# AAAI-2025 SPARTA Workshop Abstract

## NeuroHydra: A Generalizable DINOv3–Mamba Framework for Multimodal Biomedical AI

Multimodal integration is central to biomedical AI, yet current approaches often treat imaging and clinical data as independent streams or rely on computationally expensive 3D architectures. We present **NeuroHydra**, a generalizable framework that bridges 2D self-supervised vision models (DINOv3) with 3D medical imaging and structured clinical variables. NeuroHydra introduces a **Structure-Aware Visual Slice Fusion (AS-VSF)** module that reconstructs volumetric context by learning deformable relationships across MRI slices, maintaining anatomical continuity without requiring full 3D supervision. Clinical and tabular features are encoded and fused with imaging representations through a **Mamba state-space integration layer**, enabling sequential multimodal reasoning over spatially distributed pathology patterns. Applied to epilepsy surgical outcome prediction, NeuroHydra demonstrates improved performance over late-fusion and transformer-based baselines while remaining computationally efficient. **Grad-CAM and SHAP** support multi-level attribution, illustrating how imaging and clinical features jointly influence predictions. The framework is extensible to segmentation, reconstruction, and broader translational applications. Future work will include multi-site validation and expanded explainability analyses. NeuroHydra offers a scalable, interpretable, and modality-aware approach to multimodal biomedical AI.

**Keywords:** multimodal biomedical AI; medical image fusion; MRI deep learning; self-supervised vision models; DINOv3; state-space models; Mamba architecture; structure-aware slice fusion; volumetric representation learning; clinical data integration; epilepsy outcome prediction; neuroimaging biomarkers; interpretable AI in healthcare; Grad-CAM explainability; SHAP feature attribution; hierarchical attribution; sequential multimodal reasoning; computationally efficient medical AI; translational neuroimaging; clinical decision support systems.

**Title:** NeuroHydra: A Generalizable DINOv3–Mamba Framework with Structure-Aware Visual Slice Fusion for Multimodal Biomedical AI

**Use Case:** Proof-of-concept integrating multimodal MRI and clinical data to predict surgical outcomes in epilepsy.

## Background

Multimodal integration is essential in biomedical AI—imaging, clinical, and structured data capture complementary information. However, existing models often process modalities independently or fuse them late, limiting fine-grained cross-modal interaction, interpretability, and computational efficiency.

## Objective

- Develop NeuroHydra, a framework that:
    - **Bridges** 2D self-supervised vision models (e.g., DINOv3) and 3D medical data via a Structure-Aware Visual Slice Fusion (AS-VSF) module.
    - **Fuses** imaging and structured clinical data through a Mamba state-space layer for joint, sequential reasoning.
    - **Maintains** computational efficiency, interpretability, and scalability across diverse biomedical tasks.

## Methods

**Imaging:**

- 2D DINOv3 backbone extracts slice-level embeddings.
- AS-VSF learns deformable alignment (“token warp”) along the z-axis to recover 3D anatomical continuity.
- Unlike 3D transformers, AS-VSF adaptively models cross-slice relationships without assuming uniform spacing, remaining robust to variable slice gaps and missing slices.

**Clinical / Tabular:**

- Encoded via a lightweight MLP and synchronized or case-aligned with imaging embeddings.

**Fusion & Prediction:**

- Pipeline: DINOv3 → AS-VSF → Token Merging → Mamba (Linear Attention) → Task Head • Supports lesion segmentation, surgical outcome prediction, and potential 3D reconstruction tasks.

**Explainability (ongoing):**

- Grad-CAM for imaging tokens
- SHAP for tabular features
- Integrated visualization of modality contributions
- **Multi-Level Attribution:** Combined Grad-CAM and SHAP visualizations illustrate how spatial patterns and clinical variables jointly influence predictions.
- **Sequential Multimodal Reasoning:** The Mamba fusion layer models dependencies across slices and structured variables, supporting physiologically grounded inference rather than isolated feature scoring.

**Comparative Baselines:**

- ConvNeXt + LGBM (late fusion)
- UNETR + TabCat (transformer shallow fusion)
- DINOv3 + MLP (without AS-VSF or Mamba)

## Significance

- **Bridges** 2D and 3D domains: learns volumetric context from slice-level data without full 3D supervision.
- **Achieves** computational efficiency: avoids exponential 3D attention cost while preserving spatial coherence.
- **Enables** interpretable fusion: visualizes which slices and clinical factors drive predictions.
- **Ensures** robustness and generalizability: applicable beyond epilepsy, including segmentation and reconstruction.

## Conclusion

- NeuroHydra integrates self-supervised vision transformers, learnable slice fusion (AS-VSF), and state-space modeling (Mamba) into a scalable, interpretable, and efficient multimodal framework.
- The epilepsy surgical outcome experiment serves as a proof-of-concept, demonstrating potential for broader biomedical and clinical integration.
- Future work will expand explainability, validate across multi-site datasets, and extend to additional multimodal and translational applications.