

# A Disease State Fingerprint for Evaluation of Alzheimer's Disease

Jussi Mattila<sup>a,\*</sup>, Juha Koikkalainen<sup>a</sup>, Arho Virkki<sup>a</sup>, Anja Simonsen<sup>b</sup>, Mark van Gils<sup>a</sup>,  
Gunhild Waldemar<sup>b</sup>, Hilikka Soininen<sup>c</sup>, Jyrki Lötjönen<sup>a</sup> and for The Alzheimer's Disease  
Neuroimaging Initiative\*\*

<sup>a</sup>*VTT Technical Research Centre of Finland, Tampere, Finland*

<sup>b</sup>*Department of Neurology, Section 2082, The Copenhagen Memory Clinic & The Memory Disorders  
Research Group, Rigshospitalet, Copenhagen University Hospital, Copenhagen, Denmark*

<sup>c</sup>*Department of Neurology, Kuopio University Hospital, Kuopio, Finland*

Accepted 26 May 2011

**Abstract.** Diagnostic processes of Alzheimer's disease (AD) are evolving. Knowledge about disease-specific biomarkers is constantly increasing and larger volumes of data are being measured from patients. To gain additional benefits from the collected data, a novel statistical modeling and data visualization system is proposed for supporting clinical diagnosis of AD. The proposed system computes an evidence-based estimate of a patient's AD state by comparing his or her heterogeneous neuropsychological, clinical, and biomarker data to previously diagnosed cases. The AD state in this context denotes a patient's degree of similarity to a previously diagnosed disease population. A summary of patient data and results of the computation are displayed in a succinct Disease State Fingerprint (DSF) visualization. The visualization clearly discloses how patient data contributes to the AD state, facilitating rapid interpretation of the information. To model the AD state from complex and heterogeneous patient data, a statistical Disease State Index (DSI) method underlying the DSF has been developed. Using baseline data from the Alzheimer's Disease Neuroimaging Initiative (ADNI), the ability of the DSI to model disease progression from elderly healthy controls to AD and its ability to predict conversion from mild cognitive impairment (MCI) to AD were assessed. It was found that the DSI provides well-behaving AD state estimates, corresponding well with the actual diagnoses. For predicting conversion from MCI to AD, the DSI attains performance similar to state-of-the-art reference classifiers. The results suggest that the DSF establishes an effective decision support and data visualization framework for improving AD diagnostics, allowing clinicians to rapidly analyze large quantities of diverse patient data.

**Keywords:** Alzheimer's disease, automatic, biomarkers, computer-assisted, decision making, information processing, projections and predictions

Supplementary data available online: <http://dx.doi.org/10.3233/JAD-2011-110365>

## INTRODUCTION

Diagnosing Alzheimer's disease (AD) is a non-specific, subjective, and error-prone process, especially in the early phases of the disease [1]. Because of their inherent difficulty, diagnoses often come late, taking up to two years after initial memory problems occur [2]. Current criteria for AD require early and dominating decline in episodic memory supported by abnormal biomarkers [3, 4]. If a patient with objective evidence of cognitive impairment does not yet meet the criteria for AD or for other dementia, he or she

\*\* Data used in preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (<http://www.loni.ucla.edu/ADNI>). As such, the investigators within the ADNI contributed to the design and implementation of ADNI and/or provided data but did not participate in analysis or writing of this report. A complete listing of ADNI investigators can be found at: [http://adni.loni.ucla.edu/wp-content/uploads/how\\_to\\_apply/ADNI\\_Authorship\\_List.pdf](http://adni.loni.ucla.edu/wp-content/uploads/how_to_apply/ADNI_Authorship_List.pdf).

\*Correspondence to: Jussi Mattila, VTT Technical Research Center of Finland, Tekniikankatu 1, FIN-33101, Tampere, Finland. Tel.: +358 40 592 7979; Fax: +358 20 722 3499; E-mail: [jussi.mattila@vtt.fi](mailto:jussi.mattila@vtt.fi)

is usually diagnosed as having mild cognitive impairment (MCI) [5]. MCI is a heterogeneous state with several possible outcomes and is associated with an increased risk of developing AD, particularly when memory impairment is the predominant symptom [6]. For early diagnosis of AD, a key issue is finding tests and biomarkers that determine which subjects with MCI will develop AD. Here, the term ‘biomarker’ is used in a broad sense, encompassing biologic features obtained by any and all detection modalities providing information about the disease.

Diverse sets of neuropsychological tests and biomarkers have been investigated for their efficacy to predict conversion from MCI to AD [7–11], and several studies have shown that combining results can yield even better predictions [12–14]. Increased knowledge about cognitive tests and biomarkers has influenced a recent proposal for a new lexicon, where the term AD encompasses the whole spectrum of the disease from predementia to dementia phases and further emphasizes the combination of clinical and biomarker data [15]. However, results from many of the studies are not easily applied in daily diagnostic work. They may require a specific test pattern that is not available or is incompatible due to local or national differences in execution. Occasionally, the statistical analysis methods lack transparency, making them hard to incorporate into local decision making processes. Ultimately, despite all attempts, there have not yet been findings that would comprehensively differentiate MCI subjects who develop AD (progressive MCIs, PMCI) from those who do not (stable MCIs, SMCI).

New approaches for improving the diagnostic process in AD are needed. Computer-based analyses of patient data can quantify information with good diagnostic accuracy, in some cases comparable to experienced clinicians [16]. Tools that help manage the constantly increasing amounts of complex patient data can increase the quantity of information clinicians can examine, and can reveal subtle aspects of information that are buried under a wealth of clinical data [17, 18]. Clinical decision support systems (CDSS) have shown their potential in reducing medical errors and increasing health care quality and efficiency [19–21]. Visualization techniques for analyzing biomedical and temporal data are already commonplace [22, 23], and novel clinical information visualization solutions are constantly being developed [24–26]. Consequently, a statistical Disease State Index (DSI) method is proposed for deriving a scalar value denoting the AD state or progression of AD in suspected AD patients. In this context, AD state mea-

asures similarity of patient data to previously-diagnosed healthy and AD populations. While the DSI provides yet another piece of information to clinicians, its goal is to distill existing patient data to a few parameters at a high abstraction level, allowing them to quickly find relevant information and disregard irrelevant information. A Disease State Fingerprint (DSF) visualization technique is also proposed for displaying patient data and DSI values in a concise and interpretable format, extended from earlier research in another biomedical domain [27]. Together, they offer a decision support system that allows clinicians to rapidly extract knowledge from large quantities of heterogeneous patient data and combine them with personal expertise for making the diagnosis.

The main contributions of this work are the proposal of a novel patient data visualization technique (DSF) and the definition of an underlying statistical method for modeling progressing disease state (DSI). The DSI is evaluated against state-of-the-art classifiers using baseline data from the Alzheimer’s Disease Neuroimaging Initiative (ADNI); its ability to discriminate healthy elderly controls, SMCI, PMCI, and ADs and its capability to predict conversion from MCI to AD are considered. Interpretation of the resulting DSF visualizations and characteristics of the proposed system are reviewed to assess their clinical applicability.

## MATERIALS AND METHODS

### *Alzheimer’s disease neuroimaging initiative (ADNI)*

ADNI is a longitudinal 5-year study of AD conducted in the USA and Canada, with the goal of developing and validating surrogate markers for early detection and monitoring of AD progression. After launching in late 2004, approximately 800 participants, ranging in age from 55 to 90 years, were recruited for the study: 200 healthy elderly controls, 400 patients with diagnosed MCI, and 200 with early diagnosed AD. Follow-ups of ADNI participants were done by telephone or in person every 6 to 12 months for a period of two to three years. All participants underwent repeated cognitive and neuropsychological testing and magnetic resonance imaging (MRI) scanning. Other tests, including positron emission tomography (PET) and lumbar puncture providing cerebrospinal fluid (CSF) samples, were done more infrequently and not necessarily for all participants. Data from the study are freely available to researchers in an online database at the UCLA Laboratory of Neuroimaging (LONI)

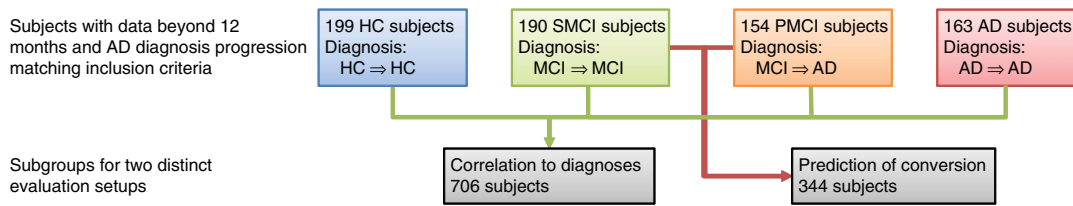


Fig. 1. Patients were divided into two overlapping subgroups. Correlation to actual diagnoses was evaluated using all subjects, capability to predict conversion from MCI to AD using only SMCI and PMCI subjects.

website (<http://www.loni.ucla.edu/ADNI/>). The site also provides exact information regarding ADNI neuroimaging instrumentation, procedures, and data processing.

