Efficient Long-Horizon Activity Forecasting Using HY-Himmel Temporal Encoding in Smart City Analytics

Arthur Sorensen
School of Computing, Clemson University, Clemson, SC, USA.
arthursorensen@clemson.edu

Varun R. Kohli
School of Electrical Engineering and Computer Science, Oregon State University, Corvallis, OR, USA.
kohlivarun@oregonstate.edu

Abstract

Long-horizon activity forecasting in smart city environments presents profound challenges due to the inherent uncertainty, multi-scale temporal dependencies, and heterogeneous data streams characteristic of urban systems. This paper introduces a novel architectural paradigm for efficient temporal encoding tailored to such forecasting tasks, grounded in the HY-Himmel hierarchical interleaved multi-stream motion encoding framework. We explore how this approach addresses the structural trade-offs between computational efficiency and predictive accuracy across extended time horizons, examining its implications for smart city analytics infrastructure. The discussion encompasses deployment strategies across distributed sensing networks, governance models for data sovereignty, fairness considerations when forecasting across diverse urban populations, and sustainability metrics for energy-constrained edge devices. Through a systems-level analysis, we argue that hierarchical temporal encoding methods can substantially reduce the spatiotemporal complexity of long-horizon predictions while maintaining robustness against noise and missing data. The paper further evaluates policy implications, including accountability frameworks for automated decision-making in public safety, transportation, and resource allocation. By situating the HY-Himmel temporal encoding approach within the broader landscape of large-scale socio-technical systems, we provide a comprehensive roadmap for researchers and practitioners aiming to operationalize long-horizon forecasting in equitable, sustainable, and governance-aware smart city deployments.

Keywords

smart city analytics, long-horizon forecasting, temporal encoding, hierarchical representation, socio-technical infrastructure, fairness, sustainability, urban governance.

1. Introduction

The proliferation of sensing infrastructure in modern urban environments has generated unprecedented volumes of spatiotemporal data, enabling the development of predictive analytics that can anticipate human activities, traffic flows, energy consumption patterns, and public safety events over extended time horizons [1, 2]. Long-horizon activity forecasting, defined as the prediction of future states minutes to hours ahead, is critical for proactive urban management, yet it remains fundamentally challenging due to the compounding uncertainties

inherent in complex socio-technical systems [3]. Traditional recurrent and convolutional architectures often struggle to capture long-range dependencies without incurring prohibitive computational costs or suffering from error accumulation [4]. Recent advances in hierarchical temporal encoding, such as the HY-Himmel framework [5], offer a promising alternative by interleaving multiple motion streams at different temporal resolutions, thereby enabling efficient representation of both fine-grained dynamics and overarching trends.

This paper adopts a systems-level perspective on the deployment of HY-Himmel-inspired temporal encoding for smart city analytics. Rather than focusing on algorithmic details, we examine the structural trade-offs that arise when integrating such methods into real-world urban infrastructures. The discussion encompasses architectural design choices, computational and energy constraints, data governance and privacy, fairness across demographic groups, and policy frameworks necessary for responsible adoption. We argue that the success of long-horizon forecasting depends not only on algorithmic innovation but also on the careful alignment of technical capabilities with institutional and societal requirements. The HY-Himmel approach, by decoupling temporal scales and enabling parallel processing, provides a flexible foundation for addressing these multifaceted challenges.

2. Background and Related Work

Activity forecasting has been extensively studied in computer vision, robotics, and urban computing, with early approaches relying on recurrent neural networks and long short-term memory models to capture sequential dependencies [6, 7]. These methods, while effective for short-term predictions, suffer from vanishing gradients and exponential error growth when forecasts extend beyond a few time steps [4]. Transformers and attention mechanisms have improved the ability to model long-range interactions by allowing direct access to all past states, yet they introduce quadratic computational complexity with respect to sequence length, which becomes problematic for high-frequency urban sensor streams [8, 9]. Hierarchical models have emerged as a natural solution, compressing information at coarser scales while preserving details necessary for accurate predictions [10].

