Feature optimization method for machine learning-based diagnosis of schizophrenia using magnetoencephalography

Jieun Kim a,c,1, Min-Young Kim b, Hyukchan Kwon b, Ji-Woong Kim b, Woo-Young Im b, Sang Min Lee d, Kiwoong Kim a,c,**, Seung Jun Kim b,*

a Advanced Instrumentation Institute, Korea Research Institute of Standards and Science, Daejeon, Republic of Korea
b Department of Psychiatry, Konyang University College of Medicine, Konyang University Hospital, Daejeon, Republic of Korea
c Department of Medical Physics, University of Science and Technology, Daejeon, Republic of Korea
d Department of Psychiatry, Kyung Hee University School of Medicine, Kyung Hee University Hospital, Seoul, Republic of Korea

ARTICLE INFO

Keywords:
Machine learning
Feature optimization
Diagnosis
Schizophrenia
Magnetoencephalography

ABSTRACT

Background: When many features and a small number of clinical data exist, previous studies have used a few top-ranked features from the Fisher’s discriminant ratio (FDR) for feature selection. However, there are many similarities between selected features. New method: To reduce the redundant features, we applied a technique employing FDR in conjunction with feature correlation. We performed an attention network test on schizophrenic patients and normal subjects with a 152-channel magnetoencephalograph. P300m amplitudes of event-related fields (ERFs) were used as features at the sensor level and P300m amplitudes of ERFs for 500 nodes on the cortex surface were used as features at the source level. Features were ranked using FDR criterion and cross-correlation measure, and then the highest ranked 10 features were selected and an exhaustive search was used to find combination having the maximum accuracy.

Results: At the sensor level, we found a single channel of the occipital region that distinguished the two groups with an accuracy of 89.7 %. At source level, we obtained an accuracy of 96.2 % using two features, the left superior frontal region and the left inferior temporal region.

Comparison with existing method: At source level, we obtained a higher accuracy than traditional method using only FDR criterion (accuracy = 88.5 %). We used only the P300 m amplitude (not latency) on a single channel and two brain regions at a fairly high rate.

1. Introduction

Patients with schizophrenia are diagnosed primarily with the official diagnostic manuals, such as the Diagnostic and Statistical Manual of Mental Disorders (DSM-5) or International Statistical Classification of Diseases and Related Health Problems (ICD-10). Schizophrenia is diagnosed through a series of questions regarding symptoms or duration of illness to the patient or the patient’s family (American Psychiatric Association, 2013; Brämer, 1988). Positive and negative syndromic measures (PANSS) are used to measure the severity of schizophrenia symptoms (Kay et al., 1988). This traditional clinical diagnosis may not be accurate because schizophrenic patients sometimes intentionally hide their symptoms, and even experts may have difficulty differentiating schizophrenia from other mental illnesses with similar symptoms (McGorry et al., 1995). The precise diagnosis of schizophrenia patients is essential to establish treatment direction and increase treatment effectiveness. Therefore, researchers are looking for biomarkers to accurately differentiate schizophrenia patients based on objective and quantitative data using neuroimaging techniques—such as electroencephalography (EEG), magnetoencephalography (MEG), and functional MRI (fMRI). Using these techniques, many studies have reported disrupted cerebral activities in schizophrenia patients, such as changed event-related potential/field (ERP and ERF) wave forms for EEG and MEG, respectively, (Oribe et al., 2014; Mathalon et al., 2000; Neuhaus et al., 2011a, 2011; Souza et al., 1995; McCarley et al., 1993; Souza et al., 1995) as well as altered oscillatory power (Gallinat et al., 2004; Herrmann and Demiralp, 2005) and altered functional connectivity patterns (Hinkley et al., 2010; Ikezawa et al., 2011; Shen...
et al., 2010; Yu et al., 2015; Gandal et al., 2012; Shim et al., 2014).

Existing biomarkers are mostly focused on univariate analysis. The traditional univariate method focuses on the total difference at the group level. On the other hand, multivariate approaches can verify subtle effects based on information that are spatially distributed in the brain otherwise undetectable using traditional univariate methods (Haynes and Rees, 2006). Neuroimaging results should be inferred at the single-subject level rather than in groups to be useful in a clinical assessment since hospitals diagnose on an individual basis. Machine learning is characterized at the single-subject level and is inherently multivariate, making it a useful diagnostic method. Many studies have been conducted to distinguish schizophrenia patients by applying machine learning to EEG or MEG results. Neuhaus et al. investigated 144 schizophrenia patients and 144 matched controls with the auditory click-conditioning/oddball paradigm. P50 and N1 gating ratios as well as target-locked N1 and P3 components were submitted to conventional general linear models and to machine learning algorithms. They found that machine learning was superior to univariate analysis for classifying schizophrenia patients in the range of about 10 % as quantified by receiver operating characteristics (Neuhaus et al., 2013). In the same group, the best feature combination was obtained through an exhaustive searching method on feature sets—such as visual ERP P3 amplitude and latency and N1 latency—with an accuracy of 79 % using the Attention Network Test (ANT) (Neuhaus et al., 2011a). Shim et al. found 15 features that best distinguish the two groups by extending the features to sensor level (P3 amplitude and latency at 62 electrodes) as well as source level (amplitude at 314 nodes); the accuracy was 88.24 % using the auditory oddball paradigm (Shim et al., 2016). Dvey-Aharon et al. found 88.7 % accuracy with no false positives in the contextual sequence paradigm by finding the electrode, time window, and frequency interval that best distinguished schizophrenia (Dvey-Aharon et al., 2015).

One of the characteristics of schizophrenia patients is that their attention and concentration are lower than those of normal people (Carter et al., 2010). Based on these characteristics, the ANT was utilized. The ANT is based on a visual stimulus developed by Posner and Petersen (1990) and is known to activate three areas of attention—alerting, orienting, and executive attention neural networks (Rueda et al., 2015; Petley et al., 2018; Ishigami et al., 2013; Konrad et al., 2006; Neuhaus et al., 2007; Oberlin et al., 2005; Johnson et al., 2008). Neuhaus et al. showed in the ANT that the EEG signal amplitude modulation deficit between targets in the parietal region was seen in schizophrenia patients (Neuhaus et al., 2011a).

Generally, when creating features for machine learning with EEG or MEG, a few, a few hundred or even a few thousand features are generated. For example, Shim et al. generated total 376 features by combining the features from source level (amplitudes of 314 nodes on cortex surface) and sensor level (P3 amplitudes and latencies at 62 electrodes) (Shim et al., 2016). To find optimal feature subsets out of many features, the exhaustive search method is the most accurate and reliable method (Haidri et al., 2006). However, exhaustive search requires a lot of computation time compared to other search methods—such as sequential forward selection, sequential backward selection, sequential floating forward search, and Greedy search—because it requires searching all combinations in a large feature set. For example, the number of all possible combinations for n features is 2^n. And when selecting d of n, the number of possible combinations is n!/((n-d)!*d!). On the other hand, any non-exhaustive search methods for finding optimal feature sets are not guaranteed.

