ADAPTIVE SHRINKAGE ESTIMATION FOR PERSONALIZED DEEP KERNEL REGRESSION IN MODELING BRAIN TRAJECTORIES

Anonymous authors
Paper under double-blind review

ABSTRACT

Longitudinal biomedical studies track individuals over time to capture dynamics in brain development, disease progression, and treatment effects. Estimating trajectories of brain measurements in such studies is challenging due to biological variability and inconsistencies in measurement protocols (e.g. MRI scanner variations and upgrades). Herein, we introduce a novel personalized Deep Kernel Regression framework for forecasting longitudinal regional brain volumetric changes. Our approach integrates two key components: a population model that captures brain volume trajectories from a large and diverse cohort, and a personalization step that generates subject-specific models for individual trajectories. To optimally combine these predictive distributions, we propose the Adaptive Posterior Shrinkage Estimation technique, which effectively balances population-level trends with individual-specific data. We evaluate model's performance through predictive accuracy metrics, uncertainty quantification, and validation against external clinical studies. Benchmarking against state-of-the-art statistical and machine learning models—including Linear Mixed Effects models, Generalized Additive Models, and deep learning methods—demonstrates the superior predictive performance of our approach across a variety of experiments. Additionally, we apply our method on predicting trajectories of composite neuroimaging biomarkers, e.g. machine-learning patterns of brain structure related to aging and Alzheimer's Disease, which highlights the generalizability of our approach to model the progression of longitudinal neuroimaging biomarkers. Furthermore, validation on three external neuroimaging studies confirms the generalizability and applicability of our method across different clinical contexts. These results highlight the versatility and robustness of our framework for predicting longitudinal brain volume changes. Overall, this framework effectively addresses the inherent challenges in longitudinal biomedical studies, providing a valuable predictive tool that can inform patient management decisions, clinical trial design and treatment effect estimation.

1 INTRODUCTION

Accurately predicting the temporal progression of brain atrophy is critical for tracking disease progression and determining optimal intervention points (Maheux et al., 2023). To ensure robust predictions over time, models must adapt dynamically as new subject-specific data becomes available. This task falls within the domain of longitudinal prediction from multivariate data, where traditional methods like Linear Mixed Effects models struggle in high-dimensional settings. Various approaches have been proposed within the framework of Gaussian Processes (GPs) for personalization (Rasmussen & Williams, 2006). Liu & Vasconcelos (2015) and Eleftheriadis et al. (2017) introduced the concept of domain-adaptive GPs, which provide tractable and intuitive model updates and integrate information from both training and adaptation data. This idea was further developed by (Rudovic et al., 2019), who proposed a meta-weighting scheme to combine two personalized GP models to achieve a final personalized prediction. This approach was applied to predict the cognitive score of ADAS-Cog13 (Mohs et al., 1997) up to two years in the future. Chung et al. (2019) proposed a Deep Mixed Effects approach for personalization in electronic health record (EHR) time-series data, utilizing an LSTM (Hochreiter & Schmidhuber, 1997) as a deep mean function to learn the population trend. Subsequently, an individual GP was employed to capture the prediction shift due to the specific individual's trend.

In this paper, we introduce a composite framework for personalized Deep Kernel Regression to model the longitudinal volumetric trajectories of brain Regions of Interest (ROIs) for subjects with follow-up acquisitions. We adapt the shrinkage estimator idea from statistical learning, commonly seen in Bayesian statistics and penalized inference (Shou et al., 2014; Lindquist & Gelman, 2009; James & Stein, 1961), and extend it to achieve optimal personalization by adaptively learning and balancing the strengths and weaknesses of each predictor.

Diverging from previous works, our method focuses on the longitudinal prediction of the biomarker at any time in the future and not the prediction in a specific time-window as suggested in (Rudovic et al., 2019). Additionally, our method does not require equal time intervals as (Rudovic et al., 2019), we rather utilize all the available longitudinal information across subjects even in the presence of temporal unalignment. From the personalization perspective, our method, learns the optimal weights to combine the population DKGP prediction with the individual-level deep kernel probabilistic predictor, while taking into account the time of observation and prediction uncertainty of each component for each subject. By balancing the strengths and weaknesses of each predictor through shrinkage, we achieve to optimally combine the two to provide a personalized outcome.

Our contributions are as follows: **(i)** We present a novel Deep Kernel Regression framework specifically designed for sparse longitudinal data. This framework employs a transformation function that maps multimodal and high-dimensional imaging and clinical features into a lower-dimensional space that is predictive of biomarker progression. **(ii)** We introduce a novel approach for personalized Deep Kernel Regression through Adaptive Shrinkage Estimation. This method optimally and intuitively provides personalized predictions through posterior correction of both the population and subject-specific predictive distributions **(ii)** We showcase that this method can be applied to composite neuroimaging biomarkers and in general longitudinal neuroimaging markers from high-dimensional multivariate input data. **(iii)** We demonstrate the generalizability of our method in different clinical contexts, showing its ability to generalize in external clinical studies.

2 Method

2.1 Problem Statement and Notations

In this work, we address the problem of predicting the temporal evolution of neuroimaging biomarkers based on the observed trajectories of individual subjects and baseline imaging and clinical information. We model a biomarker trajectory as a one-dimensional time series spanning multiple years, formalized by the function $f : U \to Y$, where $U \in \mathbb{R}^K$ and $Y \in \mathbb{R}$. We represent the input as $\mathbf{U} = (X, M, T)$, where X denotes the imaging features, M denotes the clinical covariates at the baseline visit of each subject, and T represents the temporal variable. The temporal variable T signifies the age increase in months between any subsequent scan and the baseline, primarily used for extrapolating ROI values over time and reconstructing the ROI trajectory. An ROI trajectory is denoted as $Y = (y_0, y_1, \ldots, y_n)$, corresponding to the ROI volumes at time points $T = (t_0, t_1, \ldots, t_n)$.

We aim to learn smooth functions of brain changes using multimodal imaging and clinical data. The Deep Kernel fuses the imaging and clinical modalities together and learns a lower dimensional representation that is informative for the progression of each ROI, whereas the GP models the dependency between time points and provides the smooth ROI temporal trajectories. Thus, the backbone model is described with the equation $f(\mathbf{U}) \sim \mathcal{GP}(\mu, \mathbf{K}(\Phi(\mathbf{U}), \Phi(\mathbf{U})))$, where Φ is a transformation function.

Fundamental to the DKGP exact inference is the dataset used for training. In our study, we assume a heterogeneous population dataset with longitudinal MRI acquisitions from multiple subjects over several years.

2.2 Population Model (p-DKGP)

The population model leverages data from the population dataset $D_p = \{\mathbf{U}_p, \mathbf{Y}_p\}$. It applies the transformation $\Phi(u; \mathbf{W}, \mathbf{b})$, which is a Multi-Layer Perceptron (MLP) that maps the input data $\mathbf{U}_p = (X, M, T)$ into a latent representation:

$$\mathbf{Z}_p = \Phi(\mathbf{U}_p; \mathbf{W}, \mathbf{b}). \tag{1}$$

This transformation is subsequently used by a Gaussian Process (GP), which models the ROI function f using an RBF kernel as the covariance function and a zero mean: $f(\mathbf{Z}_p) \sim \mathcal{GP}\left(0, K(\mathbf{Z}_p, \mathbf{Z}'_p)\right)$.

