Multikernel Gaussian Processes for patient stratification from imaging biomarkers with heterogeneous patterns

Liane S. Canas[1]∗**, Benjamin Yvernault**[1]**, Carole H. Sudre**[1,5]**, M Jorge Cardoso**[1]**,**
John Thornton[2]**, Frederik Barkhof** [1,3]**, Sebastien Ourselin**[1]**, Simon Mead**[4]**, Marc Modat**[1]
[1]Translational Imaging Group, Centre for Medical Image Computing, University College of London, UK
[2]Lysholm Department of Neuroradiology, National Hospital for Neurology and Neurosurgery, London, UK
[3]Institute of Neurology, University College London, UK
[4]MRC Prion Unit, Department of Neurodegenerative Disease UCL Institute of Neurology, London, UK
[5]Dementia Research Centre, UCL Institute of Neurology, London, UK
∗liane.canas.15@ucl.ac.uk

Abstract

Prion diseases are a group of progressive neurodegenerative conditions, which cause cognitive impairment and neurological deficits. The approaches used to study other types of neurodegenerative diseases, such as Alzheimer's disease, are not appropriate to capture the progression of the human form of Prion disease. This is largely due to the heterogeneity of the phenotypes associated with Prion disease. This heterogeneity, combined with the rarity of the disease and thus the limited amount of available data, hampers the ability of state of art models to stratify patients in disease stages accurately.

In this paper, we focus on the development of a novel framework based on Gaussian Process (GP) to stratify subjects with the inherited human form of Prion disease. Our framework tackles the small number of training data as well as the presence of heterogeneity among subjects' biomarkers. This is achieved by firstly using a subject-specific multi-modal feature extraction scheme in order to normalise the data across subjects. Secondly an additive GP classifier is used to correlate subjects with disease severity. Through a simulated dataset we highlight the rationale and added value of our technique before applying it to real data.

1 Introduction

Neurodegenerative diseases (NDD) are characterised by a progressive alteration of the physiology and morphology of the brain, which leads to irreversible functional impairments. In order to characterise the progression of such diseases, researchers rely on biomarkers, extracted for example from wet samples or imaging. Biomarkers are then used, through modelling, for the purpose of diagnosis, prognosis or stratification of subjects. Different approaches have been proposed in the literature to model disease progression, including event-based models [5], mixed effect models [6] or continuous Bayesian framework [4]. Gaussian processes (GP) have also been used [9; 7] as they have the advantage of modelling the interaction between multiple biomarkers [14].

All proposed approaches for disease modelling rely on the hypothesis that biomarkers follow a similar pattern of change across patients. For example, in typical Alzheimer's disease the hippocampus is affected early in the disease, followed by other areas of the brain in a systematic order [5; 14].

The human form of Prion disease, Creutzfeldt-Jakob disease (CJD), a rare NDD, does not seem to follow a singular pattern of disease progression. The clinical diagnosis of CJD is challenging while

31st Conference on Neural Information Processing Systems (NIPS 2017), Long Beach, CA, USA.

the patient is alive, due to the heterogeneity of observed phenotypes, particularly in the early stages of the disease. To this date, definite diagnosis of CJD is only possible *post mortem* (brain autopsy). Nonetheless, several approaches have been investigated to improve the sensitivity and specificity of CJD diagnosis during life. MRI images show signal abnormalities such as hyperintensities in FLAIR and DWI images [2; 8]. Being able to diagnose CJD at the early stages of the disease would enable the patients to be involved in clinical trials, which is currently challenging as patients can die in less that 12 months from diagnosis [1]. Researcher groups are focusing on the extremely rare cases of Inherited human form of Prion disease (IPD) as it is possible to follow subjects years before they develop any clinical symptoms. IPD occurs as a result of one of more than 30 mutations in the prion protein gene (*PRNP*), hence conferring highly heterogeneous phenotypes to the disease. Variablity has also been reported among affected individuals within families carrying the same mutation but this is yet to be explained [15; 11].

