LEARNING AMYLOID PATHOLOGY PROGRESSION FROM LONGITUDINAL PIB-PET IMAGES IN PRECLINICAL ALZHEIMER'S DISEASE

Wei Hao[1] *Nicholas M. Vogt*[1] *Zihang Meng*[1] *Seong Jae Hwang*[2]
Rebecca L. Koscik[1] *Sterling C. Johnson*[1,3] *Barbara B. Bendlin*[1] *Vikas Singh*[1]

[1]University of Wisconsin-Madison [2]University of Pittsburgh
[3]William S. Middleton Memorial Veterans Hospital

ABSTRACT

Amyloid accumulation is acknowledged to be a primary pathological event in Alzheimer's disease (AD). The literature suggests that propagation of amyloid occurs along neural pathways as a function of the disease process (prion-like transmission), but the pattern of spread in the preclinical stages of AD is still poorly understood. Previous studies have used diffusion processes to capture amyloid pathology propagation using various strategies and shown how future time-points can be predicted at the group level using a population-level structural connectivity template. But connectivity could be different between distinct subjects, and the current literature is unable to provide estimates of individual-level pathology propagation. We use a trainable network diffusion model that infers the propagation dynamics of amyloid pathology, conditioned on an individual-level connectivity network. We analyze longitudinal amyloid pathology burden in 16 gray matter (GM) regions known to be affected by AD, measured using Pittsburgh Compound B (PiB) positron emission tomography at 3 different time points for each subject. Experiments show that our model outperforms inference based on group-level trends for predicting future time points data (using individual-level connectivity networks). For group-level analysis, we find parameter differences (via permutation testing) between the models for APOE positive and APOE negative subjects.

Index Terms— Alzheimer's disease, Network diffusion, Differential equations, PiB PET image, MRI connectivity

1. INTRODUCTION

Understanding the cause/mechanism of Alzheimer's disease (AD) progression is critical for early treatment and prevention of AD. Recent neuropathological findings suggest that the progressive pattern of AD could be characterized by the transmission of disease pathologies along neuronal pathways in the brain networks [1, 2, 3]. Specifically, as the prion-like disease agents such as misfolded beta amyloid and/or tau protein starts accumulating in the gray matter, the disease pathologies may *propagate* along the fiber pathways [4].

Motivated in part by these findings, several strategies have been developed to model such propagation patterns using brain imaging data. For instance, the Non-linear neural mass model (NMM) takes the neural activity in localized populations (mini-columns) into account to match with functional connectivity of the brain [5]. However, it is a second order statespace differential equation form, so no closed-form solution exists for these formulations and the model construction involves computationally expensive large-scale simulations among the brain signals acquired at thousands of time points. On the other hand, the Linear network model with a closed-formed solution was formulated as a discretized multivariate auto-regressive linear system for efficiently estimating the brain network signal at different time points [6]. Later, the Laplacian of the brain network with physically interpretable parameters was used to model the propagation behavior as a heat diffusion process [7]. Recently, a network diffusion model also used a group level Laplacian that captures the group level propagation information of the full cohort to model the propagation of amyloid, a type of AD pathology measurable in PiB PET images [8]. However, the group level Laplacian derived from the model cannot easily be used to characterize the individual level propagation patterns based on both the baseline imaging scan and the individual's brain connectivity – propagation routes are based on the fiber connectivity of the full group.

A model which can characterize both the *group level trend* originating from the underlying disease progression in the full cohort as well as permit the *individual level variability* is needed. This may enable understanding the pathological progression of AD not only within the observed timeline but also predict the future disease trends of individuals at risk of AD – informed by both the baseline PiB PET imaging scans and the subject's (structural or functional) connectivity.

In this paper, we propose a machine learning based model which (1) estimates how to map (via a learned function) baseline PiB PET imaging based measurements to measurements acquired at a follow-up visit (or even a third visit), appropriately informed by the individual's connectivity information; (2) identifies group level progression similar to other

important studies [7, 8] but can offer individual level prediction, useful for evaluating risk for Alzheimer's disease. On the technical side, we formulate a simple ordinary differential equation that integrates a data-driven heat diffusion equation and a deep neural network model. We then evaluate the propagation prediction results on a neuroimaging dataset and show (promising although preliminary) group difference results using a genetic risk factor of AD as a dichotomous variable.

