Digital Neurophysiology of Sleep: Explainable Machine Learning Frameworks for Modeling Proton-Driven Brain Homeostasis and Sleep Regulation

Keepak Mhah
School of Electrical Engineering and Computer Science, Oregon State University, Corvallis, OR, USA.
hellodeepak@oregonstate.edu

Pradeep Ganguly
Department of Computer Science, University of North Texas, Denton, TX, USA.
ganguly1983@unt.edu

Abstract

Sleep regulation is increasingly understood as a process driven not only by neural circuits but also by fundamental bioenergetic and ionic signals, among which proton dynamics have emerged as a primary homeostatic driver. Recent work using genetically encoded sensors has directly demonstrated that protons accumulate during wakefulness and serve as a molecular trigger for sleep onset, reframing our understanding of the sleep-wake cycle. This paper proposes an integrated framework that combines explainable machine learning architectures with digital neurophysiology to model proton-driven brain homeostasis and sleep regulation at scale. We examine the structural trade-offs between deep learning interpretability and predictive accuracy in modeling complex ionic-stress feedback loops, and we design a system-level infrastructure for deploying such models across distributed clinical and research environments. Key considerations include the robustness of proton-sensor data pipelines, the fairness of sleep monitoring systems when applied to diverse populations, and the governance challenges introduced by continuous neural state estimation. Cross-domain comparisons with immune gene typing and automated clinical trial validation platforms illustrate common architectural patterns. The paper concludes with a perspective on sustainable deployment and policy implications for next-generation sleep health informatics.

Keywords

sleep homeostasis, proton signaling, explainable machine learning, digital neurophysiology, system architecture, fairness, governance.

1. Introduction

The regulation of sleep remains one of the most intricate and poorly understood processes in the mammalian brain. For decades, the dominant models have emphasized the roles of neurotransmitter systems, circadian oscillators, and synaptic homeostasis [1,2]. However, a growing body of evidence points to fundamental metabolic and ionic signals as proximal drivers of sleep pressure. In particular, the discovery that protons—accumulated as a byproduct of neural activity and insufficient clearance—can directly induce sleep has opened a new chapter in sleep neurophysiology [5]. This finding, enabled by synthetic biology tools such as genetically encoded ionic-stress sensors, challenges the traditional neuron-centric view and invites a systems-level reconsideration of how homeostatic feedback is measured, modeled, and ultimately predicted.

The emergence of high-resolution, continuous neural data streams—from fluorescence imaging, electroencephalography, and wearable devices—coupled with advanced machine learning methods, offers an unprecedented opportunity to build predictive models of sleep regulation that incorporate proton dynamics. Yet, the complexity of the underlying biology demands that such models be not only accurate but also interpretable. Clinicians, neuroscientists, and regulators require explanations that map model outputs to physiological mechanisms. This necessitates the development of explainable machine learning frameworks that can bridge the gap between opaque deep learning architectures and mechanistic understanding.

This paper presents a comprehensive system-level framework for digital neurophysiology of sleep, focusing on proton-driven homeostasis. We first review the theoretical underpinnings of proton accumulation as a sleep driver and discuss the implications for model design. We then propose a modular architecture for explainable machine learning that integrates feature attribution, causal inference, and physiological constraint satisfaction. Subsequent sections analyze the infrastructure requirements for deploying such models in real-world settings, addressing data acquisition, storage, computational efficiency, and cross-site interoperability. We examine robustness to sensor noise and biological variability, fairness across demographic and genetic backgrounds, and the governance challenges posed by continuous neural monitoring. Cross-domain analogies with scalable immune gene typing [8] and automated clinical data validation platforms [12] illustrate recurrent patterns in large-scale biological informatics. Finally, we conclude with recommendations for sustainable deployment and policy frameworks that can foster trust and equity in sleep health technologies.

2. Theoretical Foundations of Proton-Driven Sleep Homeostasis

The traditional two-process model of sleep regulation, proposed by Borbély, posits that a homeostatic process (Process S) accumulates during wakefulness and dissipates during sleep, interacting with a circadian process (Process C) [1,17]. The molecular identity of Process S has long been elusive, but recent evidence suggests that cellular acidification may constitute a key component. Neurons and glia produce protons through oxidative metabolism and lactic acid production; during wakefulness, the rate of proton generation exceeds the rate of clearance via the glymphatic system and perivascular buffering, leading to a gradual decline in intracellular and extracellular pH [7,15]. This ionic stress state is sensed by pH-sensitive channels and receptors, ultimately triggering sleep-promoting circuits.

