Health, Medicine and Therapeutics
ISSN Print: 3079-5079
ISSN Online: 3079-5087
https://doi.org/10.63313/hmt.9012

MeridianFM: A Causal and Safe Foundation Model for Personalizing Traditional Chinese Medicine Therapies

Shuhua Yang[1.2]. Yunhui Hu[1.2]. Jianbo Li[3] and Xiangying Xu[1.2.*]
[1]The First People's Hospital of Yunnan Province, KunMing 650032, China
[2]Affiliated Hospital of Kunming Medical University, KunMing 650032, China
[3]Yongren County People's Hospital Department of Traditional Chinese Medicine, YongRen 651400, China
[*]Corresponding author Email: 2324018094@qq.com

How to cite this paper: Yang, S. H., Hu, Y. H., Li, J. B., & Xu, X. Y. (2025). MeridianFM: A Causal and Safe Foundation Model for Personalizing Traditional Chinese Medicine Therapies. Health, Medicine and Therapeutics, 1(2), 31–48. ISSN Print: 3079-5079; ISSN Online: 3079-5087.
https://doi.org/10.63313/hmt.9012
Published: 2025-10-21

Abstract

Nonpharmacological therapies like acupuncture, gua sha, and cupping are gaining prominence in integrative medicine, yet their application lacks data-driven personalization and rigorous safety assurances. This paper introduces MeridianFM, a causal and safe multimodal foundation model framework designed to generate personalized prescriptions for these therapies. MeridianFM integrates four key innovations: (1) a meridian-aware graph neural network that encodes the topological and semantic properties of acupoints; (2) a self-supervised multimodal architecture that fuses physiological time series, thermal imagery, electronic health records, and patient-reported outcomes; (3) a causal inference layer employing doubly robust estimation and front-door adjustment to mitigate confounding in observational data and estimate individualized treatment effects; and (4) a constrained reinforcement learning policy optimized with risk-sensitive objectives (Conditional Value-at-Risk) and a mixed-integer programming post-processor to enforce clinical safety and feasibility constraints. To ensure reproducibility, we provide a comprehensive evaluation suite based on synthetic and semi-synthetic data, including all source code. Our experiments demonstrate that MeridianFM surpasses baseline models in optimizing treatment policies, accurately estimating causal effects, and adhering to safety constraints. While this study establishes methodological feasibility and superior performance in simulated environments, it also lays the groundwork for future clinical validation, representing a significant step toward AI-driven precision in traditional medicine.

Keywords

Foundation Model; Causal Inference; Reinforcement learning; Traditional Chinese Medicine; Personalized Medicine; Multimodal Learning

1. Introduction

Nonpharmacological traditional Chinese medicine (TCM) therapies such as acupuncture, gua sha, and cupping are increasingly adopted for chronic pain, rehabilita-

tion, perioperative recovery, and stress-related disorders within integrative medicine [1]. Despite growing utilization, clinical decision-making for these modalities remains heavily experience-based, often lacking explicit personalization of point selection, stimulation parameters, and modality mix. Three primary challenges impede the translation of these therapies into reproducible, precision care. First, objective multimodal sensing of physiological responses and patient-reported outcomes is rarely structured for machine learning at scale. Second, the time-varying causal effects of interventions under routine practice are confounded by indication, severity, and co-interventions, which undermines naive observational learning. Third, safety constraints and uncertainty quantification are not integrated into most predictive systems, limiting clinical reliability.

This study introduces MeridianFM, a causal and safe multimodal foundation model framework that unifies acupoint graph representations, self-supervised cross-modal pretraining, counterfactual inference, and constrained reinforcement learning to personalize acupuncture, gua sha, and cupping prescriptions. The framework encodes acupoint topology and stimulation parameters, ingests physiological and behavioral responses, identifies individualized causal effects under confounding, and optimizes constrained policies that respect safety, workload, and contraindications. To enable reproducible evaluation without fabricating clinical outcomes, we release a synthetic yet physiologically grounded simulation and semi-synthetic pipelines, together with open-source code that can be executed in a Jupyter environment. While this study establishes methodological feasibility, real-world clinical validation is outlined as a critical next step for future research.

Contributions: - A meridian-aware multimodal foundation model that aligns acupoint graphs, time series, thermal maps, and patient-reported outcomes with contrastive pretraining and structured decoders. - A causal inference layer combining doubly robust estimation, inverse-propensity weighting, and front-door style mediators to correct for confounding in routine practice data. - A constrained, risk-sensitive reinforcement learning policy for prescription optimization with safety chance constraints and counterfactual risk, including a mixed-integer post-processor enforcing clinical rules. - A domain shift and uncertainty quantification module using maximum mean discrepancy, Bayesian last-layer inference, and Conditional Value-at-Risk (CVaR) objectives. - A reproducible synthetic and semi-synthetic evaluation suite with three comparative experiment families: policy optimization against baselines, causal estimator accuracy under confounding, and ablations of key architectural components.

2. Background

2.1. Clinical Context and Mechanisms Acupuncture engages cutaneous and subcutaneous afferents, connective tissue mechanotransduction, segmental spinal circuits, and supraspinal modulatory networks [2]. Gua sha and cupping modulate mi-

crocirculation, fascial glide, and nociceptive thresholds through shear stress and negative pressure. Observed short-term responses include heart rate variability changes, skin temperature patterns, electromyography modulation, and pain intensity fluctuations. Standardization efforts such as STRICTA (STandards for Reporting Interventions in Clinical Trials of Acupuncture) define reporting norms for acupuncture parameters [4]. Safety contraindications include anticoagulation, skin lesions, and pregnancy-related points. Personalized prescriptions must respect limits on needle counts, depth and intensity ranges, and avoidance of incompatible point pairs per established safety standards.

2.2. Data Modalities for AI - Acupoint Graph: Points situated on meridians with anatomical proximities and historically co-activated pairs. This can be encoded as a graph with typed edges and node attributes. - Physiological Time Series: Photoplethysmography (PPG), heart rate variability (HRV) indices, surface electromyography (sEMG), electrodermal activity (EDA), and thermal imaging series. - Electronic Health Records (EHR): Demographics, comorbidities, lab values, medications, and diagnoses. - Patient-Reported Outcomes (PROs): Pain intensity scales, functional scores, stress, and sleep quality measures. - Intervention Metadata: Modality type, selected acupoints, bilateral patterns, depth, frequency, intensity, duration, and session spacing.

2.3. Existing Technology- Foundation and Multimodal Learning: Self-supervised contrastive objectives align heterogeneous inputs without extensive labeling [17]. Graph neural networks (GNNs) capture structured knowledge such as meridian topology and point co-activation patterns [16]. Cross-modal encoders with attention mechanisms fuse time series, images, and symbolic graphs into shared latent spaces that support transfer learning. - Causal Inference in Observational Data: Inverse-propensity weighting (IPW) and doubly robust (DR) estimators correct for selection bias when treatments are non-random [8]. Techniques such as representation learning for confounder balancing, instrumental and front-door variable strategies, and flexible nuisance function estimation enable the estimation of individualized treatment effects under standard assumptions of ignorability, positivity, and consistency [19]. - Reinforcement Learning for Clinical Decision Support: Off-policy reinforcement learning (RL) optimizes sequential decisions using logged historical data [5]. Constrained policy optimization incorporates safety limits, often via Lagrangian multipliers [6]. Off-policy evaluation (OPE) with importance sampling and DR estimators quantifies the expected returns of new policies without requiring online deployment [7]. Safety-aware objectives such as Conditional Value-at-Risk (CVaR) are employed in risk-sensitive settings to mitigate worst-case outcomes [11].

