

Learning the Latent Heat Diffusion Process through Structural Brain Network from Longitudinal Amyloid- β Data

Md Asadullah Turja¹Guorong Wu²Defu Yang²Martin Styner^{1,2}

MTURJA@CS.UNC.EDU

GUORONG_WU@MED.UNC.EDU

DEFU@EMAIL.UNC.EDU

STYNER@EMAIL.UNC.EDU

¹ *Department of Computer Science, University of North Carolina at Chapel Hill, USA.*² *Department of Psychiatry, University of North Carolina at Chapel Hill, USA.*

Abstract

The excessive deposition of misfolded proteins such as amyloid- β ($A\beta$) protein is an aging event underlying several neurodegenerative diseases. Mounting evidence shows that the spreading of neuropathological burden has a strong association to the white matter tracts in the brain which can be measured using diffusion-weighted imaging and tractography technologies. Most of the previous studies analyze the dynamic progression of amyloid using cross-sectional data which is not robust to the heterogeneous $A\beta$ dynamics across the population. In this regard, we propose a graph neural network-based learning framework to capture the disease-related dynamics by tracking the spreading of amyloid across brain networks from the subject-specific longitudinal PET images. To learn from limited (2 – 3 timestamps) and noisy longitudinal data, we restrict the space of amyloid propagation patterns to a latent heat diffusion model which is constrained by the anatomical connectivity of the brain. Our experiments show that restricting the dynamics to be a heat diffusion mechanism helps to train a robust deep neural network for predicting future time points and classifying Alzheimer’s disease brain.

Keywords: Heat Diffusion, Florbetapir PET, Structural Brain Network, Alzheimer’s Disease, Graph Neural Network, Graph Embedding, Graph u-net, Amyloid Progression Pattern, Amyloid- β Pathology.

1. Introduction

Accumulation of amyloid- β ($A\beta$) peptide into extracellular plaques is known to be a central event in Alzheimer’s disease brain (Lue et al., 1999; Palop and Mucke, 2010; Jack et al., 2013b; Fan et al., 2018; Canter et al., 2017). An increasing number of studies (Seeley et al., 2009; Jonkman et al., 2020) suggest that amyloid deposition expands into areas that receive neuronal projections from other brain regions already exhibiting $A\beta$, spreading between interconnected neurons through large-scale networks. Advancement in $A\beta$ -PET imaging enables us to quantify the spatial distribution of amyloid in the brains of living humans, allowing for the study of disease progression before symptoms manifest (Villemagne et al., 2018). Raj et al. (2012) proposes a network diffusion model to characterize this progression pattern as a diffusive mechanism mediated by the brain’s connectivity network which finds characteristic sub-modules in Alzheimer’s (AD) brain. However, their approach doesn’t consider misfolded protein (MP) burdens such as amyloid or tau. In this regard, Vogel et al. (2020) proposes an epidemic spreading model based on anatomical connectivity of the

brain to explain the observed MP burdens in healthy and diseased (AD) brains. Both of these approaches show the importance of the large-scale brain network in modeling disease progression based on MP such as $A\beta$. However, these approaches use cross-sectional data which is not sufficient to uncover the true dynamics of amyloid because of the subject-wise heterogeneity across the population. Although the authors of the epidemic spreading model reconstruct the MP burdens in a subject-specific manner, they have to tune their model for each individual during test time which might lead to overfitting and potentially distort the true dynamics.

We propose *Diffusion u-net*: a graph u-net (Gao and Ji, 2019) based architecture that uses longitudinal data to learn the disease dynamics in a subject-wise manner. However, the complex mechanics of $A\beta$ through brain networks is difficult to learn from a dataset that is noisy and has limited timestamps per subject (< 3 on average). We tackle this problem by constraining our model to learn a latent heat diffusion process that explains the neurodegenerative progression by the intercellular transfer of $A\beta$. We assume the observed longitudinal amyloid distributions are future instances of the heat diffusion mechanism constrained by the structural brain network. This way our model learns in a subset of possible disease processes which makes it possible to infer patterns from limited temporal data. Moreover, while previous approaches assume a predefined set of epicenter brain regions (For example, Vogel et al. (2020) fit their model using the left and right entorhinal cortex as the epicenters) from where $A\beta$ spreads to other brain areas, we show our model can find these epicenters from higher-order dynamics of $A\beta$ deposition. Extensive experimental results show that our model is quite effective in learning a robust embedding from only a few time point data and outperforms more complex models in terms of future $A\beta$ prediction, Alzheimer’s classification, and disease epicenter estimation.

