



Multi-Hub Sheaf-ML for Early Alzheimer’s Detection

Integrating Multi-Domain Biomarkers via Sheaf Learning

Presenter: **Duc Thanh Nguyen**^{1,4}, Authors: Lokendra S Thakur^{1,2}, Gurpreet Bharj^{1,3}, Duc Thanh Nguyen^{1,4}, Manish Saroya^{1,5}, Bushra Malik^{1,6}

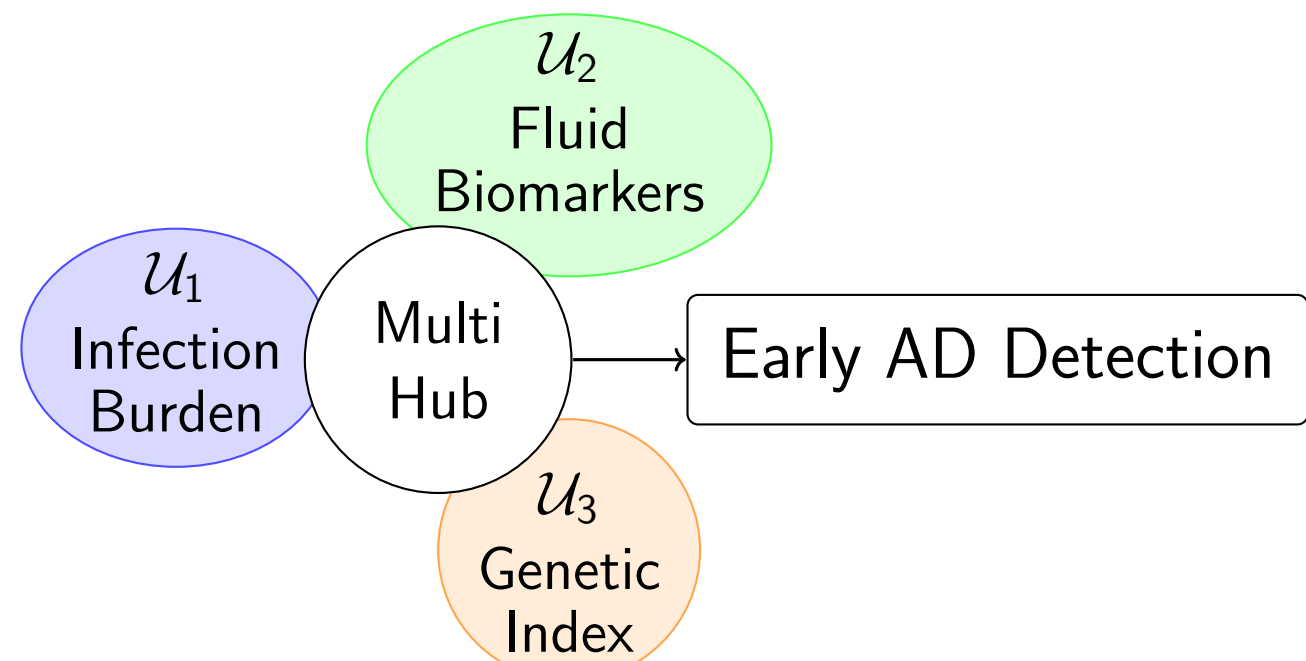
¹Innovation Engine Lab (iEL), MDtRI; ²Massachusetts General Hospital, Harvard Medical School; ³Stanford University; ⁴Bucknell University; ⁵Honda Research Institute; ⁶University of Toronto



HARVARD
MEDICAL SCHOOL

MASSACHUSETTS
GENERAL HOSPITAL

Clinical Motivation



- ▶ Late diagnosis limits intervention effectiveness^[1]
- ▶ Single-modality models underperform^[2,3]
- ▶ Multi-Hub Sheaf-ML extends our earlier study Sheaf-ML with multiple topology-aware hubs for richer cross-modal fusion^[4].

