

Machine Learning Approaches for Early Detection of Neurodegenerative Disorders Using Multimodal Biomarker Data

Internal Seed Grant Proposal

📅 Submitted: 15 September 2026

Principal Investigator

👤 Dr. Sarah Chen
✉ s.chen@meridian.edu
☎ (555) 234-8901
🏢 Dept. of Computational Biology
🏫 School of Engineering & Applied Science

Co-Principal Investigator

👤 Dr. Marcus Webb
✉ m.webb@meridian.edu
☎ (555) 234-8912
🏢 Dept. of Neurology
🏫 School of Medicine

Funding Program: Internal Seed Grant
Project Period: 1 Jan – 31 Dec 2027

Amount Requested: \$48,750
Keywords: neurodegeneration, machine learning, multimodal biomarkers

Project Summary. Neurodegenerative disorders including Alzheimer’s and Parkinson’s disease affect over 50 million people worldwide, yet early detection remains a critical clinical challenge. This proposal seeks \$48,750 in seed funding to develop a machine learning pipeline that integrates multimodal biomarker data — neuroimaging, cerebrospinal fluid proteomics, and wearable sensor time series — for early risk stratification. The 12-month project leverages existing clinical datasets from the Meridian University Hospital system ($n \approx 1,200$ patients) and will establish preliminary data and computational infrastructure for a competitive NIH R01 submission in 2028. The proposed architecture combines transformer-based sequence models with cross-modal attention mechanisms to exploit complementary biomarker signals. Anticipated outcomes include a validated predictive model, a curated multimodal benchmark dataset, two peer-reviewed publications, and a fully drafted R01 proposal.

Research Plan & Objectives

Neurodegenerative diseases impose a growing societal and economic burden. Current diagnostic criteria for Alzheimer’s and Parkinson’s rely predominantly on clinical symptom presentation, which emerges only after substantial irreversible neural loss has occurred. While individual biomarker modalities — structural MRI, amyloid-PET, CSF protein assays, and actigraphy-derived motor assessments — each carry predictive signal, no existing computational framework systematically integrates these heterogeneous data streams for early, pre-symptomatic risk assessment.

The central hypothesis is that a multimodal transformer architecture with learned cross-modal attention can identify subtle, interacting biomarker signatures preceding clinical diagnosis by 2–5 years, outperforming single-modality baselines and current clinical screening tools.

Specific Aims

- Aim 1. Data Integration and Harmonization.** Curate and harmonize three IRB-approved biomarker datasets from the Meridian University Hospital system: structural MRI ($n = 980$), CSF proteomic panels ($n = 720$), and 6-month continuous actigraphy recordings ($n = 410$). Develop standardized preprocessing pipelines with documented quality control metrics, producing a combined, anonymized multimodal dataset with $n \approx 1,200$ unique patients and longitudinal clinical outcome labels over 3–7 year follow-up.
- Aim 2. Model Architecture and Training.** Design a multimodal transformer architecture (NEUROFUSE) with modality-specific input encoders, cross-modal attention layers, and a shared latent representation space. Train under a survival analysis framework predicting time-to-diagnosis, with systematic hyperparameter optimization and five-fold cross-validation stratified by clinical site and disease subtype.
- Aim 3. Validation and Interpretability.** Evaluate predictive performance against gold-standard clinical diagnoses using concordance index, time-dependent AUC, and net reclassification improvement versus established clinical risk scores. Characterize feature importance via integrated gradients and attention weight analysis to identify which biomarker combinations drive the earliest, most reliable predictions.

Methodology

Data Sources. All data derive from the Meridian University Hospital Electronic Health Record-linked research registry (IRB#2025–M0784). T1-weighted structural MRI scans are processed through FreeSurfer v7.4 to extract regional volumetrics and cortical thickness. CSF proteomics were assayed using the Nimbus NeuroPlex 96-analyte panel on the Luminex xMAP platform. Actigraphy data were collected via Apex MotionTrack wrist-worn devices sampled at 50 Hz over 180-day continuous recordings, yielding gait, tremor, and sleep architecture features.

Model Architecture. NEUROFUSE employs three parallel encoders: a vision transformer with 3D patch embedding for imaging, a set transformer with learned per-analyte positional encodings for proteomic features, and a temporal convolutional network followed by a causal-masked transformer for actigraphy time series. Cross-modal fusion uses learnable latent query tokens that attend jointly to all three modality representations, producing a fixed-dimensional patient embedding fed into a Cox proportional hazards output head.