By reducing the number of features by removing unnecessary or less informative features, it may be possible to apply the exhaustive searching method. Traditionally, in neuroscience, researchers have used the following methods to perform feature reduction or feature selection. The traditional way to reduce features is to examine each feature individually using one of the relevant criteria—such as t-test, receiver operating characteristic (ROC), or Fisher’s discriminant ratio (FDR)—and then discard or rank them in descending order to identify the top features. In the Shim et al. study, FDR-score filtering was used for feature reduction. They calculated the discrimination accuracies for 1–20 feature sets and obtained a feature set consisting of 15 feature combinations with the greatest accuracy. Feature selection based on FDR-score filtering may select only those features having a large FDR value, however, features with low FDR values can also have a large effect on distinguishing groups. In addition, there are many similarities and correlations between selected high ranked features. In the Shim et al. study, the regions corresponding to the 15 feature combinations were concentrated in the same region in sensor (8 electrodes in the right frontal area) and source (7 sources in the left temporal area). Generally, EEG and MEG measures the electrical voltage or magnetic field generated from the brain source, spreading across multiple channels. At source level, signal leakage occurs when estimating the brain source through an inverse problem (Gobel et al., 2017; Brookes et al., 2012; Colclough et al., 2016). Therefore, similar signals are measured between adjacent sensors or adjacent brain source location. By eliminating features with similar signals, the dimension can be reduced, the computation time can be greatly improved, and, in some cases, can improve classification performance. To reduce features by eliminating less informative, redundant features, we applied the feature subset selection method, scalar feature-ranking technique, which employs FDR in conjunction with feature correlation introduced by Theodoridis et al. (2010). In this study, we generated 149 features and 500 features at sensor level and source level, respectively. We have confirmed that the computation time is reduced and the classification performance is improved by removing the highly correlated features.

There are critical issues when creating a classification model for many features (or dimensions) in machine learning. According to the curse of dimension, as the dimension increases, the amount of required data increases exponentially (Bellman, 1961). For example, if the amount of data having an acceptable performance in one dimension is N, then N^2 in two dimensions and N^3 in three dimensions are required. When a small amount of data is scattered over a large dimensional space, the data may be sparsely represented in some regions in the feature space. Fig. 1 shows the changes in the error rate according to the number of features. With finite N training data, the error rate decreases at first when the number of dimension increases, but the error rate increases at certain threshold points, which is called the peaking phenomenon. The minimum point occurs at l = N/α, where α typically ranges from 2 to 10 (Theodoridis and Koutroumas, 2008). Consequently, for a small amount of training data, a small number of features must be used. In particular, neurophysiological data measured in hospitals are limited in obtaining a large amount of data. We used the following method to apply the optimal machine learning algorithm in a limited number of subjects. We have searched for an optimum dimension for the classification model that minimizes (or maximizes) the classification error rate (or accuracy) according to the number of features.
features (Fig. 1). In this study, at sensor and source levels, we obtained the peaking point for 12 and 11 schizophrenia patients and 17 and 15 controls, respectively.

EEG is widely used among a variety of neuroimaging modalities for the diagnosis of schizophrenic patients. However, EEG signals can be distorted when they pass through tissues with different volume conduction properties (i.e., scalp, brain, and skull) (Nolte et al., 2004; Nunez et al., 1997). Conversely, MEG has better spatial resolution than EEG at sensor level because there is no signal distortion due to volume conductance differences (Quandt et al., 2012). For the purpose of this study, we used patient-friendly MEG that provides whole-head coverage with minimal set-up time.

2. Methods

2.1. Participants

The subjects of this study consisted of 14 schizophrenic patients (9 males and 5 females) and 23 healthy controls (10 males and 13 females) who were between 19 and 60 years of age. All the subjects were right-handed and reported normal or corrected-to-normal vision. The schizophrenic patients were recruited from the Psychiatry Department of Konyang University Hospital. The control group was recruited through internet job advertisements, and the age and gender were matched with the patient group. Patients with medical histories of alcohol or drug abuse within the past year, total IQ of less than 70, education duration of fewer than 9 years, other neurological disorders, or head injuries with loss of consciousness were excluded from the study. The schizophrenia patients were diagnosed based on the Structured Clinical Interview for Diagnostic and Statistical Manual of Mental Disorders (DSM-IV) for Axis I Psychiatric Disorders (Hahn et al., 2000). The PANSS was performed to evaluate the clinical symptoms of patients (Kay et al., 1987). The control group consisted of participants who had no psychiatric history or had no family history of major psychosis in their immediate family. The Korean-Wechsler Adult Intelligence Scale (K-WAIS) was used to assess the intelligence of the subjects (Yeom et al., 1997). All subjects gave written informed consent before entering the study. The experimental protocol was approved by the Institutional Review Board (IRB) of Konyang University Hospital. The demographic and clinical characteristics of the participants are shown in Table 1.

Subjects were excluded from the behavioral analysis if they did not concentrate or squeeze throughout the task. As a result, 14 schizophrenic patients and 19 controls were selected for behavior analysis. For MEG signal analysis, subjects were excluded if they had a lot of head movements or electromyography (EMG) signals. Subjects without MRI measurements were also excluded from source-level analysis. As a result, 12 and 11 schizophrenia patients and 17 and 15 controls were analyzed at sensor and source levels, respectively.

Table 1

|                        | Schizophrenia | Controls | P     |
|------------------------|---------------|----------|-------|
| N(female/male)         | 14(5/9)       | 23(13/10)| n.s.  |
| Age(years)             | 38.6 ± 11.5   | 32.9 ± 8.4| n.s.  |
| Education(years)       | 13.1 ± 1.8    | 15.4 ± 2.9| <0.05b|
| DOI(years)             | 14.1 ± 9.2    | --       | --    |
| CPZ eq.(mg/d)          | 639.3 ± 535.9| --       | --    |
| PANSS positive         | 12.3 ± 3.6    | --       | --    |
| PANSS negative         | 14.3 ± 5.0    | --       | --    |
| PANSS general          | 25.8 ± 6.1    | --       | --    |
| IQ                     | 98.6 ± 9.8    | 110.3 ± 8.3| <0.00b|

a: χ2-test for independent samples, DOI: duration of illness, PANSS: Positive and Negative Syndrome Scale, CPZ eq.: chlorpromazine equivalents.

2.2. Attention network test

Other papers have provided detailed descriptions of the ANT (Neuhaus et al., 2011b; Posner and Petersen, 1990; Fan et al., 2002). Fig. 2 provides an overview of the ANT. A fixation cross was shown during the entire experiment in the center of the screen. One of four types of cue stimuli (spatial, dual, center, no cue) is presented for 100 ms randomly and with equal probability. After 400 ms, one of three target stimuli—incongruent, congruent, and neutral—are presented for 100 ms randomly and with equal probability. The spatial cue appears as an asterisk above or below the cross mark and indicates the location of the next target. Dual cues are displayed at the top and bottom of the cross mark, and the center cue appears at the middle cross; otherwise, the cue signal does not appear (no cue). Targets are displayed randomly above or below the middle cross mark. The direction of the middle arrow is the same as the rest (congruent target), incongruent (incongruent target), and the remaining arrows are straight (neutral target). Subjects were asked to press the left or right button of the control box as quickly and accurately as possible by looking at the direction of the middle arrow of the target with the right hand. The target is presented for a maximum of 1700 ms and each trial was fixed to a total of 4 s. Each block contained 32 target stimuli and took 6 min. After a short break, the same experiment was run twice more. For three block, each target stimuli were presented 96 trials respectively.