Key Insight

- ▶ Capture latent cross-modality dependencies.
- ▶ Improve robustness under missing or incomplete data, and different datasets.
- ▶ Provide interpretable patient-level and population-level subtypes.
- ▶ Multi-Hub Sheaf-ML transforms heterogeneous multimodal data into interpretable patient- and population-level insights for personalized early risk assessment.

Datasets

Cohort	Samples	Features	Modalities
ADNI ^[8]	11,149	44	Demographics (3): age, sex, education years Cognition (15): MMSE, MoCA, CDR-SB, ADAS11/13, FAQ, RAVLT measures Imaging (12): hippocampus, whole brain, ventricles, entorhinal, fusiform, FDG, PIB, AV45 Biomarkers (6): A/ β , tau, p-tau, CSF availability, amyloid/tau pathology scores Pathology (1): overall disease burden Clinical Scores (2): neurodegeneration score, modality completeness score
GLEND ^[9]	302	33	Demographics (2): image width, image height Color & Texture (8): RGB statistics, entropy, contrast, homogeneity, energy Lesion Morphology (9): lesion count, area, eccentricity, solidity, perimeter, circularity Pathology (10): fibrosis, inflammation, structural distortion, disease burden, organ involvement Clinical Scores (3): inflammatory activity index, anatomical complexity score, endometriosis probability

Figure: Summary of multimodal cohorts used for evaluating the Multi-Hub Sheaf-ML model. Feature counts correspond to the variables used as model inputs after preprocessing and modality organization.

Method Overview (Multi-Hub Sheaf-ML Model)

Formal Construction of the PBI

Let the set of domain-specific indices be:

$$\mathcal{I} = \{I_{\text{fluid}}, I_{\text{gen}}, I_{\text{img}}, I_{\text{cog}}, I_{\text{inf}}\}$$

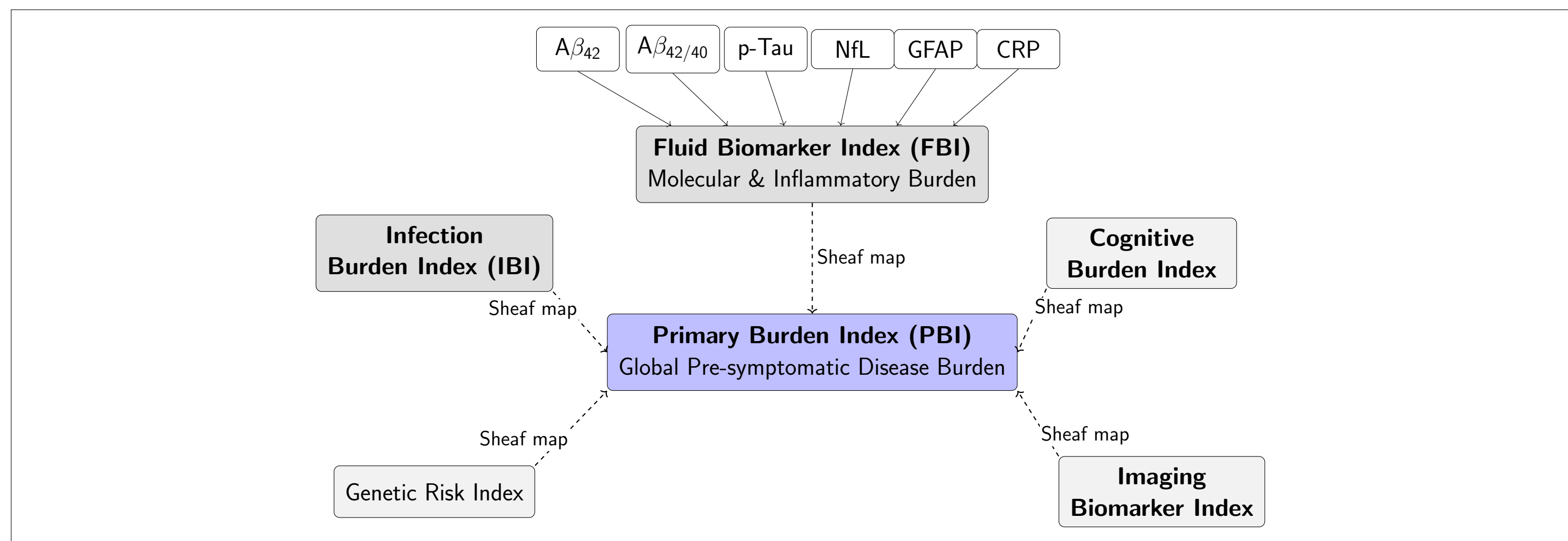
The PBI is defined as a global latent variable:

