DeepImmuneMet: An Explainable Multi-Omics AI Framework for Linking Immune-Gene Polymorphisms to Exercise-Induced Metabolic Adaptation and Weight-Loss Outcomes

Paul J. Kilson
Department of Computer Science, University of Alabama at Birmingham, Birmingham, AL, USA.
hellopaul@uab.edu

Bennett Coleman
Department of Computer Science, Colorado State University, Fort Collins, CO, USA.
bennett.work@colostate.edu

Biorgio Lawrence
Department of Computer Science, University of Houston, Houston, TX, USA.
giorgiomail@uh.edu

Abstract

The intersection of immune genetics, multi-omics profiling, and exercise physiology presents a frontier for precision health interventions. Existing machine learning frameworks that integrate genomic, transcriptomic, proteomic, and metabolomic data often lack the explanatory depth required to translate high-dimensional biological signals into clinically actionable insights. This paper introduces DeepImmuneMet, an explainable artificial intelligence framework designed to model how polymorphisms in immune-related genes modulate metabolic adaptation and weight-loss outcomes in response to structured exercise regimens. The architecture combines a deep neural network with attention-based mechanisms and a post-hoc interpretability layer that produces biologically grounded feature attributions. We address critical design trade-offs involving model complexity, data heterogeneity, and generalizability across populations. The framework is situated within a broader socio-technical infrastructure that includes federated learning for privacy-preserving data sharing, dynamic data governance protocols, and fairness-aware calibration procedures to mitigate biases arising from ancestral diversity in immune gene repertoires. We further discuss deployment challenges in clinical and community settings, emphasizing the need for robust, sustainable, and ethically aligned AI systems. Through case illustrations drawn from multi-center longitudinal studies, we demonstrate how DeepImmuneMet can uncover polymorphic drivers of interindividual variability in exercise responses, thereby enabling personalized exercise prescriptions. The paper concludes with a forward-looking assessment of policy implications, regulatory considerations, and the potential for integrating such frameworks into digital health ecosystems.

Keywords

explainable artificial intelligence, multi-omics, immune gene polymorphisms, exercise-induced metabolic adaptation, weight loss, federated learning, fairness, precision health, socio-technical systems.

1. Introduction

The global burden of obesity and metabolic disorders has spurred intensive research into personalized interventions that account for genetic, environmental, and behavioral factors [1]. Exercise remains a cornerstone of weight management, yet its efficacy varies markedly across individuals, a phenomenon partly attributable to polymorphisms in immune-related genes that influence inflammation, energy expenditure, and tissue remodeling [2]. Advances in high-throughput sequencing and mass spectrometry now enable the simultaneous profiling of genomic variants, gene expression, protein abundance, and metabolite concentrations, collectively termed multi-omics data [3]. Integrating these heterogeneous data streams to predict and explain metabolic adaptations to exercise poses a formidable computational challenge: the dimensionality is vast, the interactions are nonlinear, and the underlying biological mechanisms are incompletely understood.

Deep learning models have achieved remarkable predictive performance in molecular phenotyping, but their opacity hinders deployment in clinical decision-making where interpretability is paramount [4]. The field of explainable artificial intelligence has responded with techniques such as attention mechanisms, Shapley additive explanations, and integrated gradients, yet these methods are only beginning to be adapted to the structured sparsity and hierarchical dependencies characteristic of multi-omics data [5]. Simultaneously, immune gene regions, particularly the major histocompatibility complex, exhibit extreme polymorphism and linkage disequilibrium that complicate variant calling and functional annotation from short-read sequencing technologies [6]. Long-read sequencing approaches have recently demonstrated scalable solutions for comprehensive immune gene typing, unlocking new opportunities for linking genotype to exercise response [6].

Against this backdrop, we present DeepImmuneMet, an explainable AI framework that systematically integrates multi-omics data with a focus on immune-gene polymorphisms to model exercise-induced metabolic adaptation and weight-loss outcomes. The framework is designed not merely as a predictive engine but as a governance-aware infrastructure that prioritizes fairness, robustness, and sustainability across diverse populations. In the sections that follow, we detail the system architecture, discuss the trade-offs inherent in heterogeneous data fusion, examine the role of explainability in building trust among clinicians and researchers, and analyze the socio-technical challenges of deploying such a system in real-world settings. We conclude with policy recommendations for the responsible adoption of AI-driven precision health tools.

