Machine Learning–Driven Prediction of Glycolipid Metabolism Modulation by Plant-Derived Polysaccharides: A Case Study of Phyllostachys nigra

Neavin Kanna
Department of Computer Science and Engineering, University at Buffalo, Buffalo, NY, USA.
navinkhanna81@buffalo.edu

Feiqian Qin
Department of Computer Science, George Mason University, Fairfax, VA, USA.
feiqian301@gmu.edu

Abstract

The discovery that plant-derived polysaccharides can modulate glycolipid metabolism through multi-target interactions with gut microbiota and host signaling pathways has opened new therapeutic frontiers for metabolic disorders. However, the structural complexity of these macromolecules and the high dimensionality of metabolomic response spaces render conventional experimental screening prohibitively slow and expensive. This paper presents a systems-level analysis of machine learning–driven prediction frameworks designed to address this bottleneck, using polysaccharides from Phyllostachys nigra as a representative case. Rather than proposing a single model, we examine the architectural trade-offs, data infrastructure requirements, model governance, fairness considerations, and deployment challenges inherent in constructing a reliable prediction pipeline. We discuss how heterogeneous data sources including glycomic profiles, metagenomic sequencing, and clinical metabolic markers can be integrated within a federated, privacy-preserving infrastructure. The paper further interrogates the interpretability–performance tension, the necessity of cross-population fairness in metabolic predictions, and the regulatory pathways for clinical decision support tools. By situating the prediction of glycolipid modulation within a broader socio-technical ecosystem, we identify structural vulnerabilities such as data drift, algorithmic monoculture, and infrastructural lock-in. Policy recommendations are offered to guide the sustainable translation of such systems from laboratory research into equitable nutritional and therapeutic interventions. The analysis underscores that predictive accuracy alone is insufficient; robustness, transparency, and institutional preparedness form the pillars of responsible deployment.

Keywords

machine learning, glycolipid metabolism, Phyllostachys nigra, polysaccharides, systems architecture, model fairness, interpretability, data governance.

1. Introduction

The global escalation of metabolic disorders, particularly type 2 diabetes mellitus and dyslipidemia, has intensified the search for safe, multi-target modulators of glycolipid homeostasis. Plant-derived polysaccharides have emerged as promising candidates because they exhibit pleiotropic effects that include reshaping the gut microbiome, regulating intestinal hormone secretion, and directly interfering with hepatic glucose and lipid

metabolism. Despite this therapeutic promise, the discovery of bioactive polysaccharides is severely constrained by the vast combinatorial space of monosaccharide compositions, glycosidic linkages, molecular weights, and branching patterns. Conventional pipelines, which progress from extraction and purification to in vitro assays and animal models, are both resource-intensive and ill-suited for the systematic exploration of structure-function relationships. Machine learning offers a paradigm shift by learning complex mappings between polysaccharide structural descriptors and in vivo metabolic outcomes, thereby prioritizing candidates for experimental validation. This paper investigates the system-level requirements for such a predictive platform through the lens of a specific botanical source, Phyllostachys nigra, a bamboo species whose polysaccharide fraction has recently been characterized for its glycolipid-modulating capacity. The central argument is that a meaningful prediction system cannot be reduced to a model architecture but must be conceived as an integrated socio-technical infrastructure in which data provenance, algorithmic fairness, interpretability, and deployment governance are treated as first-class design constraints. Throughout this discussion, the focus remains on structural trade-offs and system resilience rather than on incremental quantitative gains. The analysis draws on scholarship from cheminformatics, machine learning systems engineering, and translational bioinformatics to frame a holistic view of what it means to build a trustworthy prediction pipeline for nutritional pharmacology.

The motivation for selecting Phyllostachys nigra as the case study is twofold. First, the species provides a concrete, experimentally grounded dataset that includes intestinal microbiome shifts and host serum metabolic profiles, enabling a realistic assessment of what a machine learning system must ingest. Second, bamboo polysaccharides represent an underutilized resource in functional food science, and species-level variability raises significant generalization and fairness concerns that are emblematic of broader challenges in natural product informatics. By explicitly addressing the architecture, data governance, and evaluation frameworks needed to move from raw spectroscopic and sequencing data to a validated prediction, the present work aims to bridge the gap between the computational biology community, which often focuses narrowly on model performance, and the systems engineering community, which emphasizes resilience, maintainability, and ethical alignment. The discussion is structured to first delineate the background and previous computational attempts, then to unpack the system architecture and data infrastructure, followed by modeling considerations, a dedicated evaluation lens that incorporates fairness and robustness, deployment governance, and finally a synthesis of lessons drawn from the Phyllostachys nigra case.

