



## OPEN ACCESS

## EDITED BY

Elisabetta C. del Re,  
Harvard Medical School, United States

## REVIEWED BY

Mohammad S. E. Sendi,  
Harvard Medical School, United States  
Rebekah Trotti,  
Beth Israel Deaconess Medical Center  
and Harvard Medical School, United States

## \*CORRESPONDENCE

Ruben Perellón-Alfonso  
✉ ruben.perellon@ub.edu  
Jurij Bon  
✉ jurij.bon@mf.uni-lj.si

RECEIVED 13 April 2023

ACCEPTED 04 September 2023

PUBLISHED 25 September 2023

## CITATION

Perellón-Alfonso R, Oblak A, Kuclar M, Škrlić B, Pileckyte I, Škodlar B, Pregelj P, Abellaneda-Pérez K, Bartrés-Faz D, Repovš G and Bon J (2023) Dense attention network identifies EEG abnormalities during working memory performance of patients with schizophrenia.

*Front. Psychiatry* 14:1205119.

doi: 10.3389/fpsy.2023.1205119

## COPYRIGHT

© 2023 Perellón-Alfonso, Oblak, Kuclar, Škrlić, Pileckyte, Škodlar, Pregelj, Abellaneda-Pérez, Bartrés-Faz, Repovš and Bon. 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.

# Dense attention network identifies EEG abnormalities during working memory performance of patients with schizophrenia

Ruben Perellón-Alfonso<sup>1,2\*</sup>, Aleš Oblak<sup>3</sup>, Matija Kuclar<sup>4</sup>, Blaž Škrlić<sup>5</sup>, Indre Pileckyte<sup>6</sup>, Borut Škodlar<sup>3,4</sup>, Peter Pregelj<sup>3,4</sup>, Kilian Abellaneda-Pérez<sup>1,2,7</sup>, David Bartrés-Faz<sup>1,2</sup>, Grega Repovš<sup>8</sup> and Jurij Bon<sup>3,4\*</sup>

<sup>1</sup>Faculty of Medicine and Health Sciences, and Institute of Neurosciences, University of Barcelona, Barcelona, Spain, <sup>2</sup>Institute of Biomedical Research August Pi i Sunyer (IDIBAPS), Barcelona, Spain, <sup>3</sup>University Psychiatric Clinic Ljubljana, Ljubljana, Slovenia, <sup>4</sup>Department of Psychiatry, Faculty of Medicine, University of Ljubljana, Ljubljana, Slovenia, <sup>5</sup>Jožef Stefan Institute, Ljubljana, Slovenia, <sup>6</sup>Center for Brain and Cognition, Pompeu Fabra University, Barcelona, Spain, <sup>7</sup>Institut Guttmann, Institut Universitari de Neurorehabilitació Adscrit a la UAB, Barcelona, Spain, <sup>8</sup>Department of Psychology, Faculty of Arts, University of Ljubljana, Ljubljana, Slovenia

**Introduction:** Patients with schizophrenia typically exhibit deficits in working memory (WM) associated with abnormalities in brain activity. Alterations in the encoding, maintenance and retrieval phases of sequential WM tasks are well established. However, due to the heterogeneity of symptoms and complexity of its neurophysiological underpinnings, differential diagnosis remains a challenge. We conducted an electroencephalographic (EEG) study during a visual WM task in fifteen schizophrenia patients and fifteen healthy controls. We hypothesized that EEG abnormalities during the task could be identified, and patients successfully classified by an interpretable machine learning algorithm.

**Methods:** We tested a custom dense attention network (DAN) machine learning model to discriminate patients from control subjects and compared its performance with simpler and more commonly used machine learning models. Additionally, we analyzed behavioral performance, event-related EEG potentials, and time-frequency representations of the evoked responses to further characterize abnormalities in patients during WM.

**Results:** The DAN model was significantly accurate in discriminating patients from healthy controls, ACC = 0.69, SD = 0.05. There were no significant differences between groups, conditions, or their interaction in behavioral performance or event-related potentials. However, patients showed significantly lower alpha suppression in the task preparation, memory encoding, maintenance, and retrieval phases  $F(1,28) = 5.93$ ,  $p = 0.022$ ,  $\eta^2 = 0.149$ . Further analysis revealed that the two highest peaks in the attention value vector of the DAN model overlapped in time with the preparation and memory retrieval phases, as well as with two of the four significant time-frequency ROIs.

**Discussion:** These results highlight the potential utility of interpretable machine learning algorithms as an aid in diagnosis of schizophrenia and other psychiatric disorders presenting oscillatory abnormalities.

#### KEYWORDS

schizophrenia, working memory (WM), contralateral delay activity (CDA), electroencephalography (EEG), dense attention network (DAN)

## 1. Introduction

Schizophrenia is a severe neuropsychiatric disorder with a global prevalence of 0.28% and a significant socioeconomic burden (1). The symptoms of schizophrenia can be divided into positive (i.e., hallucinations, delusions and disorganized thinking) and negative [i.e., decreased emotional expression, social withdrawal, and cognitive impairments of memory and executive functions; (2, 3)]. Schizophrenia is thought to be a neurodevelopmental disorder caused by interaction of genetic and early environmental risk factors (4–6), resulting in impaired large-scale connectivity (7, 8) and aberrant brain activity (9, 10). Pathophysiological changes include altered dopamine and glutamate neurotransmission, which is thought to be related to a disruption in the balance of excitation and inhibition in cortical microcircuits, contributing to altered synchronization of neuronal oscillations (11).

Impairment of working memory (WM) is a core cognitive deficit in schizophrenia that significantly correlates with functional capacity and outcome (12), and has been proposed as a warning sign of conversion to psychosis (13). WM is often defined as a system with limited capacity for the temporary storage and manipulation of representations of information necessary to guide behavior in complex goal-directed tasks such as comprehension, learning, and reasoning (14), and it overlaps with other cognitive domains such as attention and executive function (15). In schizophrenia, deficits can be observed in all WM subprocesses and stimulus types (16), and are associated with impairments in proactive cognitive control [i.e., the ability to actively represent goal information in working memory to guide behavior (16)] or attention hyperfocus [i.e., an abnormally narrow and intense focusing of processing resources; (17)]. Deficits have also been detected, in high-functioning patients with preserved WM performance, in the form of increased reaction time variability (18), which has been interpreted as impaired information processing. The visual modality of WM is particularly relevant in schizophrenia, as it strongly correlates with measures of higher cognitive functions and, according to some estimates, may account for up to 40% of the cognitive deficit in patients with schizophrenia (19).

Working memory tasks can be constructed to engage different WM subprocesses either simultaneously [e.g., N-back tasks; (20, 21)] or sequentially [e.g., verbal span tasks, visuospatial change detection tasks; (22, 23)]. Sequential tasks are particularly useful to probe behavioral performance and brain activity during separate time periods of the WM task corresponding to task preparation, encoding, maintenance, and retrieval of information, all of which have been shown to be affected in schizophrenia (24–26).

Electroencephalographic (EEG) studies of event-related potentials (ERPs) elicited during working memory tasks, have shown abnormalities in electrical activity during early evoked responses and late, cognition-related components of schizophrenia patients (27). In visual WM, a lateralized change detection task (23) elicits a corresponding ERP component, the contralateral delay activity (CDA), which has been shown to be closely related to WM capacity and is modulated by load (28). CDA studies in schizophrenia have shown that visual WM capacity is lower, relative to healthy controls, and that patients also show specific impairments in attention control during the task (29). In addition to ERP abnormalities, studies also found changes on synchronized neuronal oscillations in several frequency bands. Specifically, gamma (>30 Hz), which is involved in sensory processing (30) and maintenance of WM information (31), shows lack of synchronization in schizophrenia patients during WM tasks [e.g., (32)]. Theta (4–7 Hz), which supports long range connectivity and coordination of WM items (33), has been reported to be abnormally high during resting state (34) and decoupled from gamma during WM performance (35). Finally, alpha (8–12 Hz) desynchronization (also known as alpha suppression), which reflects the active inhibition of task-irrelevant information (36, 37), has been shown to be impaired in schizophrenia patients and individuals at risk of psychosis during working memory and oddball tasks (24, 38–41).