In this paper, we focus on the development of a novel framework based on GP to stratify subjects with IPD. Our framework tackles the insufficient number of training data and the presence of heterogeneity among subjects' biomarkers. We designed a subject specific biomarkers extraction scheme and model the interaction between biomarkers from different MRI pulse-sequences.

To validate our framework we apply it to synthetic data before applying it to a dataset of IPD.

2 Methods

A non-parametric kernel-based model $\mathcal{M} : y = f(X) + \varepsilon,\ f \sim \mathcal{GP}(\mu_f; K),\ \varepsilon \sim \mathcal{N}(\mu_\varepsilon; \sigma)$ was used to infer the subjects status, y, given a set of imaging biomarkers $\mathbf{X} \in \mathcal{X}$, where $\mathcal{X}$ is the features space.The selected biomarkers include volumetric features extracted from structural Magnetic Resonance (MR) images, intensity-based features from FLAIR scans and mean diffusivity (MD) measures from diffusion weighted imaging (DWI). Our approach follows four main sections: (i) features extraction followed by (ii) features selection and normalisation; (iii) model training — estimation of the hyperparameters associated to the covariance kernel functions defined for the modalities under consideration —, and (iv) stratification of subjects based on the stages of the disease.

2.1 Features extraction, selection and normalisation

2.1.1 Features Extraction

Structural features consist of regional volumetric measurements and are extracted from T1-weighted MRI scans (T1w) using the Geodesical Information Flows [3] algorithm. We regress the impact of confounding effects, such as age and the total intracranial volume, by comparison with a healthy population.

To mimic clinical practice, we extract abnormal signal intensities in the grey matter (GM) from FLAIR images using the Bayesian Model Selection algorithm [12] approach. In order to robustly characterise the signal intensities of each brain region and across all subjects we compute the Mahalanobis distance between the regional GM median and the overall healthy white matter intensity distribution.

The mean diffusivity (MD) is computed from the diffusion tensors estimated from diffusion weighted imaging (DWI). We use the median MD per brain region as biomarker.

2.1.2 Features selection and normalisation

For most NDD, group-wise approaches are used to select the relevant features. This can be done as the features are consistent and homogeneous across subjects and throughout the disease stages. Such methods however cannot be applied to IPD due to the phenotype heterogeneity and the lack of consistent pattern of disease progression. A common approach could be to use Automatic Relevance Determination (ARD) as it aims at optimising different weights to automatically extract relevant features. However, such approach requires a number of sample considerably larger than the number of features, which setup is unlikely to occur with real data of patients with Prion disease.

To overcome this issue, we here hypothese that the disease does not follow a geometrical pattern but instead that imaging biomarkers can become abnormal in any location in the brain. The quantity of abnormality rather than its location is thus to quantify the progression of the disease. To quantify this

abnormality, we extract z-scores against a normative database of healthy controls for all extracted imaging biomarkers. The z-score values are then ranked per modality (structural, FLAIR, diffusion) and only the highest values for each modality are considered for subsequent learning and inference stages.

2.2 Model learning

We assume that the feature relationship across modalities can be modelled as a multi-task paradigm, a contribution of independent functions that explain the biomarkers progression. We implement an Additive GP to perform the stratification of the subjects.

Considering a GP with $\mu_f = 0$ and covariance kernel function *K*, we determine the pattern of the inductive generalisation of the features. We implement a multiclass classification GP based on individualised likelihood factors computed for the target classes defined by $y_i \in \{1, ..., 7\}$ for the subject i. These classes correspond to the seven stages of the disease: 0 – healthy control, 1 – asymptomatic subjects, 2– subjects at clinical onset , 3 to 6 – symptomatic subjects divided in 4 groups according to their clinical scores (MRC Scale) [13]. Due to the number of classes under consideration, f_i is a vector $\mathbf{f}_i = [f_i^1, ..., f_i^7]^T$. The multi-class classification is performed by means of a multinominal probit likelihood, as defined by equation 1, where the auxiliary variable u_i is distributed as $p(u_i) = \mathcal{N}(u_i|0, 1)$ and $\Phi(x)$ denotes the cumulative density function of the standard normal distribution.

$$p(y_i|\mathbf{f}_i) = \mathrm{E}_{p(u_i)} \left\{ \prod_{j=1, j\neq y_i}^{7} \mathbf{\Phi}(u_i + f_i^{y_i} - f_i^j) \right\} \tag{1}$$

Inference is achieved via a nested expectation propagation (EP), which does not require numerical quadratures or sampling to estimate the predictive probabilities, as detailed in [10].