*Study cohorts and data selection*

The analyses in this paper included ADNI subjects who *a)* had follow-up data available beyond the 12-month visit period and *b)* belonged to one of four diagnostic groups based on the baseline diagnosis and the latest diagnosis available in the database (accessed on September 2, 2010). The first three diagnostic groups included subjects whose latest diagnosis was the same as the baseline diagnosis, particularly elderly healthy controls (HC,  $n = 199$ ), stable MCIs (SMCI,  $n = 190$ ), and Alzheimer’s disease (AD,  $n = 163$ ). The fourth group was a progressive MCI group (PMCI,  $n = 154$ ), whose diagnosis at the baseline of the ADNI study was MCI, but had converted to AD (on average after 19 months) over the course of the study. Patients whose diagnosis had changed otherwise, such as from MCI or AD to healthy subjects, were excluded from this study. Study cohort selection is illustrated in Fig. 1 and demographic data for the diagnostic groups are presented in Table 1.

All analyses were made using baseline measurement data readily available from the ADNI database. Specifically, patient data obtained from six baseline tests were used; Alzheimer’s Disease Assessment Scale-cognitive subscale (ADAS), Mini-Mental State Examination (MMSE), Trail making test from Neuropsychological Battery (TMT), MRI derived volumes (MRI), amyloid- $\beta$  and total tau from CSF, and apolipoprotein E (APOE). Altogether, analyses were run with 66 unique patient variables distributed into the ten distinct datasets, illustrated in Fig. 2. Sparse and incomplete data were intentionally included to parallel a realistic clinical setting where not all tests are administered to all patients. In ADNI, automated volumetric segmentation of MRI was performed with the Freesurfer image analysis suite [28], which is

Table 1

Demographic and clinical data of the healthy control (HC), stable mild cognitive impairment (SMCI), progressive mild cognitive impairment (PMCI, average conversion time 19 months from baseline), and Alzheimer’s disease (AD) groups

	HC	SMCI	PMCI	AD
Subjects	199	190	154	163
Diagnosis				
Baseline	HC	MCI	MCI	AD
Latest	HC	MCI	AD	AD
Gender				
Male	104 (52%)	125 (66%)	93 (60%)	87 (53%)
Female	95 (48%)	65 (34%)	61 (40%)	76 (47%)
Demographics, years				
Age	75.5 ( $\pm 5.1$ )	74.8 ( $\pm 7.6$ )	74.2 ( $\pm 6.9$ )	74.7 ( $\pm 7.5$ )
Education	16.1 ( $\pm 2.8$ )	15.8 ( $\pm 3.1$ )	15.6 ( $\pm 2.9$ )	14.9 ( $\pm 3.1$ )
Available baseline data				
MMSE	199 (100%)	190 (100%)	154 (100%)	163 (100%)
ADAS	199 (100%)	189 (99%)	152 (99%)	160 (98%)
TMT	199 (100%)	186 (98%)	153 (99%)	156 (96%)
MRI	190 (95%)	171 (90%)	135 (88%)	137 (84%)
CSF	102 (52%)	94 (49%)	83 (54%)	90 (55%)
APOE	199 (100%)	190 (100%)	154 (100%)	163 (100%)

The data are expressed as counts and (percentages) of available data except for age and education, which are expressed as mean ( $\pm$ standard deviation).

documented and freely available for download online (<http://surfer.nmr.mgh.harvard.edu/>). Composite variables and summaries of test patterns, e.g., total MMSE score and ADAS 13 point total, were excluded from the datasets, since the same information was contained within the individual variables.

*Disease state index*

To improve interpretability of heterogeneous patient data, a statistical DSI method has been developed, deriving a scalar index value indicating the state of AD in a patient. The rationale of the DSI is to provide additional evidence-based information by comparing patient data as a whole to a high number of other cases with or without the disease. It is principally intended to

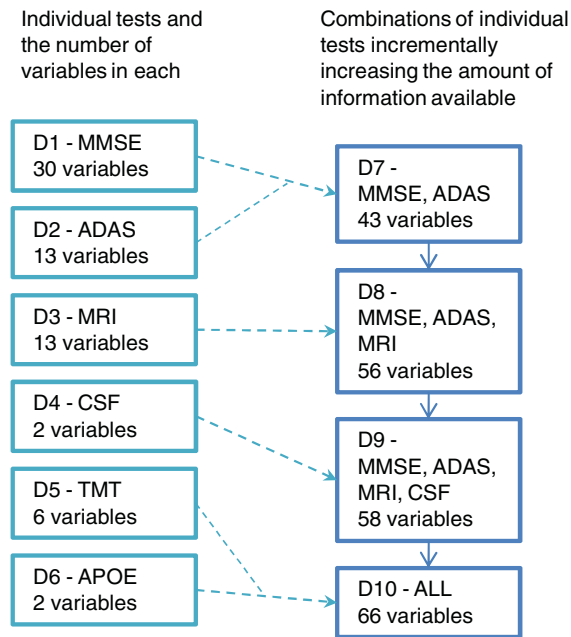


Fig. 2. Analyses were run with ten distinct datasets (D1–D10), formed using variables from six individual tests. The combinations of tests emulate the effect of having incremental tests done, gradually increasing knowledge about the patient. The tests and variable counts included in each dataset are presented in the diagram.

be used with quantitative patient data, such as standardized neuropsychological tests, laboratory test results, and computer-based analyses of medical imaging data. Applying the DSI to patient data results in a value between zero and one, indicating the patient's disease state or progression of the disease. The DSI values are assumed to lie on an interval scale, i.e., one unit on the scale represents the same magnitude across the whole range of the scale. Increasing values of DSI indicate an increasing similarity to AD population, based on the available data. More specifically, DSI measures how individual measurement values and patient data as a whole match the disease profile as defined from a large number of known disease cases.

DSI is data agnostic and can be used with any data available. It can determine the disease state between healthy and typical AD, healthy and atypical AD, MCI and AD, and potentially between other dementias and diseases, as long as the training data are available. DSI is also designed to be highly dynamic, not requiring particular tests but using any data acquired and available for the patient being studied. Together, these properties facilitate application of the method at various clinics and re-evaluation of patient data as more test results become available. Choosing a decision

boundary allows the DSI to be considered a supervised classifier, discriminating between healthy and diseased patients. Several design requirements were imposed on the DSI according to five classifier performance categories defined by Han and Kamber [29], listed in Table 2.