2. METHOD

2.1. MODEL CONSTRUCTION

We construct a brain network roughly similar to [7] where each connectivity network is denoted as a graph $G = \{V, E\}$ of vertices V where node $v_i \in V$ represents the i-th cortical or subcortical gray matter region, and edges E where edge $e_{ij} \in E$ denotes the connectivity between v_i and v_j. Each e_{ij} has a weight c_{ij} denoting the connectivity strength (described in Section 3.1). Consider a network which only contains an affected region R_1 that is connected with an unaffected region R_2. Let us denote the level of a measurement for the i-th region as ϕ_i. Then, the propagation from R_1 to R_2 will be influenced by both the levels ϕ_1 in R_1 and ϕ_2 in R_2 and the inter-region connection strength, c_{12}. Likewise, we may write down the propagation between R_2 to R_1 assuming a bidirectional connection. During a period of time δt, the pathology load in R_2 may increase by $k \cdot (\phi_1 - \phi_2) \cdot \delta t$ where k is a diffusion constant denoting the speed of diffusion. As the limit $\delta t \to 0$, we will have a first-order differential equation:

$$\frac{d\Phi_2}{dt} = k \cdot (\phi_1 - \phi_2) \tag{1}$$

Using a Laplacian, we get the *heat diffusion* equation on a network:

$$\frac{d\Phi(t)}{dt} = -k \cdot L \cdot \Phi(t) \tag{2}$$

At a time point t, $\Phi(t)$ is a vector with the pathology burden measurements for all regions of interest in an individual's PET scan. Here, L is the graph Laplacian constructed with the individual subject's connectivity matrix (see Section 3.1). To obtain a propagation pattern that is consistent across the full training dataset, our model learns a function modulating the diffusion process whose start and end points are known when we have longitudinal image measurements of the pathology:

$$\frac{d\Phi(t)}{dt} = f(L) \cdot \Phi(t) \tag{3}$$

Here, f is the function mapping an subject's L, denoted by L_s, to $f(L_s)$ through an element-wise product with $\frac{(M+M^T)}{2}$ where M is a matrix of trainable parameters of the same size as the Laplacian and M^T is its transpose:

$$f(L) = \frac{L \times (M + M^T)}{2} \tag{4}$$

This product adjusts the individual Laplacian while keeping it symmetric. Notice that the constant "$-k$" is subsumed within $\frac{(M+M^T)}{2}$. Thus, given $\phi(t_1)_s$ and L_s for a subject s at time point t_1, $\phi(t_2)_s$ at time point t_2 is a solution to an initial value problem by solving this ODE by integration through a differential equation solver [9]:

$$\Phi(t_2) = \Phi(t_1) + \int_{t_1}^{t_2} f(L) \cdot \Phi(t) dt \tag{5}$$

Finally, our model will be estimated for the entire group of subjects to derive a group level M that summarizes how connectivity informs the propagation trend over the entire training data, where the the differential equation denotes the trend.

Training the Model. Given $\phi(t_1)_s$, L_s and a *integration time* for a subject s at t_1, one can solve the differential equation using standard schemes. One issue is that such a mechanism does not easily allow us to backpropagate through the differential equation solver and update the parameters of the function we want to estimate. But various strategies have suggested solution schemes, e.g., see [9], which yields a differentiable module to produce the predicted $\hat{\phi}(t_k)_s$ at multiple time points t_ks. We trained the model to minimize the ℓ_2 norm loss between the predicted $\hat{\phi}(t_k)_s$ and the corresponding real $\phi(t_k)_s$. Backpropagation on the loss is performed based on the average gradient descent for all training subjects. The backward algorithm is described in [9].