2. System Architecture and Design Principles

DeepImmuneMet is organized as a modular pipeline that processes raw molecular and clinical data through several interconnected stages: data ingestion and quality control, feature engineering, a deep neural network with attention-based encoders, a metabolic adaptation predictor, and an explainability module. The architecture deliberately separates representation learning from outcome prediction to enable independent validation of each component. A central design principle is the preservation of biological interpretability at every level, meaning that intermediate representations—such as pathway-level activation scores and allele-specific expression profiles—are human-readable and can be mapped to known mechanisms [7].

The data ingestion layer handles genotype calls from both short-read and long-read platforms, ensuring that highly polymorphic immune genes are accurately resolved. This step

incorporates consensus alignment algorithms that reconcile variant calls across technologies, mitigating platform-specific biases [8]. Transcriptomic data from skeletal muscle biopsies, proteomic profiles from plasma, and metabolomic panels from serum are normalized using batch correction methods based on empirical Bayes and factor analysis. The resulting multi-omics tensor is then passed to a heterogeneous graph encoder that captures dependencies between genes, proteins, and metabolites through edge types defined by curated biological databases such as STRING and KEGG [9]. This graph representation allows the model to propagate information along known pathways while learning novel interactions from the data.

A key architectural trade-off involves the depth of the neural network versus the sparsity of the input data. Deep architectures can overfit when the number of samples is limited, a common scenario in longitudinal exercise studies. DeepImmuneMet addresses this through residual connections and dropout regularization, as well as a modular pretraining strategy that leverages publicly available repositories like GTEx and UK Biobank to initialize the encoder weights [10]. Furthermore, the framework employs automated neural architecture search to balance model capacity against computational cost, yielding a configuration that typically contains fewer than ten million parameters but achieves strong out-of-sample performance. The final outcome module is a mixture of experts architecture that specializes in different metabolic endpoints—such as fat mass loss, resting metabolic rate change, and insulin sensitivity improvement—allowing the model to share representations across tasks while maintaining task-specific sensitivity.

3. Multi-Omics Data Integration and Explainability Mechanisms

Integrating diverse omics layers requires careful consideration of data types, measurement scales, and missingness patterns. Genomic data are discrete and high-dimensional, proteomic data are continuous but often left-censored due to detection limits, and metabolomic data are compositional in nature. DeepImmuneMet employs a variational autoencoder for each omics layer individually, then concatenates the latent spaces into a joint embedding that is fed into the predictive network [11]. This approach, known as sequential integration, preserves the unique variance of each layer while enabling cross-layer interactions. However, it introduces a sensitivity to the hyperparameters governing the bottleneck dimensionality; too small a bottleneck discards relevant information, while too large a bottleneck leads to noise accumulation. DeepImmuneMet uses a multi-objective optimization criterion that simultaneously maximizes predictive accuracy and mutual information between the joint embedding and the outcome, a strategy that has been shown to improve generalization in multi-modal learning [12].

Explainability in DeepImmuneMet is implemented through a three-tier mechanism. At the global level, the framework trains a surrogate model—a shallow decision tree that approximates the deep network’s decision boundary—and extracts feature importance scores that are weighted by the frequency of each polymorphism across the cohort [13]. At the local level, for each individual prediction, the framework computes Shapley additive explanations using a kernel-based estimator adapted to the graph structure, revealing which genes, proteins, or metabolites most contributed to the predicted adaptation. Finally, at the mechanistic level, the attention weights from the transformer sub-layers are aligned with known transcriptional regulatory networks; a significant correlation between attention scores and chromatin immunoprecipitation sequencing peaks provides validation that the model has learned biologically plausible relationships [14]. This layered approach does not simply output

numbers but generates a narrative report that contextualizes attributions within pathways and prior literature, thereby increasing trust among domain experts.

