Multi-modal Synthesis of ASL-MRI Features with KPLS Regression on Heterogeneous Data

Toni Lassila[1](✉), Helena M. Faria[1], Ali Sarrami-Foroushani[1],
Francesca Meneghello[2], Annalena Venneri[3], and Alejandro F. Frangi[1]

[1] Centre for Computational Imaging and Simulation Technologies in Biomedicine
(CISTIB), Department of Electronic and Electrical Engineering,
University of Sheffield, Sheffield, UK
t.lassila@sheffield.ac.uk
[2] IRCCS Fondazione Ospedale San Camillo, Lido Venice, Italy
[3] Department of Neuroscience, University of Sheffield, Sheffield, UK

Abstract. Machine learning classifiers are frequently trained on heterogeneous multi-modal imaging data, where some patients have missing modalities. We address the problem of synthesising arterial spin labelling magnetic resonance imaging (ASL-MRI) - derived cerebral blood flow (CBF) - features in a heterogeneous data set. We synthesise ASL-MRI features using T1-weighted structural MRI (sMRI) and carotid ultrasound flow features. To deal with heterogeneous data, we extend the kernel partial least squares regression (kPLSR) - method to the case where both input and output data have partial coverage. The utility of the synthetic CBF features is tested on a binary classification problem of mild cognitive impairment patients vs. controls. Classifiers based on sMRI and synthetic ASL-MRI features are combined using a maximum probability rule, achieving a balanced accuracy of 92% (sensitivity 100 %, specificity 80 %) in a separate validation set. Comparison is made against support vector machine-classifiers from literature.

1 Introduction

Arterial spin-labelling magnetic resonance imaging (ASL-MRI) is a non-invasive blood flow imaging modality that can improve the diagnosis of Alzheimer's disease (AD) by providing estimates of cerebral blood flow (CBF) and identifying regions of chronic cerebral hypoperfusion in individuals. However, ASL-MRI is still not part of clinical routine; many research databases used to train dementia classifiers, such as ADNI (Alzheimer's Disease Neuroimaging Initiative), do not include ASL-MRI as part of their data collection protocol.

Image synthesis refers to the simulation of missing image modalities with machine learning algorithms using image modalities that are available. Synthetic image modalities can in some cases provide additional predictive value beyond the original data used for synthesis [7]. We use kernel partial least squares regression (kPLSR) for synthesising ASL-MRI-based CBF maps using structural MRI

(sMRI) features and carotid ultrasound flow measurements as regressors. Using partial volumes of cortical and sub-cortical regions as features allows the relation between cerebral volume loss and reduction in CBF in dementia patients to be learned by the model. The synthetic CBF maps are used to generate CBF maps of patients for whom no ASL-MRI images are available (*CBF imputation*).

The challenge of building models using multi-modal data is that multiple cohorts may be required to achieve enough coverage and some cohorts will have some modalities missing. We refer to this as the *heterogeneous data problem*. To address it, we modify the NIPALS algorithm for training the kPLSR model to work on heterogeneous data, where some features are missing in part of the input data X, and some of the output data Y are also missing. The synthetic CBF maps are utilised as classification features in discriminating mild cognitive impairment-patients (MCIs) from cognitively healthy controls (CHCs).

The MCI vs. CHC-problem is less studied than the AD vs. CHC-problem (see e.g. the review [1]) because sMRI-derived partial volume-features are less informative in the prodromal stage of AD. We apply a simultaneous feature selection and classification strategy based on: (i) use of regional CBF values averaged over anatomical subregions (instead of voxelwise values), and (ii) elastic net regression. The proposed classifiers are compared to MCI vs. CHC-classifiers from literature using different imaging modalities as features.

2   Methods

2.1   Acquisition and Pre-processing of Imaging Data

Data from two clinical centres and three different cohorts were included to increase the number of cases available for training models (Table 1). The combined data set was heterogeneous with respect to operator, MR field strength, and modalities available for each case. Three different sets of features were used.

sMRI Features: T1-weighted sMRI were acquired and volumes of 141 cortical and sub-cortical regions were computed by propagating anatomical labels with the geodesic information flows-algorithm [2]. These features encapsulated grey matter (GM) atrophy but did not contain direct information about CBF.

