

Bayesian Gaussian Process Classification from Event-Related Brain Potentials in Alzheimer's Disease

Wolfgang Fruehwirt^{1,2}, Pengfei Zhang², Matthias Gerstgrasser³,
Dieter Grossegger⁴, Reinhold Schmidt⁵, Thomas Benke⁶,
Peter Dal-Bianco⁷, Gerhard Ransmayr⁸, Leonard Weydemann¹,
Heinrich Garn⁹, Markus Waser⁹, Michael Osborne²,
and Georg Dorffner¹(✉)

¹ Section for AI and Decision Support,
Medical University of Vienna, Vienna, Austria
georg.dorffner@meduniwien.ac.at

² Department of Engineering Science, University of Oxford, Oxford, UK

³ Department of Computer Science, University of Oxford, Oxford, UK

⁴ Dr. Grossegger & Drbal GmbH, Vienna, Austria

⁵ Department of Neurology, Medical University of Graz, Graz, Austria

⁶ Department of Neurology, Medical University of Innsbruck,
Innsbruck, Austria

⁷ Department of Neurology, Medical University of Vienna, Vienna, Austria

⁸ Department of Neurology, Linz General Hospital, Linz, Austria

⁹ AIT Austrian Institute of Technology GmbH, Vienna, Austria

Abstract. Event-related potentials (ERPs) have been shown to reflect neurodegenerative processes in Alzheimer's disease (AD) and might qualify as non-invasive and cost-effective markers to facilitate the objectivization of AD assessment in daily clinical practice. Lately, the combination of multivariate pattern analysis (MVPA) and Gaussian process classification (GPC) has gained interest in the neuroscientific community. Here, we demonstrate how a MVPA-GPC approach can be applied to electrophysiological data. Furthermore, in order to account for the temporal information of ERPs, we develop a novel method that integrates interregional synchrony of ERP time signatures. By using real-life ERP recordings of a prospective AD cohort study (PRODEM), we empirically investigate the usefulness of the proposed framework to build neurophysiological markers for single subject classification tasks. GPC outperforms the probabilistic reference method in both tasks, with the highest AUC overall (0.802) being achieved using the new spatiotemporal method in the prediction of rapid cognitive decline.

Keywords: Machine learning · Gaussian process classification · Event-related potentials · Alzheimer's disease · Single subject classification

1 Introduction

Due to its degenerative nature, early detection and accurate evaluation of Alzheimer's disease (AD) are crucial. However, when it comes to routine clinical practice, AD diagnosis is most commonly based on subjective clinical interpretations at a progressed stage of the disease, i.e. when symptoms are already apparent. Thus, there is a strong need to develop affordable and thereby widely available markers that facilitate the objectivization of AD assessment. Event-related potentials (ERPs), as measured by non-invasive and cost-effective electroencephalography (EEG), have been shown to reflect neurodegenerative processes in AD [1–3] and might therefore qualify as such markers.

Lately, the combination of multivariate pattern analysis (MVPA) and Gaussian process classification (GPC), a machine learning technique, has gained interest in the neuroscientific community. While MVPA incorporates interactions between multiple brain structures or function patterns, GPC allows for an easy adjustment of predictions to compensate for variable class priors (e.g. variations in diagnostic setting or disease prevalence), and - most importantly for routine practice - provides probabilistic predictions quantifying predictive uncertainty.

Only recently, a resting-state functional magnetic resonance imaging (fMRI) study showed the applicability of a MVPA-GPC approach for single subject classification tasks in AD [4]. Here, we demonstrated how this technique can be applied to electrophysiological data. Furthermore, to account for the temporal information of ERPs, we developed a novel method that integrates interregional synchrony of ERP time signatures.

To the best of our knowledge, this is the first electrophysiological ERP study to use GPC and therefore the first to use MVPA-GPC.