While these studies have significantly advanced our understanding of the neurophysiological basis of schizophrenia, they typically rely on univariate statistical methods that, while suitable for group-level comparisons, are insufficient for the purposes of individual diagnosis within the framework of precision psychiatry (42, 43). Moreover, these studies highlight the fact that schizophrenia exhibits heterogenic symptoms and intricate neurophysiological foundations that cannot be attributed to a single brain area or neural process and that might be shared across psychiatric disorders (44). This complexity makes precise differential diagnosis and neurophysiological characterization of individual patients challenging. To confront these challenges, the field of psychiatry has increasingly turned to machine learning, a class of artificial intelligence approaches where algorithms are designed to make successful predictions without explicit programming (45). A growing number of studies have used EEG data to successfully classify patients and controls with high accuracy (46–52). These results have the potential to yield clinically translatable improvements in diagnosis. However, the best performance is often achieved by deep convolutional neural network models, which are said to be “black boxes,” since there is no straightforward solution to disentangle how the

algorithm transforms the input data to model a particular output (53). This characteristic limits their utility for investigating the neurophysiological substrate of schizophrenia and identifying suitable biomarkers for early detection, consequently, more transparent and interpretable deep learning models are needed to fill this gap (54).

A promising alternative is the dense attention network (DAN), a type of deep learning model based on the attention model (55), a simple mechanism that scatters input signals and highlights only the parts of the feature space that are relevant to the task at hand. Crucially, the attention layers can output a probability distribution over the input space, thus providing an insight into the inner workings of the neural network, in the form of a one-to-one mapping of the relative contribution of each feature in the input space (56, 57).

Here, we took a data-driven machine learning approach to determine the distribution of EEG signatures specific to schizophrenia patients over the time course of a visuospatial change detection task. Based on previous encouraging reports (46–52), we hypothesized that machine learning could be used to successfully classify patients from controls based on EEG alone. We chose an interpretable subtype of machine learning based on the attention model (55), with the hypothesis that specific temporal signatures would be most discriminative of patients and controls. We hypothesized that differences between schizophrenia patients and control subjects would also be evident using univariate statistical methods, particularly on oscillatory activity related to attention control (i.e., in the alpha frequency band), which has been consistently reported to be impaired in patients with schizophrenia, and would be most prominent during the task preparation, encoding, maintenance or memory retrieval phases of the task time-course (24, 39, 41). Finally, we expected this significantly different task segments to overlap with the features found to be most discriminative by the DAN model. This correspondence is of crucial importance if machine learning is to become not just a diagnostic aid, but also a tool capable of proving the neurophysiological substrate of schizophrenia (54).

## 2. Materials and methods

### 2.1. Study participants

Fifteen patients with a mean age of 28.1 years,  $SD = 3.9$ , and an average of 13.4 years of education,  $SD = 1.1$ , were recruited from the Department for Psychotherapy of Psychotic Disorders at the University Psychiatric Clinic Ljubljana (see [Supplementary Table 1](#) for descriptive statistics on demographics). All participants included in the study were male, due to a lack of a representative number of female participants available at the time of recruitment. All patients had a diagnosis of schizophrenia (12 subjects) or schizoaffective disorder (3 subjects). The diagnoses were confirmed according to the DSM-IV criteria by experienced clinicians (BŠ and JB) involved in the study. At the time of the experiment, all patients were taking second generation antipsychotic medication and were in stable symptomatic remission, with an average PANSS score (58) of 77.1 ( $SD = 15.3$ ), and were cleared for inclusion in psychodynamic group psychotherapy. The patients' mean duration

of illness was 6.1 years ( $SD = 3.3$ ), and the mean number of hospitalizations was 2.9 ( $SD = 2.1$ ). For further clinical details, see [Supplementary Table 1](#).

Additionally, we recruited a control group of 15 male participants of comparable age,  $M = 26.8$  years,  $SD = 5.5$ , and years of education,  $M = 14.4$ ,  $SD = 1.2$ . The study was approved by the Medical Ethics Committee of the Republic of Slovenia and all participants signed an informed consent form according to the Declaration of Helsinki.

### 2.2. Visual working memory task and EEG recording

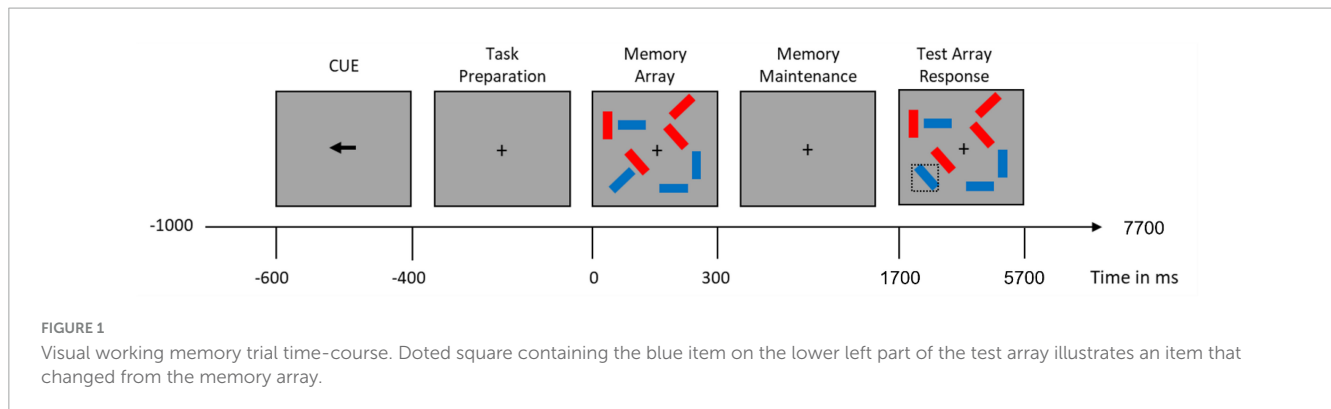
A lateralized change detection task with distractors (23, 59) was implemented using PsychoPy (60). First, participants were shown an arrow cue for 200 ms, the direction of which indicated to which half of the visual field they should direct their attention. This was followed first by a fixation cross shown for 400 ms and then by a memory array shown for 300 ms. The memory array consisted of 2 or 4 rectangles (in each half of the visual field). The rectangles were colored either blue, or blue and red (in the distractor condition), and were shown in one of 4 possible orientations ( $0^\circ$ ,  $45^\circ$ ,  $90^\circ$ , or  $135^\circ$ ). Participants were asked to remember the orientations of the blue rectangles shown on the cued side of the visual field. The presentation of the memory array was followed by a delay of 1400 ms before the presentation of a test array, that remained for 4 s and was then followed by 2 s without any stimulus. The test array was either identical to the memory array or with only one of the randomly selected (blue) rectangles on the cued side changing its orientation in half of the trials. The participants' task was to indicate whether any of the target items had changed by pressing the corresponding button on a response box ([Figure 1](#)).

There were three task conditions, differing in the number of target items and the presence of a distractor:

1. A condition with two blue rectangles shown on each side (low memory load; condition 2),
2. A condition with four blue rectangles shown on each side (high memory load: condition 4), and
3. A condition with two blue and two red rectangles shown on each side (distractor condition; condition 2+2).

In the 2+2 condition, participants had to successfully inhibit the two red distractor rectangles presented together with the two blue memory rectangles. The trials belonging to the different conditions were interleaved within a block. Participants were familiarized with the task during the practice trials, and the experimenter ensured that they all performed with at least 70% accuracy.

Participants performed 200 trials for each of the three conditions in an electrically shielded and soundproofed room while seated in a comfortable chair in front of a cathode ray monitor. Throughout the task, the EEG signal was recorded using four BrainAmp amplifiers connected to a 128-channel actiCAP system with active electrodes in a standard montage (Brain Products GmbH, Munich, Germany). The EEG was recorded with a 2000 Hz low-pass filter and digitized at a sampling rate of 500 Hz.



## 2.3. Working memory task performance metrics

To compare behavioral task performance between groups we computed memory capacity index  $K$  (61) and intra-individual reaction time variability (see [Supplementary Table 2](#) for detailed descriptive statistics of task performance).

### 2.3.1. Working memory capacity

WM capacity index  $K$  was calculated for each subject and condition using the Pashler variant of the formula appropriate for a whole-display variant of a change detection task (61):

$$K = N \left( \frac{HR - FAR}{1 - FAR} \right)$$

Where  $HR$  is the hit rate,  $FAR$  is the false alarm rate, and  $N$  is the number of to-be-remembered items.

