Period-Aware Inductive Bias Versus Scale on Influenza-like Illness Forecasting

YongKyung Oh [1] **Alex Bui** [1]

Abstract

Forecasting influenza-like illness (ILI) is important for public-health planning, but weekly surveillance series combine short-lag persistence with annual recurrence and season-to-season shifts in peak timing and amplitude. Foundation models (FMs) and large language model (LLM)-adapted forecasters are increasingly used for time-series forecasting, yet on the standard ILI benchmark they do not show a clear average-error advantage over strong compact baselines. We therefore propose EPIC (Efficient Periodic Inductive Convolution), a compact attention-free forecaster that combines temporal convolution, period projection, and latent horizon projection. Under the same benchmark setting, EPIC achieves lower four-horizon average error than published compact and LLM-based comparators, and ablations show that the temporal and periodic branches contribute complementary information. This advantage is concentrated in the average and shorter-to-medium horizons, while the longest horizon remains a limitation. On this weekly ILI benchmark, a small period-aware inductive bias is more effective than additional model scale.

1. Introduction

Forecasting influenza-like illness (ILI) informs vaccine planning, hospital preparedness, and public-health response (Biggerstaff et al., 2016; McGowan et al., 2019; Reich et al., 2019; Lutz et al., 2019). Weekly surveillance series combine short-lag persistence, annual recurrence, drifting peak timing and amplitude, and coupled multivariate channels. This regime is better described as structured periodic nonstationarity than as stationary seasonality (Brooks et al., 2015; Wu et al., 2018). A general-purpose mixer must recover both effects implicitly from a short training history, whereas a model that encodes them separately can exploit each as a distinct inductive bias. This study examines whether encoding these dynamics as separate inductive biases improves weekly ILI forecasting.

Figure 1 shows that the ILI target has strong short-lag persistence, recurring annual power with drifting peak timing, and only a short training history. The first property favors causal temporal convolutions with a small dilated receptive field. The second favors spectral period features, because the magnitude spectrum is insensitive to global phase shifts within a window. The third makes explicit periodic and temporal components more data-efficient than a single general-purpose mixer. These observations motivate **EPIC** (Efficient Periodic Inductive Convolution) for ILI forecasting. Our work makes three main contributions.

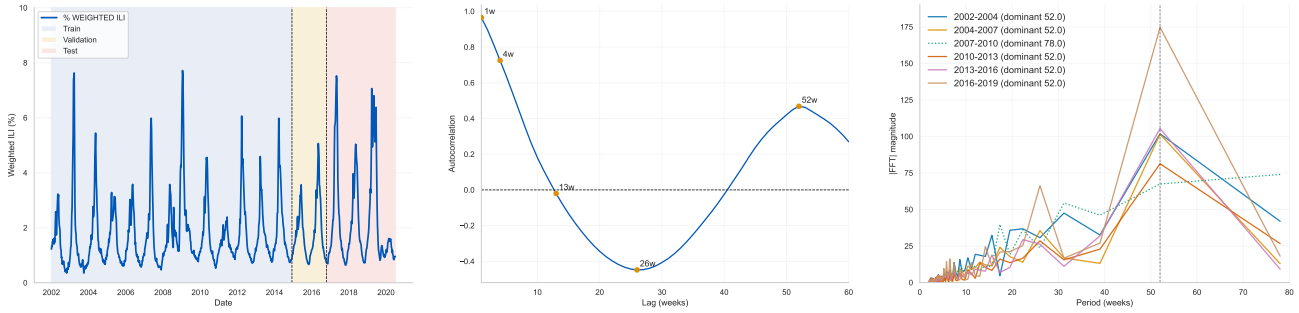
- We introduce EPIC, an attention-free ILI forecaster that combines a causal temporal-convolution backbone, an explicit period branch, and a latent horizon projector.
- We show that EPIC reaches lower four-horizon average error than published compact and LLM-based forecasters on the `national_illness` benchmark, while the same published results show no consistent average-error advantage for LLM-based comparators over compact baselines.
- We provide branch and projector ablations showing that the proposed combination is best on most individual horizons and on the average.

2. Related Work

Recent time-series forecasting has split between broad pretrained or LLM-adapted backbones and compact supervised architectures. Surveys and benchmark studies emphasize the diversity of deep forecasting designs and the continuing strength of specialized architectures (Kim et al., 2025; Wang et al., 2026). Within this landscape, transformer-style and patch-based models remain important compact references, including decomposition, non-stationary, temporal-variation, patch, multiscale, and channel-aligned designs (Wu et al., 2021; Liu et al., 2022; Wu et al., 2023; Nie et al., 2023; Chen et al., 2024; Wang et al., 2024).

[1]Medical & Imaging Informatics (MII) Group, University of California, Los Angeles (UCLA), Los Angeles, CA, USA. Correspondence to: YongKyung Oh <yongkyungoh@mednet.ucla.edu>, Alex Bui <buia@mii.ucla.edu>.

Accepted to the Forecasting as a New Frontier of Intelligence workshop at ICML 2026.