$$Z_{\text{PBI}} = \arg \min_z \sum_{k \in \mathcal{I}} |\phi_k(I_k) - z|^2 + \lambda \Omega(z)$$

where: ϕ_k denotes a learned sheaf morphism mapping index I_k into the global disease-burden space, $\Omega(z)$ enforces temporal smoothness, biological plausibility, and robustness to missing modalities, λ controls the trade-off between cross-domain agreement and regularization.

Model Architecture

Explicit representation of the Primary Burden Index (PBI) within the Sheaf-ML model^[4]. Raw fluid biomarkers first integrate into the Fluid Biomarker Index (FBI), a domain-specific biological hub. The PBI is a higher-order latent variable that emerges from consistency constraints across all domain indices, including molecular (FBI), genetic, imaging, cognitive, and infection (IBI)-related burden. This structure enables robust detection of pre-symptomatic AD/ADRD by amplifying weak but concordant cross-domain signals.



Results

Sheaf Multi-Hub Graph and Modality Contributions^[4]

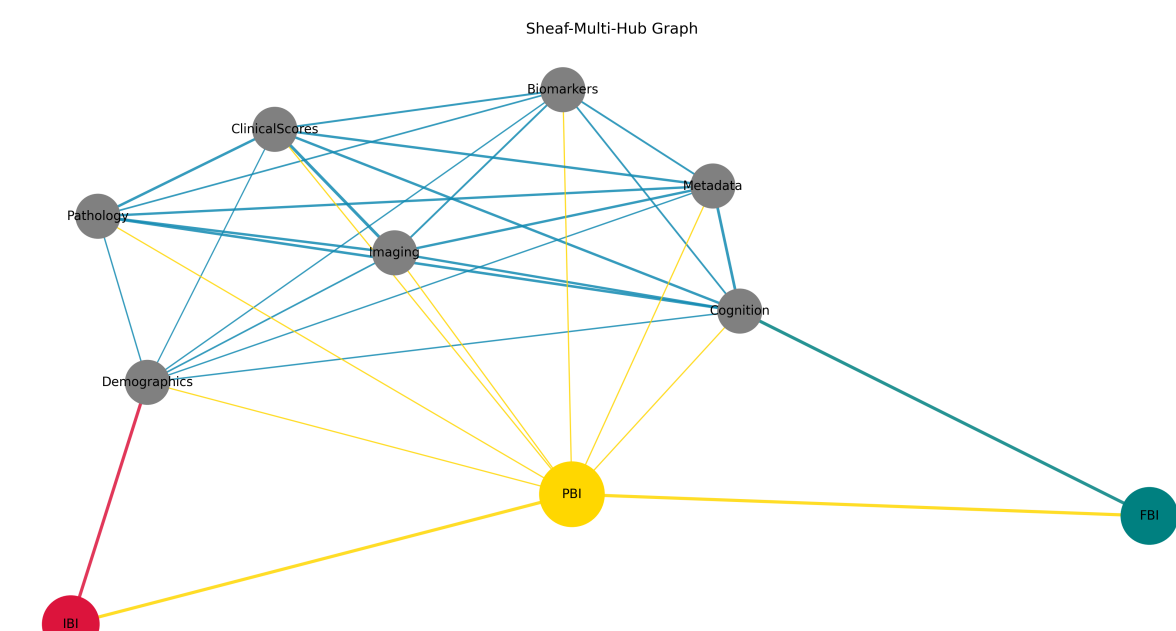


Figure: ADNI Modality Sheaf Network