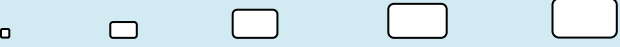
DSI values are computed from patient data in three relatively simple steps. First, each individual patient measurement value, e.g., a single answer in ADAS or the volume of a brain structure derived from MRI, is compared to previously known training data using a *fitness* function. A *fitness* function computes the DSI value for a single patient measurement revealing which population, healthy or diseased, the value fits best. Second, observing only values from known control and disease populations, the *relevance* of each variable is computed, independent of the patient measurement. *Relevance* indicates how well a variable is able to discriminate between the known healthy and diseased populations. Evaluation of *relevance* results in a value between zero and one, obtaining larger values as the separation between control and disease populations increases. Interpretations for different values of DSI and *relevance* are listed in Table 3 and they are derived in full detail in supplementary material (available online: <http://www.j-alz.com/issues/27/vol27-1.html#supplementarydata07>). Third, DSI and *relevance* values are combined as a weighted arithmetic mean, where DSI values of individual patient measurements are weighted by the variable *relevancies*, to obtain composite DSI values for tests done with the

Table 2  
Design goals for the Disease State Index method

Category	Goals for Disease State Index
Interpretability	Provides well-behaving index values that concur with severity of disease state Uses original measurement values in analysis and for reporting the results Facilitates development of interpretable visualizations for expert analyses Accommodates varying clinical and research questions
Prediction accuracy	Classification performance should be comparable to state-of-the-art classifiers
Robustness	Not all patients need to have the same set of tests performed Must be able to use any quantifiable data and all types of variables Missing data should not impose problems for using the method
Computational speed	Allow refinement of parameters and updating of results at interactive rates
Scalability	Enable computation of the model on the fly or beforehand as necessary

Table 3

Interpretations and visualizations of DSI and *relevance*. DSI is computed by comparing the patient values to training data, *relevance* is computed from the known control and diseases population values alone

DSI	0.0	0.5	1.0
<b>Interpretation</b>	Patient value matches the healthy controls perfectly	Patient value falls between control and disease populations, matching both equally well	Patient value matches the AD population perfectly
<b>Visualization</b>	Blue color	White color	Red color
			
Relevance	0.0	0.5	1.0
<b>Interpretation</b>	Not relevant for estimating disease state; variable does not differentiate between known control and disease populations	Relevant for estimating disease state; discrimination capability is halfway between random and perfect discrimination	Very relevant for estimating disease state; variable discriminates perfectly between control and disease populations
<b>Visualization</b>	Excluded from visualization	Intermediate box size	Large box size
			

patient, such as for ADAS and MRI imaging. Correlations between variables can be accounted for at this step, e.g., by applying principal component analysis (PCA) [30].

To obtain a total DSI value representing the combination of all data from multiple tests, the three steps described above are repeated recursively. In lieu of raw measurement values, the DSI values from the previous step are now used for evaluating *relevance* and *fitness*, and merged into a total DSI value (see Fig. 3).

The combination of DSI and *relevance*, schematically depicted in Fig. 4, capture the essence of patient data in relation to the disease. DSI values indicate which patient data are similar to the AD population and *relevance* specifies how important that information should be considered based on previously diagnosed cases. A large DSI value and large *relevance* for a neuropsychological test, for example, indicate that the patient performed similarly to known AD population and that the test has previously been able to discriminate between healthy and AD patients with high accuracy. On the other hand, a test with a large DSI value but little or no *relevance* may usually be ignored,

since the test is unable to differentiate between the populations.

#### Disease state fingerprint

In an analogy to the unique human fingerprints and DNA fingerprints, DSF visualization forms patterns, enabling quick visual inspection of unique disease and patient data at multiple levels of abstraction. In DSF, the patterns emerge from a tree of nodes rendered according to the DSI organization, using shapes and colors to quickly identify the patient’s disease state. Specifically, shades of colors indicate DSI values while *relevance* is indicated by node sizes (see Table 3).

The DSF tree allows rapid but detailed reviewing of raw patient measurement data, DSI values, *relevance* values, and the study of their relationship to the disease profile (see Fig. 4). Measures that have zero *relevance* are by default hidden from the DSF visualization. Interactive implementation of the DSF allows visualizations of data distributions (see Fig. 5) and ‘drill-down’/‘roll-up’ operations common to data mining and visual analytics [29]. These operations can be

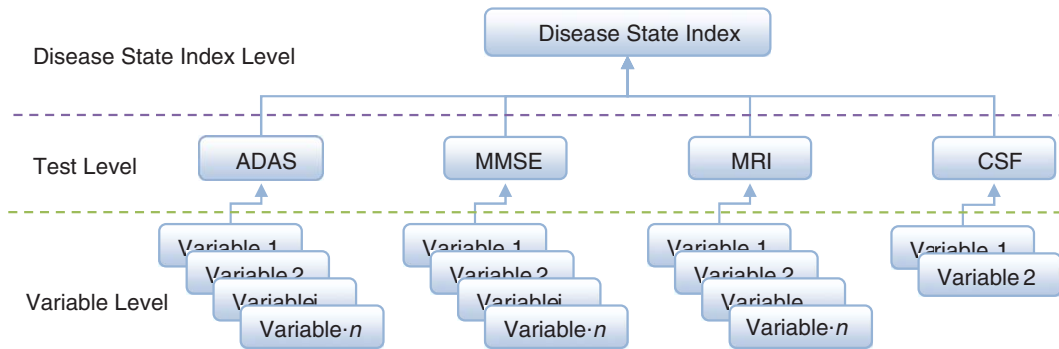


Fig. 3. Organization of a DSI / DSF visualization tree. The tree structure follows organization of patient data consisting of individual variable values (leaf nodes at *Variable Level*), performed tests (internal nodes at *Test Level*), and the resulting total Disease State Index (root node at *Disease State Index Level*). Additional levels can be employed to modify the granularity of the tree.

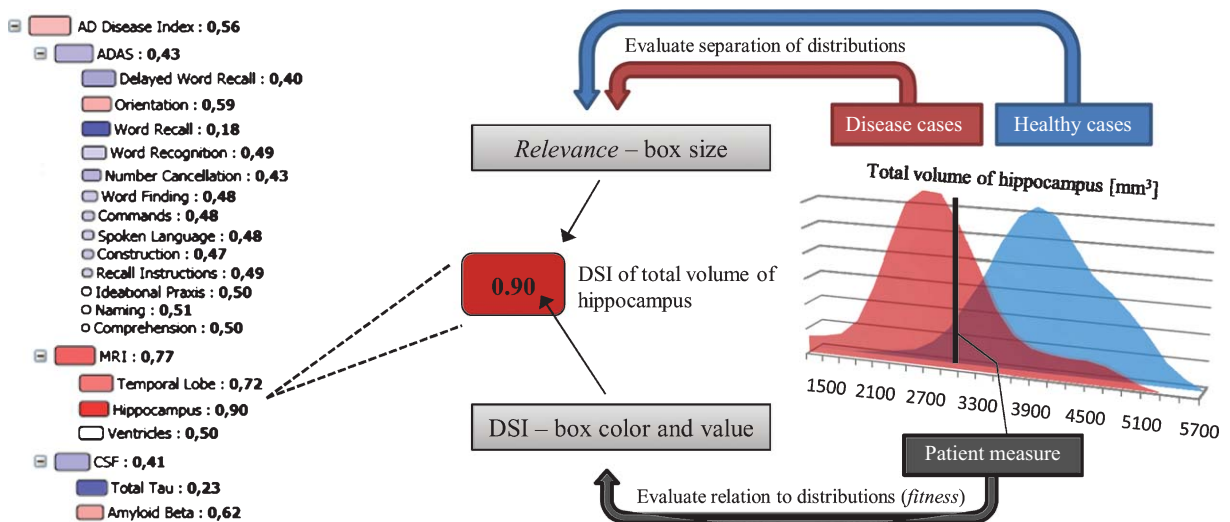


Fig. 4. DSI values of a patient with subtle indication of AD (total DSI value = 0.56). Name of the test and DSI value is shown next to each node. Larger nodes discriminate better between healthy and diseased patients (visualization of *relevance*). ‘Hot’, i.e., red, nodes highlight patient data that fits AD profile (visualization of DSI). Here, ADAS and MRI contribute the most to the AD Disease State Index, indicated by largest node size. MRI variables, especially volume of hippocampus, whose computation is schematically depicted on the right hand side, push the total DSI value towards AD population.

used for hiding or revealing extra details and for inclusion or exclusion of variables. User initiated changes to DSI model selection can give more control over the study of the patient’s disease state, making possible personalized comparison of patient data to previous cases that are of the same gender, age group, ethnicity, and educational degree.

### Evaluation

Objectives of the evaluations were to

1. compare the performance of the DSI to state-of-the-art classifiers,

2. evaluate the relationship between the index values and the actual diagnoses,
3. investigate the DSI’s capability to predict conversion from MCI to AD, and
4. visually inspect patient DSFs to evaluate their clinical practicality.

