



OPEN ACCESS

EDITED BY

Jinping Xu,
Chinese Academy of Sciences (CAS), China

REVIEWED BY

Bo Tan,
University of Electronic Science and
Technology of China, China
Yi Yang,
Beijing Tiantan Hospital Affiliated to Capital
Medical University, China

*CORRESPONDENCE

Jae-Sung Lim
✉ jaesunglim@amc.seoul.kr
Seung Wan Kang
✉ seungwkang@imedisync.com

[†]These authors have contributed equally to this work

RECEIVED 11 June 2023

ACCEPTED 11 September 2023

PUBLISHED 28 September 2023

CITATION

Lee M, Hong Y, An S, Park U, Shin J, Lee J, Oh MS, Lee B-C, Yu K-H, Lim J-S and Kang SW (2023) Machine learning-based prediction of post-stroke cognitive status using electroencephalography-derived brain network attributes.
Front. Aging Neurosci. 15:1238274.
doi: 10.3389/fnagi.2023.1238274

COPYRIGHT

© 2023 Lee, Hong, An, Park, Shin, Lee, Oh, Lee, Yu, Lim and Kang. This is an open-access article distributed under the terms of the [Creative Commons Attribution License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Machine learning-based prediction of post-stroke cognitive status using electroencephalography-derived brain network attributes

Minwoo Lee^{1†}, Yuseong Hong^{2†}, Sungsik An³, Ukeob Park², Jaekang Shin², Jeongjae Lee¹, Mi Sun Oh¹, Byung-Chul Lee¹, Kyung-Ho Yu¹, Jae-Sung Lim^{4*} and Seung Wan Kang^{2*}

¹Department of Neurology, Hallym University Sacred Heart Hospital, Hallym Neurological Institute, Hallym University College of Medicine, Anyang, Republic of Korea, ²iMedisync, Inc., Seoul, Republic of Korea, ³Department of Neurology, Hwahong Hospital, Suwon, Republic of Korea, ⁴Department of Neurology, Asan Medical Center, University of Ulsan College of Medicine, Seoul, Republic of Korea

Objectives: More than half of patients with acute ischemic stroke develop post-stroke cognitive impairment (PSCI), a significant barrier to future neurological recovery. Thus, predicting cognitive trajectories post-AIS is crucial. Our primary objective is to determine whether brain network properties from electroencephalography (EEG) can predict post-stroke cognitive function using machine learning approach.

Methods: We enrolled consecutive stroke patients who underwent both EEG during the acute stroke phase and cognitive assessments 3 months post-stroke. We preprocessed acute stroke EEG data to eliminate low-quality epochs, then performed independent component analysis and quantified network characteristics using iSyncBrain®. Cognitive function was evaluated using the Montreal cognitive assessment (MoCA). We initially categorized participants based on the lateralization of their lesions and then developed machine learning models to predict cognitive status in the left and right hemisphere lesion groups.

Results: Eighty-seven patients were included, and the accuracy of lesion laterality prediction using EEG attributes was 97.0%. In the left hemispheric lesion group, the network attributes of the theta band were significantly correlated with MoCA scores, and higher global efficiency, clustering coefficient, and lower characteristic path length were associated with higher MoCA scores. Most features related to cognitive scores were selected from the frontal lobe. The predictive powers (*R*-squared) were 0.76 and 0.65 for the left and right stroke groups, respectively.

Conclusion: Estimating EEG-based network properties in the acute phase of ischemic stroke through a machine learning model has a potential to predict cognitive outcomes after ischemic stroke.

KEYWORDS

ischemic stroke, cognition, electroencephalography, functional network, machine learning

Introduction

Stroke is the major cause of mortality and disability worldwide (Wolfe, 2000; Langhorne et al., 2011). Among disabilities, cognitive impairment after stroke exerts an important effect on patients, functional recovery, and long-term prognosis. Currently, clinicians evaluate the risk of post-stroke cognitive impairment using structural information, such as the location and size of the stroke lesion, from magnetic resonance imaging (MRI) or computed tomography (CT). An international large-scale consortium for lesion-symptom mapping research recently presented important findings on the location in the brain of lesions that cause cognitive impairment following stroke (Weaver et al., 2021b). However, these methods using structural brain imaging are based on the premise that brain function is selectively impaired in the localized area where the stroke lesion occurred. However, in practice, lesions at several sites might result in the same symptoms and signs, which is accounted for by the disruption of a shared neural network (Fox, 2018). As a result, it would be difficult to predict cognitive impairment after stroke based only on lesion location as identified in MRI, and it is necessary to evaluate the widespread effect of a lesion on the entire brain (Stinear, 2010; Carter et al., 2012). Tools for measuring functional connectivity, which evaluates not only a specific brain region's activity but also the interaction between different regions, may be essential in resolving this issue (Bressler and Menon, 2010; Stinear, 2010; Aerts et al., 2016).

Among the available modalities to explore the functional activity of the human brain, disconnectome studies using functional MRI (fMRI) and diffusion tensor imaging (DTI) are being actively investigated in post-stroke cognitive impairment studies (Biesbroek et al., 2021; Lim et al., 2021). However, fMRI and DTI are difficult to apply in stroke patients in clinical practice due to their high cost and vulnerability to motion artifacts. As an alternative, electroencephalography (EEG) is advantageous, as EEG can detect functional changes in the entire brain caused by lesions, and it is less expensive than MRI (Chen et al., 2013). Though the low spatial resolution of EEG is caused by the limited number of electrodes and by the effect of the electromagnetic field on volume conduction, several methods have been proposed to overcome these shortcomings. Standardized low-resolution brain electromagnetic tomography (sLORETA) is a mathematical tool for source localization to map real brain regions from recorded EEG signals (Jatoi et al., 2014). Imaginary coherence (iCOH) calculated from the imaginary part of coherence (Sanchez Bornot et al., 2018) also helps to overcome the low spatial resolution of EEG even though iCOH is not a direct source localization method. It can be used to determine the functional connectivity between two regions of interest (ROIs) and efficiently reduce volume conduction (Nolte et al., 2004).

Several studies have reported that quantitative analysis of EEG may be employed to predict prognosis after acute ischemic stroke. Specific frequency bands, power ratios between different frequency bands, and symmetric index were strongly correlated with patients' stroke severity index using the National Institutes of Health Stroke Scale (NIHSS) and with the functional status assessed by the modified Rankin Scale (Sheorajpanday et al., 2011; Finnigan and van Putten, 2013). Further, several QEEG indices, such as relative

theta frequency (Schleiger et al., 2017) or irregularity of spectral power of frontal lobes (Hadiyoso et al., 2022), predicted post-stroke cognitive impairment in the previous studies with limited explanatory power. As incorporating brain network attributes may provide more details about cognitive prognosis after stroke, QEEG using a machine learning approach to incorporate functional connectivity along with lesion characteristics would be a powerful tool that could reveal the network vulnerability for post-stroke cognitive impairment in stroke patients.