2.3. MEG acquisition and preprocessing

The measurements were conducted with a whole-head helmet MEG system (152 channels axial gradiometer), which was developed by the Korea Research Institute of Standards and Science in South Korea (Kim et al., 2014; Lee et al., 2009). All experiments were conducted in a magnetically shielded room (MSR). The data were sampled at 1024 Hz. The stimulus is sent to the projector from a self-created (LabVIEW) stimulus program on the personal computer outside the MSR, and the subject sees the image projected from behind on the screen inside the MSR (Fig. 3). The distance between the subject’s eye and the screen was 1 m.

Participants were given a detailed description of the procedure before the experiment and were familiar with the experimental environment and the test stimulus. Four head positioning coils estimated each participant’s head position in the MEG helmet. The location of the positioning coils was recorded before and after each session. Anatomical fiducial points (i.e., nasion, left and right preauricular), the center point of each head positioning coil, and head surface (about 80 points) were measured using a 3D digitizer (ISOTRAK, Polhemus Navigation, Colchester, USA). This was necessary to match the head surface with a personal MRI. The nasal bridge and the under-eye points were included to make the MRI better matching.

For preprocessing of the MEG data, we used Fieldtrip, which is an open program (Oostenveld et al., 2011). First, MEG data were filtered at a 0.1 Hz high frequency (high-pass filter order = 4) and a 100 Hz low frequency, and a bandpass filter (59 Hz–61 Hz) was performed to remove the 60 Hz line noise. After that, eye blinking artifacts, noise by heartbeat, and big drift by breathing were eliminated through independent component analysis (“Jader” implemented in FieldTrip, www.sccn.ucsd.edu/eeglab/). Only the data measured from 149 sensors were used because two channels were off due to large noise during the recording.

2.4. Individual MRI acquisition

Structural MRI was obtained using a 3 T Philips Ingenuity with a T1-weighted 3D volumetric magnetization prepared rapid gradient echo (MPRAGE) sequence, with TR/TE of 9.8/4.6 ms, field of view (FOV) of 210, 8° flip angle, and 0.9 × 0.9 × 1.0 mm³ spatial resolution with 340 continuous coronal slices.
2.5. Sensor-level feature set

For sensor-level analysis, cleaned MEG data cleared of artifacts were band-pass filtered at 1–20 Hz and epoched from 2 s before the target stimulus onset to 2 s after the target stimulus onset. To remove muscle artifacts, epochs with variance greater than $1.5 \times 10^{-25}$ were removed. The wrong response trials were removed and then down-sampled at 512 Hz. The relative head position of the MEG sensor varies from person to person, from session to session. MEG data were interpolated to the standard sensor location obtained by averaging the sensor positions in all sessions across subjects using Fieldtrip’s "ft_megrealign" function.

To obtain an ERF, we averaged all the experiments in three consecutive sessions for each participant and averaged over all participants (grand-average). The baseline correction was done using values between 0.9 s and 0.6 s prior to the target onset. The P300 signal is usually defined as the maximum value between 250 and 500 ms after stimulation (Bharath et al., 2000; Lazzaro et al., 1997). We defined the P300 for MEG (P300 m) as averaged values at time band which distinguish maximally the two groups. We used the following pointwise biserial correlation coefficient to find the best time to distinguish the two groups (Blankertz et al., 2011). The pointwise biserial correlation coefficient (R-value) is defined as

$$r(x) = \frac{\sqrt{N_1 \cdot N_2}}{N_1 + N_2} \frac{\text{mean}(x; y = 1) - \text{mean}(x; y = 2)}{\text{std}(x)}$$

$$\text{signed-} r^2 = \text{values are } \text{sgn} \cdot r(x)^2.$$ 

Here, $N_1$ and $N_2$ is the number of healthy controls and schizophrenia...
Here, we divided the time between 0 s and 1 s after the target onset into 50 time bands (time window = 20 ms). We obtained the biserial correlation coefficient for the ERF amplitude for each of the three targets. Here, the signal with the largest value of the time-averaged (0.2 s - 0.5 s) coefficient values was the neutral target locked ERFs (singed- $r^2$: incongruent, 0.29; congruent, 0.37; neutral, 0.39). Fig. 4 (a) shows the singed- $r^2$ for 149 sensors and 50 time bands of neutral target ERFs. A positive value indicates that the ERF amplitude of the normal group is higher than that of the patient group. Here, the time band in which the two groups are well distinguished is between 0.3 s and 0.5 s in the dotted area. We selected time-averaged 149 ERF amplitudes (neutral target locked ERFs) of this time period (0.3 s - 0.5 s) as the sensor-level features. The neutral target locked ERFs were selected as features because of the highest singed- $r^2$ value among the three targets.

2.6. Source-level feature set

MEG cortical sources were estimated using weighted minimum norm estimation (wMNE) (Hämäläinen and Ilmoniemi, 1994; Hauk, 2004; Lin et al., 2006) implemented in Brainstorm toolbox (Tadel et al., 2011). We imported individual MRI reconstructed by FreeSurfer (Fischl, 2012) and set 15,000 vertices on the cortex surface. We manually set the fiducial points nasion, left preauricular, right preauricular, anterior commissure, posterior commissure, and interhemispheric point on the MRI. We also imported artifact-removed MEG data in each block recordings. The MEG data were epoched from 2 s before the target stimulus onset to 2 s after the target stimulus onset. Sensor and anatomy registration was done by using digitizer head point and MRI. We used an overlapping-sphere head model to compute the lead field matrix. The baseline period 0.9 s to 0.6 s before target onset was used to estimate noise covariance for each session of each subject. We computed unconstrained sources on the cortex surface using wMNE for each session of each subject. We averaged sources in three sessions, and then performed the baseline normalization 0.9 s to 0.6 s before target onset, DC offset correction in the Brainstorm toolbox. After that, averaged sources were bandpass filtered at 1–20 Hz. We calculated the norm values from three-axis directional brain sources (unconstrained to the cortex surface). Finally, the reconstructed sources were smoothed (FWHM = 6 mm) before performing group analysis and projected on default anatomy (Colin27, 2016 in the Brainstorm toolbox).

To extract source-level features, 15,000 cortical current dipoles were down-sampled to 500 cortical dipoles. We selected dipoles by simply skipping a certain number in the original vertices. The singed- $r^2$ (congruent, 0.327; incongruent, 0.296; neutral, 0.349) was calculated to find the best time window to distinguish between the two groups. As shown in Fig. 4 (b), the two groups were most distinguished between 0.3 s and 0.5 s (boxed with dashed line). A positive value indicates that the ERF value of the normal group is higher than that of the patient group. As sensor level, we selected time-averaged 500 ERF amplitudes (neutral target locked ERFs) of this time period (0.3 s - 0.5 s) as the source-level features. The neutral target locked ERFs were selected as features because of the highest singed- $r^2$ value among the three targets. Fig. 4 (c) shows the 500 nodes at source level.