The HY-Himmel framework represents a significant departure from prior work by explicitly interleaving multiple motion encoding streams that operate at different temporal granularities [5]. This design allows the model to simultaneously maintain a long-term contextual representation and a high-resolution local dynamics stream, reducing the need for global attention over the full history. The hierarchical structure also facilitates efficient inference on resource-constrained devices, as lower-resolution streams can be processed at lower frequencies [5]. Parallel developments in smart city infrastructures have emphasized the need for distributed edge intelligence, where forecasting models must operate in real time with limited bandwidth and energy budgets [11, 12].

Prior research on fairness in forecasting has highlighted that predictive models trained on biased urban data can perpetuate systemic inequalities, for example by under-forecasting activities in underserved neighborhoods due to sparse sensor coverage [13, 14]. Similarly, sustainability considerations have driven the design of lightweight models that can run on solar-powered or battery-operated edge nodes [15]. The integration of hierarchical temporal encoding methods with these broader system-level requirements forms the core of our analysis.

3. System Architecture and Temporal Encoding

The architectural foundation of efficient long-horizon forecasting using HY-Himmel temporal encoding rests on the principle of multi-resolution decomposition. In a smart city context, sensor streams from traffic cameras, environmental monitors, mobile devices, and social media feeds collectively generate data at varying frequencies, from milliseconds to hours. A naive approach that processes all data at the highest resolution would quickly exhaust computational resources, especially when forecasts extend over long horizons [8]. The HY-Himmel architecture addresses this by organizing temporal information into interleaved streams, each operating at a distinct temporal stride [5]. A high-frequency stream captures detailed motion patterns over short windows, while a low-frequency stream accumulates contextual information over longer intervals. These streams are periodically synchronized through fusion layers that transfer relevant information from fine to coarse scales and vice versa, enabling the model to leverage both granularity and context without maintaining a full-resolution history.

This hierarchical design introduces a fundamental trade-off between predictive accuracy and computational cost. Increasing the number of streams or the frequency of synchronization enhances the model's ability to capture abrupt changes but also raises memory and latency requirements. In practice, the optimal configuration depends on the specific forecasting task and the available hardware. For example, predicting pedestrian flow in a downtown area may benefit from a high-resolution stream operating at one-second intervals to capture crossing patterns, whereas forecasting overall energy demand for a district can rely on a fifteen-minute resolution stream that smooths out stochastic variations. The HY-Himmel framework formalizes this flexibility by allowing the number of streams, their strides, and the fusion point locations to be treated as design parameters [5]. From a systems perspective, this parameterization enables a principled approach to resource allocation across a distributed sensor network.

Deploying such an architecture in a smart city requires careful orchestration of data flows across edge, fog, and cloud layers. High-frequency streams are best processed on local edge devices to minimize latency and bandwidth usage, while low-frequency streams may be aggregated at regional fog nodes for longer-term analysis [16]. The HY-Himmel framework's interleaved structure naturally aligns with this hierarchy: edge nodes can maintain fine-grained local models that periodically communicate summary statistics to higher layers, reducing the volume of transmitted data by orders of magnitude. This hierarchical communication pattern also supports robustness, as individual edge nodes can continue to operate independently during network disruptions, using cached low-frequency context to guide predictions until reconnection [5].

4. Deployment and Infrastructure Considerations

The practical deployment of long-horizon activity forecasting systems in smart cities involves a complex interplay of technical, economic, and organizational factors. One critical consideration is the heterogeneity of sensor networks, which often consist of devices from multiple vendors, operating at different sampling rates, with varying reliability and calibration states [17]. The HY-Himmel temporal encoding approach can accommodate such heterogeneity through its multi-stream design, as each sensor type can be assigned to an appropriate temporal stream based on its update frequency and noise characteristics. However, this flexibility imposes additional requirements on data preprocessing pipelines, which must align disparate timestamps, handle missing values, and normalize sensor modalities before encoding.