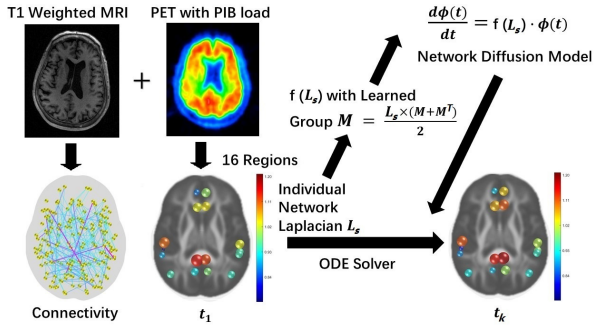

Fig. 1: Future time point prediction via a trained model

3. EXPERIMENTS

3.1. Dataset

Neuroimaging data used for these experiments was obtained from an ongoing large-scale study focused on understanding preclinical Alzheimer's disease (AD). The University of Wisconsin Health Sciences Institutional Review Board approved all study procedures and all participants provided written informed consent. Participants who had at least three time points of PiB PET and T1-weighted MRI ($n = 112$, baseline age 67.6 ± 6.0 years, 68.6% female) were included. PiB DVR

Algorithm 1 Training Process

Input: Model (3); $\phi(t_k)_s$ (DVR data for subject s at time point t_k); T (artificial integration interval for solving an ODE); L_s (Laplacian for subject s);
Loss (the ℓ_2 norm loss between prediction and real data); `odesolver` from [9]).

Loop:
if Loss does not converge **then**
    Initialize **Loss** to zero for each iteration;
    For each subject s in training set, compute prediction $\phi(\hat{t_k})_s$ at time point t_k:
    $\phi(\hat{t_k})_s = \texttt{odesolve}(\textbf{Model}, \phi(t_k)_s, T, L_s)$;
    Total Loss = $\sum_s \|\phi(\hat{t_k})_s, \phi(t_k)_s\|_2$;
    Perform average gradient descent on *Total Loss* ;
    Update group parameter M;
close;

values were extracted from 16 regions (8 bilateral regions) known to accumulate amyloid pathology [10] and used as regions of interest in the analysis. T1-weighted images were processed using Computational Anatomy Toolbox (CAT12) and 'graynet' [11] in order to construct single-subject level connectivity matrices, which provided the strength of connections between the 16 regions. We used T1-weighted images to keep the initial processing simple – connectivity based on a tractography procedure on diffusion weighted images (DWI) will likely provide better estimates of connectivity but will need a more involved processing pipeline. Using T1-weighted images allows directly using a common atlas to extract PIB values as well as calculating connectivity using `graynet` for the same regions without any additional co-registration or tractography. The single-subject level Laplacian is obtained using graynet estimated connectivity:

$$L(i,j) = \begin{cases} -c_{ij} & \text{if } i \neq j \\ \sum_{(i,j');e_{ij'} \in E} c_{ij'} & \text{otherwise} \end{cases} \quad (6)$$

where c_{ij} is the connectivity strength between region i and j, and $L(i,j)$ is the entry in row i/column j of L for the subject

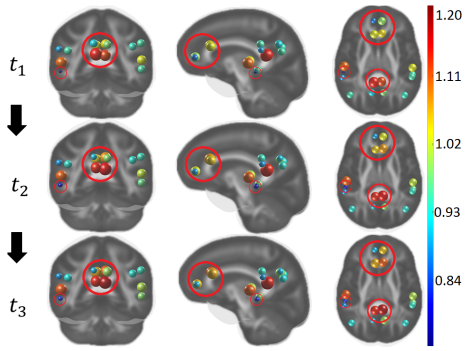

Fig. 2: Amyloid propagation over time ($t_1 \rightarrow t_2 \rightarrow t_3$)

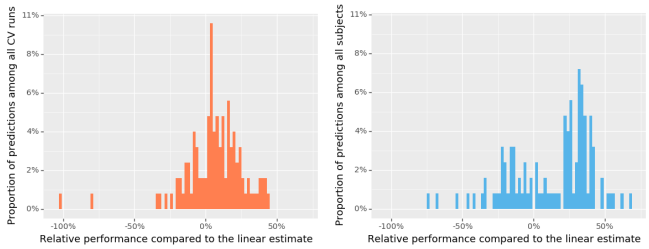

Fig. 3: Relative performance w.r.t. the linear estimate of average group change. A point on the x-axis shows how much better (in percentage) our method is, the y-axis shows the frequency. (Left) predicting $\phi(t_2)$, (Right) predicting $\phi(t_3)$.

s. Thus, for each subject s, we have a L_s (size 16×16) for the baseline time point t_1 and three Φ_s (vectors of size 1×16) for three ordered visits. Finally, for group analysis, we use a well-known AD-related genetic risk factor called APOE [12] as a dichotomous variable, available for each participant in our dataset. We designate participants as APOE+ with a high risk of AD if they had at least one APOE $\epsilon 4$ allele, and APOE- with a low risk of AD if they had no APOE $\epsilon 4$ alleles.