Utilizing real-life ERP recordings of a prospective AD cohort study, we aim to build neurophysiological markers for two crucial AD classification problems:

First, we intend to predict rapid cognitive decline. The rate of cognitive decline in AD strongly correlates with mortality and shows profound variability between individuals [5]. Hence, early prediction of individual trajectories of cognitive function is essential for treatment and care, as it allows for personalized interventions and appropriate forehanded planning of support services.

Then, we try to identify patients who test positive for the Apolipoprotein E (ApoE) $\epsilon 4$ allele ($\epsilon 4+$), the strongest genetic risk factor for AD [6]. As ApoE $\epsilon 4$ expression has been shown to alter ERP waveforms [7, 8] as well as functional EEG connectivity [9, 10], neurophysiological markers incorporating both aspects - as done by our framework - may allow for such estimates.

Distinct parameter settings are tested and, using the identical preselected MVPA features, GPC performance is compared to a probabilistic reference method (logistic regression classification, LRC).

2 Materials and Methods

2.1 Subjects

Sixty-three AD patients (31 with possible, 32 with probable AD diagnosis according to NINCDS-ADRDA criteria; 39 APOE $\epsilon 4$ carriers; 38 females; mean age 75.92 ± 8.82 standard deviation (SD); mean MMSE score 23.25 ± 3.60 SD; mean years of education 10.46 ± 2.26 SD; mean duration of illness (months) 22.89 ± 14.65 SD) were considered for this investigation. They were recruited prospectively at the tertiary-referral memory clinic of the Medical University of Innsbruck as part of the cohort study Prospective Dementia Registry Austria (PRODEM). PRODEM is a longitudinal multicenter study of AD and other dementias in a routine clinical setting by the Austrian Alzheimer Society. Ethics committee approval was obtained and patients and their caregivers gave written informed consent.

Inclusion criteria encompassed: (I) diagnosis of Alzheimer-type dementia according to NINCDS-ADRDA criteria, (II) minimum age 40 years, (III) non-institutionalization and no need for 24-hour care, (IV) availability of a caregiver who agrees to provide information on the patient's condition. Patients with comorbidities likely to preclude termination of the study were excluded.

For a maximum of 18 months patients revisited for follow-up assessments (FU) every 6 months. Prediction of rapid cognitive decline was performed for subjects who returned at least for the 12-month FU ($N = 48$; 22 with possible, 26 with probable AD; 18 with rapid cognitive decline; 33 APOE $\epsilon 4$ carriers; 29 females; mean age 75.90 ± 8.61 SD; mean MMSE score 23.71 ± 3.13 SD; mean years of education 10.83 ± 2.28 SD; mean duration of illness (months) 23.17 ± 16.02 SD).

2.2 Assessment of Cognitive Decline and Apolipoprotein E Genotyping

Assessment of cognitive decline was done using the Mini-Mental State Examination (MMSE, [11]). Rapid cognitive decline was defined as a decrease of 3 or more points on the MMSE between baseline and 12-month FU [12].

We used ApoE genotyping to determine $\epsilon 4$ expression. ApoE is the principal cholesterol carrier protein in the brain [13] and supports lipid transport and injury repair. Carriers of the ApoE $\epsilon 4$ allele ($\epsilon 4+$) are known to be of heightened risk to develop AD. Consequently, the frequency of the $\epsilon 4$ allele is dramatically increased in patients with AD as compared to the overall population [6].

2.3 Recording and Pre-processing of Event-Related Potentials

Auditory ERPs were elicited using the “oddball” paradigm, a simple discrimination task. Frequent (141) standard tones (1000 Hz) and infrequent (57) target tones (2000 Hz) appeared in a quasi-random sequence held constant across subjects. The tone duration being 100 ms, interstimulus intervals varied between 1 and 1.5 s.

