CoralMD: A Multi-Modal Dashboard for Personalized Medicine

Tanner DeGrazia

Pomona College

Data Science Capstone Project

DS 190: Senior Seminar

November 2025

Abstract

Modern healthcare systems often remain reactive, addressing disease only after symptoms appear. CoralMD is a data-driven prototype dashboard that integrates genomic, wearable, and electronic health record (EHR) data to enable interpretable, proactive healthcare insights. By combining machine learning and ethical data practices, the system aims to make precision medicine transparent, equitable, and actionable. This paper outlines the motivation, data architecture, and proposed design for CoralMD, along with a review of current literature and ethical challenges surrounding personalized healthcare technologies.

1 Introduction

Modern healthcare is a system that is focused on reacting to disease. Cancer, Heart Disease, Neurodegenerative disease, and Diabetes are four of the leading causes of death in the United States. None of them are treated in a manner where prevention is the goal. Our current healthcare system is largely reactive, intervening only after illness occurs, despite an extreme growth in accessible biological and lifestyle data in recent years. The goal of personalized medicine is to reverse that trend by using data to predict and prevent disease. So while most current systems fail to bridge the gap between genomic insights, real-time physiological tracking like apple watches, and clinical health records, I propose a multi-modal system to act as a home base for the outlets to interact. CoralMD is an interpretable, multi-modal dashboard designed to bring these domains together and make personalized, proactive care accessible for both clinicians and patients.

While genomics, wearables, and electronic health records (EHRs) have individually transformed medical data collection, integrating them into a cohesive, interpretable model remains difficult. Research highlights issues in datasets like bias and lack of explainability in clinical AI systems. Existing machine learning approaches often look to prioritize accuracy at the expense of transparency, creating black-box models that clinicians cannot trust entirely as they are being told what to do. Another problem that exists is the lack of a dataset that encompasses all modes of data, making a related model across each avenue very hard. Thus, generalization is difficult when population-level genomic data overrepresents certain ethnicities.

CoralMD will address these challenges through a machine learning framework that integrates genomic (GRCh38, 1000 Genomes, ClinVar, gnomAD), wearables (Apple Watch, OhioT1DM), and Electronic Health Record (MIMIC-IV) datasets. We can bring these streams together in a model that works together with each outlet, despite not having one dataset that is all encompassing CoralMD's modular architecture will allow clinicians to see how genetic predispositions interact with real-time information and medical history to shape health risk.

The project must overcome several technical hurdles: aligning multi-modal data of differing time scales and formats, preserving privacy, and ensuring fairness in model predictions, all while delivering an accurate product. To validate the system, CoralMD will benchmark predictive models using metrics such as accuracy, SHAP-based visualizations are popular among other personalized medicine literature, so the aim is to use this to expose model reasoning. A prototype dashboard built in Streamlit will demonstrate how interpretability can create clinician trust.

If successful, CoralMD will illustrate a scalable pathway toward **Medicine 3.0**, where prediction, prevention, and patient participation become standard practice. Scalability will be achieved hopefully in the future, by switching it to AWS in order to make it a fully usable platform. The project's integration of ethical design, explainable AI, and multi-modal data could advance how clinicians and patients interact with health information. Future work will explore scaling the system with larger datasets and integrating real clinician feedback to refine both usability and fairness.

2   Literature Survey

2.1   AI and Data-Driven Personalized Medicine

Studies such as [1] and [2] explore frameworks combining genomic, wearable, and EHR data for individualized treatment planning. These works report high predictive performance (e.g., accuracy $\approx 94\%$, AUC ≈ 0.98) but often sacrifice interpretability. CoralMD builds upon these efforts by prioritizing model transparency and clinician usability by not directing the clinician, but actually letting them interpret the model by their knowledge.

2.2   Explainability and Trust in Clinical AI

Research by [3] and [4] emphasize the importance of interpretable AI and user-centered design for clinical decision-making. Their findings support CoralMD's design choice to integrate SHAP visualizations and transparent model logic to promote trust among healthcare providers.