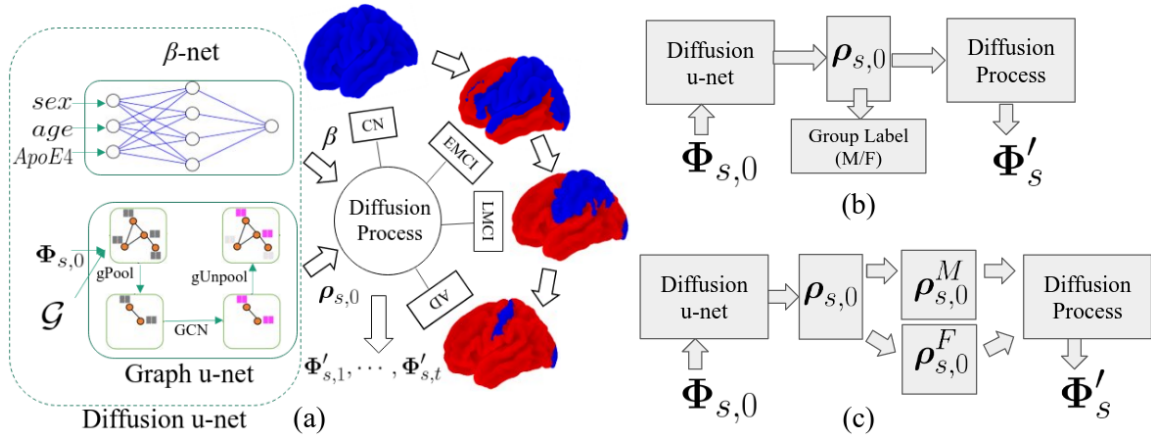


Figure 1: (a) Our proposed model **Diffusion u-net** consists of two neural networks — graph u-net and β -net to estimate future $A\beta$ depositions using Equation 4, (b) **Cluster u-net** learns group specific dynamics using an addition classifier to classify sex (Male/Female) based on $\rho_{s,0}$, and, (c) **Adaptive u-net** learns separate group specific representation $\rho_{s,0}^M$ and $\rho_{s,0}^F$ using a multi-task learning paradigm.

2. Method and Materials

2.1. Dataset

Our dataset consists of longitudinal amyloid data processed from 1312 subjects from Alzheimer’s Disease Neuroimaging Initiative (ADNI) from five different diagnosis group — Normal Control (CN), Significant Memory Concern (SMC), Early Mild Cognitive Impairment (EMCI), Late Mild Cognitive Impairment (LMCI), and, Alzheimer’s Disease (AD). with 316, 157, 357, 260, and, 222 subjects respectively. Each of the subjects has PET images at 1-6 time points (2.3 on average) with a 6 - 30 months gap between consecutive time points. The PET scans are first registered to T1-weighted MR images and then standardized uptake values (SUV) are computed for each voxel. Region-specific SUV values are computed by partitioning the cortical surface into 148 regions by using Destrieux atlas in FreeSurfer (Destrieux et al., 2012) and computing the average SUV for all the voxels of that brain region. The ratio of the SUV values (SUVR) is computed using the SUV of the cerebellum as the reference. We then normalize the SUVR based on the batch-wise min-max normalization approach. We compute the structural connectivity matrices from dMRI data of 103 subjects using FSL’s probtrackx (Behrens et al., 2007) (two-fiber model, 10000 tracks per seed) to obtain 148×148 connectivity matrices. We use the average white matter connectivity networks from the CN group in our experiments which is more robust to spurious noises in our data compared to subject-wise networks.

2.2. Method

Let $\Phi_s = [\Phi_{s,0}, \dots, \Phi_{s,T_s}]$ be a matrix for the subject s where each column $\Phi_{s,t} = [\phi_{s,t}^v]$ represent the observed longitudinal $A\beta$ values for the brain region v at different time-points t and T_s is the number of observations of s . We consider $\Phi_{s,t}$ as a graph signal on the structural brain network $\mathcal{G}(\mathcal{V}, \mathcal{E})$ which is an undirected weighted graph with $v \in \mathcal{V}$ corresponding to the region of interests (ROI) from destriux parcellation (Destrieux et al., 2012) and $e_{ij} \in \mathcal{E}$ corresponding to the edge that connects two ROIs i and j with weight w_{ij} proportional to the number of white matter fibers between i and j . Let $\rho_s = \frac{d\Phi_s}{dt} = [\rho_{s,0}, \dots, \rho_{s,T_s}]$ be the net amyloid generation rate at each ROI which characterizes the dynamics of the disease progression within the brain. ρ_s exhibits a non-linear dynamics due to inter-regional transfer of amyloid between two regions connected by white matter pathways. We model this dynamics by assuming that ρ_s follows the heat diffusion equation in Equation 1 which is in accordance with the prion-like hypothesis in neurodegenerative diseases (Palop and Mucke, 2010; Frost and Diamond, 2010).

$$\frac{d\rho_s}{dt} = -\beta\mathcal{L}\rho_s \tag{1}$$

Here, β is a scalar which controls speed of the amyloid progression through the structural network and \mathcal{L} is the graph laplacian operator with,

$$\mathcal{L}_{ij} = \begin{cases} \sum_{j'} w_{ij'}, & \text{if } i = j \\ -w_{ij}, & \text{if } i \neq j \text{ and } i \text{ adjacent to } j \\ 0, & \text{otherwise} \end{cases} \tag{2}$$

Prediction of future $A\beta$, Φ'_s : To estimate the future amyloid depositions $\Phi'_{s,t}$ at time t , we first compute $\rho_{s,t}$ from a learned $\rho_{s,0}$ using the heat diffusion equation (Equation 1) and aggregate $\rho_{s,t}$ over time using Equation 3.