Subjects were instructed to press a reaction time button, with the dominant hand, to target stimuli only. Horizontal and vertical electrooculogram (EOG) electrodes detected eye movements. The system employed was a 32-channel AlphaEEG amplifier with

NeuroSpeed software (alpha trace medical systems, Vienna, Austria). EEG electrode placement (Au-plated cups; Grass F-E5GH, Grass Technologies, West Warwick, RI, USA) was in accordance with the international 10–20 system. The electrodes were referenced to connected mastoids, the ground being positioned at FCz. The EEG amplifier had a bandpass of 0.3 to 70 Hz (3 dB points) with a 50 Hz notch filter and a sampling rate set at 256 Hz. Impedance levels were held below 10 k Ω .

After automatic horizontal and vertical regression-based EOG correction in the time domain [14] the individual sweeps to targets were visually screened for artefacts before being accepted into the average. Then, the data were bandpass filtered at 1–16 Hz using the EEGLAB toolbox [15].

2.4 Spatial Synchrony Measures

For multichannel covariance estimation, ERP time-courses of individual patients (1 s, starting at stimulus onset) were considered as N by T matrices X_z , N being the number of electrodes, T being the number of time samples, and z being the patient.

$$X_z \in \mathbb{R}^{N \times T} \quad (1)$$

Covariance for patient z was estimated using the sample covariance matrix (SCM) C_z .

$$C_z = \frac{1}{T-1} X X^T \quad (2)$$

All unique elements of the SCM, i.e., the diagonal entries representing the spatial variance and one set of the off-diagonal entries representing the spatial covariance, were then combined into a feature vector F_z of dimension $d = N(N+1)/2$. Using 19 electrodes d equaled 190.

$$F_z \in \mathbb{R}^{N(N+1)/2} \quad (3)$$

2.5 Spatiotemporal Synchrony Measures

Regular covariance estimation between brain areas - as for instance used by Challis et al. [4] - comprises signal variance at each of the individual sites as well as the covariance between all site pairs. This spatial information can be embodied in the matrix form described above. ERPs, however, represent a time- and phase-locked response to a stimulus. Therefore, their unfolding in time, i.e., their time signature or pattern, contains specific temporal information putatively valuable for classification purposes.

To take these distinct ERP time signatures into account, we adopt a special type of matrix recently designed [16, 17]. Used in brain-computer interface applications to distinguish between subjects' single responses to stimuli, we adapt the construct to classify subjects themselves. We do so by utilizing averaged instead of single trial

potentials. First, the averaged trial data of a given unlabeled patient is vertically concatenated with the grand-average waveforms (temporal prototypes) $\bar{X}_{(1)}$ and $\bar{X}_{(2)}$ of the two classes of patients. Holding one second of data, the respective matrix X_z^{ST} has a dimensionality of 57×256 .

$$X_z^{ST} = \begin{pmatrix} \bar{X}_{(1)} \\ \bar{X}_{(2)} \\ X_z \end{pmatrix} \in \mathbb{R}^{3N \times T} \quad (4)$$

Then, the covariance is estimated resulting in a 57×57 spatiotemporal SCM C_z^{ST} .

$$C_z^{ST} = \frac{1}{(T-1)} \left(X_z^{ST} (X_z^{ST})^T \right) = \frac{1}{(T-1)} \begin{pmatrix} \bar{X} \bar{X}^T & (X_z \bar{X})^T \\ X_z \bar{X}^T & X_z X_z^T \end{pmatrix} \in \mathbb{R}^{3N \times 3N} \quad (5)$$

$$\text{where } \bar{X} \bar{X}^T = \frac{1}{(T-1)} \begin{pmatrix} \bar{X}_{(1)} \bar{X}_{(1)}^T & \bar{X}_{(1)} \bar{X}_{(2)}^T \\ \bar{X}_{(2)} \bar{X}_{(1)}^T & \bar{X}_{(2)} \bar{X}_{(2)}^T \end{pmatrix} \in \mathbb{R}^{2N \times 2N} \quad (6)$$

$$\text{and } X_z \bar{X}^T = \begin{pmatrix} X_z \bar{X}_{(1)}^T & X_z \bar{X}_{(2)}^T \end{pmatrix} \in \mathbb{R}^{N \times 2N} \quad (7)$$

