Predicting tissue outcome in acute ischemic stroke using projection pursuit regression

K. Nagenthiraja[1], K. Y. Jonsdottir[1], L. Østergaard[1], and K. Mouridsen[1]
[1]Dept. of Neuroradiology, Århus University Hospital, CFIN, Århus C, Denmark

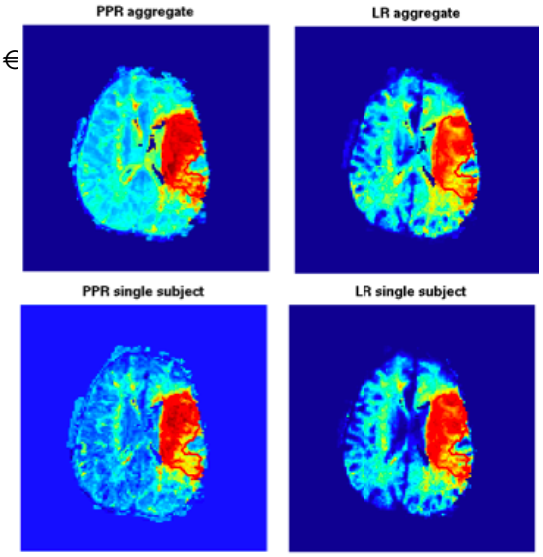
Introduction The early identification of tissue at risk of infarction after acute ischemic stroke may aid clinical decision-making and potentially improve long-term patient outcome. The perfusion-diffusion mismatch is known to overestimate final infarct size [1], representing an over-simplified dichotomization based on diffusion-weighted imaging (DWI) and perfusion-weighted imaging (PWI) maps of transit time delays. Instead, Generalized linear models such as logistic regression (LR), combining voxel-wise outcome information (infarct or non-infarct based on structural imaging) with multiple DWI and PWI indices as predictor variables have been proposed as a means of quantifying regional risk of infarct progression [2]. An increasing number of perfusion (MTT, Tmax, TTP, First moment, FWHM etc) and diffusion metrics (DWI signal intensity, ADC, FA etc) have been hypothesized to signal risk of subsequent risk. There is therefore a growing need for reducing redundancy in predictive model input maps while retaining the most important features of the input space. Here we combine these aims by proposing a method, which simultaneously identifies optimal projections of input data to obtain a reduced set of predictors, and models the relation between the resulting predictors and outcome using flexible non-linear ridge functions.

Theory Projection pursuit regression (PPR) is a supervised extension of the projection pursuit (PP) proposed by Friedman and Stuetzle [3]. The PP is a dimension reduction method identifying non-Gaussian, low dimensional projections. The PPR method estimates the relation between these projections and outcome using smooth non-linear ridge functions [4], i.e.

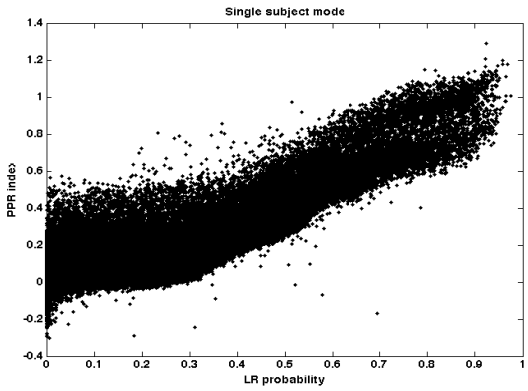
$$f(x) = \alpha_0 + \sum_{i=1}^{I} \beta_i f_i(\alpha_i^T x),$$

where the $\alpha_i^T x$ terms represent the low dimensional projection of the original predictors, and $f_i()$ are the ridge functions.

Methods Acute (< 12 hours) diffusion- and perfusion-weighted images were acquired in 16 acute stroke patients. Maps of isotropic DWI, ADC, T2, CBF, CBV and MTT maps were calculated and co-registered. The PPR and LR models were trained according to two strategies; a single-subject model fitted to the data of each patients and an aggregate model based on pooled data from all patients. For training we used a balanced set of data from each patient containing an equal number of voxels from DWI lesion + mismatch and contra-lateral unaffected brain. The aggregate model was tested by a boot-strapping method. In all tests, performance was quantified by means of AUC scores. PPR and LR were performed in R, using the built-in function `ppr`.



Results For the single-subject models, PPR AUC scores were significantly better than LR. The median difference, AUC_{PPR}-AUC_{LR} was 0.0448 (IQR: 0.0332 - 0.0647: P<0.0001). When testing the model with the bootstrapping method no significant differences were observed in AUC scores (Median 0.0064; IQR 0.0061 - 0.0253). However PPR seems to outperform LR for lesions larger than 50 ml. For the aggregate models, AUC scores were 0.8067 for PPR and 0.7792 for LR, respectively. In figure, the prediction maps of the aggregate model and single-subject model are shown. The maps indicate variation in predicted probabilities in normal tissue for LR, while PPR does not show this variation, which may imply a more clear delineation of the actual infarct. The AUC scores for the maps were 0.8349 for PPR aggregate, 0.8130 for LR aggregate, 0.8578 for PPR single-subject and 0.8238 for LR single-subject, respectively. In the bottom figure, we see a non-linear relationship of the LR probabilities and the PPR indices, which indicate that the PPR indices are not a linear scaling of LR probabilities.



Discussion We have presented a flexible procedure for predicting final tissue outcome combining a dimension reduction method with efficient regression modeling of the resulting low-dimensional projections using smooth non-linear ridge functions. Increase in performance was observed for PPR notably in single-subject models, while performance in data pooled across patients was similar to LR, possibly owing to increased heterogeneity (reperfusion status, time of scan). We speculate that PPR in combination with techniques for identifying homogenous patient subgroups could further improve quantification of regional risk of infarct progression.

References [1] Sorensen et al, Radiology, 1999. [2] Wu et al, Stroke, 2001. [3] Friedman et al. 1981, JASA. [4] Friedman, 1984, SMART User's Guide.