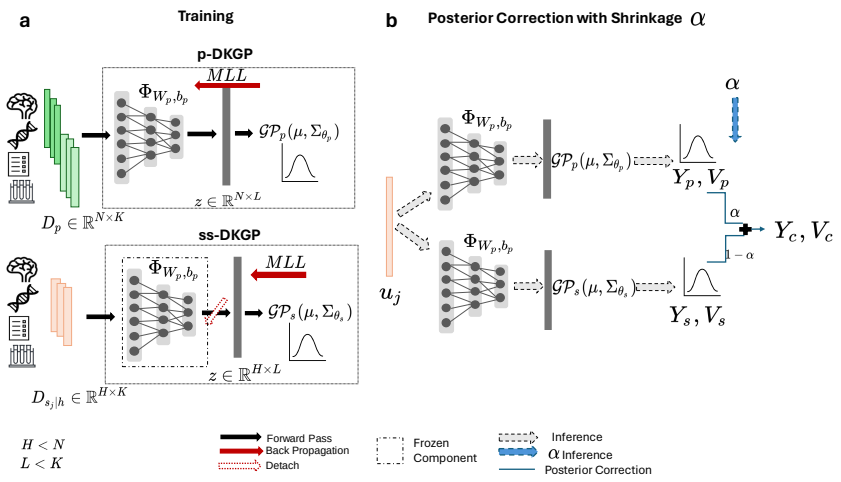

Figure 1: Overview of the proposed method. In Fig.1(**a**), we illustrate the training process of the two probabilistic models, p-DKGP and ss-DKGP. The population dataset D_p contains multiple longitudinal acquisitions across subjects, where N is the total number of samples from all subjects, and L is the latent dimension obtained after the transformation Φ. Different shades of green in the population dataset indicate different subjects in D_p. We denote the observed trajectory of subject j with h samples as $D_{s_j|h}$. These samples are utilized to train the ss-DKGP. During the training of the ss-DKGP, the transformation Φ is fixed, and only the subject-specific Gaussian Process (GP_s) is optimized. In Fig.1(**b**), we describe the personalization process through the shrinkage parameter α. For subject j, we extrapolate ROI values over time using both the p-DKGP and ss-DKGP models. These extrapolated values are then used to infer the optimal α for posterior correction, yielding the personalized posterior predictive mean Y_c variance V_c of the subject's trajectory.

The population parameters $\gamma_p = \{\mathbf{W}_p, \mathbf{b}_p, l_p, \sigma_p\}$ are learned jointly through the Marginal Log Likelihood (MLL) of the GP (Wilson et al., 2015; Rasmussen & Williams, 2006).

For a test subject j with input $u_j = (x_j, m_j, t)$, we denote the transformed input as $z_j = \Phi(u_j; \mathbf{W}_p, \mathbf{b}_p)$.

The posterior predictive distribution of the ROI function at point $u_j = (x_j, m_j, t)$ is:

$$f_{p_j} \mid (\mathbf{Z}_p, \mathbf{Y}_p), z_j \sim \mathcal{N}(\bar{\mathbf{f}}_{p_j}, \mathrm{cov}(\mathbf{f}_{p_j})). \tag{2}$$

The mean and variance of the predictive posterior distribution provide the predictions and their uncertainties, respectively, and are calculated as follows:

$$\bar{\mathbf{f}}_{p_j} = \mathbb{E}[\mathbf{f}_* \mid \mathbf{Z}_p, \mathbf{Y}_p, z_j] = K(z_j, \mathbf{Z}_p)[K(\mathbf{Z}_p, \mathbf{Z}_p) + \sigma_n^2 I]^{-1}\mathbf{Y}_p, \tag{3}$$

$$\mathrm{Var}(\mathbf{f}_{p_j}) = K(z_j, z_j) - K(z_j, \mathbf{Z}_p)[K(\mathbf{Z}_p, \mathbf{Z}_p) + \sigma_n^2 I]^{-1}K(\mathbf{Z}_p, z_j), \tag{4}$$

where σ_n^2 is the additive independent identically distributed Gaussian noise ϵ.

For simplicity, we refer to the predictive mean and predictive variance of a ROI of a test subject j from the p-DKGP as y_p and v_p, respectively. We can reconstruct the ROI trajectory of subject j by prompting the p-DKGP model with different time intervals t.This results in the predicted trajectory and predictive uncertainty across time, represented as $Y_p = (y_{p_1}, y_{p_2}, \ldots, y_{p_T})$ and $V_p = (v_{p_1}, v_{p_2}, \ldots, v_{p_T})$.

2.3 Subject-Specific Model (ss-DKGP)

For a new test subject, let h denote the number of observations and T_{obs} the time of observation from the initial acquisition. We represent the observed data of the subject as $D_s = \{(X_s, M_s, T_s), Y_s\}$. The ss-DKGP model is trained on this dataset D_s to exclusively capture the subject's trajectory. The transformation $\Phi(\cdot; \mathbf{W}_p, \mathbf{b}_p)$, learned via the p-DKGP, is employed to initialize the Deep Kernel of the subject-specific model.

We initialize a new Gaussian Process (GP) with an RBF kernel and a zero mean. During the training of the ss-DKGP, we use only the observed trajectory of the subject. Specifically, we update only the hyperparameters of the new GP, which include the length-scale l_s and the signal variance σ_s, while keeping the weights of the function $\Phi(\cdot; \mathbf{W}_p, \mathbf{b}_p)$ frozen during backpropagation.

For the subject j with input $u_j = (x_j, m_j, t)$, we denote their transformation as $z_j = \Phi(u_j; \mathbf{W}_p, \mathbf{b}_p)$.

The posterior predictive distribution of the ROI function at time point t is:

$$f_{s_j} \mid (\mathbf{Z}_s, \mathbf{Y}_s), z_j \sim \mathcal{N}(\bar{\mathbf{f}}_{s_j}, \mathrm{cov}(\mathbf{f}_{s_j})), \tag{5}$$

where $z_j = \Phi(u_j; \mathbf{W}_p, \mathbf{b}_p)$ and σ_n^2 is the additive independent identically distributed Gaussian noise ϵ.

The predictive mean and variance, representing the predictions and their associated uncertainties respectively, are computed as follows:

$$\bar{\mathbf{f}}_{s_j} = \mathbb{E}[\mathbf{f}_{s_j} \mid \mathbf{Z}_s, \mathbf{Y}_s, z_j] = K(z_j, \mathbf{Z}_s)[K(\mathbf{Z}_s, \mathbf{Z}_s) + \sigma_n^2 I]^{-1}\mathbf{Y}_s, \tag{6}$$

$$\mathrm{Var}(\mathbf{f}_{s_j}) = K(z_j, z_j) - K(z_j, \mathbf{Z}_s)[K(\mathbf{Z}_s, \mathbf{Z}_s) + \sigma_n^2 I]^{-1}K(\mathbf{Z}_s, z_j). \tag{7}$$

For simplicity, we denote the predictive mean and predictive variance of the ss-DKGP as y_{sj} and v_{sj}, respectively. The ROI trajectory of subject j can be reconstructed by querying the ss-DKGP model at different time intervals t.This process yields the predicted trajectory and predictive uncertainty over time, represented as $Y_s = (y_{s_1}, y_{s_2}, \ldots, y_{s_T})$ and $V_s = (v_{s_1}, v_{s_2}, \ldots, v_{s_T})$.