However, not all the information encoded in this special matrix is useful. The covariance block of the unlabeled patient $X_z X_z^T$ can be considered useful, as it includes spatial information of the patient's signal. $X_z \bar{X}^T$ represents the cross-covariances between the unlabeled patient z and the two pattern prototypes and therefore contains the sought after temporal information (note how shuffling of the columns of X_z changes C_z^{ST}). The covariance and cross-covariance blocks of the templates are not informative and are therefore not included in the feature vector. For this reason, we built the feature vector F_z^{ST} by dividing three non-redundant blocks of C_z^{ST} , i.e., $X_z X_z^T$, $X_z \bar{X}_{(1)}^T$, and $X_z \bar{X}_{(2)}^T$. As $X_z X_z^T$ is symmetric, only the upper triangle was included into F_z^{ST} resulting in 912 dimensions.

$$F_z^{ST} \in \mathbb{R}^{2N^2 + N(N+1)/2} \quad (8)$$

2.6 Machine Learning Classifiers

Feature Selection. Two methods for feature selection were carried out. In order to identify features with high discriminative power we computed the Kendall rank correlation coefficient versus the binary class label [4, 18]. Subsequently, only the variables with the highest absolute tau coefficients were used in the GPC or LRC model. Since this feature selection step was included in a leave-one-out cross-validation (LOOCV) algorithm, the selected features differed slightly from iteration to iteration. In line with Challis et al. [4] we examined the model performance using the

highest-ranking 5, 10, 15, and 20 features. In addition we added a run with the 3 best features. Moreover, to determine implicitly the relevance of dimensions within GPC and weight feature contribution accordingly, we utilized automatic relevance determination (ARD, see Gaussian process classification).

Gaussian Processes. In the following, we provide a formal definition of the GPC which is a Bayesian classification method for non-linear, stochastic and complex classification problems. We start with the general definition of a 2-D Gaussian process (GP), and then define GPC.

Definition 1. *Gaussian process* [19, 20]: Denote by $f(x, t) : \mathcal{X} \mapsto \mathbb{R}$ a stochastic process parameterized by $\{x, t\} \in \mathcal{X}$, where $\mathcal{X} \in \mathbb{R}^2 \times \mathbb{R}^+$. Then, the random function $f(x, t)$ is a Gaussian process if all its finite dimensional distributions are Gaussian, where for any $m \in \mathbb{N}$, the random variables $(f(x_1, t_1), \dots, f(x_m, t_m))$ are jointly normally distributed.

We can therefore interpret a GP as formally defined by the following class of random functions:

$$\begin{aligned} \mathcal{F} &:= \{f(\cdot) : \mathcal{X} \mapsto \mathbb{R} \\ &\text{s.t. } f(\cdot) \sim \mathcal{GP}(\mu(\cdot; \theta), \mathcal{C}(\cdot, \cdot; \Psi)), \text{ with} \\ \mu(\cdot; \theta) &:= \mathbb{E}[f(\cdot)] : \mathcal{X} \mapsto \mathbb{R}, \\ \mathcal{C}(\cdot, \cdot; \Psi) &:= \mathbb{E}[(f(\cdot) - \mu(\cdot; \theta))(f(\cdot) - \mu(\cdot; \theta))^T] : \mathcal{X} \times \mathcal{X} \mapsto \mathbb{R}^+ \}. \end{aligned} \quad (9)$$

Before receipt of data, at each point the (prior) mean of the function is $\mu(\cdot; \theta)$ parameterized by θ , and the (prior) spatial dependence between any two points is given by the covariance function (Mercer kernel) $\mathcal{C}(\cdot, \cdot; \Psi)$, parametrized by Ψ (see detailed discussion in Rasmussen and Williams [19]).