### 2.3.2. Intra-individual reaction time variability

Rentrop and colleagues (18) reported schizophrenia patients with relatively well-preserved WM performance still showed higher intraindividual variability in reaction times. Therefore, we compared the coefficient of variation of reaction times between the two groups, which was defined as the ratio of the standard deviation to the mean of the reaction times.

## 2.4. EEG preprocessing

Electroencephalographic data were preprocessed using EEGLAB functions (62) and custom-made MATLAB (The MathWorks Inc., Massachusetts, USA) scripts. Data were first filtered with a high-pass filter with a 0.5 Hz frequency cutoff, then the line frequency noise was removed from the signal using the CleanLine algorithm (63). Visual inspection was aided by statistical thresholding based on variance and Kurtosis to identify bad channels,  $M = 15.3$ ,  $SD = 4.6$ . Next data were referenced to the average of the mastoid channels (i.e., TP9 and TP10) and segmented into epochs around the onset of the memory array (−1,000 ms to 4,500 ms). At this point, epoched data were visually inspected, and epochs that contained obvious artifacts (e.g., high-frequency or muscular artifacts) were removed. Because lateral eye movements would impact the magnitude of the CDA,

electrooculogram channels were visually inspected for the time period from the presentation of arrow cue to the presentation of memory array, and all epochs with eye blinks or horizontal eye movements in this period were also discarded, bringing the total average of epochs removed to 47.6.3,  $SD = 29.5$ . Next, the AMICA algorithm (64) was used to identify and then remove any remaining artifactual independent components  $M = 6.4$ ,  $SD = 2.4$ . Last, the channels previously removed from the data were spline interpolated based on the signal from the neighboring electrodes.

## 2.5. Machine learning methods and empirical evaluation

The aim of this analysis was to investigate the potential of machine learning methods to discriminate patients from controls. Given the heterogeneous nature of schizophrenia, our aim was to produce a model capable of discriminating between patients and controls without relying on any specific clinical data, leveraging only EEG data that has been preprocessed using relatively simple and well-established procedures. The dense attention network model was deliberately chosen because it retains a sufficient level of interpretability to explain which events were most important in distinguishing patients from controls over the time course of the experiment (57). The dimensionality reduction of the data, the construction of the DAN architecture, and its evaluation, were performed using in-house methods (a detailed description can be found in [Supplementary material](#); scripts and data used to design, train, and evaluate the different machine learning models<sup>1</sup>). Briefly, the dimensionality of the preprocessed data was reduced from 4d to 1d by incremental stepwise averaging of three of the four original dimensions (i.e., 3 conditions left and 3 right, 128 channels, 2,750 time-points and 153 trials on average):

$$S_{\gamma} = \frac{1}{RNZ} \sum_{\mu=1}^R \left( \sum_{v=1}^N \left( \sum_{\sigma=1}^Z M_{\mu v \gamma \sigma} \right) \right)$$

Where  $\mu$ ,  $v$ ,  $\gamma$ , and  $\sigma$  stand for condition, channel, time, and trial, respectively. In this way, a 1d array was created for each subject while preserving the temporal characteristics of the data. This simplified dimensionality of the data allowed us to train a dense

<sup>1</sup> <https://gitlab.com/MaticKu/shizo>

attention network model (DAN). The final input dataset used in the model was numeric and consisted of 30 instances (i.e., number of subjects), each described by 2,750 features (i.e., corresponding to the preserved time dimension after the incremental stepwise averaging of the original 4d EEG data set).

Empirical evaluation of model performance consisted of leave-one-out cross-validation repeated ten times for each model. For the DAN model there were 160 possible configurations evaluated. We used the Adam optimization algorithm (65) and we considered the following parameters: dropout rate (0.01, 0.05, 0.2, and 0.5) hidden layer size (16, 32, 64, and 128), number of epochs (2, 4, 8, 16, and 32) and learning rate (0.001 and 0.0001).

The performance of the model was compared with the performance of other simpler architectures (i.e., linear regression, radio frequency machine learning, support vector machine, radial basis function and k-nearest neighbor), and common deep learning models (convolutional neural network and feed forward neural network).

We report the average and standard deviation of the resulting accuracy, precision and recall from these iterations. We also report the  $F_1$  scores, computed as follows:

$$F_1 = 2 \frac{\text{precision} \cdot \text{recall}}{\text{precision} + \text{recall}}$$

All models were implemented using the PyTorch deep learning library (66) and evaluated on a Tesla graphics card accelerator (Nvidia Corp. Santa Clara, USA).

Throughout training of the DAN model, a bijection is maintained with the input space (i.e., the attention layer corresponds to the input space in a one-to-one relationship). Therefore, we were able to use the attention layer's output directly as a probability distribution over the input space. This attention value vector quantifies the contribution of each feature (EEG time-point in the WM task time-course) in the distinction between patients and controls.

## 2.6. Event related potential analysis

In order to capture the electrophysiological correlate of WM capacity, we computed the contralateral delayed activity (CDA) as the difference between the contralateral and ipsilateral (relative to the cued side for the memory array) ERP waveforms using an established procedure (23, 28). For each subject, the mean amplitude of the resulting CDA difference curves was measured for the average of all parieto-occipital electrodes and the time segment from 500 ms to 900 ms after the presentation of the memory array (Figure 2). The resulting mean CDA amplitude data were used for further statistical analysis.

## 2.7. Time-frequency analysis

To compare the oscillatory dynamics between patients and controls, throughout the trial time course, we performed a time-frequency analysis of total power (i.e., comprising induced and evoked power) for the epoched data. Data was decomposed into the time-frequency domain by convolving a set of complex Morlet wavelets from 1 Hz to 60 Hz, in steps of 1 Hz, with a logarithmically

spaced wavelet width of 4–10 cycles. The resulting time-frequency maps were normalized as the decibel (db) change from baseline (i.e., –850 to –650 ms from the memory array presentation). Time-frequency regions of interest (ROIs) were then determined based on the main WM task phases and frequency bands. Specifically, the time segments of interest (i.e., ROIs x-axis) were, *preparation* for the task (–400 to 0 ms), *encoding* (0–300 ms), *maintenance* (300–1,700 ms) and *retrieval* of the memory array (1,700–3,060 ms). The end of the time window for the memory retrieval phase was chosen based on the average reaction time plus one standard deviation in the slower group (i.e., patients; Supplementary Table 2). We included four frequency bands of interest (i.e., ROIs y-axis), theta (4–7 Hz), alpha (8–12 Hz), beta (13–29 Hz) and gamma (30–60 Hz) to explore possible group differences across the frequency spectrum. For further statistical analysis, the average of all time-frequency data points within each ROI for each subject and condition was used.

## 2.8. Statistical analysis

For behavioral analysis a mixed design ANOVA was used for memory capacity K and another for reaction-time variability. For ERP analysis one mixed design ANOVA was used. For time-frequency total power four mixed design ANOVAs were used, one for each frequency band. All statistical tests were implemented in R (67). Greenhouse-Geisser correction was used in case of sphericity violations. All *post-hoc* comparisons were performed with paired or Welch's *t*-tests with Bonferroni corrections.

## 3. Results

### 3.1. Memory capacity K

To test for differences in memory capacity between groups and conditions we used a mixed design ANOVA with a within-subject factor *condition* (condition 2, condition 2+2 and condition 4) and a between-subject factor *group* (patient vs. control). The test revealed no significant main effect of *group*,  $F(1,28) = 2.21$ ,  $p = 0.149$ ,  $\eta^2 = 0.043$ , but there was a significant effect of *condition*,  $F(2,56) = 33.74$ ,  $p < 0.001$ ,  $\eta^2 = 0.344$ . *Post-hoc* analysis with a pairwise *t*-test revealed that in condition 4,  $M = 2.57$ ,  $SD = 0.86$ , memory capacity K was significantly higher than in conditions 2,  $M = 1.81$ ,  $SD = 0.14$ ,  $p < 0.001$ ,  $d = 1.236$  and 2+2,  $M = 1.76$ ,  $SD = 0.32$ ,  $p < 0.001$ ,  $d = 1.245$ . There was no interaction between *group* and *condition*,  $F(2,56) = 1.13$ ,  $p = 0.329$ ,  $\eta^2 = 0.017$ .