2.4 Predictive Posterior Correction

Given predictions y_p and y_s from two probabilistic models (p-DKGP and ss-DKGP, respectively), the combined prediction is expressed as a linear combination:

$$y_c = \alpha y_p + (1 - \alpha) y_s. \tag{8}$$

Here, α is the shrinkage parameter reflecting the relative confidence in each model. For simplicity, we assume independence between the models. Since the sum of Gaussian random variables remains Gaussian, y_c naturally retains Gaussian properties. Therefore, the combined variance is given by:

$$v_c = \alpha^2 v_p + (1 - \alpha)^2 v_s. \tag{9}$$

The weights α and $1-\alpha$ indicate the credibility of each model, resulting in a new posterior predictive mean Y_c and variance V_c. Values of α close to 1 indicate higher confidence in the p-DKGP model, while values close to 0 demonstrate greater trust in the ss-DKGP model. We refer to α as the shrinkage parameter.

2.5 Shrinkage Estimation: α

Determining the optimal shrinkage parameter α is a critical problem for constructing the personalized posterior predictive mean and variance of the ROI trajectory. The parameter α quantifies the degree of confidence placed in the population-level model versus the subject-specific trajectory.

To estimate α, we use a held-out validation set of subjects, unknown to the population model, with known trajectories. Predictions for these subjects are generated using the p-DKGP model. For each subject, the ss-DKGP component is trained by progressively increasing the length of the observed trajectory, starting with 4 observed ROI values.

We then reconstruct the entire ROI trajectory from the baseline time ($t = 0$) to the last time point of the subject (t_n). Using both models, p-DKGP and ss-DKGP, we obtain two estimates of the ROI trajectory along with their predictive variances. The term "reconstruct" is used here because the ground truth ROI values for the validation subjects are known. Denote the p-DKGP model predictive mean and variance as Y_p and V_p, and the ss-DKGP model predictive mean and variance as Y_s and V_s. Let Y represent the ground truth ROI values over time. Our goal is to estimate the optimal α. To achieve this, we minimize the following criterion:

$$J_{s|h}(\alpha) = \sum_{t=0}^{t_n} \left(y_t - (\alpha \cdot y_{p_t} + (1 - \alpha) \cdot y_{s_t})\right)^2. \tag{10}$$

We use the notation $J_{s|h}$ because we optimize this for a subject s, given h observed acquisitions. The algorithm for calculating the optimal shrinkage estimates on the validation set is outlined in Algorithm 1. The algorithm processes each subject's data individually, applying the optimization to each sequence of observations recorded for the subject. This process is repeated for every subject in the validation set.

In Supplementary Section D.1 we attach detailed theoretical justification on the criterion 10.

Algorithm 1 Shrinkage Estimation

Require: Validation set $V = \{(U^s, Y^{(s)}) \mid s \in S\}$, where $Y^{(s)} = \{y_t^{(s)}\}_{t=1}^T$ is the ground truth trajectory for subject s

Ensure: Optimal shrinkage parameters $\hat{\alpha}_{s,h}$ for each $s \in S$ and $h \in H$

for each $s \in S$ **do**
  Initialize list $L^{(s)} \leftarrow [\,]$
  for each $h \in H$ **do**
    Obtain P-DKGP trajectory: $Y_p^{(s,h)} = \{y_{p,t}^{(s,h)}\}_{t=1}^T$
    Obtain ss-DKGP trajectory: $Y_s^{(s,h)} = \{y_{s,t}^{(s,h)}\}_{t=1}^T$
    Define objective function:

$$J_{s,h}(\alpha) = \sum_{t=1}^{T} \left(y_t^{(s)} - \left(\alpha y_{p,t}^{(s,h)} + (1-\alpha) y_{s,t}^{(s,h)} \right) \right)^2$$

    Compute:

$$\hat{\alpha}_{s,h} = \arg\min_{\alpha \in [0,1]} J_{s,h}(\alpha)$$

    Append $\hat{\alpha}_{s,h}$ to $L^{(s)}$
  end for
  Store list $L^{(s)}$ for subject s
end for

2.6 Learning the shrinkage α

The shrinkage parameter α represents the trust factor between the two components (p-DKGP, ss-DKGP), and we model it as a function of the input variables $q = \{y_p, y_s, v_p, v_s, T_{\text{obs}}\}$, where $q \in \mathbb{R}^5$ and T_{obs} represents the time of the last observed acquisition from the baseline. We have collected all the optimal values of α from the Shrinkage Estimation (Section 2.5) in the validation set. Our objective is to learn a mapping function, with parameters θ, that transforms the input space $Q = \{Y_p, Y_s, V_p, V_s, T_{\text{obs}}\}$, where $Q \in \mathbb{R}^5$, to the output space $A \in \mathbb{R}$, as $\hat{\alpha} = g(Q; \theta)$.

We employ XGBoost regression to find the function g that minimizes the difference between the predicted α and the actual values. We denote the learned function as g_α. In the Supplementary (Section D.2.1), we provide results from additional models we experimented with. Our experiments demonstrate that the best α function is the XGBoost.

2.7 Personalization through Adaptive Shrinkage

For a new test subject with h observations and T_{obs} the time of the last observation from the baseline, we train the ss-DKGP model with the process described in Section 2.3. We call the posterior corrected predictive distribution as pers-DKGP. The Personalization through α-Shrinkage is described

Algorithm 2 Personalization through α-Shrinkage

Require: p-DKGP model, ss-DKGP model, and f_α

Ensure: Adapted predictive mean and variance: Y_c, V_c

Y_p, V_p trajectory from p-DKGP model
Y_s, V_s trajectory from ss-DKGP model
$\alpha_h = g_\alpha(Y_p, Y_s, V_p, V_s, T_{obs})$
$Y_c = \alpha_h \cdot Y_p + (1 - \alpha_h) \cdot Y_s$
$V_c = \alpha_h^2 \cdot V_p + (1 - \alpha_h)^2 \cdot V_s$
return Y_c, V_c

in algorithm 2.

3 RESULTS

3.1 DATASETS

We use T1-weighted (T1w) MR imaging (MRI) and clinical variables from the iSTAGING (Imaging-based coordinate SysTem for AGIng and NeurodeGenerative diseases) consortium for cognitive impairment and dementia (Habes et al., 2021). All MRIs are corrected for intensity inhomogeneities (Sled et al., 1998). Then, 145 ROIs of brain volumes are extracted using a multi-atlas label fusion method (Doshi et al., 2016) and harmonized (Pomponio et al., 2020) to remove site effects. Prior to incorporating the ROIs into our model, we normalize them. From iSTAGING, we extract data from five longitudinal studies—ADNI (Weiner et al., 2017), BLSA (Ferrucci, 2008), OASIS, AIBL (LaMontagne et al., 2019), PreventAD (Tremblay-Mercier et al., 2021). ADNI and BLSA provide a cohort of total 2,200 subjects. The 1560 subjects are used for training the p-DKGP model, the 200 subjects are used for learning the Adaptive Shrinkage function and the rest 440 subjects are used for testing any personalization approach. To ensure generalizability, OASIS, AIBL and PreventAD studies, are exclusively reserved as external clinical studies for testing. Please see Section B.1 for a comprehensive description of the datasets.

3.2 EVALUATION METRICS AND BASELINES

We evaluate model performance from two perspectives: predictive accuracy and uncertainty quantification (UQ). Predictive accuracy is assessed using Absolute Error (AE) and Mean Absolute Error (MAE) per subject. For uncertainty quantification, we compute interval width and coverage, defined as the range between two standard deviations above and below the predictive mean and the proportion of true values that fall within this range, respectively. Effective UQ is characterized by narrow intervals and high coverage, reflecting confidence in the predictions without sacrificing accuracy. Importantly, all evaluation metrics—including AE, interval width, and coverage—are calculated over the entire unseen trajectory of the test subjects, rather than focusing on the next time point alone. This allows for a more comprehensive assessment of model generalization over time.

