

Diagnostic Assessment & Prognosis

# Random forest to differentiate dementia with Lewy bodies from Alzheimer's disease

Meenakshi Dauwan<sup>a,b,\*</sup>, Jessica J. van der Zande<sup>c,1</sup>, Edwin van Dellen<sup>a,b,c</sup>, Iris E. C. Sommer<sup>b</sup>, Philip Scheltens<sup>c</sup>, Afina W. Lemstra<sup>c</sup>, Cornelis J. Stam<sup>a</sup>

<sup>a</sup>Department of Clinical Neurophysiology and MEG Center, Neuroscience Campus Amsterdam, VU University Medical Center, Amsterdam, The Netherlands

<sup>b</sup>Department of Psychiatry, Brain Center Rudolf Magnus, University Medical Center Utrecht, Utrecht, The Netherlands

<sup>c</sup>Alzheimer Center and Department of Neurology, Neuroscience Campus Amsterdam, VU University Medical Center, Amsterdam, The Netherlands

## Abstract

**Introduction:** The aim of this study was to build a random forest classifier to improve the diagnostic accuracy in differentiating dementia with Lewy bodies (DLB) from Alzheimer's disease (AD) and to quantify the relevance of multimodal diagnostic measures, with a focus on electroencephalography (EEG).

**Methods:** A total of 66 DLB, 66 AD patients, and 66 controls were selected from the Amsterdam Dementia Cohort. Quantitative EEG (qEEG) measures were combined with clinical, neuropsychological, visual EEG, neuroimaging, and cerebrospinal fluid data. Variable importance scores were calculated per diagnostic variable.

**Results:** For discrimination between DLB and AD, the diagnostic accuracy of the classifier was 87%. Beta power was identified as the single-most important discriminating variable. qEEG increased the accuracy of the other multimodal diagnostic data with almost 10%.

**Discussion:** Quantitative EEG has a higher discriminating value than the combination of the other multimodal variables in the differentiation between DLB and AD.

© 2016 The Authors. Published by Elsevier Inc. on behalf of the Alzheimer's Association. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

## Keywords:

Alzheimer's disease; Dementia with Lewy bodies; EEG; Random forest; Diagnostic accuracy; Beta power; Machine learning

## 1. Introduction

Alzheimer's disease (AD) and dementia with Lewy bodies (DLB) are the two most common forms of dementia in the aging population [1,2]. DLB and AD have several overlapping characteristics, making differential diagnosis in clinical practice at times difficult [3]. Compared to AD, consensus criteria [1] in DLB have moderate sensitivity [4,5]. Accurate diagnosis of DLB and AD is essential for patient guidance and appliance of possible early treatment and prevention strategies [6]. Therefore, disease-specific

biomarkers from cerebrospinal fluid (CSF) and neuroimaging are increasingly used, but these diagnostic tests can be costly and are not always available [5,7]. Furthermore, the frequent presence of concomitant AD pathology in DLB patients renders amyloid markers and magnetic resonance imaging (MRI) less discriminative [5,8]. In contrast, electroencephalography (EEG) has been proposed as a low-cost and readily available diagnostic tool to distinguish between DLB and AD [9,10]. At present, in a clinical setting, data from patient history and above-mentioned diagnostic tests are weighted differently in each individual patient to make a diagnosis [11]. The exact contribution of the (combinations of) EEG and other diagnostic tests to the differential diagnosis of DLB and AD remains unclear.

Automated classification algorithms can directly provide the most relevant diagnostic variables and estimate their

The authors declared no conflict of interest.

<sup>1</sup>Both authors contributed equally to the study.

\*Corresponding author. Tel.: +31-88-75-57468.

E-mail address: [m.dauwan-3@umcutrecht.nl](mailto:m.dauwan-3@umcutrecht.nl)

<http://dx.doi.org/10.1016/j.dadm.2016.07.003>

2352-8729/© 2016 The Authors. Published by Elsevier Inc. on behalf of the Alzheimer's Association. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

relative importance in classifying cognitive impairment, which can improve diagnostic efficiency [12,13]. Ensemble-learning methods construct automated classification algorithms that can learn from and predict data by building a model in the form of input-output relationships of variables (i.e., features in classification algorithms) [14]. Random forest is one such algorithm, developed by L. Breiman, and based on the principle of decision tree learning [15]. In the field of dementia, ensemble-learning methods have mainly been studied to classify patients with AD [13], whereas very little evidence is available on the automated discrimination between DLB and AD [12] or on the combination of different diagnostic modalities in an automated classifier.

This study aimed to build a random forest classifier to discriminate between DLB, AD, and controls and to quantify the importance of (combinations of) different types of diagnostic features (i.e., clinical, neuropsychological, EEG, CSF, and neuroimaging data), with a specific focus on the role of EEG.

## 2. Methods

### 2.1. Study population

A total of 66 probable DLB patients, 66 probable AD patients, and 66 subjects with subjective cognitive decline (SCD) were selected from the Amsterdam Dementia Cohort [11]. The groups were matched on group level for age and gender. All subjects were referred to the Alzheimer Center of the VU University Medical Center (VUmc) in Amsterdam, The Netherlands, between September 2003 and June 2010. Standardized dementia diagnostic workup included neuropsychological assessment, lumbar puncture, brain MRI, and resting-state EEG. All subjects gave written informed consent for storage and use of their clinical data for research purposes. The Medical Ethics Committee of the VUmc approved this study. A clinical diagnosis and treatment plan was made by consensus in a weekly multidisciplinary meeting [11]. Probable AD was diagnosed according to the revised NINCDS-ADRDA criteria [2], and probable DLB was diagnosed according to consensus guidelines [1]. Subjects were labeled as SCD when they experienced and presented with cognitive complaints, but diagnostic workup was not abnormal and no other neurological or psychiatric disorder known to cause cognitive problems could be diagnosed [11]. These subjects were included as controls.