$$\Phi'_{s,t} = \int_0^t \rho_{s,\tau} d\tau \tag{3}$$

Equation 1 has a closed form solution which is, $\rho_{s,t} = e^{-\beta \mathcal{L}t} \rho_{s,0} = \mathcal{U} e^{-\beta \Lambda t} \mathcal{U}^T \rho_{s,0}$. Here $\mathcal{U} = [u_1, \dots, u_{|\mathcal{V}|}]$ and $\Lambda = \text{diag}(\lambda_1, \dots, \lambda_{|\mathcal{V}|})$ come from eigendecomposition of $\mathcal{L} = \mathcal{U} \Lambda \mathcal{U}^T$. Replacing $\rho_{s,t}$ on Equation 3 and doing some simplification we get our longitudinal estimate:

$$\Phi'_{s,t} = \sum_{i=1}^{|\mathcal{V}|} \frac{1}{\beta \lambda_i} (1 - e^{-\beta \lambda_i t}) u_i^T \rho_{s,0} u_i \tag{4}$$

Equation 4 has two unknown variables — β and $\rho_{s,0}$. We train **Diffusion u-net** which consists of the following two neural network models to estimate those variables from baseline $A\beta$ values $\Phi_{s,0}$.

Graph u-net: We train a graph u-net model which takes $\Phi_{s,0}$ as inputs and predicts its derivative $\rho_{s,0}$. The u-net architecture has a depth of 4 with graph convolution layers of 64 units. The pooling layers in u-net select the top 50% of the ROIs based on learned ranking.

β -net: The inter-regional amyloid transfer speed β can be affected by several risk factors of Alzheimer’s disease (Raffaitin et al., 2009; Solfrizzi et al., 2010; Campbell et al., 2013) such as 1. age — amyloid progression slows down over time towards a plateau (Villemagne et al., 2013), 2. sex — interaction between tau and amyloid is stronger in women resulting in a higher risk of developing symptoms of the serious brain disease (Buckley et al., 2019), apolipoprotein E (APOE) $\epsilon 4$ allele — $A\beta$ deposition rate is higher with APOE- $\epsilon 4$ carriers (Kanekiyo et al., 2014), and education level — people with a lower level of education are known to be less resistant against Alzheimer’s (Ngandu et al., 2007). In this regard, we train a multilayer perceptron with 2 hidden layers of 256 units to predict β from the aforementioned risk factors which account for the heterogeneous dynamics of $A\beta$ across the population.

Finally, we propose two more models (Figure 1(b)) which account for the heterogeneity between female and male groups in the population by learning sex-specific representations $\rho_{s,0}^F$ and $\rho_{s,0}^M$ respectively. The two models are —

Cluster u-net: Here we aim to cluster the embedding $\rho_{s,0}$ based on the sex information (Male/Female). We train a multilayer perceptron (3 hidden layers with 256 units each) to classify $\rho_{s,0}$ based on sex. In this way, the u-net learns an embedding that is informative of the group. The u-net and the classifier are then jointly trained to reconstruct Φ'_s .

Adaptive u-net: We then took a different approach based on the multi-task learning paradigm which learns an adaptive embedding for each group. Here we consider the male and female group as a separate task and learn separate embeddings ($\rho_{M,0}$, $\rho_{F,0}$ respectively) for them. To do this, we use our graph u-net as a base architecture and fork two different branches for each of the groups. These branches consist of a multilayer perceptron with 1 hidden layer of 64 units which are trained using the data from the corresponding groups.

Training: With $\rho_{s,0}$ from graph u-net and β from β -net, we estimate the future amyloid depositions, $\Phi'_{s,0}, \dots, \Phi'_{s,t}$ using equation 4. We then jointly train both of these neural networks by minimizing the following loss function:

$$\mathcal{L}_{recon} = \sum_{s \in S} \left(\sum_{v \in \mathcal{V}} \sum_{t=0}^{T_s} (\phi'_{s,t}{}^v - \phi_{s,t}{}^v)^2 + (I + \mathcal{L})^{-1} \rho_{s,0} \right) \quad (5)$$

The first term is the mean-squared error between Φ'_s and Φ_s . The second term is a regularization term on $\rho_{s,0}$ (Huang et al., 2015) which is used to stabilize the training during the initial epochs by enforcing the model to learn embedding with lower entropy, i.e., a higher imbalance between neighboring regions.

2.3. Experimental Setup

We evaluate our method by doing a series of experiments on three different tasks — reconstruction, classification, and estimation of $A\beta$ epicenters.

2.3.1. RECONSTRUCTION

We compare the accuracy of the reconstructed $A\beta$ values from our model with the following models:

Stationary model: We assume the $A\beta$ values are constant, i.e, there no change in amyloid deposition over time.

Linear model: Here we assume a linear progression of amyloid without any diffusion through the structural brain network. This is equivalent to assuming constant $\rho_{s,t}$ in Equation 3, i.e., $\rho_{s,t} = \rho_{s,0} = \mathbf{m}_s$ which results in $\Phi'_{s,t} = \mathbf{m}_s t + \Phi_{s,0}$. We consider two models in this regard — **Linear-shared** where \mathbf{m}_s represents the shared slope for each ROI for all the subjects, and **Linear-subject** where we learn subject-specific \mathbf{m}_s from $\Phi_{s,0}$ as a linear combination, i.e., $\mathbf{m}_s = \mathbf{W} \Phi_{s,0}$. Here \mathbf{W} is a $|\mathcal{V}| \times |\mathcal{V}|$ weight matrix which is learned using stochastic gradient descent.