(a) Full history *(b)* Short-lag ACF *(c)* Period diagnostic

Figure 1. Diagnostic views of the weekly `national_illness` target. The full-history panel shows recurrent epidemic waves with changing amplitude and peak timing. The autocorrelation panel shows that predictive dependence is strongest at short lags. The period diagnostic shows that annual power remains visible across most non-overlapping three-year windows.

In parallel, time-series foundation models (FMs) and LLM-adapted forecasters aim to reuse pretrained representations across domains (Liang et al., 2024; Zhou et al., 2023; Das et al., 2024; Jin et al., 2024; Liu et al., 2024; 2025). Recent healthcare-domain evaluations caution that such transfer is not uniformly reliable across clinical time-series settings (Gu et al., 2025; Oh & Bui, 2025). ILI provides a targeted testbed for assessing whether the scale of cross-domain foundation models offers advantages over compact architectures designed with disease-specific inductive biases.

ILI forecasting is especially pointed because the target combines local persistence with recurring but phase-shifting annual structure. Prior ILI-focused forecasting work has used both deep sequence models and white-box or multimodal designs (Wu et al., 2020; Shen et al., 2024; Moon et al., 2025). The study instead evaluates whether explicit temporal and periodic components are better matched to the ILI surveillance regime. The full comparison set is collected in Appendix A.

3. Methodology

Problem Formulation. Let $\mathbf{X} = [\mathbf{x}_1, \ldots, \mathbf{x}_T]^\top \in \mathbb{R}^{T\times C}$ denote a multivariate influenza-like illness (ILI) history with C channels and look-back window T. The task is to predict the next H steps, $\hat{\mathbf{Y}} \in \mathbb{R}^{H\times C}$.

Model Overview. EPIC comprises four components: input encoding with horizon projection, causal temporal convolution, a learnable period branch, and additive fusion with linear readout.

3.1. Input Encoding and Horizon Projection

We first normalize each channel of the input sequence using instance statistics,

$$\tilde{\mathbf{X}} = \frac{\mathbf{X} - \boldsymbol{\mu}}{\boldsymbol{\sigma}}, \tag{1}$$

where $\boldsymbol{\mu}, \boldsymbol{\sigma} \in \mathbb{R}^{1\times C}$ are computed over the temporal dimension. The normalized sequence is embedded into a latent representation

$$\mathbf{Z}^{(0)} = \mathcal{E}(\tilde{\mathbf{X}}), \qquad \mathbf{Z}^{(0)} \in \mathbb{R}^{T\times d}. \tag{2}$$

Before applying the forecasting blocks, EPIC expands the latent sequence from length T to length $T + H$ through independent latent-dimension projectors. For each latent coordinate $k \in \{1, \ldots, d\}$,

$$\bar{\mathbf{z}}^{(0)}_{:,k} = \mathcal{P}_k\left(\mathbf{z}^{(0)}_{:,k}\right), \qquad \mathcal{P}_k : \mathbb{R}^T \to \mathbb{R}^{T+H}. \tag{3}$$

This operation produces $\bar{\mathbf{Z}}^{(0)} \in \mathbb{R}^{(T+H)\times d}$. Projecting to $T + H$ lets all forecast positions be processed in one forward pass and avoids autoregressive error accumulation, a relevant constraint when H exceeds T.

3.2. Temporal Convolutional Interaction

EPIC stacks L identical forecasting blocks. Given the block input $\bar{\mathbf{Z}}^{(\ell-1)}$, the first stage applies a causal temporal-convolution interaction module,

$$\mathbf{U}^{(\ell)} = \mathrm{LN}\left(\mathcal{F}_{\mathrm{TCN}}\left(\bar{\mathbf{Z}}^{(\ell-1)}\right)\right), \tag{4}$$

where $\mathcal{F}_{\mathrm{TCN}}$ is a residual temporal convolutional network with three residual blocks, two kernel-3 dilated convolutions per block, and dilations $\{1, 2, 4\}$. The convolution is causal over the projected latent sequence of length $T + H$, which covers short- and medium-range temporal dependence within a single forward pass. Because each forecast position is a learned function of the observed window through the projector $\mathcal{P}_k$, the model is non-autoregressive in observed time.

We use fixed local structure instead of a temporal self-attention operator, and the temporal feature is

$$\mathbf{H}^{(\ell)}_{\mathrm{T}} = \mathbf{U}^{(\ell)}. \tag{5}$$

3.3. Learnable Period Branch

ILI sequences recur annually, but their peaks are not phase-locked across years. To capture this component without assuming a fixed seasonal template, EPIC augments the temporal feature with a learnable period branch. Given $\mathbf{U}^{(\ell)}$, we first compute its magnitude spectrum,

$$\mathbf{S}^{(\ell)} = \left|\mathrm{FFT}\left(\mathbf{U}^{(\ell)}\right)\right|. \tag{6}$$