The EEG data set of the present study population has been previously analyzed focusing on functional and directed connectivity and network topology in DLB and AD [16,17].

### 2.2. Feature selection

All the non-EEG features (Table 1) for the classification algorithm were manually selected from the diagnostic workup based on availability, and their correspondence with the clinical criteria of DLB and AD [1,2].

#### 2.2.1. Clinical features

Visual hallucinations were assessed with the neuropsychiatric inventory (NPI) [18]. Extrapyramidal signs were assessed by a preformatted checklist and defined as the presence of bradykinesia, rigidity, or tremor. Cognitive functions were assessed using a standardized test battery [11]. From this, the mini-mental state examination (MMSE) was used as a measure of global cognitive function [19], trail making test part A (TMT-A) as a measure of motor speed [20], the visual association test (VAT) as a measure of episodic memory [21], and the forward and backward condition of the Digit Span as a measure of attention [22].

#### 2.2.2. Biomarkers

CSF was collected by lumbar puncture [11]. Amyloid- $\beta$  1–42 ( $A\beta_{42}$ ), total tau, phosphorylated tau (p-tau), and a ratio of tau to  $A\beta_{42}$  were included as features [23]. From neuroimaging, medial temporal lobe (MTA) atrophy, global cortical atrophy, and white-matter hyperintensities on MRI were included as features [11].

#### 2.2.3. EEG recordings

As part of the diagnostic workup, all subjects underwent a 20-minute no-task, resting-state EEG recording with OSG digital equipment (Brainlab; OSG B.V. Belgium), according to the international 10–20 system [17].

EEGs of all subjects were rated according to a standard visual rating scheme [24]. The visual rating includes the severity of EEG abnormalities on a 4-point rating scale and the presence of focal, diffuse, and epileptiform abnormalities [11,24]. In addition, all EEGs were assessed for the presence of frontal intermittent rhythmic delta activity (FIRDA) [9,10].

Subsequently, four artifact-free epochs, recorded in an awake state with eyes closed, were visually selected for each subject. Data were converted to American Standard Code for Information Interchange (ASCII) format, and 4 epochs of 4096 samples per subject (i.e., approximately  $4 \times 8$  s EEG data per subject, sufficient to perform qEEG analyses [25]) were loaded into the BrainWave software for further analysis (BrainWave version 0.9.152.2.17, C. J. Stam; available for download at <http://home.kpn.nl/stam7883/brainwave.html>).

The machine-learning module of BrainWave was used to create a data file containing all the qEEG features shown in Table 1. Phase transfer entropy (PTE) was used as a measure for effective connectivity between EEG channels. PTE measures the strength and direction of phase-based functional connectivity between interacting oscillations [26]. In addition, minimum spanning tree (MST) measures (i.e., highest degree, leaf number, and tree hierarchy) were used as a representation of functional network topology. MST is a unique acyclic subnetwork that connects all nodes in a network such that only the strongest connections in the network are included without forming loops [27].

Table 1  
Overview selected features

Feature number	Feature name	Feature number	Feature name
Quantitative EEG			
1	Lowest delta power	36	Mean PTE alpha1 band
2	Mean delta power	37	Highest PTE alpha1 band
3	Highest delta power	38	Lowest PTE alpha2 band
4	Lowest theta power	39	Mean PTE alpha2 band
5	Mean theta power	40	Highest PTE alpha2 band
6	Highest theta power	41	Lowest PTE beta band
7	Lowest alpha1 power	42	Mean PTE beta band
8	Mean alpha1 power	43	Highest PTE beta band
9	Highest alpha1 power		
10	Lowest alpha2 power	Clinical data	
11	Mean alpha2 power	44	Hallucinations
12	Highest alpha2 power	45	Extrapyramidal signs
13	Lowest beta power	Neuropsychological data	
14	Mean beta power	46	MMSE score
15	Highest beta power	47	VAT total score
16	Lowest peak frequency	48	TMT-A score
17	Mean peak frequency	49	Digit span forward
18	Highest peak frequency	50	Digit span backward
19	Theta/alpha ratio		
20	MST: highest degree theta band	Neuroimaging (MRI) biomarkers	
21	MST: leaf number theta band	51	MTA score
22	MST: tree hierarchy theta band	52	GCA score
23	MST: highest degree alpha1 band	53	Fazekas score
24	MST: leaf number alpha1 band		
25	MST: tree hierarchy alpha1 band	CSF biomarkers	
26	MST: highest degree alpha2 band	54	A $\beta_{42}$
27	MST: leaf number alpha2 band	55	Tau
28	MST: tree hierarchy alpha2 band	56	p-Tau
29	MST: highest degree beta band	57	Tau/A $\beta_{42}$ ratio
30	MST: leaf number beta band		
31	MST: tree hierarchy beta band	Visual EEG	
32	Lowest PTE theta band	58	Severity of EEG abnormalities
33	Mean PTE theta band	59	Diffuse abnormalities
34	Highest PTE theta band	60	Focal abnormalities
35	Lowest PTE alpha1 band	61	FIRDA

Abbreviations: FIRDA, frontal intermittent rhythmic delta activity; MST, minimum spanning tree; PTE, phase transfer entropy; MMSE, mini mental state examination; VAT, visual association test; TMT-A, trail making test part A; MTA, medial temporal lobe atrophy; GCA, global cortical atrophy; A $\beta_{42}$ , amyloid- $\beta$  1–42; tau, total tau; p-tau, tau phosphorylated at threonine 181.

