Neural Networks 171 (2024) 242–250

Contents lists available at ScienceDirect

Neural Networks

journal homepage: www.elsevier.com/locate/neunet

A deep learning model for the detection of various dementia and MCI pathologies based on resting-state electroencephalography data: A retrospective multicentre study

Yusuke Watanabe [a,#], Yuki Miyazaki [b,#], Masahiro Hata [b,#], Ryohei Fukuma [a,c], Yasunori Aoki [b,d], Hiroaki Kazui [e], Toshihiko Araki [f], Daiki Taomoto [b], Yuto Satake [b], Takashi Suehiro [b], Shunsuke Sato [b], Hideki Kanemoto [b], Kenji Yoshiyama [b], Ryouhei Ishii [b,g], Tatsuya Harada [h,i], Haruhiko Kishima [c], Manabu Ikeda [b], Takufumi Yanagisawa [a,c,*]

[a] *Institute for Advanced Co-creation Studies, Osaka University, Osaka, Japan*
[b] *Department of Psychiatry, Osaka University Graduate School of Medicine, Osaka, Japan*
[c] *Department of Neurosurgery, Osaka University Graduate School of Medicine, Osaka, Japan*
[d] *Department of Psychiatry, Nippon Life Hospital, Osaka, Japan*
[e] *Department of Neuropsychiatry, Kochi Medical School, Kochi University, Kochi, Japan*
[f] *Department of Medical Technology, Osaka University Hospital, Osaka, Japan*
[g] *Department of Occupational Therapy, Graduate School of Rehabilitation Science, Osaka Metropolitan University, Habikino, Japan*
[h] *Research Center for Advanced Science and Technology, The University of Tokyo, Tokyo, Japan*
[i] *RIKEN, Tokyo, Japan*

ARTICLE INFO

Keywords:
Dementia screening
Mild cognitive impairment
Alzheimer's disease
Dementia with Lewy bodies
Idiopathic normal-pressure hydrocephalus
Multiple centres

ABSTRACT

Dementia and mild cognitive impairment (MCI) represent significant health challenges in an aging population. As the search for noninvasive, precise and accessible diagnostic methods continues, the efficacy of electroencephalography (EEG) combined with deep convolutional neural networks (DCNNs) in varied clinical settings remains unverified, particularly for pathologies underlying MCI such as Alzheimer's disease (AD), dementia with Lewy bodies (DLB) and idiopathic normal-pressure hydrocephalus (iNPH). Addressing this gap, our study evaluates the generalizability of a DCNN trained on EEG data from a single hospital (Hospital #1). For data from Hospital #1, the DCNN achieved a balanced accuracy (bACC) of 0.927 in classifying individuals as healthy (n = 69) or as having AD, DLB, or iNPH (n = 188). The model demonstrated robustness across institutions, maintaining bACCs of 0.805 for data from Hospital #2 (n = 73) and 0.920 at Hospital #3 (n = 139). Additionally, the model could differentiate AD, DLB, and iNPH cases with bACCs of 0.572 for data from Hospital #1 (n = 188), 0.619 for Hospital #2 (n = 70), and 0.508 for Hospital #3 (n = 139). Notably, it also identified MCI pathologies with a bACC of 0.715 for Hospital #1 (n = 83), despite being trained on overt dementia cases instead of MCI cases. These outcomes confirm the DCNN's adaptability and scalability, representing a significant stride toward its clinical application. Additionally, our findings suggest a potential for identifying shared EEG signatures between MCI and dementia, contributing to the field's understanding of their common pathophysiological mechanisms.

Introduction

Dementia, which affects an estimated 50 million people worldwide, is a significant health challenge characterized by the progressive decline of memory, attention, and executive functions, which are pivotal for maintaining daily independence (Prince et al., 2015). Accurate identification of its various subtypes, such as Alzheimer's disease (AD), dementia with Lewy bodies (DLB), and idiopathic normal-pressure hydrocephalus (iNPH), is crucial for appropriate clinical management (Barker et al., 2002; Williams and Malm, 2016; McKeith et al., 2017;

* Corresponding author at: Department of Neurosurgery, Osaka University Graduate School of Medicine, 2-2 Yamadaoka, Suita, Osaka 565-0871, Japan.
E-mail address: tyanagisawa@nsurg.med.osaka-u.ac.jp (T. Yanagisawa).
# These authors contributed equally to this work.

https://doi.org/10.1016/j.neunet.2023.12.009
Received 21 January 2023; Received in revised form 12 November 2023; Accepted 4 December 2023
Available online 6 December 2023

Arvanitakis et al., 2019; Leuzy et al., 2022).

The pursuit of early and precise differentiation of dementia subtypes is hindered by several factors. Traditional diagnostic techniques, such as magnetic resonance imaging (MRI), positron emission tomography (PET), and cerebrospinal fluid (CSF) tests, while effective, are costly, invasive, and not universally accessible, limiting their utility in certain regions (Garn et al., 2017; Hata et al., 2023). Moreover, the mild cognitive impairment (MCI) stage, often a precursor to dementia, presents a diagnostic challenge due to its subtle symptom presentation, complicating its diagnosis in clinical practice (Ieracitano et al., 2020). These issues underscore the necessity for a more accessible, noninvasive, and standardized screening tool capable of discerning the nuanced differences between dementia subtypes.

Further highlighting the urgency for early and accurate diagnosis, the recent identification of an antibody targeting amyloid beta protofibrils opens a promising avenue for decelerating the progression of AD in its nascent stages (Swanson et al., 2021). Furthermore, iNPH presents a unique case wherein early detection could lead to significant symptom relief through interventions such as CSF shunting, emphasizing the importance of precise early diagnostic techniques (Williams & Malm, 2016).

Responding to the need for noninvasive, cost-effective, and widely available diagnostic methods, our research has pivoted toward the utilization of electroencephalography (EEG). This neurophysiological recording technique, renowned for its noninvasiveness and cost effectiveness, holds untapped potential for enhancing the early detection and differentiation of dementia. The EEG methods proposed by existing studies, however, have often been less successful in practical application due to limited comparative analyses being available in the literature, misapplication of machine learning techniques, e.g., the failure to appropriately segregate datasets from external institutions for independent testing compromises the assessment of a model's true classification accuracy across various clinical settings, and a dearth of generalizable findings (Rossini et al., 2008; Aoki et al., 2015; Bonanni et al., 2016; Ieracitano et al., 2020; Chatzikonstantinou et al., 2021; Sánchez-Reyes et al., 2021; Micchia et al., 2022). A particularly promising yet underexplored domain is the use of deep convolutional neural networks (DCNNs) with EEG data for dementia classification; notably, this technique may bypass the intricacies of feature engineering and reveal the nuanced patterns indicative of various dementia subtypes.

In this study, we are committed to testing three pivotal hypotheses, each aimed at overcoming the challenges of early diagnosis of dementia and its diverse subtypes. Our first hypothesis contends that the proposed DCNN using EEG data can accurately distinguish between healthy volunteers (HVs) and patients with dementia across multiple institutions; verifying this hypothesis would indicate the robustness and wide applicability of our model. Our second hypothesis posits that this EEG-driven DCNN will be able to adeptly classify dementia subtypes (AD, DLB, or iNPH); evidence in support of this hypothesis would showcase the model's diagnostic precision. Last, our third hypothesis is that identifiable EEG patterns associated with dementia subtypes are anticipated to be present in the preceding MCI stages; if this is the case, it would suggest the existence of potential indicators that may provide insights into the progression from MCI to overt dementia. Our findings could mark a significant advancement in predictive diagnostics and tailor-made therapeutic interventions in the realm of neurodegenerative diseases.