Gaussian Process Classification. Given a model which is based on GP, we can classify an unknown field at unobserved locations x_* .

Definition 2. *Gaussian Process Classification* [19]: A GP prior is placed over a latent function $f(x)$, then is “squashed” through the logistic function to obtain a prior on $\pi(x) \triangleq p(y = +1|x) = \sigma(f(x))$.

To define our model and its approximate inference scheme, we make design decisions as per the following three sections.

- (1) **Square Exponential Kernel:** The model for $f(x)$ is specified by the kernel of the GP. One of the mostly widely used kernels in GP is the square exponential kernel (also known as the exponential quadratic, RBF, or Gaussian kernel), which specifies the covariance between any two locations in space in the following equation:

$$K(x_1, x_2) = \exp\left(-\frac{(x_1 - x_2)^2}{2l_1^2}\right) \quad (10)$$

- ARD is a popular method for determining the hyperparameters in the square exponential kernels (namely l_1). It is implemented in the Gaussian process for machine learning (GPML) package in MATLAB [21].
- (2) Expectation Propagation: For the GPC, the posterior density of a latent process $p(f_*|X, y, x_*)$ is intractable due to the fact that the likelihood $p(y|X)$ is intractable. Therefore, approximation techniques need to be proposed to get the posterior. One widely used technique is expectation propagation (EP). The EP method has been discussed in detail in Rasmussen and Williams [19] and the implementation for the study at hand is done via the GPML package [21].
 - (3) Prediction: The estimated class label at the unknown location is denoted by \hat{f}_* . The expectation propagation framework gives a Gaussian approximation to the posterior distribution. The approximate predictive mean for the latent variable f_* denoted as $\mathbb{E}[f_*|X, y, x_*]$ and the uncertainty (statistical error) denoted as $\sigma_*^2|X, y, x_*$ is given by Rasmussen and Williams [19]. The approximate predictive distribution for the binary target becomes where $q[f_*|X, y, x_*]$ is the approximate latent predictive Gaussian with mean and variance given by $\mathbb{E}[f_*|X, y, x_*]$ and $\sigma_*^2|X, y, x_*$.

There are various ways of fitting parametric models for μ (defined in formula 9). In this work, we select from two mean functions. The first is a combination of a linear and a constant mean function (sum mean). The second is a constant mean function (constant mean) alone. Performance of these two functions was compared in the classification experiments.

Classification Experiments. To evaluate the real added value of GPC in terms of classification performance, we compared results achieved by GPC to those achieved by a probabilistic reference technique, namely, LRC.

LRC was implemented using the MATLAB package Logistic Regression for Classification (Pattern Recognition and Machine Learning Toolbox), as developed by Chen [22]. Importantly, the LRC model had the same preselected input features as the GPC model. We used LOOCV to assess the ability of the classifiers to generalize to independent data sets.

GPC and LRC were applied to two distinct single subject classification problems. Patients were classified into (I) carriers ($\epsilon 4+$) and non-carriers ($\epsilon 4-$) of the $\epsilon 4$ allele, and (II) future rapid cognitive decliners (RCD) and future non-rapid cognitive decliners (nRCD).

In order to assess the degree to which unlabeled patients were identified with their correct class labels in the LOOCV, several classifier performance indices were computed (Table 1). Primary outcome measure was the area under the curve (AUC) of the receiver operating characteristic (ROC) curve, for which we calculated test statistics (null hypothesis = AUC of 0.5). A major advantage of AUC as performance measure is that it incorporates the various threshold settings inherent to the ROC curve. Remaining indices were derived from symmetric threshold values.

As opposed to conventional accuracy, balanced accuracy does not lead to an optimistic estimate when a biased classifier is tested on an imbalanced dataset [23]. Thus, it is reported additionally.

Potential model differences were evaluated by statistically comparing the entire areas under the ROC curves, while accounting for the paired nature of the data [24].