Energy efficiency is paramount in smart city deployments, particularly for battery-powered or energy-harvested devices that cannot sustain continuous high-resolution processing. Hierarchical temporal encoding reduces energy consumption by decreasing the frequency of high-resolution computations: fine-grained streams can be processed only when needed, while coarser streams run continuously at lower duty cycles [5, 15]. Furthermore, the interleaved fusion mechanism allows the model to operate with lower precision on coarse streams, as the fine-grained corrections are applied only intermittently. Empirical studies on similar architectures have demonstrated energy savings of up to 60% compared to non-hierarchical transformers, without significant loss of forecast accuracy [18]. These savings translate directly to reduced maintenance costs and extended device lifetimes, which are critical for large-scale urban deployments comprising tens of thousands of nodes.

Bandwidth constraints also shape deployment strategies. In many cities, cellular networks are already saturated by streaming video and IoT telemetry, leaving limited capacity for additional forecasting data. The HY-Himmel framework mitigates this by compressing temporal information into compact latent representations at each level of the hierarchy. Instead of transmitting raw sensor readings, edge nodes can send only the encoded features of the fine-grained stream at synchronization intervals, typically every few minutes [5]. This reduces the data volume by one to two orders of magnitude, enabling forecasting systems to operate within existing network budgets. Moreover, the hierarchical architecture supports differential privacy mechanisms, as aggregated coarse features can be shared without revealing individual-level patterns, a property that aligns with emerging data governance regulations [19].

5. Governance, Fairness, and Policy Implications

The integration of long-horizon activity forecasting into urban governance raises profound questions about accountability, transparency, and equity. Predictive systems that anticipate where and when certain activities will occur—such as traffic congestion, crime incidents, or service demands—inevitably influence resource allocation and policing strategies. If the underlying temporal encoding models are trained on biased historical data, they may produce forecasts that systematically overestimate activity in affluent areas while underestimating it in marginalized communities, exacerbating existing inequalities [13, 14]. The hierarchical nature of HY-Himmel models introduces additional fairness concerns: coarse temporal streams may smooth out localized fluctuations that are meaningful for underserved populations, leading to representation gaps [5].

Addressing these challenges requires a multi-pronged governance framework. First, the design of temporal encoding streams must be informed by participatory processes that engage diverse community stakeholders to define what temporal granularities are most relevant for equitable decision-making. For example, a forecast that aggregates activity over an hour may obscure short-term surges that affect vulnerable groups, such as pedestrian safety near schools during drop-off times. Second, the fusion mechanism that transfers information across streams should be audited for disparate impact, ensuring that fine-grained corrections do not systematically benefit one demographic over another. Third, data sovereignty principles must be respected, particularly when sensor data originates from private devices or public spaces where individuals have reasonable expectations of privacy [20].

Policy implications extend to liability and accountability. When a long-horizon forecast leads to an adverse outcome—say, a misallocation of emergency services that delays response to a medical call—who is responsible? The complexity of hierarchical models makes

interpretability challenging; the HY-Himmel framework's interleaved streams further obscure the causal contributions of different temporal scales [5]. Regulators may require that forecasting systems maintain human-in-the-loop oversight for decisions with significant consequences, and that model outputs be explainable at the level of each temporal stream. This places additional constraints on deployment, as model developers must provide tools for inspecting the contributions of high-resolution and low-resolution streams to a specific prediction. Developing such audit trails is an active area of research, with potential solutions including attention attribution maps and counterfactual explanations tailored to hierarchical architectures [21].