Carotid Flow Features: Carotid ultrasound measurements were performed in two of the cohorts. Flow velocity signals were extracted from DICOM images and used to compute the mean flow rate and flow pulsatility indices (for both ICA-L and ICA-R separately), for a total of four features. These features encapsulated the baseline total CBF, but did not contain region-specific effects.

ASL Features: Pseudo-continuous ASL-MRI parameters were: TR/TE, 4,000 ms/14 ms; flip angle, 40°; FOV, 240 mm × 240 mm; matrix size, 80 × 80; 17 slices; thickness, 7 mm; labelling duration, 1.65 s; post-labelling delay, 1.525 s; and labelling gap, 20 mm. The ASL-MRI CBF maps were registered against the sMRI using SPM12 and equipped with maximum probability tissue labels defined on the MNI152 atlas, provided by Neuromorphometrics, Inc. from data
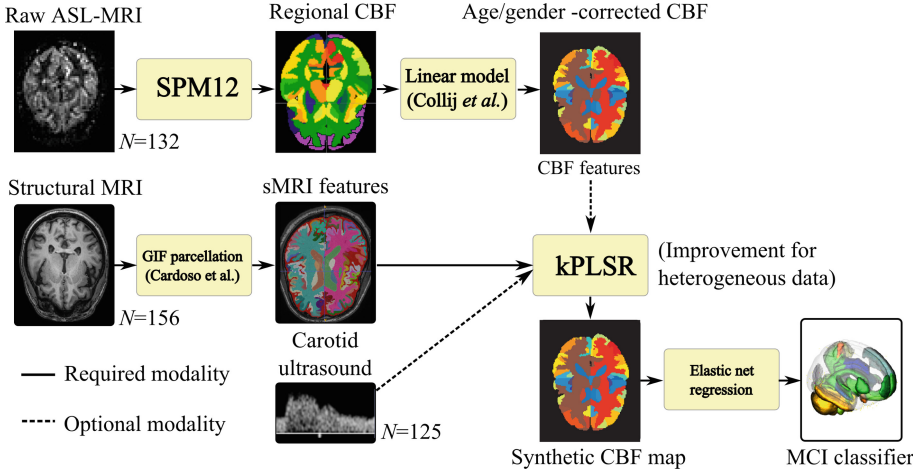
Fig. 1. Workflow for extracting features from heterogeneous medical imaging data and training a kPLSR model for CBF feature synthesis. An MCI classifier is trained with both synthetic and real ASL-MRI features for comparison.

collected in the OASIS project (http://www.oasis-brains.org/), and CBF was estimated in 144 anatomical regions. Regional CBF values were normalised for age and sex by using the w-scores method of [3] to obtain the final ASL features:

$$w_n = \frac{\text{CBF}_n - (\beta_0 + \text{gender}_n \cdot \beta_1 + \text{age}_n \cdot \beta_2)}{\text{SD of residuals}}, \tag{1}$$

where $\text{CBF} = \beta_0 + \text{gender} \cdot \beta_1 + \text{age} \cdot \beta_2 + \varepsilon$ is a linear model trained separately for all anatomical regions on the sub-population of cognitively healthy controls.

Table 1. Study cohorts contributing data to this work. AD cases not used in classification were included in the kPLSR-model training to increase coverage.

Cohort	Field strength	sMRI	Carotid	ASL	AD	MCI	Controls	Age
Cohort 1	3.0T	x		x	13	14	28	65 ± 11
Cohort 1	3.0T			x	13	14	13	66 ± 12
Cohort 2	1.5T	x	x		0	53	48	74 ± 8
Cohort 3	3.0T		x	x	0	10	15	75 ± 6
Cohort 3	3.0T			x	0	8	4	77 ± 9
Total cases		156	126	132	26	99	108	

2.2   Feature Synthesis of the Regional CBF Maps

Given a matrix of inputs $X \in \mathbb{R}^{n \times p}$ and a matrix of outputs $Y \in \mathbb{R}^{n \times m}$, partial least squares-regression (PLSR) attempts to find a lower-dimensional representation of the input-output map using only $\ell \ll p$ latent variables. This is achieved by simultaneous approximate decomposition of the two matrices as:

$$X \approx TP^T, \quad Y \approx UQ^T, \tag{2}$$

where $T, U \in \mathbb{R}^{n \times \ell}$ are the loading matrices for X and Y respectively, and the scores $P \in \mathbb{R}^{p \times \ell}$, $Q \in \mathbb{R}^{m \times \ell}$ maximise the covariance $\mathrm{Cov}(T^T X, U^T Y)$. The feature space for synthesising ASL-MRI maps consisted of 3 demographic features (age, weight, height), 141 sMRI features, and 4 carotid flow-features. Thus the maximum number of input features was $p = 148$. A total of $n = 249$ cases were available for learning a model to predict the $m = 137$ CBF features.