In all of the analyses, index values from DSI were compared to the probability of having AD obtained with three reference classifiers: logistic regression (LR) [31], probability estimates from support vector machines (SVM) [32] and Naïve Bayes classifier [33]. These classifiers were chosen as being representative of commonly used classification methods in

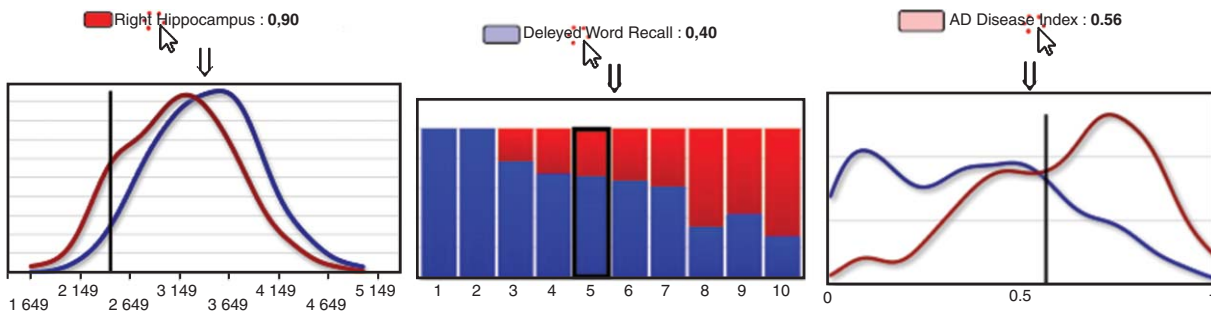


Fig. 5. If a node in the DSF tree is clicked, a comparison of patient data in relation to control and disease population distributions is displayed. Values of the AD population are rendered in red. The healthy population is rendered in blue. Black bars denote the values of the patient being studied. On the left, right hippocampal volume of the patient (2455 mm<sup>3</sup>) is overlaid on the distributions. In the center, an ordinal variable (Delayed Word Recall from ADAS) is displayed as bar chart. On the right, total AD Disease State Index of the patient (0.56) is overlaid on the control and AD population DSI values.

practical applications. Like the reference methods, the DSI is a predictor of having AD, and they are all in congruence, with increasing disease probabilities generally resulting in increasing DSI values. Comparisons between values resulting from the DSI and the reference methods are appropriate if one considers the values being used by human readers for decision support. All methods were evaluated using the same training and test data. For LR and SVM, variables not significant between the control and disease populations (Student’s *t*-test result of  $p > 0.05$ ) were excluded. For all reference methods, missing values were handled appropriately. For DSI, this type of pre-processing of the data was not required due to its design.

Comparison to actual diagnoses was performed by training the methods with HC and AD subjects and testing with all patients. The methods’ ability to assign values that have a relation with interval-level diagnoses ( $HC = \frac{0}{3}$ ,  $SMCI = \frac{1}{3}$ ,  $PMCI = \frac{2}{3}$ , and  $AD = \frac{3}{3}$ ) was evaluated using Kruskal-Wallis non-parametric test, Pearson’s linear correlation test, and visual inspection. Capability to predict conversion of MCI patients to AD was evaluated by determining area under curve (AUC) measures from receiver-operator curves (ROC) using SMCI-PMCI datasets. MCI patients who obtained index/probability values within the upper or lower ranges of the scale were pooled together to determine classification accuracy for these subsets separately. The patients included in each subset were selected from both ends of the index/probability value range [0, 1], extending to a distance of 0.02, 0.05, 0.1, 0.2, 0.3, and finally 0.4 from either end.

In all analyses, ten iterations of stratified (with same proportions of class labels) 10-fold cross validation were performed to produce robust estimates of performance metrics associated with the methods. Using

such a large number of iterations is especially important for data where the differences between classes are subtle and results can vary considerably over consecutive iterations. All analyses were implemented and executed within Matlab version R2010a, using libsvm [32] implementation of SVM and MathWorks® Statistics toolbox implementations of LR and Bayes classifier.

## RESULTS

### Correlation between disease state index and diagnosis

DSI, LR, SVM, and Bayes classifier were evaluated using baseline data from the ADNI database to determine how they relate to the diagnostic classes of 199 healthy controls, 190 SMCI, 154 PMCI, and 163 ADs. Figure 6 shows the box-plots and distributions of values assigned to the patients using the best performing dataset, best individual test dataset, and the worst dataset (ALL, ADAS, and TMT respectively).

The graphs clearly illustrate that DSI is different in nature from the reference methods, distributing index values evenly over the whole scale. The significance of DSI’s evident linearity can be appreciated by comparing results from two example patients whose total scores from ADAS differ only slightly (17 vs. 19). With data from ADAS alone, DSI gave to these patients indices of 0.36 and 0.57 (a moderate difference of 0.21), respectively. For the same patients, the probability of having AD estimated by LR were 0.39 and 0.73 (difference of 0.34), by SVM 0.23 and 0.95 (difference of 0.72), and by Bayes 0.0 and 0.75 (difference of 0.75). Especially with SVM and Bayes, the inflated proba-