GRU: Here we learned the non-linear dynamics from a gated recurrent unit (GRU) (Cho et al., 2014) based architecture instead of our heat diffusion equation. We use the same graph u-net based architecture as ours to learn a hidden embedding $\rho_{s,t}$ for each time point t and then feed them to a GRU to predict the future amyloid depositions.

2.3.2. CLASSIFICATION

We show the effectiveness of our predicted Φ'_s by a series of classification experiments. We train 5 different classifiers which classify between CN and AD group based on Φ'_s from Adaptive u-net. We compare the classification result by training all of these models on the ground truth $A\beta$ deposition Φ_s . The models are:

\mathbf{v}_1 : This is a simple random forest classifier trained using only the baseline $A\beta$ depositions ($\Phi_{s,0}$ and $\Phi'_{s,0}$).

\mathbf{v}_2 : This model is a graph convolution based classifier (**GCNClassifier**) with 2 graph convolutional blocks with size 64 (Kipf and Welling, 2017) followed by a multilayer perceptron with 3 hidden layers of size 256. This model is trained using the baseline $A\beta$ depositions

($\Phi_{s,0}$ and $\Phi'_{s,0}$) to minimize the cross-entropy loss between predicted and ground-truth diagnosis label.

v_3 : Unlike v_2 , in this model we use the longitudinal data (Φ_s and Φ'_s) to predict the diagnosis label (CN/AD). First, we learn an embedding from the longitudinal data using a recurrent neural network with 2 stacked gated recurrent units (GRU) with hidden unit size 148 and then feed this embedding through our GCNClassifier.

v_4 : In this version, instead of using a GRU, we utilize the longitudinal data by taking the minimum (Φ_s^{min}) and maximum (Φ_s^{max}) A β values across all the time points for each brain region and feed them to the GCNClassifier (Min-Max Classifier). The input Φ_s^{mnx} to the model is the concatenation of the minimum and maximum values, i.e., $\Phi_s^{mnx} = [\Phi_s^{min}, \Phi_s^{max}]$.

v_5 : In all the previous models, the input of the classifier $\Phi'_{s,0}$ comes from the pretrained Adaptive u-net. In this model, we first pretrain Adaptive u-net with data from all the diagnosis groups and then jointly train this Adaptive u-net and v_4 for only CN and AD group. We minimize the sum of reconstruction loss from Equation 5 and cross-entropy loss for joint training.

2.3.3. EPICENTER ESTIMATION

A β deposition starts at certain epicenters in the Alzheimer’s brain from where it spreads to neighbouring brain regions. It is known that A β load increases almost in a linear fashion until it plateaus exhibiting non-linear dynamics (Jack et al., 2013a; Villemagne et al., 2013). This means the epicenters should exhibit higher order dynamics ($\Phi_s^1 = \frac{d\Phi_{s,0}}{dt}$, $\Phi_s^2 = \frac{d\Phi_s^1}{dt}$ etc.) compared to other brain regions. In this experiment, we use these higher order dynamics to find the epicenters from where the A β spreads to the brain. Our diffusion u-net model learns to estimate a graph signal by reconstructing $\Phi_{s,0}$ from $\rho_{s,0}$ using Equation 3 which enforces $\rho_{s,0} = \frac{d\Phi_{s,0}}{dt}$. This means we can treat our model as a differential operator of a graph signal. We recursively compute higher order dynamics using $\Phi_s^{l+1} = \frac{d\Phi_s^l}{dt}$ where $\Phi_s^0 = \Phi_{s,0}$ by feeding Φ_s^l to our model until $\Phi_s^l = \mathbf{0}$. We then define a diffusion distance measure $d_s^v = [d_s^v]$ for each brain region v by simply counting how many higher order derivatives v has and normalizing it using min-max normalization approach across all subjects. Higher d_s^v value will indicate higher likelihood of being an epicenter.

3. Results

3.1. Reconstruction

Table 1 shows the r^2 values between predicted A β values $\Phi'_{s,t}$ and ground truth A β values $\Phi_{s,t}$ across all the ROIs for all our reconstruction models. Our experiments show that the stationary model can explain 79% of the variance of the longitudinal A β data. Linear-shared and Linear-subject models improve the performance by 2% and 7% respectively which shows the effectiveness of learning subject-specific dynamics. Surprisingly, the non-linear GRU model performs worse ($r^2 = 0.84$) compared to the Linear-subject model ($r^2 = 0.86$). We hypothesize that it’s challenging for complex neural networks to learn the hidden dynamics due to the low signal-to-noise ratio in our dataset. In this regard, our model Diffusion u-net

replaces the GRU with a much simpler non-linear process (Equation 4) which outperforms the aforementioned models with $r^2 = 0.88$. Moreover, our results in Table 1 show that the performance of all the models is significantly worse for the female subjects in Alzheimer’s group. Cluster u-net and Adaptive u-net attempt to solve this problem by learning sex-specific representation. While Cluster u-net doesn’t show any improvement, Adaptive u-net significantly improves the performance for both males and females across all the diagnosis groups with 2% improvement ($r^2 = 0.90$) overall and 5% improvement ($r^2 = 0.56$) for the female-AD group compared to our original Diffusion u-net.