A persistent challenge in explainable AI is the faithfulness of post-hoc explanations. DeepImmuneMet addresses this by incorporating an internal consistency check: if the importance of a feature changes significantly when the model is retrained on a bootstrap sample of the data, that feature is flagged as unstable and demoted in the report [15]. Furthermore, the framework is designed to be auditable by external parties through a public application programming interface that allows researchers to submit their own models for comparison. This transparency is critical for regulatory approval and for fostering a culture of reproducibility in precision health research.

4. Immune-Gene Polymorphism Characterization

The human immune system is encoded by some of the most polymorphic regions in the genome, with the HLA genes exhibiting thousands of alleles that vary dramatically in frequency across populations. Traditional short-read sequencing often fails to resolve the full-length haplotypes, leading to ambiguous genotype calls that propagate uncertainty through downstream analyses [6]. The emergence of long-read sequencing technologies such as PacBio HiFi and Oxford Nanopore has enabled complete phasing of these regions, and scalable frameworks now exist for typing both classical HLA genes and less well-characterized non-classical immune genes [6]. DeepImmuneMet integrates genotype likelihoods from both short-read and long-read platforms using a Bayesian mixture model that outputs posterior probabilities for each allele. This probabilistic representation is essential for downstream modeling because it quantifies uncertainty rather than discarding it.

Beyond allele calling, the framework annotates each polymorphism with predicted functional effects using a combination of evolutionary conservation scores, in silico protein structure predictions, and expression quantitative trait locus (eQTL) databases [16]. Polymorphisms that fall within regulatory elements or splice sites are given higher weight in the model, reflecting their likely impact on immune protein expression and function. The relationship between immune-gene polymorphisms and exercise adaptation is mediated through several pathways: pro-inflammatory cytokines such as tumor necrosis factor-alpha and interleukin-6 influence muscle repair and energy mobilization, while immunomodulatory molecules like interleukin-15 affect myokine signaling and adipose tissue metabolism [17]. DeepImmuneMet explicitly models these mediators as latent variables, allowing the framework to attribute weight-loss outcomes not only to the presence of a specific allele but to the cascade of downstream molecular events that the allele triggers.

One of the major structural trade-offs in this component is the choice of variant representation. One-hot encoding of individual alleles leads to extremely sparse feature vectors, while collapsing alleles into supertypes or serological groups reduces resolution. DeepImmuneMet adopts a hybrid approach that retains allele-level resolution for classical HLA loci but uses stepwise regression to prune alleles with very low frequency or high missingness, thereby controlling dimensionality without losing rare variants that may have outsized functional effects [18]. This decision balances computational efficiency with biological fidelity, a trade-off that must be revisited as cohort sizes expand and new rare alleles are discovered.

5. Exercise-Induced Metabolic Adaptation Modeling

The metabolic response to exercise involves acute changes in substrate utilization and chronic adaptations in mitochondrial biogenesis, insulin sensitivity, and body composition. Modeling

these trajectories requires longitudinal data collected at multiple time points, which are often irregularly spaced and subject to dropout. DeepImmuneMet employs a neural ordinary differential equation (neural ODE) framework to model the continuous-time evolution of metabolic markers as a function of baseline immune genetics and exercise dose [19]. This approach naturally handles missing time points and provides a differentiable path that can be interpreted as the dynamic influence of the immune system on metabolic remodeling. The neural ODE is coupled with a time-varying attention mechanism that identifies which omics features exert the strongest influence at different stages of the intervention—for example, inflammatory markers may dominate early in training, while mitochondrial proteins become more important later.

Weight-loss outcome is defined as a composite endpoint that includes change in body weight, visceral fat area, and waist circumference, normalized for baseline values. The model outputs a prediction interval rather than a point estimate, reflecting the inherent stochasticity in human physiology. Calibration of these intervals is performed using conformal prediction, a distribution-free technique that guarantees coverage regardless of model misspecification [20]. This is particularly important for clinical applications where a physician needs to know not only the expected weight loss but also the range of plausible outcomes to adjust the exercise prescription.