Therefore, we aimed to investigate which EEG-derived brain networks are associated with short-term cognitive status after stroke and to assess whether the brain network can accurately predict cognitive status after acute ischemic stroke using a machine learning approach.

Methods

Study population

Of the 1,959 consecutive ischemic stroke patients admitted between September 2016 and February 2020, 1,240 patients had an anterior circulation stroke and were admitted to the hospital within 7 days of symptom onset. Out of these, 223 underwent a 3-month MoCA. After excluding 18 patients with a pre-morbid cognitive impairment, as indicated by an Informant Questionnaire on Cognitive Decline in the Elderly (IQCODE) score (Lee et al., 2005) of ≥ 3.6 , a total of 87 patients who had an EEG during the acute stroke phase were selected for our study.

This study was approved by the Institutional Review Board of the Hallym University Sacred Heart Hospital and was not required to seek additional consent because of its retrospective nature and due to the minimal risk to participants.

Neuropsychological and clinical variables

The participants underwent neuropsychological tests, including the Korean version of the Mini-Mental State Examination (MMSE) and the Korean version of the Montreal Cognitive Assessment (MoCA) during admission (baseline) and after 3 months. A trained neuropsychologist performed the tests. Standardized scores for each test were calculated based on age, sex, and education-adjusted norm (Kang et al., 2009). Change values were calculated as "3-month test score - baseline test score."

From a stroke registry database (Bae et al., 2022), demographic factors, including age, sex, and years of education, were collected. Vascular risk factors, including history of hypertension, diabetes, hyperlipidemia, atrial fibrillation, smoking status, coronary heart disease, and previous stroke or transient ischemic attack, were also collected. The evaluation of stroke characteristics involved the assessment of initial stroke severity based on the NIHSS score and stroke subtypes according to the Trial of ORG 10172 in Acute Stroke Treatment (TOAST) classification. Pre-morbid functional status was evaluated using the modified Rankin scale. Time intervals between index stroke, EEG, and neuropsychological evaluations (baseline and follow-up) were also determined.

Brain imaging

Neuroimaging analysis was performed using an MRI acquired at the time of the stroke for medical purposes. The participants underwent 3.0-Tesla MRI scanning (Achieva, Philips Healthcare, Eindhoven, The Netherlands). The protocols consisted of fluid-attenuated inversion recovery imaging (FLAIR), axial T1- and T2-weighted spin echo, gradient-echo imaging, coronal T1-weighted spin echo imaging, and diffusion-weighted imaging (DWI). As for acute lesions, multiplicity, left hemisphere involvement, and cortical involvement of stroke lesions were investigated. As for chronic lesions, white matter hyperintensities (Fazekas scale), lacunes and cerebral microbleeds (presence and count), and medial temporal lobe atrophy (Scheltens visual scale) (Scheltens et al., 1992) were evaluated using the STRIVE criteria (Wardlaw et al., 2013).

EEG data acquisition

We retrospectively collected EEG data taken within 2 weeks of onset in patients with acute ischemic stroke. The EEG was recorded at rest using 19 channels of the international 10–20 system according to the laboratory's internal standardized guidelines. To enhance the data quality and mitigate artifact effects, the EEG data underwent preprocessing. In the initial stage, the signals were sampled at a rate of 250 Hz and subjected to bandpass filtering in the frequency range of 1~45.5 Hz. Further, notch filtering was applied to eliminate unwanted noise. Subsequent steps involved re-referencing using common average referencing (CAR), the rejection of bad epochs through artifact subspace reconstruction (ASR), and the removal of stationary noise using adaptive mixture independent component analysis (AMICA) (Mullen et al., 2015; Blum et al., 2019) in iSyncBrain[®]. Additionally, artifacts attributed to electromyogram (EMG) and electrooculogram (EOG) were eliminated to generate reliable QEEG data. All preprocessing steps, including sensor-level data handling and source-level data computation and extraction, were performed using the cloud-based EEG analysis platform, iSyncBrain[®] provided by iMediSync, Inc. Korea¹ (Subasi and Ismail, 2010). Each component was analyzed, and noise components, such as heartbeat and muscle behavior, were excluded in this process. The automatically cleaned data were inspected manually once again to filter out the unfiltered noise and check whether the cleaned data were analyzable or not.

After the signal processing, spectrum powers were measured from 19 channels and 8 frequency bands. The 8 frequency bands were delta (1–4 Hz), theta (4–8 Hz), alpha1 (8–10 Hz), alpha2 (10–12 Hz), beta1 (12–15 Hz), beta2 (15–20 Hz), beta3 (20–30 Hz), and gamma (30–45 Hz). All the power values were also transformed into their relative values, which is the ratio of a specific band power to the entire band power. Also, the power ratios between two spectral bands, namely, TAR (Theta/Alpha), TBR (Theta/Beta), TBR2 (Theta/Beta2), and DAR (Delta/Alpha), were calculated at each channel. The next step was source power estimation. Source cortical activity was mathematically estimated by using sLORETA (Pascual-Marqui,

2002). Subsequently, the spectrum powers were expanded from 19 channels to 68 cortical ROIs based on the Desikan–Killiany atlas (Desikan et al., 2006), providing us with more detail on brain activity and the source of spatial information. Following the Desikan–Killiany atlas, the 68 ROIs are arranged in frontal, temporal, parietal and occipital lobe. Even and odd numbers were used for the ROIs in the left and right hemispheres, respectively (Supplementary Table S1). Based on each source power, the functional connectivity between two ROIs was estimated using the iCOH method, which is calculated as

$$iCOH = Im(COH) = Im\left(\frac{S_{xy}(f)}{\sqrt{S_{xx}(f)S_{yy}(f)}}\right) \quad (1)$$

where $S_{xy}(f)$ is the cross-power spectral density of ROI x , y and $S_{xx}(f)$, $S_{yy}(f)$ are the power spectral density of ROI x , y . The iCOH method is robust in volume conduction so that we can separate the brain's functional connectivity from the electromagnetic field effect. Finally, the network was constructed using each ROI as a node and the iCOH between ROIs as an edge. Four network features (global efficiency, characteristic path length, clustering coefficient, and modularity) (Liu et al., 2017) were calculated based on the iCOH values. Considering that each node had the iCOH values for all other nodes, we retained only the top 25% of the iCOH values and used them to construct a network having only a quarter of the edges being connected. The detailed information and the number of network features used in this study are shown in Supplementary Table S2.