Methods

Study population

Expanding upon our previous study (Hata et al., 2023), we retrospectively gathered EEG data from 482 subjects across three recording sites: Osaka University Hospital (Hospital O), Kochi University Hospital (Hospital K), and Nippon Life Hospital (Hospital N). In addition, 69 healthy subjects from Osaka University were enrolled. All patients underwent baseline assessments to collect demographic characteristics and evaluate cognitive function, neuropsychiatric symptoms, and brain structure (using MRI or computed tomography [CT]). Additionally, laboratory tests (such as blood cell count and blood chemistry measurements) and assessments of thyroid hormones and vitamin B1, vitamin B12, and folic acid levels were conducted. Furthermore, EEG recordings from April 2009 to September 2022 were obtained. In addition, a subset of subjects underwent the following optional assessments: single photon emission computed tomography (SPECT), measurement of CSF markers, ^{123}I-metaiodobenzylguanidine myocardial scintigraphy, dopamine transporter SPECT, and polysomnography (PSG). Based on these examinations, we excluded patients with any physical disorders that could impact cognition, including endocrine disorders, cerebral structural lesions, a history of brain injury, and drug/alcohol abuse.

Based on the Clinical Dementia Rating (CDR; Morris, 1993) and Mini-Mental State Examination (MMSE; Folstein et al., 1975), the clinical subjects were stratified into MCI and dementia groups. MCI was operationally defined as having a CDR value of 0.5 and an MMSE score of 24 or higher. Dementia was defined as having a CDR value of 0.5 and an MMSE score lower than 24 or a CDR value of 1 or higher. Only clinical subjects with AD, DLB, or iNPH (refer to Table 1 & Figure S1) were included in this study; subjects with other subtypes of dementia were excluded. The diagnoses were made independently by a panel of psychiatrists specializing in dementia at each hospital using the aforementioned assessments as diagnostic tools.

Data were excluded from analysis if the subjects met any of the following conditions: being under 40 years old ($n = 15$), having missing age data ($n = 1$) or missing MMSE scores ($n = 4$), or obtaining a diagnosis of iNPH while displaying normal cognition (CDR = 0 and MMSE score $\geq$ 27) ($n = 3$). HVs were recruited from the general population attending Osaka University through the distribution of posters and internet advertisements. The inclusion criteria were as follows: (1) Japanese ethnicity; (2) an age of 40 years or older; (3) no prior history of psychiatric disorders or severe head injury, and no prior consultation with a psychiatrist, psychotherapist, neurologist, or neurosurgeon; (4) no history of alcohol or drug abuse; and (5) no impairments affecting daily life activities or global cognitive impairment (MMSE score $\geq$ 27).

The study procedures adhered to the guidelines of the Declaration of Helsinki and were performed in accordance with protocols approved by the Research Ethics Committee of the Osaka University Clinical Trial Centre (no. 14,448, UMIN000022957, Dementia 17,441–6), Kochi University (no. 31–52) and Nippon Life Hospital (2020–081). Each participant provided written informed consent after being informed of the purpose of the study. In instances where a patient was unable to provide informed consent, consent was obtained from their legal guardian.

EEG recording

EEG data were recorded continuously for a minimum duration of 20 min using a digital 19-channel scalp EEG device (EEG-1200, Nihon Koden, Tokyo, Japan, at Hospitals O and K; EEG-1260 or EEG-1290, Nihon Koden, Tokyo, Japan, at Hospital N), following the International 10–20 system. The following montages were employed: FP1–A1, F3–A1, C3–A1, P3–A1, O1–A1, FP2–A2, F4–A2, C4–A2, P4–A2, O2–A2, F7–A1, T7–A1, P7–A1, F8–A2, T8–A2, P8–A2, Fz–A1, Cz–A1, and Pz–A1. The bilateral ears were used as references. The impedance of the electrodes was maintained below 50 kΩ at Hospitals O and K and below 30 kΩ at Hospital N. The sampling rate was 500 Hz for all subjects, except for 44 subjects at Hospital K with an EEG sampling rate of 400 Hz. These data were then resampled to 500 Hz using the scipy.signal.resample function (Virtanen et al., 2020). The recording rooms at each hospital were electrically shielded, acoustically insulated, and dimly illuminated. During the recordings, each subject assumed a supine

Table 1
Demographic characteristics of participants.

Disease class	Cognitive function	Hospital	# of subjects	Female/male ratio	Age, years (median [IQR])	MMSE score (median [IQR])
HV	Normal	O	69	0.58	64 [11]	30 [0]
		K	3	0.67	47 [20]	30 [1]
		N	0	–	–	–
AD	MCI	O	34	0.50	78·5 [10]	26 [2]
	Dementia	O	101	0.73	75 [14]	19 [6]
		K	45	0.78	77 [9]	23 [6]
		N	119	0.66	83 [7]	20 [5]
DLB	MCI	O	22	0.45	76.5 [10]	26 [2]
	Dementia	O	59	0.59	79 [10]	20 [4]
		K	10	0.60	75 [9]	24 [6]
		N	17	0.59	83 [6]	20 [8]
iNPH	MCI	O	27	0.26	76 [8]	26 [2]
	Dementia	O	28	0.46	80.5 [8]	21 [5]
		K	15	0.60	78 [7]	22 [6]
		N	3	0.33	81 [8]	18 [3]
			Total = 551			

Note that sampling biases were present (e.g., a smaller sample size and smaller proportion of females in the iNPH group, younger age and higher MMSE scores in the HV group). Abbreviations: HV, healthy volunteer; AD, Alzheimer's disease; DLB, dementia with Lewy bodies; iNPH, idiopathic normal-pressure hydrocephalus; MCI, mild cognitive impairment; IQR, interquartile range.

position on a bed in the shielded room. All EEG recordings were conducted by experienced EEG technicians. An offline bandpass filter with a passband from 0.53 Hz to 100 Hz was applied.

Before the EEG recording sessions began, all subjects were given instructions to relax with their eyes closed and remain conscious throughout the recording. An experienced electrophysiologist monitored the EEG signals in real time. When the EEG signals indicated stage I or II sleep, the subjects were either instructed to awaken verbally or were stimulated auditorily (e.g., with a beep). If the recorded EEG data exhibited contamination from muscle artifacts, appropriate measures were taken. Additionally, to regulate their level of vigilance, the subjects were directed to open and close their eyes ten times consecutively during the middle of the EEG recording session.

EEG preprocessing

The analysis utilized the eyes-closed EEG data. EEG periods containing significant artifacts were identified through visual inspection and subsequently excluded. A notch filter centered at 60 Hz was utilized to reduce line noise. To compare EEG recordings across the dementia subtypes, a one-way analysis of variance (ANOVA) was performed for each frequency band, electrode position, and hospital (Figure S2).

Data splitting

All experiments were conducted using a between-subjects design. A fivefold cross-validation approach was employed for experiments at Hospital O. Within each fold, the dataset was divided into training, validation, and test subsets at proportions of 0.6, 0.2, and 0.2, respectively. Notably, each subject's data were included in only one of the three subsets to prevent potential data leakage.

Forward model

To perform the classification task, we deployed the MNet model, a deep convolutional neural network previously developed by our team (Aoe et al., 2019). The architecture of MNet is depicted in Figure S3, with layer configurations in Tables S1 & S2. Our model implementation is also available as open-source code on our GitHub repository (https://github.com/yanagisawa-lab/eeg-dementia-classification/tree/main/models/MNet). MNet begins by standardizing EEG data along the temporal axis in its first layer. It then employs two-dimensional convolutional layers to extract localized spatiotemporal features. The temporal resolution of MNet, calculated based on the size of the receptive field, is 106 ms (refer to the Methods in the supplementary file for detailed calculations).