Note that our model also accounts for the individualised pattern of each genetic mutation of IPD. In order to reduce the bias introduced by the high number of genetic mutations, we group the subjects in three clusters according to the rate of disease progression associated with each mutation. This information is included in our model using a fourth kernel matrix, K_c. The matrix K_I, which encodes the imaging biomarker, is obtained by the addition of the three kernel matrices computed individually using the information extracted from the three MR images. The final matrix K considered in the GP model results from the Hadamard product, $K_c \odot K_I$.

3 Experiments

3.1 Synthetic Dataset

To demonstrate the efficacy of the proposed approach as well as its rationale, we create a synthetic dataset. The dataset is derived from three functions designed to simulate the three MRI modalities used in this study. For each one of them, referred to as modality A, B and C in figure 1, we generate 10 different biomarkers.

Biomarkers from the same modality are set to follow a common evolution over time but all deviate differently from the main function. We alter the three main functions to also simulate the three different rate of disease progression previously mentioned. For each individual i we obtain the biomarkers $\tau \in [1, 10]$ corresponding to class y_i as a function $\mathbf{f}_i(\tau) = (f^i_{M_A}(\tau) + \epsilon_i, f^i_{M_B}(\tau) + \epsilon_i, f^i_{M_C}(\tau) + \epsilon_i), \epsilon_i \sim \mathcal{N}(0; 0.25)$. f_{M_A}, f_{M_B} and f_{M_C} correspond respectively to a monotonically increasing sigmoid function, a second order polynomial function and a monotonically decreasing sigmoid function. The number of subjects per class is uniformly distributed. In order to simulate the spatial heterogeneity of features among subjects, we selected randomly between one to three biomarkers τ to deviate from the controls samples. The estimation of y_i, j, $\hat{y}_{i,j}$ requires to find the best hyperparameters for our regression task and for each kernel. The hyperparameters θ are estimated via the maximisation of the marginal likelihood of the model, $p(y|X, \theta)$; i.e., the marginalisation over the kernel parameters is performed by *maximum a posteriori* algorithm (MAP). A 10-fold cross-validation scheme is used.