### 3.2. Reaction time's coefficient of variation

A mixed design ANOVA with a within-subject factor of *condition* (condition 2, condition 2+2 and condition 4) and a between-subject factor of *group* (patient vs. control) was used to test for differences in the coefficient of variation. The test revealed no significant main effect *condition*,  $F(2,56) = 1.08$ ,  $p = 0.345$ ,  $\eta^2 = 0.007$ , *group*,  $F(1,28) = 2.49$ ,  $p = 0.125$ ,  $\eta^2 = 0.068$ , or their interaction  $F(2,56) = 1.14$ ,  $p = 0.326$ ,  $\eta^2 = 0.007$ .

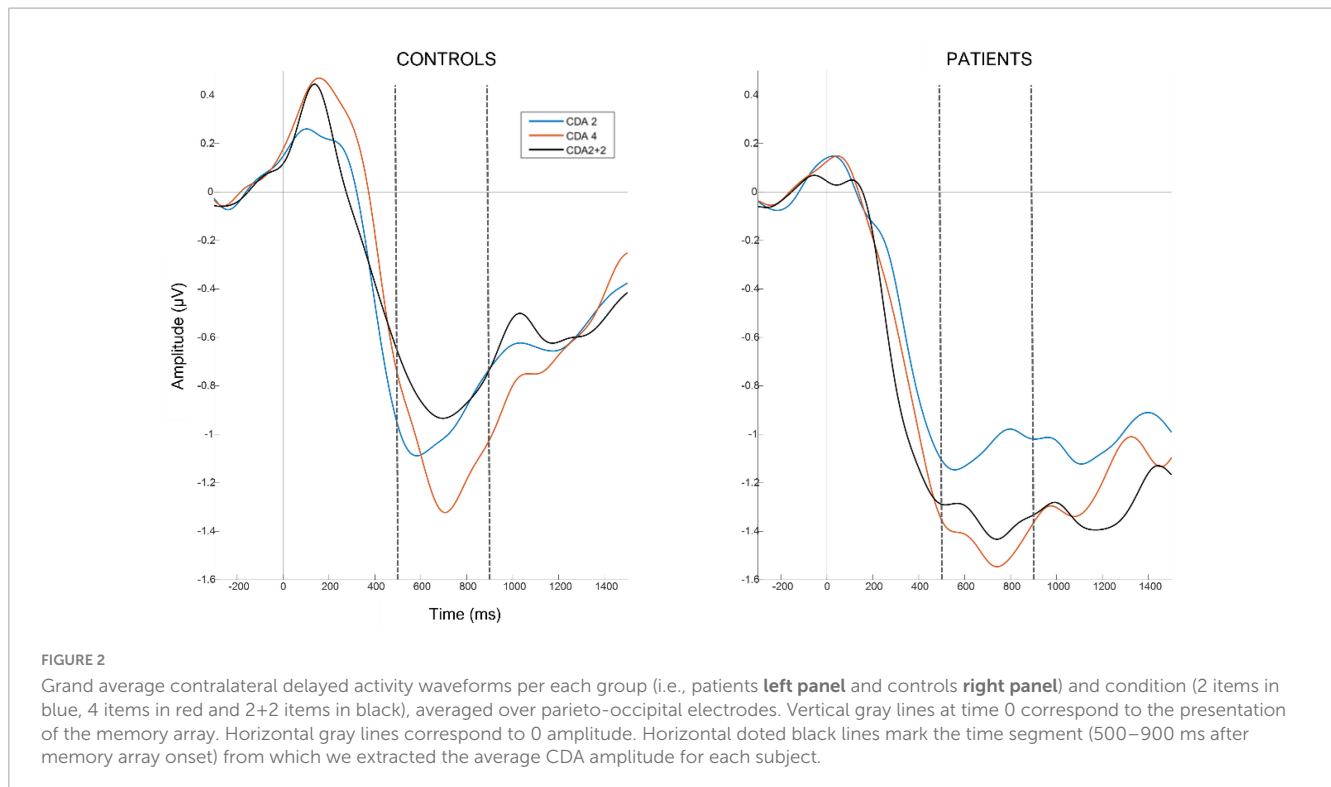


FIGURE 2

Grand average contralateral delayed activity waveforms per each group (i.e., patients **left panel** and controls **right panel**) and condition (2 items in blue, 4 items in red and 2+2 items in black), averaged over parieto-occipital electrodes. Vertical gray lines at time 0 correspond to the presentation of the memory array. Horizontal gray lines correspond to 0 amplitude. Horizontal dotted black lines mark the time segment (500–900 ms after memory array onset) from which we extracted the average CDA amplitude for each subject.

### 3.3. Contralateral delay activity's mean amplitude

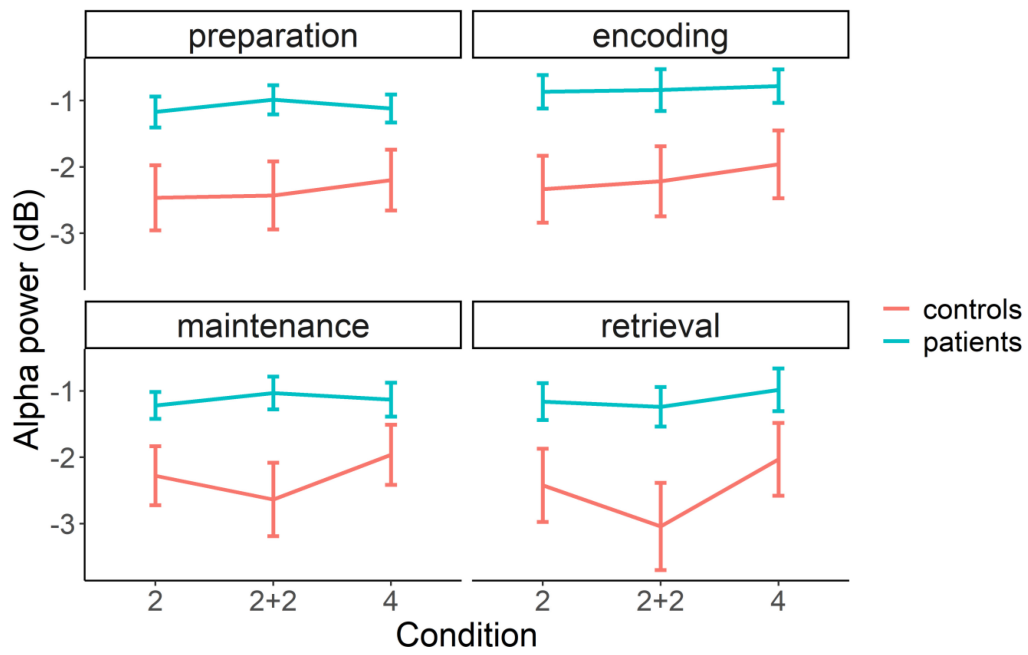
To examine the differences in mean CDA amplitude between the two groups and the three experimental conditions, we used a mixed design ANOVA with a within-subject factor *condition* (condition 2, condition 2+2, and condition 4) and a between-subject factor of *group* (patient vs. control). This analysis revealed no significant main effect of *group*  $F(1,28) = 0.22, p = 0.642, \eta^2 = 0.004$ , *condition*,  $F(2,56) = 3.12, p = 0.065, \eta^2 = 0.048$ , or their interaction  $F(2,56) = 0.41, p = 0.617, \eta^2 = 0.007$ .