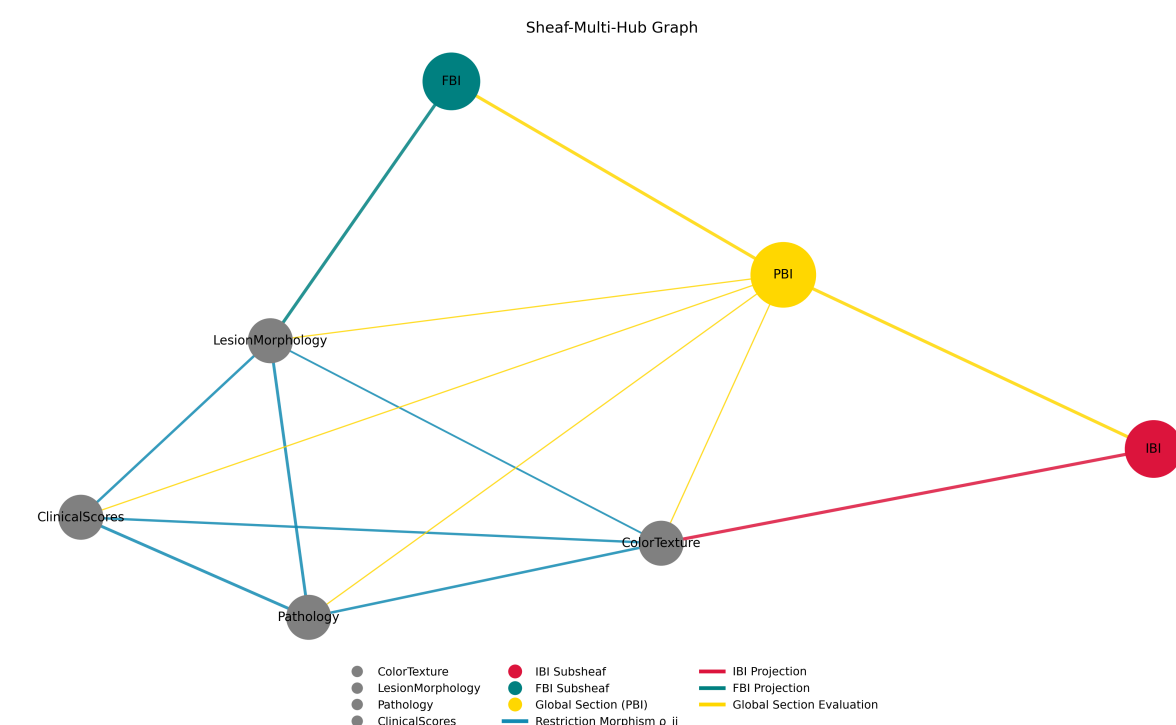


Figure: GLEND^[9] Modality Sheaf Network

Figure: Robustness check applying to different datasets

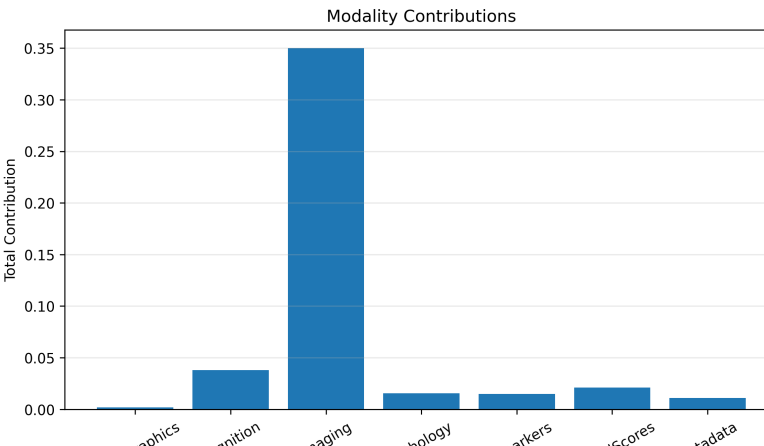


Figure: ADNI Top Drivers

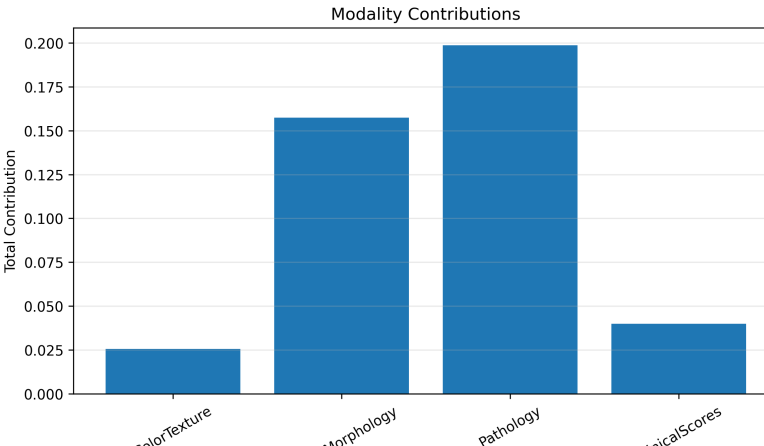


Figure: GLEND^[9] Top Drivers

- ▶ Sheaf-ML extends feature concatenation by learning restriction morphisms between modalities and enforcing cross-modal consistency during integration graph construction.^[2,3,4]
- ▶ Learned restriction morphisms explicitly encode cross-domain compatibility, linking biologically related signals within a unified sheaf approach.^[4,5,6,10]
- ▶ The integration graph extends beyond correlation by encoding compatibility constraints through restriction morphisms, enabling structured cross-modal coupling.^[4]
- ▶ **Critical note:** Higher modality contribution does not imply biological dominance. Sheaf consistency-based edge weighting emphasizes cross-modal agreement, preventing high-dimensional modalities from disproportionately influencing the inferred biological network.