2. Background and Related Work

Glycolipid metabolism is regulated by a distributed network of organs and signaling cascades, with the gut microbiota acting as an essential metabolic interface that transforms otherwise indigestible polysaccharides into short-chain fatty acids and other signaling molecules [1,4]. The therapeutic potential of polysaccharides is rooted in their ability to simultaneously modulate multiple nodes of this network: enhancing glucagon-like peptide-1 secretion, inhibiting carbohydrate-hydrolyzing enzymes, and altering bile acid profiles [1,5]. However, translating a polysaccharide structure into a predicted metabolic phenotype is difficult because the biological readout is a convoluted function of host genetics, baseline microbiota composition, and dietary context. Previous computational efforts in natural product bioactivity prediction have largely relied on molecular fingerprints and quantitative structure-

activity relationship models developed for small molecules [2]. These techniques do not transfer seamlessly to polysaccharides, which are polymers with polydispersity and conformation-dependent epitopes. There have been isolated attempts to encode polysaccharide features using monosaccharide composition vectors, glycosidic linkage matrices, and molecular weight distributions, yet the field lacks a standardized, machine-readable representation that captures the hierarchical nature of glycan structures.

Systems biology frameworks have long advocated for data-driven models that integrate multi-omics layers to predict physiological outcomes [3]. In parallel, the machine learning community has produced highly scalable supervised learning algorithms, such as gradient-boosted trees, that handle heterogeneous tabular data effectively [6,7]. Ensemble methods and deep learning architectures have been applied to metabolomics and metagenomics data to predict metabolic syndrome markers, but the explicit coupling of polysaccharide chemical spaces with host metabolic trajectories remains underexplored. The broader artificial intelligence literature has simultaneously advanced critical discussions on interpretability and fairness that are especially pertinent to biomedical predictions. Explainable artificial intelligence methods now offer granular feature attribution, making it possible to identify which structural motifs drive a given prediction, while fairness research alerts designers to the risk that models trained on homogeneous cohorts will fail on underrepresented populations [8,9]. The intersection of these streams provides the intellectual foundation for our systems-level inquiry.

Existing computational pipelines in natural product discovery are often monolithic scripts that run on local workstations and are difficult to audit, reproduce, or update. They rarely incorporate the principles of continuous integration, versioned data, or model cards that have become standard in high-stakes machine learning systems [11]. The case of Phyllostachys nigra polysaccharide, whose broad metabolic effects have been experimentally documented, offers a concrete testbed to examine how these fragmented best practices can be unified into a coherent architecture.

3. System Architecture and Data Infrastructure

Designing a prediction system for glycolipid modulation demands an architecture that decouples data ingestion, feature engineering, model training, serving, and monitoring while maintaining strict provenance chains. We conceptualize the platform as a layered architecture comprising a data lake, a feature store, a model registry, an evaluation service, and a deployment gateway. The data lake must accommodate heterogeneous data modalities: nuclear magnetic resonance spectra and chromatography-derived monosaccharide profiles of polysaccharides, 16S ribosomal RNA or shotgun metagenomic sequencing data of gut microbiomes, and clinical tables containing fasting glucose, insulin, lipid panels, and inflammatory cytokines. Ingesting these streams requires robust schema enforcement and metadata annotation aligned with the FAIR principles, so that downstream consumers can assess dataset suitability without ambiguous manual curation [12].