NOTE. Features 1–43 represent quantitative EEG features. Power is the relative power per frequency band (delta [0–4 Hz], theta [4–8 Hz], alpha1 [8–10 Hz], alpha2 [10–13 Hz], beta [13–30 Hz]). Theta/alpha ratio is calculated as theta/(theta + alpha1 + alpha2). MST highest degree is the maximum degree (i.e., number of links for a given node) within the MST. MST leaf number is the number of nodes in the MST with only one link (i.e., degree). MST tree hierarchy is a measure of optimal network organization. Fazekas score is a measure of white-matter hyperintensities on T2-weighted fluid-attenuated inversion recovery (FLAIR) imaging.

### 2.3. Data handling

All the measures from clinical and neuropsychological data, CSF and neuroimaging biomarkers, and visual EEG rating (Table 1) were added to the data file with qEEG data per subject. Missing data were imputed by the average value over the two tested diagnostic groups for a particular feature. Features with  $\geq 33\%$  missing values were excluded from analyses.

### 2.4. Classification algorithm

The classification algorithm was built in the machine-learning module of BrainWave. The random forest approach

was used to build a classifier to differentiate between DLB and AD, DLB and controls, and AD and controls. Each decision tree in the random forest is built using a bootstrap sample (i.e., new training set), with replacement, from the original data (i.e., training set). Each new training set of features is randomly drawn from the original data set of features. This bootstrap aggregating (i.e., bagging) and random feature selection help in reducing the variance of the model, avoid overfitting, and result in uncorrelated trees [15]. Consequently, in random forest, the cross-validation is done internally, and there is no need for a separate test set to estimate the generalization error of the training set [15].

The two random forest parameters, namely mTry (i.e., the number of input variables randomly chosen at each split

calculated by the square root of number of features) and nTree (i.e., the number of trees to grow for each forest) were set to 8 (square root of 61 features) and 500, respectively. Interestingly, the classification outcome is not highly sensitive to the choice of these parameters [28].

In every classification, each feature receives a variable importance (VIMP) score between 0 and 1. For each analysis, it was possible to manually include and exclude features from being used in the tree. By doing so, it was possible to determine the performance of the classifier for various combinations of clinical and EEG features.

Three performance metrics, accuracy, sensitivity, and specificity, were used to assess the performance of the random forest in discriminating DLB, AD, and controls.

Additional methodologic details are provided in the [Supplementary Material](#).

### 3. Results

#### 3.1. Baseline characteristics

The baseline characteristics of the three groups are shown in [Table 2](#). DLB patients more frequently used medication affecting the central nervous system (CNS), compared to AD and controls ( $P < .05$ ). EEG power in theta and alpha1 band and peak frequency differed between the three groups ( $P < .05$ ), whereas power in the alpha2 and beta band differed between DLB and controls, and between DLB and AD, but not between AD and controls.

#### 3.2. Classifier results

For all three data sets (DLB vs. AD, DLB vs. controls, AD vs. controls), performance of the classifier using different combinations of clinical and/or EEG features is shown in [Table 3](#). An example of the machine-learning output is shown in [Fig. 1](#).

It was possible to discriminate between DLB and AD with an accuracy of 87%, a sensitivity of 88%, and a specificity of 86%, when all features were included. qEEG increased the accuracy of the combination of clinical, biomarker, and visual EEG features with 7% ([Table 3](#)). Differentiation between DLB and controls was possible with an accuracy of 94% and a sensitivity and specificity of 95% and 93%, respectively. For differentiation between AD and controls, an accuracy of 91% with a sensitivity of 92% and a specificity of 91% was achieved.

#### 3.3. Feature importance

[Fig. 2](#) shows the VIMP scores of the features in the group analyses when all features were included. For discrimination between DLB and AD, EEG highest beta power was the most important feature, followed by mean beta power. Using highest beta power as the only feature resulted in an accuracy, sensitivity, and specificity of 71%. Clinical features, MRI, and CSF biomarkers were of limited value in the

discrimination between DLB and AD ([Table 3](#)). qEEG solely had a higher diagnostic value in the differentiation between DLB and AD than the combination of the other multimodal variables without qEEG.

For DLB and controls, the qEEG measure theta/alpha ratio was the most important discriminating feature, followed by the visual EEG feature “diffuse abnormalities”. Using theta/alpha ratio as the only feature resulted in an accuracy and specificity of 83% with a sensitivity of 82%.

For discrimination between AD and controls, the clinical feature MMSE showed the highest VIMP score, followed by VAT, TMT-A, and CSF Tau/A $\beta_{42}$  ratio. MMSE, solely, provided an accuracy of 87% with a sensitivity of 93% and specificity of 81%.