3 Results

3.1 Prediction of Rapid Cognitive Decline

Overall, the best classification in terms of AUC was achieved by spatiotemporal GPC ($\text{GPC}_{\text{ST-RCD}}$, 0.802, $p < 0.001$, best 20 features, GP function = constant mean). This model had 75.6% balanced-accuracy, 77.1% accuracy, 66.7% sensitivity, and 83.3% specificity. The best competing LRC model was of spatiotemporal nature as well ($\text{LRC}_{\text{ST-RCD}}$, best 5 features) and yielded an AUC of 0.561 ($p = 0.482$). The ROC curves of the two classifiers differed significantly at $p = 0.012$ and are depicted in Fig. 1. The best model using spatial information only was a GPC model ($\text{GPC}_{\text{S-RCD}}$, best 5 features, GP function = sum mean) and yielded an AUC of 0.607 ($p = 0.217$). Difference of $\text{GPC}_{\text{S-RCD}}$ to the superior spatiotemporal GP classifier $\text{GPC}_{\text{ST-RCD}}$ was significant ($p = 0.046$). Performance measures for the best models per feature and classifier category are given in Table 1.

3.2 Apolipoprotein E $\epsilon 4$ Classification

Again, GPC performed better than LRC in both feature categories, the corresponding models significantly rejecting the null hypothesis at $p = 0.002$ ($\text{GPC}_{\text{S-}\epsilon 4+}$, best 3 features, GP function = constant mean) and $p = 0.046$ ($\text{GPC}_{\text{ST-}\epsilon 4+}$, best 3 features, GP function = constant mean) respectively. However, the model based on spatial information ($\text{GPC}_{\text{S-}\epsilon 4+}$) ranked higher ($p = 0.017$) than the one based on spatiotemporal information ($\text{GPC}_{\text{ST-}\epsilon 4+}$). The difference between the best GPC and the best LRC ($\text{LRC}_{\text{S-}\epsilon 4+}$, best 20 features) model was significant ($p = 0.041$). Performance measures for the best models per feature and classifier category are given in Table 1.

Table 1. Performance measures of the best spatial and spatiotemporal GPC and LRC classifiers for cognitive decline ($N = 48$) and ApoE $\epsilon 4$ expression ($N = 63$)

Model	AUC (p-value)	B. accuracy	Accuracy	Sensitivity	Specificity
$\text{GPC}_{\text{ST-RCD}}$	0.802 (<0.001)	75.6%	77.1%	66.7%	83.3%
$\text{GPC}_{\text{S-RCD}}$	0.607 (0.217)	59.4%	62.5%	44.4%	73.3%
$\text{LRC}_{\text{ST-RCD}}$	0.561 (0.482)	52.8%	45.8%	77.8%	26.7%
$\text{LRC}_{\text{S-RCD}}$	0.476 (0.782)	47.2%	43.8%	61.1%	33.3%
$\text{GPC}_{\text{ST-}\epsilon 4+}$	0.651 (0.046)	64.3%	66.7%	76.9%	50.0%
$\text{GPC}_{\text{S-}\epsilon 4+}$	0.735 (0.002)	69.6%	71.4%	79.5%	58.3%
$\text{LRC}_{\text{ST-}\epsilon 4+}$	0.510 (0.899)	47.9%	42.9%	25.6%	70.8%
$\text{LRC}_{\text{S-}\epsilon 4+}$	0.563 (0.404)	57.1%	52.4%	38.5%	75.0%