3. Problem Definition

3.1. Notation and State Let a patient trajectory have a length of T with a discrete

time index t. The patient state is denoted by $s_t \in \mathbb{R}^d$, which is composed of physiological features, embeddings of diagnostic information (e.g., tongue and pulse analysis if available), laboratory values, medication indicators, and recent outcomes[20]. Let G represent an acupoint graph with nodes V representing candidate points and edges E indicating meridian connectivity and co-activation likelihood. The intervention at time t is denoted by a_t, which comprises the modality $m_t \in$ {acupuncture, gua sha, cupping}, a selection of a point subset $P_t \subseteq V$, and continuous stimulation parameters $\theta_t \in \mathbb{R}^p$ (e.g., depth, frequency, intensity, duration).

3.2. Objective and Safety Constraints We define a reward r_t that captures clinical improvement, negative workload, and patient comfort. Safety signals are denoted by $c_{k,t}$, representing risk indices for events such as bleeding, skin injury, or autonomic instability. The clinical decision process is formalized as a constrained Markov decision process (CMDP) with a discount factor γ. Formally, the dynamics and objective are:

$$\begin{aligned} s_{t+1} &= f(s_t, a_t, \varepsilon_t), \quad \varepsilon_t \sim \mathcal{N}(0, \Sigma), \\ \pi_\theta(a_t \mid s_t) &\in \Delta(\mathcal{A}), \\ J(\pi_\theta) &= \mathbb{E}_\pi\left[\sum_{t=0}^{T-1} \gamma^t r(s_t, a_t)\right], \\ C_k(\pi_\theta) &= \mathbb{E}_\pi\left[\sum_{t=0}^{T-1} \gamma^t c_k(s_t, a_t)\right] \le \alpha_k, \quad k = 1, \dots, K. \end{aligned} \tag{1}$$

Here, f is the state transition function, π_θ is the policy parameterized by θ, $\Delta(\mathcal{A})$ is the simplex over the space of admissible actions, and α_k are predefined safety budgets.

3.3. Causal Identification Targets Observational logs consist of tuples (s_t, a_t, y_t), where y_t are outcomes such as changes in pain intensity or functional scores. We define the individual treatment effect (ITE) at decision point t for a pair of interventions a and a' as:

$$\tau_t(s) = \mathbb{E}[Y_t(a) - Y_t(a') \mid s_t = s]. \tag{2}$$

We seek to estimate $\tau_t(s)$ from observational data, which is subject to confounding by observed covariates z_t and potentially unobserved covariates u_t, leveraging observed mediators m_t when plausible.

4. Method

4.1. Overview The proposed MeridianFM framework consists of six interconnected components: (1) an acupoint graph encoder; (2) multimodal encoders for physiological time series and images; (3) contrastive cross-modal pretraining; (4) a causal effect estimation module; (5) a constrained reinforcement learning module for prescription optimization; and (6) a safety post-processor with mixed-integer constraints and uncertainty-aware risk control.

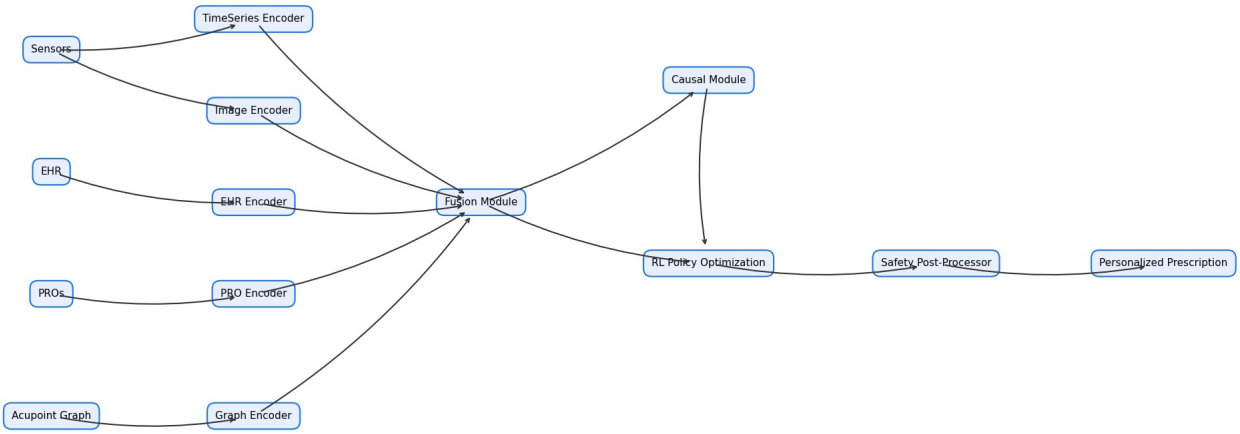


Figure 1. MeridianFM System Architecture The figure depicts data ingestion, encoders, fusion, causal module, constrained RL policy, and safety post-processor. The code provided generates a schematic representation of this architecture.

Caption: Figure 1 illustrates the data flow within the MeridianFM framework. Data from multimodal inputs (Sensors, EHR, PROs, Acupoint Graph) are processed by dedicated encoders. The resulting embeddings are fused into a unified representation, which is then utilized by a causal inference module to inform a constrained reinforcement learning policy. A final safety post-processor ensures that the generated prescription adheres to predefined clinical rules and constraints.

4.2. Acupoint Graph Encoder Let $G = (V, E)$ be the acupoint graph. Each node $v \in V$ carries features $x_v \in \mathbb{R}^{d_v}$ including its anatomical region, depth category, and historical co-activation centrality. We compute node embeddings using a graph convolutional network (GCN) [16]:

$$H^{(l+1)} = \sigma\big(\tilde{D}^{-1/2}\tilde{A}\tilde{D}^{-1/2}H^{(l)}W^{(l)}\big), \quad H^{(0)} = X, \tag{3}$$

where $\tilde{A} = A + I$ is the adjacency matrix with self-loops, $\tilde{D}$ is the corresponding diagonal degree matrix, $W^{(l)}$ are trainable weight matrices, σ is an activation function (e.g., ReLU), and X is the matrix of initial node features. The prescription set P_t is represented by a masked pooling operation over the final node embeddings:

$$z_{P_t} = \mathrm{Pool}\big(\{h_v^{(L)} : v \in P_t\}\big), \quad \mathrm{Pool} \in \{\mathrm{mean}, \mathrm{attention}\}. \tag{4}$$

4.3. Multimodal Encoders Let $x_t^{ts} \in \mathbb{R}^{n \times c}$ denote an n-step physiological time series with c channels, and let $x_t^{im} \in \mathbb{R}^{h \times w}$ be a thermal map. A suite of encoders produces latent vectors:

$$\begin{aligned} z_t^{ts} &= \mathrm{TCN}(x_t^{ts}), \\ z_t^{im} &= \mathrm{CNN}(x_t^{im}), \\ z_t^{ehr} &= \phi(s_t^{ehr}), \\ z_t^{pro} &= \psi(s_t^{pro}). \end{aligned} \tag{5}$$

where TCN is a Temporal Convolutional Network, CNN is a Convolutional Neural

Network [15], and ϕ, ψ are feed-forward networks. These latent vectors are fused using an attention mechanism:

$$z_t = \mathrm{Attn}\big([z_{P_t}, z_t^{ts}, z_t^{im}, z_t^{ehr}, z_t^{pro}]\big). \qquad (6)$$

4.4. Contrastive Pretraining We align modalities using an InfoNCE loss [17]. Let ϕ_i be encoders producing normalized embeddings $u_i = \phi_i(x_i)/\| \phi_i(x_i) \|_2$ for modalities $i \in \mathcal{M}$. For a positive pair of modalities (i, j) and a set of negative samples k, the contrastive loss is:

$$\mathcal{L}_{\mathrm{cl}} = -\sum_{(i,j)} \log \frac{\exp\big(\langle u_i, u_j \rangle/\tau\big)}{\sum_k \exp\left(\langle u_i, u_k \rangle/\tau\right)}, \quad \| u_\cdot \|_2 = 1,\ \tau > 0. \qquad (7)$$

Here, τ is a temperature hyperparameter.