2.7. Data normalization

There is a wide range of values between features, and if data normalization is not performed, features with large values can have a large influence when creating a classification model. We performed Z-score standardization with a mean of 0 and a variance of 1 for each of the features at sensor and source levels, respectively.

2.8. Feature selection (or reduction) and classification

As mentioned in introduction, data measured with EEG or MEG generate few to many hundreds of features. Most features do not affect the distinction between classes, they are unnecessary and need to be removed. By eliminating features with similar signals, the dimension can be reduced and the computation time can be greatly improved. In sensor spatial analysis, the most affected sensors in MEG (and EEG) are not located directly above the source, and the fields generated by a single brain region are spread across many sensors (Schoffelen and Gross, 2009). For example, as shown in Fig. 5, when a visual stimulus-induced signal is measured with MEG, a similar signal is measured by nearby sensors due to volume conduction or spread effects. Fig. 5(a) shows the 152-channel congruent target locked grand averaged ERF signal. Fig. 5 (b) shows the grand averaged ERF signal measured by the four sensors in the red circle in Fig. 5(a). Fig. 5(c) shows magnetic fields by four sensors for all subjects with the mean value of the shaded area in Fig. 5(b). We found that the correlations between the sensors were very high (Fig. 5(d)).

We applied the method for reducing redundant features introduced by Theodoridis et al. (2010). They call this method as a scalar feature-ranking technique, which employs FDR in conjunction with feature correlation. The method of eliminating features with high cross-correlation is as follows. Before finding the correlation between features, first, the FDR for each feature is calculated, then, ranked in descending order. The FDR value is expressed as $C_{ij}$ in Eq. (1). The first feature with the largest FDR is called $i_1$ and the second-best feature is obtained by Eq. (1). The correlation between the first feature and the remaining
features is obtained using Eq. (2). Here, the two variables are multiplied by \( a_1 \) (related to the FDR) and \( a_2 \) related to the correlation. The parameters \( a_1 \) and \( a_2 \) are user-defined. Here, \( a_1 + a_2 = 1 \). Select the feature that maximizes Eq. (1) for the second-best feature.

\[
i_i = \max_j \{ |a_1 C_j - a_2 \rho_{ij}| \}, \quad j \neq i
\]  

(1)

\( i_1 \): second most important feature

\( C_j \): FDR values

\( \rho_{ij} \): correlation coefficient among the first feature with each of the remaining features.

\( a_1, a_2 \): user-defined values

\[
\rho_{ij} = \frac{\sum_{m=1}^{N} x_{im} x_{jm}}{\sqrt{\sum_{m=1}^{N} x_{im}^2} \sum_{m=1}^{N} x_{jm}^2}
\]  

(2)

To select the third best feature, follow Eq. (3).

\[
i_k = \max_j \{ a_1 C_j - a_2 \frac{1}{k-1} \sum_{r=1}^{k-1} |\rho_{ij}| \}, \quad j \neq i, \quad r = 1, 2, ..., k-1, \quad k = 3, 4, ..., m.
\]  

(3)

Next, calculate the correlation between the first best feature and the rest of the features other than the first and second best features. Calculate the correlation between the second-best feature and the remaining ranked features. Two correlations are obtained and then averaged. The remaining best feature selection also follows Eq. (3). Fig. 6(a) shows the selected sensors when filtering in order of high FDR value (10, 20, 30) without consideration of cross-correlation. Fig. 6(b) shows the selected sensors when cross-correlation is considered. For example, here, \( a_1 = 0.2 \) and \( a_2 = 0.8 \).

At sensor and source levels, when the correlation was not considered, the selected areas were close to each other. When the regions with high correlation are removed, we confirmed the selected areas are scattered.

To find the best feature subset for the classification model, we applied the feature subset selection method introduced by Theodoridis et al. (2010). The two stages for feature subset selection are as follows. First, we used the scalar feature-ranking technique. As mentioned before, scalar feature selection ranks features using FDR criterion as well as cross-correlation measure between pairs of features. The scalar feature selection technique can eliminate less informative ones and reduce the number of features. Second, 10 features with the highest ranking are selected and exhaustive search is used to find the combination set having maximum accuracy. The classification accuracy was evaluated using leave-one-out cross-validation (LOO-CV) with a support vector machine (SVM) classifier for each feature set (Pereira et al., 2009; Vapnik, 2013; Cherkassky and Mulier, 2006). Fig. 7 shows the overall analysis procedures adopted in this study.

In the scalar feature selection step, we mentioned the two parameters \( a_1 \) and \( a_2 \) that are related to FDR value and the cross-correlation (Eq. (1)). We examined the classification accuracies when the two parameters \( a_1 \) and \( a_2 \) are changed. We also investigated the classification accuracies according to the change in the number of features (or dimensions) for classification model at sensor and source levels, respectively.

2.9. Statistical analyses

Statistical calculations were conducted with MATLAB and Statistics Toolbox Release R2013b (The MathWorks, Inc., Natick, Massachusetts,
United States). Demographic data were analyzed with χ² test and t-test for independent samples. In the behavioral analysis, two sample t-tests were applied to analyze the average response time, response times for each condition, and three attention network effects between the two groups. A repeated measures analysis of variance (ANOVA) was applied to analyze the group as between-subjects factors and the cue and target stimuli as within-subjects factors. For univariate analysis between groups and between conditions at sensor and source levels, cluster-based permutation tests were conducted. Classification analysis between groups was performed as described in the previous section.

3. Results

3.1. Behavior analysis

There was a significant difference in the mean response time between the two groups. The mean response time was 520.09 ms (SD 68.6 ms) in the control group and 609.99 ms (SD 133 ms) in the patient group (Two-sample t-test; p < 0.016). Fig. 8 shows the mean response time and error rate for schizophrenia patients (a) and healthy controls (b) by cue and target condition.

The statistical results between the two groups according to each condition are shown in Table 2. Significant differences between the two groups were noted in all conditions except spatial cue.

For three attention network effects—alerting (dual cue – no cue), orienting (spatial cue – center cue), and inhibition (incongruent – congruent)—there were no significant differences between groups (controls: 33.8 ± 15.7 ms, 34.9 ± 25.1 ms, 91.5 ± 27.2 ms; patients: 33.6 ± 15.8 ms, 51.2 ± 31.0 ms, 88.8 ± 44.1 ms). Using a repeated measured ANOVA, in healthy controls significant main effects of RT were found for targets (F_{1,22} = 45.1; p < 0.001) and cues (F_{1,10} = 6.14; p < 0.001) and no significant difference was found for the interaction of targets and cues (F_{1,2} = 0.15; p = 0.98). In post-hoc analysis, significant differences were found between spatial cue and center cue (p < 0.05), spatial cue and no cue (p < 0.001), the neutral and incongruent target (p < 0.001), and the congruent and incongruent target (p < 0.001). In schizophrenia patients, significant main effects of RT were found for targets (F_{1,2} = 9.21; p < 0.01), however no significant difference was found for cues (F_{1,7} = 2.50; p = 0.061) and for the interaction of targets and cues (F_{1,12} = 0.06; p = 0.99). In post-hoc analysis, significant differences were found between spatial cue and no cue (p < 0.05), neutral and incongruent target (p < 0.001), and congruent and incongruent target (p < 0.05).