Prediction of a stroke lesion's laterality

First, we categorized patients into two groups based on lesion laterality for further analysis to uncover novel predictors independent of lesion laterality, as previous research has indicated that left-sidedness of the infarct site is a robust predictor of cognitive impairment (Weaver et al., 2021a). Based on prior studies showing that slow waves predominate close to the stroke (Sheorajpanday et al., 2011), we trained the machine learning model to classify the laterality of the infarct regions and used the classified results for the next step. According to the predictions of the machine learning method built using data from patients with clearly left- or right-sided lesions, individuals with both hemispheric lesions were divided into left or right laterality.

Prediction of post-stroke cognitive status

For each of the two groups dichotomized by the laterality of the lesion, we constructed a regression model with cognitive level as the outcome variable. Among various EEG features, including spectral power, power ratio, source power, connectivity, and network measures, we only utilized network variables as explanatory variables because our hypothesis was to predict cognitive prognosis from functional network characteristics and to reduce the possibility of overfitting by limiting the number of explanatory variables. We first calculated the correlation coefficient between the MoCA standardized score and each network feature

¹ <https://isyncbrain.com>

extracted from the EEG. For further analysis, we extracted only those network features with a correlation coefficient higher than 0.3. The extracted features were arranged in their order of importance as calculated using a machine learning model. The importance means how much each feature is used to predict the output variable in each node. It is calculated by the ratio between the importance of each feature to the summation of all the importance. In this model, the top 20 features were selected. In the final process, machine learning-based regression models for post-stroke MoCA standardized scores were generated using Ridge, Lasso, ElasticNet, support vector regression, and AdaBoost regression. The flow chart of the model including the prediction of the lesion's laterality is shown in Figure 1 and each model's explanatory power was calculated by R-squared as follows:

$$R^2 = 1 - \frac{SS_{res}}{SS_{tot}} = 1 - \frac{\sum (y_i - \hat{y}_i)^2}{\sum (y_i - \bar{y})^2} \quad (2)$$

where SS_{res} is the residual sum of squares, SS_{tot} is the total sum of squares, y_i is the actual value, \hat{y}_i is the predicted value, and \bar{y} is the mean value of the total actual values.

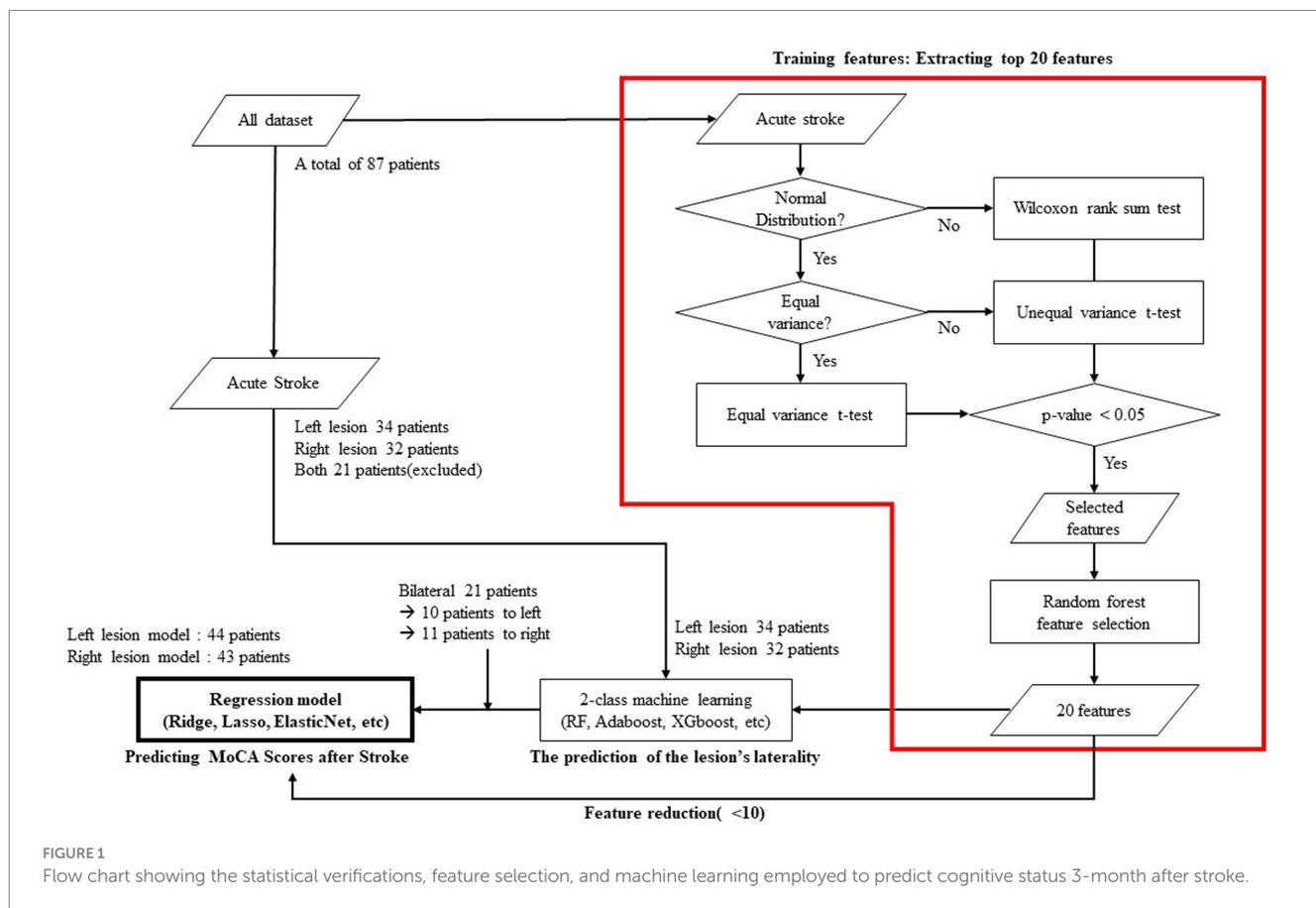
All statistical analyses were performed using R version 4.0.5, and statistical significance was set at $p < 0.05$.

Leave-one-out (LOO) validation technique

For the validation of our machine learning models, we employed the Leave-one-out (LOO) method. This approach is especially suited for limited datasets. Our dataset comprises data from 87 patients, and given the study's objective, separate models were developed for the left and right brain regions, further restricting the available training data. The LOO technique, a variant of k-fold cross-validation, involves using each data point in the dataset for both training and testing (Cawley and Talbot, 2004). Given the limited size of our training data, this method allows for an in-depth sensitivity analysis on individual data points, offering insights into the model's responsiveness to specific observations.