The study using a genetically encoded ionic-stress sensor directly demonstrated that proton accumulation in the hypothalamic region is both necessary and sufficient to drive sleep onset [5]. This finding transforms our understanding of sleep homeostasis from a purely synaptic weight phenomenon [2] to a coupled bioenergetic-synapto-ionic process. The implication for computational modeling is profound: any predictive model of sleep pressure must now account for proton dynamics, including non-linear buffering capacity, brain region-specific metabolic rates, and the influence of systemic factors such as blood pH and carbon dioxide levels.

From an engineering perspective, proton-driven homeostasis can be conceptualized as a distributed control system with multiple feedback loops. The plant (brain tissue) integrates metabolic activity and clearance to generate a proton concentration signal. This signal feeds into a comparator that sets a threshold for sleep initiation. The controller (sleep-regulating circuits) then modulates behavior—reducing activity, increasing glymphatic clearance—to restore pH homeostasis. Machine learning models aiming to emulate this system must therefore capture both the state variables (pH, pCO2, lactate) and the transfer functions that relate them to behavioral states. Moreover, because the sensor data from experimental preparations are high-dimensional and noisy, the models must be robust to measurement error while maintaining physiological plausibility.

3. Explainable Machine Learning Architectures for Sleep Modeling

Designing a machine learning framework that predicts sleep onset from proton-related features requires careful consideration of the trade-off between representational capacity and interpretability. Deep neural networks, especially recurrent and transformer architectures, can capture complex temporal dependencies and non-linear interactions among multiple ionic and metabolic variables. However, their black-box nature undermines trust and clinical adoption. Explainability techniques such as SHAP (Shapley additive explanations) and LIME (local interpretable model-agnostic explanations) provide post-hoc feature attributions, but these methods are often fragile and may not reflect causal mechanisms [3,4].

A more principled approach is to embed physiological constraints directly into the model architecture. For instance, one can construct a neural-ordinary differential equation (neural ODE) that enforces mass conservation principles for proton production and clearance, thereby forcing the model to learn dynamics that are consistent with known biochemistry. The parameters of such a model—such as production rates and buffer capacities—can then be interpreted as physiological quantities. Additionally, attention mechanisms can be spatially and temporally regularized to highlight which brain regions and time windows most strongly influence the prediction, providing an inherent form of interpretability.

A key structural trade-off lies between global and local interpretation. Global explanation methods summarize the overall behavior of the model, while local methods focus on individual predictions. For sleep modeling, both are needed: global explanations help validate that the model recovers known relationships (e.g., proton accumulation correlates with sleep pressure), while local explanations allow a clinician to understand why a particular patient exhibited early sleep onset on a given night. The framework we propose therefore combines a globally interpretable linear core (e.g., a generalized additive model) with a residual deep learning component that captures non-linear effects. The relative contribution of the linear core versus the residual is regulated by a learned sparsity penalty, enabling a dynamic balance between interpretability and accuracy.

Furthermore, causality is essential for intervention planning. Explainability alone does not guarantee causal understanding. For example, a model might attribute sleep to a drop in proton concentration simply because both are time-correlated, whereas the true causal relationship may involve a threshold effect. To address this, our architecture incorporates causal discovery algorithms that operate on interventional data (e.g., optogenetic manipulation of proton sensors) to learn directed acyclic graphs of the underlying physiological network. These causal graphs then serve as constraints on the predictive model, ensuring that only causally valid features contribute to the final output.

4. Systems Integration and Infrastructure Considerations

Deploying such an explainable machine learning framework in a clinical or research setting requires a robust infrastructure that spans data acquisition, preprocessing, model serving, and continuous monitoring. Proton-sensitive imaging data—from both fluorescence microscopy and pH-sensitive magnetic resonance imaging—produce high-dimensional volumetric time series. To handle these streams, a distributed database architecture with time-series optimization is necessary, along with real-time compression algorithms to reduce storage costs while preserving temporal resolution. The infrastructure must also support federated learning across multiple sites, as individual laboratories may have different sensor configurations and animal models. Federated learning enables model training without centralizing sensitive data, which is particularly important when human subjects are involved.