We benchmark our method against several baselines: Linear Mixed Effects (LMM) models, Generalized Additive Models (GAMs), Deep Regression, and the Deep Mixed Effects (DME) method (Chung et al., 2019). Further details on the baselines are provided in Supplementary E.2.

3.3 EVALUATION OF PREDICTED TRAJECTORIES OF ATROPHY ON BRAIN ROIS

In this section, we evaluate the Adaptive Shrinkage Estimation on 6 Brain ROIs: Hippocampus (R/L), Thalamus, Amygdala, Parahippocampal Gyrus, and Lateral Ventricle. The evaluation of the predictive performance is performed on the 440 test subjects from the ADNI and BLSA cohorts. For details on the training of the ROI deep kernel models (p-DKGP and ss-DKGP) please see Section C.3.

Our model demonstrates the capability to accurately predict long-term longitudinal trajectories. In Figure 2**(a)**, we visualize the mean AE across different lengths of observed trajectories and plot the errors relative to the time from the last observation. Our model achieves progressively lower Mean AE over time, indicating improved precision in both long-term and short-term predictions. We also present the Mean MAE across progression status for the 6 ROIs in Figure 2**(a)**. Notably, the largest Mean MAE difference between the two models was observed in participants with AD and AD progressors, who exhibit non-linear and steeper trends that the competing baselines where not able to capture. Specifically, the LMM that is only constrained to linear patterns in ROI volumes evidences by the high percentage mean AE differences in AD 177.66% and AD progressors 22.05%. In healthy controls, the LMM also shows critical percentage difference from our method, with a percentage MAE difference of 29.78%. This indicates that even for subjects with relatively stable brain atrophy, the LMM is unable to capture trajectories.

We proceed with qualitative evaluation of the predicted trajectories. The efficacy of our model is further illustrated by the predicted trajectories in Figure 3. For the brain ROIs of Hippocampus R, Thalamus Proper R, and Lateral Ventricle, our model successfully adapts to the observed trajectories of test subjects, resulting in more accurate long-term predictions. Specifically, for the second subject in Figure 3, the population prediction significantly deviates from the actual trajectory. However, as the number of observations increases, the pers-DKGP trajectory shifts towards the observed trajectory, adapting to the subject-specific trend. The third subject in this panel, a progressor from Mild Cognitive Impairment (MCI) to Alzheimer's Disease (AD), shows a more abrupt increase in

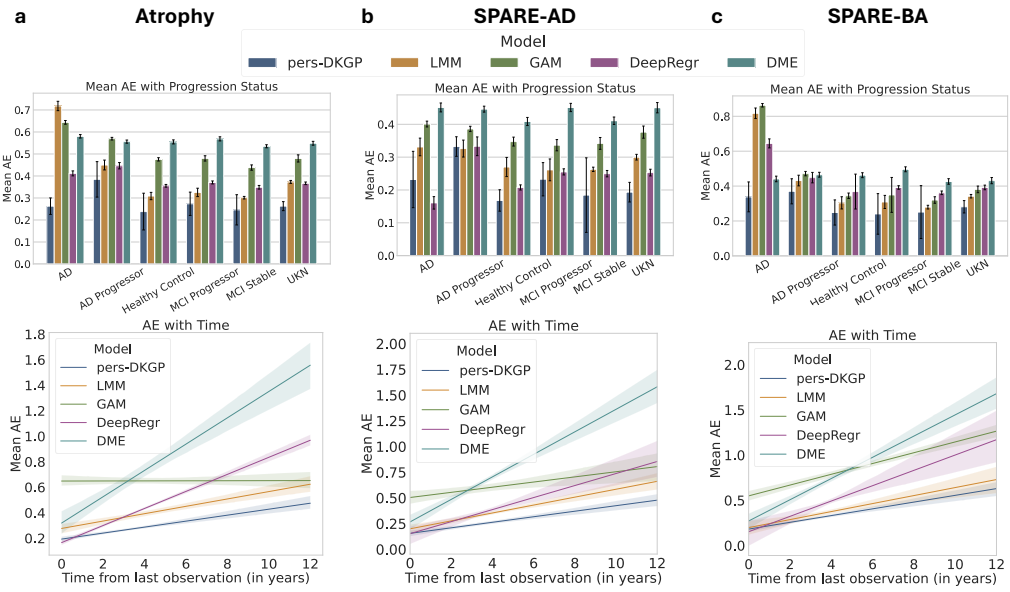

Figure 2: Quantitative Comparison with Baseline methods: Mean AE across all ROIs categorized by progression status (top). Error bars represent 95% confidence intervals and Mean AE over predicted time horizon from last observation (bottom) for: **(a)** Atrophy on 6 ROIs, **(b)** SPARE-AD and **(c)** SPARE-BA scores. Our method is denoted as pers-DKGP.

ventricular volume. This trend is captured with few observations by the pers-DKGP model, while the Linear Mixed Model (LMM) underestimates the ventricular volume in the long term.

Overall, the LMM exhibits limited flexibility in capturing non-linear patterns in ROI volumes, rendering it inadequate for long-term trend prediction. While it performs reasonably well in short-term forecasts and lower-dimensional settings, its expressiveness falls short for complex, high-dimensional inputs. Deep regression, though capable of learning from observed data, often yields non-smooth or non-monotonic trajectories that deviate from biologically plausible biomarker progression trends. The deep mixed effects model, with a shared deep mean function and subject-specific Gaussian processes, struggles to achieve personalization in high-dimensional input spaces, with persistent errors across time and diagnostic categories. These issues stem from the limitations of the RBF kernel in managing multivariate, high-dimensional data. In contrast, our method effectively approximates non-linear mixed effects models, demonstrating flexibility in handling multivariate, high-dimensional data and capturing diverse temporal patterns. Through Adaptive Shrinkage of the two GP probabilistic models, it consistently produces smooth, biologically plausible trajectories, delivering robust predictions for neuroimaging biomarkers and brain ROIs.

3.4 Application to Neuroimaging Biomarkers: SPARE Scores

Having shown that our framework effectively adapts longitudinal predictions of atrophy on Brain ROIs on Section 3.3, we also demonstrate that our method is applicable to other neuroimaging biomarkers apart from atrophy. Specifically, we apply it to composite biomarkers, the SPARE-AD (Davatzikos et al., 2009) and SPARE-BA (Habes et al., 2016) scores. SPARE-AD quantifies atrophy patterns associated with Alzheimer's disease, focusing on regions vulnerable to early AD pathology. SPARE-BA reflects brain aging patterns, capturing structural decline typically linked to aging. By modeling these biomarkers, we explore the intersection between pathological aging, as seen in AD, and normative brain aging trajectories. We use the population dataset from ADNI and BLSA so as to train the p-DKGP and the held-out validation dataset to train the Adaptive Shrinkage function for the SPARE-AD and SPARE-BA scores. Details on these models are attached on the Supplementary Section C.4. The evaluation of the Adaptive Shrinkage function for SPARE scores is done on the ADNI and BLSA 440 test subjects.