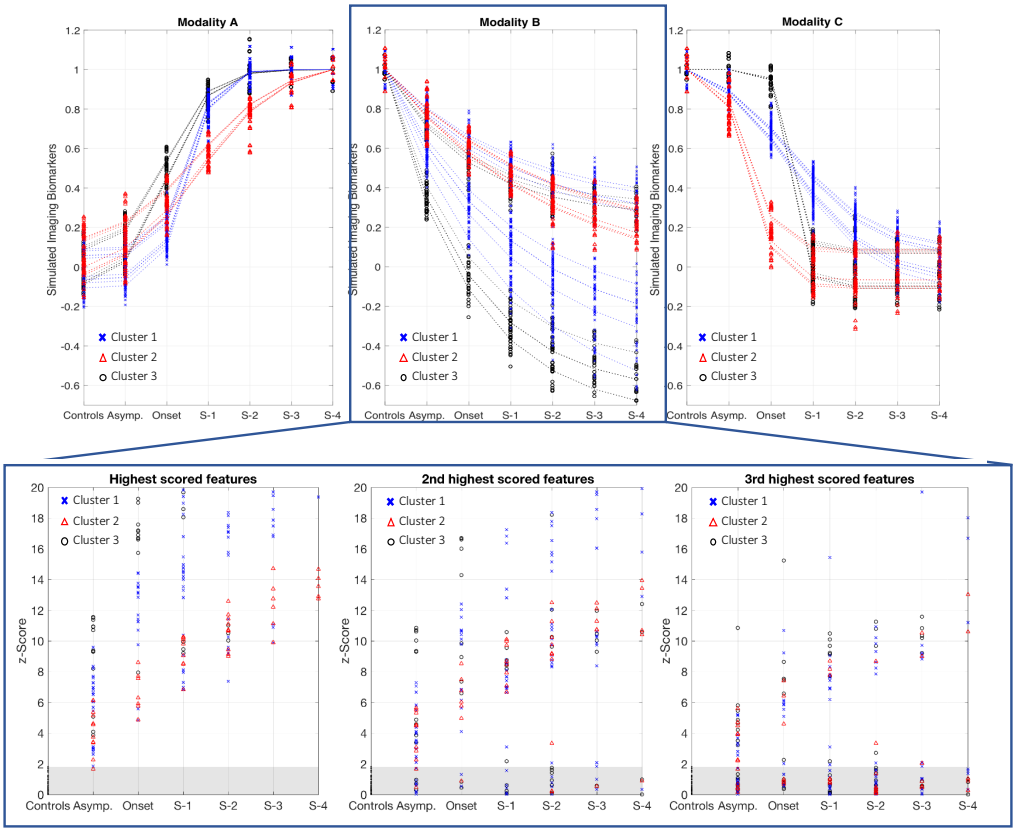

Figure 1: Synthetic dataset. Upper row: Synthetic data generated to simulate the three different modalities. The three different colour define the clusters of subjects according to the rate of progression of IPD. Lower row: Example of the data extracted from modality B after being normalised and ranked by the most significant features for each subject.

The model was used to estimate the probability of a subject to belong to each class. The same experiment is repeated with several sample size $N \in \{100, 500, 1000, 2500\}$ and different GP schemes for comparison. We divided our synthetic samples in two sub-samples: a training set with 75% of subjects, and a testing set with the remaining 25%. To ensure stability and accuracy of the presented results with the synthetic data we use bootstrapping.

Results. Table 1 reports the results of the proposed approach on synthetic data as well as the results obtained using an additive GP using all available features and an additive GP with an ARD scheme for feature selection. The mean percentages of misclassified subjects per class indicate that the proposed method is able to better capture the granularity of the disease stages than the standard approaches. Our method is only outperformed by the GP-ARD when the sample size is very large, as GP-ARD is able to find correlations between subjects even using heterogeneous biomarkers. Note that in this case, the number of samples is considerably higher (2500) than the number of features (30).

3.2 Inherited human form of Prion disease dataset

Using real patient data, we perform the stratification of IPD subjects based on their clinical diagnosis, using both imaging and genetic data. The dataset includes the baseline scans of control, symptomatic and asymptomatic subjects, as well as scans of subjects who have shown clinical symptoms within a year after their scan. The whole sample consists of 25 controls, 29 asymptomatic, 5 close to clinical onset and 30 symptomatic subjects, yielding unbalanced classes. We randomly divided our sample in two sub-samples, training and testing.

Table 1: Percentage of misclassified subjects per class. *N* defines the sample size (number of subjects). The first models corresponds to an additive GP; second model is an additive GP, as implemented before, with ARD scheme for feature selection; finally, third corresponds to the approach proposed in this paper. The bold values correspond to the best approach in each test.

N	Model	Healthy Controls	Asymp. Subjects	Clinical Onset	MRC S. (20-16)	MRC S. (15-11)	MRC S. (10-6)	MRC S. (5-0)	Mean Percentage
	GP	1.38	4.63	5.00	9.63	10.13	13.75	**4.75**	7.04
100	*GP-ARD*	2.25	6.50	6.13	9.75	10.63	12.88	5.38	7.65
	Proposed	**1.00**	**1.13**	**1.38**	**5.13**	**8.50**	**7.88**	5.50	**4.36**
	GP	**0.40**	1.36	2.00	8.00	11.44	10.96	7.12	5.90
500	*GP-ARD*	0.72	2.16	3.28	7.04	10.08	10.08	7.44	5.83
	Proposed	0.56	**0.80**	**1.20**	**3**.28	**6.72**	**9.52**	**5.60**	**3.95**
	GP	**0.40**	1.60	2.67	6.93	10.27	13.33	**6.67**	5.98
1000	*GP-ARD*	0.93	4.40	6.27	8.80	10.67	10.27	6.93	6.90
	Proposed	**0.40**	**1.20**	**1.60**	**3.33**	**6.80**	**9.73**	6.80	**4.27**
	GP	**0.08**	**0.08**	0.48	3.76	11.04	10.48	4.96	4.41
2500	*GP-ARD*	0.24	0.24	0.48	1.92	**5.20**	**6.96**	4.48	**2.79**
	Proposed	**0.08**	0.16	**0.08**	**1.84**	6.24	8.16	**4.40**	2.99