2.3   Applications in Clinical Decision Support and Visualization

[5], [6], and [7] illustrate how visualization-driven systems enhance clinical interpretation. Their focus on user interface design and model explainability directly informs CoralMD's dashboard layer, which will visualize genomic, metabolic, and clinical data in real time.

2.4 Ethical and Generalization Challenges

Across studies, recurring concerns include fairness, data security, and equity in clinical AI. CoralMD addresses these by adopting fairness metrics, population-level normalization, and a privacy-first data handling strategy informed by [8] and [9].

3 Proposal for New Contribution

3.1 Motivation, Aims, and Objectives

- Build a prototype that integrates multi-modal health data.

- Provide interpretable ML outputs rather than black-box scores.

- Support clinician and patient decision-making ethically.

- Demonstrate the feasibility of equitable personalized medicine tools.

3.2 Research Design and Methods

Data Sources: Genomic: GRCh38, 1000 Genomes, ClinVar, gnomAD.

Wearable: Apple Watch/Fitbit dataset, OhioT1DM CGM data.

EHR: MIMIC-IV de-identified hospital dataset.

Technology Stack: Python (Pandas, PyTorch, Scikit-learn), PostgreSQL, Streamlit dashboard, AWS for scalability.

Modeling and Evaluation: Multi-modal feature integration, SHAP interpretability, metrics including Accuracy, Precision, Recall, F1, and Fairness indices.

Deliverables: Interactive CoralMD prototype, data integration pipeline, and visualization module.

4 Ethical Considerations

4.1 The Ethics of Representation

When creating this project, a lot of thought went into ethics and consideration of the effect of this tool. In turn, when designing CoralMD's dashboard, ethical questions came from not only what data is used but from how results are represented. Mahony and Hulme's analysis of the IPCC's *"burning embers"* diagram [10] does a good job describing this project's intention of handling how visualizations can hold power or directly influence based on the wrong factors. In the paper, a color gradient intended to summarize scientific

consensus became a persuasive tool that shaped public and policy understanding of climate risk. Their study demonstrates that risk visualizations are never neutral, they embody aesthetic and moral choices that can direct interpretation. If you colored your risk score yellow at 75, but green at 76 the human brain affilitates the green result with a much healthier being, than someone who may only be one 'overall point' lower, but presented in a different color range. Thus, in the context of personalized medicine, this warns against dashboards that translate probabilistic risk into simple scores or red-flag icons that might implicitly tell clinicians what to do, whether this is intentional or not.

CoralMD, therefore, avoids "traffic-light" style cues or deterministic scores. Instead, it uses transparent model explanations that focus on integrating visualizations. This is done to help practitioners explore why certain variables influence predictions instead of just demanding they do so. Following Mahony and Hulme, the goal is not to produce unanimity through visualization but to invite informed interpretation. This allows the clinician to remain the reasoning agent, not the system's passive recipient. This ethical stance shows the value of interpretability as a form of humility, meaning the interface must reveal the model's logic while resisting the temptation to declare a single truth.

4.2 The Ethics of Comparison and Normalization

Dumit and de Laet's *"Curves to Bodies"* [11] traces how statistical methods such as growth charts, calorie tables, and biomedical curves do not merely describe human variation. Instead, they produce societal norms about what counts as healthy, normal, or deviant. The presentation of data can prevent users from framing the interpretation in reality, and instead forces them into tidy categories, which inevitably alters social assumptions and beliefs. When done incorrectly with health data, these practices can make certain presentations appear as deficiency, reinforcing gendered, racial, or socioeconomic stereotypes.

CoralMD confronts this problem directly. By integrating genomic, wearable, and EHR data, it risks reproducing existing inequities if normalization is applied in a non critical manner. The platform is then responsible of avoiding these inequities and will include fairness checks across demographic subgroups and directly mentions uncertainty ranges rather than "ideal" baselines. Once again putting the power to the interpreter, in an effort to mitigate bias in the tool. Instead of comparing a patient to an universal model applied to everyone, CoralMD wishes to emphasize trends and contextual explanations focused solely on the person at hand, which is what Dumit and de Laet might call plural curves rather than a single normative one. The ethical goal here is to make data comparison truthful, and give it the ability to acknowledge variability rather than shortening the scope and forcing it into misleading averages.