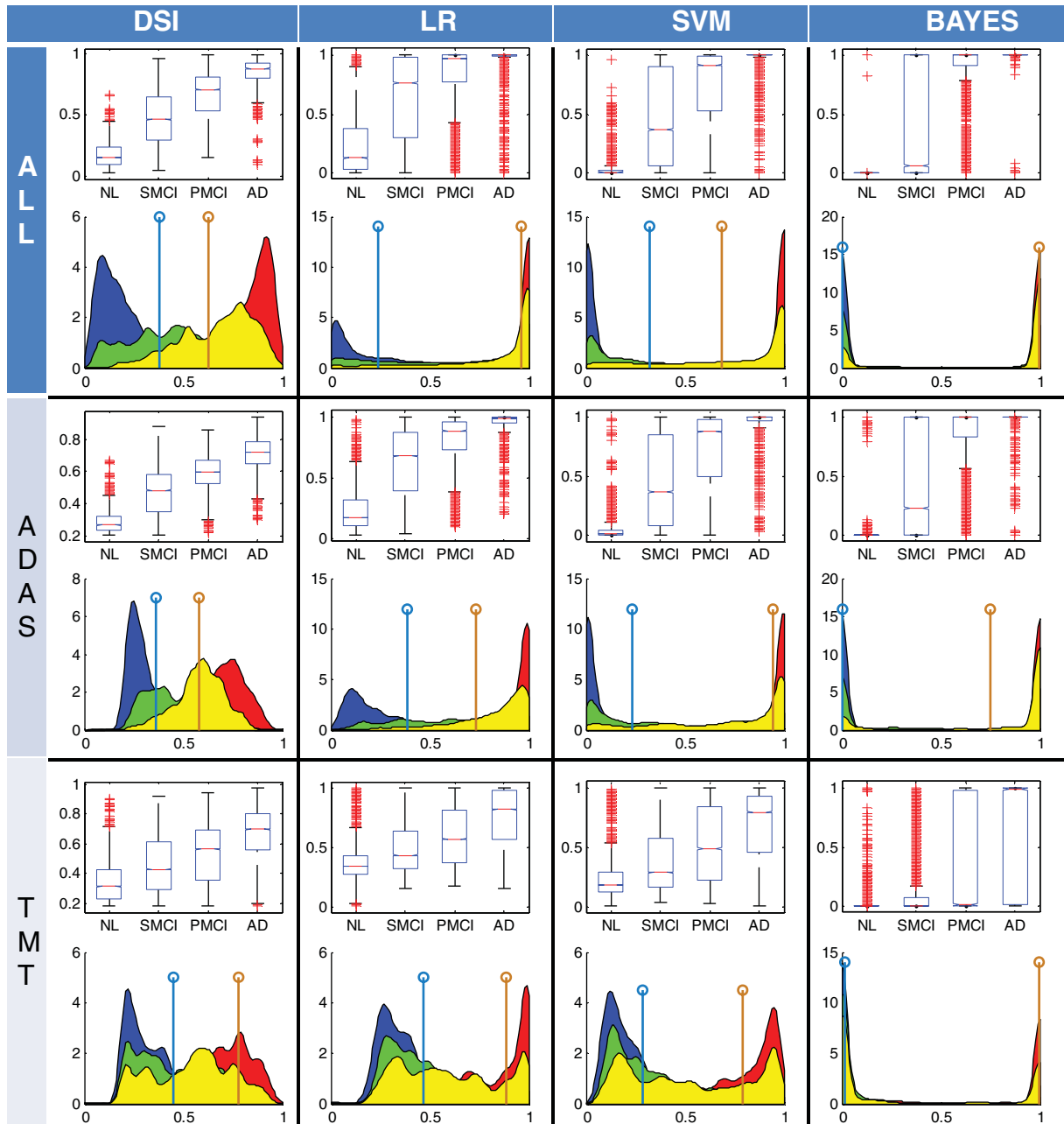


Fig. 6. Index and probability values obtained from evaluating datasets ALL, ADAS, and TMT with DSI, LR, SVM, and Bayes. Results are displayed as box plots and probability density estimates of patient classes NL (blue), SMCI (green), PMCI (yellow), and AD (red) according to index/probability values assigned to them by the methods. In box plots, the line in the middle is the median, the upper and lower ends of the box are the 75% and 25% percentiles, and the whiskers give an indication of the range. Values of two arbitrarily chosen SMCI (light blue) and PMCI (brown) patients with relatively similar clinical test results and biomarker discoveries are visualized on top of each distribution graph. Locations of the stems demonstrate the differences between the methods when assessing individual patients.

bilities obscure what in reality is a small difference in cognitive performance between the patients.

All methods distinguished between the diagnostic categories with high significance ( $p < 0.001$  in Kruskal-Wallis test) using all datasets. Linear corre-

lation with interval-level diagnoses also attained high significance ( $p < 0.001$  in Pearson) using all datasets. Table 4 shows the eight best and eight poorest performing method/dataset combinations from both statistical tests.



Table 4

Results from the Kruskal-Wallis and Pearson tests using DSI, LR, SVM, and Bayes for discriminating between the diagnostic classes of NL, SMCI, PMCI, and AD and for linear correlation with the interval-level diagnoses, respectively

Kruskal-Wallis					Pearson				
Rank	Method	Dataset	$\chi^2$	$p$	Rank	Method	Dataset	$r$	$p$
1	DSI	ALL	117.6 (5.8)	8.67E-24	1	DSI	ALL	0.56 (0.01)	8.53E-30
2	Bayes	MMSE + ADAS + MRI + CSF	116.4 (6.0)	1.39E-23	2	DSI	MMSE + ADAS + MRI + CSF	0.54 (0.01)	2.70E-26
3	Bayes	MMSE + ADAS + MRI	113.1 (6.4)	1.05E-22	3	DSI	ADAS	0.53 (0.02)	2.88E-26
4	SVM	ALL	113.3 (6.8)	1.45E-22	4	DSI	MMSE + ADAS + MRI	0.52 (0.01)	6.10E-25
5	Bayes	ALL	118.2 (7.1)	1.51E-22	5	SVM	ALL	0.50 (0.02)	2.23E-22
6	Bayes	MMSE + ADAS	114.5 (6.6)	2.50E-22	6	LR	ADAS	0.51 (0.02)	2.52E-22
7	DSI	MMSE + ADAS + MRI + CSF	108.4 (5.6)	1.02E-21	7	SVM	MMSE + ADAS + MRI	0.49 (0.01)	7.75E-22
8	SVM	MMSE + ADAS + MRI	107.7 (5.7)	1.31E-21	8	DSI	MMSE + ADAS	0.49 (0.01)	9.92E-22
:	:	:	:	:	:	:	:	:	:
33	LR	TMT	40.3 (6.5)	1.25E-06	33	DSI	APOE	0.28 (0.03)	1.72E-06
34	SVM	TMT	39.7 (6.4)	2.49E-06	34	LR	APOE	0.28 (0.03)	1.78E-06
35	Bayes	TMT	35.7 (6.6)	4.49E-06	35	DSI	CSF	0.38 (0.03)	2.28E-06
36	DSI	CSF	29.5 (3.0)	5.01E-06	36	Bayes	APOE	0.28 (0.03)	2.39E-06
37	DSI	TMT	37.3 (6.5)	5.59E-06	37	Bayes	CSF	0.37 (0.03)	2.43E-06
38	SVM	CSF	26.9 (3.5)	2.42E-05	38	Bayes	TMT	0.28 (0.03)	4.16E-06
39	Bayes	CSF	26.4 (3.2)	2.45E-05	39	SVM	CSF	0.35 (0.03)	4.63E-06
40	SVM	APOE	27.0 (5.7)	6.23E-05	40	SVM	APOE	0.27 (0.03)	8.71E-06

The table shows method/dataset performance ordered by the mean of  $p$ -values over  $10 \times 10$ -fold cross-validation iterations. The Kruskal-Wallis test statistic  $\chi^2$  and Pearson test statistic  $r$  shown are the mean and standard deviation over  $10 \times 10$ -fold cross-validation iterations.

### Prediction of MCI to AD conversion

Capability to predict conversion from MCI to AD was evaluated with 190 SMCI and 154 PMCI cases from the ADNI database. Figure 7 shows results from two of the best individual tests and from the four increasingly complete combinations of tests. In general, AUC improves and standard deviation decreases through having better or more patient data available.

Relevance parameters obtained from DSI indicate that within the ADNI database, ADAS is the most relevant single test for predicting conversion from MCI to AD, followed by MRI, APOE, CSF, MMSE, and finally TMT (see Table 5). Within ADAS, relevance values are very similar to weights of a recently introduced ADAS composite [34]. Between all individual variables from all tests, DSI considers the most relevant to be Delayed Word Recall from ADAS (relevance of 0.294), Left Middle Temporal Lobe from MRI (0.262), and Total Tau from CSF (0.258).

### Levels of confidence for predicting conversion from MCI to AD

Based on data alone, there are no machine learning methods that can predict conversion from MCI to AD for all cases reliably. Therefore, clinicians always need to consider all available evidence. Nevertheless, index/probability values obtained with the complete dataset (ALL) were examined to determine if the methods studied here could provide more confidence for

diagnosing certain subsets of patients. From Table 6 it can be seen that extreme value ranges provide considerably better prediction accuracies and there is a small subset of patients where the classification methods attain perfect prediction accuracy.

### Visual inspection of disease state fingerprints

The DSFs of several SMCI and PMCI patients were inspected to confirm that they quickly reveal the state of the patient data in relation to AD population and highlight the tests and variables contributing to the results. Fig. 8 shows example DSFs for clear SMCI, subtle SMCI, subtle PMCI, and clear PMCI cases. With the clear cases, nearly all variables point towards AD (shades of red) or against it (shades of blue). With the subtle cases, there is a mix of colors that show which patient data indicate AD and which do not.

## DISCUSSION

The DSF provides a quickly interpretable visual overview of patient state, obtained from evidence-based statistical analysis of patient data. It draws the clinician to data that are the most relevant, omitting the need to go over tens or hundreds of data points individually. DSF clearly discloses the factors contributing to the results, highlights the important measures, and thus supports application of clinical judgment. In its design, equal emphasis was given to prediction accuracy and