## 4. Discussion

This study is the first to combine visual and qEEG measures with multimodal diagnostic tests in a machine learning-based classifier. For all three groups (DLB vs. AD, DLB vs. controls, and AD vs. controls), reasonable diagnostic accuracies ( $>85\%$ ) could be achieved when all preselected diagnostic variables were used. However, when studying variable importance, different “profiles” were found. (q)EEG features were identified as the most important for discrimination between DLB and AD and DLB and controls. Interestingly, the accuracy of the classifier for discrimination between DLB and AD was higher when only qEEG features were used, than with a combination of clinical features (including MRI and CSF analysis) and visual EEG. When discriminating AD from controls, cognitive tests (e.g., MMSE) were more valuable. (q)EEG did not have additional value for this discrimination.

#### 4.1. DLB vs. AD and controls

To date, little evidence is available on machine-learning classifiers for the diagnosis of DLB [12]. Nonautomated (q)EEG as a diagnostic modality has been more extensively studied and seems valuable for this diagnosis [9,10,29,30]. Visual EEG abnormalities are a supportive feature in the clinical criteria for DLB [1]. The present findings strongly support the potential of (q)EEG for the differentiation between DLB and AD.

Although the relative power in the delta, alpha2, and beta band was significantly different between DLB and AD, power in the beta band turned out to be the most important differentiating feature. This finding is remarkable, as beta power has not been reported previously to have specific discriminative value between these two forms of dementia. Medication use does not seem to be a likely underlying cause, because most medication that was used in the DLB group (e.g., cholinesterase inhibitors) would be expected to cause an increase in beta power [31], while instead a decrease was found. Likewise,

Table 2  
Patient characteristics

	DLB	AD	Control
N	66	66	66
Age, y	70 (9)	70 (9)	70 (7)
Sex, female	14 (21%)	14 (21%)	14 (21%)
Disease duration, y	2.9 (2.2)	3.3 (2.2)	3.6 (4.8)
CNS medication*†	16 (24.2%)	6 (9.1%)	6 (9.1%)
Rivastigmine	6 (9.1%)	4 (6.1%)	1 (1.5%)
Haloperidol	1 (1.5%)	1 (1.5%)	1 (1.5%)
Clozapine	2 (3%)	0 (0%)	0 (0%)
Quetiapine	2 (3%)	0 (0%)	0 (0%)
AED	3 (4.5%)	1 (1.5%)	2 (3%)
Other CNS medication	3 (4.5%)	0 (0%)	2 (3%)
MMSE‡	23 (5) (n = 59)	21 (5) (n = 63)	28 (1) (n = 66)
VAT‡	7.9 (3.5) (n = 47)	5.6 (4.3) (n = 60)	11.5 (.8) (n = 62)
TMT-A‡, sec	123 (86) (n = 47)	87 (63) (n = 54)	43 (15) (n = 63)
Digit span forward§	11.5 (2.5) (n = 50)	10.5 (3.2) (n = 61)	12.4 (3.0) (n = 64)
Digit span backward	6.5 (2.8) (n = 49)	6.6 (3.0) (n = 60)	9.3 (2.9) (n = 64)
Hallucinations‡	16 (37.2%) (n = 43)	3 (5.8%) (n = 52)	0 (0%) (n = 40)
Extrapyramidal signs	32 (72.7%) (n = 44)	7 (13.5%) (n = 52)	4 (9.1%) (n = 44)
Bradykinesia‡	26 (59.1%) (n = 44)	2 (3.8%) (n = 52)	1 (2.3%) (n = 44)
Rigidity‡	26 (59.1%) (n = 44)	2 (3.8%) (n = 52)	3 (6.8%) (n = 44)
Tremor	6 (13.6%) (n = 44)	4 (7.8%) (n = 51)	2 (4.5%) (n = 44)
RBD	23 (88.5%) (n = 26)	NA	NA
Cognitive fluctuations	42 (91.3%) (n = 46)	NA	NA
CSF			
A $\beta$ <sub>42</sub> ‡	677.7 (236.7) (n = 47)	503.6 (218.2) (n = 48)	835.0 (245.0) (n = 37)
Tau‡§	341.4 (187.9) (n = 47)	601.7 (338.1) (n = 48)	326.2 (156.2) (n = 37)
p-Tau‡§	56.7 (26.4) (n = 47)	86.9 (39.7) (n = 48)	52.1 (19.0) (n = 37)
Neuroimaging			
MTA score‡	1.0 (0.25–1.5) (n = 45)	1.5 (1.0–2.0) (n = 59)	0.5 (0.0–1.0) (n = 59)
GCA score‡	1.0 (1.0–2.0) (n = 45)	1.0 (1.0–2.0) (n = 59)	1.0 (0.0–1.0) (n = 59)
Fazekas score	1.0 (0.0–1.0) (n = 45)	1.0 (0.0–2.0) (n = 59)	1.0 (0.0–1.0) (n = 59)
Power			
Delta band*†	0.42 (0.16)	0.29 (0.12)	0.27 (0.11)
Theta band‡	0.32 (0.12)	0.22 (0.11)	0.14 (0.07)
Alpha1 band‡	0.11 (0.07)	0.17 (0.10)	0.23 (0.14)
Alpha2 band*†	0.05 (0.03)	0.11 (0.07)	0.12 (0.07)
Beta band*†	0.08 (0.04)	0.16 (0.07)	0.18 (0.07)
Peak frequency‡	7.02 (0.91)	8.06 (1.17)	8.84 (0.91)
Theta/alpha ratio‡	0.67 (0.15)	0.45 (0.18)	0.30 (0.13)

Abbreviations: NA, not available; AD, Alzheimer's disease; DLB, dementia with Lewy bodies; MMSE, mini mental state examination; VAT, visual association test; TMT-A, trail making test part A; MTA, medial temporal lobe atrophy; GCA, global cortical atrophy; RBD, REM sleep behavior disorder; A $\beta$ <sub>42</sub>, amyloid- $\beta$  1–42; Tau, total tau; p-Tau, tau phosphorylated at threonine 181; MST, minimum spanning tree; AED, anti-epileptic drugs; CNS, central nerve system.