In addition to the EEG signals, age, sex, and MMSE score were incorporated into the classification modules, although the default model originally did not account for these three variables (Eq. S9–10). The classification modules are described by the following function:

$$\left(\mathbf{y}_{\text{disease}},\ \mathbf{y}_{\text{subject}}\right)^{\mathrm{T}} = \mathrm{f}(\mathbf{X},\ \mathbf{a},\ \mathbf{s}, \mathbf{m}) \tag{1}$$

where $\mathbf{y}_{\text{disease}} \in \mathbb{R}^{n_{EEG}}$ (n_{EEG}: number of EEG segments in a task) and $\mathbf{y}_{\text{subject}} \in \mathbb{R}^{n_s}$ (n_s: number of subjects in the training dataset) are the posterior probabilities for the disease-type and subject classification tasks, respectively; $\mathbf{X} \in \mathbb{R}^{n_{EEG} \times 1{,}000 \times 19}$ represents 2-s windows of bandpass filtered EEG signals (0.53–100 Hz); and $\mathbf{a} \in \mathbb{R}^{n_{EEG}}$, $\mathbf{s} \in \mathbb{R}^{n_{EEG}}$, and $\mathbf{m} \in \mathbb{R}^{n_{EEG}}$ are transformed age, sex, and MMSE score (Figure S4), respectively. Notably, the second output for the subject classification task was only used for model training to improve training efficiency. In addition, we evaluated the classification accuracy of MNet without using the subject classification module.

The mean of $\mathbf{y}_{\text{disease}}$ for each subject was calculated to determine the disease prediction for the subject as follows:

$$\mathbf{y}'_{\text{disease}}\ |S = \overline{\mathbf{y}_{\text{disease}}|\mathrm{S}} \tag{2}$$

where S represents a subject. The disease type with the maximum value of $\mathbf{y}'_{\text{disease}}\ |\mathrm{S}$ was defined as the model diagnosis.

Model training

The training parameters, including the model architecture parameters, hyperparameters, and optimization method parameters, were determined using the training and validation datasets from the first fold of the four-disease classification task (HV vs. AD vs. DLB vs. iNPH) at Hospital O. For each classification task (i.e., HV vs. AD, HV vs. DLB, HV vs. iNPH, HV vs. AD+DLB+iNPH, AD vs. DLB, DLB vs. iNPH, or HV vs. AD vs. DLB vs. iNPH; Table 1), the model was trained and evaluated independently using the corresponding data acquired at Hospital O (i.e., some of the data labeled HV [$n = 69$], AD [$n = 101$], DLB [$n = 59$], and iNPH [$n = 28$]; for instance, data labeled HV [$n = 69$] or AD [$n = 101$] for the HV vs. AD classification task). The data were partitioned into training, validation, and test datasets at a ratio of 0.6:0.2:0.2. Importantly, data obtained from individuals with MCI and data collected at Hospital K and Hospital N were not used for training and were included only in the test sets.

An undersampling strategy (Kubat & Matwin, 1997) was

implemented to ensure that an equal number of subjects from each disease category were represented in the training set. For each subject, $n_{EEG_per_sub}\times$EEG segments (4 s) were randomly sampled, allowing for overlap. At the beginning of every epoch, a 2-s EEG segment was sampled from every 4-s EEG segment. The number of EEG segments per subject ($n_{EEG\_per\_sub}$) was determined for each training dataset, with approximately 95% of subjects yielding $n_{EEG_per_sub}\times$2-s EEG samples without overlap, resulting in approximately 100 EEG segments per subject.

A multitask loss approach (Kendall et al., 2018) was used to combine the disease and subject classification tasks. The architecture included two fully connected layers consisting of 1024 and 256 neurons, and both layers had a dropout rate of 0.85. Training proceeded for up to 50 epochs, with early stopping criteria in place to prevent overfitting.

Model inference

During the prediction phase, only novel data were input into the model to avoid any potential data leakage. An ensemble model was employed when applicable.

During five-fold cross-validation, five different MNets were trained on overlapping subsets of data from Hospital O. These five models can be expressed as follows:

$$\left(\mathbf{y}_{\mathrm{i,disease}},\ \mathbf{y}_{\mathrm{i,subject}}\right)^{\mathrm{T}} = \mathrm{f_i}(\mathbf{X},\ \mathbf{a},\ \mathbf{s},\mathbf{m})(i=1,\ 2,\ 3,\ 4,\ \text{or}\ 5) \tag{3}$$

where i is the index of a fold. The ensemble model was determined by computing the mean of the posterior probabilities for the disease-type classification task:

$$\mathbf{y}_{\mathrm{ensemble,disease}} = \overline{\sum_{i=1}^{5}\mathbf{y}_{\mathrm{i,disease}}}. \tag{4}$$

As shown in Eq. (2), the mean posterior probability of EEG data segments for a subject was used to obtain predictions from the ensemble model at the subject level:

$$\mathbf{y}^{'}_{\mathrm{ensemble,disease}}\ |S = \overline{\mathbf{y}_{\mathrm{ensemble,disease}}|\mathrm{S}}. \tag{5}$$

Statistical evaluation

The accuracy of the classification task was assessed using the balanced accuracy (bACC) and the area under the receiver-operating characteristic curve (ROC AUC; macro average). ROC curves were drawn in a one-versus-rest manner. Scalar metrics, the bACC and ROC AUC, precision, recall (= sensitivity), F1-score, and specificity, are expressed as the mean ± SD and were calculated using the scikit-learn package (Abraham et al., 2014). Kruskal–Wallis and Pearson's chi-squared tests were conducted using the pingouin package (Vallat, 2018). The Kolmogorov–Smirnov test was executed using the scipy.stats package (Virtanen et al., 2020).

Frequency band-stopping experiments

In the frequency band-stopping experiments, portions of the six frequency bands (delta, 0.5–4 Hz; theta, 4–8 Hz; low-alpha, 8–10 Hz; high-alpha, 10–13 Hz; beta, 13–32 Hz; and gamma, 32–75 Hz) were attenuated before being fed into the model. Band stopping was conducted only during the prediction phase, and all 2^6 $(=64)$ patterns of band attenuation were evaluated.

Channel masking experiments

In the channel-masking experiments, the signals of a subset of the 19 channels were substituted with random values drawn from a standard Gaussian distribution. This channel masking was performed only during the prediction phase. From the total of 2^{19} $(524,288)$ possible patterns, 0.1% (5243) were evaluated due to computational constraints.

Feature importance as indicated by Shapley values

The feature importance of our models was estimated as follows. First, by conducting frequency band-stopping or channel-masking experiments, we determined the bACC ($\mathbf{y}$) according to which elements were intact or mitigated (e.g., 1 or 0; $\mathbf{Z}$) from the original EEG signals ($\mathbf{X}$):

$$\mathrm{f_{MNet}}(\mathbf{X}|\mathbf{Z}) \overset{\text{yields}}{\rightarrow} \mathbf{y} \tag{6}$$

where $\mathrm{f_{MNet}}(\cdot)$ is our model, MNet.

To calculate the Shapley value (Štrumbelj & Kononenko, 2014), we approximated our model as an XGBoost regression model ($\boldsymbol{y} = f_{xgboost}(\mathbf{Z})$). Then, based on the SHAP package (Lundberg & Lee, 2017), the absolute Shapley values for each condition ($\mathbf{Z}$) were determined.

Data availability

The source code is available at https://github.com/yanagisawa-lab/eeg-dementia-classification.

Role of the funding source

The funders of the study had no role in the study design, data collection, data analysis, data interpretation, or writing of the report.