   As the relation between CBF and partial volumes of cortical sub-regions in the brain was likely nonlinear, the kernel version of PLSR [6] was used. In this approach, the feature samples X are mapped using a nonlinear map, $\Psi(X)$, and then the standard linear PLSR is performed in the mapped feature space $(\Psi(X), Y)$. The NIPALS algorithm [8] can be formulated in such a way that only inner products of the type $\Psi(x_i)^T \Psi(x_j)$, for $i, j = 1, \ldots, n$ are required. These can then be obtained using the kernel trick as: $K(x_i, x_j) = \Psi(x_i)^T \Psi(x_j)$.

   In the case of partially missing input data, we divided the features into two parts $X = [X_1 \, X_2]$, where X_1 contained the features that are present for all samples, and $X_2 = \emptyset$ whenever the remaining features were missing in the sample X. We then defined the modified kernel function:

$$[\widetilde{K}]_{i,j}(X, X) := \begin{cases} \Psi(x_{i,1})^T \Psi(x_{j,1}), & \text{if } x_i^2 = \emptyset \text{ or } x_j^2 = \emptyset \\ \Psi([x_{i,1} \, x_{i,2}])^T \Psi([x_{j,1} \, x_{j,2}]), & \text{otherwise} \end{cases}, \tag{3}$$

i.e. in the case of partially missing features the kernel function operated only on the subset of available features. Similarly, we divided the output matrix as $Y = [Y_1; Y_2]$ such that $Y_2 = \emptyset$ for all of the cases where the output data was missing, and defined the matrix $S \in \mathbb{R}^{n \times n_1}$ as having ones on the diagonal and zero otherwise. It was used to extend the outputs Y_1 from the restricted space to the full space, SY_1. The rest of the NIPALS algorithm remained the same, as shown in Algorithm 1. The kPLSR estimator $\widehat{X}_n^{\mathrm{CBF}}$ for the CBF features in the nth patient, learned from the demographic variables ($X_{\mathrm{train}}^{\mathrm{demo}}$), sMRI features ($X_{\mathrm{train}}^{\mathrm{sMRI}}$), and carotid flow features ($X_{\mathrm{train}}^{\mathrm{carotid}}$), was then given by the formula:

$$\widehat{X}_n^{\mathrm{CBF}} = \widetilde{K}\left(\mathcal{X}_n; \mathcal{X}_{\mathrm{train}}\right) B, \tag{4}$$

where $\mathcal{X}_{(\cdot)} = [X_{(\cdot)}^{\mathrm{demo}} X_{(\cdot)}^{\mathrm{sMRI}} X_{(\cdot)}^{\mathrm{carotid}}]$ was the combined feature vector. Figure 1 represents the workflow for extracting features, training a kPLSR model for CBF feature synthesis, and training a MCI vs. CHC binary classifier.

Algorithm 1. NIPALS-kPLS algorithm with heterogeneous data

1: $\widetilde{K}^{(0)} = \widetilde{K}(X, X)$ $\triangleright$ compute kernel matrix for mapped features

2: $Y^{(0)} = [Y_1; Y_2]$

3: **for** $\ell = 1, \ldots, \ell_{\max}$ **do**

4:      $u_0 = SY_1^{(\ell-1)}(\,:\,,1)$ $\triangleright$ initialise loading vector in restricted output space

5:      **while** $\|\Delta u_\ell\| > tol$ **do**

6:          $t_\ell = \widetilde{K}^{(\ell-1)} u_\ell, \quad t_\ell \leftarrow t_\ell / \|t_\ell\|$ $\triangleright$ iterate in feature space

7:          $q_\ell = [Y_1^{(\ell-1)}]^T S^T t_\ell$ $\triangleright$ update output score vector

8:          $u_\ell = SY_1^{(\ell-1)} q_\ell, \quad u_\ell \leftarrow u_\ell / \|u_\ell\|$ $\triangleright$ iterate in restricted output space

9:      $\widetilde{K}^{(\ell)} = (\mathbb{I} - t_\ell t_\ell^T) \widetilde{K}^{(\ell-1)} (\mathbb{I} - t_\ell t_\ell^T)$ $\triangleright$ deflate K