Table 1: r^2 values across different diagnosis groups

Model	Sex	CN	SMC	EMCI	LMCI	AD	Overall
Stationary	Female	0.74	0.61	0.73	0.60	0.51	0.79
	Male	0.83	0.55	0.79	0.77	0.82	
Linear-shared	Female	0.76	0.64	0.72	0.61	0.52	0.81
	Male	0.84	0.62	0.80	0.78	0.82	
Linear-subject	Female	0.82	0.71	0.85	0.72	0.52	0.86
	Male	0.86	0.70	0.84	0.82	0.86	
GRU	Female	0.79	0.65	0.83	0.60	0.21	0.84
	Male	0.86	0.54	0.84	0.73	0.85	
Diffusion u-net	Female	0.84	0.73	0.88	0.74	0.52	0.88
	Male	0.88	0.78	0.88	0.82	0.88	
Cluster u-net	Female	0.84	0.74	0.88	0.75	0.53	0.88
	Male	0.87	0.79	0.87	0.81	0.89	
Adaptive u-net	Female	0.87	0.77	0.91	0.78	0.56	0.90
	Male	0.92	0.89	0.93	0.86	0.96	

3.2. Classification

The results of the classification experiments are listed in Table 2. From the results we find that v_1 and v_2 don’t show any significant difference in performance between Φ_s and Φ'_s . In v_3 , the performance for both Φ_s and Φ'_s are again similar and they outperform our previous models v_1 and v_2 . This shows that our model doesn’t lose any temporal information pertinent for diagnosis. Interestingly, in v_4 we see a significant difference in performance where Φ'_s beats Φ_s by more than 2% in terms of ROC value. Minimum and maximum $A\beta$ values are highly sensitive to spurious noises in data. We maintain that our proposed method reduces such noises by restricting the dynamics in a heat diffusion mechanism. However, jointly training the Adaptive u-net with classifier in v_5 , doesn’t show any performance boost compared to v_4 .

3.3. Epicenter estimation

The group-wise average of d_s for different diagnosis groups (CN, AD, etc.) shows that the amyloid starts at the temporal lobe (*Anterior transverse collateral sulcus*, *Temporal*

Table 2: Micro ROC values for CN/AD classification

Model	Model Description	Input	ROC
v_1	Random Forest Classifier	$\Phi_{s,0}$	0.8855
		$\Phi'_{s,0}$	0.8857
v_2	GCNClassifier	$\Phi_{s,0}$	0.8853
		$\Phi'_{s,0}$	0.8872
v_3	GRU + GCNClassifier	Φ_s	0.9304
		Φ'_s	0.9288
v_4	Min-Max + GCNClassifier	Φ_s^{mnx}	0.9061
		Φ'_s^{mnx}	0.9241
v_5	Adaptive u-net + v_4	Φ'_s	0.9289

pole) and then spreads through limbic lobe (*Subcallosal area*), and, frontal lobe (*Medial orbital sulcus, Orbital sulci*) and finally expand into the parietal lobe and the occipital lobe(Figure 2). These results align with the current findings of cortical AD progression (Smith et al., 2002; Dickerson et al., 2008; Wu et al., 2021).

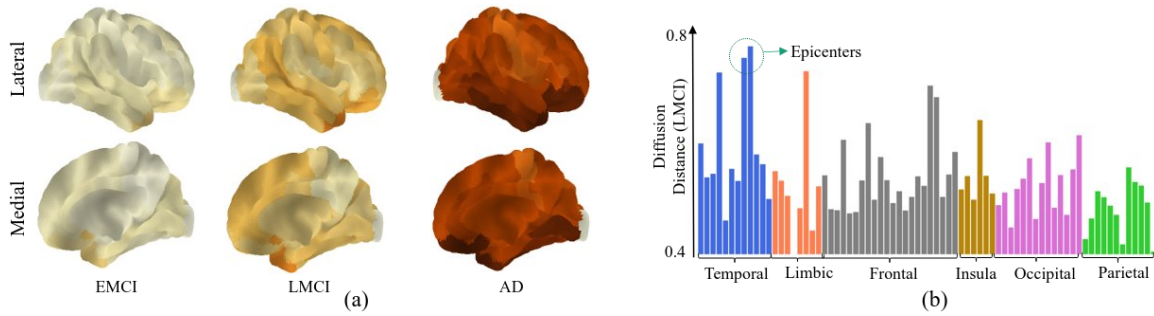


Figure 2: (a) The dynamics of diffusion distance (d_s) of the right hemisphere from EMCI to AD group are shown where darker color indicates higher diffusion distances. (b) Average diffusion distances for the right hemisphere of the LMCI group are organized by the lobes (marked with different color bars) of the brain. The peak distances in the temporal lobe indicate epicenter brain regions from where $A\beta$ spreads to other brain regions.

4. Conclusion

Understanding the progression dynamics of Alzheimer’s disease is a major factor in the preclinical stages of the disease. In this paper, we propose a graph neural network-based model to learn these dynamics from longitudinal amyloid data in a subject-wise manner. By constraining these dynamics to a heat diffusion process, our model can learn robust dynamics from a few time point data which shows effectiveness in future amyloid prediction, Alzheimer’s disease classification, and, disease epicenter prediction.