The magnitude spectrum is invariant to global shifts of a windowed signal, so it emphasizes recurring periodic content while suppressing exact within-window phase. The learnable period prototypes operate on this representation, allowing the period branch to track annual structure even when peak timing drifts across years. Let $\mathbf{W} \in \mathbb{R}^{N_p \times d}$ denote N_p learned period prototypes. We obtain period assignment scores and aggregate periodic features as

$$\boldsymbol{\alpha}^{(\ell)} = \mathrm{softmax}\left(\mathbf{S}^{(\ell)}\mathbf{W}^{\top}\right), \qquad \mathbf{V}^{(\ell)} = \boldsymbol{\alpha}^{(\ell)}\mathbf{W}, \tag{7}$$

where $\boldsymbol{\alpha}^{(\ell)} \in \mathbb{R}^{(T+H)\times N_p}$ and $\mathbf{V}^{(\ell)} \in \mathbb{R}^{(T+H)\times d}$. The final period representation is

$$\mathbf{H}_{\mathrm{P}}^{(\ell)} = \mathrm{LN}\left(\mathcal{W}_{\mathrm{p}}\mathbf{V}^{(\ell)} + \mathbf{U}^{(\ell)}\right). \tag{8}$$

The branch introduces frequency-domain information without adding an attention module.

3.4. Fusion and Forecasting

The default EPIC block fuses the temporal and period branches additively,

$$\mathbf{G}^{(\ell)} = \mathbf{H}_{\mathrm{T}}^{(\ell)} + \mathbf{H}_{\mathrm{P}}^{(\ell)}. \tag{9}$$

The block output is then formed by a residual update,

$$\bar{\mathbf{Z}}^{(\ell)} = \mathrm{LN}\left(\mathbf{G}^{(\ell)} + \bar{\mathbf{Z}}^{(\ell-1)}\right). \tag{10}$$

Section B compares this default fusion with two single-branch variants that retain only the temporal feature $\mathbf{H}_{\mathrm{T}}^{(\ell)}$ or only the periodic feature $\mathbf{H}_{\mathrm{P}}^{(\ell)}$. Because $\mathbf{H}_{\mathrm{P}}^{(\ell)}$ already includes the temporal residual $\mathbf{U}^{(\ell)}$, the latter variant is a period-enhanced branch rather than a pure frequency-only model.

After the final block, a linear readout maps latent states back to the target space,

$$\hat{\mathbf{X}} = \mathcal{W}_{\mathrm{out}}\left(\bar{\mathbf{Z}}^{(L)}\right), \tag{11}$$

and the forecast is taken from the last H steps,

$$\hat{\mathbf{Y}} = \hat{\mathbf{X}}_{T+1:T+H}. \tag{12}$$

Finally, the predictions are de-normalized using the instance statistics of the input sequence. The overall architecture is compact and explicitly period-aware, without a general-purpose temporal attention backbone.

4. Experiments

4.1. Experimental Setup

We evaluate on the standard weekly `national_illness` benchmark used in multivariate long-horizon forecasting (Wu et al., 2020). The dataset has seven channels. We use a look-back window of 36 and horizons $H \in \{24, 36, 48, 60\}$. We report MSE and MAE at each horizon and their four-horizon average. Unless otherwise stated, EPIC denotes the reference configuration averaged over three random seeds. In experiments, we use $L = 2$ forecasting blocks, latent dimension $d = 128$, and $N_p = 5$ learnable period prototypes.

Table 1 makes the benchmark setup explicit. Across the full retrospective panel we observe 966 weekly rows, split into 676/97/193 train/validation/test rows. Under the fixed 36-step look-back, carrying boundary context expands the validation and test segments to 133 and 229 effective rows, yielding 74/62/50/38 validation windows and 170/158/146/134 test windows for $H = 24/36/48/60$, respectively.

Table 1. Split sizes and horizon-specific sliding-window counts on `national_illness` under the fixed 36-step look-back and 18-step label context. 'Effective rows' count the context-extended validation and test segments after carrying the preceding look-back history across split boundaries.

Split	Rows	Effective rows	W24	W36	W48	W60
TRAIN	676	676	617	605	593	581
VAL	97	133	74	62	50	38
TEST	193	229	170	158	146	134
TOTAL	966	1038	861	825	789	753

The full-history, ACF, and spectral panels of Figure 1 confirm three features of the target series: short-lag persistence, annual periodicity, and an exception in the 2007–2010 interval coinciding with the 2009 H1N1 pandemic[1].

4.2. Baselines on the Benchmark

The main benchmark tables compare EPIC against two empirical regimes: compact supervised forecasters trained for the ILI benchmark, and pretrained or LLM-adapted forecasters intended to transfer across domains. The comparison tests whether a compact, disease-matched architecture can outperform published compact and LLM-adapted forecasters under the same ILI benchmark. The qualitative examples add representative transformer-style baselines and two zero-shot TimesFM variants. Appendix A lists the individual methods and citations.

Tables 2 and 3 compare published benchmark results with EPIC. MSCformer and Time-LLM are nearly tied on Avg MSE, 1.434 versus 1.435, and MIFlu is close at 1.436. This

[1] https://www.who.int/emergencies/situations/influenza-a-(h1n1)-outbreak

does not show a clear average-error advantage for the LLM-based comparator over the compact baseline. EPIC reaches Avg MSE 1.173, an 18.2% reduction relative to MSCformer and Time-LLM. Table 3 shows a similar margin in MAE. Gains are largest on the four-horizon average and on shorter-to-medium horizons.