Results

Classification of healthy volunteers and dementia patients across multiple hospitals

MNet was trained on EEG data from Hospital O to classify individuals into the categories of HV, AD, DLB, or iNPH (Tables S3 and S4). The bACC and ROC AUC values in the classification of healthy volunteers (HVs) and dementia patients (AD, DLB, or iNPH) in the Hospital O dataset were 0.927 ± 0.015 and 0.987 ± 0.010, respectively (Figure 1; refer to Figure S5 for representative EEG traces classified as HVs and patients with AD, DLB, or iNPH using the trained MNet, and Figure S6 for the relationship between the classification score (bACC) and the number of segments). When applying the trained MNet to EEG data from Hospitals K and N, the bACC and ROC AUC values were 0.805 and 0.895 (Hospital K) and 0.920 and N/A (not available) (Hospital N), respectively (Fig. 1, Tables S5–S8).

Similarly, the model successfully classified the EEG data in the following tasks: HV vs. AD (bACC and ROC AUC values; Hospital O, 0.934 ± 0.014 and 0.986 ± 0.004; Hospital K, 0.778 and 0.904; and Hospital N, 0.950 and N/A), HV vs. DLB (Hospital O, 0.951 ± 0.071 and 0.993 ± 0.016; Hospital K, 0.833 and 0.967; Hospital N, 1.000 and N/A) and HV vs. iNPH (Hospital O, 0.890 ± 0.069 and 0.982 ± 0.017; Hospital K, 0.833 and 0.978; Hospital N, 1.000 and N/A) (Fig. 1).

Additionally, classification results using a support vector machine (SVM) classifier based on the amplitudes of the six frequency bands (delta, 0.5–4 Hz; theta, 4–8 Hz; low alpha, 8–10 Hz; high alpha, 10–13 Hz; beta, 13–32 Hz; and gamma, 32–75 Hz) are reported in Table S9. Furthermore, the classification accuracy of the MNet models without the subject identification module is summarized in Table S10. Notably, the power spectrum density (PSD) values of the EEG data significantly differed among frequency bands, disease types, and three hospitals (Figure S7; $ps < 0.001$, a multivariable linear model, Table S11). Even when accounting for differences in data acquisition sites, the classification results exceeded chance levels not only for the EEG data from Hospital O, which were used to train MNet, but also for the EEG data from Hospitals N and K, which were not used during MNet training (refer to Figure S8 for the activation maps of MNet during the processing

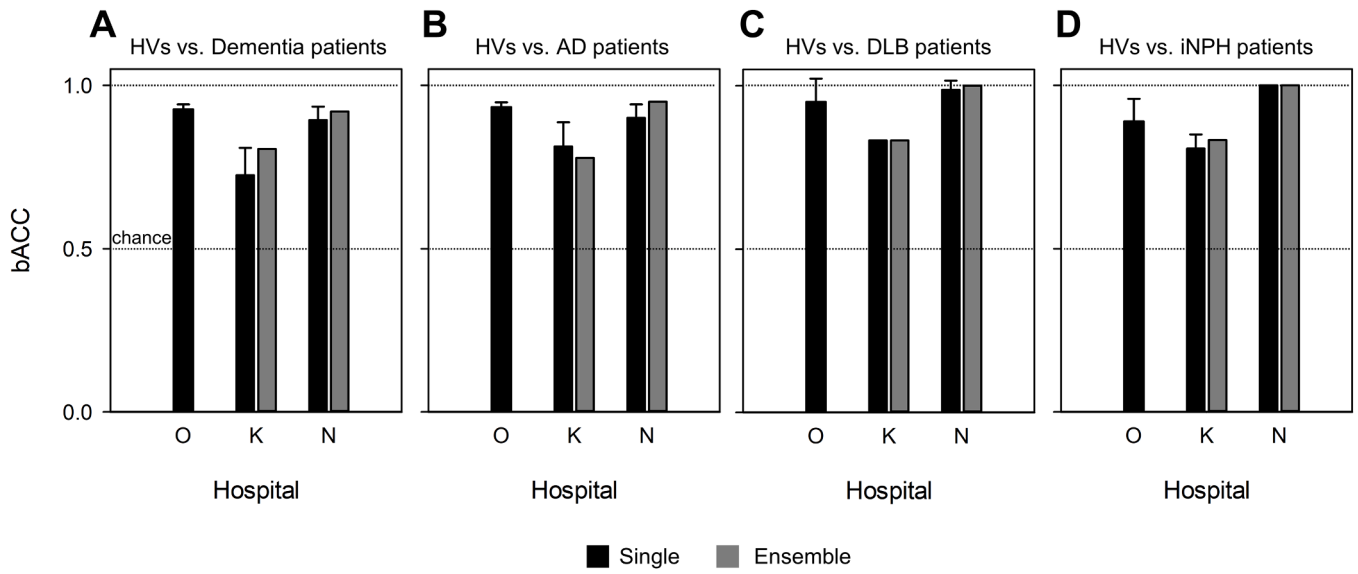

Fig. 1. bACC values of HVs vs. patients with dementia (subtypes) across multiple hospitals.
The bACC values of HVs vs. patients with dementia (subtypes: AD, DLB, and iNPH) (***A***), HVs vs. AD patients (***B***), HVs vs. DLB patients (***C***), and HVs vs. iNPH patients (***D***) at Hospitals O (internal), K and N (external). Abbreviations: HVs, healthy volunteers; AD, Alzheimer's disease; DLB, dementia with Lewy bodies; iNPH, idiopathic normal-pressure hydrocephalus; bACC, balanced accuracy; Single, individual models of the fivefold cross-validation (CV) method; Ensemble, combined model from the fivefold CV method Eqs. (4)–(5).

of EEG data acquired at Hospitals K and N).

Effects of age, sex, and MMSE score on the classification results

We evaluated the effects of age, sex, and MMSE score on the results of the classification task involving HVs vs. patients with dementia (AD, DLB, or iNPH) at Hospital O. Among HVs, the age of subjects correctly classified as HVs (true positives; median [IQR] = 64 [11] years) was younger than that of those incorrectly classified as HVs (false positives; 73 [3.5] years) (Fig. 2A & Table S12; $p = 0.011$, $D = 0.501$, Kolmogorov–Smirnov test, number of true positives [false negatives] = 55 [14]). However, among patients with dementia, the age of correctly predicted patients (true positives) did not significantly differ from that of incorrectly predicted patients (false negatives) (Table S12; $p = 0.313$, $D = 0.054$, Kolmogorov–Smirnov test, number of correct [incorrect] predictions = 169 [19]). These significant differences in age, sex, and MMSE score were observed in only three cases among all 63 classification results (Figure S9 & Table S12).

In addition, we incorporated age, sex, and MMSE score into the MNet classification module following the transformation of these variables (refer to Eq. S1–4, Figure S4, and Table S13). The combined model yielded a bACC of 0.519 ± 0.087 and an ROC AUC of 0.826 ± 0.045, which were lower than the EEG-only model's metrics (bACC: 0.647 ± 0.126, ROC AUC: 0.849 ± 0.057). This suggests that the inclusion of these variables might result in overfitting, possibly due to the limited sample size.