Our model demonstrates robust performance in predicting long-term longitudinal trajectories for both SPARE-AD and SPARE-BA biomarkers, as illustrated in Figure 2**(b, c)**. Notably, the model

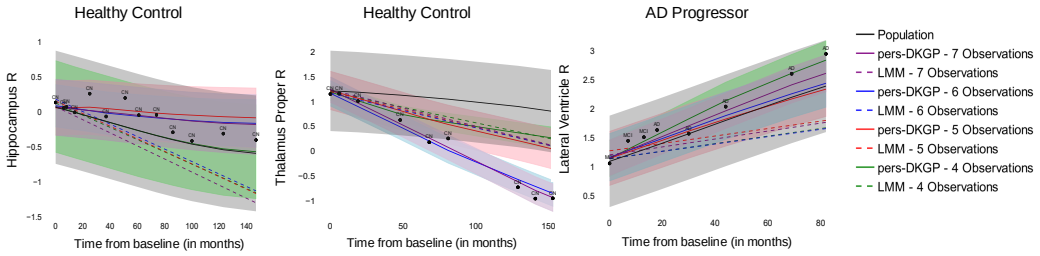

Figure 3: Adaptation of the predicted personalized ROI trajectories for three different test subjects, going from using 4 to 7 observations as training to learn the optimal shrinkage parameters. The dashed lines are the results using LMM as a comparison. The first two panels show the Hippocampus and Right Thalamus Proper volumes from two stable Cognitively Normal individuals. Last panel shows the Right Lateral Ventricle volumes from one subject who progressed from MCI to AD. The shaded bands represent the predictive uncertainty over time.

achieves progressively lower Mean Absolute Error (MAE) over time, indicating improved precision in forecasting long-term outcomes. Compared to other approaches, the LMM is the second most competitive in the majority of the diagnosis status. For SPARE-BA, model performance differences are minimal in stable subjects and healthy controls, but more pronounced in Alzheimer's disease (AD) subjects, where SPARE-BA exhibit steeper progression trends due to accelerated brain aging.

With this experiment, we demonstrate that our method effectively models the progression of composite neuroimaging longitudinal biomarkers using high-dimensional multivariate inputs for subjects with follow-up biomarker measurements.

3.5 Generalization to External Clinical Studies

In this section, we demonstrate the **generalizability** of our method to previously unseen clinical neuroimaging datasets. After training the p-DKGP and Adaptive Shrinkage function on the population and validation datasets from the ADNI and BLSA cohorts, we personalize from the second follow-up point on each subject and predict the rest of the trajectory. This process is repeated for all follow-up points, leaving the very last follow-up for testing purposes. We assess the performance of our framework on subjects from three independent clinical studies: OASIS, AIBL, and PreventAD.

These datasets differ from the training population. The variations in demographics, diagnosis composition, site, and follow-up intervals across the three studies underscore the challenge of generalizing across diverse clinical populations. Thus, they provide a robust test of the model's generalizability. The differences in demographics and follow-up intervals between the three studies emphasize the challenge of generalizing across diverse clinical populations:

- **AIBL**: Includes individuals with a mean age of 75 years, which is slightly older than the ADNI cohort and comparable to BLSA. It is predominantly composed of Alzheimer's disease (AD) patients (91 subjects), followed by MCI and controls. On average, each subject has approximately 3 follow-up visits, with a mean interval of 24 months between visits.
- **OASIS**: Includes individuals younger on average (67.8 years) compared to both ADNI and BLSA. It is primarily composed of healthy controls, with smaller representations of MCI and AD cases. The average number of follow-ups is ~ 3 per subject, with a mean interval of 32 months.
- **PreventAD**: PreventAD focuses on pre-symptomatic early detection of AD in a healthier and younger population (mean age 65.3 years) with an average of 4 follow-up visits per subject and a shorter mean interval of 10 months.

Our method demonstrates consistent superiority over baseline methods across three independent clinical studies—AIBL, OASIS, and PreventAD—underscoring its robustness and reliability in diverse real-world scenarios (Figure 4). The model achieves significantly lower Mean Absolute Error (Mean AE) compared to traditional baselines, with narrow confidence intervals reflecting its precision and stability.

In the AIBL study, pers-DKGP achieves a Mean AE of *0.197 ± 0.009*, substantially outperforming the baselines. This trend is similarly observed in the OASIS study, where pers-DKGP attains a Mean AE of *0.259 ± 0.006*. Notably, in the PreventAD study, our method achieves the lowest Mean AE of *0.139 ± 0.004*, outperforming LMM and GAM. The narrow CIs associated with pers-DKGP across all datasets highlight its reliability, offering consistently precise predictions that are resilient to data variability.

The superior performance of our method is attributed to its ability to effectively combine population-level and subject-specific predictions via the Adaptive Shrinkage function, enabling it to generalize across diverse neuroimaging populations. These results establish our model as a robust and reliable framework for personalized forecasting of neurodegeneration, with significant potential for deployment in clinical trials and neuroimaging studies.

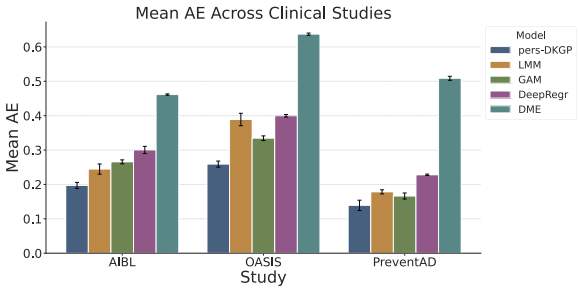

Figure 4: Comparison of Mean Absolute Errors (MAE) and 95% confidence intervals across three independent external clinical studies, evaluating the proposed method against competing baselines.

3.6 Explaining Adaptive Shrinkage: An Ablation Study on the α Function

In this section, we aim to demonstrate the **effectiveness** and **interpretability** of the Adaptive Shrinkage function. First, we compare it to other posterior correction approaches. Then, through explainability analysis, we illustrate how Adaptive Shrinkage learns in a data-driven manner to optimally balance the two posterior predictive distributions, making its decision-making process intuitive.

We explore different strategies for selecting the α parameter to evaluate their performance in the personalization process. Initially, we experiment with a constant $\alpha = c$, where $c \in (0, 1)$. This approach serves as an **uninformative** way to correct the posterior. Next, we employ a **semi-informative** (deterministic) approach, where we determine the optimal α for each test subject by optimizing the objective in Equation 10 using only the observed trajectory. Finally, we use our **Adaptive Shrinkage** method (Section 2.6) to determine α.

We conduct this experiment for seven distinct ROIs: Hippocampus R/L, Lateral Ventricle, Thalamus Proper, Amygdala R/L, and the Parahippocampal Gyrus (PHG R). The deterministic method result in the worst outcomes in terms of both predictive performance and uncertainty quantification. This suggests that the observed trajectory of the subject alone is insufficient to determine the optimal α correction for future predictions. Herein, we present the results for Hippocampus R, Lateral Ventricle, and Thalamus Proper under the two settings of constant α and the XGBoost α. The results for the remaining ROIs are provided in Table 6 of the Supplementary E.1 for completeness. In the constant α section of Table 1, we present the performance of the best α values. This demonstrates that optimal performance is not achieved through simple averaging, and that the optimal α for each ROI can vary significantly. For example, for Hippocampus the optimal α is 0.5 while for Ventricles and Thalamus is 0.3 and 0.7 respectively.The results in Table 1 show that the best constant α values for Hippocampus R, Lateral Ventricle, and Thalamus Proper differ from each other.This variation highlights the inadequacy of a one-size-fits-all approach and underscores the necessity for a more sophisticated method. The presented evidence suggests that Adaptive Shrinkage α estimation is a more informed approach for determining the optimal shrinkage parameted, leading to improved predictive performance and uncertainty quantification.

To elucidate the decision-making process of Adaptive Shrinkage, we conducted explainability analysis. We focus on the impact of each input variable—y_{pp}, y_{ss}, V_{pp}, V_{ss}, and T_{obs}—and their interactions on the prediction of the shrinkage parameter α. Specifically, we aimed to understand how the deviation between the population and subject-specific predictive means ($\delta_y = y_{\text{pp}} - y_{\text{ss}}$) and the observation time T_{obs} influence the model's predictions.