A critical architectural choice is whether to centralize or federate the data storage layer. Centralization simplifies model training but introduces privacy risks, regulatory hurdles under frameworks such as the General Data Protection Regulation, and data sovereignty concerns when cohorts span multiple jurisdictions. A federated infrastructure, in which raw data remain at the institution of origin and only model gradients or encrypted summary statistics are exchanged, aligns better with the ethical imperative of protecting participant privacy while enabling large-scale learning [10]. This trade-off, however, introduces significant systems

complexity. Federated learning demands synchronization protocols, straggler mitigation, and differential privacy guarantees, all of which increase the latency of training cycles and require dedicated network middleware. The system architect must also decide whether to adopt a cloud-native stack with elastic compute or rely on on-premise high-performance computing clusters, a decision that carries implications for cost predictability, data locality, and vendor lock-in [10].

The feature store serves as the linchpin that transforms raw observations into reusable, time-stamped feature vectors. For polysaccharide prediction, features fall into several categories: molecular descriptors, such as molar ratios of rhamnose, galactose, glucose, and uronic acids; topological indices derived from inferred glycan graphs; in silico predictions of solubility and viscosity; and aggregated metagenomic characteristics, including alpha diversity indices, enterotype assignments, and pathway abundance profiles. The feature store must support point-in-time correctness to avoid data leakage, a common pitfall where future information inadvertently contaminates training sets. This requires immutable, append-only feature tables and automated pipeline lineage tracking. Operationalizing such a store demands a dedicated governance board that defines feature definitions, monitors drift, and deprecates obsolete features in a controlled manner, mirroring the data product management practices seen in large-scale industrial machine learning systems [11].

4. Machine Learning Modeling and Feature Engineering

Once the infrastructure is in place, the core predictive challenge becomes mapping the high-dimensional polysaccharide–microbiome feature space onto continuous outcomes such as the homeostatic model assessment of insulin resistance or serum triglyceride area-under-the-curve. Gradient-boosted decision tree ensembles have proven remarkably effective on tabular data with mixed variable types and missing values, and they provide native feature importance measures that facilitate first-pass interpretability [6]. Random forests offer an alternative that can capture higher-order interactions without the need for extensive hyperparameter tuning, though they may suffer from higher bias in extrapolation regimes [7]. While deep learning models, including graph convolutional networks that operate directly on polysaccharide molecular graphs, hold theoretical appeal for automatically learning hierarchical representations, their practical utility is constrained by the limited size of labeled polysaccharide datasets [15]. The structural trade-off is clear: representationally powerful models demand more data and yield less transparent decision boundaries, whereas simpler models risk underfitting the intricate biology but afford easier regulatory acceptance.

Feature engineering exerts an outsized influence on system performance. Encoding polysaccharide branching patterns as graph kernels or persistent homology barcodes captures three-dimensional information that simple composition vectors miss. Integrating aggregated microbiome co-abundance groups as features introduces another dimension but raises the specter of spurious associations if confounders such as diet and physical activity are not properly controlled. The system must therefore incorporate causal discovery modules that go beyond correlation, using instrumental variable techniques or directed acyclic graph constraints to isolate the effect of polysaccharide interventions from background covariates. The associated computational cost can be managed by scheduling such analyses as asynchronous batch jobs that inform, but do not block, the online prediction pipeline.

Interpretability is not a monolithic property but a spectrum of stakeholder needs. A bench biologist requires explanations at the molecular motif level to guide synthesis or extraction optimization, while a clinician needs patient-specific risk drivers, and a regulatory reviewer

demands formal guarantees of stability under input perturbations. Post-hoc explanation frameworks such as SHAP can decompose predictions into additive feature contributions, pinpointing, for example, that a high mannose-to-glucose ratio is primarily driving a predicted reduction in fasting insulin [14]. Building such explanations into the model registry as versioned artifacts creates an audit trail that connects predictions to evidence, a prerequisite for both scientific reproducibility and legal accountability [8].

5. Evaluation Framework, Robustness, and Fairness

Conventional evaluation metrics such as root-mean-square error or area under the receiver operating characteristic curve, while necessary, are insufficient for a system intended to guide nutritional or clinical decisions. The evaluation framework must include distributional robustness checks that assess performance across meaningful subpopulations defined by sex, age, ethnicity, dietary background, and baseline gut microbiome enterotype. A model that performs well on average but fails on individuals with low Bifidobacterium abundance, for instance, could exacerbate health disparities. Fairness-aware evaluation should employ metrics such as equalized odds and demographic parity, while acknowledging the limitations of statistical fairness criteria in the face of domain-specific causal structures [9]. The system must be instrumented to emit these subgroup metrics continuously so that degradation caused by data drift or concept drift can be detected before it affects decision-making.