Identification of underlying pathological conditions associated with dementia

We then classified patients with dementia or MCI into subgroups of AD, DLB, and iNPH (Fig. 3). For the cohort from Hospital O, the MNet classifier exhibited bACC and ROC AUC values of 0.572 ± 0.098 and 0.745 ± 0.072 within the dementia group, respectively. The MNet

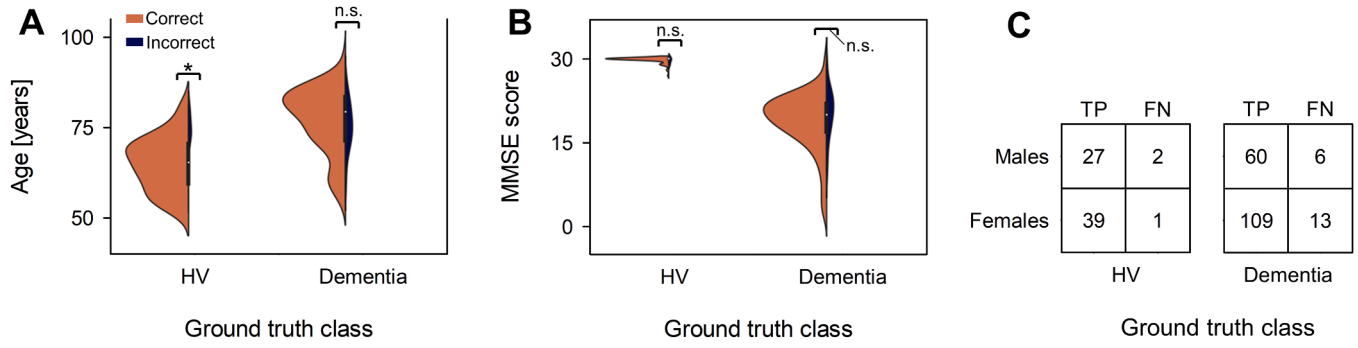

Fig. 2. Negligible effects of sampling bias on classification.
The results of the HV vs. dementia (AD, DLB, or iNPH) task at Hospital O (internal) are shown. ***A***. Age distributions of correct (TNs for HV and TPs for dementia; tan) and incorrect (FPs for HV and FNs for dementia; *navy*) predictions. Note that in the HV group, TNs (median [IQR] = 64 [11] years) and FPs (73 [3.5] years) had different distributions ($p = 0.011$, $D = 0.501$, Kolmogorov–Smirnov test, number of correct [incorrect] predictions = 66 [3]). As can be observed in Panel A, even if the population were classified only by age (by drawing any horizontal line), the classification accuracy (the proportion of the tan area) would not increase. ***B***. Distributions of MMSE scores for correct and incorrect predictions. ***C***. Cross-tabulation of sex distributions for correct and incorrect predictions. There were no significant deviations from the expected ratios according to Pearson's chi-squared test. Abbreviations: TN, true negative; TP, true positive; FP, false positive; FN, false negative.

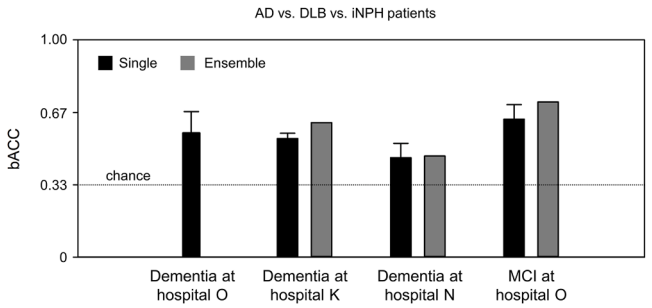

Fig. 3. bACC values of dementia subtypes for identification tasks across multiple sites.
The panel shows the bACC values for the AD vs. DLB vs. iNPH task among the dementia or MCI patients at Hospitals O (internal), K and N (external). Abbreviations: AD, Alzheimer's disease; DLB, dementia with Lewy bodies; iNPH, idiopathic normal-pressure hydrocephalus; bACC, balanced accuracy; Single, individual models of the fivefold cross-validation (CV) method; Ensemble, combined model from the fivefold CV method Eqs. (4)–(5).

classifier was also able to categorize patients at Hospital K and Hospital N, achieving bACC values of 0.619 and 0.508 and ROC AUC values of 0.782 and 0.710, respectively. Notably, when the MNet classifier was trained with the dementia group instead of the MCI group, it predicted the underlying pathologies of the MCI group at Hospital O, the bACC and ROC AUC values were 0.715 and 0.847, respectively (Table S14 & S15).

Feature importance analysis

We computed the significance of the frequency bands and electrodes needed by MNet to accomplish the tasks of classifying the four disease types. Initially, we attenuated the six frequency band components with band-stop filters and then observed the input−output responses. Based on these results (Table S16), the Shapley value (Štrumbelj & Kononenko, 2014) was calculated (Fig. 4A) to assess feature importance. Among the six frequency bands, the beta band was identified as the most valuable. Subsequently, we randomly masked several of the 19 channels and calculated the Shapley value (Fig. 4B); this method indicated that O1–A1 was the most important electrode montage.

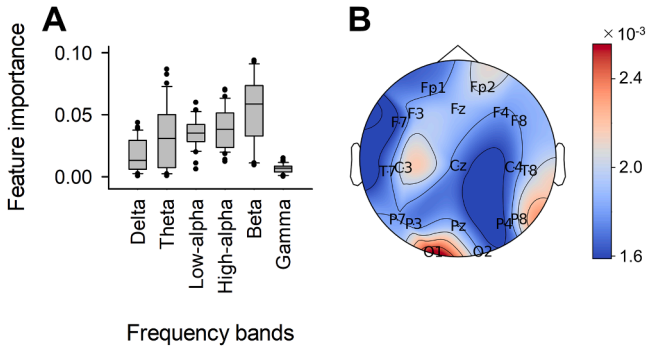

Fig. 4. Feature importance of frequency bands and channels.
A. Feature importance (Shapley value) of frequency bands calculated for the four-class classification task by attenuating band components with bandpass filters. The beta band was the most important band for successful model classification. The frequency bands were defined as follows: delta, 0.5–4 Hz; theta, 4–8 Hz; low alpha, 8–10 Hz; high alpha, 10–13 Hz; beta, 13–32 Hz; and gamma, 32–75 Hz. ***B***. The feature importance (Shapley value) of channels, calculated by masking signals from some channels, is visualized as a topographical map showing the median values. The O1–A1 electrode montage was the most important for classification in the model.

Discussion

Our findings effectively validated MNet's proficiency in classifying HVs and differentiating dementia subtypes (AD, DLB, and iNPH) using EEG data, even surmounting interinstitutional variability (Figure S7 & Table S11). Despite the dataset presenting biases in age, sex, and MMSE scores (Table 1 & Figure S1), their impact on the classification accuracy was negligible (Fig. 2 & Table S12). Moreover, MNet's ability to detect the pathology associated with dementia or MCI was demonstrated (Fig. 3). Notably, for MNet classification, the beta band (13–32 Hz) emerged as the most significant frequency band, and the O1–A1 electrode pair proved to be the most crucial among the standard 10–20 system electrodes (Fig. 4).

While previous studies have shown high accuracy for dementia classification based on EEG data, they often overlook the challenge of interinstitutional reproducibility, an issue our study addressed with an externally validated model. For example, HVs and patients with AD or DLB were previously classified with bACC scores of approximately 0.9 (Sánchez-Reyes et al., 2021; Ieracitano et al., 2020; Chatzikonstantinou et al., 2021), aligning with our results. However, such studies typically validate models with data from a single center, neglecting the issue of interinstitutional reproducibility, which could lead to overestimated performance due to unaccounted site-specific effects (Yamashita et al., 2019). Bonanni et al. (2016) highlighted this by demonstrating a drop in accuracy from 100% on an internal dataset to 64% and 90% on external datasets. Our findings further underscore this issue — substantial differences in the PSD of EEG signals recorded at the three institutions were observed (Figure S7 & Table S11). To address this, we applied our internally validated model to external datasets, which, notably, showed bACC values consistent with those of the model on data from the original site. For example, the bACC for differentiating HVs from AD+DLB+iNPH achieved scores of 0.927 ± 0.015 at Hospital O (Table S3), while at Hospital K, the bACC was 0.725 ± 0.084, improving to 0.805 with the ensemble model (Table S5). At Hospital N, the bACC reached 0.903 ± 0.033, rising to 0.914 with the ensemble model (Table S7), even given the notable intersite variability. (Figure S7 & Table S11).