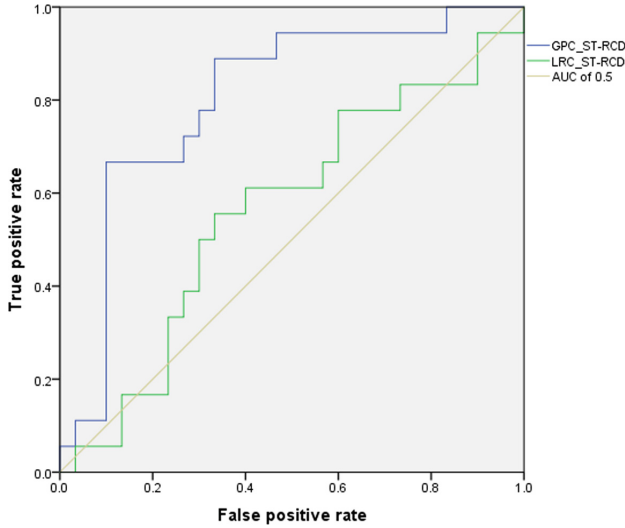


Fig. 1. ROC curves of the best GPC ($\text{GPC}_{\text{ST-RCD}}$) and LRC ($\text{LRC}_{\text{ST-RCD}}$) classifiers for predicting rapid cognitive decline.

4 Summary and Discussion

Although potential advantages of GPC for ERP analysis have already been stressed in the literature [25], this is - to the best of our knowledge - the first electrophysiological ERP study to report the use of GPC.

We demonstrated the applicability of MVPA-GPC for electrophysiological data and proposed a method that takes the synchrony of temporal ERP signatures into account. To examine potential advantages of GPC in classification performance we compared it to a probabilistic reference method (LRC). Using ERP data of a prospective cohort study we aimed at building cheap and non-invasive MVPA markers for crucial AD classification problems, i.e., the prediction of rapid cognitive decline and the distinction between carriers and non-carriers of the ApoE $\epsilon 4$ allele. GPC significantly outperformed LRC in both tasks, with the highest AUC overall (0.802) being achieved using the newly developed spatiotemporal method in the prediction of rapid cognitive decline.

Although the number of AD patients included in this examination is relatively large compared to other ERP studies, the modest sample size - in absolute terms - constitutes a limitation to the results. Even though single subject classification was cross-validated using a leave-one-out strategy, further studies with larger sample sizes, including extensive external validation sets, should follow.

Acknowledgment. The PRODEM study has been supported by the Austrian Research Promotion Agency FFG, project no. 827462, including financial contributions from Dr. Grosseegger and Drbal GmbH, Vienna, Austria.