3.2. Univariate analysis at sensor level

In the control group, there was a significant difference between congruent and incongruent target, as shown in Fig. 9 (a) (*p < 0.025, in detail, negative cluster: p < 0.001, time range: 0.30 s - 0.47 s; positive cluster: p = 0.016, time range: 0.35 s - 0.47 s; cluster-based
permutation test provided by Fieldtrip, two-sided test, number of permutations = 1,000). Magnetic field distribution presents the grand averaged ERF (incongruent target minus congruent target) at time 0.4 s after target onset in healthy controls. We also found that there is a significant difference between neutral and incongruent targets in the control group, as shown in Fig. 9 (b) (*p < 0.025, in detail, negative cluster: p < 0.001, time range: 0.30 s - 0.48 s, positive cluster: p = 0.003, time range: 0.35 s - 0.45 s, cluster-based permutation test provided by Fieldtrip, two-sided test, number of permutation = 1,000). There were no significant differences among the three targets in the schizophrenia group. Fig. 9(c) and (d) show the grand averaged ERFs for healthy controls and schizophrenia patients, respectively, corresponding to the channel indicated by the white circle in Fig. 9 (a) and (b). The target stimulus is presented at 0 s. MEG data were interpolated to the standard sensor positions obtained by averaging the sensor positions in all sessions across subjects.

Significant differences between groups were found only in the neutral target ERF signal among the three target types (neutral, p < 0.005; congruent, p = 0.057; incongruent target, p > 0.3; two-sided test, number of permutation = 1000, cluster-based corrected). Significant differences between the two groups were found at 0.27 to 0.44 s, as shown in Fig. 10(a) (p < 0.005, cluster-based permutation test). Fig. 10 (b) shows the grand averaged ERFs of the neutral target, and Fig. 10(c) shows the grand averaged ERF signals of two groups for the highlighted channel of the occipital region.

3.3. Univariate analysis at source level

We found the greatest difference between groups at the left medial cortex and the left inferior temporal cortex for a very short time for congruent and incongruent target stimuli (congruent: time = 0.340 s - 0.350 s, p = 0.008; incongruent: time = 0.350 s - 0.359 s, p = 0.016; no significant difference for neutral, cluster-based correction) as shown in Fig. 11. Significant cortical areas differing between schizophrenia and control groups for the congruent target included the left parahippocampal, entorhinal, medial orbitofrontal, anterior fusiform, and anterior part of the inferior temporal cortex. The significant cortical regions differing between subject groups for the incongruent target were the same as for the congruent target and included the left temporal pole, pars orbitalis, and lateral orbitofrontal region. These results are in agreement with the Kawasaki et al. paper in that schizophrenia's information processing impairment indicated by reduced P300 amplitude is related to structural abnormalities of the medial temporal lobe (Kawasaki et al., 1997).

3.4. Classification using machine learning

We evaluated the classification accuracy changes according to the number of features (or dimensions) and parameters (a1, a2) using an SVM classifier at sensor and source levels, respectively. The soft margin SVM algorithm was implemented in MATLAB Version 9.6.0.1072779 (R2019a). The built-in Matlab function “fitcsvm” was applied to the training dataset. Linear kernels were used and Sequential Minimal Optimization (SMO) was used to solve the quadratic programming (QP) problem. The performance of SVM was evaluated by calculating a leave-one-out cross validation, and a confusion matrix, which is a table showing the number of true positives and negatives as well as false positives and negatives after prediction.

Fig. 12 shows the classification accuracies with respect to the number of features as well as the changes of parameters (a1, a2) used for the classification at sensor level. Here, we observed a peaking phenomenon in which the classification accuracy value decreases with increasing number of dimensions. We confirmed that the classification accuracy decreases when the dimension is larger than 5th dimension. As mentioned in the introduction, if the amount of training data is small, it is preferable that the dimension of the peaking point is low. We selected the dimension with the highest classification accuracy and lowest dimension. Regardless of the parameter values, a maximum classification accuracy of 89.7 % (sensitivity = 83.3 %, specificity = 94.1 %) in 2nd dimension has been reported.

| Schizophrenia vs. healthy controls | No cue | Dual cue | Center cue | Spatial cue | Incongruent target | Congruent target | Neutral target |
|-----------------------------------|-------|---------|-----------|------------|-------------------|-----------------|--------------|
| Mean response time for patients (ms) | 645   | 612    | 616      | 565       | 673               | 585             | 572          |
| Mean response time for controls (ms) | 548   | 513    | 526      | 491       | 584               | 493             | 483          |
| p-value                          | 0.01  | 0.008  | 0.02     | 0.051     | 0.022             | 0.015           | 0.018        |

Fig. 8. Averaged button response time according to three targets and four cues. (a) Schizophrenic patients. (b) Healthy controls.
Fig. 13(b) shows the normalized P300 amplitudes of all subjects in both groups for two channels (Fig. 13(a)) representing the best feature obtained at sensor level. The black line is the classifier obtained by SVM.

At source level, we obtained the accuracies according to the parameter \((a_1, a_2)\) and dimension number change as shown in Fig. 14. When the parameters are \(a_1 = 0.2\) and \(a_2 = 0.8\), the schizophrenia were classified with maximum accuracy 96.2 % (sensitivity = 90.9 %, specificity = 100 %) in 7th dimension.

Fig. 15 shows brain regions corresponding to seven features (left prefrontal, left temporal, left parietal, right prefrontal, right temporal, right parietal, right precuneus). At source level, the accuracy is higher than sensor level (96.2 % and 89.7 % in source and sensor levels, respectively).

Additionally, we evaluated the classification accuracy changes according to the number of features (or dimensions) and parameters \((a_1, a_2)\) using an SVM classifier at behavioral data. For each target and cue, we built a feature set with the mean response time, kurtosis, skewness, variance, and mean error rate of the response time. That is, total 29 features ((three targets + four cues)*4 response time features + mean error rate) were made.

At behavioral data, we obtained the accuracies according to the parameter \((a_1, a_2)\) and dimension number change as shown in Fig. 16. When the parameters are \(a_1 = 1\) and \(a_2 = 0\), the schizophrenia were classified with maximum accuracy 78.8 % (sensitivity = 71.4 %, specificity = 84.2 %) in 3rd dimension. The selected features were mean of spatial cue, mean of dual cue, and variance of incongruent target. In the behavioral data, the correlation between features was low, so the correlation removal effect could not be seen.

Finally, we evaluated the classification accuracy changes according to the number of features (or dimensions) and parameters \((a_1, a_2)\) using an SVM classifier at fused data. Fused data includes behavioral data, sensor level, and source level data. That is, total 678 features (behavior, 29 features; sensor level, 149 features; source level, 500 features) were made.