3.2. Experimental Setup and Results

We test our model in two scenarios, each with a different dataset split. For the first experiment, the model is trained using $\phi(t_1)$, $\phi(t_2)$, and L. The training data includes 90% of the subjects. The model is evaluated on the remaining 10% subjects to estimate their $\phi(t_2)$ from $\phi(t_1)$ (test). We perform 10-fold cross-validation. For the second setting, the model is trained using $\phi(t_1)$, $\phi(t_2)$, and L of all subjects (i.e., scans at t_3 are held out). We estimate the accuracy of estimated amyloid loads at t_3 (i.e., $\phi(t_3)$) given only the baseline $\phi(t_1)$.

Comparison with inference based on group trends. Without knowing the actual PiB DVR values at future time points, a simple evaluation for future PiB DVR estimates obtained from our model is to compare them to those obtained from using a simple group-level trend (a linear estimate), i.e., the predicted PiB DVR at time t_k is:

$$\hat{\phi}'(t_k)_s = \phi(t_1)_s + \frac{\sum_{s'=1, s' \neq s}^{n} (\phi(t_k)_{s'} - \phi(t_1)_{s'})}{n-1} \quad (7)$$

where n is the total number of subjects.

We compare errors in predictions from our model ($\hat{\phi}$) versus those made by the linear estimate ($\hat{\phi}'$). The result is plotted as a histogram in Fig. 3. We found that our model consistently makes better predictions where 70.5% and 67.0% are better than the linear estimate predictions when predicting $\phi(\hat{t_2})$ and $\phi(\hat{t_3})$ using $\phi(t_1)$ respectively. We also calculated how often (in percentage) relative to the group-level prediction when our model is better (positive value) or worse

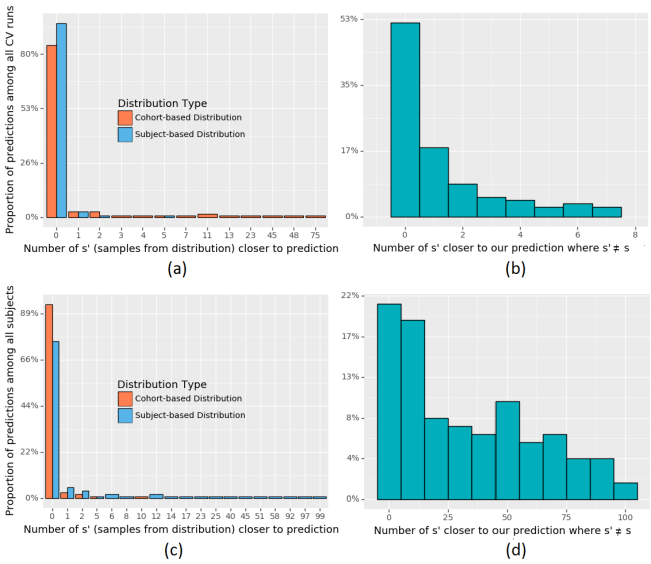

Fig. 4: Evaluating if the model's prediction is closest to the amyloid measurements of the correct subject. (Top) predicting $\phi(t_2)$, (Bottom) predicting $\phi(t_3)$.

(negative value) as,

$$\frac{1}{n}\sum_{s=1}^{n}\frac{\|\phi(t_k)_s - \hat{\phi}'(t_k)_s\|_2 - \|\phi(t_k)_s - \hat{\phi}(t_k)_s\|_2}{\|\phi(t_k)_s - \hat{\phi}'(t_k)_s\|_2} \quad (8)$$

We found that our model captures most of the propagation patterns and performs better when we seek to predict DVR measures at t_3 when only the baseline scans of each subject are available (average precision gain of 30.1% and 16.0% for predicting DVR measures at t_3 from t_1 and t_2 from t_1 respectively).