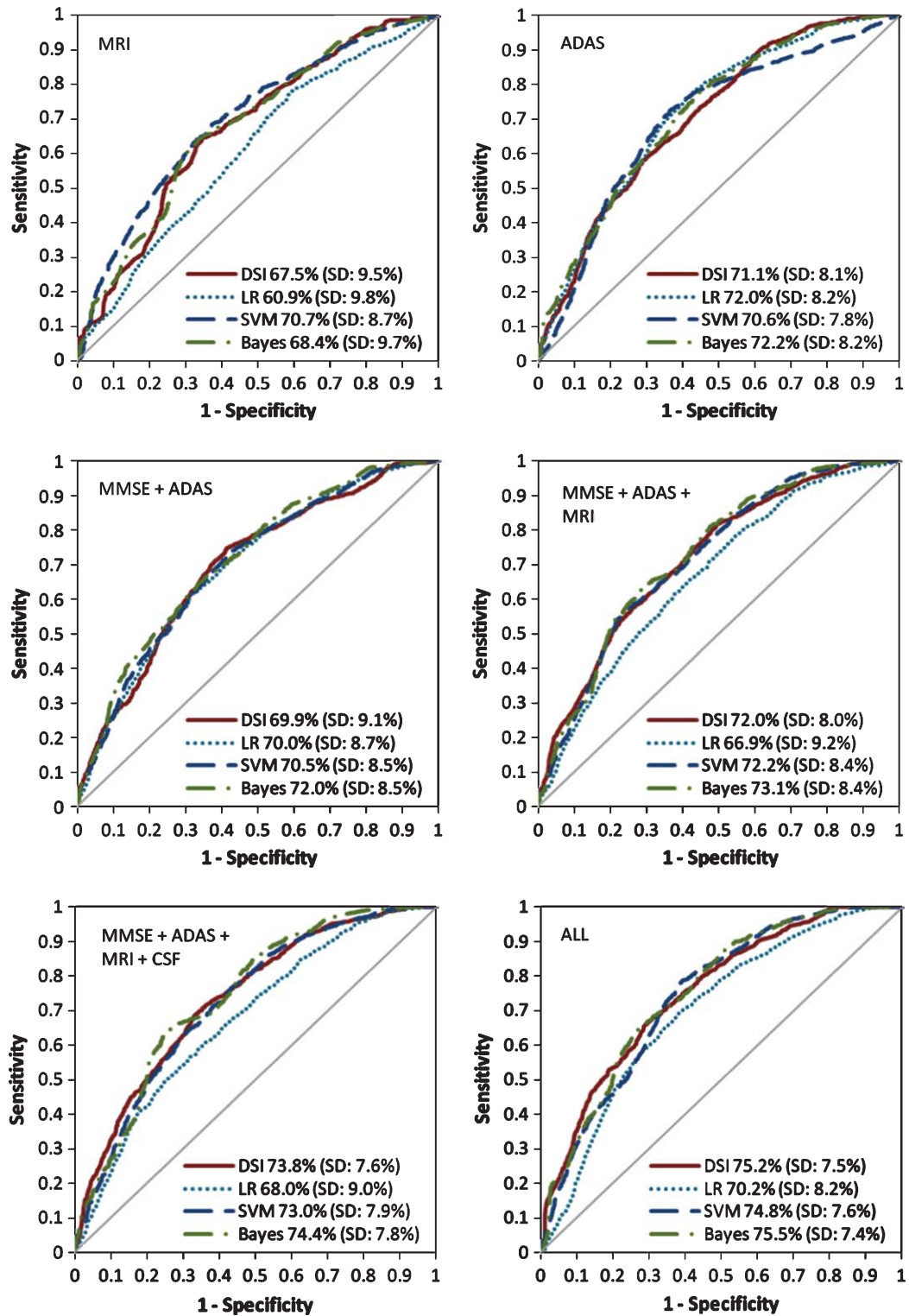


Fig. 7. ROC curves for individual tests and combinations of tests for predicting conversion from MCI to AD. Numbers denote AUC and standard deviations of AUC with the respective datasets over the  $10 \times 10$ -fold cross-validation iterations.

Table 5

Relevance values for all data, individual tests, and best individual features (where *relevance* >0.200) for predicting conversion from MCI to AD based on ADNI data

Test	Relevance (SD)	
Disease State Index	0.420	(0.019)
ADAS	0.333	(0.022)
Delayed Word Recall	0.294	(0.018)
Orientation	0.256	(0.017)
Word Recall	0.256	(0.017)
Word Recognition	0.203	(0.017)
MRI	0.300	(0.021)
Left Middle Temporal Lobe	0.262	(0.018)
Right Middle Temporal Lobe	0.246	(0.020)
Left Inferior Temporal Lobe	0.221	(0.021)
Left Hippocampus	0.207	(0.022)
Right Hippocampus	0.201	(0.020)
Right Entorhinal Cortex	0.201	(0.020)
APOE	0.256	(0.016)
Allele B of genotype A/B	0.256	(0.016)
CSF	0.249	(0.029)
Total Tau	0.258	(0.025)
Amyloid-β	0.221	(0.025)
MMSE	0.249	(0.024)
no individual features with relevance >0.200		
TMT	0.207	(0.022)
Time to Complete Trail B	0.202	(0.022)

The table shows mean *relevance* values and their standard deviation over 10 × 10-fold cross-validation.

to clinical practicality. To the authors’ knowledge there exists no other evidence-based data visualization methods developed with a similar philosophy. Several established machine learning methods were considered for the foundation of DSF, but none were found satisfactory. For example, regression analysis cannot be capitalized fully when working with existing discrete AD diagnoses that do not include much information about the stage or severity of the disease. SVM, with

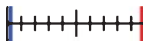

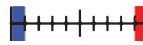
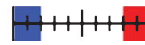


its high dimensional decision boundary, is too abstract for human interpretation. Naïve Bayes works well as a classifier, but results in very unrealistic and unpractical disease probabilities. Thus, DSI was developed to provide a good foundation for visual expert analysis of progressing disease state.

The DSI model of progressing disease state was able to discriminate well between the diagnostic classes of healthy, SMCI, PMCI, and AD and attained good levels of linear correlation, superior to the reference classifiers. Improved linearity is clearly evident with visual inspection of the value distribution graphs in Fig. 6, in which reference methods lean heavily on the head and tail values of the scale even when source data differs only slightly. Thus, DSI is truly indicative of patient state between healthy and AD and appears to correspond well with clinical practice. Even though maximizing classification accuracy was not the only goal, DSI’s capability to predict conversion from MCI to AD was similar to the reference classifiers. Analysis of the relevance values reinforced the view that combinations of tests are required for reliable early diagnoses. Interestingly, the relatively simple and computationally low-cost method for computing *relevance* produced almost the same weighting factors as a novel method employed for prediction of 12 months conversion from MCI to AD [34].

Currently, clinicians are forced to browse test results one by one, possibly losing track of the big picture. Analysis of extreme DSI values indicates that there are MCI cases where data leaves little doubt as to whether a patient has AD or not. Particularly those clinicians with less experience might be more confident to diagnose AD at an early stage if they were

Table 6

Classification accuracies of DSI, LR, SVM, and Bayes when observing subgroups of SMCI and PMCI patients based on index/probability values assigned to them

Distance from end of scale [0, 1]	< 0.02	< 0.05	< 0.1	< 0.2	< 0.3	< 0.4
Allowed value ranges						
DSI	None assigned	None assigned	100% (0.7%)	93.6% (9.1%)	84.0% (30.2%)	75.6% (63.7%)
LR	52.5% (1.2%)	73.6% (5.1%)	71.9% (13.2%)	72.9% (31.7%)	71.4% (52.3%)	67.5% (75.1%)
SVM	100% (1.0%)	95.1% (4.2%)	90.7% (13.1%)	84.3% (32.3%)	77.3% (53.0%)	72.0% (76.0%)
BAYES	77.6% (57.0%)	74.2% (66.1%)	72.1% (74.2%)	71.1% (82.7%)	70.7% (88.3%)	70.0% (94.2%)

In parentheses is the percentage of patients assigned to the subgroup over 10 × 10-fold cross-validation iterations. For example, DSI assigned an index value <0.2 or >0.8 to 9.2% of the patients, which was a correct prediction for 93.7% of cases, i.e., classification accuracy for the subgroup was 93.7%.