The horizon profile matters for interpreting the result. EPIC is most effective when the observed window still contains enough local persistence and annual signal to anchor the forecast: the temporal branch propagates recent dynamics, while the period branch supplies a phase-tolerant seasonal cue. At the longest horizon this cue weakens and the comparison becomes more mixed, so our claim centers on four-horizon average performance rather than uniform pointwise dominance.

Table 2. ILI comparison on the retrospective benchmark in MSE. Lower is better. Published baselines are taken from reported ILI results under the matching benchmark and horizon setting.

Method	H=24	H=36	H=48	H=60	Avg
EPIC	**0.818**	**1.009**	**1.235**	1.630	**1.173**
MSCformer	1.215	1.382	1.411	1.726	1.434
Time-LLM	1.285	1.404	1.523	1.531	1.435
MIFlu	1.542	1.422	1.414	**1.364**	1.436
WhiteTST	1.069	1.382	1.461	1.863	1.443
PatchTST	1.319	1.579	1.553	1.470	1.480
Autoformer	3.483	3.103	2.669	2.770	3.006
Non-Stationary	2.294	1.825	2.010	2.178	2.077
TimesNet	2.317	1.972	2.238	2.027	2.139
xLSTMTime	1.514	1.519	1.500	1.418	1.488
Pathformer	1.587	1.429	1.505	1.731	1.563
TimeCMA	1.996	1.906	1.867	1.920	1.922
OFA	2.063	1.868	1.790	1.979	1.925
CARD	2.043	2.300	1.899	1.993	2.059
UniTime	2.460	1.998	1.979	2.109	2.137

Table 3. ILI comparison on the retrospective benchmark in MAE. Lower is better. Published baselines are taken from reported ILI results under the matching benchmark and horizon setting.

Method	H=24	H=36	H=48	H=60	Avg
EPIC	**0.573**	**0.653**	**0.710**	0.823	**0.690**
MSCformer	0.728	0.743	0.767	0.835	0.768
Time-LLM	0.727	0.814	0.807	0.854	0.801
MIFlu	0.726	0.779	0.757	0.719	0.745
WhiteTST	0.634	0.732	0.771	0.878	0.753
PatchTST	0.754	0.870	0.815	0.788	0.807
Autoformer	1.287	1.148	1.085	1.125	1.161
Non-Stationary	0.945	0.848	0.900	0.963	0.914
TimesNet	0.934	0.920	0.940	0.928	0.931
xLSTMTime	0.694	0.722	0.725	**0.715**	0.714
Pathformer	0.758	0.711	0.742	0.799	0.752
TimeCMA	0.998	0.915	0.868	0.904	0.921
OFA	0.881	0.892	0.884	0.957	0.903
CARD	0.863	0.917	0.846	0.876	0.875
UniTime	0.954	0.912	0.912	0.938	0.929

Figure 2 complements the aggregate results with four qualitative examples at horizon $H = 24$ using 36 observed weeks. The panels are drawn from different parts of the test period to illustrate distinct forecasting scenarios rather than near-duplicate trajectories. Across these examples, EPIC generally tracks the rise-fall dynamics, whereas several baselines exhibit larger phase shifts or amplitude errors. This qualitative comparison complements the benchmark tables under the same evaluation setting.

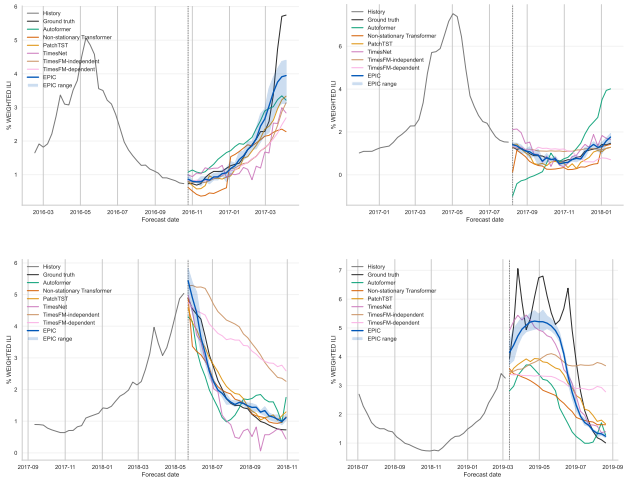


Figure 2. Qualitative forecast comparisons at horizon $H = 24$ using 36 observed weeks. Each panel compares the proposed EPIC configuration with PatchTST, Autoformer, Non-stationary Transformer, TimesNet, and two zero-shot TimesFM variants under the same seven-variable forecasting setting.

The figure also includes the two zero-shot TimesFM variants analyzed in Appendix C and Figures 3 to 6. `TimesFM-independent` forecasts each variable separately, whereas `TimesFM-dependent` incorporates the remaining variables as auxiliary inputs.

5. Conclusion

We presented EPIC, a compact attention-free forecaster for weekly influenza-like illness. The proposed lightweight model separates short-range temporal propagation from slowly drifting annual structure through a causal convolutional path, a learnable period branch, and a latent horizon projector. On the published `national_illness` benchmark, EPIC improves the four-horizon average over compact and LLM-based comparators.