MNet's design, which was tailored to the duration of EEG microstates, may underpin its high classification accuracy and allow it to capture the transient neural dynamics disrupted in neurological disorders. EEG microstates are transient, quasistable phases of brain activity that typically last from 60 to 120 milliseconds and are essential for understanding various neurological disorders (Mishra et al., 2020). MNet's temporal receptive field size, calculated to be approximately 106 milliseconds (as detailed in our supplementary Methods), falls within the critical range for EEG microstate duration. This enables MNet to effectively detect spatial patterns corresponding to EEG microstates, which are pivotal yet ephemeral snapshots of neural dynamics often disrupted in brain diseases.

The challenge of distinguishing between AD and DLB in patients is reflected in the lower bACCs of MNet for these subtypes, suggesting an overlap in pathological features. Specifically, MNet showed relatively low performance in classifying AD and DLB with bACCs of approximately 0.7 across different institutions (Tables S4 & S6). Although this performance is on par with that reported by Bonanni et al. (2016) and Engedal et al. (2015), it remains significantly lower than the discrimination performance achieved between HVs and dementia patients. One contributing factor to this challenge could be the coexistence of AD and DLB pathologies within patients; Dugger et al. (2014) reported that 89% of individuals with neuropathologically confirmed DLB also presented with AD pathology. Even with our stringent selection criteria for patients who were each exclusively diagnosed with one subtype, the overlapping nature of these conditions presents a substantial obstacle to accurate classification. Nonetheless, the model's consistent accuracy, comparable to results from contemporary studies, attests to its potential clinical value in distinguishing between these closely related dementia

subtypes.

Our study highlights MNet's capabilities for predicting MCI pathologies (Fig. 3), suggesting that EEG features of MCI may mirror those found in more advanced dementia stages. These observations are partially consistent with those of Bonanni et al. (2015), namely, that there were common characteristics between patients with MCI progressing to DLB and established DLB cases. The integration of EEG and machine learning methods could, therefore, be pivotal in detecting dementia in their early periods and identifying the pathologies shared by dementia and MCI.

Feature importance analysis within our study reveals MNet's emphasis on beta band frequency and the O1–A1 electrode pair, aligning with known neurological distinctions of dementia subtypes. For instance, regarding the frequency bands, our result is consistent with some previous reports demonstrating that beta power was an important variable for characterizing patients with AD, DLB, and iNPH (Aoki et al., 2015; Dauwan et al., 2016), although numerous studies have demonstrated the importance of other frequency bands as well. Notably, the importance of frequency bands is influenced by the model characteristics and classification tasks. First, MNet tends to learn higher frequency components because its learnable filters have short temporal windows. Second, adjacent frequency bands are highly correlated, and hence, a machine-learning model can satisfy one of those variables — the multicollinearity problem. Indeed, the bACC tended to decrease when the low-alpha, high-alpha, and beta bands were all attenuated using band-stop filters (Table S16). On the other hand, regarding electrode location, differences in the EEG data from the O1–A1 electrode pair might help characterize dementia subtypes (Fig. 4B). Compared with AD patients, patients with DLB experience visual hallucinations and show deficits in the occipital regions (Mak et al., 2014). Moreover, it was shown that iNPH patients have decreased alpha power in the occipital lobe (Aoki et al., 2019). Our successful classification suggests that MNet succeeded in learning these characteristics of the dementia subtypes.

MNet's versatility extends to various EEG recording systems, suggesting that with appropriate calibration, its utility is not limited to a single device or referencing standard. In fact, in this study, the data were recorded using three different devices: (i) EEG-1200 (Nihon Koden, Tokyo, Japan) at Hospitals O and K, and (ii) EEG-1260 (Nihon Koden, Tokyo, Japan) and (iii) EEG-1290 (Nihon Koden, Tokyo, Japan) at Hospital N. In the case of alternative referencing systems, such as those incorporating FCz, rereferencing could reconcile any discrepancies with our bipolar settings. Additionally, transfer learning techniques (Weiss et al., 2016) could effectively optimize the computations of MNet for various channel settings. Even considering these factors, the MNet model trained in this study might still be applicable to diverse EEG recording devices and reference systems, albeit with additional calibration steps.

The integration of MNet into existing diagnostic frameworks could enhance the precision and timeliness of dementia diagnoses. Leveraging its high classification accuracy, MNet can serve as an adjunctive, noninvasive screening modality that synergizes with conventional cognitive assessments and imaging. Early detection, pivotal for mitigating disease progression and optimizing patient outcomes, may be more attainable by applying MNet. Furthermore, MNet's nuanced capability to discern between dementia subtypes fosters a tailored approach to management, allowing for interventions that are finely tuned to individual pathologies. Additionally, the elucidation of shared EEG features between MCI and dementia by MNet could be instrumental in monitoring disease trajectory and prognosticating conversion to dementia, thereby informing clinical intervention strategies at critical junctures.

Nonetheless, the study is not without limitations. First, the 'black box' nature of the MNet model complicates its interpretation, particularly when analyzing complex neural activity signals such as beta-band oscillations in relation to MCI and dementia. Enhancing the model's transparency is an ongoing challenge. Second, MNet currently overlooks the dynamics of neural microstate sequences, which are crucial for cognitive function. Introducing recurrent architectures or attention mechanisms in future iterations may improve the modeling of these temporal dynamics. Furthermore, while our cross-institutional approach establishes a foundation for effective data use, the variable progression of cognitive decline in MCI and dementia warrants longitudinal studies to validate the model's long-term predictive capacity. Addressing these concerns will be pivotal in advancing MNet's utility and in elucidating the neurophysiological underpinnings of cognitive impairments.

Conclusions

In conclusion, our study confirms that MNet effectively classifies dementia and its subtypes across different cognitive levels and institutions. Its robust performance endorses its potential for clinical adoption as a noninvasive, cost-effective, and widely accessible screening tool for early dementia detection. Furthermore, the common EEG features identified across MCI and dementia establish a basis for further investigation into their underlying pathologies.

Declaration of generative AI and AI-assisted technologies in the writing process

During the preparation of this work the authors used ChatGPT 4, provided by OpenAI, in order to improve readability and language. After using this tool, the authors reviewed and edited the content as needed and take full responsibility for the content of the publication.

CRediT authorship contribution statement

Yusuke Watanabe: Formal analysis, Writing – original draft, Writing – review & editing. **Yuki Miyazaki:** Supervision, Writing – review & editing. **Masahiro Hata:** Conceptualization, Supervision, Writing – review & editing. **Ryohei Fukuma:** Supervision, Writing – review & editing. **Yasunori Aoki:** Writing – review & editing. **Hiroaki Kazui:** Writing – review & editing. **Toshihiko Araki:** Writing – review & editing. **Daiki Taomoto:** Writing – review & editing. **Yuto Satake:** Writing – review & editing. **Takashi Suehiro:** Writing – review & editing. **Shunsuke Sato:** Writing – review & editing. **Hideki Kanemoto:** Writing – review & editing. **Kenji Yoshiyama:** Writing – review & editing. **Ryouhei Ishii:** Writing – review & editing. **Tatsuya Harada:** Writing – review & editing. **Haruhiko Kishima:** Writing – review & editing. **Manabu Ikeda:** Conceptualization, Writing – original draft. **Takufumi Yanagisawa:** Conceptualization, Writing – original draft, Writing – review & editing.