### 3.4. Time-frequency power representation

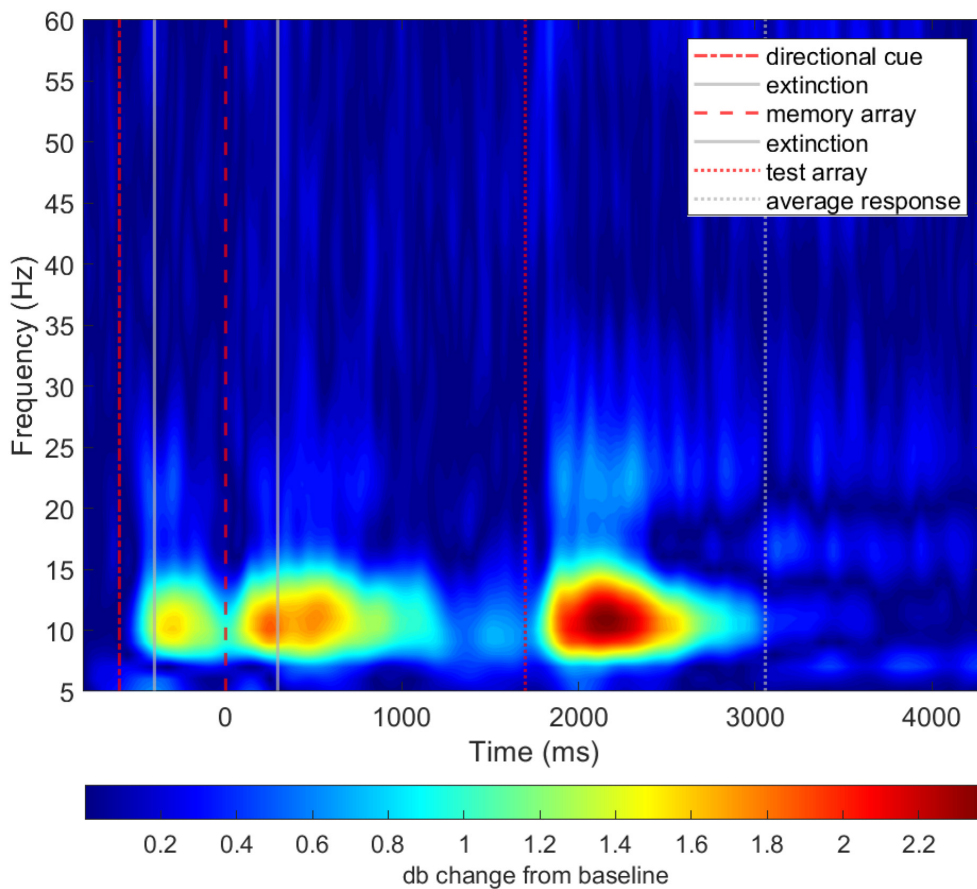
To investigate possible differences between patients and controls in overall power at four frequency bands, we used a mixed design ANOVA for each frequency band with a within-subject factor *condition* (condition 2, condition 2+2, and condition 4), a within-subject factor WM task *phase* (preparation, encoding, maintenance, retrieval), and a between-subject factor *group* (patient vs. control). This analysis revealed a significant *group* effect,  $F(1,28) = 5.93, p = 0.022, \eta^2 = 0.149$ , and interaction between *group*, *condition* and task *phase*,  $F(6,168) = 2.20, p = 0.046, \eta^2 = 0.002$ , only for the alpha frequency band (for complete results in the theta, beta, and gamma bands see [Supplementary Section 5](#) and [Supplementary Figure 3](#)). *Post-hoc* analysis with Welch's *t*-test revealed that patients had overall higher alpha power (i.e., less alpha suppression),  $M = -1.04, SD = 0.89$ , than controls,  $M = -2.33, SD = 1.84, t(20.24) = 2.43, p = 0.024, d = 0.889$ .

To unravel the triple interaction, we ran an additional mixed design ANOVA for each of the four WM task phases, with a within-subject factor *condition* (condition 2, condition 2+2, and condition 4) and a between-subject factor *group* (patients vs. controls). We found that groups differed in each of the 4 WM task phases ([Figure 3](#)). During task preparation,  $F(1,28) = 6.16, p = 0.019, \eta^2 = 0.167$ , patients,  $M = -1.09, SD = 0.81$ , had higher alpha power than controls,  $M = -2.36, SD = 1.81, t(19.41) = 2.48, p = 0.022, d = 0.906$ . During memory encoding,  $F(1,28) = 5.73, p = 0.024, \eta^2 = 0.159$ , patients,  $M = -0.83, SD = 1.00$ , had higher alpha power than controls,  $M = -2.17, SD = 1.93, t(21.04) = 2.39, p = 0.026, d = 0.874$ . During memory maintenance, we found a significant interaction between group and condition,  $F(2,56) = 3.28, p = 0.045, \eta^2 = 0.013$ . A follow-up analysis revealed that the interaction was driven by the decrease in alpha power from condition 2+2,  $M = -1.03, SD = 0.96$ , to condition 4,  $M = -1.13, SD = 0.99$ , in patients, and increase in alpha power from condition 2+2,  $M = -2.63, SD = 2.14$ , to condition 4,  $M = -1.96, SD = 1.75$ , in controls. Finally, during memory retrieval,  $F(1,28) = 4.73, p = 0.038, \eta^2 = 0.133$ , patients,  $M = -1.13, SD = 1.13$ , had higher alpha power than controls,  $M = -2.50, SD = 2.16, t(21.12) = 2.17, p = 0.041, d = 0.794$ .

To summarize, there was a significant difference between patients and controls in the alpha frequency band, but not in other frequency bands. This difference was observed in all four task phases. In support to these univariate results, the difference between patients and controls can also be visually evaluated in time-frequency difference maps, where it can be seen that the differences in alpha band peak after the presentation of the main task stimuli, namely, directional cue, memory array and test array ([Figure 4](#)).



**FIGURE 3** Average alpha power baseline normalized in patient (blue line) and control (red line) groups in three different conditions (2, 2+2, and 4) and four WM task phases (preparation, encoding, maintenance, and retrieval). The error bars represent standard error of the mean.



**FIGURE 4** Unthresholded average time-frequency representation of total power difference of patients and controls (i.e., patients–controls).

TABLE 1 Summary of the empirical evaluation average results for the different machine learning architectures used.

Model	Accuracy	F1	Recall	Precision
DAN	0.69 ± (0.05)	0.71 ± (0.07)	0.77 ± (0.11)	0.67 ± (0.03)
FFNN	0.71 ± (0.03)	0.72 ± (0.03)	0.73 ± (0.05)	0.70 ± (0.02)
CNN	0.69 ± (0.05)	0.72 ± (0.05)	0.81 ± (0.05)	0.66 ± (0.04)
LR	0.65 ± (0.0)	0.67 ± (0.0)	0.69 ± (0.0)	0.64 ± (0.0)
SVM linear	0.62 ± (0.0)	0.62 ± (0.0)	0.62 ± (0.0)	0.62 ± (0.0)
SVM poly	0.62 ± (0.0)	0.58 ± (0.0)	0.54 ± (0.0)	0.64 ± (0.0)
KNN	0.69 ± (0.0)	0.71 ± (0.0)	0.77 ± (0.0)	0.67 ± (0.0)
RF	0.56 ± (0.08)	0.58 ± (0.09)	0.62 ± (0.11)	0.55 ± (0.07)
SVM rbf	0.58 ± (0.0)	0.62 ± (0.0)	0.69 ± (0.0)	0.56 ± (0.0)

Standard deviations in parenthesis. CNN, convolutional neural network; DAN, dense attention network; FFNN, feed forward neural network; KNN, K-nearest neighbor; LR, linear regression; RF, radio frequency machine learning; SVM, support vector machine; rbf, radial basis function.

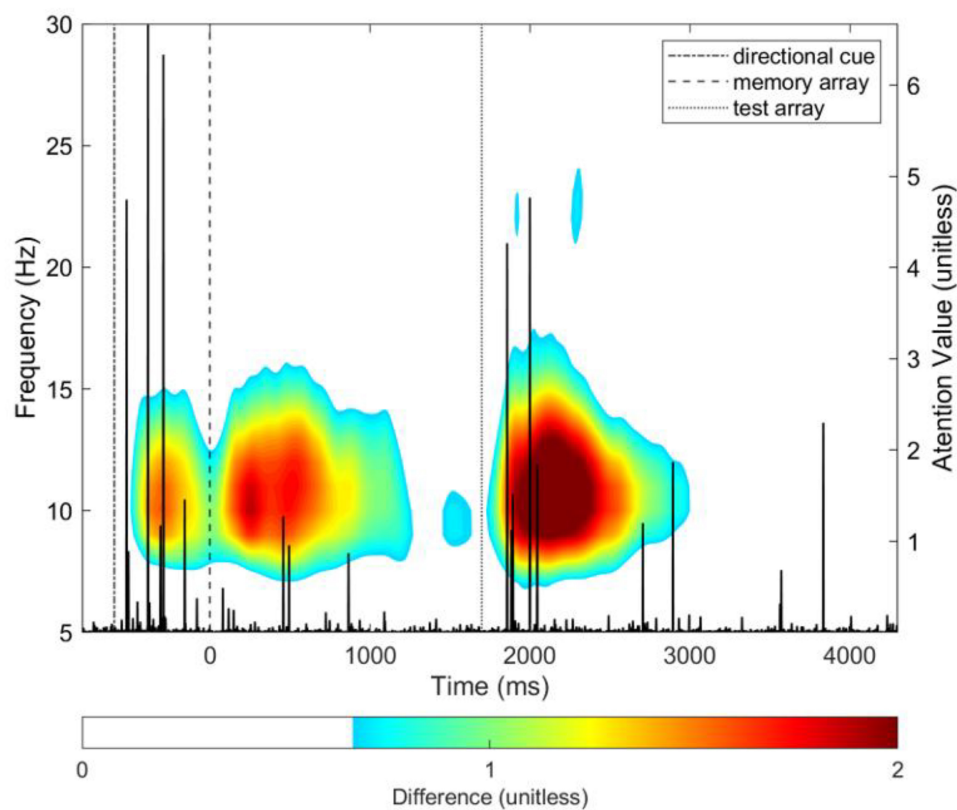


FIGURE 5

DAN attention value vector (in black; y-axis scale on the right side of plot) overlaid on the group difference time-frequency total power time-frequency map for the average of all parieto-occipital electrodes. The difference was computed by subtracting the grand average map of total power of controls from patients (i.e., patients–controls). The colormap is thresholded, for visualization purposes only, to show no color for values lower than 30% of the maximum difference value of 2. See Figure 4 for unthresholded map.

