NeuroHealth-Vision: A Volumetric Perception Module for Multimodal Clinical Reasoning

UCSC OSRE 2026 / Google Summer of Code Proposal

Project Mentor(s) Dr. Linsey Pang and Dr. Bin Dong
Contributor Antonio Tan

1. Introduction

The NeuroHealth project, led by Dr. Linsey Pang and Dr. Bin Dong, is a large-scale initiative aimed at bridging the critical gap between patients and the healthcare system. By developing an LLM-based clinical assistant, NeuroHealth seeks to transform triage support, appointment scheduling, and patient education through intelligent health inquiry and stronger patient engagement. At present, however, the NeuroHealth framework is still largely limited to NLP-based symptom checking. My project, NeuroHealth-Vision, serves as a foundational module within this broader ecosystem by extending its capabilities from a purely conversational agent to a multimodal diagnostic platform.

Despite the growing power of clinical LLMs, they remain effectively "blind" to the high-dimensional 3D imaging data that often drives patient outcomes. To move NeuroHealth toward a truly comprehensive assistant, it must be equipped with a perception engine capable of interpreting complex volumetric imaging results. Yet current 3D medical foundation models face a substantial specialization gap. For example, CT-CLIP [1] is trained on the CT-RATE dataset [2] of non-contrast chest CT volumes, while Merlin [3] focuses on abdominal imaging. Even specialized models such as Cardiac-CLIP [4] are typically limited to specific anatomical regions, and broader generalist models like CT-FM [5], despite being trained on more diverse CT data, still lack rigorous validation of their transfer ability on downstream tasks in specialized clinical domains.

A central challenge is that foundation models are rarely exposed to the unique physics, acquisition protocols, and contrast dynamics of specialty imaging studies. This project aims to build a robust evaluation framework for benchmarking state-of-the-art models on unseen specialty tasks for prognosis prediction. By investigating the transfer capabilities of models such as CT-FM and Cardiac-CLIP through zero-shot, few-shot, and fine-tuning settings, I will evaluate their performance on selected specialized protocols, including possibly multiphasic liver CT (arterial vs. venous washout kinetics), CT urography (excretory-phase urinary tract analysis), dual-energy CT (material decomposition), coronary CT angiography (high-resolution gated vascular imaging), or CT colonography (air-distended hollow-organ geometry).

By systematically identifying where these "generalist" models fail in "specialist" settings, this project will help select and refine the optimal perception engine for the NeuroHealth reasoning core. This would enable the central LLM to integrate imaging-based prognostic risk with patient text inquiries, offering a promising solution for complex clinical use cases and helping establish a new standard for multimodal health assistance.

2. Project Goals

The goal of NeuroHealth-Vision is to provide the "perception engine" for the broader NeuroHealth clinical assistant by developing a standardized evaluation and transfer-learning framework for 3D medical foundation models. This project aims to bridge the "Protocol Gap" in current frontier models by benchmarking their transferability to one or more underexplored specialty CT types through zero-shot, few-shot, and fine-tuning lenses. By the conclusion of the project, I expect to deliver a thorough evaluation of the transfer capabilities of current frontier medical foundation models, identifying exactly where "generalist" architectures fail when applied to high-stakes, specialized downstream tasks for prognosis prediction.

The expected deliverables include open-source data preprocessing pipelines for specialized 3D volumes (e.g., contrast-phase alignment and normalization), as well as transfer-training scripts specifically designed to adapt large foundational CT models to these specialty scenarios. Finally, I will provide optimized post-training model weights and performance benchmarks that demonstrate the effectiveness of these adapted "specialist" models. This work will equip the community with new open-source training pipelines for specialty CT types that remain largely underexplored in medical AI today.

Future work will focus on the deep integration of these imaging embeddings into the NeuroHealth LLM core, transforming the system from a conversational agent into a multimodal reasoning assistant capable of interpreting complex clinical imaging.

3. Implementation Plan

3.1 Project Methodology and Technical Plan

I will work closely with the mentors throughout the project through regular progress updates, milestone reviews, and technical discussion of model adaptation strategies, dataset feasibility, and evaluation design. The project will proceed in the following stages.

Step 1: Literature Review & Architectural Analysis
I will conduct a comprehensive audit of existing 3D medical foundation models to identify structural bottlenecks for specialty transfer. Table 1 provides a non-exhaustive set of representative examples that will guide this analysis.