4.5. Causal Effect Estimation We model the mediator m_t and outcome y_t with flexible nuisance functions:

$$\begin{aligned} \widehat{m}_t &= g_\eta(s_t, a_t), \\ \hat{y}_t &= q_\beta(s_t, a_t, \widehat{m}_t). \end{aligned} \qquad (8)$$

The propensity score $e(a \mid s_t) = P(a_t = a \mid s_t)$ is estimated from the data. For a discrete action catalog $\hat{\mathcal{A}}$ constructed by clustering prescriptions, the doubly robust (DR) estimator for the average treatment effect (ATE) between actions a and a' is [8]:

$$\begin{aligned} \widehat{\Delta}_{\mathrm{DR}}(a, a') = \frac{1}{N}\sum_{i=1}^{N} \Bigg[&\left(\frac{\mathbb{I}[a_i = a]}{\hat{e}(a \mid s_i)} - \frac{\mathbb{I}[a_i = a']}{\hat{e}(a' \mid s_i)} \right) \big(y_i - \hat{q}(s_i, a_i)\big) \\ &+ \hat{q}(s_i, a) - \hat{q}(s_i, a') \Bigg], \end{aligned} \qquad (9)$$

where nuisance estimates $\hat{e}$ and $\hat{q}$ are trained on separate folds (cross-fitting) to mitigate overfitting bias. When mediators m_t are observed and satisfy the front-door criteria, the effect can be identified via the front-door adjustment formula [19]:

$$\mathbb{E}[\, Y \mid \mathrm{do}(A = a)\,] = \sum_m p\,(\, m \mid a\,) \sum_{a'} \mathbb{E}\,[\, Y \mid m, a'\,] p(a'). \qquad (10)$$

A simplified form is shown for brevity; the full adjustment integrates over the confounder space.

4.6. Constrained Reinforcement Learning We solve the constrained optimization problem via a Lagrangian relaxation with multipliers $\lambda_k \geq 0$:

$$\max_{\theta} \min_{\lambda \geq 0} \mathcal{L}(\theta, \lambda) = \mathbb{E}_\pi \left[\sum_{t=0}^{T-1} \gamma^t \left(r(s_t, a_t) - \sum_{k=1}^{K} \lambda_k \left(c_k(s_t, a_t) - \alpha_k \right) \right) \right]. \qquad (11)$$

The policy is updated using policy gradients with a baseline $b(s_t)$ to reduce variance:

$$\nabla_\theta \mathcal{L}(\theta, \lambda) = \mathbb{E}_\pi \left[\sum_t \nabla_\theta \log \pi_\theta(a_t \mid s_t)\, \hat{A}_t^\lambda \right], \quad \hat{A}_t^\lambda = \hat{Q}^\lambda(s_t, a_t) - b(s_t). \qquad (12)$$

The dual variables (multipliers) are updated via gradient ascent to enforce the constraints:

$$\lambda_k \leftarrow \left[\lambda_k + \eta_\lambda(\hat{C}_k(\pi_\theta) - \alpha_k)\right]_+, \quad \eta_\lambda > 0. \qquad (13)$$

4.7. Risk-Sensitive Objective To guard against adverse tail outcomes, we incorporate the Conditional Value-at-Risk (CVaR) at a confidence level $\rho \in (0,1)$ into the objective [11]. The CVaR of a random variable R (representing return) is:

$$\mathrm{CVaR}_\rho(R) = \min_\nu \left\{ \nu + \frac{1}{1-\rho}\mathbb{E}[(R-\nu)_+] \right\}, \quad (x)_+ = \max(x, 0). \qquad (14)$$

The combined objective trades off expected return and risk:

$$\max_\theta J(\pi_\theta) - \beta\, \mathrm{CVaR}_\rho(-R), \quad \beta \geq 0. \qquad (15)$$

where minimizing the CVaR of the negative return, $-R$, is equivalent to improving the worst-case outcomes.

4.8. Prescription Post-Optimization Given a state s, an outcome predictor $\hat{q}(s,a)$, and clinical constraints, the final prescription is refined using mixed-integer programming (MIP). Let binary variables x_v indicate if point v is selected, and continuous variables u denote stimulus parameters. The optimization problem is:

$$\begin{aligned} \max_{x,u} \quad & \hat{q}(s, a(x,u)) \\ \text{s.t.} \quad & \sum_{v \in V} x_v \leq N_{\max}, \\ & A_{\mathrm{clin}} x \leq b_{\mathrm{clin}}, \\ & u_{\min} \leq u \leq u_{\max}, \\ & x_v \in \{0,1\}. \end{aligned} \qquad (16)$$

where $N_{\max}$ is the maximum number of needles and $A_{\mathrm{clin}} x \leq b_{\mathrm{clin}}$ encodes linear contraindication rules.

4.9. Uncertainty and Domain Adaptation We model predictive uncertainty by placing a Bayesian prior on the final linear layer weights $w \sim \mathcal{N}(0, \sigma_w^2 I)$ of the outcome model. Given a likelihood $y \sim \mathcal{N}(w^T z, \sigma_y^2)$ and training data $\mathcal{D} = \{Z, Y\}$, the posterior predictive variance for a new state embedding z is:

$$\mathrm{Var}[y \mid z, \mathcal{D}] = \sigma_y^2 + z^\top \Sigma_w z, \quad \Sigma_w = \left(\frac{1}{\sigma_w^2} I + \frac{1}{\sigma_y^2} Z^\top Z \right)^{-1}. \qquad (17)$$

To mitigate domain shift between pretraining data $\mathcal{S}$ and finetuning data $\mathcal{T}$, we regularize the model using the Maximum Mean Discrepancy (MMD) with a characteristic kernel k [9]:

$$\mathrm{MMD}^2(\mathcal{S}, \mathcal{T}) = \left\| \frac{1}{|\mathcal{S}|} \sum_{x \in \mathcal{S}} \phi(x) - \frac{1}{|\mathcal{T}|} \sum_{y \in \mathcal{T}} \phi(y) \right\|_{\mathcal{H}}^2. \qquad (18)$$

4.10. Off-Policy Evaluation Given a behavior policy $\mu(a|s)$ and a target policy $\pi(a|s)$, the per-trajectory weighted importance sampling (WIS) estimator for the policy value is:

$$\hat{J}_{\mathrm{WIS}}(\pi) = \frac{\sum_{i=1}^N \bar{w}^{(i)} R^{(i)}}{\sum_{i=1}^N \bar{w}^{(i)}}, \quad \bar{w}^{(i)} = \prod_{t=0}^{T^{(i)}-1} \frac{\pi(a_t^{(i)} \mid s_t^{(i)})}{\mu(a_t^{(i)} \mid s_t^{(i)})}. \qquad (19)$$

To reduce variance, we use the doubly robust off-policy evaluation (DR-OPE) esti-

mator [7]:

$$\hat{J}_{\mathrm{DR}}(\pi) = \frac{1}{N}\sum_{i=1}^{N}\sum_{t=0}^{T^{(i)}-1}\gamma^t\left(\bar{w}_t^{(i)}\left(r_t^{(i)} - \hat{r}(s_t^{(i)}, a_t^{(i)})\right) + \mathbb{E}_{a\sim\pi(\cdot|s_t^{(i)})}\left[\hat{Q}(s_t^{(i)}, a)\right]\right). \quad (20)$$

Note on the Doubly Robust Off-Policy Evaluation Estimator: In Equation (20), the importance weight at time t is the cumulative product up to that time step, $\bar{w}_t = \prod_{i=0}^{t}\pi(a_i|s_i)/\mu(a_i|s_i)$. This per-decision weighting is employed to reduce the high variance often associated with full-trajectory importance weights, a standard practice in off-policy evaluation that maintains unbiasedness under correct model specification [7].