Table 1: Ablation on α functions

	Best Constant				XGB		
ROI	α	Mean AE (CI)	Mean Cov.	Mean Int.	Mean AE (CI)	Mean Cov.	Mean Int.
Hippocampus R	0.5	0.257 (±0.007)	0.808	0.843	**0.243** (±0.003)	0.795	0.902
Lateral Ventricle R	0.3	0.143 (±0.006)	0.853	0.507	**0.131** (±0.002)	0.855	0.626
Thalamus Proper R	0.7	0.241 (±0.007)	0.934	1.127	**0.219** (±0.003)	0.849	0.911

We employed SHAP (SHapley Additive exPlanations) values (Lundberg & Lee, 2017) to interpret the contribution of each feature to individual predictions. Figure 8 in Supplementary D.3 reveals that T_{obs} is the variable that most affects the decision-making process. This is further validated by observing that the distribution of the predicted α decreases as we increase the number of follow-up observations, and thus T_{obs}. In Figure 7 in Supplementary D.3, we demonstrate the distribution of α with the number of observations for the six ROIs and the SPARE scores, as well as the predicted α values obtained from the application to external studies. The consistent trend of decreasing α as the number of observations increases highlights the biomarker-agnostic ability of Adaptive Shrinkage to optimally combine population and subject-specific trends. Also, the same behavior of alpha as the number of observations increases, highlights the consistent behavior of the Adaptive Shrinkage function to external clinical studies.

Additionally, correlation analysis (Supplementary D.3, Table 5) reveals that the model consistently demonstrates a negative relationship between the observation time T_{obs} and the predicted α when the deviation δ_y is large. This indicates that, in the presence of significant deviations between the two predictors, Adaptive Shrinkage decreases the weight assigned to the population-level model (p-DKGP) for longer observation periods. This aligns with the intuition that the more follow-up observations we have for a subject, the more we trust their subject-specific predictive distribution.

4 Discussion

In this study, we introduce Adaptive Shrinkage Estimation for personalized Deep Kernel Regression via posterior correction. We show that our method learns a shrinkage parameter that effectively combines the two posterior predictive distributions, adapting the predictive trajectory to each subject's follow-up acquisitions. We evaluate the predictive performance of our model across multiple scenarios, demonstrating superior accuracy over baselines in both short-term and long-term predictions, and across all diagnosis statuses from Healthy Controls to AD Progressors. Additionally, we show that our method generalizes to other neuroimaging biomarkers. It effectively models the progression of monotonic longitudinal biomarkers using high-dimensional multivariate inputs in subjects with any follow-up measurements. The accurate prediction of SPARE-AD and SPARE-BA trajectories suggests broader applicability to biomarkers such as cognitive measures (MMSE, ADAS-Cog13) and blood biomarkers (Amyloid-β, Tau protein).

Importantly, our approach exhibits strong generalization capabilities when applied to external clinical studies with diverse demographics and follow-up intervals. Across the OASIS, AIBL, and PreventAD datasets, our model consistently outperforms baseline methods, highlighting its robustness and adaptability in varying clinical settings. This generalization is particularly valuable for real-world clinical scenarios, where models must perform reliably across different populations.

However, we acknowledge some limitations in our approach. In the posterior correction part of our algorithm (Section 2.4), we assumed independence between the α parameter and the posterior distributions. While this assumption simplifies the correction process, as we elaborate in Supplementary Section D.1.4, it affects only the uncertainty quantification and leaves the posterior-corrected predictive mean unaffected. This ensures that our method remains accurate in its primary predictions.

Our method's optimal and intuitive design positions it as a valuable tool for clinical trial design, disease progression modeling, treatment effect estimation, and clinical neuroimaging research. The use of personalized predicted brain ROIs and neuroimaging biomarkers, such as SPARE-AD, as endpoint for selecting trial subjects, showcases its potential for real-world application. The Adaptive Shrinkage mechanism balances population-level and subject-specific insights, ensuring both accuracy and clinical relevance. With robust performance across diverse external datasets and varying clinical settings, our method stands out as a reliable, generalizable, and transparent solution for advancing clinical research in neurodegeneration.

REFERENCES

Marilyn Albert, Yishan Zhu, Abhay Moghekar, Susumu Mori, Michael I. Miller, Anja Soldan, Corinne Pettigrew, Ola Selnes, Shanshan Li, and Marie-Ann C. Wang. Predicting progression from normal cognition to mild cognitive impairment for individuals at 5 years. *Brain: A Journal of Neurology*, 141:877–887, 2018. doi: 10.1093/brain/awx365.

Anna Caroli and Giovanni B. Frisoni. The dynamics of alzheimer's disease biomarkers in the alzheimer's disease neuroimaging initiative cohort. *Neurobiology of Aging*, 31:1263–1274, 2010. doi: 10.1016/j.neurobiolaging.2010.04.024.

Tianqi Chen and Carlos Guestrin. Xgboost: A scalable tree boosting system. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 785–794. ACM, 2016.

Ingyo Chung, Saehoon Kim, Juho Lee, Kwang Joon Kim, Sung Ju Hwang, and Eunho Yang. Deep mixed effect model using gaussian processes: A personalized and reliable prediction for healthcare, 2019. URL `https://arxiv.org/abs/1806.01551`.

Jeffrey Cummings, Carla Reiber, and Pranesh Kumar. The price of progress: funding and financing alzheimer's disease drug development. *Alzheimer's Dementia: Translational Research Clinical Interventions*, 4:330–343, 2018. doi: 10.1016/j.trci.2018.04.008.

Christos Davatzikos, Feng Xu, Yang An, Yong Fan, and Susan M. Resnick. Longitudinal progression of alzheimer's-like patterns of atrophy in normal older adults: the spare-ad index. *Brain*, 132(8): 2026–2035, August 2009. doi: 10.1093/brain/awp091. URL `https://doi.org/10.1093/brain/awp091`.

Yu Ding, Jee H. Sohn, Matthew G. Kawczynski, Harish Trivedi, Robert Harnish, Nicholas W. Jenkins, Denis Lituiev, Tyler P. Copeland, Mariam S. Aboian, Celina Mari Aparici, Sterling C. Behr, Richard R. Flavell, Shuo-Yen Huang, Kyle A. Zalocusky, Luca Nardo, Youngho Seo, Richard A. Hawkins, Miguel Hernandez Pampaloni, David Hadley, and Benjamin L. Franc. A deep learning model to predict a diagnosis of alzheimer's disease by using 18f-fdg pet of the brain. *Radiology*, 290:456–464, 2018. doi: 10.1148/radiol.2018180958.

Jimit Doshi, Guray Erus, Yangming Ou, Susan M. Resnick, Ruben C. Gur, Raquel E. Gur, Theodore D. Satterthwaite, Susan Furth, and Christos Davatzikos. Muse: Multi-atlas region segmentation utilizing ensembles of registration algorithms and parameters, and locally optimal atlas selection. *NeuroImage*, 127:186–195, 2 2016. ISSN 10959572. doi: 10.1016/j.neuroimage.2015.11.073. URL `https://doi.org/10.1016/j.acra.2013.09.010`.

S. Duchesne, A. Caroli, C. Geroldi, D.L. Collins, and G.B. Frisoni. Relating one-year cognitive change in mild cognitive impairment to baseline mri features. *NeuroImage*, 47:1363–1370, 2009. doi: 10.1016/j.neuroimage.2009.04.023.