Declaration of Competing Interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests:

Takufumi Yanagisawa reports financial support was provided by Japan Agency for Medical Research and Development. Takufumi Yanagisawa reports financial support was provided by Japan Science and Technology Agency. Takufumi Yanagisawa reports financial support was provided by Japan Society for the Promotion of Science. Takufumi Yanagisawa reports financial support was provided by Council of State Secretariats for Science Technology and Innovation. Takufumi Yanagisawa reports financial support was provided by National Institute of Biomedical Innovation Health and Nutrition. Takufumi Yanagisawa reports a relationship with PGV inc that includes: consulting or advisory. Takufumi Yanagisawa has patent #Information processing device, determination method, and determination program (WO 2020/218,013 A1) pending to Osaka University. The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Funding

This research was funded by a grant from the Japan Agency for Medical Research and Development (AMED) (JP19de0107001). This work was also supported in part by the Japan Science and Technology Agency (JST), Core Research for Evolutional Science and Technology (JPMJCR18A5), the Exploratory Research for Advanced Technology (JPMJER1801), and the Moonshot R&D-MILLENNIA Program (JPMJMS2012), as well as Grants-in-Aid for Scientific Research from the Japan Society for the Promotion of Science (JSPS) (JP20H05705, JP18H04085 and JP18H05522); AMED (JP19dm0307103, 19dm0207070h0001 and JP19dm0307008); the Council for Science, Technology and Innovation (CSTI); the Cross-Ministerial Strategic Innovation Promotion Program (SIP): and the "Innovative AI Hospital System" (Funding Agency: National Institute of Biomedical Innovation, Health and Nutrition (NIBIOHN)).

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.neunet.2023.12.009.

References

Abraham, A., Pedregosa, F., Eickenberg, M., Gervais, P., Mueller, A., Kossaifi, J., Gramfort, A., Thirion, B., & Varoquaux, G. (2014). Machine learning for neuroimaging with scikit-learn. *Frontiers in Neuroinformatics, 8*. https://www.frontiersin.org/articles/10.3389/fninf.2014.00014.

Aoe, J., Fukuma, R., Yanagisawa, T., Harada, T., Tanaka, M., Kobayashi, M., Inoue, Y., Yamamoto, S., Ohnishi, Y., & Kishima, H. (2019). Automatic diagnosis of neurological diseases using MEG signals with a deep neural network. *Scientific Reports, 9*(1), 1. https://doi.org/10.1038/s41598-019-41500-x

Aoki, Y., Kazui, H., Pascual-Marqui, R. D., Ishii, R., Yoshiyama, K., Kanemoto, H., Suzuki, Y., Sato, S., Azuma, S., Suehiro, T., Matsumoto, T., Hata, M., Canuet, L., Iwase, M., & Ikeda, M. (2019). EEG resting-state networks responsible for gait disturbance features in idiopathic normal pressure hydrocephalus. *Clinical EEG and Neuroscience, 50*(3), 210–218. https://doi.org/10.1177/1550059418812156

Aoki, Y., Kazui, H., Tanaka, T., Ishii, R., Wada, T., Ikeda, S., Hata, M., Canuet, L., Katsimichas, T., Musha, T., Matsuzaki, H., Imajo, K., Kanemoto, H., Yoshida, T., Nomura, K., Yoshiyama, K., Iwase, M., & Takeda, M. (2015). Noninvasive prediction of shunt operation outcome in idiopathic normal pressure hydrocephalus. *Scientific Reports, 5*, 7775. https://doi.org/10.1038/srep07775

Arvanitakis, Z., Shah, R. C., & Bennett, D. A. (2019). Diagnosis and management of dementia: Review. *JAMA, 322*(16), 1589–1599. https://doi.org/10.1001/jama.2019.4782

Barker, W. W., Luis, C. A., Kashuba, A., Luis, M., Harwood, D. G., Loewenstein, D., Waters, C., Jimison, P., Shepherd, E., Sevush, S., Graff-Radford, N., Newland, D., Todd, M., Miller, B., Gold, M., Heilman, K., Doty, L., Goodman, I., Robinson, B., & Duara, R. (2002). Relative frequencies of Alzheimer disease, Lewy body, vascular and frontotemporal dementia, and hippocampal sclerosis in the State of Florida Brain Bank. *Alzheimer Disease and Associated Disorders, 16*(4), 203–212. https://doi.org/10.1097/00002093-200210000-00001

Bonanni, L., Franciotti, R., Nobili, F., Kramberger, M. G., Taylor, J. P., Garcia-Ptacek, S., Falasca, N. W., Famá, F., Cromarty, R., Onofrj, M., Aarsland, D., & E-DLB study group. (2016). EEG markers of dementia with lewy bodies: A multicenter cohort study. *Journal of Alzheimer's Disease: JAD, 54*(4), 1649–1657. https://doi.org/10.3233/JAD-160435

Bonanni, L., Perfetti, B., Bifolchetti, S., Taylor, J. P., Franciotti, R., Parnetti, L., Thomas, A., & Onofrj, M. (2015). Quantitative electroencephalogram utility in predicting conversion of mild cognitive impairment to dementia with Lewy bodies. *Neurobiology of Aging, 36*(1), 434–445. https://doi.org/10.1016/j.neurobiolaging.2014.07.009

Chatzikonstantinou, S., McKenna, J., Karantali, E., Petridis, F., Kazis, D., & Mavroudis, I. (2021). Electroencephalogram in dementia with Lewy bodies: A systematic review. *Aging Clinical and Experimental Research, 33*(5), 1197–1208. https://doi.org/10.1007/s40520-020-01576-2

Dauwan, M., van der Zande, J. J., van Dellen, E., Sommer, I. E. C., Scheltens, P., Lemstra, A. W., & Stam, C. J. (2016). Random forest to differentiate dementia with Lewy bodies from Alzheimer's disease. Alzheimer's & Dementia: Diagnosis. *Assessment & Disease Monitoring, 4*(1), 99–106. https://doi.org/10.1016/j.dadm.2016.07.003

Dugger, B. N., Adler, C. H., Shill, H. A., Caviness, J., Jacobson, S., Driver-Dunckley, E., Beach, T. G., & Arizona Parkinson's Disease Consortium. (2014). Concomitant pathologies among a spectrum of parkinsonian disorders. *Parkinsonism & Related Disorders, 20*(5), 525–529. https://doi.org/10.1016/j.parkreldis.2014.02.012

Engedal, K., Snaedal, J., Hoegh, P., Jelic, V., Andersen, B. B., Naik, M., Wahlund, L. O., & Oeksengaard, A. R. (2015). Quantitative EEG applying the statistical recognition pattern method: a useful tool in dementia diagnostic workup. *Dementia and Geriatric Cognitive Disorders, 40*(1–2), 1–12. https://doi.org/10.1159/000381016

Folstein, M. F., Folstein, S. E., & McHugh, P. R. (1975). Mini-mental state". A practical method for grading the cognitive state of patients for the clinician. *Journal of Psychiatric Research, 12*(3), 189–198. https://doi.org/10.1016/0022-3956(75)90026-6

Garn, H., Coronel, C., Waser, M., Caravias, G., & Ransmayr, G. (2017). Differential diagnosis between patients with probable Alzheimer's disease, Parkinson's disease dementia, or dementia with Lewy bodies and frontotemporal dementia, behavioral variant, using quantitative electroencephalographic features. *Journal of Neural Transmission, 124*(5), 569–581. https://doi.org/10.1007/s00702-017-1699-6