Sensitivity analysis

To bolster the robustness of our findings, we undertook additional machine learning analyses with a focus on minimizing potential overfitting by adjusting the number of incorporated features. Our approach involved: (1) Prioritizing features based on their effect size importance and (2) integrating alpha values from the top 5 features with beta values sourced from the bottom 15 features. Models were then selected based on their R-squared values. By employing a comprehensive strategy, which included



the selection of 3 to 5 features from the top quintile and varying numbers from the bottom 15 (with a cumulative count not exceeding 10 features), we were able to discern models that exhibited the optimal *R*-squared values.

Results

Baseline characteristics and stroke features of the study population

A total of 87 patients with acute stroke who underwent MRI, EEG, and the MoCA were enrolled in this study. The mean age of the patients was 65.4 years (SD 11.6), and 54 (61.4%) of them were men. The median scores for NIHSS and educational attainment were 3 (IQR 1–6) and 9.5 years (IQR 6–12), respectively (Table 1). The time intervals from stroke onset to the EEG and cognitive tests had a median of 4 days (IQR 2–6) and 97 days (IQR 90.0–103.0), respectively. The 3-month MoCA score had a median of 23 points (IQR 19–26).

Lesion laterality prediction

During the prediction of the lesion laterality process, 21 patients were excluded because they had infarct regions in both hemispheres. Thus, they were not labeled or used to train the laterality prediction model. The model predicting lesion laterality showed high accuracy, sensitivity, and specificity at 96.97, 97.01, and 96.88%, respectively. The 21 patients excluded during the training due to bilateral lesions were included according to the predicted lesion sides. More specifically, ten patients were classified into the left lesion group, and the remaining 11 patients were in the right lesion group. Finally, 44 patients were categorized into the left and 43 patients into the right lesion group. The model performance according to the number of features from 5 to 10 are demonstrated in the Supplementary Table S3.

Distribution of the selected features over the frequency band and ROIs, and summation of feature importance

The distribution of the selected features over the frequency band and the summation of their feature importance as calculated from the feature selection model are shown in Table 2. Theta, alpha2, and beta1-related features were chosen as significant explanatory factors in the model for the left stroke group, but only theta-related features were chosen for the right stroke group. Regarding feature importance, theta and beta1 exhibited the most significant values in the models for the left and right stroke groups, respectively. For the cortical region distribution, the left and right stroke models were similar in selecting most features from the frontal lobe. The number of important features from the frontal lobe was 65% in both models. Specifically, 13 and 11 features were located in the ROI 1–20 for the left and right models, respectively. The summation of feature importance of the features located in the frontal lobe was 0.50 and 0.46 for the left and right models, respectively, which is 73.47 and 68.06% of the total importance.

TABLE 1 Baseline characteristics of the study population.

| N = 87 | |
|----------------------------------------------------------|-------------------------------------------------------|
| <i>Demographics</i> | |
| Age | 65.4 ± 11.7 |
| Male | 53 (60.9) |
| Education, years | 10 [6–12] |
| <i>Vascular risk factors</i> | |
| Hypertension | 51 (58.6) |
| Diabetes mellitus | 25 (28.7) |
| Hyperlipidemia | 23 (26.4) |
| Smoking | 30 (34.5) |
| Atrial fibrillation | 14 (16.1) |
| Coronary heart disease | 7 (8.1) |
| Previous stroke/transient ischemic attack | 15 (17.2) |
| <i>Index-stroke characteristics</i> | |
| Premorbid modified Rankin scale >0 | 3 (3.4%) |
| TOAST | |
| LAA | 32 (36.8) |
| SVO | 24 (27.6) |
| CE | 13 (14.9) |
| Other-determined | 2 (2.3) |
| Undetermined | 16 (18.4) |
| Initial NIHSS | 3.0 [1.0–6.0] |
| <i>Acute lesions characteristics</i> | |
| Multiple lesions | 41 (48.8) |
| Left hemispheric lesions | 51 (60.7) |
| Cortical lesions | 37 (44.1) |
| <i>Chronic lesions</i> | |
| Medial temporal lobe atrophy | |
| 0 / 1 / 2 / 3 / 4 | 10 (12.1) / 39 (47.0) / 29 (35.0) / 4 (4.8) / 1 (1.2) |
| PVWMH | |
| 1 / 2 / 3 | 58 (69.1) / 17 (20.2) / 9 (10.7) |
| SCWMH | |
| 1 / 2 / 3 | 56 (66.7) / 19 (22.6) / 9 (10.7) |
| Lacunes | |
| Number of lacunes | 0 [0–1] |
| CMB | |
| Number of CMBs | 0 [0–1] |
| <i>Time intervals</i> | |
| Index-stroke to EEG | 3 [2–5] |
| Index-stroke to baseline neuropsychological evaluations | 4 [2–6] |
| Index-stroke to follow-up neuropsychological evaluations | 97 [90–103] |

Continuous variables were presented as mean ± standard deviation for normally distributed data or as median with IQR for skewed data; categorical variables were presented as frequencies (proportions).

TOAST (Trial of ORG 10172 in Acute Stroke Treatment), LAA (large-artery atherosclerosis), SVO (small vessel occlusion), CE (cardioembolism), NIHSS (National Institutes of Health Stroke Scale), PVWMH (periventricular white matter hyperintensities), SCWMH (subcortical white matter hyperintensities), CMB (cerebral microbleeds), EEG (electroencephalography).

TABLE 2 Distribution of the selected features over the frequency band and ROIs, and summation of feature importance.

| | Number of features | | Summation of feature importance | |
|---------------------------------|--------------------|--------------------|---------------------------------|--------------------|
| | Left stroke group | Right stroke group | Left stroke group | Right stroke group |
| <i>Frequency bands</i> | | | | |
| Delta | 1 (5%) | 2 (10%) | 0.04 (14.09%) | 0.05 (8.53%) |
| Theta | 10 (50%) | 1 (5%) | 0.05 (17.54%) | 0.03 (4.74%) |
| Alpha1 | 1 (5%) | 2 (10%) | 0.02 (6.15%) | 0.04 (7.00%) |
| Alpha2 | 5 (25%) | 4 (20%) | 0.11 (42.43%) | 0.17 (27.94%) |
| Beta1 | 2 (10%) | 4 (20%) | 0.03 (12.76%) | 0.13 (21.71%) |
| Beta2 | 1 (5%) | 6 (30%) | 0.02 (7.03%) | 0.16 (25.36%) |
| Beta3 | 0 (0%) | 0 (0%) | 0 (0%) | 0 (0%) |
| Gamma | 0 (0%) | 1 (5%) | 0 (0%) | 0.03 (4.71%) |
| Total | 20 (100%) | 20 (100%) | 0.26 (100%) | 0.61 (100%) |
| <i>Region of interests</i> | | | | |
| Frontal lobe | 13 (65%) | 13 (65%) | 0.50 (73.47%) | 0.46 (68.06%) |
| Temporal lobe | 4 (20%) | 2 (10%) | 0.10 (15.48%) | 0.04 (5.32%) |
| Parietal lobe | 1 (5%) | 1 (5%) | 0.03 (4.12%) | 0.02 (3.54%) |
| Occipital lobe | 1 (5%) | 3 (15%) | 0.02 (3.60%) | 0.07 (9.79%) |
| Isthmus (excluded from 4 lobes) | 1 (5%) | 0 (0%) | 0.02 (3.33%) | 0 (0%) |
| Others | 0 (0%) | 1 (5%) | 0 (0%) | 0.03 (3.87%) |
| Total | 20 (100%) | 20 (100%) | 0.6751 (100%) | 0.61 (100%) |