### 3.5. Machine learning performance and interpretation

The results from the empirical evaluation show that the DAN model consistently demonstrated accuracy in discriminating patients from controls significantly above chance,  $ACC = 0.69$ ,  $SD = 0.05$ ,  $F_1 = 0.71$ ,  $SD = 0.07$ ,  $Recall = 0.77$ ,  $SD = 0.11$ ,  $PRC = 0.67$ ,  $SD = 0.03$  (Table 1). This model outperformed simpler architectures, such as the support vector machine, while

performing similarly to other common deep neural network models, such as the convolutional neural network. Full results for all models tested under different subsets of the feature space (i.e., task conditions) can be found in Supplementary Figure 1 and Supplementary Table 3.

Projecting the DAN attention value over the WM task timeline and onto the time-frequency map of group differences shows that the time points most relevant to the model's discrimination between groups (i.e., attention value) overlap with time-frequency ROIs that were found to be significantly different between groups (Figure 5),



and correspond to the task preparation and memory retrieval phases of the WM task.

## 4. Discussion

In this study, we investigated the WM performance and associated EEG signatures of schizophrenia patients and compared them to healthy controls. The results show that in our sample neither WM performance, measured by memory index K and reaction time variability, nor CDA amplitude showed a significant difference between patients and controls. However, statistical analysis in the time-frequency domain revealed, a significant group effect in all time segments of interest (task preparation, memory encoding, maintenance and retrieval) in the alpha-band range (8–12 Hz). We demonstrated that a simple dimensionality reduction procedure consisting of incremental stepwise averaging, that preserves the temporal characteristics of the EEG signal, can be used as input to train a DAN machine learning model capable of successfully discriminating patients from control subjects based on the EEG signal after standard preprocessing alone, with accuracy significantly above chance ( $ACC = 0.69$ ). We then compared the model's performance with simpler machine learning architectures, as well as more common deep neural network models, showing similar performance. Finally, direct mapping of the attention value vector with the WM task trial time course, revealed that the most discriminative time points for the classification overlapped with the task preparation and memory retrieval phases, as well as with the identified time-frequency regions of interest that show significant group differences in alpha suppression, with patients showing less suppression than controls at these ROIs.

### 4.1. Normal WM performance and contralateral delay activity in schizophrenia patients

In our study, behavioral and CDA results did not differ significantly between patients and controls. This is in contrast with previous studies that generally find working memory performance deficits in schizophrenia (68, 69). Recent studies have found associations between poor performance and deficits in consolidation or early maintenance of stimuli (24), deficits in attention and executive control (70) or less efficient allocation of memory resources (71). Previous studies also report CDA amplitude differences between patients and controls, with amplitude being larger than that of control subjects at low memory load but smaller at high memory load (29), even when their maximum visual WM capacity is equal to that of control subjects. This pattern of impairment may support the theory of inefficient attention hyperfocus on a small number of items, especially when they are salient (17).

Normal behavioral and CDA results in our sample suggest that patients performed well on this particular visual WM task. These results are consistent with previous research showing no differences between high-functioning individuals with schizophrenia and healthy controls in task performance (18, 72) and working memory related ERPs (73, 74). Thus, given the preserved working

memory performance and lack of significant CDA abnormalities, our findings may be more representative of high-functioning patients. In this context, it is worth noting that, at the time of recruitment and throughout the data gathering phase, our patients were asymptomatic and engaged in psychodynamic group psychotherapy. While this criterion alone need not imply high-functioning status, given the known associations between engagement in psychodynamic psychotherapy and functional outcome (75), it is reasonable to suspect that our patients might potentially be close to high-functioning.

### 4.2. Schizophrenia patients exhibit decreased suppression of alpha spectral power during visual WM task

Our analysis revealed significantly lower suppression of non-lateralized parietal alpha spectral power during the task preparation, memory encoding, maintenance, and retrieval phases of the visual WM change detection task. Given the recognized role of oscillations in the alpha frequency band in long-range synchronization (38), top-down control (76, 77), attention (78) and cortical inhibition (79, 80), our time-frequency results may reflect a deficit that makes it difficult for patients to inhibit task-irrelevant brain regions and processes and to maintain efficient attention control, regardless of the experimental condition. These results are consistent with existing literature reporting alpha suppression abnormalities in schizophrenia during working memory tasks (24, 38–41). In the presence of these potential inhibitory and attention deficits, patients might have been able to maintain behavioral performance through various compensatory strategies, such as greater attention effort (17) reflected by less alpha suppression.

### 4.3. Deep attention networks can discriminate high performing individuals with schizophrenia from healthy controls in EEG data after dimensionality reduction

The dense attention network model implemented in this study was able to classify patients and controls with an accuracy significantly above chance,  $ACC = 0.69$  outperforming simpler machine learning architectures, while achieving similar performance to more commonly used deep network models. Moreover, our attention model revealed the relative importance of each feature in the input space for the successful classification of patients and controls. This was possible owing to our proposed data aggregation technique (i.e., incremental stepwise averaging), that allowed us to reduce each patient's preprocessed EEG data to a one-dimensional vector, while preserving the temporal characteristics of the signal.

The results show that the time points that were most discriminative for the machine learning algorithm, overlapped with both the preparatory and memory retrieval phases during the task, as well as with ROIs selected from the time-frequency maps. Furthermore, the two highest peaks in The DAN attention

value vector were found to overlap with the significant main effect of *group* found in time-frequency ROIs in the alpha band during task preparation and memory retrieval phases. Based on this congruence, we can conclude with high *confidence* (81) that the DAN model's decision to classify subjects as patients or controls is based on the same aspects of the data that were revealed by the time-frequency analysis. Given that the detected abnormalities are oscillatory in nature and the DAN algorithm partially operates by convolution (82), it might have been specially suited to detect oscillatory signatures in the EEG. Because, similarly, decomposition of the EEG into the time-frequency domain is often accomplished by convolution of the EEG signal with complex Morlet wavelets, which was our method of choice for the time-frequency analysis in this study.

These results add to the rapidly growing body of literature reporting encouraging results in the use of machine learning to classify patients and controls in schizophrenia (46–52), with the ultimate goal of aiding and improving the challenging diagnosis of such heterogeneous disorder (83). Furthermore, the demonstrated interpretability of our model highlights that machine learning can be designed to serve not only as a diagnostic aid in classification, but also to probe the neurophysiological correlates of schizophrenia and, potentially, other psychiatric disorders.

#### 4.4. Limitations and future directions

This study has some limitations. Our sample size was small, which may have affected statistical power. However, this is a consequence of the challenging goal of recruiting a homogeneous group of schizophrenia patients. Furthermore, our sample is constituted exclusively by males, which may limit the translational value of the study. Finally, although the accuracy of our DAN machine learning model is significant and provides additional information about the differences between patients and controls, it is not robust enough to support the direct diagnosis or classification of patients on its own.