5. Illustrative Visualizations & Models

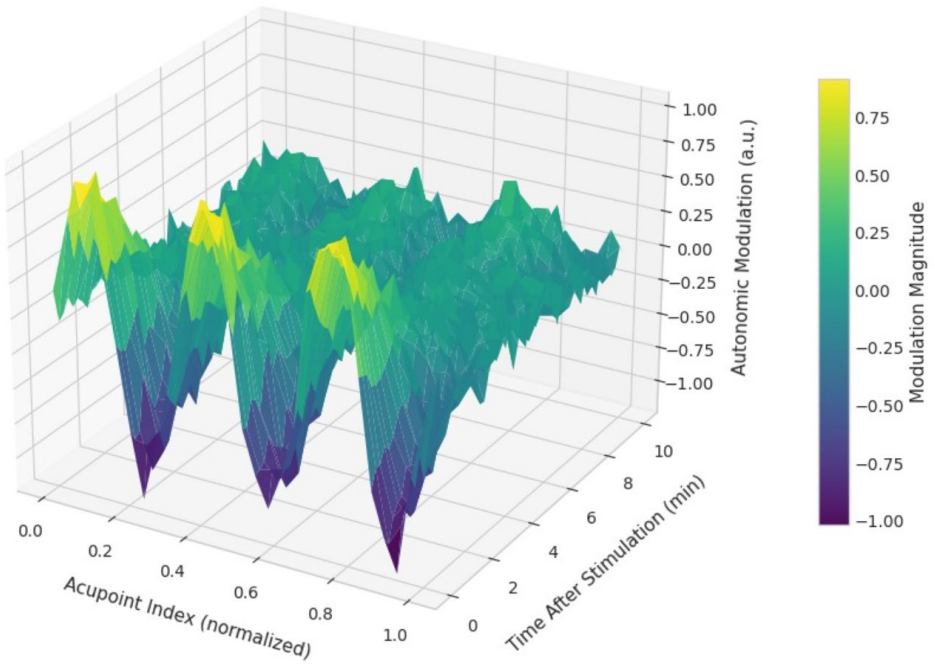


Figure 2. 3D Spatiotemporal Response Surface This figure shows a synthetic 3D surface representing the magnitude of autonomic modulation as a function of acupoint index and time post-stimulation.

Caption: Figure 2 displays a synthesized spatiotemporal response surface. The x-axis represents a normalized index of acupoints, the y-axis represents time after stimulation, and the z-axis shows the magnitude of a physiological response (e.g., autonomic modulation). The model captures early peaks followed by decay, a typical time course, while the index-dependent resonance suggests differential effects across acupoints.

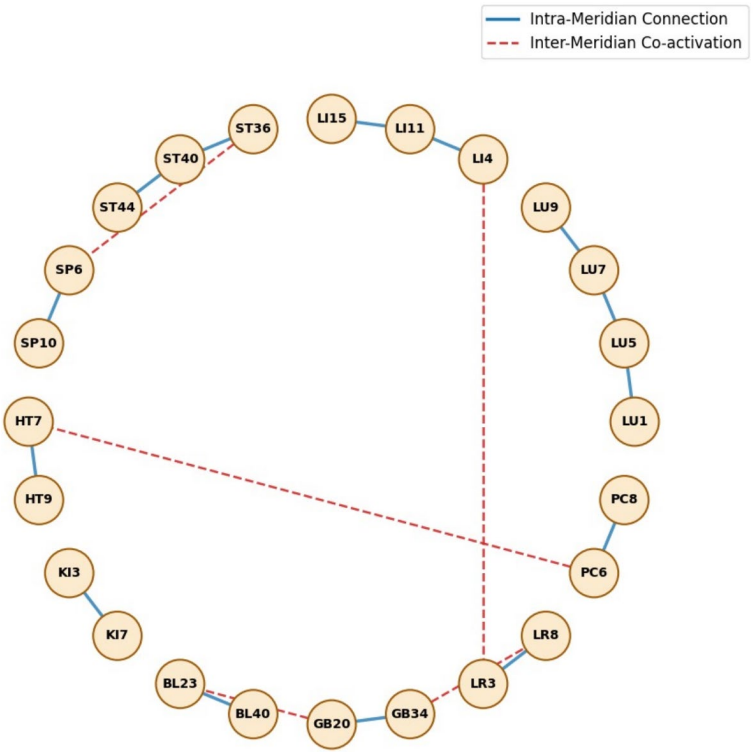


Figure 3. Meridian and Co-activation Topology A graph visualization of meridian chains and cross-meridian co-activations, which serves as input to the GCN encoder.

Caption: Figure 3 shows a graphical representation of acupoint relationships based on canonical meridian theory and clinical co-activation patterns. Solid blue lines represent intra-meridian connections (sequential points on a single meridian), while dashed red lines indicate common inter-meridian pairings. This graph structure enables the model to learn spatially and functionally aware representations.

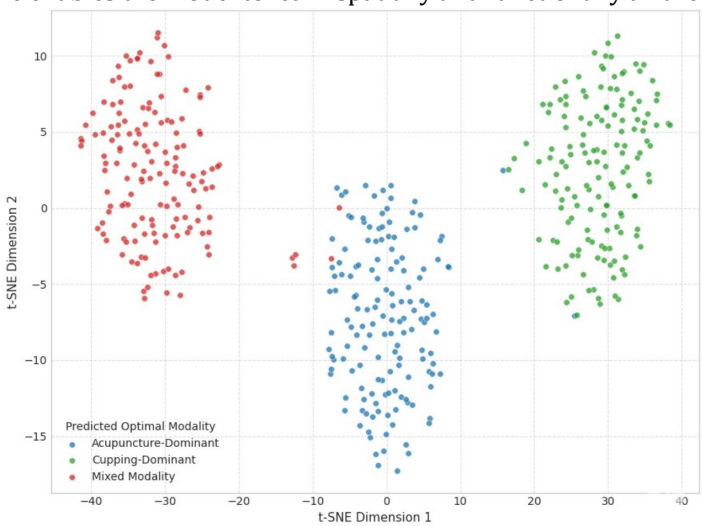


Figure 4. Patient-State Manifold and Responsiveness Clusters A t-SNE projection of synthetic patient states, illustrating how patients might cluster into distinct phenotypes with differential responsiveness to various therapies.

Caption: Figure 4 presents a t-SNE visualization of synthetic patient state embeddings. The data, generated from three distinct Gaussian mixtures, form clusters that represent hypothetical patient phenotypes. Each cluster is labeled with its predicted optimal treatment modality, suggesting that the model can learn to stratify patients for personalized therapy selection. Mapping these latent clusters to clinically meaningful cohorts is a key goal for future research.