Acknowledgments

This work is supported by the National Institutes of Health (NIH) (grant numbers AG068399, AG059065, AG049089), and Foundation of Hope.

References

- T.E. Behrens, H. Johansen Berg, S. Jbabdi, M.F. Rushworth, and M.W. Woolrich. Probabilistic diffusion tractography with multiple fibre orientations: What can we gain? *Neuroimage*, 34(1):144—55, 2007. doi: 10.1016/j.neuroimage.2006.09.018.
- Rachel F. Buckley, Elizabeth C. Mormino, Jennifer S. Rabin, Timothy J. Hohman, Susan Landau, Bernard J. Hanseeuw, Heidi I. L. Jacobs, Kathryn V. Papp, Rebecca E. Amariglio, Michael J. Properzi, Aaron P. Schultz, Dylan Kirn, Matthew R. Scott, Trey Hedden, Michelle Farrell, Julie Price, Jasmeer Chhatwal, Dorene M. Rentz, Victor L. Villemagne, Keith A. Johnson, and Reisa A. Sperling. Sex Differences in the Association of Global Amyloid and Regional Tau Deposition Measured by Positron Emission Tomography in Clinically Normal Older Adults. *JAMA Neurology*, 76(5):542–551, 05 2019. ISSN 2168-6149. doi: 10.1001/jamaneurol.2018.4693. URL <https://doi.org/10.1001/jamaneurol.2018.4693>.
- Noll L. Campbell, Fred Unverzagt, Michael A. LaMantia, Babar A. Khan, and Malaz A. Boustani. Risk factors for the progression of mild cognitive impairment to dementia. *Clin Geriatr Med.*, 29(4):873–93, 2013. doi: 10.1016/j.cger.2013.07.009.
- R.G. Canter, H. Choi, J. Wang, L.A. Watson, C.G. Yao, F. Abdurrob, S.M. Bousleiman, I. Delalle, K. Chung, and L-H. Tsai. 3d mapping reveals network-specific amyloid progression and subcortical susceptibility. *bioRxiv*, 2017. doi: 10.1101/116244. URL <https://www.biorxiv.org/content/early/2017/03/19/116244>.
- Kyunghyun Cho, Bart van Merriënboer, Caglar Gulcehre, Dzmitry Bahdanau, Fethi Bougares, Holger Schwenk, and Yoshua Bengio. Learning phrase representations using RNN encoder–decoder for statistical machine translation. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 1724–1734, Doha, Qatar, October 2014. Association for Computational Linguistics. doi: 10.3115/v1/D14-1179. URL <https://www.aclweb.org/anthology/D14-1179>.
- Christophe Destrieux, Bruce Fischl, Anders Dale, and Eric Halgren. Automatic parcellation of human cortical gyri and sulci using standard anatomical nomenclature. *Neuroimage*, 78(6):387—395, 2012. URL <https://doi.org/10.1212/WNL.0b013e318245d295>.
- Bradford C. Dickerson, Akram Bakkour, S. Jbabdi, David H. Salat, Eric Feczko, Jenni Pacheco, Douglas N. Greve, Fran Grodstein, Christopher I. Wright, Deborah Blacker, H. Diana Rosas, Reisa A. Sperling, Alireza Atri, John H. Growdon, Bradley T. Hyman, John C. Morris, Bruce Fischl, and Randy L. Buckner. The cortical signature of alzheimer’s disease: regionally specific cortical thinning relates to symptom severity in very mild to mild ad dementia and is detectable in asymptomatic amyloid-positive individuals. *Cereb Cortex*, 19(3):497—510, 2008. doi: 10.1093/cercor/bhn113.

