MULTIVARIATE BEHAVIORAL RESEARCH
2026, VOL. 61, NO. 3, 447–448

When Deep Learning Outperforms Conventional Models in Long-Term Forecasting

Young Won Cho
Department of Human Development and Family Studies, The Pennsylvania State University, University Park, PA, USA
Sy-Miin Chow
Department of Human Development and Family Studies, Social Science Research Institute, The Pennsylvania State University, University Park, PA, USA
Danielle Symons Downs
Departments of Kinesiology and Obstetrics and Gynecology, Social Science Research Institute, The Pennsylvania State University, University Park, PA, USA
Steriani Elavsky
Department of Human Movement Studies, University of Ostrava, Ostrava, Czech Republic

1. Introduction

Forecasting long-term health trajectories from inten-sive longitudinal data (ILD) is essential for timely interventions. Growth Curve Models (GCMs) provide a conventional multilevel framework that offers inter-pretable and sample-efficient forecasts when trajecto-ries exhibit smooth structural change. However, when trajectories show nonlinear or idiosyncratic departures from assumed functional forms, they require extensive specification of interactions or nonlinearities. In con-trast, deep learning (DL) models offer an assumption-light approach. By learning temporal dynamics through nonlinear transformations, DL models can capture complex, emergent patterns without prespeci-fying functional structures (Kontopoulou et al., 2023) and thus, have impact on advancing early and person-alized interventions in health sciences.

The present study evaluated whether, and in what ways, DL offers practical advantages in behavioral forecasting with ILD compared to GCMs. We com-pared approaches across two datasets. In Healthy-Mom-Zone (HMZ) 1.0 study (Downs et al., 2018), we forecasted 2nd and 3rd trimester weights with 1st tri-mester data. These trajectories show a gradual increase over pregnancy with small day-to-day fluctuations and individual differences in overall patterns. The partici-pant sample is relatively small (N 25) despite the dense time series (T 200). In the second dataset, Healthy-Aging-in-Industrial-Environment (4HAIE) study (N 197, T 250) (Elavsky et al., 2021), pre-lockdown ILD were used to forecast daily step counts during and after the COVID-19 lockdown. Daily steps show high day-to-day variability and between-person heterogeneity, Limited theory exists on which factors drove severe declines in physical activity during and after the lockdown.

Two modeling families were compared: (1) quadratic GCMs and (2) sequence-to-sequence (Seq2Seq) neural networks with either LSTM (Long Short-Term Memory) or Transformer encoder–decoder cells, which capture nonlinear and long-range temporal dependencies. Seq2Seqs were adopted to incorporate person-level covariates into the decoder's initial state to prevent individual information from fading during long-term forecasts and to capture individual-by-tra-jectory interactions. Performance was assessed by fore-casting accuracy. As baseline forecasts, HMZ used the Institute of Medicine's mean recommended weight gain based on pre-pregnancy Body Mass Index cat-egory, and 4HAIE used the pre-lockdown median step count. We used a cross-individual

forecasting para-digm: models were trained on full trajectories from training participants and then forecasted outcomes for new individuals using their early data, leveraging shared population-level patterns.

Overall, GCMs excelled when trajectories followed stable, smooth growth with limited heterogeneity, whereas DL models, particularly Transformers, gained advantage under abrupt changes, heterogeneous responses, and high-variance temporal patterns. In HMZ, GCM outperformed LSTM. For third-trimester weight, GCM achieved a mean absolute error (MAE) of 3.46 lbs, compared with 7.04 for the baseline and 5.18 for LSTM. Transformers were not tested in HMZ due to the small sample. In 4HAIE, forecasting during lockdown was more difficult than post-lockdown. During lockdown, LSTM (MAE: 3224) and Transformer (3254) outperformed GCM (3417), reflecting their ability to model abrupt and heteroge-neous shifts. Post-lockdown, Transformers (2818) matched or exceeded GCM (2824), followed by the median baseline (3083) and LSTM (3161). Due to high day-to-day variability, the Transformer's atten-tion mechanism handled volatility best. GCM showed competitive MAE but weaker trajectory tracking.

2. Funding

The preparation of the abstract was supported from the project, Research of Excellence on Digital Technologies and Wellbeing CZ.02.01.01/00/22_008/0004583 "which is co-financed by the European Union. The author(s) reported there is no funding associated with the work featured in this article.

References

Downs, D. S., Savage, J. S., Rivera, D. E., Smyth, J. M.,

Rolls, B. J., Hohman, E. E., … Guo, P. (2018). Individually tailored, adaptive intervention to manage gestational weight gain: Protocol for a randomized con-trolled trial in women with overweight and obesity. *JMIR Research Protocols*, *7*(6), e9220.

Elavsky, S., Janda˘ckov�a, V., Knapov�a, L., Va˘sendov�a, V., Sebera, M., Ka˘stovsk�a, B., Blaschov�a, D., Ku€hnov�a, J., Cimler, R., Vil�ımek, D., Bosek, T., Koenig, J., &

Janda˘cka, D. (2021). Physical activity in an air-polluted environment: Behavioral, psychological and neuroimaging protocol for a prospective cohort study (Healthy Aging in Industrial Environment study–Program 4). *BMC Public Health*, *21*(1), 126. https://doi.org/10.1186/s12889-021-

10166-4

Kontopoulou, V. I., Panagopoulos, A. D., Kakkos, I., & Matsopoulos, G. K. (2023). A review of ARIMA vs. machine learning approaches for time series forecasting in data driven networks. *Future Internet*, *15*(8), 255. https://doi.org/10.3390/fi15080255

Symons Downs, D., Savage, J. S., Rivera, D. E., Smyth,

J. M., Rolls, B. J., Hohman, E. E., McNitt, K. M., Kunselman, A. R., Stetter, C., Pauley, A. M., Leonard, K. S., & Guo, P. (2018). Individually Tailored, Adaptive Intervention to Man age Gestational Weight Gain: Protocol for a Randomized Controlled Trial in Women With Overweight and Obesity. *JMIR research protocols*, *7*(6), e150. https:// doi.org/10.2196/resprot.9220