NOTE. Data are mean (SD), median (interquartile range), or n (%). Disease duration measured as years since onset of complaints. TMT-A scores are presented as time needed to complete the task; higher scores mean worse performance. Diagnoses, including "subjective cognitive decline" for the control group, were made in a consensus meeting after clinical workup; therefore, some control subjects were using medication affecting the central nerve system. Hallucinations were assessed using the NPI. Cognitive fluctuations, extrapyramidal signs, and RBD were qualitatively assessed on their presence or absence at the first clinical presentation. Fazekas score is a measure of white-matter hyperintensities on T2-weighted fluid-attenuated inversion recovery (FLAIR) imaging. Power is the relative power per frequency band (delta [0–4 Hz], theta [4–8 Hz], alpha1 [8–10 Hz], alpha2 [10–13 Hz], and beta [13–30 Hz]). Peak frequency is the frequency with highest power in range between 4 and 13 Hz. Theta/alpha ratio is an index that shows the percentage of theta versus alpha spectral potential during resting state, computed as theta/(theta + alpha1 + alpha2).

\*Significantly different between DLB and controls.

†Significantly different between AD and DLB.

‡Significantly different between all groups ( $P < .05$ ).

§Significantly different between AD and controls.

||Significantly different between the two dementia groups and controls ( $P < .05$ ).

muscle artifacts cannot explain the difference since motor symptoms in DLB would have resulted in increased beta power. A more plausible explanation is that defective dopaminergic networks in DLB, that are intact in

AD, could be related to the lower beta power found in this group and therefore be a cause for the high discriminative value of this EEG measure. Previous studies have linked changes in beta power, and beta peak frequency to



Table 3  
Classifier results

Group and feature selection	Accuracy (%)	Sensitivity (%)	Specificity (%)
<b>DLB vs. AD</b>			
All features	87	88	86
Only clinical features	66	65	67
Clinical features + biomarkers	71	71	70
Clinical features + biomarkers + visual EEG	78	76	80
Quantitative and visual EEG	85	86	84
Only quantitative EEG	85	86	85
<b>DLB vs. controls</b>			
All features	94	95	92
Only clinical features	89	92	86
Clinical features + biomarkers	86	87	85
Clinical features + biomarkers + visual EEG	90	87	94
Quantitative and visual EEG	91	93	89
Only quantitative EEG	92	95	89
<b>AD vs. controls</b>			
All features	91	92	91
Only clinical features	90	93	88
Clinical features + biomarkers	93	94	92
Clinical features + biomarkers + visual EEG	93	93	92
Quantitative and visual EEG	63	62	64
Only quantitative EEG	62	63	62

Abbreviations: AD, Alzheimer's disease; DLB, dementia with Lewy bodies; EEG, electroencephalography.

NOTE. Clinical features include hallucinations, extrapyramidal signs, and neuropsychological test results (mini mental state [MMSE] score; visual association test (VAT) score; Trail-making-test (TMT)-A score; and digit span forward and backward); Biomarkers include MRI (medial temporal lobe atrophy [MTA] score; global cortical atrophy [GCA] score; Fazekas score), and cerebrospinal fluid (CSF) (amyloid-β 1-42 (Aβ<sub>42</sub>); total Tau, phosphorylated Tau (p-Tau), and tau to Aβ<sub>42</sub> ratio) data.

PD, PD dementia, and dopaminergic medication [32–34]. Furthermore, beta power can also be influenced by the cholinergic system [35,36]. Both the cholinergic system and the beta band have been related to the processes of attention [16,37]. The cholinergic system is more severely impaired in DLB brains than in AD brains [38,39], and this cholinergic deficit and associated attentional deficits might also be a discriminating aspect between the two types of dementia. Finally, the lower beta power in DLB could be caused by the overall shift in EEG activity from higher to lower frequency bands in DLB that has been shown by previous work [9,10,29].

For DLB and controls, theta/alpha ratio was the most important discriminating feature. In DLB, theta power is higher, and alpha power is lower than in AD and controls (Table 2). Therefore, EEG slowing seems to be more remarkable in DLB, which is in line with previous results [9,30]. The greater EEG slowing in DLB makes theta/alpha ratio a potentially important discriminating factor between DLB and controls.