- Ling-Yun Fan, Kai-Yuan Tzen, Ya-Fang Chen, Ta-Fu Chen, Ya-Mei Lai, Ruoh-Fang Yen, Ya-Yao Huang, Chyng-Yann Shiue, and Ming-Jang Chiu. The relation between brain amyloid deposition, cortical atrophy, and plasma biomarkers in amnesic mild cognitive impairment and alzheimer’s disease. *Front Aging Neurosci*, 10:175, 2018. doi: 10.3389/fnagi.2018.00175. URL <https://pubmed.ncbi.nlm.nih.gov/29967578>.
- Bess Frost and Marc I. Diamond. Prion-like mechanisms in neurodegenerative diseases. *Nature Reviews Neuroscience*, 11:155–159, 2010. doi: 10.1038/nrn2786. URL <https://doi.org/10.1038/nrn2786>.
- Hongyang Gao and Shuiwang Ji. Graph u-nets. In *ICML*, 2019.
- Weiyu Huang, Santiago Segarra, and Alejandro Ribeiro. Diffusion distance for signals supported on networks. In *2015 49th Asilomar Conference on Signals, Systems and Computers*, pages 1219–1223, 2015. doi: 10.1109/ACSSC.2015.7421335.
- C. R. Jack, H. Wiste, T. G. Lesnick, S. D. Weigand, D. Knopman, P. Vemuri, V. S. Pankratz, M. L. Senjem, J. L. Gunter, M. Mielke, V. Lowe, B. Boeve, and R. Petersen. Brain β -amyloid load approaches a plateau. *Neurology*, 80:890 – 896, 2013a.
- Clifford R. Jack, David S. Knopman, William J. Jagust, Michael W. Weiner, Paul S. Aisen, Leslie M. Shaw, Prashanthi Vemuri, Heather J. Wiste, Stephen D. Weigand, Timothy G. Lesnick, Vernon S. Pankratz, Michael C. Donohue, and John Q. Trojanowski. Tracking pathophysiological processes in alzheimer’s disease: an updated hypothetical model of dynamic biomarkers. *The Lancet Neurology*, 207(16), 2013b. doi: 10.1016/S1474-4422(12)70291-0. URL <https://www.sciencedirect.com/science/article/pii/S000294401065184X>.
- Laura E. Jonkman, Martijn D. Steenwijk, Nicky Boesen, Annemieke J.M. Rozemuller, Frederik Barkhof, Jeroen J.G. Geurts, Linda Douw, and Wilma D.J. van de Berg. Relationship between β -amyloid and structural network topology in decedents without dementia. *Neurology*, 95(5):e532–e544, 2020. ISSN 0028-3878. doi: 10.1212/WNL.0000000000009910. URL <https://n.neurology.org/content/95/5/e532>.
- Takahisa Kanekiyo, Huaxi Xu, and Guojun Bu. Apoe and $a\beta$ in alzheimer’s disease: accidental encounters or partners? *Neuron*, 81(4):740–54, 2014. doi: 10.1016/j.neuron.2014.01.045.
- Thomas N. Kipf and Max Welling. Semi-Supervised Classification with Graph Convolutional Networks. In *Proceedings of the 5th International Conference on Learning Representations*, ICLR ’17, 2017. URL <https://openreview.net/forum?id=SJU4ayYgl>.
- Lih-Fen Lue, Yu-Min Kuo, Alex E. Roher, Libuse Brachova, Yong Shen, Lucia Sue, Thomas Beach, Janice H. Kurth, Russel E. Rydel, and Joseph Rogers. Soluble amyloid β peptide concentration as a predictor of synaptic change in alzheimer’s disease. *The American Journal of Pathology*, 155(3):853–862, 1999. ISSN 0002-9440. doi: [https://doi.org/10.1016/S0002-9440\(10\)65184-X](https://doi.org/10.1016/S0002-9440(10)65184-X). URL <https://www.sciencedirect.com/science/article/pii/S000294401065184X>.

- T. Ngandu, E. von Strauss, E-L. Helkala, B. Winblad, A. Nissinen, J. Tuomilehto, H. Soininen, and M. and Kivipelto. Education and dementia: what lies behind the association? *Neurology*, 71(19):1442–50, 2007. doi: 10.1212/01.wnl.0000277456.29440.16.
- Jorge J. Palop and Lennart Mucke. Amyloid- β -induced neuronal dysfunction in alzheimer’s disease: from synapses toward neural networks. *Nature Neuroscience*, 13:812–818, 2010. doi: 10.1038/nn.2583. URL <https://doi.org/10.1038/nn.2583>.
- Christelle Raffaitin, Henri Gin, Jean-Philippe Empana, Catherine Helmer, Claudine Berr, Christophe Tzourio, Florence Portet, Jean-François Dartigues, Annick Alperovitch, and Pascale Barberger-Gateau. Metabolic syndrome and risk for incident alzheimer’s disease or vascular dementia: the three-city study. *Diabetes care*, 32(1):169–74, 2009. doi: 10.2337/dc08-0272.
- A. Raj, A. Kuceyeski, and M. Weiner. A network diffusion model of disease progression in dementia. *Neuron*, 73(6):1204–15, 2012. doi: 10.1016/j.neuron.2011.12.040.
- William W. Seeley, Richard K. Crawford, Juan Zhou, Bruce L Miller, and Michael D. Greicius. Neurodegenerative diseases target large-scale human brain networks. *Neuron*, 62(1):42–52, 2009. doi: 10.1016/j.neuron.2009.03.024.
- A. David Smith, Yun Peng, Ming Hong, and Yingchun Zhang. Imaging the progression of alzheimer pathology through the brain. *Proc Natl Acad Sci U S A*, 99(7):4135–7, 2002. doi: 10.1073/pnas.082107399.
- Vincenzo Solfrizzi, Emanuele Scafato, Cristiano Capurso, Alessia D’Introno, Anna Maria Colacicco, Vincenza Frisardi, Gianluigi Vendemiale, Marzia Baldereschi, Gaetano Crepaldi, Antonio Di Carlo, Lucia Galluzzo, Claudia Gandin, Domenico Inzitari, Stefania Maggi, Antonio Capurso, and Francesco Panza. Metabolic syndrome and the risk of vascular dementia: the italian longitudinal study on ageing. *Journal of Neurology, Neurosurgery & Psychiatry*, 81(4):433–440, 2010. ISSN 0022-3050. doi: 10.1136/jnnp.2009.181743. URL <https://jnnp.bmj.com/content/81/4/433>.
- Victor L. Villemagne, Samantha Burnham, Pierrick Bourgeat, Belinda Brown, Kathryn A. Ellis, Olivier Salvado, Cassandra Szoeki, S Lance Macaulay, Ralph Martins, Paul Maruff, David Ames, Christopher C. Rowe, and Colin L. Masters. Amyloid β deposition, neurodegeneration, and cognitive decline in sporadic alzheimer’s disease: a prospective cohort study. *The Lancet. Neurology*, 12(4):357–67, 2013. doi: 10.1016/S1474-4422(13)70044-9.
- Victor L. Villemagne, Vincent Doré, and Christopher C. Rowe Samantha C. Burnham, Colin L. Masters. Imaging tau and amyloid- β proteinopathies in alzheimer disease and other conditions. *Nature reviews. Neurology.*, 14(4):225–236, 2018. doi: 10.1038/nrneurol.2018.9.
- Jacob W. Vogel, Yasser Iturria-Medina, Olof T. Strandberg, Ruben Smith, Elizabeth Levitis, Alan C. Evans, Oskar Hansson, Alzheimer’s Disease Neuroimaging Initiative, and the Swedish BioFinder Study. Spread of pathological tau proteins through communicating neurons in human alzheimer’s disease. *Nature Communications*, 11(1):2612, 2020. doi: 10.1038/s41467-020-15701-2. URL <https://doi.org/10.1038/s41467-020-15701-2>.