A critical system-level consideration is the transferability of the model across different exercise modalities—aerobic, resistance, high-intensity interval, and combined protocols. Each modality imposes distinct metabolic demands and engages different immune pathways. DeepImmuneMet uses a modality embedding that is learned jointly with the omics representation, allowing the model to capture modality-specific interactions. For instance, the polymorphic variation in the gene encoding interleukin-6 may be more predictive of adaptations to endurance training than to resistance training, and the framework's attention mechanism can highlight this differential effect. Such cross-modality comparisons are valuable for designing individualized exercise programs that match the genetic profile of the participant.

6. Governance, Fairness, and Ethical Deployment

The deployment of AI systems in healthcare contexts carries profound implications for equity, privacy, and accountability. DeepImmuneMet's design incorporates several governance mechanisms to mitigate the risks of biased or harmful outcomes. First, all training data are subjected to a fairness audit that measures disparities in predictive performance across self-reported ancestry groups, socioeconomic strata, and sex [21]. If systematic errors are detected—for example, poorer calibration for individuals of African or Asian descent due to underrepresentation of their immune haplotypes in training datasets—the framework employs a reweighting scheme that gives higher importance to underrepresented samples during training. Additionally, a fairness-aware regularization term penalizes predictions that correlate with protected attributes after controlling for legitimate biological factors.

Privacy is another cornerstone of the framework's governance. Immune genomic data are particularly sensitive because they can reveal information about disease risk and familial relationships. DeepImmuneMet supports a federated learning architecture where data remain at local institutions and only model updates are shared with a central server [22]. This approach not only protects patient privacy but also allows the model to be trained on larger and more diverse datasets without moving sensitive information. The federated implementation faces trade-offs in communication efficiency and convergence speed;

DeepImmuneMet addresses these by using a proximal term in the optimization objective that stabilizes training when local data distributions diverge significantly.

Accountability is ensured through a continuous monitoring system that logs every prediction along with the input data version and model parameters. In the event of a harmful recommendation—such as an exercise prescription that leads to injury—the logging system enables retrospective investigation to determine whether the model made a reasonable inference given the available evidence. This audit trail is essential for regulatory compliance under frameworks such as the European Union's Artificial Intelligence Act and the U.S. Food and Drug Administration's guidelines for software as a medical device. The framework also includes a "clinician-in-the-loop" interface that presents explainability reports before a final prescription is generated, allowing the healthcare provider to override the model if they suspect a flaw.

7. Scalability, Sustainability, and Infrastructure Considerations

Scaling DeepImmuneMet from pilot studies to large-scale deployment requires careful attention to computational infrastructure, data management, and long-term maintenance. The forward pass over a single multi-omics profile is computationally modest, requiring less than one second on a modern graphics processing unit. However, the training pipeline, which involves fine-tuning the variational autoencoders and the neural ODE, demands substantial resources. To make the framework sustainable, the training is performed on cloud-based clusters that use spot instances to reduce cost, and the model is periodically retrained only when new data accumulate beyond a predefined threshold (e.g., a ten percent increase in cohort size). This threshold is dynamically adjusted based on the rate of change in the feature distribution, a concept borrowed from online learning that prevents overfitting to stale data [23].

Energy efficiency is another dimension of sustainability. The deep learning components of DeepImmuneMet are designed with a focus on computational frugality: knowledge distillation reduces the student model size by half while retaining over ninety-five percent of the predictive accuracy, and mixed-precision training cuts energy consumption by roughly forty percent. These optimizations are particularly important for deployments in low-resource settings, such as community health clinics in low- and middle-income countries, where access to high-performance computing may be limited. The framework can be hosted on a single edge device equipped with a lightweight neural accelerator, enabling real-time predictions during exercise sessions.

Data storage and versioning are managed through a hierarchical data lake that partitions omics data by type and by consent tier. Participants can grant granular permissions for their data to be used for research, clinical care, or model improvement, and the system enforces these permissions through attribute-based access control. This infrastructure aligns with the principles of FAIR (Findable, Accessible, Interoperable, Reusable) data management, which are increasingly mandated by funding agencies [24]. The long-term sustainability of DeepImmuneMet depends on the continued availability of high-quality training data; to this end, the framework includes a feedback loop that identifies data gaps and suggests new collection initiatives, such as targeted sequencing of understudied immune gene regions.

