

أكاديمية كاوست
KAUST ACADEMY



جامعة الملك عبدالله
للعلوم والتقنية
King Abdullah University of
Science and Technology

**Project Title:** Validation of AI-Generated Cerebrospinal Fluid Flow from Resting-State fMRI Data: The Effect of Age

**Mentor:** Dr. Jean Chen

**Student:** Mohammad Niazy

**Final report**

During the initial weeks, I focused on establishing a strong foundational understanding of the PINN-based model developed in the reference study, which estimates cerebrospinal fluid (CSF) velocity from resting-state fMRI signals. This week's primary objective was to deeply dissect the paper's methodology, trace its core assumptions, and build a well-structured visual representation of the pipeline in the form of a slide deck. This deck will later evolve into the final poster for project presentation.

The week began with an intensive reading of the bioRxiv preprint that our work is based on, focusing specifically on the theory and methods sections to understand how the signal equation (Eq. 1)

$$M_T(n,\,\Delta t) = M_{eq}\sin\theta e^{\frac{-TE}{T_2}} \prod_{k=1}^{n-1} e^{-\frac{\Delta t_k}{T_1}} \left[ \sum_{m=0}^{n-1} \left( \cos\theta^m \frac{1-e^{-\frac{\Delta t_{n-m-1}}{T_1}}}{\prod_{k=1}^{n-m-1} e^{-\frac{\Delta t_k}{T_1}}} \right) + \cos\theta^{n-1} \right] \qquad \text{Eqn. 1}$$

models the relationship between CSF velocity and observed fMRI brightness. The paper assumes that inflow effects, increased brightness caused by fresh protons entering the imaging volume, can be exploited to infer CSF movement. To capture this effect, the authors construct a forward model using Bloch simulations and train a convolutional neural network (CNN) to perform the inverse mapping: recovering velocity curves v(t) from brightness signals M(t). My current role involves adapting this model to our own dataset, validating its outputs, and identifying where assumptions in the original paper might fail to generalize.

I created a full-stage summary of the pipeline in my slide deck, breaking it into six key stages: (1) extracting CSF voxels from fMRI using anatomical segmentation, (2) preprocessing signals, (3) modeling inflow physics with assumed parameters (T1, T2, TR, flip angle, etc.), (4) using the CNN to learn the inverse mapping from brightness to velocity, (5) validating with PC-MRI or age-dependent trends, and (6) interpreting results. Each of these steps includes major assumptions that I highlighted clearly in the pipeline and in a dedicated "Assumptions" slide, which was my main work in week 1. For instance, the assumption of plug flow ignores spatial velocity gradients, and the initialization of v(t) as small random values between -0.1 and 0.1 cm/s could constrain learning dynamics if applied to real data with different flow patterns.

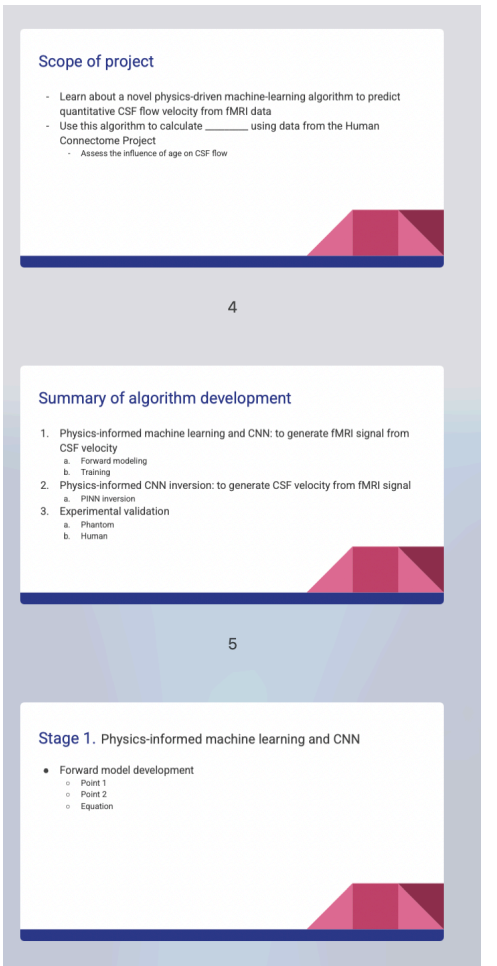
A major challenge this week was navigating and interpreting assumptions that are mathematically simple in the paper but complex in implication. For example, I examined the assumption that TR and flip angle are constant throughout the fMRI acquisition, which simplifies the simulation of M(t), but doesn't always reflect scanner variability. Another point is the variable RF exposure due to bidirectional CSF flow, where protons may exit and re-enter the imaging slab, gaining signal strength due to longer relaxation. These types of assumptions, if mismatched with our acquisition settings or subject physiology, could cause model outputs to diverge from reality.