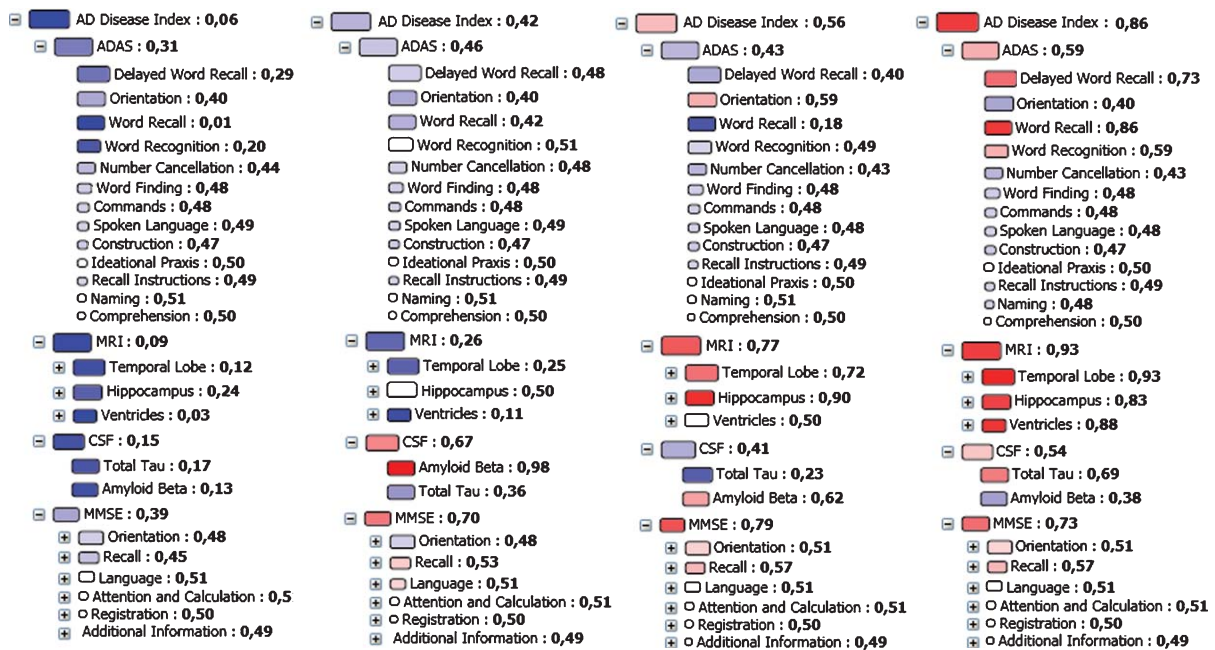


Fig. 8. Four patients visualized using the DSF. Starting from the left, the figure shows two stable MCI (SMCI) patients and two progressive MCI (PMCI) patients. Box sizes (denoting *relevance*) indicate capability of a variable or test to discriminate between SMCI and PMCI cases. The nodes are reordered top to bottom according to this measure. Colors indicate which group the patient data fits better; blue color equals SMCI, red color equals PMCI. A unique disease state fingerprint emerges from the node sizes and color codes for each patient, allowing quick evaluation of patient state and reviewing of individual tests and variables contributing to the results.

able to see all data at once, and also see how patient data relate with previously diagnosed disease population at their clinic. While the DSI and DSF increase the amount of information available to a clinician, they also allow clinicians to concentrate on what is important and ignore irrelevant information, making the most of existing data.

When compared to many other machine learning methods, the benefits of DSI and DSF are numerous. Due to linearity, small changes in patient data cause only small changes in DSI, making interpretation of DSF easier and longitudinal follow-ups consistent. The methods are data agnostic, able to work with any tests or variables in use at a particular clinic. They work with raw test and measurement values, increasing familiarity and requiring no pre-processing of data, feature selection, or data cleanup. All data acquisition modalities are quantified both in isolation and as a part of the whole, providing additional context to the results. It is very easy to support different types of tests and variables (scalar, nominal, ordinal, even textual with text mining methods) with suitable fitness functions. If desired, the probability of a patient having AD can be computed using the DSI values obtained during its evaluation. Unlike as is the case with many other machine learning methods, sparse data creates no prob-

lems. Each variable is initially treated individually, and only used if the data exists. Additionally, as long as the training set patients are representative of the control and disease populations, there is no need to have very large quantities of data.

Further potential is anticipated from interactive implementation of DSF, which can provide a quick path to personalized healthcare. Limiting comparison of patient data to cases that are of same gender, age, ethnicity, or educational degree provides personalized results for that patient. A clinical application could also employ *relevance* measures to suggest additional tests to be done, based on their ability to discriminate between healthy and diseased cases. Interactive visualizations of disease population distributions with patient values overlaid on them are an expressive way of comparing patient data to previously diagnosed cases.

Further studies are being planned to cover aspects of DSI and DSF not reported here. Non-linear dependencies between the variables, e.g., differences in cognitive tests due to varying levels of education, were not considered in this work. Stratification of training data will be studied to see if regression of variables based on demographics, such as age and education, further improves the results. Bootstrapping would allow better relevance estimates and, more importantly,

would also provide statistical measures that could improve analysis of patient data [35]. Robustness of the DSI will be evaluated with data from several longitudinal studies of AD and other neurodegenerative diseases. Performance of the proposed system will also be examined when there are heavy correlations and other adverse conditions within the data. Utility of an interactive DSF tool is being evaluated with clinicians using existing data from ADNI. There are also plans to take part in upcoming longitudinal studies where an implementation of DSF is provided to the clinicians.

Diagnostic guidelines for AD emphasize the congruence of neuropsychological test results and biomarkers. DSF was designed to enable quick visual analysis of all patient data as a whole. It is a versatile decision support system that uses locally available patient data, presents a synthesis of the information in an understandable manner, and allows an expert to interpret and report the results within the diagnostic process. The proposition is that the DSF can be a clinically relevant tool which enables clinicians to make better and more consistent decisions in daily practice.

## ACKNOWLEDGMENTS

The authors thank all participants for their contribution to this project.

This work was supported by the 7th Framework Programme by the European Commission (<http://cordis.europa.eu/ist>); EU-Grant-224328-PredictAD; Name: From Patient Data to Personalised Healthcare in Alzheimer's Disease). Additional grants have been awarded to the medical doctors involved in this study.

Data collection and sharing for this project was funded by the Alzheimer's Disease Neuroimaging Initiative (ADNI) (National Institutes of Health Grant U01 AG024904). ADNI is funded by the National Institute on Aging, the National Institute of Biomedical Imaging and Bioengineering, and through generous contributions from the following: Abbott, AstraZeneca AB, Bayer Schering Pharma AG, Bristol-Myers Squibb, Eisai Global Clinical Development, Elan Corporation, Genentech, GE Healthcare, GlaxoSmithKline, Innogenetics, Johnson and Johnson, Eli Lilly and Co., Medpace, Inc., Merck and Co., Inc., Novartis AG, Pfizer Inc, F. Hoffman-La Roche, Schering-Plough, Synarc, Inc., as well as non-profit partners the Alzheimer's Association and Alzheimer's Drug Discovery Foundation, with participation from the U.S. Food and Drug Administration. Private sec-

tor contributions to ADNI are facilitated by the Foundation for the National Institutes of Health ([www.fnih.org](http://www.fnih.org)). The grantee organization is the Northern California Institute for Research and Education, and the study is coordinated by the Alzheimer's Disease Cooperative Study at the University of California, San Diego. ADNI data are disseminated by the Laboratory for Neuro Imaging at the University of California, Los Angeles. This research was also supported by NIH grants P30 AG010129, K01 AG030514, and the Dana Foundation.

The funding partners had no active role in this research.

Authors' disclosures available online (<http://www.jalz.com/disclosures/view.php?id=899>).