How close is our prediction to the real data? Among all test subjects, the ℓ_2-norm is calculated between our prediction $\hat{\phi}(t_k)_s$ and real DVR vector $\phi(t_k)_{s'}$ for each test subject s'. The ideal case is that even if we are slightly off in terms of absolute error in our predictions, nonetheless, this prediction $\hat{\phi}(t_k)_s$ is closer to $\phi(t_k)_s$ relative to other subjects' $\phi(t_k)_{s'}$ ($s' \neq s$). The results are shown in Fig. 4. The x axis denotes how many other subjects $s' \neq s$ lie between our prediction of measurements for a subject and the real ground truth measurements at that time point for that specific subject s. In Fig. 4 (b) and (d), we include only the subjects in our test set for this evaluation (for predicting $\phi(t_2)$ and $\phi(t_3)$), and find that the model performs well. A simulation experiment is described below.

Model prediction versus simulated samples from cohort/individual distribution. In Fig. 4 (a) and (c), we use the cohort distribution or the individual's measurement distribution (orange and blue bars) to artificially "simulate" test subjects, again for predicting $\phi(t_2)$ and $\phi(t_3)$. In both cases, we find that the model performs well. Among all test subjects, the ℓ_2 norm is calculated between our prediction $\hat{\phi}(t_k)_s$ and the real DVR vector $\phi(t_k)_s$ along with 99 randomly generated DVR vectors at that time point t_k of that subject (see Fig. 4 (a) and (c)). Each random DVR vector is generated from the distribution with the mean and standard deviation of the real $\phi(t_k)_s$. We expect that our prediction $\hat{\phi}(t_k)_s$ is the closest to the real DVR vector $\phi(t_k)_s$ rather than most simulated samples. Results are shown in Fig. 4(a) and (c). We see that 95% of our predictions for a subject is the closest to the real DVR value of *that* subject. Next, we assume that the DVR in each region i of a randomly generated DVR vector at t_k for a subject has the distribution with the mean and standard deviation of the real *region-wise* measurements at t_k in the test set such that the DVR data in region i itself, i.e., $\phi_i(t_k)$, is generated across all $\phi_i(t_k)_s$ for every subject s in the test cohort. Overall, the model performs well and the results are promising.

Group difference analysis: APOE+ versus APOE-. A group analysis is performed between subjects in the APOE positive group ($n = 45$) and APOE negative group ($n = 67$). We split the dataset and train two models, one each for the APOE positive and negative groups, separately. The ℓ_2 norm between the estimated parameters of the function, M_1 and M_2 acquired from these two models is computed: this is the statistical summary for the correct group. Our null hypothesis H_0 is that there is no group level difference between the parameters within these two groups. To test the hypothesis, we also split the dataset randomly but keep the sizes of two groups and calculate the distance between Ms from the two models within a permutation testing setup. Permutation testing is performed with 10000 draws which gives the null distribution of the summary statistic: difference between the Ms. The p-value is 0.00467 which is smaller than the significance threshold $\alpha = 0.05$ (area plot in red in Fig. 4(c)). The null hypothesis is rejected: there is indeed a group difference in terms of the parameters learned by our model to characterize amyloid progression pattern in APOE+ and APOE- groups.

4. CONCLUSIONS

Recent studies have shown that the early amyloid accumulation patterns may be a strong indicator of AD-related cognitive decline [13, 14]. Our formulation gives a mechanism to model amyloid pathology propagation using longitudinal PiB PET imaging data as well as brain connectivity. Our model could also serve as a way to estimate the early amyloid accumulation trajectory in the unobserved past (e.g., $t < 1$) by "reversing" the direction of the process (sign change of $f(L)$), and estimating the start time of amyloid accumulation.

Acknowledgments. This project was supported in part by RF1 AG059312, R01 AG037639, R01 AG027161, R01 AG021155, P50 AG033514, UW CPCP (U54AI117924) and NSF CAREER award RI 1252725.