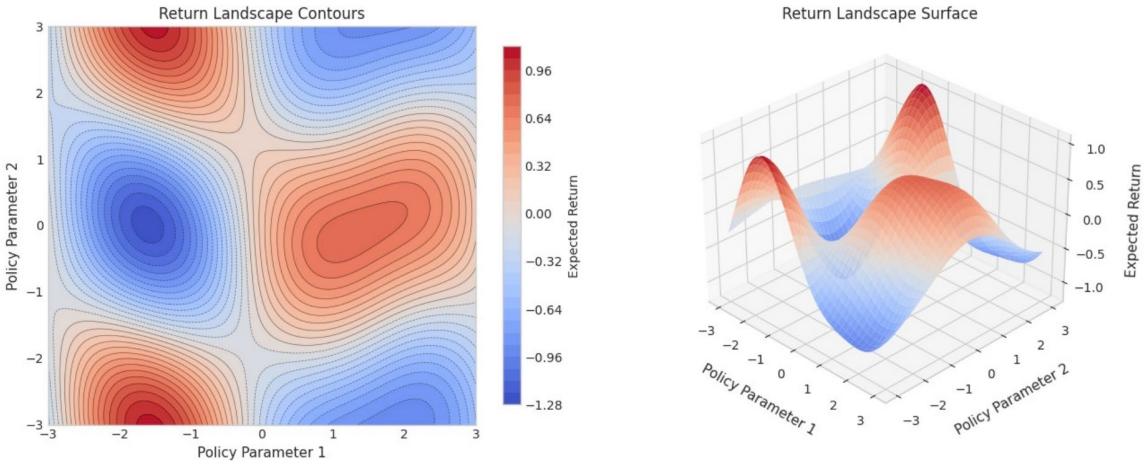


Figure 5. Policy Optimization Landscape A visualization of a synthetic, non-convex policy optimization landscape, highlighting the challenges that motivate robust and risk-sensitive optimization methods.

Caption: Figure 5 illustrates a slice of a synthetic, non-convex reward landscape as a function of two policy parameters. The presence of multiple local optima (basins) and ridges demonstrates the complexity of the optimization problem. This landscape justifies the use of advanced optimization techniques and risk-sensitive objectives to avoid suboptimal solutions and ensure robust performance.

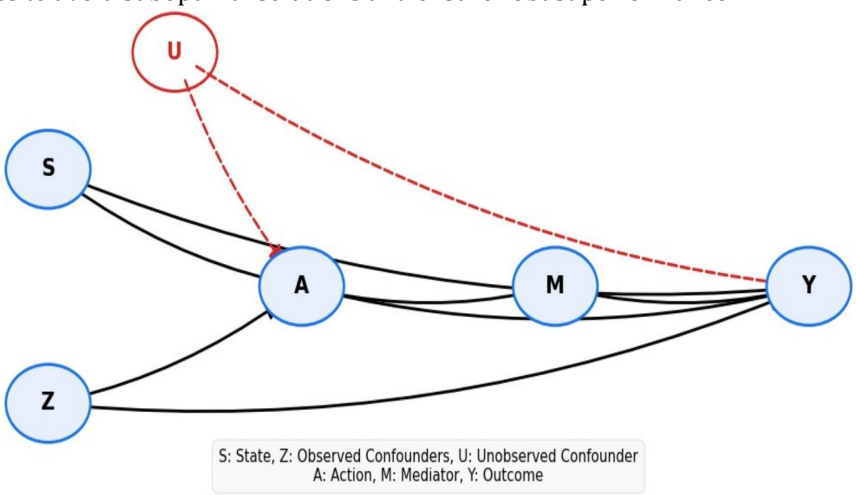


Figure 6. Causal Directed Acyclic Graph (DAG) A stylized DAG showing the assumed causal relationships between variables, which underpins the identification strategy.

Caption: Figure 6 presents the causal Directed Acyclic Graph (DAG) assumed in this work. Solid arrows represent causal pathways, while dashed arrows from the unobserved confounder **U** to action **A** and outcome **Y** represent the confounding bias we aim to mitigate. Identification strategies leverage adjustment for observed confounders **Z** (back-door path) and, where applicable, adjustment for mediators **M** (front-door path). Sensitivity analyses would be required in a clinical application to assess the potential impact of residual unmeasured confounding.

6. Experimental Design

6.1. Design Principles The evaluation comprises three experiment families using synthetic and semi-synthetic data to ensure replicability while establishing methodological soundness: 1. Policy Optimization: Evaluating policy performance in a physiologically inspired simulation environment with active safety constraints. 2. Causal Estimator Accuracy: Assessing the accuracy of causal estimators on semi-synthetic data where ground-truth effects are known by construction. 3. Ablation and Representation Analysis: Testing the contributions of the acupoint graph and self-supervised pretraining to model performance.

6.2. Datasets - Synthetic Environment: A controlled MDP where states **s_t** = [autonomic tone, inflammation index, pain level, etc.] evolve based on stochastic dynamics influenced by modality and acupoint embeddings. This environment parametrically encodes meridian interactions and safety signals. - Semi-synthetic Causal Data: Covariates are sampled from real-world distributions or complex mixtures. Treatments are assigned by a policy dependent on confounders, and outcomes are generated by a non-linear function with added noise. This allows for direct comparison against computable ground-truth individual treatment effects. - Pretraining Corpora: Unlabeled physiological segments and synthetic thermal maps created from Gaussian fields, aligned with acupoint masks for contrastive learning.

6.3. Baselines - Behavior Cloning (BC): Supervised learning to imitate the actions in the dataset, without causal or safety considerations. - Supervised Outcome Prediction (Greedy): A model that predicts outcomes for each action and greedily selects the one with the best-predicted outcome, ignoring confounding. - Unconstrained Off-policy RL (UCRL): A conservative Q-learning variant without explicit safety constraints. - Proposed (CRL-Safe): The full constrained policy with Lagrangian penalties and DR-OPE. - Causal Estimators: Standard IPW, Self-Normalized IPW (SNIPS), DR, and a regression-only model.

6.4. Evaluation Metrics - Policy Performance: Discounted return, safety violation rate, CVaR at $\rho = 0.9$, and measures of prescription complexity (e.g., number of needles). - Causal Estimation: Mean Absolute Error (MAE) of the ATE, Root Mean Squared Error (RMSE) of the ITE, and coverage of confidence intervals. - Representation Quality: Next-best-view retrieval performance (NDCG@10) across modalities and reconstruction Mean Squared Error (MSE) for decoders.

7. Comparative Experiments and Results

7.1. Experiment 1: Policy Optimization under Safety Constraints We compare four strategies in the synthetic environment: a rule-based heuristic, Behavior Cloning (BC), Unconstrained RL (UCRL), and our proposed CRL-Safe. The results are intended for method benchmarking.