Zhanxiong Wu, Yun Peng, Ming Hong, and Yingchun Zhang. Gray matter deterioration pattern during alzheimer’s disease progression: A regions-of-interest based surface morphometry study. *Frontiers in Aging Neuroscience*, 13:23, 2021. ISSN 1663-4365. doi: 10.3389/fnagi.2021.593898. URL <https://www.frontiersin.org/article/10.3389/fnagi.2021.593898>.

Appendix A. Implementation Details

This section provides the details of the neural network architectures in this work.

A.1. Diffusion u-net

We have two different neural networks in our model: 1. Graph u-net (Figure 3), and, 2. β -net. The graph u-net has a depth of 4. The graph convolution layers inside the u-net have 64 units and the graph pooling layer selects top 50% of the nodes based on the learned ranking. The β -net is a multilayer perceptron with two hidden layers each having 256 units. For training we split the dataset into training (70%), validation (15%), and, evaluation (15%) folds and jointly train both of these networks for 700 epochs with a learning rate = 1e-6, dropout rate = 0.01, activation = Exponential linear unit (ELU) and batch size = 32.

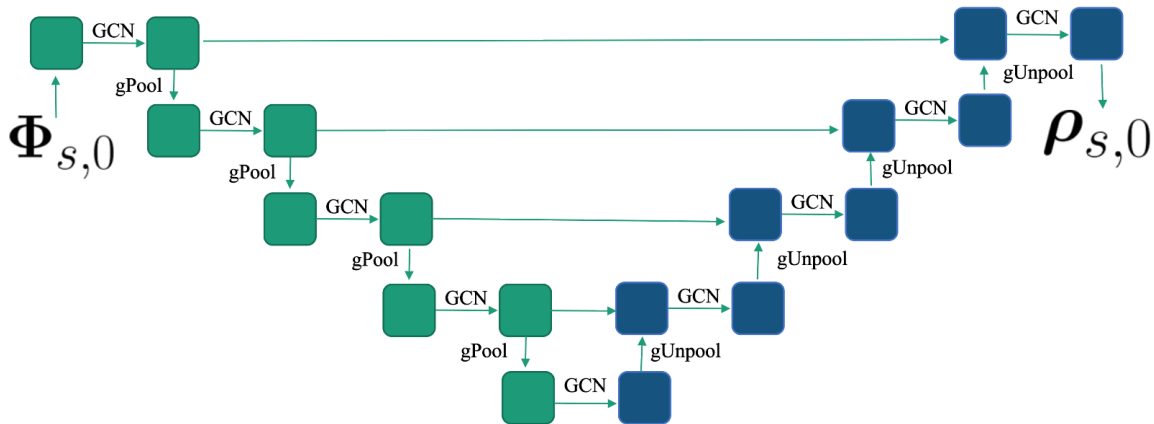


Figure 3: Graph u-net architecture

A.2. GCNClassifier

The GCNClassifier has two convolutional blocks from (Kipf and Welling, 2017) each having 64 units. The output of these convolutional blocks is then concatenated and fed into a multilayer perceptron with 3 hidden layers each having 256 units. The classifier is trained for 500 epochs using the hyper-parameters: learning rate = 1e-5, dropout rate = 0.01, activation = Exponential linear unit (ELU) and batch size = 32.

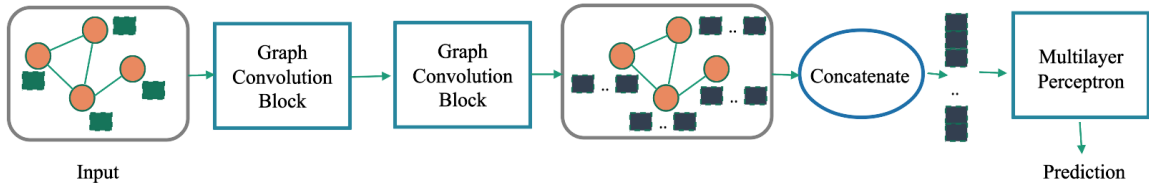


Figure 4: Graph convolution based classifier