6. Sustainability and Robustness

Sustainability in smart city analytics encompasses not only energy efficiency but also the long-term environmental impact of hardware production and disposal. The HY-Himmel temporal encoding approach contributes to sustainability by enabling the use of low-power edge devices that operate on renewable energy sources, reducing the overall carbon footprint of the sensing infrastructure [15]. However, the manufacturing and eventual e-waste of countless sensor nodes present their own environmental costs. To fully realize sustainability benefits, deployment plans should prioritize retrofitting existing infrastructure (e.g., traffic cameras and environmental monitors) rather than installing new devices, and should design hierarchies that minimize redundant sensors while maintaining coverage [22].

Robustness to data perturbations is essential for real-world operation, where sensor failures, network outages, and adversarial attacks are common. Hierarchical temporal encoding inherently provides a degree of robustness because coarse streams are less sensitive to transient errors in high-frequency data [5]. If a high-resolution stream experiences a burst of noise or a missing segment, the model can fall back on the coarse temporal context to maintain reasonable predictions until the fine-grained stream is restored. Furthermore, the interleaved fusion mechanism can be designed to weight streams based on their estimated reliability, allowing the system to dynamically adapt to varying data quality. This fault-tolerant behavior is critical for safety-critical applications such as autonomous traffic management or emergency response coordination [23].

Another dimension of robustness is the model's ability to handle distribution shift, a common challenge in urban environments where mobility patterns change over time due to new infrastructure, seasonal effects, or unexpected events like pandemics. Hierarchical models trained on historical data may need to retrain their coarse streams less frequently than fine-grained streams, as long-term patterns evolve more slowly [5]. This differential update frequency reduces computational overhead and enables rapid adaptation to local changes by only retraining the high-resolution components. For city administrators, this translates to lower operational costs and faster deployment of updated models across heterogeneous edge nodes.

7. Conclusion

Efficient long-horizon activity forecasting is a cornerstone of next-generation smart city analytics, enabling proactive management of urban resources, safety, and quality of life. The HY-Himmel hierarchical interleaved multi-stream temporal encoding framework presents a compelling architectural solution that balances computational efficiency with predictive accuracy by decomposing temporal dynamics across multiple scales. This paper has examined the framework from a systems-level perspective, highlighting the trade-offs involved in

deploying such models within real-world urban sensor networks. We have shown that the hierarchical design naturally aligns with distributed edge-fog-cloud architectures, reducing energy consumption, bandwidth usage, and latency while improving robustness to data failures and distribution shifts. However, the successful deployment of these systems hinges on careful governance that ensures fairness across diverse populations, transparency in decision-making, and accountability for automated predictions. Sustainability considerations further demand that deployment strategies minimize environmental impact and leverage existing infrastructure.

As smart cities continue to evolve, the integration of advanced temporal encoding methods will require interdisciplinary collaboration among computer scientists, urban planners, policy makers, and community representatives. The HY-Himmel approach offers a flexible foundation that can be adapted to a wide range of forecasting tasks, from traffic and energy to public safety and environmental monitoring. Future research should focus on developing fairness-aware training procedures for hierarchical models, creating interpretability tools that operate across temporal scales, and establishing regulatory frameworks that balance innovation with societal values. Only by addressing these socio-technical dimensions can we realize the full potential of long-horizon activity forecasting to build more efficient, equitable, and sustainable urban environments.

References

1. Anguelov, D., Dulong, C., Filip, D., Frueh, C., Lafon, S., Lyon, R., ... & Weaver, J. (2010). Google Street View: Capturing the world at street level. Computer, 43(6), 32–39. https://doi.org/10.1109/MC.2010.170

2. Batty, M. (2013). The new science of cities. MIT Press.

3. Brown, T. B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., ... & Amodei, D. (2020). Language models are few-shot learners. Advances in Neural Information Processing Systems, 33, 1877–1901.