Patient-to-Population Subtypes

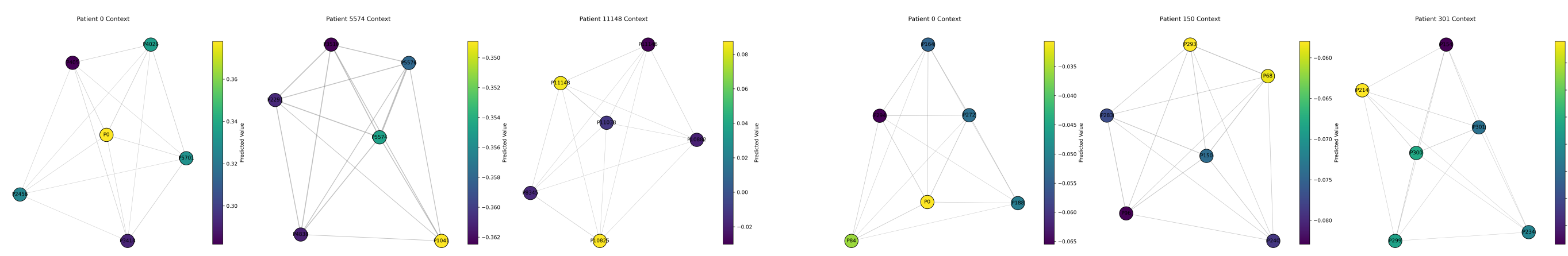


Figure: ADNI Patient-Level Subtype

Figure: GLEND^[9] Patient-Level Subtype

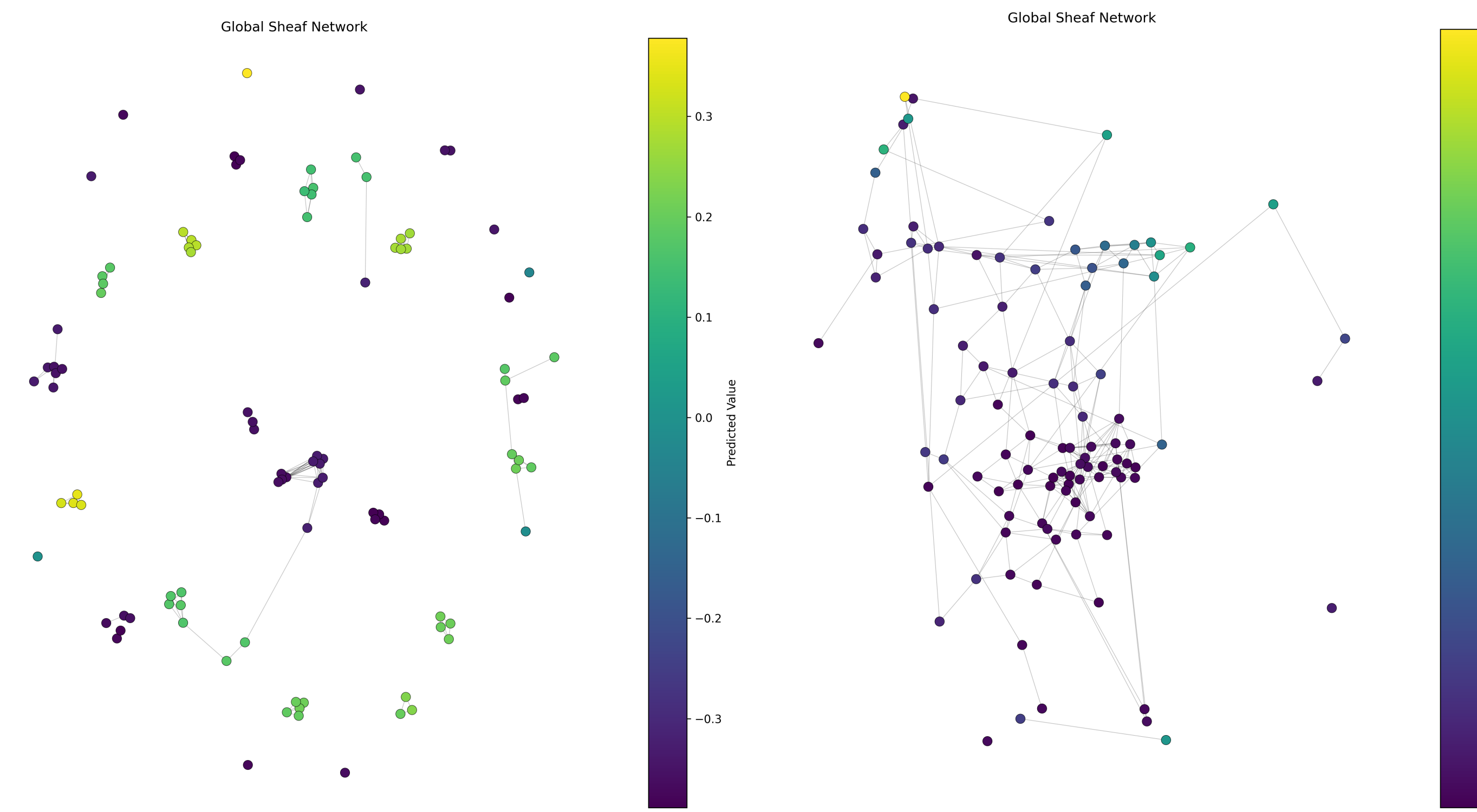


Figure: ADNI Population-Level Subtype

Figure: GLEND^[9] Population-Level Subtype

Figure: Patient-to-Population Level Subtype Graphs

Patient-Level Context Networks: Patients are visualized within neighborhoods of biologically similar individuals derived from multimodal data.^[4] Node color indicates predicted disease burden, highlighting stable versus heterogeneous disease profiles.