Robustness extends to adversarial resilience and input uncertainty. Polysaccharide characterization measurements carry inherent error from analytical instruments, and microbiome profiles vary with sequencing depth and batch effects. The prediction system should therefore undergo stress testing through systematic perturbation of inputs within their uncertainty envelopes, documenting the variance of the output distributions. Techniques from robust optimization and conformal prediction can equip the model with calibrated prediction intervals, ensuring that end-users are not misled by overconfident point estimates. Regular retraining schedules, governed by a monitoring service that flags distributional shifts in input features using divergence measures like maximum mean discrepancy, close the feedback loop between deployment and model maintenance [11].

Validation of predictive systems for metabolic outcomes presents unique statistical challenges because the ground truth is often derived from relatively small, short-duration animal experiments followed by limited human pilot trials. External validity must be demonstrated across independent cohorts that were not used for feature engineering or hyperparameter selection. Reporting should follow standardized frameworks originally developed for clinical prediction models, such as the TRIPOD statement, adapted for the multi-omics setting, thereby enabling transparent comparison and meta-analysis [16]. The system’s design must resist the temptation to cherry-pick positive results by pre-registering analysis plans and version-locking the entire computational environment, ensuring that evaluations are reproducible by third parties.

6. Deployment and Operational Governance

Transitioning a validated model from a research environment into operational service introduces an entirely new class of system requirements around latency, availability, security, and regulatory compliance. A machine learning model intended to predict metabolic responses for personalized nutrition must respond to queries within milliseconds if embedded in a consumer-facing application, whereas batch predictions for population-level screening can tolerate longer latencies. The deployment gateway, therefore, needs to support multiple

serving modes: a low-latency representational state transfer endpoint for interactive use and an asynchronous, high-throughput interface for research cohorts. Containerization and orchestration through platforms such as Kubernetes facilitate horizontal scaling and rolling updates, but they also introduce a dependency on the underlying cloud infrastructure, raising concerns about cost transparency and the risk of proprietary lock-in that could stifle academic and public health innovation [10].

Operational governance demands a clear delineation of roles and responsibilities. A cross-functional board comprising computational scientists, domain biologists, ethicists, and legal advisors should oversee the model lifecycle, from version approval to decommissioning. Every model that reaches production must be accompanied by a model card that documents its intended use, performance characteristics across subpopulations, known failure modes, and the composition of the training data. Such transparency is not merely a best practice but is increasingly becoming a regulatory expectation as agencies like the U.S. Food and Drug Administration develop frameworks for software as a medical device. The architectural implication is that the model registry must be tightly coupled with the evaluation service and the monitoring stack, so that any update to the model is automatically cross-referenced against the latest fairness and drift reports before being cleared for production [17].

Data privacy and security form a cross-cutting governance layer. Personal metabolic and microbiome data are sensitive by nature, and their leakage could lead to genetic discrimination or psychological harm. The system must enforce attribute-based access control, encrypt data at rest and in transit, and support the right to erasure without corrupting the feature store's lineage tracking. Federated and split learning architectures reduce the attack surface by minimizing raw data movement, yet they introduce new attack vectors such as gradient inversion and membership inference that must be modeled and mitigated through differential privacy mechanisms and secure multi-party computation. The governance framework should mandate periodic privacy impact assessments and pen-testing, treating security not as a static checklist item but as a dynamic capability that evolves with the threat landscape.

The broader policy implications cannot be overlooked. As machine learning predictions begin to influence dietary recommendations and supplement formulation, there is a risk that commercial interests could exploit algorithmic opacity to market unsubstantiated health claims. Regulatory bodies will need to develop pathways that validate the entire prediction pipeline rather than just the final output, inspecting data provenance, feature selection procedures, and model update cadence. International coordination is necessary to prevent a race to the bottom in which prediction systems are deployed in jurisdictions with lax oversight, accumulating data that later influences global markets. The infrastructure must therefore be architected to output tamper-evident audit logs that can be reviewed by independent inspectors, embedding accountability into the digital fabric of the system [18,19].