EEG-based network features and machine learning models for predicting MoCA scores after stroke

The correlation between the 20 selected features and the standardized scores of the MoCA after 3 months for the left and right stroke groups are demonstrated in [Supplementary Tables S4, S5](#), respectively. In the case of the left stroke group, all features related to the theta band showed a positive correlation coefficient with the MoCA standardized scores. By contrast, the correlation in the right stroke group does not show a clear relationship between the selected features and the target value. Using the 20 selected features, we trained the aforementioned machine learning regression models, and the results are shown in [Figure 2](#) and [Table 3](#). As shown in [Table 3](#), Ridge regression demonstrated the best performance in both groups (R -squared 0.76 for the left stroke group, 0.65 for the right stroke group). Our model did not incorporate patients' clinical variables that were not derived from the EEG data. Nonetheless, certain variables, like age and baseline MoCA scores, could potentially influence the MoCA score at 3 months. As part of a sensitivity analysis, we also evaluated the model's performance when it included age, baseline MoCA scores, and baseline NIHSS scores. However, this addition did not enhance the model's predictive capability, with R -squared values of 0.6174 and 0.6207 for the left and right stroke groups, respectively.

Sensitivity analysis with reduced variable count

In our sensitivity analyses with a reduced number of variables, the left lesion model, incorporating the top 8 features, achieved an

R -squared value of 0.57. Conversely, the right lesion model, leveraging the top 10 features, reached an R -squared value of 0.51 (refer to [Supplementary Figure S1](#)). We recognized that simplifying features based purely on their significance could potentially diminish the model's explanatory capability. To circumvent this, we crafted a composite model, opting for 3 to 5 of the top 5 features, and complementing with a selection from the bottom 15, ensuring the total did not surpass 10 features. This intricate procedure entailed evaluating close to 9,000 models to pinpoint those with the most robust R -squared values.

As evidenced in the [Supplementary Table S6](#), three models from the left lesion exceeded the performance of the top 20 feature-oriented models in terms of their explanatory capacity. In contrast, for the right lesion, a singular model outperformed the foundational model. The superior model for the left lesion amalgamated the top 5 features with an added 5 from the lower end, registering an R -squared value of 0.77. As for the right lesion, the eminent model incorporated the top 3 features and a supplementary 7 from the bottom tier. The definitive variables chosen by these exemplary models, along with the regression plots, are detailed in [Supplementary Figures S2, S3](#).

Discussion

In our study, the brain network attributes, which were analyzed using the EEG data obtained at the acute stage after stroke through machine learning, were useful in predicting the MoCA score at 3 months after acute stroke. Furthermore, this prediction was made based only on EEG analysis results without considering other

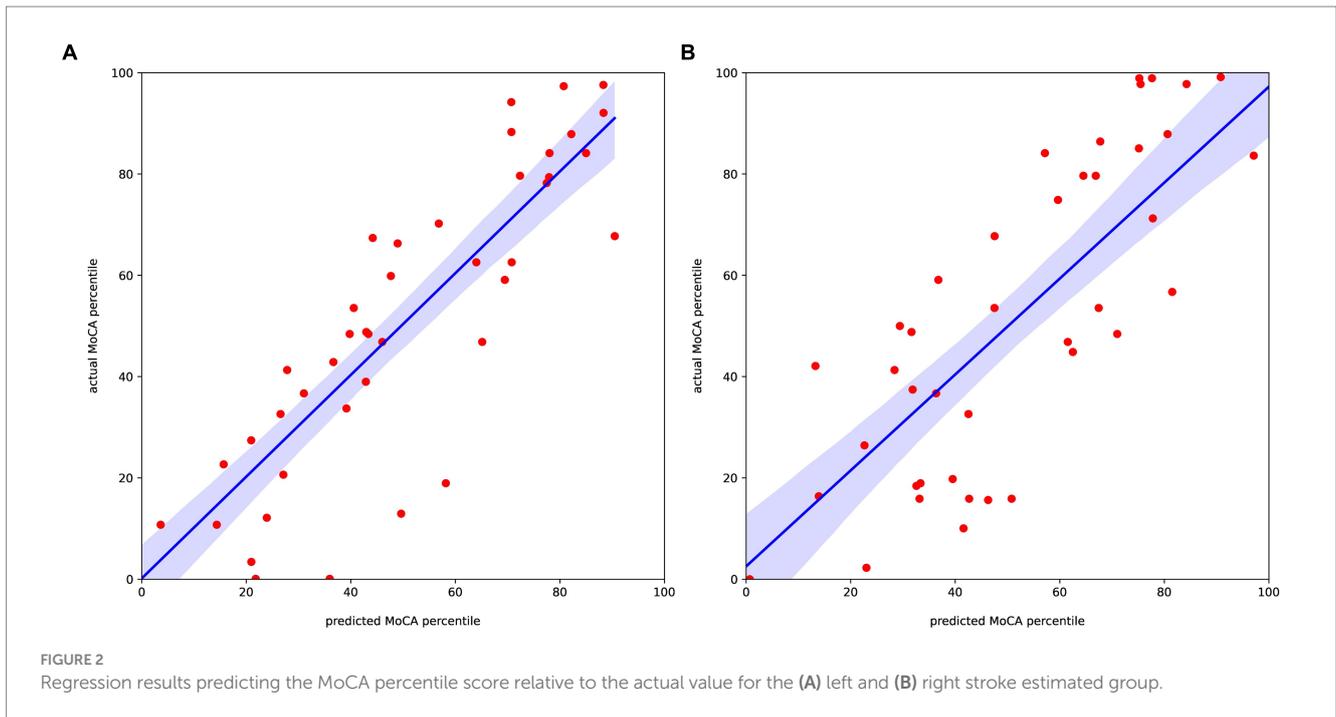


TABLE 3 R-squared results of the machine learning regression models.

| Network type | Network features only (left/right) | All EEG features (left/right) |
|---------------------------|------------------------------------|-------------------------------|
| Ridge | 0.76/0.65 | 0.64/0.48 |
| Lasso | 0.71/0.57 | 0.61/0.35 |
| ElasticNet | 0.63/0.55 | 0.44/0.37 |
| Support Vector Regression | 0.17/0.26 | 0.07/0.01 |
| AdaBoost Regression | 0.54/0.54 | 0.62/0.63 |

clinical information, and this approach may be useful in clinical practice.

