Artificial neural networks in radiological predictive models

Nikolaos Dikaios[1], Taiki Fujiwara[2], David Atkinson[3], and Shonit Punwani[2]
[1]Department of Medical Physics and Bioengineering, University College London, London, Greater London, United Kingdom, [2]Department of Radiology, University College London Hospital, [3]University College London, Centre for Medical Imaging

Purpose: Predictive models are being increasingly employed in radiology as diagnostic aids for cancer detection. A variety of model types exist. Linear discriminant analysis (LDA) models assume linearity, normality and that the input variables are independent, assumptions which may affect classification accuracy. Neural networks (NN) whilst less intuitive, do not make these assumptions and can detect complex non-linear relationships between the input variables. Both LDA and NN are prone to over-fitting [1, 2]. In this work we compared the performance of multilayer perceptron (MLP) artificial NN [3] and LDA models for prediction of transition zone (TZ) prostate cancer (based on quantitative multi-parametric MRI variables) using a leave-one-out (LOO) and a 2-fold cross validation analysis [4].

Methods: *Data selection:* Thirty-one individual input variables were recorded in each of 73 men. The target (dependent) variable was the presence or absence of cancer confirmed by post-MRI template mapping biopsy of the prostate. Twenty-eight input variables were derived from the multi-parametric MRI datasets composed of T2 weighted (T2w) images, apparent diffusion coefficient (ADC) maps and arterial contrast enhanced (aCE) images. Prostate specific antigen, age of the patient, and gland volume were also recorded as input variables.
Predictive models: The MLP neural network had three layers (input, hidden and output layer), with a feed-forward architecture. Only one hidden layer was selected to make the model less complex and avoid over-fitting. The MLP model was trained by the back propagation and conjugate gradient descent algorithms in a batch manner. The data for the MLP algorithm was randomly partitioned into a training (70%) and a testing (30%) dataset. The testing dataset was independent of the training dataset and was used to avoid over-fitting of the model.
Model comparison: The performance of the two models was evaluated with the area under curve (AUC) of a receiver operating characteristic (ROC) curve. The AUCs were estimated by bootstrap resampling (1000 replicates) with the patient as the unit of analysis. As data was from a moderate number of patients a LOO and a 2-fold cross-validation was chosen. For the 2-fold the model was trained with 52 observations and tested with the other 21 observations. The 2-fold cross validation was randomly repeated 100 times, maintaining the ratio between training and test observations.

Results: The classification accuracies for the original dataset were 92% for the LDA model and 93% for the MLP model in the training set (and 79% for both in the testing set). Figure 1 illustrates the performance of the two models on independent samples. The average classification accuracy over the 100 repetitions for the 2-fold cross validation is 67% for the LDA and 74% for the MLP. The minimum/maximum classification accuracy for the two models was 43/90% (LDA) and 48/95% (MLP). The LOO cross validation had 72% classification accuracy for the LDA model and 78% for the MLP model. Both models achieved a high AUC for the bootstrap sample (Table I).

Table I: Mean AUC of the 1000 bootstrap sub-samples for the LDA and MLP predictive model. CI is the confidence interval.

			Asymptotic 95% CI	
	AUC	Std Error	Lower Bound	Upper Bound
LDA	0.971	0.016	0.942	0.993
MLP	0.969	0.019	0.971	1.000

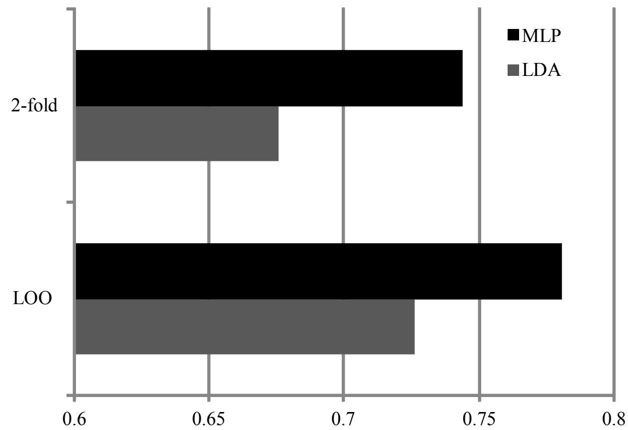

Figure 1: Classification accuracy of the LOO and 2-fold cross validation for the MLP and the LDA predictive model. An observation is assigned to a class if the predicted probability is higher than 0.5.

Conclusion: Linear models such as LDA are popular because they are not computationally demanding and are easy to interpret but they make certain assumptions, whereas neural networks do not but are more complex and computationally demanding. Both models examined in this work were built on 31 input variables, and achieved high classification accuracy and AUC on their training datasets. The MLP neural network model was 6% more accurate than the LDA model for the LOO and 7% more accurate for the 2-fold validation suggesting better performance.

References: [1] Luo D et al Linear Discriminant Analysis: New Formulations and Overfit Analysis 25th AAAI Conference 2011 417-422 [2] Tu J Advantages and disadvantages of using artificial neural networks versus logistic regression for predicting medical outcomes J Clin Epidemiol 49 1225-1331 (1996) [3] Biganzoli E et al Artificial neural network for the joint modelling of discrete cause-specific hazards Artificial Intelligence in Medicine 37119-130 (2006) [4] Efron B Cross-validation and the bootstrap: Estimating the error rate of a prediction rule J Am Stat Assoc 92 548-560 (1997)