7. Case Study Discussion: Phyllostachys nigra Polysaccharide

The polysaccharide fraction derived from Phyllostachys nigra serves as a rich exemplar for the system described above because it has been systematically characterized in terms of its monosaccharide composition, molecular weight distribution, and in vivo effects on glycolipid metabolism and gut microbiome structure [13]. The experimental data provide labeled outcomes in mouse models that include reductions in serum triglycerides and low-density lipoprotein cholesterol, alongside increases in beneficial bacterial genera such as Akkermansia and Lactobacillus. Mapping this dataset onto the proposed infrastructure reveals

several instructive tensions. First, the monosaccharide profile of the polysaccharide—dominated by glucose, galactose, and mannose—is shared by many other botanical polysaccharides, meaning that a model relying solely on composition would exhibit high epistemic uncertainty unless augmented with higher-order structural features such as linkage patterns and branching degree. This observation justifies the investment in graph-based representations and the corresponding computational graph kernel machinery [15].

Second, the gut microbiome sequencing data in the reference study were generated from a single cohort under controlled laboratory conditions. Deploying a model trained on such data into human populations with vastly more diverse dietary habits and genetic backgrounds would almost certainly encounter distributional shift. The system's evaluation framework, with its emphasis on subgroup audits and fairness metrics, was explicitly designed to detect this degradation before clinical harm ensues. Had the model been built without these monitoring capabilities, a subsequent real-world test might have shown that the predicted reduction in postprandial glucose was accurate only for individuals with a baseline microbiota resembling that of the laboratory mouse colony, a failure mode that could erode trust in polysaccharide-based interventions at large.

Third, the translational pathway from animal data to human application demonstrates the criticality of the deployment governance structures we have outlined. A model developed on murine data must be clearly labeled as such, with prominent warnings against extrapolation to human physiology without explicit transfer learning or domain adaptation steps. The model registry must gate access so that clinicians or nutritionists querying the system are presented with predictions that are either calibrated for the target species or accompanied by quantified domain discrepancy estimates. The Phyllostachys nigra case thus crystallizes the often-underestimated distance between a scientifically interesting experimental finding and a safe, equitable, and reliable prediction service [20].

Finally, the case highlights the sustainability dimension of these systems. The bamboo plant itself is a rapidly renewable resource, positioning its polysaccharides as an environmentally favorable candidate for large-scale nutritional applications. Yet, the computational infrastructure required to run continuous model monitoring, federated retraining, and secure query serving has a non-trivial carbon footprint. Sustainable deployment would necessitate energy-efficient hardware, intelligent model compression, and policies that weigh the marginal health benefit against the environmental cost. This multi-objective optimization is emblematic of the trade-offs that mature prediction systems must navigate, and it underscores that the ultimate measure of success is not a single accuracy metric but a balanced scorecard that integrates clinical effectiveness, fairness, privacy, interpretability, and ecological responsibility.

8. Conclusion

This paper has examined the problem of predicting glycolipid metabolism modulation by plant-derived polysaccharides through a systems lens, using Phyllostachys nigra as a concrete case study. The analysis has demonstrated that building a trustworthy prediction platform extends well beyond the selection of a machine learning algorithm. It requires a carefully architected data infrastructure that honors the heterogeneity of polysaccharide chemistry and multi-omics readouts, a federated learning framework that reconciles the global need for diverse training data with the local imperative of privacy, and a robust evaluation ecosystem that continuously audits fairness and robustness across shifting populations. The incorporation of post-hoc interpretability layers and causal reasoning modules is not an optional luxury but

a structural necessity for regulatory acceptance and scientific credibility. Deployment governance, encompassing model cards, cross-functional oversight boards, and comprehensive security protocols, constitutes the organizational scaffolding that prevents technical systems from drifting into ethically hazardous territory.

The systemic perspective advanced here carries several practical implications. Research funders should require machine learning projects in the nutritional and metabolic sciences to include explicit infrastructure and governance plans rather than focusing exclusively on model performance. Open-source feature stores and model registries tailored to glycomics and metagenomics should be developed as community goods to avoid vendor lock-in and reduce duplicated effort. Regulatory science must evolve to evaluate prediction pipelines as dynamic, evolving entities rather than as static products, a shift that will necessitate new standards for auditability and continuous validation. The Phyllostachys nigra case illustrates that even a single well-characterized polysaccharide can expose the fault lines of ad hoc machine learning implementations, reinforcing the imperative for a disciplined, system-level approach.