10:     $Y^{(\ell)} = Y^{(\ell-1)} - t_\ell t_\ell^T SY_1^{(\ell-1)}$ $\triangleright$ deflate Y in restricted output space

11: $T = [t_1 \, t_2 \, \ldots \, t_{\ell_{\max}}]$ $\triangleright$ assemble X loadings

12: $U = [u_1 \, u_2 \, \ldots \, u_{\ell_{\max}}]$ $\triangleright$ assemble Y loadings

13: $B = \widetilde{K}(X,X)U \left(T^T \widetilde{K}(X,X)U\right)^{-1} T^T SY_1$ $\triangleright$ assemble regression vector

14: **return** (T, U, B)

2.3 Simultaneous Classification and Feature Selection

As the amount of available training data was modest and pre-selected anatomical regions were used instead of voxelwise CBF values, standard elastic net regression (ENR) - classifier techniques were used to train three different classifier:

(i) In **Model A**, the sMRI features X^{sMRI} were used to train an ENR-model:

$$\min_{\beta_0, \beta} \left\{ \frac{1}{2N} \sum_{n=1}^{N} \left(Y_n - \beta_0 - X_n^{\text{sMRI}}\beta\right)^2 + \lambda R(\beta; \alpha) \right\} \tag{5}$$

with the elastic net regularisation term defined as $R(\beta; \alpha) := \frac{(1-\alpha)}{2}\|\beta\|_2^2 + \alpha\|\beta\|_1$. Here $Y_n \in \mathbb{R}$ is the binary MCI diagnosis for the n'th patient, X_n^{sMRI} denotes the sMRI features, β_0 is the model intercept, and β are the regression weights. The continuous model prediction $\widehat{Y}^A = \beta_0 + X^{\text{sMRI}}\beta$ was thresholded to a binary prediction to obtain the standard ROC-curve. The hyperparameters $\lambda > 0$ and $\alpha \in (0,1]$ were chosen to maximise the area under the ROC-curve.

(ii) In **Model B**, the synthesised CBF-features were used to train the model:

$$\min_{\beta_0, \beta} \left\{ \frac{1}{2N} \sum_{n=1}^{N} \left(Y_n - \beta_0 - \widehat{X}_n^{\text{CBF}}\beta\right)^2 + \lambda R(\beta; \alpha) \right\}, \tag{6}$$

where $\widehat{X}_n^{\text{CBF}}$ is the kPLSR estimator (5). In order to measure the effect of using synthetic vs. ASL-MRI-derived CBF values, Model B was trained using two different sets of data. In one case, when CBF features were missing we simply used synthetic CBF features in their place (*MRI + synthetic*). In another case, only the synthetic CBF features were used even if ASL-MRI was available (*synthetic only*). Again the continuous probability was thresholded to a binary prediction.

(iii) In **Model C**, the feature selection was performed simultaneously on both sMRI and CBF features:

$$\min_{\beta_0,\beta_1,\beta_2} \left\{ \frac{1}{2N} \sum_{n=1}^{N} \left(Y_n - \beta_0 - \widehat{X}_n^{\mathrm{CBF}}\beta_1 - X_n^{\mathrm{sMRI}}\beta_2 \right)^2 + \lambda R([\beta_1;\beta_2];\alpha) \right\}.$$

(7)

(iv) In **Model D**, we combined Models A and B by using the maximum probability rule, $\widehat{Y}_n^C = \max\{\widehat{Y}_n^A, \widehat{Y}_n^B\}$. The rationale for this was that a combination of two diagnostic tests with high specificity but lower sensitivity (typical for AD classifiers) may provide more sensitive diagnostic tests while avoiding the problem that simultaneous feature selection favours one set of features over the other. Models C and D were likewise trained using both synthetic CBF features alone and by combining ASL-MRI and synthetic CBF features.