Hata, M., Watanabe, Y., Tanaka, T., Awata, K., Miyazaki, Y., Fukuma, R., Taomoto, D., Satake, Y., Suehiro, T., Kanemoto, H., Yoshiyama, K., Iwase, M., Ikeda, S., Nishida, K., Takekita, Y., Yoshimura, M., Ishii, R., Kazui, H., Harada, T., & Yanagisawa, T. (2023). Precise discrimination for multiple etiologies of dementia cases based on deep learning with electroencephalography. *Neuropsychobiology*, 1–10. https://doi.org/10.1159/000528439

Ieracitano, C., Mammone, N., Hussain, A., & Morabito, F. C. (2020). A novel multi-modal machine learning based approach for automatic classification of EEG recordings in dementia. *Neural Networks, 123*, 176–190. https://doi.org/10.1016/j.neunet.2019.12.006

Kendall, A., Gal, Y., & Cipolla, R. (2018). Multi-Task Learning Using Uncertainty to Weigh Losses for Scene Geometry and Semantics. 7482–7491. https://openaccess.thecvf.com/content_cvpr_2018/html/Kendall_Multi-Task_Learning_Using_CVPR_2018_paper.html.

Kubat, M., & Matwin, S. (1997). Addressing the curse of imbalanced training sets: One-sided selection. *Icml, 97*(1), 179.

Leuzy, A., Smith, R., Cullen, N. C., Strandberg, O., Vogel, J. W., Binette, A. P., Borroni, E., Janelidze, S., Ohlsson, T., Jögi, J., Ossenkoppele, R., Palmqvist, S., Mattsson-Carlgren, N., Klein, G., Stomrud, E., & Hansson, O. (2022). Biomarker-Based prediction of longitudinal Tau positron emission tomography in Alzheimer disease. *JAMA Neurology, 79*(2), 149–158. https://doi.org/10.1001/jamaneurol.2021.4654

Lundberg, S. M., & Lee, S. I. (2017). A unified approach to interpreting model predictions. *Advances in Neural Information Processing Systems, 30*. https://proceedings.neurips.cc/paper/2017/hash/8a20a8621978632d76c43dfd28b67767-Abstract.html.

Mak, E., Su, L., Williams, G. B., & O'Brien, J. T (2014). Neuroimaging characteristics of dementia with Lewy bodies. *Alzheimer's Research & Therapy, 6*(2), 18. https://doi.org/10.1186/alzrt248

McKeith, I. G., Boeve, B. F., Dickson, D. W., Halliday, G., Taylor, J. P., Weintraub, D., Aarsland, D., Galvin, J., Attems, J., Ballard, C. G., Bayston, A., Beach, T. G., Blanc, F., Bohnen, N., Bonanni, L., Bras, J., Brundin, P., Burn, D., Chen-Plotkin, A., & Kosaka, K. (2017). Diagnosis and management of dementia with Lewy bodies: Fourth consensus report of the DLB consortium. *Neurology, 89*(1), 88–100. https://doi.org/10.1212/WNL.0000000000004058

Micchia, K., Formica, C., De Salvo, S., Muscarà, N., Bramanti, P., Caminiti, F., Marino, S., & Corallo, F. (2022). Normal pressure hydrocephalus. *Medicine, 101*(9), e28922. https://doi.org/10.1097/MD.0000000000028922

Mishra, A., Englitz, B., & Cohen, M. X. (2020). EEG microstates as a continuous phenomenon. *NeuroImage, 208*, Article 116454. https://doi.org/10.1016/j.neuroimage.2019.116454

Morris, J. C. (1993). The Clinical Dementia Rating (CDR): Current version and scoring rules. *Neurology, 43*(11). https://doi.org/10.1212/WNL.43.11.2412-a

Prince, M.J., Wimo, A., Guerchet, M.M., Ali, G.C., Wu, Y.T., & Prina, M. (2015). World Alzheimer Report 2015 - The Global Impact of Dementia: An analysis of prevalence, incidence, cost and trends. https://kclpure.kcl.ac.uk/portal/en/publications/world-alzheimer-report-2015–the-global-impact-of-dementia(ae525fda-1938-4892-8daa-a2222a672254).html.

Rossini, P. M., Buscema, M., Capriotti, M., Grossi, E., Rodriguez, G., Del Percio, C., & Babiloni, C. (2008). Is it possible to automatically distinguish resting EEG data of normal elderly vs. Mild cognitive impairment subjects with high degree of accuracy? *Clinical Neurophysiology, 119*(7), 1534–1545. https://doi.org/10.1016/j.clinph.2008.03.026

Sánchez-Reyes, L. M., Rodríguez-Reséndiz, J., Avecilla-Ramírez, G. N., García-Gomar, M. L., & Robles-Ocampo, J. B. (2021). Impact of EEG parameters detecting dementia diseases: A systematic review. *IEEE Access: Practical Innovations, Open Solutions, 9*, 78060–78074. https://doi.org/10.1109/ACCESS.2021.3083519

Štrumbelj, E., & Kononenko, I. (2014). Explaining prediction models and individual predictions with feature contributions. *Knowledge and Information Systems, 41*(3), 647–665. https://doi.org/10.1007/s10115-013-0679-x

Swanson, C. J., Zhang, Y., Dhadda, S., Wang, J., Kaplow, J., Lai, R. Y. K., Lannfelt, L., Bradley, H., Rabe, M., Koyama, A., Reyderman, L., Berry, D. A., Berry, S., Gordon, R., Kramer, L. D., & Cummings, J. L. (2021). A randomized, double-blind, phase 2b proof-of-concept clinical trial in early Alzheimer's disease with lecanemab, an anti-Aβ protofibril antibody. *Alzheimer's Research & Therapy, 13*(1), 80. https://doi.org/10.1186/s13195-021-00813-8

Vallat, R. (2018). Pingouin: Statistics in python. *Journal of Open Source Software, 3*(31), 1026. https://doi.org/10.21105/joss.01026

Virtanen, P., Gommers, R., Oliphant, T.E., Haberland, M., Reddy, T., Cournapeau, D., Burovski, E., Peterson, P., Weckesser, W., Bright, J., van der Walt, S.J., Brett, M.,

Wilson, J., Millman, K.J., Mayorov, N., Nelson, A.R.J., Jones, E., Kern, R., Larson, E., SciPy 1. 0 Contributors. (2020). SciPy 1.0: Fundamental algorithms for scientific computing in Python. Nature Methods, 17, 261–272. https://doi.org/10.1038/s41592-019-0686-2.

Weiss, K., Khoshgoftaar, T. M., & Wang, D. (2016). A survey of transfer learning. *Journal of Big Data, 3*(1), 9. https://doi.org/10.1186/s40537-016-0043-6

Williams, M. A., & Malm, J. (2016). Diagnosis and treatment of idiopathic normal pressure hydrocephalus. *Continuum: Lifelong Learning in Neurology, 22*(2 Dementia), 579–599. https://doi.org/10.1212/CON.0000000000000305

Yamashita, A., Yahata, N., Itahashi, T., Lisi, G., Yamada, T., Ichikawa, N., Takamura, M., Yoshihara, Y., Kunimatsu, A., Okada, N., Yamagata, H., Matsuo, K., Hashimoto, R., Okada, G., Sakai, Y., Morimoto, J., Narumoto, J., Shimada, Y., Kasai, K., & Imamizu, H. (2019). Harmonization of resting-state functional MRI data across multiple imaging sites via the separation of site differences into sampling bias and measurement bias. *PLOS Biology, 17*(4), Article e3000042. https://doi.org/10.1371/journal.pbio.3000042