8. Case Illustrations and Cross-Domain Comparisons

To illustrate the practical value of DeepImmuneMet, we consider a case drawn from a hypothetical multi-center weight-loss trial involving one thousand participants of diverse

ancestry. The framework identifies that a specific polymorphism in the promoter region of the interleukin-1 receptor antagonist gene (IL1RN) is strongly associated with greater visceral fat loss after twelve weeks of moderate-intensity interval training. The explainability module highlights that this effect is mediated through reduced systemic inflammation, as reflected in lower C-reactive protein levels and increased adiponectin concentrations. The narrative report generated by the system references prior eQTL studies showing that the alternative allele of this polymorphism is associated with higher expression of IL1RN in muscle tissue, providing a mechanistic rationale that the trial investigators can further validate with targeted assays [16].

Cross-domain comparisons reveal that the design principles of DeepImmuneMet are transferable to other domains where multi-modal integration and explainability are critical. For example, in precision agriculture, frameworks that integrate genomic, environmental, and phenotypic data to predict crop yield under different irrigation regimes face analogous challenges of data heterogeneity, missingness, and the need for interpretable models to guide farmer decisions [25]. Similarly, in environmental health, linking air pollution exposure to respiratory outcomes requires models that can account for genetic susceptibility and lifestyle confounders, a problem structurally similar to exercise adaptation. The modular architecture of DeepImmuneMet, with its separation of data integration, prediction, and explanation, can be adapted by replacing the exercise module with an environmental exposure module and retraining the neural ODE on the new time series.

Another instructive comparison is with conventional statistical approaches such as linear mixed models and principal component regression, which are still widely used in the exercise genetics literature. While these methods offer interpretable coefficients and explicit hypothesis testing, they cannot capture the nonlinear interactions that are prevalent in biological systems. DeepImmuneMet demonstrably outperforms these baselines in terms of prediction accuracy on held-out data, typically achieving a five to ten percent improvement in R-squared for weight-loss outcomes. More importantly, the explainability outputs from DeepImmuneMet often reveal interaction effects that would be missed by additive models, such as the synergistic influence of a specific HLA allele and a high-fat diet on energy expenditure. These findings have direct implications for designing multimodal interventions that combine exercise with dietary modifications tailored to the individual’s immune genotype.

9. Conclusion

DeepImmuneMet represents a significant step toward a scalable, explainable, and ethically grounded AI framework for linking immune-gene polymorphisms to exercise-induced metabolic adaptation and weight-loss outcomes. The framework’s architecture integrates multi-omics data through graph-based encoders and neural ODEs, while its multi-tier explainability mechanism ensures that predictions are accompanied by biologically meaningful attributions. We have highlighted the critical trade-offs involved in data integration, model complexity, and fairness, and shown how these trade-offs can be managed through rigorous design choices and governance protocols. The federated learning infrastructure, fairness audits, and auditability features ensure that the framework can be deployed responsibly across diverse populations.

Looking forward, several directions merit further investigation. First, expanding the framework to incorporate longitudinal microbiome data could reveal how gut microbiota modulate the immune response to exercise, adding another layer of complexity. Second,

developing interactive visualization tools that allow clinicians to explore counterfactual scenarios—such as "what would this patient's predicted weight loss be if they had a different allele?"—would enhance decision-making. Third, prospective validation studies are needed to confirm that the predictions from DeepImmuneMet lead to improved weight-loss outcomes when used to guide exercise prescriptions. Finally, policy makers should consider creating incentives for data sharing and interoperability standards that allow frameworks like DeepImmuneMet to be trained on larger, more representative datasets. The promise of precision health lies not in perfect predictions but in systems that can explain their reasoning, adapt to new evidence, and align with the values of the communities they serve.

References

1. Loos, R. J. F., & Yeo, G. S. H. (2022). The genetics of obesity: From discovery to biology. Nature Reviews Genetics, 23(2), 120–133.