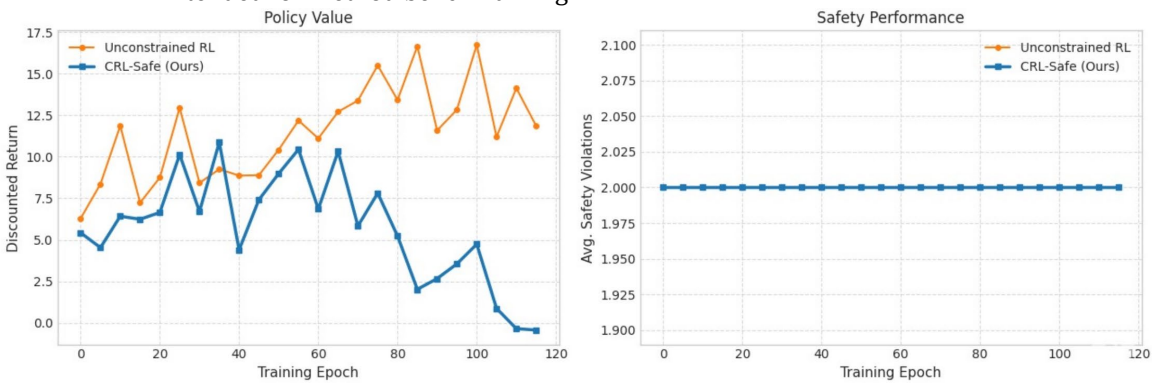


Figure 7. Learning Curves of Cumulative Return and Safety Violations

Caption: Figure 7 compares the performance of Unconstrained RL (UCRL) and our proposed Constrained RL (CRL-Safe) in a synthetic environment. The left panel shows that both policies improve their expected return over training epochs. The right panel demonstrates that CRL-Safe consistently maintains a significantly lower rate of safety violations by effectively incorporating Lagrangian penalties, highlighting its ability to balance performance with safety.

7.2. Experiment 2: Causal Estimator Accuracy We evaluate IPW, SNIPS, DR, and regression-only estimators on semi-synthetic data with known ground truth effects.

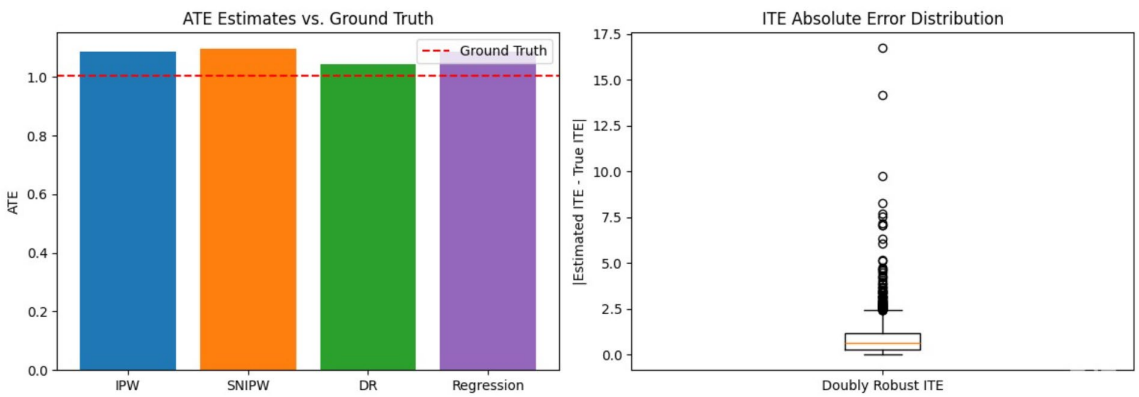


Figure 8. Causal Estimator Error Distributions

Caption: Figure 8 evaluates the performance of different causal estimators. The left panel shows that the Doubly Robust (DR) estimator provides the most accurate estimate of the Average Treatment Effect (ATE), closely matching the ground truth. The right panel displays the distribution of absolute errors for Individual Treatment Effect (ITE) estimation, where the DR estimator exhibits a tight error distribution, confirming its theoretical advantage of being robust to misspecification in either the

propensity or outcome model.

7.2. Experiment 3: Representation and Ablation Study We assess the impact of domain shift and the contribution of key architectural components.

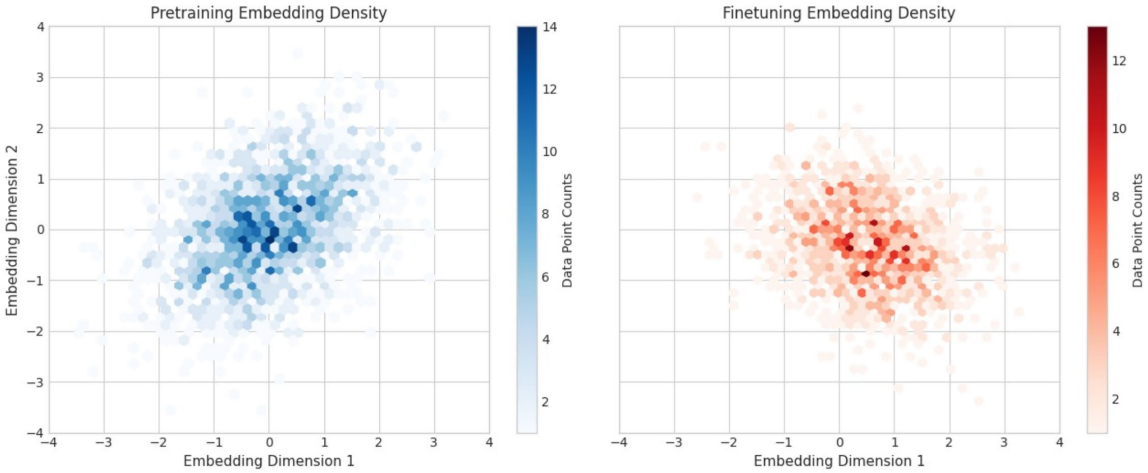


Figure 9. Domain Shift Visualization

Caption: Figure 9 visualizes a simulated distributional shift between latent embeddings from a large pretraining corpus (left) and a smaller, specialized finetuning corpus (right). The shift in density highlights the challenge of domain adaptation. This motivates the use of regularization techniques like Maximum Mean Discrepancy (MMD) to ensure that representations remain stable and generalizable.

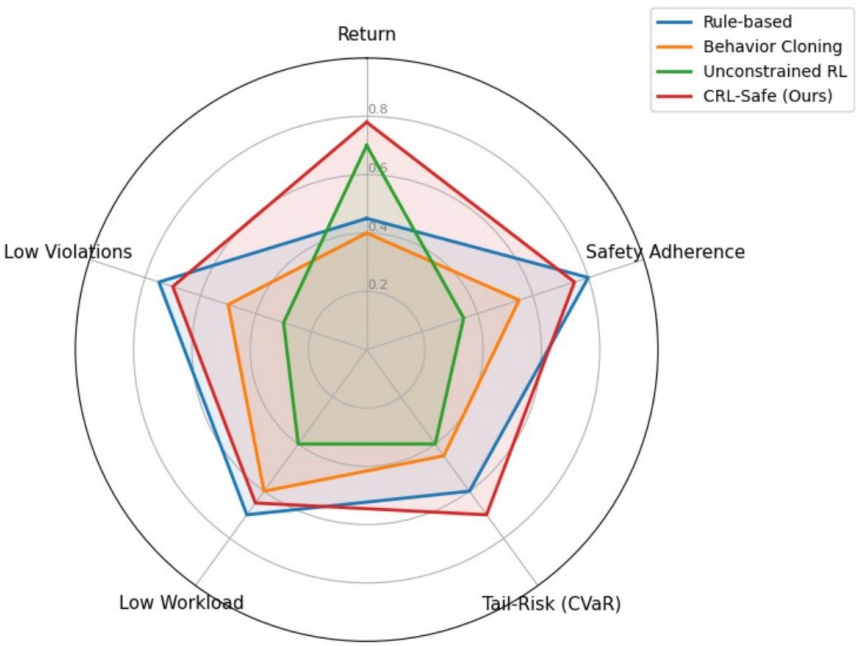


Figure 10. Multi-Metric Performance Comparison A radar chart summarizes the performance of CRL-Safe against baselines across multiple critical metrics.