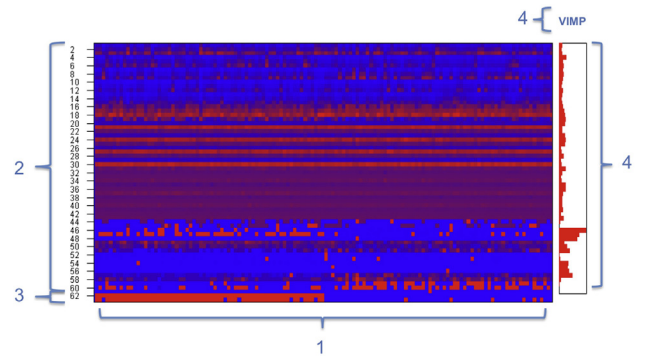


Fig. 1. Example of machine-learning output for discrimination between AD and controls. 1 = subjects on x-axis arranged by diagnosis: 1–66 = AD patients, 67–132 = controls (represented in feature 62 on y-axis); 2 = diagnostic features 1–61 (Table 1); 3 = feature 62: “true” diagnostic labels, set by authors, dividing subjects between AD patients (red) and controls (blue); feature 63: diagnostic labels set by classifier; 4 = variable importance (VIMP) score per feature.

4.2. AD vs. controls

For classification between AD and controls, an accuracy of 91% is in line with accuracies reported earlier in machine learning–based classification techniques for AD [13,40,41]. Although most of these studies were imaging based, some studies used EEG data. Established neuropsychological tests and CSF biomarkers [23] have high VIMP scores in the present classifier, which is in accordance with their importance in current clinical decision making. However, the observed accuracy of 63% when including only EEG features is low compared to previous studies on qEEG-based classifiers reporting accuracies of 80%–95% [40,41]. A possible explanation for this could be the use of healthy

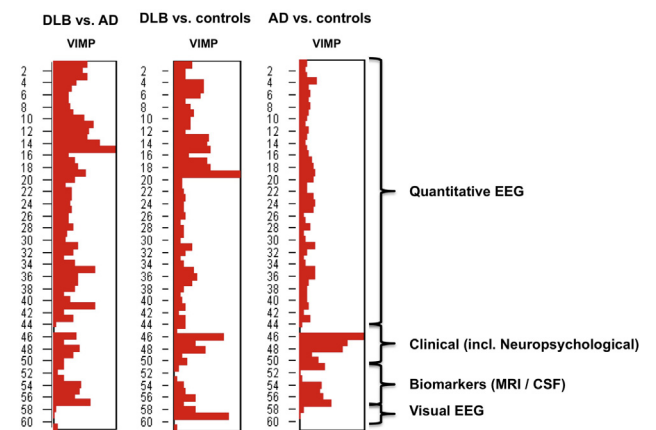


Fig. 2. Variable importance scores in the three main classifications using all features. VIMP scores showing the relative importance of different groups of features for discrimination between DLB and AD, DLB and controls, and AD and controls, respectively. Variable importance score (VIMP) on a 0–1 scale; Clinical features: hallucinations, mini mental state examination (MMSE) score; visual association test (VAT) score; trail-making-test (TMT)-A score; Digit span forward and backward. CSF, cerebrospinal fluid.

subjects as controls in these studies [40,41]. In the present study, subjects with SCD were used as controls, and possibly, they have more EEG abnormalities than healthy elderly without SCD. Furthermore, earlier studies have shown that EEG results in AD can be heterogeneous, abnormalities are less profound than in DLB, and normal EEGs frequently occur [29,39]. These factors could all be contributing to the lower accuracy found.

#### 4.3. Strengths and limitations

A strength of this study is that the random forest algorithm produces a highly accurate classifier (with cross-validation build into the method and only two required parameters, none of which is critical for the results). The method is easy to use, has a high interpretability, runs efficiently on large data sets, and gives estimates of which variables are important in the classification [14]. Moreover, an essential part of the feature selection is done internally in the random forest algorithm and thus helps in reducing the variance of the model and avoids overfitting [15]. Second, this study compared a relatively large group of DLB patients with a carefully matched group of controls and AD patients. Third, subjects with SCD were used as a control group. These subjects visited the memory clinic with subjective memory complaints, and therefore represent a heterogeneous group. In clinical practice, this group is the exact population from which patients with a diagnosis of dementia need to be distinguished.

This study has some limitations. First, not all possibly relevant variables (e.g., REM sleep behavior disorder, cognitive fluctuations, results of DAT-SPECT scans and EEG variability) were available or scored quantitatively to be included as features in the classification algorithm. Second, some features had a relatively high number of missing values, which were excluded from the analysis or imputed. For instance, cognitive fluctuations and REM sleep behavior disorder were excluded from the analyses for this reason. In the case of CSF biomarkers, hallucinations, and extrapyramidal signs, missing values were imputed. Notably, CSF biomarkers turned out to be important discriminating features between AD and controls. This not only implies that CSF biomarkers could have had a higher VIMP score if no data had been missing, but also that when a continuous feature is important in distinguishing two groups, its performance is not fully influenced by the number of missing values. In contrast, in the case of categorical features, it is more difficult to impute the missing values with a meaningful average. Therefore, missing data in these types of features can result in an underestimation of the importance of categorical features.

Finally, initial selection of available features by the authors could have resulted in circular reasoning by including features that are already known to be important discriminating variables between two groups, whereas there is no autopsy confirmed diagnosis as a gold standard. Nonetheless,

during a follow-up period of 0–7 years, none of the DLB diagnoses and only two AD diagnoses (mean follow-up of 21.2 months) were changed (the diagnosis was converted in 15 SCD subjects).