3 Experiments

3.1 Synthetic CBF vs. ASL-MRI Reconstructed CBF

The Gaussian kernel, $K(x_1,x_2) = \exp(-\|x_1 - x_2\|_2^2/d)$, was used in the kPLSR model. This resulted in two model hyperparameters, the kernel width d and the number of latent variables ℓ, that had to be tuned using leave-one-out cross-validation. Only cases where the sMRI features were available ($n = 156$) were used in model training and cross-validation. Out of these, carotid ultrasound and ASL-MRI features were present in 100 and 55 cases, respectively. Hyperparameter values optimising the R^2-statistic were found to be $d = 35$ and $\ell = 2$. Possible bias and standard deviation of the synthesised CBF from ground truth w-score values were measured using a Bland-Altman - plot of w-scores averaged across all regions, separately for the white matter (WM) and gray matter (GM),
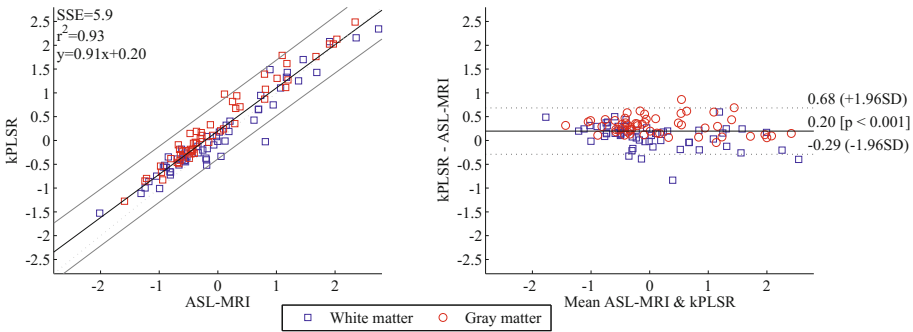

Fig. 2. Bland-Altman plot of kPLSR-modelled vs. ASL-MRI derived w-scores. Mean w-scores averaged over all WM/GM regions reported separately.

see Fig. 2. The mean bias was $\Delta w = 0.20$ ($p < 0.001$). The w-score is normalised so that its standard deviation in the normal population equals 1. The kPLS regressor slightly overestimated CBF in both the WM and GM.

Table 2. Performance of the classifier in the CBF imputation problem (top half), compared with studies in the literature with at least 100 cases (bottom half). ENR = Elastic Net Regression, SVM = Support Vector Machine.

Features used for classification	Classifier	# of cases	AUC	ACC	SENS	SPEC
sMRI	ENR	143	0.76	77%	90%	60%
ASL (MRI+synthetic)	ENR	143	0.74	77%	90%	60%
ASL (synthetic only)	ENR	143	0.65	77%	90%	60%
sMRI+ASL (joint features, MRI+synth.)	ENR	143	0.77	88%	100%	70%
sMRI+ASL (joint features, synth. only)	ENR	143	0.75	83%	100%	50%
sMRI+ASL (max probability, MRI+synth.)	ENR	143	0.77	92%	100%	80%
sMRI+ASL (max probability, synth. only)	ENR	143	0.72	92%	100%	80%
ASL (Collij *et al.* [3])	SVM	160	0.63	60%	60%	60%
sMRI (Liu *et al.* [4])	SVM	454	-	85%	82%	88%
PET (Ortiz *et al.* [5])	SVM	179	0.74	73%	70%	77%
sMRI+PET (Ortiz *et al.* [5])	SVM	179	0.91	86%	90%	82%

3.2 Utility of Synthetic ASL in the CBF Imputation Problem

The MCI classifiers using Models A, B, C, and D were trained with four-fold cross-validation (4-FCV) in a training set of $n = 123$ cases. An additional randomly selected validation set of $n = 20$ cases not included in the training was used to evaluate the balanced accuracy (ACC), sensitivity (SENS), and specificity (SPEC) of each classifier using hyperparameters and cut-offs obtained in 4-FCV. A heat map of the 46 features chosen by Model B is shown in Fig. 3.

We compared our MCI classification accuracy to results reviewed in [1] with the following selection criteria: (i) the MCI vs. CHC classification problem was addressed, (ii) the feature set consisted of sMRI, ASL, or PET features, (iii) the cohort size was at least 100, and (iv) studies that used CSF biomarkers or neurocognitive test scores as features were excluded. The study with the best reported accuracy for each feature set was chosen as representative.