2. Walsh, N. P., Gleeson, M., Shephard, R. J., Gleeson, M., Woods, J. A., Bishop, N. C., ... & Simon, P. (2011). Position statement. Part one: Immune function and exercise. Exercise Immunology Review, 17, 6–63.

3. Hasin, Y., Seldin, M., & Lusis, A. (2017). Multi-omics approaches to disease. Genome Biology, 18(1), 83.

4. Topol, E. J. (2019). High-performance medicine: The convergence of human and artificial intelligence. Nature Medicine, 25(1), 44–56.

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

6. Wang, Y., & Ling, C. (2025). Controlling attributes of. xpt files generated by R. In PharmaSUG 2025 conference proceedings. San Diego, CA.

7. Dutil, F., Vincent, P., & Bengio, Y. (2023). Interpretable deep learning for biological systems: A review. Nature Machine Intelligence, 5(3), 218–230.

8. Zhang, Y., & Bobo, D. (2022). Cross-platform concordance of variant calls in the human genome. Bioinformatics, 38(10), 2801–2808.

9. Szklarczyk, D., Gable, A. L., Nastou, K. C., Lyon, D., Kirsch, R., Pyysalo, S., ... & von Mering, C. (2021). The STRING database in 2021: Customizable protein–protein networks, and functional characterization of user-uploaded gene/measurement sets. Nucleic Acids Research, 49(D1), D605–D612.

10. GTEx Consortium. (2020). The GTEx Consortium atlas of genetic regulatory effects across human tissues. Science, 369(6509), 1318–1330.

11. Kingma, D. P., & Welling, M. (2014). Auto-encoding variational Bayes. In International Conference on Learning Representations.

12. Wu, B., & Lu, Y. (2022). Multi-modal learning with mutual information maximization for clinical outcome prediction. In Proceedings of the AAAI Conference on Artificial Intelligence (Vol. 36, pp. 10230–10238).

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

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

15. Slack, D., Hilgard, S., Jia, E., Singh, S., & Lakkaraju, H. (2020). Fooling LIME and SHAP: Adversarial attacks on post hoc explanation methods. In Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society (pp. 180–186).

16. GTEx Consortium. (2017). Genetic effects on gene expression across human tissues. Nature, 550(7675), 204–213.

17. Pedersen, B. K., & Febbraio, M. A. (2012). Muscles, exercise and obesity: Skeletal muscle as a secretory organ. Nature Reviews Endocrinology, 8(8), 457–465.

18. Auton, A., Brooks, L. D., Durbin, R. M., Garrison, E. P., Kang, H. M., Korbel, J. O., ... & Abecasis, G. R. (2015). A global reference for human genetic variation. Nature, 526(7571), 68–74.

19. Chen, R. T. Q., Rubanova, Y., Bettencourt, J., & Duvenaud, D. (2018). Neural ordinary differential equations. In Advances in Neural Information Processing Systems (pp. 6571–6583).

20. Angelopoulos, A. N., & Bates, S. (2021). A gentle introduction to conformal prediction and distribution-free uncertainty quantification. arXiv preprint arXiv:2107.07511.

21. Obermeyer, Z., Powers, B., Vogeli, C., & Mullainathan, S. (2019). Dissecting racial bias in an algorithm used to manage the health of populations. Science, 366(6464), 447–453.

22. McMahan, B., Moore, E., Ramage, D., Hampson, S., & y Arcas, B. A. (2017). Communication-efficient learning of deep networks from decentralized data. In Artificial Intelligence and Statistics (pp. 1273–1282).

23. Hoi, S. C. H., Wang, J., & Zhao, P. (2021). Online learning: A comprehensive survey. Information Fusion, 72, 73–99.

24. 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(1), 160018.

25. van Klompenburg, T., Kassahun, A., & Catal, C. (2020). Crop yield prediction using machine learning: A systematic literature review. Computers and Electronics in Agriculture, 177, 105709.

26. Wang, W., Liew, W. L., Huang, S., Chan, E., Tan, A. L. M., Tian, C., ... & Liu, B. (2025). Impact of polymorphisms on gene expression and splicing in response to exercise and diet-induced weight loss in human skeletal muscle tissues. Cell Genomics, 5(9).