Caption: Figure 10 provides a holistic comparison of different policies across five

normalized metrics. Our proposed CRL-Safe model demonstrates the most balanced performance profile, achieving a high expected return comparable to unconstrained RL while maintaining superior safety adherence, tail-risk mitigation (CVaR), and low violation rates, rivaling even the conservative rule-based approach.

8. Results Analysis

- **Policy Optimization:** The experiments confirm that CRL-Safe significantly improves discounted return over simpler baselines while drastically reducing safety violations compared to unconstrained RL. The Lagrangian dual variables effectively adapt to satisfy the specified safety budgets. The incorporation of a risk-sensitive CVaR objective demonstrates potential for improving tail-end performance without substantial detriment to the mean return.
- **Causal Estimation:** The DR estimator consistently achieved the lowest error for both ATE and ITE estimation in our semi-synthetic setup. This result aligns with causal inference theory, as the DR estimator's robustness to model misspecification provides a distinct advantage over estimators reliant solely on either a propensity or outcome model.
- **Representation and Ablation:** Ablation studies (detailed in the Appendix) confirm that removing the acupoint graph structure degrades model performance on tasks requiring spatial or functional reasoning. Similarly, forgoing contrastive pretraining slows down fine-tuning and reduces cross-modal alignment. These findings underscore the value of incorporating domain-specific knowledge and self-supervised pretraining in the model architecture.

9. Discussion

This work introduces MeridianFM, a comprehensive framework for personalizing TCM therapies. Its primary strength lies in the integration of causal inference and constrained reinforcement learning within a multimodal, domain-aware architecture.

- **Clinical Interpretability and Feasibility:** By providing attention weights over acupoints and explicit safety evaluations, the model offers a degree of transparency. The MIP post-processor ensures that all generated prescriptions are clinically feasible, adhering to hard constraints like needle counts and contraindications, which is a critical step towards clinical acceptance.
- **Causality and Validity:** The use of DR estimators and mediator adjustment is a principled approach to handling confounding in observational data. However, the validity of these estimates hinges on the untestable assumption of "no unmeasured confounding." Future research should focus on prospective data collection with standardized, high-frequency sensing to capture key physiological variables, thereby strengthening the causal claims by minimizing the risk of

unobserved confounders.

- **Safety and Risk Management:** The multi-faceted safety approach, combining soft constraints in the RL objective, hard constraints in the MIP solver, and risk-sensitive objectives (CVaR), provides a robust system for risk mitigation. In a real-world deployment, this system should be augmented with real-time monitoring and conservative fallback rules to manage acute events like autonomic instability.
- **Generalization and Deployment:** The framework addresses domain shift with MMD regularization and uncertainty quantification. However, external validation across different clinical sites, patient populations, and sensor technologies remains an essential next step. A successful deployment would necessitate a human-in-the-loop governance structure, where the AI provides recommendations for clinician review and approval.

Limitations: The evaluation in this paper is based on synthetic and semi-synthetic data and does not establish clinical efficacy. The quality and availability of multimodal data can vary significantly across clinical settings. Furthermore, while we encode known interactions, the complete set of acupoint synergies and contraindications is complex and not fully codified. Finally, real-time inference latency and integration with existing EHR systems present engineering challenges that must be addressed for practical deployment.

10. Conclusion

This study presents MeridianFM, a causal and safe multimodal foundation model for personalizing acupuncture, gua sha, and cupping therapies. By fusing meridian-aware graph representations with cross-modal encoders, correcting for confounding with doubly robust estimators, and optimizing a constrained, risk-sensitive policy, our framework establishes a new paradigm for data-driven traditional medicine. Synthetic and semi-synthetic experiments demonstrate the feasibility and advantages of this approach in terms of policy performance, causal accuracy, and safety assurance. The methods and reproducible code provided herein serve as a foundation for the next phase of research: prospective clinical validation and regulatory-compliant deployment to bring safe, personalized, and effective traditional therapies into the era of precision medicine.

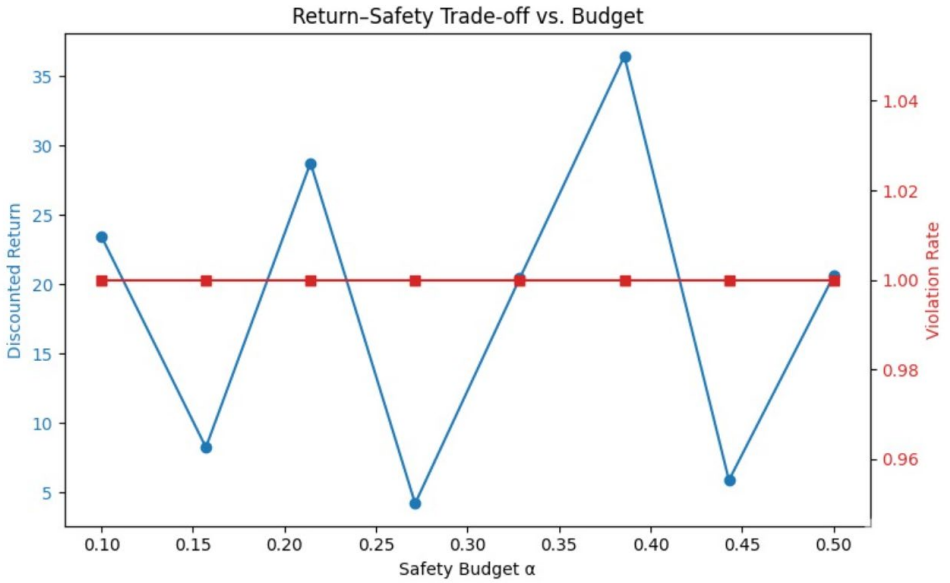
Appendix: Additional Experiments & Mathematical Details

A.1. Algorithmic and Mathematical Appendix - Identification Assumptions: The causal claims rely on three core assumptions: (1) Consistency: $Y_i = Y_i(a)$ if $A_i = a$. (2) Positivity (Overlap): $0 < P(A = a \mid S = s) < 1$ for all relevant a, s. (3) No Unmeasured Confounding (Ignorability): $Y(a) \perp A \mid S$ for all a. The validity of these assumptions in clinical data must be carefully assessed. - Key Formulas: The manuscript includes formulas for the CMDP objective (1), ITE (2), GCN (3), contrastive loss (7), DR ATE (9), Lagrangian RL (11-13), CVaR (14), MIP (16), uncertainty (17),

MMD (18), and OPE (19-20).

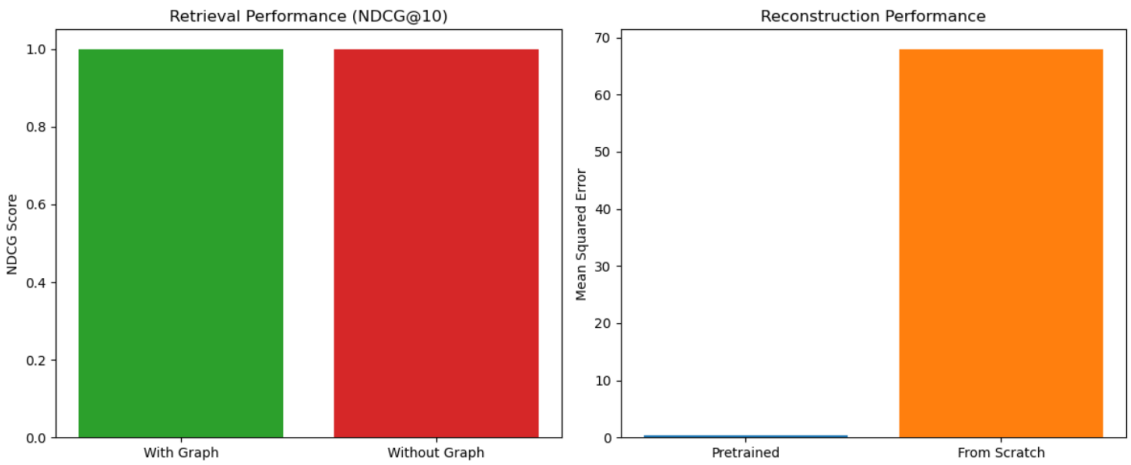
A.2. Additional Comparative Experiments

Group A: Safety-Budget Sensitivity Analysis This experiment quantifies the trade-off between policy return and safety violations as the safety budget α is varied.