Evaluation. Performance is assessed via five-fold cross-validation stratified by clinical site and diagnostic category. Primary metrics: Harrell’s C-index and time-dependent AUC at 2-, 3-, and 5-year horizons. Ablation studies quantify per-modality contribution and cross-modal attention benefit. Baselines include univariate Cox models, random survival forests on concatenated features, and single-modality transformer variants.

Computational Resources. Training uses the Synapse Cloud GPU cluster (4× Lumina H100 GPUs) funded through the requested compute allocation. Code developed in Python 3.11 with PyTorch 2.x and pycox, released under Apache 2.0 upon completion.

Budget

Category	Description	Amount
Personnel	Graduate Research Assistant (50% FTE, 12 months)	\$28,000
Computing	Synapse Cloud GPU credits (8,500 GPU-hours)	\$8,000
Data Acquisition	Patient data extraction, de-identification, processing	\$5,000
Travel & Dissemination	Conference registration and travel (2 presentations)	\$3,750
Supplies & Materials	Software licenses, data storage, archival media	\$4,000
Total Requested		\$48,750

No cost-sharing is required. Funds administered through the Meridian University Office of Sponsored Programs.

Timeline

Phase	Months	Key Activities & Deliverables
1. Data Curation	1–3	IRB renewal; data extraction from EHR registry; FreeSurfer pipeline on imaging cohort; CSF QC; actigraphy feature extraction; harmonized multimodal dataset generation
2. Model Development	4–7	NEUROFUSE architecture implementation; encoder pre-training; cross-modal attention development; hyperparameter optimization; five-fold CV training runs; baseline implementation
3. Validation	8–10	Benchmarking against clinical risk scores; ablation studies; integrated gradient analysis; sensitivity analyses by demographic subgroup; preliminary results write-up
4. Dissemination	11–12	Manuscript preparation (<i>npj Digital Medicine</i> , <i>Alzheimer's & Dementia</i>); conference abstract submission; benchmark dataset release; NIH R01 proposal drafting

Expected Outcomes & Broader Impact

Anticipated Deliverables.

- A validated, open-source multimodal transformer model (NEUROFUSE) for early neurodegenerative risk prediction
- A curated, de-identified multimodal biomarker benchmark dataset released via the Meridian Data Repository
- Two first-author manuscripts submitted to high-impact journals
- A fully drafted NIH R01 proposal with compelling preliminary data and power analyses
- Conference presentations at ICMLM and the American Academy of Neurology Annual Meeting

This seed grant will position Dr. Chen and Dr. Webb to compete successfully for extramural funding from the National Institutes of Health (target: NINDS R01, February 2028). Beyond the immediate funding trajectory, the project will train one Ph.D. student in cutting-edge multimodal machine learning methods and strengthen Meridian University's research profile at the intersection of AI and clinical neuroscience. The open-source release of NEUROFUSE and its benchmark dataset will enable independent validation and accelerate progress toward clinically deployable early screening tools for neurodegenerative disease.

References

- [1] Chen, S., Patel, R., & Kowalski, T. (2025). Transformer-based survival analysis with multimodal clinical data. *Journal of Machine Learning in Medicine*, 8(2), 112–134.
- [2] Webb, M., Okafor, N., & Chen, S. (2024). Longitudinal CSF proteomic trajectories in preclinical Alzheimer's disease. *Neurology: Biomarker Discovery*, 41(3), 201–219.
- [3] Vaswani, A., et al. (2017). Attention is all you need. *Advances in Neural Information Processing Systems*, 30, 5998–6008.
- [4] Lee, C., Yoon, J., & van der Schaar, M. (2019). Dynamic-DeepHit: A deep learning approach for dynamic survival analysis. *IEEE Trans. Biomedical Engineering*, 67(1), 122–133.
- [5] Meridian University Hospital. (2025). EHR-linked research registry: Cohort profile and data dictionary (v3.2). *Meridian Clinical Data Archive*. IRB#2025–M0784.
- [6] Kvamme, H., Borgan, Ø., & Scheel, I. (2019). Time-to-event prediction with neural networks and Cox regression. *Journal of Machine Learning Research*, 20(129), 1–30.