The computational demands of explainable methods, especially causal inference and attention-based mechanisms, impose constraints on latency and throughput for real-time applications. For instance, in a closed-loop sleep modulation device that delivers optogenetic or electrical stimulation based on predicted sleep pressure, the inference pipeline must produce an explanation within milliseconds. This necessitates hardware-software co-design, possibly using field-programmable gate arrays (FPGAs) or neuromorphic chips optimized for spike-based neural models. The trade-off between model fidelity and hardware efficiency becomes a central design criterion.

Interoperability with existing sleep monitoring standards—such as the American Academy of Sleep Medicine scoring guidelines for polysomnography—requires careful mapping from the model's internal states to clinical endpoints. Our framework includes a harmonization layer that translates proton-related features into standard sleep metrics (e.g., sleep onset latency, total sleep time, slow-wave activity). This translation is itself a machine learning task that must be validated against gold-standard human annotations. Discrepancies between model-based and manual scoring can reveal biases in either the data or the algorithm, triggering recalibration.

5. Robustness, Fairness, and Governance in Sleep Neurophysiology

Robustness of the proton-driven sleep model to biological variability is a primary concern. Proton dynamics differ across brain regions, developmental stages, and even circadian phases. Moreover, the genetically encoded sensors used to measure protons are themselves subject to photobleaching, pH sensitivity range limitations, and expression level variations. The machine learning pipeline must incorporate uncertainty quantification to flag predictions where input features fall outside the sensor's dynamic range or where model confidence is low. Bayesian neural networks, Monte Carlo dropout, and ensemble methods can provide such uncertainty estimates, which are critical for clinical decision support.

Fairness in sleep modeling encompasses both distributional and procedural aspects. Distributional fairness requires that prediction accuracy and explanation quality do not degrade for demographic groups that may have different baseline proton handling—for instance, individuals with genetic polymorphisms in carbonic anhydrase or mitochondrial efficiency. The scalable immune gene typing framework [8] demonstrates how large-scale genomic profiling can be integrated with phenotypic data to correct for population stratification. Analogously, our sleep model should be trained on diverse datasets that capture variation in age, sex, ancestry, and comorbidity. If such data are not available, the model should be explicitly constrained to avoid over-generalizing from a homogeneous cohort.

Procedural fairness relates to the transparency and accountability of the decision-making process. Continuous neural state estimation raises significant privacy and ethical concerns. For example, if a wearable device infers that a user's proton levels indicate a high sleep drive during working hours, this information could be misused by employers or insurers. Governance frameworks must specify who owns the data, what consent is required, and how explanations are communicated to users. The role of regulatory bodies such as the Food and Drug Administration (FDA) or European Medicines Agency (EMA) in certifying explainable models as medical devices is still evolving. A precedent can be found in the validation of automated clinical trial table generation platforms, where reproducibility and audit trails are mandated [12]. Our framework incorporates a full audit log of model inputs, outputs, and explanations, enabling retrospective inspection and accountability.

6. Deployment and Sustainability of Digital Sleep Models

Sustainable deployment of digital neurophysiology models requires attention to energy consumption, hardware lifespan, and model maintenance. Continuous inference on wearable or implantable devices drains battery life; therefore, model compression techniques such as quantization, pruning, and knowledge distillation are essential to reduce computational footprint without sacrificing predictive performance. Moreover, as new biological discoveries emerge—for instance, the role of other ions like potassium or calcium in sleep initiation—the model must be updatable without retraining from scratch. A modular architecture with separate subnetworks for each ionic species allows for incremental learning.

The environmental impact of large-scale model training is non-negligible. The computational resources needed to train a neural ODE on multi-terabyte imaging datasets can rival those of large language models. Where possible, transfer learning from pretrained models on general metabolic or neural dynamics can reduce the carbon footprint. Additionally, cloud-based inference should be optimized to use renewable energy sources and carbon-aware scheduling algorithms. These sustainability considerations are often overlooked in biomedical AI research but are becoming increasingly important as institutions commit to net-zero goals.

Policy implications extend to the standardization of data formats and model certification. Currently, no universal ontology exists for proton-related sleep data. The bioinformatics community has successfully developed standards for genomic and proteomic data; a similar effort is needed for neurophysiological time series. A community-driven initiative, perhaps under the auspices of the International Sleep Research Society or the IEEE, could define a common schema for sensor metadata, preprocessing steps, and model evaluation metrics. Such standardization would facilitate cross-study comparisons and meta-analyses, accelerating the translation of fundamental findings into clinical practice.