As expected, the features selected in the machine learning models showed different trends between the left and right stroke groups. In the patients with the left hemispheric lesion group, we identified many significant features related to the theta band, whereas the right lesion group had only a few features distributed in the theta band. In the left hemispheric lesion group, the network attributes in the theta band were consistently and significantly correlated with the post-stroke MoCA scores at 3 months after stroke. The higher the global efficiency or clustering coefficient, or the lower the characteristic path length, the higher the MoCA score. However, the directionality of this association was not observed in the right hemispheric lesion group.

Network properties can be categorized in terms of integration (global efficiency and characteristic path lengths) and segregation (clustering coefficients and modularity) to explore their association with post-stroke cognitive impairment (Lim et al., 2021). In a previous study, a gradual recovery of reduced modularity after an acute stroke to the level of the control group was associated with improvements in memory and attention scores (Siegel et al., 2018). Another study showed the association between reduced global efficiency and post-stroke depression (Xu et al., 2019). In our results, the global efficiency of the theta band was the most important

explanatory factor in the left stroke group, and the clustering coefficient of the alpha band was found to be important in the right stroke group (Siegel et al., 2018). Contrary to the findings of a previous study, our results did not underscore the significance of modularity. This discrepancy may be attributable to the smaller quantity of modularity features included in our analysis compared to other features. As previously discussed, modularity shares conceptual similarities with the clustering coefficient, and our analysis revealed that the clustering coefficient constituted the majority of the top 20 selected features, irrespective of lesion laterality.

The top 20 important EEG network features reflecting cognitive prognosis were also more often extracted from the opposite side of the lesion. In the case of the left stroke group, 12 features were selected from the right hemisphere, and 8 were selected from the left hemisphere. In the case of the right stroke group, 12 features were selected from the left hemisphere, and 7 features were selected from the right hemisphere. Previous studies have shown that contralesional hemispheric networks were activated for compensatory processes during functional recovery after stroke (Rehme et al., 2011). During subacute periods after stroke, a positive correlation of contralesional, homologous regions with ipsilesional regions exerts a beneficial effect on muscle strength recovery (Rehme et al., 2011). However, these correlations were dependent on time

since the stroke, and our results show that some features of the contralateral hemisphere are negatively correlated with cognition, so further research is needed. Prior to the prediction, we divided the dataset into two subsets based on the laterality of stroke infarct, as it affects the overall EEG signal. Then, we selected the 20 most important features from the network variables estimated from functional connectivity using the iCOH between 68 ROIs. The selected features almost belonged to specific frequency bands, namely, theta, fast alpha, and slow beta, even though the left and right stroke estimated groups showed different tendencies. Several studies also proposed the relationship between post-stroke cognitive impairment and these frequency bands (De Vico et al., 2009; Al-Qazzaz et al., 2018; Doerrfuss et al., 2019). In the case of the ROIs of the network features, both the left and right stroke groups showed consistent results, wherein most of the features were concentrated in the frontal lobe, especially in the ROI 1–20 (Schleiger et al., 2014). It is known that frontal executive function, attention, and processing speed are mainly impaired in stroke patients (Lo et al., 2019), and MoCA is reported to be more sensitive in detecting frontal dysfunction than other tests, such as MMSE (Duncan and Owen, 2000; Fuster, 2002; Pendlebury et al., 2012; Braakman et al., 2013; Braun et al., 2015). In this sense, the regression results support the relationship between predictive cognitive status and brain network features, predominantly of frontal origin.

Our primary analysis consisted of only network characteristics as explanatory variables, as hypothesized. We also examined whether explanatory power improved in sensitivity analyses using all functional connectivity features, including iCOH. However, the prediction results were worse than those obtained using network features only. It is not uncommon when certain variables irrelevant to the outcome of interest are included in the machine learning model that eventually adds noises and reduces the accuracy of predictability. Further, in cases with many independent variables, several statistical problems arise due to multiple comparison issues. In this case, reducing dimension is a useful approach to finding biologically relevant results (Lim et al., 2021). The iCOH values offer considerably more information, so they are highly correlated with other iCOH values obtained in the same or adjacent ROIs. By contrast, the network variables of each ROI provide relatively independent information from other ROIs, and it is assumed that the independence between input features provided better results.

Our study has several limitations. Firstly, the sample size of stroke patients was insufficient to establish an independent validation cohort for testing the derived machine learning model. Consequently, we were compelled to implement the leave-one-out technique to enhance the reliability of both the training set and validation outcomes. Secondly, our limited patient population did not allow for examining differences attributable to varying stroke mechanisms. Furthermore, we gauged cognitive function solely through MOCA, excluding comprehensive neuropsychological testing. This restricts our ability to ascertain potential variations in correlations between EEG characteristics and individual cognitive domains. Additional inherent limitation is our choice of threshold for brain network indicators, set at the 1/4 level. While this decision balanced detail and feasibility, the optimal thresholding remains an area for future refinement and validation. Finally, patients who were unable to perform neuropsychological tests were excluded. Consequently, our findings may not be universally applicable to all ischemic stroke

patients. Administering cognitive tests to every ischemic stroke patient in a real-world context presents practical challenges. This inherent bias, stemming from attrition, has been extensively discussed in our earlier sections. Readers should exercise caution and keep this context in mind while interpreting our results (Pendlebury et al., 2015).

Conclusion

The predictive accuracy of post-stroke cognitive function may be enhanced by evaluating network characteristics via a machine learning-based EEG analysis. Therefore, implementing EEG during the acute phase of a stroke could serve as a reliable method for anticipating short-term cognitive prognosis following an ischemic stroke. Further, prediction of post-stroke cognitive status after stroke may present a promising avenue for enhancing the precision medicine for post-stroke care.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Ethics statement

The studies involving humans were approved by Institutional Review Board of the Hallym University Sacred Heart Hospital. The studies were conducted in accordance with the local legislation and institutional requirements. The ethics committee/institutional review board waived the requirement of written informed consent for participation from the participants or the participants' legal guardians/next of kin because of its retrospective nature and due to the minimal risk to participants.

Author contributions

ML, YH, SA, J-SL, and SK contributed to conception and design of the study. UP organized the database. JS and YH performed the statistical analysis. SA and YH wrote the first draft of the manuscript. JL, MO, B-CL, K-HY, and ML wrote sections of the manuscript. All authors contributed to the article and approved the submitted version.