Step 2: Downstream Task Selection & Data Collection
I will select the most effective downstream specialty transfer tasks for testing and apply for access to relevant open-source and managed-access datasets. To ensure clinical compliance and feasibility, I will consider the examples highlighted in Table 2 (and others as identified) in making the final selection.

Step 3: Data Preprocessing
For each selected task and dataset, I will develop automated preprocessing pipelines to normalize the 3D volumetric data. This includes standardized voxel resampling, Hounsfield Unit (HU) windowing, and spatial alignment to ensure the data is compatible with the input requirements of the selected foundation models.

Step 4: Model Adaptation & Implementation
I will design and implement architectural adaptations to the selected foundation models so they can process specialized clinical inputs. This involves modifying the model's "perception" layers to handle unique geometric

Category	**Models / Focus**
Generalist models	CT-FM [5] and SPECTRE [6], used for baseline volumetric feature extraction
Chest-centric / thoracic models	CT-CLIP [1], M3FM [7], and TANGERINE [8]
Abdominal models	Merlin [3]
Cardiac-specific models	CORA [9] and Cardiac-CLIP [4]
Adaptation strategy	I will examine model weights and attention mechanisms to brainstorm architectural modifications required to handle altered input formats in specialty CT types.

Table 1: Non-exhaustive examples of representative 3D medical foundation models and the adaptation focus guiding this project.

Specialty Task	**Datasets**	**Links**
CCTA (Coronary)	SCOT-HEART; PROMISE; CONFIRM; UK Biobank; CRESCENT; CONVENE	SCOT-HEART [10]; PROMISE [11]; CONFIRM [12]; UK Biobank [13]; CRESCENT [14]; CONVENE [15]
CT Urography (Excretory)	CPTAC-CCRCC Cohort; KiTS23	CPTAC-CCRCC [16]; KiTS23 [17]
Multiphasic Liver CT	MCT-LTDiag; TCGA-LIHC; LiTS; HCC-TACE-Seg; TCIA-CRLM	MCT-LTDiag [18]; TCGA-LIHC [19]; LiTS [20]; HCC-TACE-Seg [21]; TCIA-CRLM [22]

Table 2: Non-exhaustive examples of candidate specialty CT tasks and datasets considered for downstream benchmarking.

structures or multi-sequence temporal data that may differ from the models' original training distributions.

Step 5: Model Training & Evaluation
The adapted perception engine will undergo a rigorous training and benchmarking phase:

- **Fine-tuning:** Applying parameter-efficient fine-tuning (PEFT/LoRA) on the selected specialty datasets.
- **Performance comparison:** Quantifying the "Protocol Gap" by comparing fine-tuned performance against zero-shot and few-shot baselines.

Step 6: Comparative Analysis (If Time Permitted)
To validate the clinical utility of the foundation-model approach, I will implement and compare results against non-ML or statistical baselines (e.g., traditional clinical risk scores) to quantify the added value of deep-learning-based prognostic prediction.

4. Project Timeline

The tentative schedule is summarized in Table 3 below.

Dates	Proposed Phase	Key Activities and Deliverables
05/01 – 05/24 (3.5 w)	Community Bonding & Literature Review	Intensive audit of volumetric foundation models, including CT-FM and SPECTRE; finalize adaptation strategy and development environment.
05/25 – 06/07 (2 w)	Task & Data Finalization	Finalize downstream task selection and secure access to datasets.
06/08 – 06/21 (2 w)	Preprocessing Development	Build automated pipelines for 3D resampling, HU windowing, and spatial alignment for specialty volumes.
06/22 – 07/05 (2 w)	Model Adaptation & Implementation	Implement architectural modifications for specialty downstream tasks.
07/06 – 07/12 (1 w)	Midterm Evaluation	Submit adaptation source code and preliminary zero-shot / few-shot results for midterm mentor review.
07/13 – 07/26 (2 w)	Model Training & PEFT	Execute fine-tuning using PEFT/LoRA techniques and related methods to minimize the "Protocol Gap" in specialized domains.
07/27 – 08/01 (1 w)	Final Evaluation & Submission	Quantify model performance through rigorous evaluation; finalize technical documentation and the "Perception Engine" framework for final submission.
08/02 – 08/24 (3 w)	Buffer Period	Buffer period.

Table 3: Tentative project timeline and planned deliverables across the OSRE period.

Commitment: ~20 hours/week

Conflict: I ensure that I will have enough time to meet the GSoC expectations.