## REFERENCES

- [1] Waldemar G, Phungh KTT, Burns A, Georges J, Hansen FR, Iliffe S, Marking C, Olde-Rikkert M, Selmes J, Stoppe G, Sartorius N (2007) Access to diagnostic evaluation and treatment for dementia in Europe. *Int J Geriatr Psych* **22**, 47-54.
- [2] Bond J, Stave C, Sganga A, Vincenzino O, O'connell B, Stanley R (2005) Inequalities in dementia care across Europe: key findings of the Facing Dementia Survey. *Int J Clin Pract* **59**, 8-14.
- [3] Waldemar G, Dubois B, Emre M, Georges J, McKeith IG, Rossor M, Scheltens P, Tariska P, Winblad B (2007) Recommendations for the diagnosis and management of Alzheimer's disease and other disorders associated with dementia: EFNS guideline. *Eur J Neurol* **14**, e1-e26.
- [4] Dubois B, Feldman HH, Jacova C, DeKosky ST, Barberger-Gateau P, Cummings J, Delacourte A, Galasko D, Gauthier S, Jicha G, Meguro K, O'Brien J, Pasquier F, Robert P, Rossor M, Salloway S, Stern Y, Visser PJ, Scheltens P (2007) Research criteria for the diagnosis of Alzheimer's disease: Revising the NINCDS-ADRDA criteria. *Lancet Neurol* **6**, 734-746.
- [5] Petersen RC (2004) Mild cognitive impairment as a diagnostic entity. *J Int Med* **256**, 183-194.
- [6] Petersen RC, Roberts RO, Knopman DS, Boeve BF, Geda YE, Ivnik RJ, Smith GE, Jack CR Jr (2009) Mild cognitive impairment: Ten years later. *Arch Neurol* **66**, 1447-1455.
- [7] Visser PJ, Verhey F, Knol DL, Scheltens P, Wahlund LO, Freund-Levi Y, Tsolaki M, Minthon L, Wallin AK, Hampel H, Bürger K, Pirttila T, Soininen H, Rikkert MO, Verbeek MM, Spira L, Blennow K (2009) Prevalence and prognostic value of CSF markers of Alzheimer's disease pathology in patients with subjective cognitive impairment or mild cognitive impairment in the DESCRIPA study: A prospective cohort study. *Lancet Neurol* **8**, 619-627.
- [8] Morris JC, Roe CM, Grant EA, Head D, Storandt M, Goate AM, Fagan AM, Holtzman DM, Mintun MA (2009) Pittsburgh compound B imaging and prediction of progression from cognitive normality to symptomatic Alzheimer's disease. *Arch Neurol* **66**, 1469-1475.
- [9] Petersen RC, Smith GE, Ivnik RJ, Tangalos EG, Schaid DJ, Thibodeau SN, Kokmen E, Waring SC, Kurland LT (1995) Apolipoprotein E status as a predictor of the development of Alzheimer's disease in memory-impaired individuals. *JAMA* **273**, 1274-1278.

- [10] Devanand DP, Pradhaban G, Liu X, Khandji A, De Santi S, Segal S, Rusinek H, Pelton GH, Honig LS, Mayeux R, Stern Y, Tabert MH, de Leon MJ (2007) Hippocampal and entorhinal atrophy in mild cognitive impairment: prediction of Alzheimer's disease. *Neurology* **68**, 828-836.
- [11] Querbes O, Aubry F, Pariente J, Lotterie JA, eacute D, monet JF, Duret V, Puel M, Berry I, Fort JC, Celsis P (2009) Alzheimer's Disease Neuroimaging Initiative early diagnosis of Alzheimer's disease using cortical thickness: Impact of cognitive reserve. *Brain* **132**, 2036-2047.
- [12] Devanand DP, Liu X, Tabert MH, Pradhaban G, Cuasay K, Bell K, de Leon MJ, Doty RL, Stern Y, Pelton GH (2008) Combining early markers strongly predicts conversion from mild cognitive impairment to Alzheimer's disease. *Biol Psychiatry* **64**, 871-879.
- [13] Vemuri P, Wiste HJ, Weigand SD, Shaw LM, Trojanowski JQ, Weiner MW, Knopman DS, Petersen RC, Jack CR, Jr. (2009) Alzheimer's Disease Neuroimaging, Initiative MRI and CSF biomarkers in normal, MCI, and AD subjects: Diagnostic discrimination and cognitive correlations. *Neurology* **73**, 287-293.
- [14] Risacher SL, Saykin AJ, West JD, Shen L, Firpi HA, McDonald BC (2009) Alzheimer's Disease Neuroimaging, Initiative Baseline MRI predictors of conversion from MCI to probable AD in the ADNI Cohort. *Curr Alzheimer Res* **6**, 347-361.
- [15] Dubois B, Feldman HH, Jacova C, Cummings JL, Dekosky ST, Barberger-Gateau P, Delacourte A, Frisoni G, Fox NC, Galasko D, Gauthier S, Hampel H, Jicha GA, Meguro K, O'Brien J, Pasquier F, Robert P, Rossor M, Salloway S, Sarazin M, de Souza LC, Stern Y, Visser PJ, Scheltens P (2010) Revising the definition of Alzheimer's disease: a new lexicon. *Lancet Neurol* **9**, 1118-1127.
- [16] Kloppel S, Stonnington CM, Barnes J, Chen F, Chu C, Good CD, Mader I, Mitchell LA, Patel AC, Roberts CC (2008) Accuracy of dementia diagnosis – a direct comparison between radiologists and a computerized method. *Brain* **131**, 2969-2974.
- [17] Huang QR, Qin Z, Zhang S, Chow CM (2008) Clinical patterns of obstructive sleep apnea and its comorbid conditions: A data mining approach. *J Clin Sleep Med* **4**, 543-550.
- [18] Mullins IM, Siadaty MS, Lyman J, Scully K, Garrett CT, Miller WG, Muller R, Robson B, Apte C, Weiss S, Rigoutsos I, Platt D, Cohen S, Knaus WA (2006) Data mining and clinical data repositories: Insights from a 667,000 patient data set. *Comput Biol Med* **36**, 1351-1377.
- [19] Bates DW, Cohen M, Leape LL, Overhage JM, Shabot MM, Sheridan T (2001) Reducing the frequency of errors in medicine using information technology. *J Am Med Inform Assoc* **8**, 299-308.
- [20] Sim I, Gorman P, Greenes RA, Haynes RB, Kaplan B, Lehmann H, Tang PC (2001) Clinical decision support systems for the practice of evidence-based medicine. *J Am Med Inform Assn* **8**, 527-534.
- [21] Kawamoto K, Houlihan CA, Balas EA, Lobach DF (2005) Improving clinical practice using clinical decision support systems: A systematic review of trials to identify features critical to success. *Evid Based Cardiovasc Med* **9**, 208-210.
- [22] Smith NB, Webb A (2010) Introduction to Medical Imaging: Physics, Engineering and Clinical Applications, Cambridge University Press, Cambridge.
- [23] Zimetbaum PJ, Josephson ME (2008) *Practical clinical electrophysiology*, Lippincott Williams & Wilkins, Philadelphia.
- [24] Aigner W, Kaiser K, Miksch S (2008) Visualization Methods to Support Guideline-Based Care Management. *Stud Health Technol Inform* **139**, 140-159.
- [25] Bade R, Schlechtweg S, Miksch S (2004) Connecting time-oriented data and information to a coherent interactive visualization. *Proc CHI* **2004**, 105-112.
- [26] Wang TD, Plaisant C, Shneiderman B (2010) Visual Information Seeking in Multiple Electronic Health Records: Design Recommendations and a Process Model. *Proc IHI* **2010**, 46-55.
- [27] Koikkalainen JR, Antila M, Lötjönen JM, Heliö T, Lauerma K, Kivistö SM, Sipola P, Kaartinen MA, Kärkkäinen ST, Reissell E, Kuusisto J, Laakso M, Oresic M, Nieminen MS, Peuhkurinen KJ (2008) Early familial dilated cardiomyopathy: Identification with determination of disease state parameter from cine MR image data. *Radiology* **249**, 88-96.
- [28] Fischl B, Salat DH, Busa E, Albert M, Dieterich M, Haselgrove C, van der Kouwe A, Killiany R, Kennedy D, Klaveness S, Montillo A, Makris N, Rosen B, Dale AM (2002) Whole brain segmentation: Automated labeling of neuroanatomical structures in the human brain. *Neuron* **33**, 341-355.
- [29] Han J, Kamber M (2006) *Data Mining: Concepts and Techniques*, 2nd ed, Morgan Kaufmann Publishers, Burlington.
- [30] Jolliffe IT (2002) *Principal Component Analysis*, Series: Springer Series in Statistics, 2nd ed, Springer, New York.
- [31] Hosmer DW, Lemeshow S (2000) *Applied logistic regression*. Wiley-Interscience, New Jersey.
- [32] Chang C, Lin C LIBSVM (2001) A Library for Support Vector Machines, <http://www.csie.ntu.edu.tw/~cjlin/>, Accessed February 17 2011 libsvm.
- [33] Mitchell TM (1997) *Machine Learning*. McGraw-Hill Series in Computer Science, McGraw-Hill, New York.
- [34] Llano DA, Laforet G, Devanarayan V (2011) Derivation of a New ADAS-cog Composite Using Tree-based Multivariate Analysis: Prediction of Conversion From Mild Cognitive Impairment to Alzheimer's Disease. *Alz Dis Assoc Dis* **25**, 73-84.
- [35] Mooney CZ, Duval RD, Duval R (1993) *Bootstrapping: A nonparametric approach to statistical inference*. Sage Publications, Inc, London.