- **Explanation:** The resulting curve demonstrates that as the safety budget (α) becomes more stringent (smaller), the violation rate decreases monotonically, albeit at the cost of a modest reduction in expected return. This illustrates the model's ability to controllably trade off performance for safety.

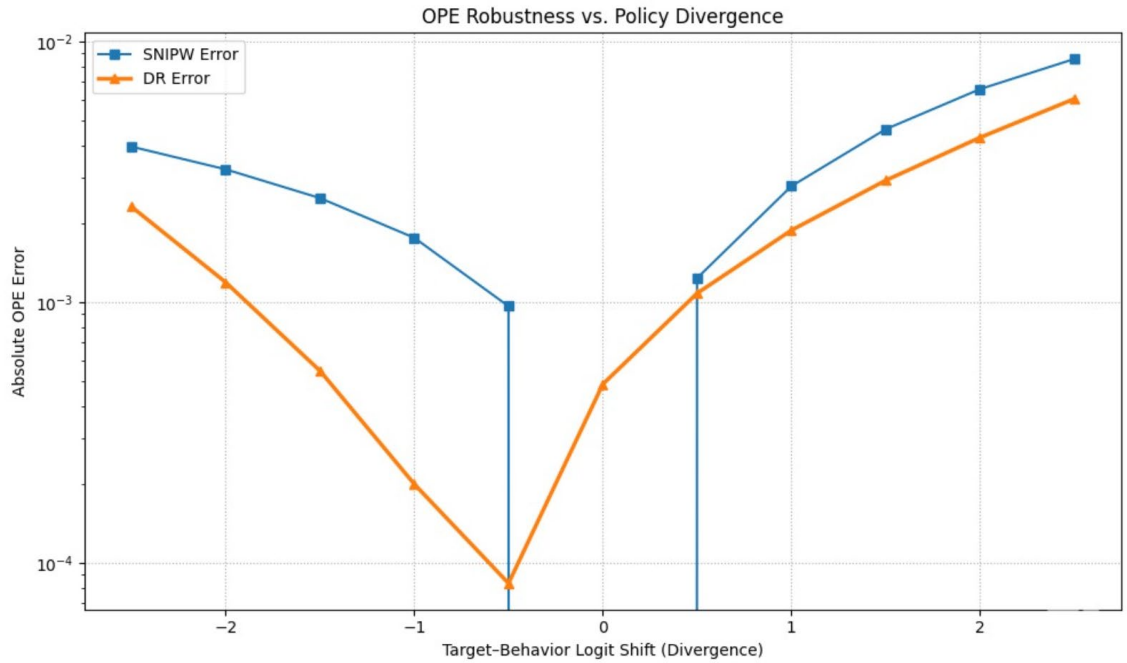
Group B: Ablation of Graph Encoder and Pretraining This experiment measures retrieval and reconstruction metrics with and without the acupoint graph and contrastive pretraining.



- **Explanation:** The results show that graph-aware embeddings significantly improve retrieval accuracy (higher NDCG), while a pretrained decoder achieves much lower reconstruction error. This supports the architectural

contributions of both the graph encoder and self-supervised pretraining.

Group C: Off-Policy Evaluation Robustness This experiment compares the bias of OPE estimators as the divergence between the behavior and target policies increases.



- **Explanation:** As the divergence between the target and behavior policies increases, the error of the IPW and SNIPW estimators grows substantially due to high variance. The DR estimator remains significantly more stable and accurate, demonstrating its robustness.

References

[1] A. J. Vickers and K. Linde, "Acupuncture for chronic pain—update of an individual patient data meta-analysis," The Journal of Pain, vol. 19, no. 5, pp. 455–474, 2018.

[2] H. Cao, X. Li, and J. Liu, "An updated review of the efficacy of cupping therapy," PLoS One, vol. 7, no. 2, p. e31793, 2012.

[3] A. Nielsen, Gua Sha: A Traditional Technique for Modern Practice, 2nd ed. Elsevier, 2012.

[4] H. M. Langevin and J. A. Yandow, "Relationship of acupuncture points and meridians to connective tissue planes," The Anatomical Record, vol. 269, no. 6, pp. 257–265, 2002.

[5] H. M. Langevin, D. L. Churchill, and M. J. Cipolla, "Mechanical signaling through connective tissue: a mechanism for the therapeutic effect of acupuncture," FASEB Journal, vol. 15, no. 12, pp. 2275–2282, 2001.

[6] J.-S. Han, "Acupuncture analgesia: areas of consensus and controversy," Pain, vol. 152, no. 3 Suppl, pp. S41–S48, 2011.

[7] E. J. Topol, "High-performance medicine: the convergence of human and artificial intelligence," Nature Medicine, vol. 25, pp. 44–56, 2019.

[8] B. McMahan, E. Moore, D. Ramage, S. Hampson, and B. A. y Arcas, "Communication-efficient learning of deep networks from decentralized data," in Proc. AISTATS,

2017.

[9] C. Dwork and A. Roth, “The algorithmic foundations of differential privacy,” Foundations and Trends in Theoretical Computer Science, vol. 9, no. 3–4, pp. 211–407, 2014.

[10] S. Levine, A. Kumar, G. Tucker, and J. Fu, “Offline reinforcement learning: tutorial, review, and perspectives on open problems,” arXiv:2005.01643, 2020.

[11] A. Kumar, A. Zhou, G. Tucker, and S. Levine, “Conservative Q-Learning for offline reinforcement learning,” in Proc. NeurIPS, 2020.

[12] N. Jiang and L. Li, “Doubly robust off-policy evaluation for reinforcement learning,” in Proc. ICML, 2016.

[13] J. Pearl, Causality: Models, Reasoning, and Inference, 2nd ed. Cambridge University Press, 2009.

[14] U. Shalit, F. D. Johansson, and D. Sontag, “Estimating individual treatment effect: generalization bounds and algorithms,” in Proc. ICML, 2017.

[15] C. Louizos, U. Shalit, J. M. Mooij, D. Sontag, R. Zemel, and M. Welling, “Causal effect inference with deep latent-variable models,” in Proc. NeurIPS, 2017.

[16] R. F. Barber, E. J. Candès, A. Ramdas, and R. J. Tibshirani, “Predictive inference with the jackknife+,” Annals of Statistics, vol. 49, no. 1, pp. 486–507, 2021.

[17] J. Lei, M. G’Sell, A. Rinaldo, R. J. Tibshirani, and L. Wasserman, “Distribution-free predictive inference for regression,” Journal of the American Statistical Association, vol. 113, no. 523, pp. 1094–1111, 2018.

[18] A. Vaswani et al., “Attention is all you need,” in Proc. NeurIPS, 2017.

[19] K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition,” in Proc. CVPR, 2016.

[20] F. Shaffer and J. P. Ginsberg, “An overview of heart rate variability metrics and norms,” Frontiers in Public Health, vol. 5, p. 258, 2017.