S. Eleftheriadis, O. Rudovic, M.P. Deisenroth, and M. Pantic. Gaussian process domain experts for modeling of facial affect. *IEEE Transactions on Image Processing*, 26(10):4697–4711, 2017. URL `https://ibug.doc.ic.ac.uk/media/uploads/documents/07961231.pdf`.

Luigi Ferrucci. The baltimore longitudinal study of aging (blsa): A 50-year-long journey and plans for the future. *The Journals of Gerontology: Series A: Biological Sciences and Medical Sciences*, 63(12):1416–1419, 2008. doi: 10.1093/gerona/63.12.1416.

Pedro J. García-Laencina, José-Luis Sancho-Gómez, and Aníbal R. Figueiras-Vidal. Pattern classification with missing data: a review. *Neural Computing and Applications*, 19:263–282, 2010. doi: 10.1007/s00521-009-0295-6.

Samuel Gruffaz, Pierre-Emmanuel Poulet, Etienne Maheux, Bruno Jedynak, and Stanley Durrleman. Learning riemannian metric for disease progression modeling. In *Advances in Neural Information Processing Systems*, 2021. URL `https://proceedings.neurips.cc/paper/2021/hash/c7b90b0fc23725f299b47c5224e6ec0d-Abstract.html`.

M. Habes, D. Janowitz, G. Erus, and et al. Advanced brain aging: relationship with epidemiologic and genetic risk factors, and overlap with alzheimer disease atrophy patterns. *Translational Psychiatry*, 6(4):e775, 2016. doi: 10.1038/tp.2016.39. URL https://doi.org/10.1038/tp.2016.39. Published 2016 Apr 5.

Mohamad Habes, Raymond Pomponio, Haochang Shou, Jimit Doshi, Elizabeth Mamourian, Guray Erus, Ilya Nasrallah, Lenore J Launer, Tanweer Rashid, Murat Bilgel, et al. The brain chart of aging: machine-learning analytics reveals links between brain aging, white matter disease, amyloid burden, and cognition in the istaging consortium of 10,216 harmonized mr scans. *Alzheimer's & Dementia*, 17(1):89–102, 2021.

S. Hochreiter and J. Schmidhuber. Long short-term memory. *Neural Computation*, 9(8):1735–1780, 1997.

K. Ito, S. Ahadieh, B. Corrigan, J. French, T. Fullerton, T. Tensfeldt, and Alzheimer's Disease Working Group. Disease progression meta-analysis model in alzheimer's disease. *Alzheimer's Dementia: The Journal of the Alzheimer's Association*, 6:39–53, 2010. doi: 10.1016/j.jalz.2009.05.665.

Clifford R. Jack, David S. Knopman, William J. Jagust, Leslie M. Shaw, Paul S. Aisen, Michael W. Weiner, Ronald C. Petersen, and John Q. Trojanowski. Hypothetical model of dynamic biomarkers of the alzheimer's pathological cascade. *Lancet Neurology*, 9:119–128, 2010. doi: 10.1016/S1474-4422(09)70299-6.

W. James and C. Stein. Estimation with quadratic loss. In *Proceedings of the Third Berkeley Symposium on Mathematics. Statistics and Probability*, volume 1, 1961.

J. K. Johnson, A. L. Gross, J. Pa, D. G. McLaren, L. Q. Park, and J. J. Manly. Longitudinal change in neuropsychological performance using latent growth models: a study of mild cognitive impairment. *Brain Imaging and Behavior*, 6:540–550, 2012. doi: 10.1007/s11682-012-9161-8.

Diederik P. Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *CoRR*, abs/1412.6980, 2014. URL https://api.semanticscholar.org/CorpusID:6628106.

Igor Koval, Alexandre Bône, Maxime Louis, Thomas Lartigue, Simona Bottani, Arnaud Marcoux, Jorge Samper-González, Ninon Burgos, Benjamin Charlier, Anne Bertrand, et al. Ad course map charts alzheimer's disease progression. *Scientific Reports*, 2021. Available at Inria - Institut national de recherche en sciences et technologies du numérique (HAL).

Pamela J. LaMontagne, Tammie LS. Benzinger, John C. Morris, Sarah Keefe, Russ Hornbeck, Chengjie Xiong, Elizabeth Grant, Jason Hassenstab, Krista Moulder, Andrei G. Vlassenko, Marcus E. Raichle, Carlos Cruchaga, and Daniel Marcus. Oasis-3: Longitudinal neuroimaging, clinical, and cognitive dataset for normal aging and alzheimer disease. *medRxiv*, 2019. doi: 10.1101/2019.12.13.19014902. URL https://www.medrxiv.org/content/early/2019/12/15/2019.12.13.19014902.

Bo Lei, Pan Yang, Tao Wang, Shuyuan Chen, and Dong Ni. Relational-regularized discriminative sparse learning for alzheimer's disease diagnosis. *IEEE Transactions on Cybernetics*, 47:1102–1113, 2017. doi: 10.1109/TCYB.2016.2644718.

M.A. Lindquist and A. Gelman. Correlations and multiple comparisons in functional imaging: a statistical perspective (commentary on vul et al., 2009). *Perspectives on Psychological Science*, 4 (3):310–313, 2009.

B. Liu and N. Vasconcelos. Bayesian model adaptation for crowd counts. In *Proceedings of the IEEE International Conference on Computer Vision*, pp. 4175–4183, 2015. URL https://openaccess.thecvf.com/content_iccv_2015/papers/Liu_Bayesian_Model_Adaptation_ICCV_2015_paper.pdf.

Mingxia Liu, Junhao Zhang, Ehsan Adeli, and Dinggang Shen. Joint classification and regression via deep multi-task multi-channel learning for alzheimer's disease diagnosis. *IEEE Transactions on Biomedical Engineering*, 66:1195–1206, 2019. doi: 10.1109/TBME.2018.2869989.

Scott M Lundberg and Su-In Lee. A unified approach to interpreting model predictions. In I. Guyon, U. Von Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett (eds.), *Advances in Neural Information Processing Systems*, volume 30. Curran Associates, Inc., 2017. URL https://proceedings.neurips.cc/paper_files/paper/2017/file/8a20a8621978632d76c43dfd28b67767-Paper.pdf.

Etienne Maheux, Igor Koval, Juliette Ortholand, Colin Birkenbihl, Damiano Archetti, Vincent Bouteloup, Stéphane Epelbaum, Carole Dufouil, Martin Hofmann-Apitius, and Stanley Durrleman. Forecasting individual progression trajectories in alzheimer's disease. *Nature Communications*, 14(1):1–15, 2023. doi: 10.1038/s41467-022-35712-5.

Razvan V. Marinescu, Neil P. Oxtoby, Alexandra L. Young, Esther E. Bron, Arthur W. Toga, Michael W. Weiner, Frederik Barkhof, Nick C. Fox, Stefan Klein, Daniel C. Alexander, and the EuroPOND Consortium. Tadpole challenge: Prediction of longitudinal evolution in alzheimer's disease, 2018. URL https://arxiv.org/abs/1805.03909.

John J. McArdle, Brent J. Small, Lars Bäckman, and Laura Fratiglioni. Longitudinal models of growth and survival applied to the early detection of alzheimer's disease. *Journal of Geriatric Psychiatry and Neurology*, 2016. doi: 10.1177/0891988705281879.

R Mohs, D Knopman, R Petersen, SH Ferris, C Ernesto, M Grundman, M Sano, L Bieliauskas, D Geldmacher, C Clark, et al. Development of cognitive instruments for use in clinical trials of antidementia drugs: additions to the alzheimer's disease assessment scale that broaden its scope. *Alzheimer Disease and Associated Disorders*, 1997.