At fused data, we obtained the accuracies according to the parameter \((a_1, a_2)\) and dimension number change as shown in Fig. 17. When the parameters are \(a_1 = 0.2\) and \(a_2 = 0.8\), the schizophrenia were classified with maximum accuracy 100 % (sensitivity = 100 %, specificity = 100 %) in 3rd dimension. The selected features were left anterior cingulate cortex (ACC) and right temporal (inferior part) at source level, occipital region (MEG 151) at sensor level (Fig. 18).

4. Discussion

4.1. Best feature selection for classification model

By eliminating the unnecessary and redundant features, we were able to obtain an optimal combination set with reduced computation time when applying an exhaustive search method. When applying the scalar feature-ranking technique, we used the FDR values of individual feature and correlation values between features. A trade-off occurred
between the coefficients $a_1$ (related FDR) and $a_2$ (related correlation) multiplied by the two parameters. We observed how the accuracy varies with the variation of the coefficients. At source level, it was more accurate when considering the feature correlation while permitting the FDR value to be somewhat than when using only the FDR. However, ignoring the FDR value too strictly ($a_1 = 0, a_2 = 1$) degraded performance. On the other hand, at sensor level, the effect of correlation was not visible as at source level. Regardless of the parameter settings, the maximum accuracy was 89.7%. In other words, correlation considerations may not be effective in some cases. However, we suggest basically considering the correlation when making feature selection.

As mentioned in introduction, when there is a small amount of training data, if the dimension is too large when setting the classification model, there may be empty space without data, and the classifier may not work properly. Therefore, when the amount of training data is small, it is important to reduce the dimension for classification model.

For this, we obtained the peaking point by calculating all the accuracies according to the number of dimensions for the classification model. We obtained five peak point values according to the parameters used in the scalar feature-ranking technique. We selected the peak point with the highest accuracy and lowest dimension. We have chosen two dimension at sensor level and seven dimension at source level.

### 4.2. Diagnosis at sensor level

We classified the schizophrenic patients with maximum accuracy 89.7% and 96.2% at sensor level and source level, respectively. Accuracy is important for patient diagnosis, but the signal processing process should be simple and take less time. Diagnosis at source level requires personal MRI measurement and is time-consuming and costly. In addition, MRI segmentation and source analysis should be added. From this point of view, the diagnostic procedure at sensor level is relatively simple and cost-effective. Shim et al. combined the sensor and source levels to distinguish schizophrenia with an accuracy of 88.2% (Shim et al., 2016). We obtained better result with 89.7% accuracy using only sensor-level features. Neuhaus et al. obtained 79% accuracy using four features (feature 1: N1 latency, P4, averaged ERPs of congruent and incongruent, feature 2: P3 latency, Cz, congruent, feature 3: P3 latency, Cz, incongruent, feature 4: P3 amplitude, Cz, incongruent) using the ANT task at sensor level (Neuhaus et al., 2011a). In contrast, we used only two features (P300 m amplitude, two channels at the posterior, neutral target) at sensor level and classified the schizophrenia with greater accuracy than study of Neuhaus et al.

### 4.3. Comparison of previous study in ANT

We found a reduction in schizophrenia patients for neutral target P300 m amplitude on the posterior region at sensor level in both machine learning and statistical analysis (89.7% for classification accuracy, $p < 0.005$ for statistical analysis). Previous ANT studies have reported that the ERP activity at the anterior cingulate cortex in the incongruent target is reduced in schizophrenia compared to normal subjects in association with executive conflict (Carter et al., 1997; Kerns et al., 2005). However, the neutral target P300 m amplitude reduction in schizophrenia was not found in the existing EEG studies. This new finding will probably be possible because MEG is the first application of ANT tasks to schizophrenia.

As a result of machine learning at source level, seven features (left prefrontal, left temporal, left parietal, right prefrontal, right temporal, right parietal, right precuneus) were selected. In previous studies,
although the measurement equipment was different and different task-based signals with this study, Kim et al. found that the activity of the schizophrenia was abnormal in these areas (Kim et al., 2015). They showed decreased fMRI activities in schizophrenia in the superior frontal gyrus, dorsolateral prefrontal cortex, ventrolateral prefrontal cortex, anterior cingulate cortex, inferior parietal gyrus, and fusiform gyrus during the delayed-response working memory task with human face distracters. However, machine learning-based analysis is fundamentally different from conventional univariate analysis, therefore, it may not be meaningful to compare the features obtained through machine learning with previous univariate results in this study.

4.4. Limitation

At source level, we obtained a two-dimensional model with maximum accuracy using an SVM classifier, but the sample pointers were distributed too close to the classifier line. New samples for diagnosis are may be located at the classifier line. It is necessary to make a more reliable classification model by using other classifiers, such as linear discriminant analysis, k-nearest neighbor, Mahalanobis, and Naïve Bayes.

IQ was significantly different between the patient group and the control group (98.6 for schizophrenia patients, 110.3 for health controls, p < 0.001). Whether the result of this study is caused by IQ
difference is of importance. Accordingly, future studies recruiting schizophrenia patients with similar IQ to the normal group are needed.

Another limitation is the small number of subjects and little training data for machine learning. Using data from meta-analysis of machine learning (ML) in recent schizophrenia imaging, low sample size (N) studies can reach more than 90 % accuracy, but it was found that the maximum accuracy achieved above N/2 = 50 steadily dropped below 70 % for N/2 > 150 (Schnack and Kahn, 2016). Schnack et al. argue that smaller N studies can reach higher prediction accuracy while lowering the cost of generalization to other samples. Higher N studies, on the other hand, have more generalization power but less accurate. As N increases, they assumed that the root cause of the drop in accuracy is sample heterogeneity. Small studies include more homogeneous groups of subjects (the strict inclusion criteria are easily met and subjects live close to the study site), but larger studies will inevitably relax criteria/adoption in large areas. Although our study has a small number of training data, we focus on analyzing the correlation between features and suggesting ways to remove unnecessary features. We would like to emphasize that our method is applicable to any number of subjects, and also to other brain imaging modalities, such as fMRI and EEG.

In this study, we applied machine learning with pre-ranked ten features. As the number of pre-ranked features increases, the calculation time becomes longer. Also, if the amount of training data increases, the calculation time may increase indefinitely. Since sensor level and source level both have fairly high accuracy, it is enough to use 10 features.

5. Conclusion

The main features of this study are as follows. First, we implemented machine learning at sensor (149 channels), source levels (500 nodes) and behavior data (29 features), lastly fused data (behavior data, sensor and source level data) by measuring MEG with superior time and spatial resolution to distinguish schizophrenia. Second, to apply optimal machine learning techniques, we reduced the number of features by eliminating less informative, highly correlative features, resulting in improved computation time and classification performance. Third, we found the optimal number of features for the classification model by calculating the peaking point to overcome the curse of dimension occurring in high-dimensional space when the amount of training data is small.