Results of the comparison are given in Table 2. Models A and B alone produced similar results in terms of accuracy, although Model B achieved better accuracy than was reported for ASL-MRI features in [3]. Model C improved the results slightly when MRI+synthetic CBF features were used, but the best results were obtained with Model D regardless or whether MRI+synthetic or synthetic only CBF features were used.
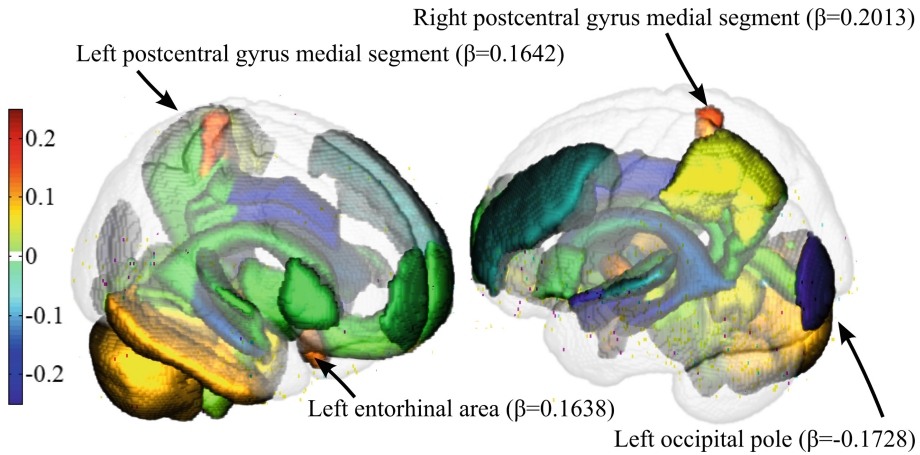
Fig. 3. Heat map of the coefficients β for the CBF imputation problem. A total of 46 regions were chosen as features. Regions with largest coefficients identified.

4   Discussion

Kernel PLS regression on heterogeneous data was used for the robust synthesis of regional CBF values in cases where no ASL-MRI images were available. As was reported in [3], CBF features alone were not particularly informative in MCI classification, but a multi-modal classifier using synthetic CBF features outperformed pure sMRI-based classifiers in a validation test. Best classifier performance (balanced accuracy 92%, sensitivity 100%, specificity 80%) was achieved when a maximum probability-rule was used to combine classifiers using different feature sets. The benefit of our proposed method is that only basic sMRI features (partial volumes of subregions) were used and, as a result, synthetic CBF features can, therefore, be generated in large-scale brain imaging databases, such as ADNI, without the need for extensive feature computation. It is possible that more informative sMRI features, e.g. ventricular and/or hippocampal shape morphometrics, could increase the accuracy of the resultant classifiers. Provided more ASL-MRI data were available, the use of convolutional neural networks on voxelwise CBF values should also be investigated to eliminate the need for pre-selecting anatomical regions for analysis.

References

1. Arbabshirani, M., Plis, S., Sui, J., Calhoun, V.: Single subject prediction of brain disorders in neuroimaging: promises and pitfalls. NeuroImage **145**, 137–165 (2017)
2. Cardoso, M., Modat, M., Wolz, R., Melbourne, A., Cash, D., Rueckert, D.: Geodesic information flows: spatially-variant graphs and their application to segmentation and fusion. IEEE Trans. Med. Imag. **34**(9), 1976–1988 (2015)
3. Collij, L., et al.: Application of machine learning to arterial spin labeling in mild cognitive impairment and Alzheimer disease. Radiology **281**(3), 865–875 (2016)
4. Liu, M., Zhang, D., Shen, D.: Hierarchical fusion of features and classifier decisions for Alzheimer's disease diagnosis. Hum. Brain Mapp. **35**(4), 1305–1319 (2014)
5. Ortiz, A., Munilla, J., Álvarez-Illán, I., Górriz, J., Ramírez, J.: Exploratory graphical models of functional and structural connectivity patterns for Alzheimer's disease diagnosis. Front. Comput. Neurosci. **9**, 132 (2015)
6. Rosipal, R., Trejo, L.J.: Kernel partial least squares regression in reproducing kernel Hilbert space. J. Mach. Learn. Res. **2**, 97–123 (2001)
7. van Tulder, G., de Bruijne, M.: Why does synthesized data improve multi-sequence classification? In: Navab, N., Hornegger, J., Wells, W.M., Frangi, A.F. (eds.) MIC-CAI 2015. LNCS, vol. 9349, pp. 531–538. Springer, Cham (2015). https://doi.org/10.1007/978-3-319-24553-9_65
8. Wold, S., Geladi, P., Esbensen, K., Öhman, J.: Multi-way principal components and PLS analysis. J. Chemom. **1**, 41–56 (1987)