Looking forward, the convergence of high-throughput glycoanalytics, large-scale human microbiome cohort studies, and scalable machine learning infrastructure brings the vision of personalized polysaccharide-based nutrition within reach. Realizing this vision responsibly, however, demands that the field resists the gravitational pull of techno-solutionism and instead embraces the slower, more deliberate work of building institutions, protocols, and oversight mechanisms that match the power of predictive algorithms. The architecture, governance, and fairness frameworks discussed here frame a research agenda that is as much about social choice as about technological design, and deserve sustained interdisciplinary attention in the coming decade.

References

1. Yu, Y., Shen, M., Song, Q., & Xie, J. (2018). Biological activities and pharmaceutical applications of polysaccharide from natural resources: A review. Carbohydrate Polymers, 183, 91–101.

2. Lo, Y. C., Rensi, S. E., Torng, W., & Altman, R. B. (2018). Machine learning in chemoinformatics and drug discovery. Drug Discovery Today, 23(8), 1538–1546.

3. Kitano, H. (2002). Systems biology: a brief overview. Science, 295(5560), 1662–1664.

4. Samuel, V. T., & Shulman, G. I. (2016). The pathogenesis of insulin resistance: integrating signaling pathways and substrate flux. Journal of Clinical Investigation, 126(1), 12–22.

5. Tremaroli, V., & Bäckhed, F. (2012). Functional interactions between the gut microbiota and host metabolism. Nature, 489(7415), 242–249.

6. Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 785–794.

7. Breiman, L. (2001). Random forests. Machine Learning, 45(1), 5–32.

8. Arrieta, A. B., Díaz-Rodríguez, N., Del Ser, J., Bennetot, A., Tabik, S., Barbado, A., … & Herrera, F. (2020). Explainable artificial intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. Information Fusion, 58, 82–115.

9. 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.

10. Armbrust, M., Fox, A., Griffith, R., Joseph, A. D., Katz, R., Konwinski, A., … & Zaharia, M. (2010). A view of cloud computing. Communications of the ACM, 53(4), 50–58.

11. Sculley, D., Holt, G., Golovin, D., Davydov, E., Phillips, T., Ebner, D., … & Dennison, D. (2015). Hidden technical debt in machine learning systems. Advances in Neural Information Processing Systems, 28, 2503–2511.

12. Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J., Appleton, G., Axton, M., Baak, A., … & Mons, B. (2016). The FAIR Guiding Principles for scientific data management and stewardship. Scientific Data, 3, 160018.

13. Zhao, K., Wu, X., Han, G., Sun, L., Zheng, C., Hou, H., ... & Shi, Z. (2024). Phyllostachys nigra (Lodd. ex Lindl.) derived polysaccharide with enhanced glycolipid metabolism regulation and mice gut microbiome. International journal of biological macromolecules, 257, 128588.

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

15. Kipf, T. N., & Welling, M. (2017). Semi-supervised classification with graph convolutional networks. International Conference on Learning Representations.

16. Collins, G. S., Reitsma, J. B., Altman, D. G., & Moons, K. G. (2015). Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD statement. BMJ, 350, g7594.

17. Rajkomar, A., Dean, J., & Kohane, I. (2019). Machine learning in medicine. New England Journal of Medicine, 380(14), 1347–1358.

18. Floridi, L., Cowls, J., Beltrametti, M., Chatila, R., Chazerand, P., Dignum, V., … & Schafer, B. (2018). AI4People—an ethical framework for a good AI society: opportunities, risks, principles, and recommendations. Minds and Machines, 28(4), 689–707.

19. He, J., Baxter, S. L., Xu, J., Xu, J., Zhou, X., & Zhang, K. (2019). The practical implementation of artificial intelligence technologies in medicine. Nature Medicine, 25(1), 30–36.

20. Zitnik, M., Agrawal, M., & Leskovec, J. (2018). Modeling polypharmacy side effects with graph convolutional networks. Bioinformatics, 34(13), i457–i466.