. M. Ljung, Time Series Analysis: Forecasting and Control, 5th ed., Hoboken, NJ, USA: Wiley, 2015.

[11] J. W. Taylor, "Forecasting daily supermarket sales using exponentially weighted quantile regression," European Journal of Operational Research, vol. 259, no. 2, pp. 707–720, 2017.

[12] T. Chen and C. Guestrin, "XGBoost: A scalable tree boosting system," in Proc. 22nd ACM SIGKDD Int. Conf. Knowledge Discovery and Data Mining (KDD), San Francisco, CA, USA, pp. 785–794, 2016.

[13] Y. Fan, Q. Li, and S. Liu, "Support vector regression based electricity load forecasting," Energy, vol. 144, pp. 48–59, 2018.

[14] Y. Qin, D. Song, H. Chen, W. Cheng, G. Jiang, and G. Cottrell, "A dual-stage attention-based recurrent neural network for time series prediction," in Proc. 26th Int. Joint Conf. Artificial Intelligence (IJCAI), Melbourne, VIC, Australia, pp. 2627–2633, 2017.

[15] H. Wu, J. Xu, J. Wang, and M. Long, "Autoformer: Decomposition transformers with auto-correlation for long-term series forecasting," in Proc. 35th Int. Conf. Neural Information Processing Systems (NeurIPS), 2021.

[16] H. Zhou et al., "Informer: Beyond efficient transformer for long sequence time-series forecasting," in Proc. AAAI Conf. Artificial Intelligence, vol. 35, no. 12, pp. 11106–11115, 2021.

[17] N. Mert, A. Akan, and H. Yildirim, "Wavelet-based hybrid forecasting models for time series," Expert Systems with Applications, vol. 159, Art. no. 113610, 2020.

[18] Y. Wang, L. Zhang, and X. Li, "Hybrid decomposition-based deep learning models for time-series forecasting," Neurocomputing, vol. 481, pp. 187–201, 2022.

[19] D. Zosso, K. Dragomiretskiy, and A. Flandrin, "Adaptive variational mode decomposition," IEEE Signal Processing Letters, vol. 26, no. 8, pp. 1150–1154, 2019.

[20] J. Gama, I. •liobaite, A. Bifet, M. Pechenizkiy, and A. Bouchachia, "A survey on concept drift adaptation," ACM Computing Surveys, vol. 46, no. 4, pp. 1–37, 2014.

[21] A. Gretton et al., "A kernel two-sample test," Journal of Machine Learning Research, vol. 13, pp. 723–773, 2007.

[22] D. Dua and C. Graff, "UCI machine learning repository," Univ. of California, Irvine, CA, USA, 2017. [Online]. Available: https://archive.ics.uci.edu

[23] Yahoo Finance, "Yahoo Finance historical market data," 2024. [Online]. Available: https://finance.yahoo.com

[24] C. Chen, Y. Wang, L. Li, J. Hu, and Z. Zhang, "Large-scale traffic speed prediction with deep learning," IEEE Transactions on Intelligent Transportation Systems, vol. 15, no. 4, pp. 1680–1689, 2015.

[25] S. Zhang, G. Li, and Y. Li, "PM2.5 air-quality data of Beijing," UCI Machine Learning Repository, 2017.