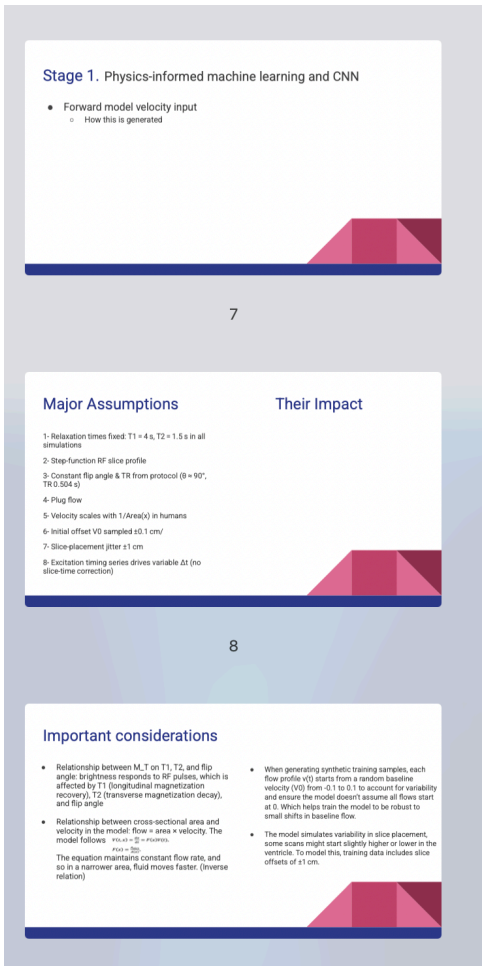
In addition to theoretical analysis, I reviewed the available dataset paths provided by our lab, including Freesurfer-based anatomical segmentation of the fourth ventricle and multiband EPI fMRI data. I outlined which files are needed, what preprocessing is expected like no slice-time correction, baseline demeaning, and confirmed that our dataset is likely appropriate in both size and age diversity to replicate the original study's findings on age-related CSF decline. I also noted the need to handle variable slice positions and temporal offsets, as the original model assumes a possible ±1 cm spatial variation at the base slice, a major assumption highlighted for future testing.

Next week, I aim to begin scripting the preprocessing pipeline and explore the model's GitHub codebase to understand how inputs are structured, how loss is computed, and how outputs are interpreted. For now, we already downloaded the "tofmodel" (the AI model named from the original paper) into the lab's servers. A key upcoming goal is to run the model on a test sample to observe whether it functions as expected on unfamiliar data. Long term, the goal remains to validate this model's use as a cost-effective tool for estimating CSF flow, potentially opening a path to early biomarkers for Alzheimer's disease.

Tools and resources used so far include the original paper, its code repository, fMRI and anatomical datasets, and slide design platforms for building the visual pipeline. The primary challenges have been domain unfamiliarity and assumptions buried deep in the methodology, but I've addressed these by breaking the paper into stages and building conceptual clarity through visual summaries and supervisor feedback. Below are a few photos of the slide to provide a general idea of what I am going for.



# Project objectives and goals (refined, building on prior plan)

Our priorities remain: (i) fully map the paper's pipeline and assumptions, (ii) prepare inputs from our dataset (HCP-Aging), (iii) document where assumptions transfer or need retuning, and (iv) set up validation vs. age. Your last report already framed this end-to-end plan (six stages from masking → preprocessing → forward model → CNN inversion → validation → interpretation) and why assumption handling matters.

Since then, we've sharpened the **operational objectives**:

- Convert "pipeline storyboard" into **slide blocks** with inputs/outputs per stage (now present in your deck: dataset & collection → preprocessing & mask → Bloch forward model → synthetic training → PINN inversion → validation).

- Define **application plan** to HCP-Aging: data sources, preprocessing recipe, and GLM for age effect (now captured on your slides).

# Current progress and milestones achieved

## 1) Pipeline documentation elevated to presentation-ready artifacts

- The pipeline is now reflected as a **succinct stage table** (1→5) with short, precise bullets that you can show publicly (and keep detailed notes hidden). It mirrors the stages in your deck ("Dataset and collection," "Pre-processing & Mask," "Bloch Forward model," "Training," "PINN inversion") and provides common language for the team.

## 2) Data definition and input construction finalized

- **Input data summary** slide now states: HCP-Aging provides anatomical masks (FreeSurfer) and resting-state fMRI; the 4th ventricle is the ROI; averaging ventricle voxels over time yields the **brightness time series M(t)** that feeds the model. This locks the definition of what we extract per subject.

- The **fMRI preparation** steps (motion-correct → bias-field → FreeSurfer mask → erode 1 voxel → time-series extraction → demean) are enumerated clearly on a dedicated slide.