Population-Level Networks: Global sheaf networks uncover disease phenotypes, progression gradients, and atypical cases across the cohort, linking individual predictions to broader population structure.

Clinical Interpretability

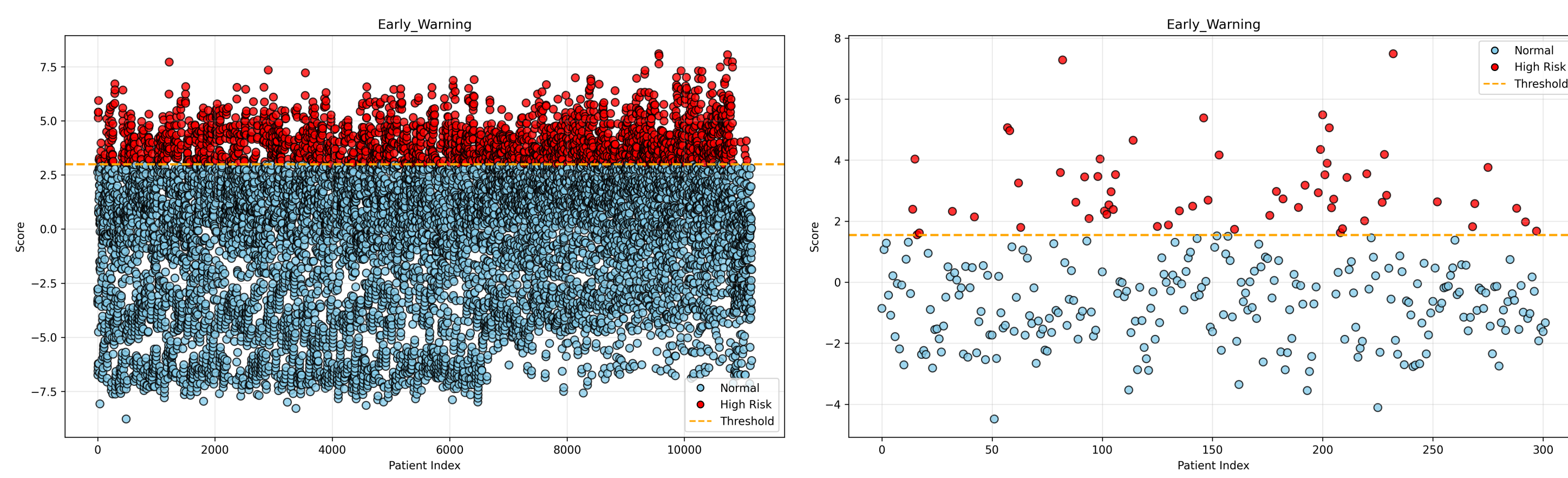


Figure: Early AD Detection

Figure: Early Endometriosis Detection

Early Warning Stratification: Patients are stratified using a multimodal risk score derived from integrated clinical, imaging, and biomarker information. Individuals exceeding the warning threshold (red) exhibit profiles associated with elevated disease burden or potential progression. Patients exceeding the 80th percentile are flagged as early-warning cases, corresponding to approximately 20% of the cohort.^[4]

- ▶ **Patient context:** High-warning patients are interpreted against neighborhoods of biologically similar individuals, linking risk alerts to stable or heterogeneous disease profiles.^[4]
- ▶ **Population context:** Cohort-level structure helps distinguish subtype-consistent risk, transitional states, and atypical presentations for targeted clinical assessment.^[4,7]
- ▶ **Robustness context:** The GLEND^[9] panel tests whether the warning-stratification view remains interpretable in a histopathology-based imaging–pathology dataset.^[9]