Funding

This research was supported by the Bio&Medical Technology Development Program of the National Research Foundation (NRF) funded by the Korean government (MSIT) (No. RS-2023-00223501), and the research grant funded by the Korean Dementia Association.

Conflict of interest

YH, UP, JS and SK were employed by iMedisync, Inc.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the

reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fnagi.2023.1238274/full#supplementary-material>

References

- Aerts, H., Fias, W., Caeyenberghs, K., and Marinazzo, D. (2016). Brain networks under attack: robustness properties and the impact of lesions. *Brain* 139, 3063–3083. doi: 10.1093/brain/aww194
- Al-Qazzaz, N. K., Ali, S. H. B. M., Ahmad, S. A., Islam, M. S., and Escudero, J. (2018). Discrimination of stroke-related mild cognitive impairment and vascular dementia using eeg signal analysis. *Med. Biol. Eng. Comput.* 56, 137–157. doi: 10.1007/s11517-017-1734-7
- Bae, H.-J., Kim, J. Y., Kang, J., Kim, B. J., Han, M.-K., Choi, K.-H., et al. (2022). Sherman lecture award: 15-year experience of the nationwide multicenter stroke registry in Korea. *Stroke* 53, 2976–2987. doi: 10.1161/STROKEAHA.122.039212
- Biesbroek, J. M., Lim, J.-S., Weaver, N. A., Arikani, G., Kang, Y., Kim, B. J., et al. (2021). Anatomy of phonemic and semantic fluency: a lesion and disconnectome study in 1231 stroke patients. *Cortex* 143, 148–163. doi: 10.1016/j.cortex.2021.06.019
- Blum, S., Jacobsen, N. S. J., Bleichner, M. G., and Debener, S. (2019). A riemannian modification of artifact subspace reconstruction for eeg artifact handling. *Front. Hum. Neurosci.* 13:141. doi: 10.3389/fnhum.2019.00141
- Braakman, H. M. H., Vaessen, M. J., Jansen, J. F. A., Debeij-van Hall, M. H. J. A., de Louw, A., Hofman, P. A. M., et al. (2013). Frontal lobe connectivity and cognitive impairment in pediatric frontal lobe epilepsy. *Epilepsia* 54, 446–454. doi: 10.1111/epi.12044
- Braun, U., Schäfer, A., Walter, H., Erk, S., Romanczuk-Seiferth, N., Haddad, L., et al. (2015). Dynamic reconfiguration of frontal brain networks during executive cognition in humans. *Proc. Natl. Acad. Sci.* 112, 11678–11683. doi: 10.1073/pnas.1422487112
- Bressler, S. L., and Menon, V. (2010). Large-scale brain networks in cognition: emerging methods and principles. *Trends Cogn. Sci.* 14, 277–290. doi: 10.1016/j.tics.2010.04.004
- Carter, A. R., Shulman, G. L., and Corbetta, M. (2012). Why use a connectivity-based approach to study stroke and recovery of function? *NeuroImage* 62, 2271–2280. doi: 10.1016/j.neuroimage.2012.02.070
- Cawley, G. C., and Talbot, N. L. (2004). Fast exact leave-one-out cross-validation of sparse least-squares support vector machines. *Neural Netw.* 17, 1467–1475. doi: 10.1016/j.neunet.2004.07.002
- Chen, J.-L., Ros, T., and Gruzelier, J. H. (2013). Dynamic changes of Ica-derived eeg functional connectivity in the resting state. *Hum. Brain Mapp.* 34, 852–868. doi: 10.1002/hbm.21475
- De Vico, F. F., Astolfi, L., Cincotti, F., Mattia, D., la Rocca, D., Maksud, E., et al. (2009). Evaluation of the brain network organization from eeg signals: a preliminary evidence in stroke patient. *Anat. Rec.* 292, 2023–2031. doi: 10.1002/ar.20965
- Desikan, R. S., Ségonne, F., Fischl, B., Quinn, B. T., Dickerson, B. C., Blacker, D., et al. (2006). An automated labeling system for subdividing the human cerebral cortex on mri scans into gyral based regions of interest. *NeuroImage* 31, 968–980. doi: 10.1016/j.neuroimage.2006.01.021
- Doerrfuss, J. I., Kilic, T., Ahmadi, M., Holtkamp, M., and Weber, J. E. (2019). Quantitative and qualitative eeg as a prediction tool for outcome and complications in acute stroke patients. *Clin. EEG Neurosci.* 51, 121–129. doi: 10.1177/1550059419875916
- Duncan, J., and Owen, A. M. (2000). Common regions of the human frontal lobe recruited by diverse cognitive demands. *Trends Neurosci.* 23, 475–483. doi: 10.1016/S0166-2236(00)01633-7
- Finnigan, S., and van Putten, M. J. A. M. (2013). Eeg in ischaemic stroke: quantitative eeg can uniquely inform (sub-)acute prognoses and clinical management. *Clin. Neurophysiol.* 124, 10–19. doi: 10.1016/j.clinph.2012.07.003
- Fox, M. D. (2018). Mapping symptoms to brain networks with the human connectome. *N. Engl. J. Med.* 379, 2237–2245. doi: 10.1056/NEJMra1706158
- Fuster, J. M. (2002). Frontal lobe and cognitive development. *J. Neurocytol.* 31, 373–385. doi: 10.1023/A:1024190429920
- Hadiyoso, S., Ong, P. A., Zakaria, H., and Rajab, T. L. E. (2022). Eeg-based spectral dynamic in characterization of poststroke patients with cognitive impairment for early detection of vascular dementia. *J. Healthc. Eng.* 2022, 1–11. doi: 10.1155/2022/5666229
- Jatoi, M. A., Kamel, N., Malik, A. S., and Faye, I. (2014). Eeg based brain source localization comparison of sloreta and eloreta. *Australas. Phys. Eng. Sci. Med.* 37, 713–721. doi: 10.1007/s13246-014-0308-3
- Kang, Y. P. J., Yu, K., and Lee, B.-C. (2009). A reliability, validity, and normative study of the korean-Montreal cognitive assessment (k-Moca) as an instrument for screening of vascular cognitive impairment(vci). *Korean J. Clin. Psychol.* 28, 549–562. doi: 10.15842/kjcp.2009.28.2.013
- Langhorne, P., Bernhardt, J., and Kwakkel, G. (2011). Stroke rehabilitation. *Lancet* 377, 1693–1702. doi: 10.1016/S0140-6736(11)60325-5
- Lee, D. W., Lee, J. Y., Ryu, S. G., Cho, S. J., Hong, C. H., Lee, J. H., et al. (2005). Validity of the korean version of informant questionnaire on the cognitive decline in the elderly(iqcode). *Ann. Geriatr Med Res.* 9, 196–202.
- Lim, J.-S., Lee, J.-J., and Woo, C.-W. (2021). Post-stroke cognitive impairment: pathophysiological insights into brain disconnectome from advanced neuroimaging analysis techniques. *J. Stroke* 23, 297–311. doi: 10.5853/jos.2021.02376
- Liu, J., Li, M., Pan, Y., Lan, W., Zheng, R., Wu, F.-X., et al. (2017). Complex brain network analysis and its applications to brain disorders: a survey. *Complexity* 2017, 1–27. doi: 10.1155/2017/8362741
- Lo, J. W., Crawford, J. D., Desmond, D. W., Godefroy, O., Jokinen, H., Mahinrad, S., et al. (2019). Profile of and risk factors for poststroke cognitive impairment in diverse ethnoregional groups. *Neurology* 93:e2257, e2271. doi: 10.1212/WNL.00000000000008612
- Mullen, T. R., Kothe, C. A., Chi, Y. M., Ojeda, A., Kerth, T., Makeig, S., et al. (2015). Real-time neuroimaging and cognitive monitoring using wearable dry eeg. *I.E.E.E. Trans. Biomed. Eng.* 62, 2553–2567. doi: 10.1109/TBME.2015.2481482
- Nolte, G., Bai, O., Wheaton, L., Mari, Z., Vorbach, S., and Hallett, M. (2004). Identifying true brain interaction from eeg data using the imaginary part of coherency. *Clin. Neurophysiol.* 115, 2292–2307. doi: 10.1016/j.clinph.2004.04.029
- Pascual-Marqui, R. D. (2002). Standardized low-resolution brain electromagnetic tomography (sloreta): technical details. *Methods Find. Exp. Clin. Pharmacol.* 24 Suppl D, 5–12.
- Pendlebury, S. T., Chen, P. J., Bull, L., Silver, L., Mehta, Z., Rothwell, P. M., et al. (2015). Methodological factors in determining rates of dementia in transient ischemic attack and stroke: (i) impact of baseline selection bias. *Stroke* 46, 641–646. doi: 10.1161/STROKEAHA.114.008043
- Pendlebury, S. T., Mariz, J., Bull, L., Mehta, Z., and Rothwell, P. M. (2012). Moca, ace-r, and mmse versus the national institute of neurological disorders and stroke-Canadian stroke network vascular cognitive impairment harmonization standards neuropsychological battery after tia and stroke. *Stroke* 43, 464–469. doi: 10.1161/STROKEAHA.111.633586
- Rehme, A. K., Eickhoff, S. B., Wang, L. E., Fink, G. R., and Grefkes, C. (2011). Dynamic causal modeling of cortical activity from the acute to the chronic stage after stroke. *NeuroImage* 55, 1147–1158. doi: 10.1016/j.neuroimage.2011.01.014
- Sanchez Bornot, J. M., Wong-Lin, K., Ahmad, A. L., and Prasad, G. (2018). Robust eeg/meg based functional connectivity with the envelope of the imaginary coherence: sensor space analysis. *Brain Topogr.* 31, 895–916. doi: 10.1007/s10548-018-0640-0
- Scheltens, P., Leys, D., Barkhof, F., Huglo, D., Weinstein, H. C., Vermersch, P., et al. (1992). Atrophy of medial temporal lobes on mri in "probable" alzheimer's disease and normal ageing: diagnostic value and neuropsychological correlates. *J. Neurol. Neurosurg. Psychiatry* 55, 967–972.
- Schleiger, E., Sheikh, N., Rowland, T., Wong, A., Read, S., and Finnigan, S. (2014). Frontal eeg delta/alpha ratio and screening for post-stroke cognitive deficits: the power of four electrodes. *Int. J. Psychophysiol.* 94, 19–24. doi: 10.1016/j.ijpsycho.2014.06.012