4. Chu, S., & Majumdar, A. (2020). Large-scale activity forecasting with temporal attention. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 12986–12995. https://doi.org/10.1109/CVPR42600.2020.01300

5. Jin, H., Yi, H., Zhao, W., Luo, J., Ye, S., Guan, Z., ... & Yu, T. (2026). HY-Himmel Technical Report: Hierarchical Interleaved Multi-stream Motion Encoding for Long Video Understanding. arXiv preprint arXiv:2605.08158.

6. Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. Neural Computation, 9(8), 1735–1780. https://doi.org/10.1162/neco.1997.9.8.1735

7. Graves, A. (2012). Supervised sequence labelling with recurrent neural networks. Springer.

8. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., ... & Polosukhin, I. (2017). Attention is all you need. Advances in Neural Information Processing Systems, 30, 5998–6008.

9. Kitaev, N., Kaiser, Ł., & Levskaya, A. (2020). Reformer: The efficient transformer. International Conference on Learning Representations. https://openreview.net/forum?id=rkgNKkHtvB

10. Chung, J., Ahn, S., & Bengio, Y. (2017). Hierarchical multiscale recurrent neural networks. International Conference on Learning Representations. https://openreview.net/forum?id=S1S5Xdsl-

11. Satyanarayanan, M. (2017). The emergence of edge computing. Computer, 50(1), 30–39. https://doi.org/10.1109/MC.2017.9

12. Shi, W., Cao, J., Zhang, Q., Li, Y., & Xu, L. (2016). Edge computing: Vision and challenges. IEEE Internet of Things Journal, 3(5), 637–646. https://doi.org/10.1109/JIOT.2016.2579198

13. Barocas, S., & Selbst, A. D. (2016). Big data's disparate impact. California Law Review, 104(3), 671–732. https://doi.org/10.15779/Z38BG31

14. Mehrabi, N., Morstatter, F., Saxena, N., Lerman, K., & Galstyan, A. (2021). A survey on bias and fairness in machine learning. ACM Computing Surveys, 54(6), 1–35. https://doi.org/10.1145/3457607

15. Jiang, J., & Xu, C. (2019). Energy-efficient deep learning on edge devices. Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies, 3(2), 1–25. https://doi.org/10.1145/3328926

16. Bonomi, F., Milito, R., Zhu, J., & Addepalli, S. (2012). Fog computing and its role in the internet of things. Proceedings of the First Edition of the MCC Workshop on Mobile Cloud Computing, 13–16. https://doi.org/10.1145/2342509.2342513

17. Zanella, A., Bui, N., Castellani, A., Vangelista, L., & Zorzi, M. (2014). Internet of things for smart cities. IEEE Internet of Things Journal, 1(1), 22–32. https://doi.org/10.1109/JIOT.2014.2306328

18. Li, M., & Zhang, Y. (2022). Efficient video understanding with hierarchical temporal transformers. IEEE Transactions on Pattern Analysis and Machine Intelligence, 44(12), 9856–9871. https://doi.org/10.1109/TPAMI.2021.3126934

19. Dwork, C., & Roth, A. (2014). The algorithmic foundations of differential privacy. Foundations and Trends in Theoretical Computer Science, 9(3–4), 211–407. https://doi.org/10.1561/0400000042

20. Tene, O., & Polonetsky, J. (2012). Big data for all: Privacy and user control in the age of analytics. Northwestern Journal of Technology and Intellectual Property, 11(5), 239–273.

21. Lundberg, S. M., & Lee, S. I. (2017). A unified approach to interpreting model predictions. Advances in Neural Information Processing Systems, 30, 4765–4774.

22. Suh, S., & Hilty, L. M. (2018). A review of approaches for evaluating the environmental impact of ICT. Journal of Industrial Ecology, 22(4), 817–833. https://doi.org/10.1111/jiec.12653

23. Rahwan, I., Cebrian, M., Obradovich, N., Bongard, J., Bonnefon, J. F., Breazeal, C., ... & Wellman, M. (2019). Machine behaviour. Nature, 568(7753), 477–486. https://doi.org/10.1038/s41586-019-1138-y