5. REFERENCES

[1] Nicolas Villain, Batrice Desgranges, Fausto Viader, Vincent De La Sayette, Florence Mezenge, Brigitte Landeau, Jean-Claude Baron, Francis Eustache, and Gal Chetelat, "Relationships between hippocampal atrophy, white matter disruption, and gray matter hypometabolism in alzheimers disease," *Journal of Neuroscience*, vol. 28, no. 24, pp. 61746181, Nov 2008.

[2] Beth Kuczynski, Elizabeth Targan, Cindee Madison, Michael Weiner, Yu Zhang, Bruce Reed, Helena C. Chui, and William Jagust, "White matter integrity and cortical metabolic associations in aging and dementia," *Alzheimers & Dementia*, vol. 6, no. 1, pp. 5462, 2010.

[3] Chun-Yi Lo, Pei-Ning Wang, Kun-Hsien Chou, Jinhui Wang, Yong He, and Ching-Po Lin, "Diffusion tensor tractography reveals abnormal topological organization in structural cortical networks in alzheimers disease," *Journal of Neuroscience*, vol. 30, no. 50, pp. 1687616885, 2010.

[4] Bill Seeley, "Neurodegenerative diseases target large-scale human brain networks," *Alzheimers & Dementia*, vol. 6, no. 4, 2010.

[5] Christopher J. Honey, Olaf Sporns, leila Cammoun, Xavier Gigandet, Jean-Philippe. Thiran, Reto Meuli, and Patric Hagmann, "Predicting human resting-state functional connectivity from structural connectivity," *Proceedings of the National Academy of Sciences*, vol. 106, no. 6, pp. 20352040, Feb 2009.

[6] Roberto F. Galn, "On how network architecture determines the dominant patterns of spontaneous neural activity," *PLoS ONE*, vol. 3, no. 5, 2008.

[7] Ashish Raj, Amy Kuceyeski, and Michael Weiner, "A network diffusion model of disease progression in dementia," *Neuron*, vol. 73, no. 6, pp. 12041215, 2012.

[8] Seong Jae Hwang, Sathya Ravi, Nagesh Adluru, Barbara B. Bendlin, Sterling C. Johnson, and Vikas Singh, "Data-driven propagation modeling of pet-derived alzheimers pathology in a preclinical cohort," *Alzheimers & Dementia*, vol. 14, no. 7, 2018.

[9] Ricky T. Q. Chen, Yulia Rubanova, Jesse Bettencourt, and David Duvenaud, "Neural ordinary differential equations," *Advances in Neural Information Processing Systems*, 2018.

[10] Sterling C Johnson, Bradley T Christian, Ozioma C Okonkwo, Jennifer M Oh, Sandra Harding, Guofan Xu, Ansel T Hillmer, Dustin W Wooten, Dhanabalan Murali, Todd E Barnhart, et al., "Amyloid burden and neural function in people at risk for alzheimer's disease," *Neurobiology of aging*, vol. 35, no. 3, pp. 576–584, 2014.

[11] Pradeep Reddy Raamana, Michael W Weiner, Lei Wang, Mirza Faisal Beg, Alzheimer's Disease Neuroimaging Initiative, et al., "Thickness network features for prognostic applications in dementia," *Neurobiology of aging*, vol. 36, pp. S91–S102, 2015.

[12] G William Rebeck, Joell S Reiter, Dudley K Strickland, and Bradley T Hyman, "Apolipoprotein e in sporadic alzheimer's disease: allelic variation and receptor interactions," *Neuron*, vol. 11, no. 4, pp. 575–580, 1993.

[13] Murat Bilgel, Yang An, Yun Zhou, Dean F Wong, Jerry L Prince, Luigi Ferrucci, and Susan M Resnick, "Individual estimates of age at detectable amyloid onset for risk factor assessment," *Alzheimer's & Dementia*, vol. 12, no. 4, pp. 373–379, 2016.

[14] Rebecca Koscik, Tobey J Betthauser, Erin M Jonaitis, Lindsay R Clark, Samantha Allison, Kimberly D Mueller, Bruce P Hermann, Jennifer D Poetter, Leah Sanson, Heather Shouel, Nathaniel A Chin, Bradley T Christian, and Sterling C Johnson, "Modeling pib pet trajectory groups identifies a subgroup with pib beta-amyloid accumulation near age 50 and predicts mk-6240 suvr," in *Human Amyloid Imaging*, 2019.