The result supports the central premise of this study: on this weekly ILI benchmark, matching the model bias to the signal structure can matter more than increasing model scale. The benchmark results do not show a clear average-error advantage for LLM-based comparators over strong compact baselines, while the ablations indicate that EPIC benefits from combining temporal and periodic components.

Future work should extend the framework to longer horizons, probabilistic forecasts, and additional epidemic datasets where seasonal recurrence, reporting shifts, and outbreak irregularity interact in different ways.

References

Alharthi, M. and Mahmood, A. xlstmtime: Long-term time series forecasting with xlstm. *AI*, 5(3):1482–1495, 2024. doi: 10.3390/ai5030071.

Biggerstaff, M., Alper, D., Dredze, M., Fox, S., Fung, I. C.-H., Hickmann, K. S., Lewis, B., Rosenfeld, R., Shaman, J., Tsou, M.-H., et al. Results from the centers for disease control and prevention's predict the 2013–2014 influenza season challenge. *BMC Infectious Diseases*, 16(1):357, 2016. doi: 10.1186/s12879-016-1669-x.

Brooks, L. C., Farrow, D. C., Hyun, S., Tibshirani, R. J., and Rosenfeld, R. Flexible modeling of epidemics with an empirical bayes framework. *PLOS Computational Biology*, 11(8): e1004382, 2015. doi: 10.1371/journal.pcbi.1004382.

Chen, P., Zhang, Y., Cheng, Y., Shu, Y., Wang, Y., Wen, Q., Yang, B., and Guo, C. Pathformer: Multi-scale Transformers with Adaptive Pathways for Time Series Forecasting. In *The Twelfth International Conference on Learning Representations, ICLR 2024, Vienna, Austria, May 7-11, 2024*, 2024.

Das, A., Kong, W., Sen, R., and Zhou, Y. A decoder-only foundation model for time-series forecasting. In *Proceedings of the 41st International Conference on Machine Learning*, volume 235 of *ICML'24*, pp. 10148–10167, Vienna, Austria, July 2024. JMLR.org.

Gu, X., Liu, Y., Mohsin, Z., Bedford, J., Thakur, A., Watkinson, P., Clifton, L., Zhu, T., and Clifton, D. Are time series foundation models ready for vital sign forecasting in healthcare? In *Proceedings of the 4th Machine Learning for Health Symposium*, volume 259, pp. 401–419. PMLR, 2025.

Jin, M., Wang, S., Ma, L., Chu, Z., Zhang, J. Y., Shi, X., Chen, P.-Y., Liang, Y., Li, Y.-F., Pan, S., et al. Time-llm: Time series forecasting by reprogramming large language models. In *The Twelfth International Conference on Learning Representations*, 2024.

Kim, J., Kim, H., Kim, H., Lee, D., and Yoon, S. A comprehensive survey of deep learning for time series forecasting: architectural diversity and open challenges. *Artificial Intelligence Review*, 58 (7):216, April 2025. doi: 10.1007/s10462-025-11223-9.

Li, A., Li, Y., Xu, Y., Li, X., and Zhang, C. Multi-scale convolution enhanced transformer for multivariate long-term time series forecasting. *Neural Networks*, 180:106745, 2024. doi: 10.1016/j.neunet.2024.106745.

Liang, Y., Wen, H., Nie, Y., Jiang, Y., Jin, M., Song, D., Pan, S., and Wen, Q. Foundation models for time series analysis: A tutorial and survey. In *Proceedings of the 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, pp. 6555–6565. Association for Computing Machinery, 2024.

Liu, C., Xu, Q., Miao, H., Yang, S., Zhang, L., Long, C., Li, Z., and Zhao, R. Timecma: Towards llm-empowered multivariate time series forecasting via cross-modality alignment. *Proceedings of the AAAI Conference on Artificial Intelligence*, 39(18):18780–18788, 2025. doi: 10.1609/aaai.v39i18.34067.

Liu, X., Hu, J., Li, Y., Diao, S., Liang, Y., Hooi, B., and Zimmermann, R. Unitime: A language-empowered unified model for cross-domain time series forecasting. In *Proceedings of the ACM Web Conference 2024*, pp. 4095–4106. Association for Computing Machinery, 2024.

Liu, Y., Wu, H., Wang, J., and Long, M. Non-stationary Transformers: Exploring the Stationarity in Time Series Forecasting. In Koyejo, S., Mohamed, S., Agarwal, A., Belgrave, D., Cho, K., and Oh, A. (eds.), *Advances in Neural Information Processing Systems 35: Annual Conference on Neural Information Processing Systems 2022, NeurIPS 2022, New Orleans, LA, USA, November 28 - December 9, 2022*, 2022.

Lutz, C. S., Huynh, M. P., Schroeder, M., Anyatonwu, S., Dahlgren, F. S., Danyluk, G., Fernandez, D., Greene, S. K., Kipshidze, N., Liu, L., et al. Applying infectious disease forecasting to public health: a path forward using influenza forecasting examples. *BMC Public Health*, 19(1):1659, 2019. doi: 10.1186/s12889-019-7966-8.