Elaheh Moradi, Alessandro Pepe, Christian Gaser, Heikki Huttunen, Jussi Tohka, and Alzheimer's Disease Neuroimaging Initiative. Machine learning framework for early mri-based alzheimer's conversion prediction in mci subjects. *Neuroimage*, 104:398–412, 2015. doi: 10.1016/j.neuroimage.2014.10.002.

Minh Nguyen, Tong He, Lijun An, Daniel C. Alexander, Jiashi Feng, and B.T. Thomas Yeo. Predicting alzheimer's disease progression using deep recurrent neural networks. *bioRxiv*, pp. 755058, 2019. doi: 10.1101/755058.

Raymond Pomponio, Guray Erus, Mohamad Habes, Jimit Doshi, Dhivya Srinivasan, Elizabeth Mamourian, Vishnu Bashyam, Ilya M Nasrallah, Theodore D Satterthwaite, Yong Fan, et al. Harmonization of large mri datasets for the analysis of brain imaging patterns throughout the lifespan. *NeuroImage*, 208:116450, 2020.

Carl Edward Rasmussen and Christopher K. I. Williams. *Gaussian Processes for Machine Learning*. MIT Press, 2006.

Ognjen (Oggi) Rudovic, Yuria Utsumi, Ricardo Guerrero, Kelly Peterson, Daniel Rueckert, and Rosalind W. Picard. Meta-weighted gaussian process experts for personalized forecasting of ad cognitive changes. In Finale Doshi-Velez, Jim Fackler, Ken Jung, David Kale, Rajesh Ranganath, Byron Wallace, and Jenna Wiens (eds.), *Proceedings of the 4th Machine Learning for Healthcare Conference*, volume 106 of *Proceedings of Machine Learning Research*, pp. 181–196. PMLR, 09–10 Aug 2019. URL https://proceedings.mlr.press/v106/rudovic19a.html.

Mert R. Sabuncu, Rahul S. Desikan, Jorge Sepulcre, B. T. Thomas Yeo, Hesheng Liu, Nicholas J. Schmansky, Martin Reuter, Michael W. Weiner, Randy L. Buckner, Reisa A. Sperling, and Bruce Fischl. The dynamics of cortical and hippocampal atrophy in alzheimer disease. *Archives of Neurology*, 68:1040–1048, 2011. doi: 10.1001/archneurol.2011.167.

Mert R. Sabuncu, Juan L. Bernal-Rusiel, Martin Reuter, Douglas N. Greve, and Bruce Fischl. Event time analysis of longitudinal neuroimage data. *Neuroimage*, 97:9–18, 2014. doi: 10.1016/j.neuroimage.2014.04.015.

Mahesh N. Samtani, Martha Farnum, Valeriy Lobanov, Eun Yang, Nandini Raghavan, Abbey DiBernardo, and Vaibhav Narayan. An improved model for disease progression in patients from the alzheimer's disease neuroimaging initiative. *Journal of Clinical Pharmacology*, 52:629–644, 2012. doi: 10.1177/0091270010390662.

H. Shou, A. Eloyan, M.B. Nebel, A. Mejia, J.J. Pekar, S. Mostofsky, B. Caffo, M.A. Lindquist, and C.M. Crainiceanu. Shrinkage prediction of seed-voxel brain connectivity using resting state fmri. *NeuroImage*, 102(Pt 2):938–944, 2014.

John G. Sled, Alex P. Zijdenbos, and Alan C. Evans. A nonparametric method for automatic correction of intensity nonuniformity in mri data. *IEEE Transactions on Medical Imaging*, 17:87–97, 1998. ISSN 02780062. doi: 10.1109/42.668698.

Cynthia M. Stonnington, Ching Chu, Stefan Klöppel, Clifford R. Jack, John Ashburner, Richard S.J. Frackowiak, and Alzheimer's Disease Neuroimaging Initiative. Predicting clinical scores from magnetic resonance scans in alzheimer's disease. *Neuroimage*, 51:1405–1413, 2010. doi: 10.1016/j.neuroimage.2010.03.051.

Rania Sukkar, Elliot Katz, Yuan Zhang, David Raunig, and Bradley T. Wyman. Disease progression modeling using hidden markov models. In *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, IEEE Engineering in Medicine and Biology Society Annual Conference, pp. 2845–2848, 2012. doi: 10.1109/EMBC.2012.6346556.

Robert Tibshirani. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B (Methodological)*, 58(1):267–288, 1996.

J. Tremblay-Mercier, C. Madjar, S. Das, et al. Open science datasets from prevent-ad, a longitudinal cohort of pre-symptomatic alzheimer's disease. *NeuroImage: Clinical*, 31:102733, 2021. doi: 10.1016/j.nicl.2021.102733.

Prashanthi Vemuri, Heather J. Wiste, Stephen D. Weigand, Leslie M. Shaw, John Q. Trojanowski, Michael W. Weiner, David S. Knopman, Ronald C. Petersen, Clifford R. Jack, and Alzheimer's Disease Neuroimaging Initiative. Mri and csf biomarkers in normal, mci, and ad subjects: predicting future clinical change. *Neurology*, 73:294–301, 2009. doi: 10.1212/WNL.0b013e3181af79fb.

Heng Wang, Feiping Nie, Heng Huang, Jun Yan, Sungeun Kim, Shannon Risacher, Andrew Saykin, and Li Shen. High-order multi-task feature learning to identify longitudinal phenotypic markers for alzheimer's disease progression prediction. In F. Pereira, C.J.C. Burges, L. Bottou, and K.Q. Weinberger (eds.), *Advances in Neural Information Processing Systems 25*, pp. 1277–1285. Curran Associates, Inc., 2012.

Xiao Wang, Dinggang Shen, and Heng Huang. Prediction of memory impairment with mri data: a longitudinal study of alzheimer's disease. In *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2016*, volume 9900 of *Lecture Notes in Computer Science*, pp. 273–281. Springer, 2016. doi: 10.1007/978-3-319-46720-7_32.

Michael W. Weiner, Dallas P. Veitch, Paul S. Aisen, Laurel A. Beckett, Nigel J. Cairns, Robert C. Green, Danielle Harvey, Clifford R. Jack, William Jagust, John C. Morris, Ronald C. Petersen, Andrew J. Saykin, Leslie M. Shaw, Arthur W. Toga, and John Q. Trojanowski. Recent publications from the alzheimer's disease neuroimaging initiative: Reviewing progress toward improved ad clinical trials. *Alzheimer's and Dementia*, 13:e1–e85, 4 2017. ISSN 15525279. doi: 10.1016/j.jalz.2016.11.007. URL https://doi.org/10.1016/j.jalz.2016.11.007.

Andrew Gordon Wilson. *Covariance kernels for fast automatic pattern discovery and extrapolation with Gaussian processes*. PhD thesis, University of Cambridge, 2014.

Andrew Gordon Wilson, Zhiting Hu, Ruslan Salakhutdinov, and Eric P. Xing. Deep kernel learning, 2015. URL https://arxiv.org/abs/1511.02222.

Andrew Gordon Wilson, Zhiting Hu, Ruslan Salakhutdinov, and Eric P. Xing. Stochastic variational deep kernel learning, 2016.

D. Zhang and D. Shen. Predicting future clinical changes of mci patients using longitudinal and multimodal biomarkers. *PLoS ONE*, 7:e33182, 2012. doi: 10.1371/journal.pone.0033182.