## 3) Pre-processing script specification (design complete)

- We documented and prototyped a minimal script that: loads the REST1 run and `aparc+aseg.mgz`; isolates 4th-ventricle voxels; **erodes** by one voxel to reduce partial volume; averages across voxels each TR; **demeans** the result; and **saves** `M(t)` to CSV (slide: “Pre-processing Script”). This is the exact I/O contract the model expects.

### 4) Assumptions register consolidated into a checkable table

- Your deck now contains a **Major Assumptions** slide: fixed T1/T2, step slice profile, protocol TR/flip, plug flow (phantom), area-velocity coupling, baseline $V_0$ sampling, ±1 cm slice-placement jitter, and keeping exact RF timing history (no slice-time correction). We use this as a living checklist when transferring to HCP-Aging.

- The “Important considerations” notes on how **T1/T2/flip** affect M_T; **area–velocity** coupling; **baseline $V_0$** robustness; and **slice placement** variability are now integrated into your deck, making these points easy to reference in discussion.

### 5) Validation narrative prepared (phantom + human)

- The deck includes concise validation bullets (phantom straight-tube ranges; human paced breathing at 0.167 Hz; errors and expected age slope). This gives you a ready-made section for the poster and for planning our own checks later.

## Methodology and approach (what we will do, now operationalized)

### A) Data → M(t): subject-level input assembly

- **Why M(t):** The model doesn’t read images; it reads a 1-D brightness signal from the CSF compartment. We take the 4th-ventricle mask from FreeSurfer and average voxel values over time; then **demean** to center the series (exactly as your slide describes).

- **Pre-processing stance:** we do not apply slice-time correction because the forward model explicitly accounts for slice-specific RF exposure history; this is recorded in your preprocessing bullets and “Important considerations.”

### B) Forward model & synthetic design (kept conceptual for now)

- The deck captures that a Bloch model (Eq. 1) converts **v(t)** to **M(t)**; training uses **45 k synthetic samples** with random $V_0$, amplitude/frequency ranges, and slice offset ±1 cm.

This is our reference when we decide if/when to expand physiological ranges for HCP-Aging.

### C) CNN inversion & physics-informed loss (documented, not executed)

- Your “Stage 2” slide records the network structure (3 conv → 3 transposed conv → FC; ReLU+BN), the **loss on simulated brightness** (not on v(t) directly), and the test-set MSE figure reported in the original pipeline—these are the specs we will use.

## Preliminary findings (non-numerical, pre-execution)

- We now have a **repeatable recipe** to produce M(t)M(t)M(t) per subject, including mask erosion and demeaning, which directly matches the pipeline’s expected input.

- The **assumption deltas** most likely to affect transfer have been identified and elevated into the slides: protocol differences (TR/TE/flip), physiological bands (slightly broader cardiac), baseline drift (potentially > ±0.1 cm/s in older adults), and slice-offset variability (keeping ±1 cm supports robustness).

## Timeline and schedule (next week focus)

- **Early week:** Batch preprocessing to **export M(t)M(t)M(t)** for an initial subset; add QC columns (motion, voxel count, SNR) to a tracking sheet following your “fMRI data preparation” plan.

- **Mid-week:** Finalize assumption table with any measured metadata deltas; lock poster figure placeholders (you already drafted the application & preprocessing panels).

- **End-week:** (After this report’s scope) generate synthetic ranges if we need modest retuning; place figure placeholders for velocity waveform and age-trend plot (the “Expected outcomes” slide already sets the narrative).

## Challenges and obstacles encountered

- **Domain transfer risk:** Differences between the paper’s simulated acquisition and HCP-Aging (TR/flip) can rescale M(t)M(t)M(t); we addressed this by codifying those differences in the assumptions table and in the preprocessing stance.

- **Partial volume contamination:** The ventricle is small; your slides explicitly include **mask erosion** to minimize tissue mixing.

- **Time constraints:** With limited time before the poster, we prioritized documentation, inputs, and analysis plan so that any subsequent runs can drop their numbers into an already-built narrative.

## Next steps and future work

- **Immediate:** Execute the preprocessing script to create subject-level M(t)M(t)M(t); finalize the assumption-vs-dataset table with any measurements (e.g., motion distributions, slice-offset proxy).

- **Near-term (post-report):** Prepare synthetic range adjustments (if needed) and placeholders for the poster figures: pipeline schematic, example M(t)M(t)M(t), assumptions table, and an age-effect panel.

## Resources and tools used (to date)

- **Primary artifacts:** the previous reports and the evolving slide deck (pipeline, assumptions, application plan, validation summaries).

- **Software/standards (design stage):** FreeSurfer parcellations (`aparc+aseg.mgz`) for ventricle masks; FSL for motion correction; NiBabel/NumPy for I/O and time-series extraction—matching the steps listed on the “fMRI data preparation” and “Pre-processing Script” slides.