McGowan, C. J., Biggerstaff, M., Johansson, M., Apfeldorf, K. M., Ben-Nun, M., Brooks, L., Convertino, M., Erraguntla, M., Farrow, D. C., Freeze, J., et al. Collaborative efforts to forecast seasonal influenza in the united states, 2015–2016. *Scientific Reports*, 9(1):683, 2019. doi: 10.1038/s41598-018-36361-9.

Moon, J., Shim, J., Kim, E., and Hwang, E. Miflu: Large language model-based multimodal influenza forecasting scheme. *IEEE Journal of Biomedical and Health Informatics*, 29(10):7790–7801, 2025. doi: 10.1109/JBHI.2025.3561214.

Nie, Y., Nguyen, N. H., Sinthong, P., and Kalagnanam, J. A time series is worth 64 words: Long-term forecasting with transformers. In *The Eleventh International Conference on Learning Representations*, 2023.

Oh, Y. and Bui, A. Multi-View Contrastive Learning for Robust Domain Adaptation in Medical Time Series Analysis. In Xu, X. O., Choi, E., Singhal, P., Gerych, W., Tang, S., Agrawal, M., Subbaswamy, A., Sizikova, E., Dunn, J., Daneshjou, R., Sarker, T., McDermott, M., and Chen, I. (eds.), *Proceedings of the sixth Conference on Health, Inference, and Learning*, volume 287, pp. 502–526, Proceedings of Machine Learning Research, 2025. PMLR.

Reich, N. G., Brooks, L. C., Fox, S. J., Kandula, S., McGowan, C. J., Moore, E., Osthus, D., Ray, E. L., Tushar, A., Yamana, T. K., et al. A collaborative multiyear, multimodel assessment of seasonal influenza forecasting in the united states. *Proceedings of the National Academy of Sciences*, 116(8):3146–3154, 2019. doi: 10.1073/pnas.1812594116.

Shen, R., Lin, Y., Wang, B., Liu, L., Guan, Y., and Jiang, J. Forecasting influenza like illness based on white-box transformers. In *2024 IEEE International Conference on Bioinformatics and Biomedicine*, pp. 3691–3694, 2024. doi: 10.1109/BIBM62325.2024.10822536.

Wang, X., Zhou, T., Wen, Q., Gao, J., Ding, B., and Jin, R. CARD: Channel Aligned Robust Blend Transformer for Time Series Forecasting. In *The Twelfth International Conference on Learning Representations, ICLR 2024, Vienna, Austria, May 7-11, 2024*, 2024.

Wang, Y., Wu, H., Dong, J., Liu, Y., Wang, C., Long, M., and Wang, J. Deep Time Series Models: A Comprehensive Survey and Benchmark. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pp. 1–20, 2026. doi: 10.1109/TPAMI.2026.3690845.

Wu, H., Xu, J., Wang, J., and Long, M. Autoformer: Decomposition Transformers with Auto-Correlation for Long-Term Series Forecasting. In Ranzato, M., Beygelzimer, A., Dauphin, Y., Liang, P. S., and Vaughan, J. W. (eds.), *Advances in Neural Information Processing Systems*, volume 34, pp. 22419–22430. Curran Associates, Inc., 2021.

Wu, H., Hu, T., Liu, Y., Zhou, H., Wang, J., and Long, M. TimesNet: Temporal 2D-Variation Modeling for General Time Series Analysis. In *The Eleventh International Conference on Learning Representations, ICLR 2023, Kigali, Rwanda, May 1-5, 2023*, 2023.

Wu, N., Green, B., Ben, X., and O'Banion, S. Deep transformer models for time series forecasting: The influenza prevalence case, 2020.

Wu, Y., Yang, Y., Nishiura, H., and Saitoh, M. Deep Learning for Epidemiological Predictions. In *The 41st International ACM SIGIR Conference on Research & Development in Information Retrieval*, SIGIR '18, pp. 1085–1088, New York, NY, USA, June 2018. Association for Computing Machinery. doi: 10.1145/3209978.3210077.

Zeng, A., Chen, M., Zhang, L., and Xu, Q. Are Transformers Effective for Time Series Forecasting? *Proceedings of the AAAI Conference on Artificial Intelligence*, 37(9):11121–11128, June 2023. doi: 10.1609/aaai.v37i9.26317.

Zhou, T., Niu, P., Wang, X., Sun, L., and Jin, R. One fits all: Power general time series analysis by pretrained lm. In *Advances in Neural Information Processing Systems*, volume 36, 2023.

A. Compared Methods

The compared methods fall into three groups: compact supervised forecasters, pretrained or LLM-adapted forecasters, and zero-shot TimesFM variants. The broader motivation comes from surveys and benchmarks of deep time-series forecasting (Kim et al., 2025; Wang et al., 2026), reusable time-series foundation or LLM-adapted forecasters (Liang et al., 2024; Jin et al., 2024; Liu et al., 2024; 2025), and compact supervised architectures that remain strong on long-horizon benchmarks (Zeng et al., 2023; Nie et al., 2023).