Table 2: Percentage of misclassified subjects per class using clinical data. *N* defines the sample size (number of subjects).

N	Model	Healthy Controls	Asymp. Subjects	Clinical Onset	MRC S. (20-16)	MRC S. (15-11)	MRC S. (10-6)	Mean Percentage
	GP	14.58	21.53	**4.66**	11.41	**2.18**	**0.20**	7.79
89	*GP-ARD*	14.48	16.17	5.36	11.31	2.68	**0.20**	7.17
	Proposed	**12.30**	**14.88**	4.96	**9.23**	2.28	**0.20**	**6.26**

Results. Table 2 reports the results of our proposed approach when applied to IPD patients. The proposed method reduces the misclassification rate per class, when compared with the other methods tested. The classification rate for class of symptomatic subjects was not evaluated due to the absence of individual in IPD dataset with MRC Scale lower than 5. Note that the results reported in Table 2 are deterministic and they do not account for the fuzziness of the classes estimation. The results also support the hypothesis that for rare diseases, like IPD, due to the reduced number of subjects the problem is ill-posed. In this specific case, the results achieved by the GP-ARD are not as accurate as the results obtained by an approach in which the number of features considered are reduced in order to avoid confounding effects raised by the inconsistency of features across subjects.

4 Discussion

We proposed a non-parametric Bayesian approach, which aimed to predict the disease stage of IPD subjects. The approach was designed to (i) account for the lack of biomarker geometrical consistency across subjects, as well as (ii) small sample size. The results obtained with our framework suggest that the model is better able to address efficiently the two aforementioned constraints by implementing a subject-specific biomarkers extraction. In fact the model obtained better results when compared with the current frameworks used in context of NDD, both on synthetic and clinical data. Note also that by using the GP as classification model, the current framework can be easily extended to predict the time to clinical onset of asymptomatic subjects. We thus plan to explore the advantages of GP to model the evolution of the biomarkers over time.

We aim to extend our framework to account for the longitudinal information available from these subjects. This will allow not only the stratification of subjects based on the extracted biomarkers, but also the subjects prognosis in a given time frame. Similarly, other neurodegenerative diseases with consistent biomarkers patterns across subjects can benefit from our approach.

Acknowledgements. This work is supported by the EPSRC-funded UCL Centre for Doctoral Training in Medical Imaging (EP/L016478/1) and the Department of Health's NIHR-funded Biomedical Research Centre at University College London Hospitals.

References

[1] Barry M. Bradford, Pedro Piccardo, James W. Ironside, and Neil A. Mabbott. Human prion diseases and the risk of their transmission during anatomical dissection. *Clinical Anatomy*, 27(6):821–832, 2014. doi: 10.1002/ca.22403.

[2] Federico Caobelli, Milena Cobelli, Claudio Pizzocaro, Marco Pavia, Silvia Magnaldi, and Ugo Paolo Guerra. The role of neuroimaging in evaluating patients affected by Creutzfeldt-Jakob disease: A systematic review of the literature. *Journal of neuroimaging : official journal of the American Society of Neuroimaging*, 6:1–12, 2014. doi: 10.1111/jon.12098.

[3] M. Jorge Cardoso, Marc Modat, Robin Wolz, Andrew Melbourne, David Cash, Daniel Rueckert, and Sebastien Ourselin. Geodesic information flows: Spatially-variant graphs and their application to segmentation and fusion. *IEEE Transactions on Medical Imaging*, 34(9):1976–1988, 2015. doi: 10.1109/TMI.2015.2418298.