At sensor level, we found two channels of the occipital region that maximally distinguishes the two groups with an accuracy of 89.7 % (sensitivity = 83.3 %, specificity = 94.1 %). At source level, we obtained maximum classification accuracy of 96.2 % (sensitivity = 90.9 %, specificity = 100 %) using seven features—left prefrontal, left temporal, left parietal, right prefrontal, right temporal, right parietal, right precuneus. At behavior data, we obtained maximum accuracy...
78.8 % (sensitivity = 71.4 %, specificity = 84.2 %) using three features—mean of spatial cue, mean of dual cue, and variance of incongruent target. Finally, at fused data, we could obtain a perfect accuracy using three features—left anterior cingulate cortex (ACC) and right temporal (inferior part) at source level, occipital region (MEG 151) at sensor level; we do not suppose the value 100 % is not so meaningful since it could be an over-fitted result due to a limited number of samples in our condition but at least it demonstrates that our method would be correctly working on these kind of data set.

We used only P300 m target locked ERF signal as a feature. In addition, schizophrenia can be classified using the ERF N100 m signal or the cue locked ERF signal as features. As features, new neurophysiological indicators—such as power oscillation changes and functional connectivity changes—can be used. Alternatively, as in the An et al. study, a machine-learning based diagnosis can be performed using neurophysiological data (gamma activity or ERFs) as well as behavioral responses (An et al., 2018). In this study, we classified schizophrenia, but other mental disorders—such as depression, obsessive-compulsive disorder, attention deficit hyperactivity syndrome (ADHD)—or degenerative brain diseases—such as Alzheimer’s and Parkinson’s disease—can also be distinguished and diagnosed by applying the proposed method.

Declaration of Competing Interest

The authors declare that they have no conflict of interest.

Acknowledgments

This research was supported by Development of Core Technology for Advanced Scientific Instrument funded by Korea Research Institute of Standards and Science (KRISS – 2019 – GP2019-0018)

References

American Psychiatric Association, 2013. Diagnostic and Statistical Manual of Mental Disorders (DSM-5®). American Psychiatric Pub.

An, K.M., Ikeda, T., Yoshimura, Y., Hasegawa, C., Saito, D.N., Kumazaki, H., et al., 2018. Development of Korean version of structured clinical interview schedule for DSM-IV axis I disorder: interrater reliability. J. Korean Neuropsychiatr. Assoc. 39 (2), 362–372.

Haindl, M., Somol, P., Verwerder, D., Kotropoulos, C., 2006. Feature Selection Based on Cross-Correlation. In: Iberoamerican Congress on Pattern Recognition. Springer, Berlin, Heidelberg, pp. 569–577.

Hämäläinen, M.S., Ilmoniemi, R.J., 1994. Interpreting magnetic fields of the brain: minimum norm estimates. Med. Biol. Eng. Comput. 32 (1), 35–42.

Hauk, O., 2004. Keep it simple: a case for using classical minimum norm estimation in the analysis of EEG and MEG data. Neuroimage 21 (4), 1612–1624.

Haynes, J.D., Rees, G., 2006. Neuroimaging: decoding mental states from brain activity in humans. Nat. Rev. Neurosci. 7 (7), 525.

Herrmann, C.S., Demiralp, T., 2005. EEG and MEG gamma oscillations in neuropsychiatric disorders. Clin. Neurophysiol. 116 (12), 2719–2731.

Hinkley, L.B., Owen, J.P., Fisher, M., Findlay, A.M., Vinogradov, S., Nagarajan, S.S., 2010. Cognitive impairments in schizophrenia as assessed through activation and connectivity measures of magnetoencephalography (MEG) data. Front. Hum. Neurosci. 3, 73.

Ikezawa, K., Ishii, R., Ivase, M., Kuriyama, R., Canuet, L., Takahashi, H., et al., 2011. Decreased alpha event-related synchronization in the left posterior temporal cortex in schizophrenia: a magnetoencephalography-beamformer study. Neurosci. Res. 71 (3), 235–245.

Ichihara, Y., Fisk, J.D., Wojtowicz, M., Klein, R.M., 2013. Repeated measured of the attention components of patients with multiple sclerosis using the Attention Network Test-Interaction (ANT-I): stability, isolaiblity, robustness, and reliability. J. Neurosci. Meth. 216, 1–9.

Johnson, K.A., Robertson, I.H., Barry, E., Mulligan, A., Dalhjin, A., Daly, M., Watchorn, A., Gill, M., Bellgrove, M.A., 2008. Impaired conflict resolution and alerting in children with ADHD evidence from the Attention Network Test (ANT). J. Child Psychol Psych. 49, 1339–1347.

Kawasaki, Y., Maeda, Y., Higashima, M., Nagasawa, T., Koshino, Y., Suzuki, M., Ide, Y., 1997. Reduced auditory P300 amplitude, medial temporal volume reduction and psychopathology in schizophrenia. Schizophr. Res. 26 (2–3), 107–115.

Kay, S.R., Fiszbein, A., Opler, L.A., 1987. The positive and negative syndrome scale. Schizophr. Res. 122 (1–2), 19–29.

Kawabata, M., Balota, D.A., Duchek, J.M., 2006. Assessing and improving the spatial accuracy in MEG source localization by depth-weighted minimum-norm estimates. Neuroimage 31 (1), 160–171.

Mathalon, D.H., Ford, J.M., Pfeifer, S.M., 2000. Trait and state aspects of P300 amplitude reduction in schizophrenia: a retrospective longitudinal study. Biol. Psychiat. 651.

McGorry, P.D., Mihalopoulos, C., Henry, L., Dakis, J., Jackson, H.J., Flaum, M., et al., 1995. Spurious precision: procedural validity of diagnostic assessment in psychotic disorders. Am. J. Psychiatry 152 (10), 1833–1838.

Kim, J., Kwon, H., Yu, K., Lee, Y., Kim, K., Kim, J., et al., 2014. Note: optical receiver system for 152-channel magnetoencephalography. Rev. Sci. Instrum. 85, 116105.

Kim, G.W., Chung, Y.C., Yang, J.C., Chung, G.H., Park, T.J., Kang, J.W., 2015. Neuroanatomical mechanism on the effect of distraction in working memory maintenance in patients with schizophrenia. J. Neuropsychiatr. Clin. Neurosci. 27 (1), e1–e9.

Kim, K., Neufang, S., Hanisch, C., Fink, G.R., Herperts-Dahlmann, B., 2006. Dysfunctional attentional networks in children with attention deficit/hyperactivity disorder: evidence from an event-related functional magnetic resonance imaging study. Biol. Psychiat. 59, 643–651.

Lazzarini, A., Anderson, J., Compton, E., Mcle, S., Leong, J., Meares, R., 1997. Single trial variability within the P300 (250–500 ms) processing window in adolescents with attention deficit hyperactivity disorder. Psychiatry Res. 73 (1), 91–101.

Lee, Y.H., Yu, K.K., Kwon, H., Kim, J.M., Kim, K., Park, Y.K., et al., 2009. A whole-head magnetoencephalography system with compact axial gradiometer structure. Supercond. Sci. Technol. 22, 045023.

Lin, F.H., Witzel, T., Ahlfors, S.P., Stufflebeam, S.M., Belliveau, J.W., Hämäläinen, M.S., 2006. Assessing and improving the spatial accuracy in MEG source localization by depth-weighted minimum-norm estimates. Neuroimage 31 (1), 160–171.