4.3 The Stakes of Precision

Both Mahony and Hulme's *burning embers* and Dumit and de Laet's calorie charts highlight how data can mislead through oversimplification. In medicine, this danger even more magnified. A visualization does more than influence decisions, it can shape treatments or mask existing bias. So, CoralMD treats every data comparison as a moral act, similar to Dumit's call for truthful comparison. Each chance, like a genomic variant vs. population frequency or a heart-rate trend vs. a baseline, must be justified and contextualized. The ethical challenge here is not just to predict accurately but also to communicate risk responsibly. Addressing this problem at all is a very important part of the process and is often overlooked in creating an unassuming model that is supposed to be helpful. Precision must always be discussed side by side with honesty about uncertainty as the stakes of misrepresentation are directly human in personalized medicine.

4.4 Predictive Power to Ethical Power

Finally, Coral must recognize that predictive power is also social power. To visualize risk is to make a claim about whose health matters, whose data counts, and how the future should be interpreted. Following Mahony and Hulme's warning against consensus driven simplification and Dumit and de Laet's critique of norm forming graphs, CoralMD's ethical stance is grounded in interpretability, diversity, and transparency. Its goal is not to persuade but to enable understanding, turning data science from an authority that directs clinical practice into a collaborator that supports it. Ultimately, ethical data science is not only about producing correct predictions, it is about empowering human reasoning through clarity and context.

References

[1] Z. Lotfi, R. Haji Hosseini, and M. Aminipour, "Artificial intelligence–driven approaches for prediction, management, and complication risk in type 2 diabetes: A systematic review," *InfoScience Trends*, vol. 2, no. 6, pp. 1–17, 2025.

[2] J. Ballard, Z. Wang, W. Li, *et al.*, "Deep learning-based approaches for multi-omics data integration and analysis," *BioData Mining*, vol. 17, no. 38, 2024.

[3] S. Sadhu, D. Solanki, L. Brick, N. Nugent, and K. Mankodiya, "Designing a clinician-centered wearable data dashboard (careportal): Participatory design study," *JMIR Formative Research*, vol. 7, p. e46866, 2023.

[4] L. H. Goetz and N. J. Schork, "Personalized medicine: Motivation, challenges, and progress," *Fertility and Sterility*, vol. 109, no. 6, pp. 952–963, 2018.

[5] M. K. Im, C. Rouphael, J. McMichael, N. Welch, and S. Dasarathy, "Challenges in and opportunities for electronic health record-based data analysis and interpretation," *Gut and Liver*, vol. 18, no. 2, pp. 201–208, 2024.

[6] E. J. Patterson, A. D. Bounds, S. K. Wagner, *et al.*, "Oculomics: A crusade against the four horsemen of chronic disease," *Ophthalmology and Therapy*, vol. 13, pp. 1427–1451, 2024.

[7] J. Wang, J. Luo, M. Ye, X. Wang, Y. Zhong, A. Chang, G. Huang, Z. Yin, C. Xiao, J. Sun, and F. Ma, "Recent advances in predictive modeling with electronic health records," in *Proceedings of the 33rd International Joint Conference on Artificial Intelligence (IJCAI)*, pp. 8272–8280, 2024.

[8] C. M. Erickson, A. Wexler, and E. A. Largent, "Digital biomarkers for neurodegenerative disease," *JAMA Neurology*, vol. 82, no. 1, pp. 5–6, 2025.

[9] H. K. Brittain, R. Scott, and E. Thomas, "The rise of the genome and personalised medicine," *Clinical Medicine (London)*, vol. 17, no. 6, pp. 545–551, 2017.

[10] M. Mahony and M. Hulme, "The colour of risk: An exploration of the ipcc's "burning embers" diagram," *Spontaneous Generations: A Journal for the History and Philosophy of Science*, vol. 6, no. 1, pp. 75–89, 2012.

[11] J. Dumit and M. de Laet, "Curves to bodies," in *Routledge Handbook of Science, Technology, and Society*, pp. 71–82, Routledge, 2014.