- Schleiger, E., Wong, A., Read, S., Rowland, T., and Finnigan, S. (2017). Poststroke qeeg informs early prognostication of cognitive impairment. *Psychophysiology* 54, 301–309. doi: 10.1111/psyp.12785
- Sheorajpanday, R. V. A., Nagels, G., Weeren, A. J. T. M., van Putten, M. J. A. M., and Deyn, D. (2011). Quantitative eeg in ischemic stroke: correlation with functional status after 6months. *Clin. Neurophysiol.* 122, 874–883. doi: 10.1016/j.clinph.2010.07.028
- Siegel, J. S., Seitzman, B. A., Ramsey, L. E., Ortega, M., Gordon, E. M., Dosenbach, N. U. F., et al. (2018). Re-emergence of modular brain networks in stroke recovery. *Cortex* 101, 44–59. doi: 10.1016/j.cortex.2017.12.019
- Stinear, C. (2010). Prediction of recovery of motor function after stroke. *Lancet Neurol.* 9, 1228–1232. doi: 10.1016/S1474-4422(10)70247-7
- Subasi, A., and Ismail, G. M. (2010). Eeg signal classification using pca, Ica, lda and support vector machines. *Expert Syst. Appl.* 37, 8659–8666. doi: 10.1016/j.eswa.2010.06.065
- Wardlaw, J. M., Smith, E. E., Biessels, G. J., Cordonnier, C., Fazekas, F., Frayne, R., et al. (2013). Neuroimaging standards for research into small vessel disease and its contribution to ageing and neurodegeneration. *Lancet Neurol.* 12, 822–838. doi: 10.1016/S1474-4422(13)70124-8
- Weaver, N. A., Kancheva, A. K., Lim, J.-S., Biesbroek, J. M., Wajer, I. M. H., Kang, Y., et al. (2021a). Post-stroke cognitive impairment on the mini-mental state examination primarily relates to left middle cerebral artery infarcts. *Int. J. Stroke* 16, 981–989. doi: 10.1177/1747493020984552
- Weaver, N. A., Kuijff, H. J., Aben, H. P., Abrigo, J., Bae, H. J., Barbay, M., et al. (2021b). Strategic infarct locations for post-stroke cognitive impairment: a pooled analysis of individual patient data from 12 acute ischaemic stroke cohorts. *Lancet Neurol.* 20, 448–459. doi: 10.1016/S1474-4422(21)00060-0
- Wolfe, C. D. A. (2000). The impact of stroke. *Br. Med. Bull.* 56, 275–286. doi: 10.1258/0007142001903120
- Xu, X., Tang, R., Zhang, L., and Cao, Z. (2019). Altered topology of the structural brain network in patients with post-stroke depression. *Front. Neurosci.* 13:776. doi: 10.3389/fnins.2019.00776