Matalon, D.H., Ford, J.M., Pfeifer, S.M., 2000. Trait and state aspects of P300 amplitude reduction in schizophrenia: a retrospective longitudinal study. Biol. Psychiat. 47, 434–449.

McFarland, R.W., Shenton, M.E., O’donnell, R.F., Faux, S.F., Kikinis, R., Nestor, P.G., Jolesz, F.A., 1993. Auditory P300 abnormalities and left posterior superior temporal gyrus volume reduction in schizophrenia. Arch. Gen. Psychiatry 50 (3), 190–197.

McGorry, P.D., Mihalopoulos, C., Henry, L., Dakis, J., Jackson, H.J., Flann, M., et al., 1995. Spurious precision: procedural validity of diagnostic assessment in psychotic disorders. Am. J. Psychiatry 152 (2), 220–223.

Neuhaus, A.H., Koehler, S., Oppen-Rhein, C., Urbanek, C., Hahn, E., Detting, M., 2007. Selective anterior cingulate cortex defect during conflict solution in schizophrenia: an event-related potential study. J. Psychiatr. Res. 41, 635–644.

Neuhaus, A.H., Hahn, E., Hahn, C., Ta, T.M.T., Oppen-Rhein, C., Urbanek, C., Detting, M., 2011a. Visual P3 amplitude modulation deficit in schizophrenia is independent of duration of illness. Schizophr. Res. 130 (1–3), 210–215.

Neuhaus, A.H., Popescu, F.C., Rentsch, J., Gallinat, J., 2013. Critical evaluation of auditory event-related potential deficits in schizophrenia: evidence from large-scale single-subject pattern classification. Schizophr. Bull. 40 (5), 1062–1071.

Nolte, G., Bai, O., Wheaton, J., Marz, J., Vorbsach, B., Hallert, M., 2004. Identifying true connectivity measures of magnetoencephalography (MEG) data. Front. Hum. Neurosci. 3, 73.

Gohel, B., Lee, P., Kim, M.Y., Kim, K., Jeong, Y., 2017. MEG based functional con- nection: application of ICA to alleviate signal leakage. Irbm 38 (3), 127–137.

Hahn, O.S., Ahn, J.H., Song, S.H., Cho, M.J., Kim, J.K., Bae, J.N., et al., 2000. Development of Korean version of structured clinical interview schedule for DSM-IV axis I disorder: interrater reliability. J. Korean Neuropsychiatr. Assoc. 39 (2), 362–372.

Gohel, B., Lee, P., Kim, M.Y., Kim, K., Jeong, Y., 2017. MEG based functional con- nection: application of ICA to alleviate signal leakage. Irbm 38 (3), 127–137.
Neurophysiol. 115 (10), 2292–2307.
Nunez, P.L., Srinivasan, R., Westdorp, A.F., Wijesinghe, R.S., Tucker, D.M., Silverstein, R.B., Cadusch, P.J., 1997. EEG coherency: I: statistics, reference electrode, volume conduction, Laplacians, cortical imaging, and interpretation at multiple scales. Electroencephalogr. Clin. Neurophysiol. 103 (5), 499–515.
Oberlin, B.G., Alford, J.L., Marrocco, R.T., 2005. Normal attention orienting but abnormal stimulus alerting and conflict effect in combined subtype of ADHD. Behav. Brain Res. 165, 1–11.
Oostenveld, R., Fries, P., Maris, E., Schoffelen, J.M., 2011. FieldTrip: open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data. Comput. Intell. Neurosci. 2011, 1.
Oribe, N., Hirano, Y., Kanba, S., Del Re, E., Seidman, L., Mesholam-Gately, R., et al., 2014. Progressive reduction of visual P300 amplitude in patients with first-episode schizophrenia: an ERP study. Schizophr. Bull. 41 (2), 460–470.
Pereira, F., Mitchell, T., Botvinick, M., 2009. Machine learning classifiers and fMRI: a tutorial overview. Neuroimage 45 (1), S199–S209.
Petley, L., Bardouille, T., Chiasson, D., Froese, P., Patterson, S., Newman, A., et al., 2018. Attentional dysfunction and recovery in concussion: effects on the P300m and contingent magnetic variation. Brain Inj. 32 (4), 464–473.
Ponser, M.I., Petersen, S.E., 1990. The attention system of the human brain. Annu. Rev. Neurosci. 13 (1), 25–42.
Quandt, F., Reichert, C., Hrinrichs, H., Heinze, H.J., Knight, R.T., Rieger, J.W., 2012. Single trial discrimination of individual finger movements on one hand: a combined MEG and EEG study. Neuroimage 59 (4), 3316–3324.
Rueda, M.R., Pozuelos, J.P., Cómbita, L.M., 2015. Cognitive neuroscience of attention. AIMS Neurosci. 2 (4), 183–202.
Schnack, H.G., Kahn, R.S., 2016. Detecting neuroimaging biomarkers for psychiatric disorders: sample size matters. Front. Psychiatry 7, 50.
Schoffelen, J.M., Gross, J., 2009. Source connectivity analysis with MEG and EEG. Hum. Brain Mapp. 30 (6), 1857–1865.
Shen, H., Wang, L., Liu, Y., Hu, D., 2010. Discriminative analysis of resting-state functional connectivity patterns of schizophrenia using low dimensional embedding of fMRI. Neuroimage 49 (4), 3110–3121.
Shim, M., Kim, D.W., Lee, S.H., Im, C.H., 2014. Disruptions in small-world cortical functional connectivity network during an auditory oddball paradigm task in patients with schizophrenia. Schizophr. Res. 156 (2–3), 197–203.
Shim, M., Hwang, H.J., Kim, D.W., Lee, S.H., Im, C.H., 2016. Machine-learning-based diagnosis of schizophrenia using combined sensor-level and source-level EEG features. Schizophr. Res. 176 (2–3), 314–319.
Souza, V.B., Mui, W.J., Walker, M.T., Glabus, M.F., Roxburgh, H.M., Sharp, C.W., et al., 1995. Auditory P300 event-related potentials and neuropsychiological performance in schizophrenia and bipolar affective disorder. Biol. Psychiatry 37 (5), 300–310.
Tadel, F., Baillet, S., Mosher, J.C., Pantazis, D., Leahy, R.M., 2011. Brainstorm: a user-friendly application for MEG/EEG analysis. Comput. Intell. Neurosci. 2011, 8.
Theodoridis, S., Koutroumbas, K., 2008. Pattern recognition. IEEE Trans. Neural Netw. 19 (2), 376.
Theodoridis, S., Pikrakis, A., Koutroumbas, K., Cavouras, D., 2010. Introduction to Pattern Recognition: a Matlab Approach. Academic Press.
Vapnik, V., 2013. The Nature of Statistical Learning Theory. Springer science & business media.
Yeom, T.H., Park, Y.S., Ob, K.J., Lee, Y.H., 1992. Korean version Wechsler adult intelligence scale. Seoul: Korean Guidance 4, 13–28.
Yu, Q., Erhardt, E.B., Sui, J., Du, Y., He, H., Hjelm, D., et al., 2015. Assessing dynamic brain graphs of time-varying connectivity in fMRI data: application to healthy controls and patients with schizophrenia. Neuroimage 107, 345–355.