1 **S1 File. Probability of correct class (PCC)**
2

3 DSI defines the patient's similarity to each diagnostic group. For example, the following DSI

4 values could have been measured for a patient: DSI=0.75 for AD, DSI=0.67 for FTD, DSI=0.45

5 for VaD and DSI=0.32 for controls. The interpretation is that AD is the most similar diagnostic

6 group, but the question is how confident we can be about this suggestion. *Probability of*

7 *correct class* (PCC) for this patient measures the probability that the diagnosis suggested is

8 correct when the highest DSI value is 0.75 and the second highest DSI value is 0.67.

9

10 Intuitively, one could take a large database of previously diagnosed patients, search all cases

11 with the highest DSI value close to 0.75 and the second highest close to 0.67, and calculate

12 the share of database cases for which the diagnostic group of the highest DSI value

13 corresponds to the ground truth diagnosis. The open question is which cases are considered

14 to be "close" and included in defining PCC.

15 Formally, PCC is defined by maximizing the likelihood function of the observed data:

16 $$\mathcal{L}(\theta|x) = PCC_\theta(x)$$

17 where θ denotes the parameters used to estimate the probability function *PCC(x)* and *x*

18 denotes a vector containing the highest DSI value and the difference to the second highest DSI

19 value, {DSI_{max}, ΔDSI). Maximizing likelihood defines a probability function that is the most likely

20 for the observed data. If we consider all observed data samples independent, the likelihood

21 of measuring all the cases is

22 $$PCC_\theta(x) = \prod_{i=1}^{n} PCC_{\theta,i}(x)$$

23 Maximizing this function is similar to maximizing its logarithm

24 $$\log PCC_\theta(x) = \sum_{i=1}^{n} \log PCC_{\theta,i}(x)$$

25 Probability density function *PCC(x)* is estimated from n d-dimensional samples x_i (observed

26 data) using a Gaussian kernel:

27 $$PCC_H(x) = \frac{1}{n}\sum_{i=1}^{n} K_H(x - x_i) \qquad (1)$$

28 where

29 $$K_H(x) = (2\pi)^{-d/2}|H|^{-1/2}e^{-\frac{1}{2}x^T H^{-1} x}.$$

30

31 In this case, θ is the width of the kernel, *H*, and *PCC(x)* estimates the probability of the class

32 "correctly classified". As we want to maximize balance accuracy, i.e., not to prefer the most

33 prevalent diagnoses, each sample is weighted based on the share of all samples in the class.

34 Log-likelihood is optimized using cross-validation: samples are divided using a random

35 selection to a training and test sets. The probability function is estimated using the training

36 set with a specific kernel size and likelihood is computed using the test set. This is repeated

37 multiple times and the kernel width maximizing the sum of likelihoods over all iterations is

38 chosen. Because the size of the database is still relatively limited, we chose to estimate a one-

39 dimensional kernel (circle – one parameter) instead of a two-dimensional kernel (ellipse – two

40 parameters). Figure S1 shows an example on estimating PCC for a database.

41

42 **Figure S1**. Example on estimating PCC.
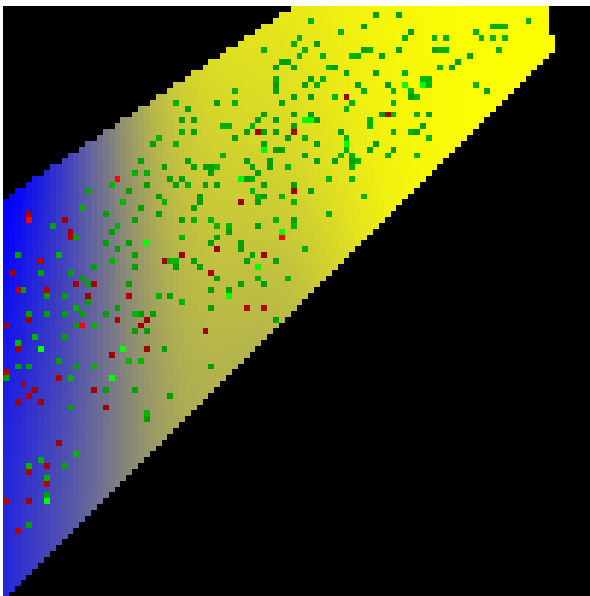
43

44 The vertical axis shows the highest DSI value (0.5-1.0) and the horizontal axis the difference between the two

45 highest DSI values (0.0-0.67). The probabilities are color coded from blue (p=0.5) to yellow (p=1.0). The dots show

46 all the samples: green are correctly classified cases and red ones are wrongly classified cases. The shade of the

47 green and red dots indicates the class: the darker dots have more cases in the class (less weight in estimating the

48 probability) than the brighter dots. The black background indicates the value pairs (DSI_{max}, ΔDSI) that are not

49 possible. PCC shows that when DSI_{max} and/or ΔDSI are high, almost all points in the neighborhood are green and

50 PCC starts to get close to one.