The main benchmark tables compare EPIC with compact supervised forecasters, including PatchTST (Nie et al., 2023), WhiteTST (Shen et al., 2024), xLSTMTime (Alharthi & Mahmood, 2024), MSCformer (Li et al., 2024), Pathformer (Chen et al., 2024), and CARD (Wang et al., 2024). They also compare against pretrained or LLM-adapted forecasters, including OFA (Zhou et al., 2023), Time-LLM (Jin et al., 2024), UniTime (Liu et al., 2024), TimeCMA (Liu et al., 2025), and MIFlu (Moon et al., 2025). The separate zero-shot TimesFM (Das et al., 2024) study in Appendix C includes both a matched $T = 36$ comparison and a longer-context sensitivity analysis.

B. Discussion and Ablation Details

Table 4. EPIC ablation on `national_illness` in MSE.

Variant	Projection	H=24	H=36	H=48	H=60	Avg
EPIC	NL	**0.818**	**1.009**	1.235	**1.630**	**1.173**
Temporal-only	NL	0.849	1.144	**1.169**	1.725	1.222
Period-only	NL	0.867	1.222	1.267	1.702	1.265
EPIC	L	0.866	1.252	1.240	1.737	1.274
Temporal-only	L	0.936	1.211	1.381	1.765	1.323
Period-only	L	0.993	1.373	1.399	1.686	1.363

Table 5. EPIC ablation on `national_illness` in MAE.

Variant	Projection	H=24	H=36	H=48	H=60	Avg
EPIC	NL	**0.573**	**0.653**	0.710	**0.823**	**0.690**
Temporal-only	NL	0.583	0.682	**0.702**	0.830	0.699
Period-only	NL	0.591	0.699	0.720	0.829	0.710
EPIC	L	0.592	0.728	0.731	0.843	0.724
Temporal-only	L	0.626	0.709	0.768	0.849	0.738
Period-only	L	0.629	0.732	0.758	0.828	0.737

In both MSE and MAE, the full model achieves the lowest four-horizon average among the ablated variants, improving Avg MSE from 1.222 and 1.265 for the temporal-only and period-enhanced variants to 1.173, and Avg MAE from 0.699 and 0.710 to 0.690. This pattern is consistent with complementary contributions from short-range propagation and recurring periodic structure. The temporal-only model remains slightly better at $H = 48$, but the full model is better on the average and on the other reported horizons.

Replacing the nonlinear projector with the linear control increases the four-horizon average error for all three variants. For the full model, Avg MSE rises from 1.173 to 1.274, and Avg MAE rises from 0.690 to 0.724. The same degradation appears for the temporal-only and period-enhanced rows. Both the joint branch design and the nonlinear projector contribute independently to the four-horizon average error.

The main exception is the 60-step horizon, where EPIC records MSE 1.630, compared with 1.531 for Time-LLM, 1.470 for PatchTST, and 1.364 for MIFlu. The MAE results show the same pattern: EPIC records 0.823, worse than PatchTST (0.788), MIFlu (0.719), and xLSTMTime (0.715). We therefore do not claim uniform superiority at the longest horizon. Accordingly, we limit our claim to average performance on the published `national_illness` benchmark.

C. Matched Zero-Shot TimesFM Comparison

We compare EPIC with zero-shot TimesFM under the same seven-variable forecasting benchmark as in the main pipeline. The comparison treats TimesFM as a zero-shot transfer model rather than a supervised model tuned on the ILI benchmark. `TimesFM-independent` forecasts each channel separately and then stacks the resulting outputs. `TimesFM-dependent` forecasts one channel at a time while using the other six channels as auxiliary inputs, with future values supplied by the corresponding independent forecasts. Because those auxiliary future values are generated rather than observed, the dependent setting should be read as a practical zero-shot adaptation rather than an oracle.

Tables 6 and 7 show that matching the context length does not make zero-shot TimesFM competitive with trained EPIC. Longer contexts help, especially for `TimesFM-independent`, but the best zero-shot row still remains above EPIC in both Avg MSE and Avg MAE. We treat the longer-context rows as a sensitivity analysis rather than as a direct substitute for the trained EPIC result.

Table 6. MSE comparison between EPIC and zero-shot TimesFM variants under multiple input context lengths.

Method	Context (T)	H=24	H=36	H=48	H=60	Avg
EPIC	36	**0.818**	**1.009**	**1.235**	**1.630**	**1.173**
TimesFM-independent	36	2.940	4.046	4.983	5.181	4.287
	54	2.478	2.781	3.578	3.295	3.033
	72	2.394	2.609	2.601	2.517	2.530
	90	2.021	1.877	1.779	1.763	1.860
	108	1.655	2.084	2.213	2.404	2.089
TimesFM-dependent	36	3.251	4.169	4.970	5.012	4.351
	54	2.783	2.697	3.948	4.112	3.385
	72	2.662	3.088	3.068	3.066	2.971
	90	2.393	2.134	2.188	2.414	2.282
	108	2.044	2.532	2.982	3.254	2.703

Table 7. MAE comparison between EPIC and zero-shot TimesFM variants under multiple input context lengths.