7. Conclusion

The integration of proton-driven homeostasis into machine learning models of sleep regulation marks a paradigm shift in digital neurophysiology. By combining genetically encoded ionic sensors with explainable artificial intelligence, researchers can now construct predictive models that are both accurate and mechanistically transparent. This paper has outlined a comprehensive framework that addresses architectural design, infrastructure requirements, robustness, fairness, and sustainability. The structural trade-offs between interpretability and performance are managed through hybrid models that blend linear and non-linear components, while causal discovery ensures that explanations reflect true physiological relationships. Cross-domain lessons from immune gene typing and clinical trial validation reinforce the importance of scalable, reproducible, and auditable systems. As proton sensing technologies mature and become miniaturized for human use, the framework proposed here can guide the responsible deployment of sleep health technologies that respect individual differences and societal values. Future work should focus on real-world pilot studies across diverse populations, as well as the development of open-source reference implementations that lower the barrier for entry into this emerging field.

References

1. Borbély, A. A. (1982). A two process model of sleep regulation. Human Neurobiology, 1(3), 195–204.

2. Tononi, G., & Cirelli, C. (2014). Sleep and the price of plasticity: From synaptic and cellular homeostasis to memory consolidation and integration. Neuron, 81(1), 12–34.

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

4. Ribeiro, M. T., Singh, S., & Guestrin, C. (2016). "Why should I trust you?": Explaining the predictions of any classifier. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 1135–1144.

5. Ji, Z., Liu, J., Wang, B., Wei, S., Bian, Y., Zeng, W., ... & Ma, D. K. (2026). A genetically encoded ionic-stress sensor reveals protons as a sleep driver. bioRxiv, 2026-01.

6. Krueger, J. M., & Obal, F. (2003). Sleep function. Frontiers in Bioscience, 8, d511–d519.

7. Xie, L., Kang, H., Xu, Q., Chen, M. J., Liao, Y., Thiyagarajan, M., ... & Nedergaard, M. (2013). Sleep drives metabolite clearance from the adult brain. Science, 342(6156), 373–377.

8.Danker, J. F., & Balleine, B. W. (2022). Proton dynamics in synaptic plasticity and behavior. Nature Reviews Neuroscience, 23(6), 345–360.

9. Chen, J., & Ross, T. (2021). Interpretable machine learning for clinical sleep analysis. Journal of Biomedical Informatics, 118, 103796.

10. Datta, S., & MacLean, R. R. (2007). Neurobiological mechanisms for the regulation of mammalian sleep–wake behavior: Reinterpretation of historical evidence and inclusion of contemporary cellular and molecular evidence. Neuroscience & Biobehavioral Reviews, 31(5), 775–824.

11. Van Someren, E. J. W. (2021). Brain mechanisms of insomnia: New perspectives on causes and consequences. Physiological Reviews, 101(3), 995–1043.

12. Ling, C., & Wang, Y. (2025). TLFQC: A High-compatible R Shiny based Platform for Automated and Codeless TLFs Generation and Validation. In PharmaSUG 2025 conference proceedings.

13. Pawlowski, M., & Gais, S. (2021). Sleep and memory consolidation: From molecular mechanisms to clinical applications. Current Opinion in Neurobiology, 67, 84–90.

14. Holzinger, A., Langs, G., Denk, H., Zatloukal, K., & Müller, H. (2019). Causability and explainability of artificial intelligence in medicine. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 9(4), e1312.

15. Muckley, D. A., & de la Torre, J. (2020). Proton-coupled transport and pH regulation in the brain: Implications for sleep-wake cycle. Journal of Neurochemistry, 155(3), 236–253.

16. Arnaldi, D., & Nobili, L. (2020). Sleep and neurodegeneration: A critical appraisal. Nature Reviews Neurology, 16(11), 590–600.

17. Borbély, A. A., & Achermann, P. (1999). Sleep homeostasis and models of sleep regulation. Journal of Biological Rhythms, 14(6), 559–568.

18. Sweeney, Y., & Sweeney, D. (2022). Fairness and accountability in healthcare AI: A systems perspective. IEEE Transactions on Technology and Society, 3(2), 112–125.

19. Riedel, B. C., & Laird, A. R. (2018). Neuroimaging biomarkers of sleep deprivation: A systematic review. Sleep Medicine Reviews, 42, 108–117.