References

1. Howe, A.S., Bani-Fatemi, A., De Luca, V.: The clinical utility of the auditory P300 latency subcomponent event-related potential in preclinical diagnosis of patients with mild cognitive impairment and Alzheimer's disease. *Brain Cogn.* **86**, 64–74 (2014)
2. Howe, A.S.: Meta-analysis of the endogenous N200 latency event-related potential subcomponent in patients with Alzheimer's disease and mild cognitive impairment. *Clin. Neurophysiol.* **125**, 1145–1151 (2014)
3. Olichney, J.M., Yang, J.C., Taylor, J., Kutas, M.: Cognitive event-related potentials: biomarkers of synaptic dysfunction across the stages of Alzheimer's disease. *J. Alzheimer's Dis.* **26**(Suppl. 3), 215–228 (2011)
4. Challis, E., Hurley, P., Serra, L., Bozzali, M., Oliver, S., Cercignani, M.: Gaussian process classification of Alzheimer's disease and mild cognitive impairment from resting-state fMRI. *NeuroImage* **112**, 232–243 (2015)
5. Hui, J.S., Wilson, R.S., Bennett, D.A., Bienias, J.L., Gilley, D.W., Evans, D.A.: Rate of cognitive decline and mortality in Alzheimer's disease. *Neurology* **61**, 1356–1361 (2003)
6. Liu, C.C., Kanekiyo, T., Xu, H., Bu, G.: Apolipoprotein E and Alzheimer disease: risk, mechanisms and therapy. *Nat. Rev. Neurol.* **9**, 106–118 (2013)
7. Rosengarten, B., Paulsen, S., Burr, O., Kaps, M.: Effect of ApoE ϵ 4 allele on visual evoked potentials and resultant flow coupling in patients with Alzheimer. *J. Geriatr. Psychiatry Neurol.* **23**, 165–170 (2010)
8. Green, J., Levey, A.I.: Event-related potential changes in groups at increased risk for Alzheimer disease. *Arch. Neurol.* **56**, 1398–1403 (1999)
9. Lee, T.-W., Yu, Y.W.-Y., Hong, C.-J., Tsai, S.-J., Wu, H.-C., Chen, T.-J.: The influence of apolipoprotein E Epsilon4 polymorphism on qEEG profiles in healthy young females: a resting EEG study. *Brain Topogr.* **25**, 431–442 (2012)
10. Canuet, L., Tellado, I., Couceiro, V., Fraile, C., Fernandez-Novoa, L., Ishii, R., Takeda, M., Cacabelos, R.: Resting-state network disruption and APOE genotype in Alzheimer's disease: a lagged functional connectivity study. *PLoS ONE* **7**, e46289 (2012)
11. Folstein, M.F., Folstein, S.E., McHugh, P.R.: "Mini-mental state". A practical method for grading the cognitive state of patients for the clinician. *J. Psychiatr. Res.* **12**, 189–198 (1975)
12. Carcaillon, L., Pérès, K., Péré, J.J., Helmer, C., Orgogozo, J.M., Dartigues, J.F.: Fast cognitive decline at the time of dementia diagnosis: a major prognostic factor for survival in the community. *Dement. Geriatr. Cogn. Disord.* **23**, 439–445 (2007)
13. Puglielli, L., Tanzi, R.E., Kovacs, D.M.: Alzheimer's disease: the cholesterol connection. *Nat. Neurosci.* **6**, 345–351 (2003)
14. Anderer, P., Semlitsch, H.V., Saletu, B., Barbanoj, M.J.: Artifact processing in topographic mapping of electroencephalographic activity in neuropsychopharmacology. *Psychiatry Res.: Neuroimaging* **45**, 79–93 (1992)
15. Delorme, A., Makeig, S.: EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *J. Neurosci. Methods* **134**, 9–21 (2004)
16. Barachant, A., Congedo, M.: A Plug&Play P300 BCI Using Information Geometry. *arXiv preprint arXiv:1409.0107* (2014)
17. Congedo, M., Barachant, A., Andreev, A.: A New Generation of Brain-Computer Interface Based on Riemannian Geometry *arXiv:1310.8115* (2013)
18. Zeng, L.-L., Shen, H., Liu, L., Wang, L., Li, B., Fang, P., Zhou, Z., Li, Y., Hu, D.: Identifying major depression using whole-brain functional connectivity: a multivariate pattern analysis. *Brain* **135**, 1498–1507 (2012)

19. Rasmussen, C.E., Williams, C.K.I.: Gaussian Processes for Machine Learning. The MIT Press, Cambridge (2006)
20. Adler, R.J., Taylor, J.E.: Random Fields and Geometry. Springer, New York (2007)
21. Rasmussen, C.E., Nickisch, H.: Gaussian processes for machine learning (GPML) toolbox. *J. Mach. Learn. Res.* **11**, 3011–3015 (2010)
22. Chen, M.: Pattern Recognition and Machine Learning Toolbox. MATLAB Central File Exchange (2016)
23. Brodersen, K.H., Ong, C.S., Stephan, K.E., Buhmann, J.M.: The balanced accuracy and its posterior distribution. In: Proceedings of the 2010 20th International Conference on Pattern Recognition, pp. 3121–3124. IEEE Computer Society (2010)
24. Hanley, J.A., McNeil, B.J.: A method of comparing the areas under receiver operating characteristic curves derived from the same cases. *Radiology* **148**, 839–843 (1983)
25. Stahl, D., Pickles, A., Elsabbagh, M., Johnson, M.H., The, B.T.: Novel machine learning methods for ERP analysis: a validation from research on infants at risk for autism. *Dev. Neuropsychol.* **37**, 274–298 (2012)