In summary, machine learning–based diagnostic classifiers show that qEEG is a valuable contribution in differentiating between DLB and AD. EEG data are easily obtained, as EEG is a low-cost and noninvasive procedure. Further research should elucidate the diagnostic value of relative beta power and theta/alpha ratio for the diagnosis of DLB. VIMP scores quantify the importance of a variable but do not provide information about the actual value, such as a cutoff, and this needs to be studied separately. Still, the current findings suggest a bigger role for (q)EEG in the diagnostic criteria for DLB.

#### Acknowledgments

The authors thank Dr. H.d.Waal for her contribution in the EEG data collection.

This study was partly supported by ZONMW TOP grant 40-00812-98-13009.

#### Supplementary data

Supplementary data related to this article can be found at <http://dx.doi.org/10.1016/j.dadm.2016.07.003>.

#### RESEARCH IN CONTEXT

1. Systematic review: We searched PubMed for studies on automatic classification of dementia. Although machine-learning techniques have been studied to classify AD patients, we found only little work on automatic differentiation between DLB and AD. These relevant articles were cited. We found no studies that used multimodal diagnostic data for this purpose.
2. Interpretation: We hypothesize that machine-learning techniques can aid and provide insight in the multimodal diagnostic process of dementia. Our findings show that (q)EEG is of important additional value for the discrimination between DLB and AD. Furthermore, some potentially important (q)EEG features were identified.
3. Future directions: Future studies should further elaborate the diagnostic value of EEG features that were identified as important by our classifier (e.g., relative beta power and theta/alpha ratio). Overall, this study suggests a bigger role for (q)EEG in the clinical diagnostic criteria of DLB.