Conclusion

Population Structure → Patient Context → Early Warning → Clinical Decision Support

- ▶ Multi-Hub Sheaf-ML unifies heterogeneous biomedical modalities into a clinically interpretable disease representation.^[4,5,6]
- ▶ The model links individual patient risk to broader population structure through patient-context and global similarity networks.^[4,7]
- ▶ Early-warning stratification highlights patients with elevated multi-hub disease burden, supporting proactive clinical assessment.
- ▶ Validation on ADNI and GLEND^[9] demonstrates the feasibility of translating multi-hub network intelligence into personalized early risk assessment and decision support.^[8,9]

Take-home: Multi-Hub Sheaf-ML transforms heterogeneous multimodal data into interpretable patient- and population-level insights for personalized early Alzheimer’s detection for early interventions and optimal clinical trial design leading to promising therapeutics.

Acknowledgements

We gratefully acknowledge the contributors and custodians of the datasets used in this study. This work utilizes data from the ADNI neuroimaging cohort,^[8] which has enabled large-scale investigation of cognitive decline, neuroimaging biomarkers, and molecular pathology in aging and Alzheimer’s disease research. We also acknowledge the GLEND^[9] dataset, which provides histopathology-based annotations essential for evaluating structured imaging–pathology relationships in our Multi-Hub Sheaf-ML model. We thank the respective study investigators, data contributors, and funding agencies for making these datasets publicly available, thereby enabling large-scale multi-modal biomedical research. We would also like to acknowledge our collaborators: Dr. Fernando Gómez-Baquero PhD, Cornell Tech, NYC, United States; Dr. Jason Flannick PhD, Boston Children’s Hospital-Harvard Medical School & Broad Institute of MIT and Harvard, United States; Dr. Eyal Y. Kimchi MD-PhD, Northwestern University, IL, United States; Dr. James Berry MD-MPH, MGH-Harvard Medical School, United States; Dr. Satrajit Ghosh MGH-Harvard Medical School & Massachusetts Institute of Technology (MIT), United States; Dr. Russell Rockne, City of Hope, LA, United States; Dr. Charlotte Teunissen PhD, Amsterdam University Medical Center, the Netherlands; Dr. Nadine Tatton PhD; Dr. Amanpreet Badhwar MSc-PhD, University of Montreal, Montreal, Canada; Dr. Kamana Porwal PhD, IIT-New Delhi, India. Finally we do acknowledge our research team at MDtRI’s iEL, administrative, compliance, and finance staff for their valuable support.

References

[1] Borchert et al., Alzheimer’s & Dementia, 2023; [2] Stahlshmidt et al., Brief. Bioinform., 2022; [3] Acosta et al., Nat. Med., 2022; [4] Thakur et al., medRxiv, 2025, doi:10.1101/2025.11.10.25339915. [5] Bredon, Sheaf Theory, 1997; [6] Mac Lane & Moerdijk, Sheaves in Geometry and Logic, 1992; [7] Allen, J. Pers. Med., 2024; [8] Petersen et al., Neurology, 2010, doi:10.1212/WNL.0b013e3181cb3c25. [9] Leibetseder et al., GLEND^[9]: Gynecologic Laparoscopy Endometriosis Dataset, LNCs/MMM, 2020, doi:10.1007/978-3-030-37734-2_36; [10] Wang et al., J. Neuroinflammation, 2024.

Contact

Dr. Lokendra S. Thakur M.B.A., M.S., Ph.D.
Affiliations: 1.MDtRI Foundation, innovation Engine Lab (iEL).
2.Neurology Department, MGH-Harvard Medical School.
Email: lthakur@mdtri.org; Mobile: +1 916-934-4112