Nevertheless, the machine learning and time-frequency results both suggest that in schizophrenia there is a significant impact on working memory processes during the task preparation and maintenance phases. Even in high performing patients that show no significant impact in behavioral performance or ERP correlates, when compared to healthy controls. Furthermore, the features studied could be combined with a broader set of features to support more accurate identification of patients. In that fashion, these techniques could be used as a diagnostic complement to more established clinical assessment methods to help in early detection or differential diagnosis of neuropsychiatric disorders with suspected oscillatory abnormalities. Moreover, the DAN model's accuracy could still be further improved by enriching its input with relevant multimodal data. For instance, as we have argued that the oscillatory abnormalities in the alpha band may indicate an inhibitory and attention deficit, future research could design experiments that would not only include EEG, but also additional techniques such as pupillometry, to measure changes in attention and arousal (84), or non-invasive stimulation, to directly probe the role of inhibitory neural circuits during the task (85). Finally, based on the neurophysiological insight provided by our

model, we further encourage the incorporation of interpretable models in schizophrenia research.

### Data availability statement

The original contributions presented in the study are publicly available. This data can be found here: <https://gitlab.com/MaticKu/shizo>.

### Ethics statement

The studies involving humans were approved by the Medical Ethics Committee of the Republic of Slovenia. The studies were conducted in accordance with the local legislation and institutional requirements. The participants provided their written informed consent to participate in this study.

### Author contributions

JB and GR: study conception and design. JB, RP-A, and IP: data collection. RP-A, MK, BŠkr, and IP: data analysis. RP-A, AO, BŠko, PP, KA-P, GR, and JB: results interpretation. RP-A, AO, and JB: manuscript writing. All authors contributed to the manuscript review and approved the submitted version.

### Funding

RP-A was supported by a fellowship from “la Caixa” Foundation (ID 100010434; Fellowship code: LCF/BQ/DI19/11730050). KA-P was financially supported by a Juan de la Cierva-Formacion research grant (FJC2021-047380-I) from the Spanish Ministry of Science and Innovation. IP is supported by a fellowship from “la Caixa” Foundation (ID 100010434; fellowship code LCF/BQ/DI18/11660026). This project has received funding from the European Union's Horizon 2020 Research and Innovation Programme under the Marie Skłodowska-Curie grant agreement No. 713673. This study was supported by Slovenian Research Agency grants P5-0110, P3-0338, J3-1763, and J3-9264.

### Acknowledgments

The authors want to thank the participants, without whom this study would not have been possible.

### Conflict of interest

GR consults for and holds equity in Neumora Therapeutics and Manifest Technologies.

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/fpsy.2023.1205119/full#supplementary-material>

## References

- Charlson FJ, Ferrari AJ, Santomauro DF, Diminic S, Stockings E, Scott JG, et al. Global epidemiology and burden of schizophrenia: Findings from the global burden of disease study 2016. *Schizophr Bull.* (2018) 44:1195–203. doi: 10.1093/schbul/sby058
- Batinic B. Cognitive models of positive and negative symptoms of schizophrenia and implications for treatment. *Psychiatria Danubina.* (2019) 31:S181–4.
- Mosolov SN, Yaltonskaya PA. Primary and secondary negative symptoms in schizophrenia. *Front Psychiatry.* (2022) 12:766692. doi: 10.3389/fpsy.2021.766692
- Robinson N, Bergen SE. Environmental risk factors for schizophrenia and bipolar disorder and their relationship to genetic risk: current knowledge and future directions. *Front Genet.* (2021) 12:686666. doi: 10.3389/fgene.2021.686666
- Seidman LJ, Mirsky AF. Evolving notions of schizophrenia as a developmental neurocognitive disorder. *J Int Neuropsychol Soc.* (2017) 23:881–92. doi: 10.1017/S1355617717001114
- Van Os J, Kenis G, Rutten BPF. The environment and schizophrenia. *Nature.* (2010) 468:203–12. doi: 10.1038/nature09563
- Howes OD, Shatalina E. Integrating the neurodevelopmental and dopamine hypotheses of schizophrenia and the role of cortical excitation-inhibition balance. *Biol Psychiatry.* (2022) 92:501–13. doi: 10.1016/j.biopsych.2022.06.017
- Liu Y, Ouyang P, Zheng Y, Mi L, Zhao J, Ning Y, et al. A selective review of the excitatory-inhibitory imbalance in schizophrenia: underlying biology, genetics, microcircuits, and symptoms. *Front Cell Dev Biol.* (2021) 9:664535. doi: 10.3389/fcell.2021.664535
- Anticevic A, Hu X, Xiao Y, Hu J, Li F, Bi F, et al. Early-course unmedicated schizophrenia patients exhibit elevated prefrontal connectivity associated with longitudinal change. *J Neurosci.* (2015) 35:267–86. doi: 10.1523/JNEUROSCI.2310-14.2015
- Anticevic A, Murray JD, Barch DM. Bridging levels of understanding in schizophrenia through computational modeling. *Clin Psychol Sci.* (2015) 3:433–59. doi: 10.1177/2167702614562041
- Hirano Y, Uhlhaas PJ. Current findings and perspectives on aberrant neural oscillations in schizophrenia. *Psychiatry Clin Neurosci.* (2021) 75:358–68. doi: 10.1111/pcn.13300
- Gold JM, Barch DM, Feuerstahler LM, Carter CS, MacDonald AW, Daniel Ragland J, et al. Working memory impairment across psychotic disorders. *Schizophr Bull.* (2019) 45:804–12. doi: 10.1093/schbul/sby134
- Tao TJ, Hui CLM, Hui PWM, Ho ECN, Lam BST, Wong AKH, et al. Working memory deterioration as an early warning sign for relapse in remitted psychosis: A one-year naturalistic follow-up study. *Psychiatry Res.* (2023) 319:114976. doi: 10.1016/j.psychres.2022.114976
- D'Esposito M, Postle BR. The cognitive neuroscience of working memory. *Annu Rev Psychol.* (2015) 66:115–42. doi: 10.1146/annurev-psych-010814-015031
- Luck SJ, Vogel EK. Visual working memory capacity: From psychophysics and neurobiology to individual differences. *Trends Cogn Sci.* (2013) 17:391–400. doi: 10.1016/j.tics.2013.06.006
- Barch DM, Ceaser A. Cognition in schizophrenia: Core psychological and neural mechanisms. *Trends Cogn Sci.* (2012) 16:27–34. doi: 10.1016/j.tics.2011.11.015
- Luck SJ, Hahn B, Leonard CJ, Gold JM. The hyperfocusing hypothesis: a new account of cognitive dysfunction in schizophrenia. *Schizophr Bull.* (2019) 45:991–1000. doi: 10.1093/schbul/sbz063
- Rentrop M, Rodewald K, Roth A, Simon J, Walther S, Fiedler P, et al. Intra-individual variability in high-functioning patients with schizophrenia. *Psychiatry Res.* (2010) 178:27–32. doi: 10.1016/j.psychres.2010.04.009
- Johnson MK, McMahon RP, Robinson BM, Harvey AN, Hahn B, Leonard CJ, et al. The relationship between working memory capacity and broad measures of cognitive ability in healthy adults and people with schizophrenia. *Neuropsychology.* (2013) 27:220–9. doi: 10.1037/a0032060
- Jaeggi SM, Buschkuhl M, Perrig WJ, Meier B. The concurrent validity of the N-back task as a working memory measure. *Memory.* (2010) 18:394–412. doi: 10.1080/09658211003702171
- Scharinger C, Soutschek A, Schubert T, Gerjets P. Comparison of the working memory load in N-back and working memory span tasks by means of EEG frequency band power and P300 amplitude. *Front Hum Neurosci.* (2017) 11:6. doi: 10.3389/fnhum.2017.00006
- Luck SJ, Vogel EK. The capacity of visual working memory for features and conjunctions. *Nature.* (1997) 390:279–84. doi: 10.1038/36846
- Vogel EK, Machizawa MG. Neural activity predicts individual differences in visual working memory capacity. *Nature.* (2004) 428:748–51. doi: 10.1038/nature02447
- Coffman BA, Haas G, Olson C, Cho R, Ghuman AS, Salisbury DF. Reduced Dorsal Visual Oscillatory Activity During Working Memory Maintenance in the First-Episode Schizophrenia Spectrum. *Front Psychiatry.* (2020) 11:743. doi: 10.3389/fpsy.2020.00743
- Driesen NR, Leung HC, Calhoun VD, Constable RT, Gueorguieva R, Hoffman R, et al. Impairment of working memory maintenance and response in schizophrenia: functional magnetic resonance imaging evidence. *Biol Psychiatry.* (2008) 64:1026–34. doi: 10.1016/j.biopsych.2008.07.029
- Huang AS, Rogers BP, Anticevic A, Blackford JU, Heckers S, Woodward ND. Brain function during stages of working memory in schizophrenia and psychotic bipolar disorder. *Neuropsychopharmacology.* (2019) 44:2136–42. doi: 10.1038/s41386-019-0434-4
- Wang B, Zartaloudi E, Linden JF, Bramon E. Neurophysiology in psychosis: The quest for disease biomarkers. *Transl Psychiatry.* (2022) 12:1–10. doi: 10.1038/s41398-022-01860-x
- Luria R, Balaban H, Awh E, Vogel EK. The contralateral delay activity as a neural measure of visual working memory. *Neurosci Biobehav Rev.* (2016) 62:100–8. doi: 10.1016/j.neubiorev.2016.01.003
- Leonard CJ, Kaiser ST, Robinson BM, Kappenman ES, Hahn B, Gold JM, et al. Toward the neural mechanisms of reduced working memory capacity in schizophrenia. *Cereb Cortex.* (2013) 23:1582–92. doi: 10.1093/cercor/bhs148
- Wang XJ. Neurophysiological and computational principles of cortical rhythms in cognition. *Physiol Rev.* (2010) 90:1195–268. doi: 10.1152/physrev.00035.2008
- Lisman JE, Jensen O. The Theta-Gamma Neural Code. *Neuron.* (2013) 77:1002–16. doi: 10.1016/j.neuron.2013.03.007
- Missonnier P, Prévot A, Herrmann FR, Ventura J, Padée A, Merlo MCG. Disruption of gamma-delta relationship related to working memory deficits in first-episode psychosis. *J Neural Trans.* (2020) 127:103–15. doi: 10.1007/s00702-019-02126-5
- Sauseng P, Griesmayr B, Freunberger R, Klimesch W. Control mechanisms in working memory: A possible function of EEG theta oscillations. *Neurosci Biobehav Rev.* (2010) 34:1015–22. doi: 10.1016/j.neubiorev.2009.12.006
- Cao Y, Han C, Peng X, Su Z, Liu G, Xie Y, et al. Correlation between resting theta power and cognitive performance in patients with schizophrenia. *Front Hum Neurosci.* (2022) 16:853994. doi: 10.3389/fnhum.2022.853994
- Barr MS, Rajji TK, Zomorodi R, Radhu N, George TP, Blumberg DM, et al. Impaired theta-gamma coupling during working memory performance in schizophrenia. *Schizophr Res.* (2017) 189:104–10. doi: 10.1016/j.schres.2017.01.044
- Jensen O, Mazaheri A. Shaping functional architecture by oscillatory alpha activity: Gating by inhibition. *Front Hum Neurosci.* (2010) 4:186. doi: 10.3389/fnhum.2010.00186