## References

- [1] McKeith IG, Dickson DW, Lowe J, Emre M, O'Brien JT, Feldman H, et al. Diagnosis and management of dementia with Lewy bodies: third report of the DLB Consortium. *Neurology* 2005;65:1863–72.
- [2] McKhann GM, Knopman DS, Chertkow H, Hyman BT, Jack CR, Kawas CH, et al. The diagnosis of dementia due to Alzheimer's disease: Recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. *Alzheimers Dement* 2011;7:263–9.
- [3] Morra LF, Donovick PJ. Clinical presentation and differential diagnosis of dementia with Lewy bodies: A review. *Int J Geriatr Psychiatry* 2014;29:569–76.
- [4] Nelson PT, Jicha GA, Kryscio RJ, Abner EL, Schmitt FA, Cooper G, et al. Low sensitivity in clinical diagnoses of dementia with Lewy bodies. *J Neurol* 2010;257:359–66.
- [5] Walker Z, Possin KL, Boeve BF, Aarsland D. Lewy body dementias. *Lancet* 2015;386:1683–97.
- [6] Norton S, Matthews FE, Barnes DE, Yaffe K, Brayne C. Potential for primary prevention of Alzheimer's disease: an analysis of population-based data. *Lancet Neurol* 2014;13:788–94.
- [7] Jack CR, Knopman DS, Jagust WJ, Petersen RC, Weiner MW, Aisen PS, et al. Tracking pathophysiological processes in Alzheimer's disease: an updated hypothetical model of dynamic biomarkers. *Lancet Neurol* 2013;12:207–16.
- [8] Nedelska Z, Ferman TJ, Boeve BF, Przybelski SA, Lesnick TG, Murray ME, et al. Pattern of brain atrophy rates in autopsy-confirmed dementia with Lewy bodies. *Neurobiol Aging* 2015; 36:452–61.
- [9] Lee H, Brekelmans GJ, Roks G. The EEG as a diagnostic tool in distinguishing between dementia with Lewy bodies and Alzheimer's disease. *Clin Neurophysiol* 2015;126:1735–9.
- [10] Roks G, Korff ES, van der Flier WM, Scheltens P, Stam CJ. The use of EEG in the diagnosis of dementia with Lewy bodies. *J Neurol Neurosurg Psychiatry* 2008;79:377–80.
- [11] Van Der Flier WM, Pijenburg YA, Prins N, Lemstra AW, Bouwman FH, Teunissen CE, et al. Optimizing patient care and research: The Amsterdam dementia cohort. *J Alzheimers Dis* 2014; 41:313–27.
- [12] Zaffalon M, Wesnes K, Petrini O. Reliable diagnoses of dementia by the naive credal classifier inferred from incomplete cognitive data. *Artif Intell Med* 2003;29:61–79.
- [13] Falahati F, Westman E, Simmons A. Multivariate data analysis and machine learning in Alzheimer's disease with a focus on structural magnetic resonance imaging. *J Alzheimers Dis* 2014;41:685–708.
- [14] Geurts P, Irthum A, Wehenkel L. Supervised learning with decision tree-based methods in computational and systems biology. *Mol Biosyst* 2009;5:1593–605.
- [15] Breiman L, University of California. Random forest. *Mach Learn* 1999;45:1–35.
- [16] Dauwan M, van Dellen E, van Bostel L, van Straaten EC, de Waal H, Lemstra AW, et al. EEG directed connectivity from posterior brain regions is decreased in Dementia with Lewy bodies: a comparison with Alzheimer's disease and controls. *Neurobiol Aging* 2016;41:122–9.
- [17] van Dellen E, de Waal H, van der Flier WM, Lemstra AW, Slooter AJ, Smits LL, et al. Loss of EEG Network Efficiency Is Related to Cognitive Impairment in Dementia With Lewy Bodies. *Mov Disord* 2015; 30:1785–93.
- [18] Cummings JL, Mega M, Gray K, Rosenberg-Thompson S, Carusi DA, Gornbein J. The Neuropsychiatric Inventory comprehensive assessment of psychopathology in dementia. *Neurology* 1994;44:2308.
- [19] Folstein MF, Folstein SE, McHugh PR. "Mini-mental state". A practical method for grading the cognitive state of patients for the clinician. *J Psychiatr Res* 1975;12:189–98.
- [20] Reitan RM. Validity of the Trail Making Test as an indicator of organic brain damage. *Percept Mot Skills* 1958;8:271–6.
- [21] Lindeboom J, Schmand B, Tulner L, Walstra G, Jonker C. Visual association test to detect early dementia of the Alzheimer type. *J Neurol Neurosurg Psychiatry* 2002;73:126–33.
- [22] Lindeboom J, Matto D. Digit series and Knox cubes as concentration tests for elderly subjects. *Tijdschr Gerontol Geriatr* 1994;25:63–8.
- [23] Duits FH, Teunissen CE, Bouwman FH, Visser PJ, Mattsson N, Zetterberg H, et al. The cerebrospinal fluid "Alzheimer profile": easily said, but what does it mean? *Alzheimers Dement* 2014;10:713–7232.
- [24] Liedorp M, Van Der Flier WM, Hoogervorst EL, Scheltens P, Stam CJ. Associations between patterns of EEG abnormalities and diagnosis in a large memory clinic cohort. *Dement Geriatr Cogn Disord* 2009;27:18–23.
- [25] Gasser T, Bacher P, Steinberg H. Test-retest reliability of spectral parameters of the EEG. *Electroencephalogr Clin Neurophysiol* 1985; 60:312–9.
- [26] Lobier M, Siebenhüner F, Palva S, Palva JM. Phase transfer entropy: A novel phase-based measure for directed connectivity in networks coupled by oscillatory interactions. *Neuroimage* 2014;85:853–72.
- [27] Stam CJ, Tewarie P, Van Dellen E, van Straaten EC, Hillebrand A, Van Mieghem P. The trees and the forest: Characterization of complex brain networks with minimum spanning trees. *Int J Psychophysiol* 2014;92:129–38.
- [28] Liaw A, Wiener M. Classification and Regression by randomForest. *R News* 2002;2:18–22.
- [29] Bonanni L, Thomas A, Tiraboschi P, Perfetti B, Varanese S, Onofrij M. EEG comparisons in early Alzheimer's disease, dementia with Lewy bodies and Parkinson's disease with dementia patients with a 2-year follow-up. *Brain* 2008;131:690–705.
- [30] Cromarty RA, Elder GJ, Graziadio S, Baker M, Bonanni L, Onofrij M, et al. Neurophysiological biomarkers for Lewy body dementias. *Clin Neurophysiol* 2016;127:349–59.
- [31] Fogelson N, Kogan E, Korczyn AD, Giladi N, Shabtai H, Neufeld MY. Effects of rivastigmine on the quantitative EEG in demented Parkinsonian patients. *Acta Neurol Scand* 2003;107:252–5.
- [32] Pezard L, Jech R, Ruzicka E. Investigation of non-linear properties of multichannel EEG in the early stages of Parkinson's disease. *Clin Neurophysiol* 2001;112:38–45.
- [33] Gu Y, Chen J, Lu Y, Pan S. Integrative Frequency Power of EEG Correlates with Progression of Mild Cognitive Impairment to Dementia in Parkinson's Disease. *Clin EEG Neurosci* 2016;47:113–7.
- [34] Melgari JM, Curcio G, Mastrolilli F, Salomone G, Trotta L, Tombini M, et al. Alpha and beta EEG power reflects L-dopa acute administration in parkinsonian patients. *Front Aging Neurosci* 2014; 6:302.
- [35] Podol'skii II, Vorob'ev VV, Belova NA. Long-term changes in EEG spectra of the hippocampus and neocortex during pharmacological action on the cholinergic system. *Zh Vyssh Nerv Deiat Im I P Pavlova* 2000;50:982–90.
- [36] Sloan EP, Fenton GW, Standage KP. Anticholinergic drug effects on quantitative electroencephalogram, visual evoked potential, and verbal memory. *Biol Psychiatry* 1992;31:600–6.
- [37] Pepeu G, Giovannini MG, Bracco L. Effect of cholinesterase inhibitors on attention. *Chem Biol Interact* 2013;203:361–4.
- [38] Onofrij M, Thomas A, Iacono D, Luciano AL, Di Iorio A. The effects of a cholinesterase inhibitor are prominent in patients with fluctuating cognition: a part 3 study of the main mechanism of cholinesterase inhibitors in dementia. *Clin Neuropharmacol* 2003;26:239–51.
- [39] Kai T, Asai Y, Sakuma K, Koeda T, Nakashima K. Quantitative electroencephalogram analysis in dementia with Lewy bodies and Alzheimer's disease. *J Neurol Sci* 2005;237:89–95.
- [40] Lehmann C, Koenig T, Jelic V, Prichep L, John RE, Wahlund LO, et al. Application and comparison of classification algorithms for recognition of Alzheimer's disease in electrical brain activity (EEG). *J Neurosci Methods* 2007;161:342–50.
- [41] Trambaiolli LR, Lorena AC, Fraga FJ, Kanda PA, Anghinah R, Nitirini R. Improving Alzheimer's disease diagnosis with machine learning techniques. *Clin EEG Neurosci* 2011;42:160–5.