[4] Edward Challis, Peter Hurley, Laura Serra, Marco Bozzali, Seb Oliver, and Mara Cercignani. Gaussian process classification of Alzheimer's disease and mild cognitive impairment from resting-state fMRI. *NeuroImage*, 112:232–243, 2015. doi: 10.1016/j.neuroimage.2015.02.037.

[5] Hubert M. Fonteijn, Matt J. Clarkson, and Marc Modat. An event-based disease progression model and its application to familial Alzheimer's disease. In *Information Processing in Medical Imaging*, pages 1–12, Kloster Irsee, Germany, 2011.

[6] Tian Ge, Thomas E. Nichols, Debashis Ghosh, Elizabeth C. Mormino, Jordan W. Smoller, and Mert R. Sabuncu. A kernel machine method for detecting effects of interaction between multidimensional variable sets: An imaging genetics application. *NeuroImage*, 109:505–514, 2015. doi: 10.1016/j.neuroimage.2015.01.029.

[7] Jung Won Hyun, Yimei Li, Chao Huang, Martin Styner, Weili Lin, and Hongtu Zhu. STGP: Spatio-temporal Gaussian process models for longitudinal neuroimaging data. *NeuroImage*, 134:550–562, 2016. doi: 10.1016/j.neuroimage.2016.04.023.

[8] Marc Manix, Piyush Kalakoti, Miriam Henry, Jai Thakur, Richard Menger, Bharat Guthikonda, and Anil Nanda. Creutzfeldt-Jakob disease: updated diagnostic criteria, treatment algorithm, and the utility of brain biopsy. *Neurosurgical Focus*, 39:1–11, 2015. doi: 10.3171/2015.8.FOCUS15328.

[9] Carl Rasmussen and Christopher Williams. *Gaussian processes for machine learning.*, volume 14. 2004. doi: 10.1142/S0129065704001899.

[10] Jaakko Riihim. Nested expectation propagation for Gaussian process classification with a multinomial probit likelihood. *Journal of Machine Learning Research*, 14:75–109, 2013.

[11] Andreas Schroter, Inga Zerr, Karten Henkel, Henriette J. Tschampa, Michael Finkenstaedt, and Sigrid Poser. Magnetic resonance imaging in the clinical diagnosis of Creutzfeldt-Jakob disease. *Journal of American Medical Association*, 57:1751–1757, 2000. doi: 10.1001/archneur.57.12.1751.

[12] Carole Sudre, M. Jorge Cardoso, Willem Bouvy, Geert Biessels, Josephine Barnes, and Sebastien Ourselin. Bayesian model selection for pathological neuroimaging data applied to white matter lesion segmentation. *IEEE Transactions on Medical Imaging*, 34(c):1–1, 2015. doi: 10.1109/TMI.2015.2419072.

[13] Andrew G B Thompson, Jessica Lowe, Zoe Fox, Ana Lukic, Marie Claire Porter, Liz Ford, Michele Gorham, Gosala S. Gopalakrishnan, Peter Rudge, A. Sarah Walker, John Collinge, and Simon Mead. The medical research council prion disease rating scale: A new outcome measure for prion disease therapeutic trials developed and validated using systematic observational studies. *Brain*, 136(4):1116–1127, 2013. ISSN 14602156. doi: 10.1093/brain/awt048.

[14] Jonathan Young, Marc Modat, Manuel J Cardoso, John Ashburner, and Sebastien Ourselin. An oblique approach to prediction of conversion to Alzheimer ' s disease with multikernel Gaussian processes. In *Machine Learning and Interpretation in Neuroimaging*, volume 1, pages 122–128, 2014. doi: 10.1007/978-3-319-45174-9.

[15] Inga Zerr and Sigrid Poser. Clinical diagnosis and differential diagnosis of CJD and vCJD with special emphasis on laboratory tests. *Acta Pathologica, Microbiologica et Immunologica Scandinavica*, 110:88–98, 2002. doi: 10.1034/j.1600-0463.2002.100111.x.