Method	Context (T)	H=24	H=36	H=48	H=60	Avg
EPIC	36	**0.573**	**0.653**	**0.710**	**0.823**	**0.690**
TimesFM-independent	36	1.124	1.356	1.522	1.552	1.388
	54	0.984	1.068	1.208	1.170	1.108
	72	0.975	1.057	1.060	1.069	1.041
	90	0.875	0.883	0.885	0.895	0.885
	108	0.827	0.934	0.973	0.990	0.931
TimesFM-dependent	36	1.198	1.414	1.581	1.605	1.450
	54	1.045	1.119	1.337	1.339	1.210
	72	1.068	1.158	1.155	1.155	1.134
	90	0.961	0.948	0.972	0.998	0.970
	108	0.911	1.021	1.111	1.116	1.040

Figures 3, 4, 5, and 6 compare the same four test cases across forecast horizons. Because the cases are fixed, differences across panels primarily reflect how each method handles longer-range phase drift and amplitude error.

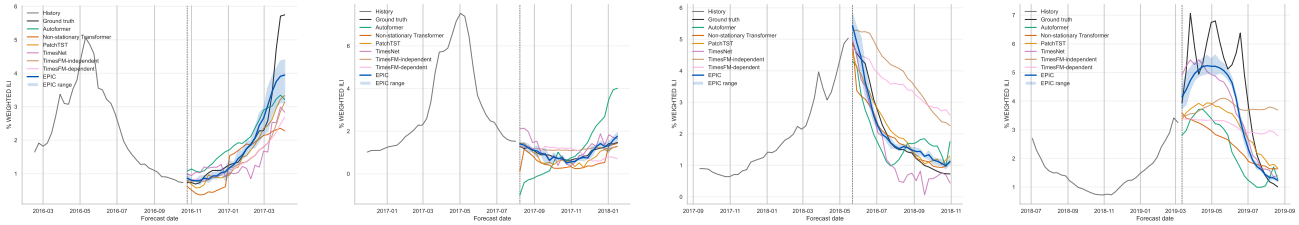


Figure 3. Qualitative forecast comparisons at $H = 24$.

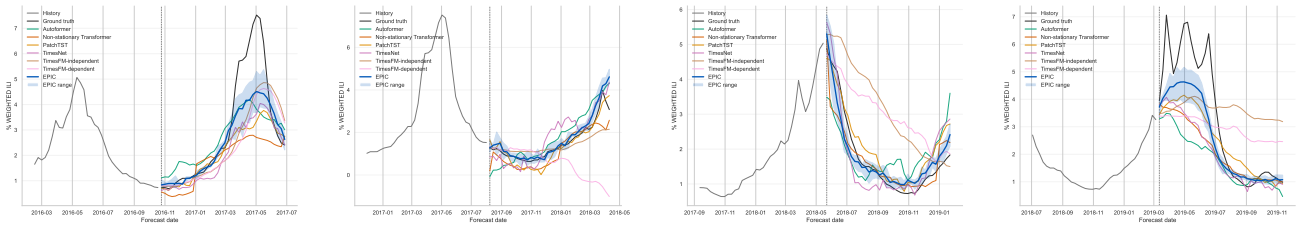


Figure 4. Qualitative forecast comparisons at $H = 36$.

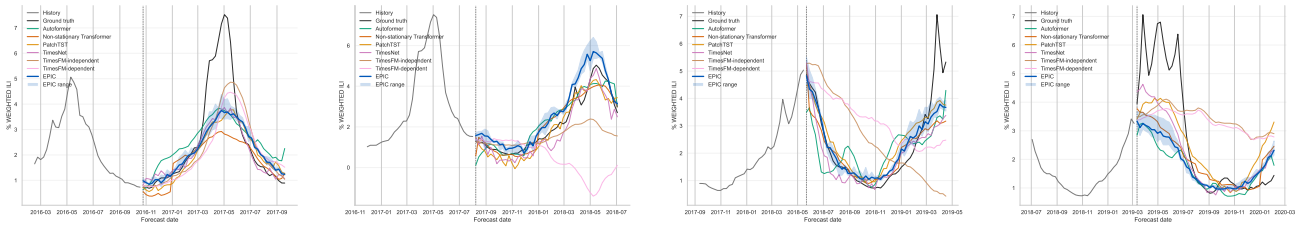


Figure 5. Qualitative forecast comparisons at $H = 48$.

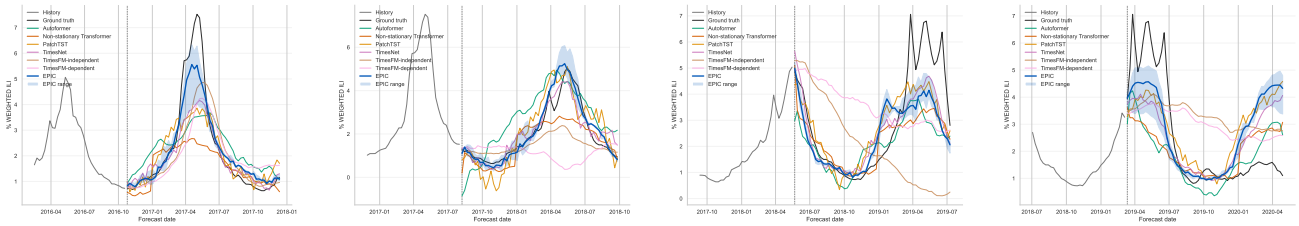


Figure 6. Qualitative forecast comparisons at $H = 60$.