37. Klimesch W. Alpha-band oscillations, attention, and controlled access to stored information. *Trends Cogn Sci.* (2012) 16:606–17. doi: 10.1016/j.tics.2012.10.007
38. Doesburg SM, Green JJ, McDonald JJ, Ward LM. From local inhibition to long-range integration: A functional dissociation of alpha-band synchronization across cortical scales in visuospatial attention. *Brain Res.* (2009) 1303:97–110. doi: 10.1016/j.brainres.2009.09.069
39. Erickson MA, Albrecht MA, Robinson B, Luck SJ, Gold JM. Impaired Suppression of Delay-Period Alpha and Beta Is Associated With Impaired Working Memory in Schizophrenia. *Biol Psychiatry.* (2017) 2:272–9. doi: 10.1016/j.bpsc.2016.09.003
40. Kustermann T, Rockstroh B, Kienle J, Miller GA, Popov T. Deficient attention modulation of lateralized alpha power in schizophrenia: Deficient lateralized alpha modulation in schizophrenia. *Psychophysiology.* (2016) 53:776–85.
41. Rameyad A, Roach B, Hamilton H, Addington J, Bachman P, Bearden C, et al. O33. EEG Alpha Event-Related Desynchronization Deficits Predict Conversion to Psychosis in Individuals With the Psychosis Risk Syndrome. *Biol Psychiatry.* (2019) 85:S119. doi: 10.1016/j.biopsych.2019.03.298.
42. Scangos KW, State MW, Miller AH, Baker JT, Williams LM. New and emerging approaches to treat psychiatric disorders. *Nat Med.* (2023) 29:317–33. doi: 10.1038/s41591-022-02197-0
43. Williams LM. Precision psychiatry: A neural circuit taxonomy for depression and anxiety. *Lancet Psychiatry.* (2016) 3:472–80. doi: 10.1016/S2215-036600579-9
44. Feczko E, Miranda-Dominguez O, Marr M, Graham AM, Nigg JT, Fair DA. The Heterogeneity Problem: Approaches to Identify Psychiatric Subtypes. *Trends Cogn Sci.* (2019) 23:584–601. doi: 10.1016/j.tics.2019.03.009
45. Martín Noguero T, Paulano-Godino F, Martín-Valdivia MT, Menias CO, Luna A. Strengths, weaknesses, opportunities, and threats analysis of artificial intelligence and machine learning applications in radiology. *J. Am. Coll. Radiol.* (2019) 16:1239–1247. doi: 10.1016/j.jacr.2019.05.047
46. Johannesen JK, Bi J, Jiang R, Kenney JG, Chen C-MA. Machine learning identification of EEG features predicting working memory performance in schizophrenia and healthy adults. *Neuropsychiatric Electrophysiol.* (2016) 2:1–21. doi: 10.1186/s40810-016-0017-0
47. Kim JW, Lee YS, Han DH, Min KJ, Lee J, Lee K. Diagnostic utility of quantitative EEG in un-medicated schizophrenia. *Neurosci Lett.* (2015) 589:126–31. doi: 10.1016/j.neulet.2014.12.064
48. Phang CR, Noman F, Hussain H, Ting CM, Ombao H. A multi-domain connectome convolutional neural network for identifying schizophrenia from EEG connectivity patterns. *IEEE J Biomed Health Inform.* (2020) 24:1333–43. doi: 10.1109/JBHI.2019.2941222
49. Ruiz de Miras J, Ibáñez-Molina AJ, Soriano MF, Iglesias-Parro S. Schizophrenia classification using machine learning on resting state EEG signal. *Biomed Signal Process Control.* (2023) 79:104233. doi: 10.1016/j.bpsc.2022.104233
50. Shim M, Hwang HJ, Kim DW, Lee SH, Im CH. Machine-learning-based diagnosis of schizophrenia using combined sensor-level and source-level EEG features. *Schizop Res.* (2016) 176:314–9. doi: 10.1016/j.schres.2016.05.007
51. Shoebai A, Sadeghi D, Moridian P, Ghassemi N, Heras J, Alizadehsani R, et al. Automatic Diagnosis of Schizophrenia in EEG Signals Using CNN-LSTM Models. *Front Neuroinform.* (2021) 15:777977. doi: 10.3389/fninf.2021.777977
52. Sun J, Cao R, Zhou M, Hussain W, Wang B, Xue J, et al. A hybrid deep neural network for classification of schizophrenia using EEG Data. *Sci Rep.* (2021) 11:1–16. doi: 10.1038/s41598-021-83350-6
53. Sheu YH. Illuminating the Black Box: Interpreting Deep Neural Network Models for Psychiatric Research. *Front Psychiatry.* (2020) 11:551299. doi: 10.3389/fpsy.2020.551299
54. Barros C, Silva CA, Pinheiro AP. Advanced EEG-based learning approaches to predict schizophrenia: Promises and pitfalls. *Artif Intell Med.* (2021) 114:102039.
55. Bahdanau D, Cho KH, Bengio Y. Neural machine translation by jointly learning to align and translate. *Proceedings of the 3rd International Conference on Learning Representations, ICLR 2015 - Conference Track Proceedings.* Vienna (2015).
56. de Santana Correia A, Colombini EL. Attention, please! A survey of neural attention models in deep learning. *Artif Intell Rev.* (2022) 55:6037–124. doi: 10.1007/s10462-022-10148-x
57. Škrlić B, Daeroski S, Lavrač N, Petković M. Feature importance estimation with self-attention networks. *Front Artif Intell Applic.* (2020) 325:1491–8. doi: 10.3233/FAIA200256
58. Kay SR, Fiszbein A, Opler LA. The positive and negative syndrome scale (PANSS) for schizophrenia. *Schizop Bull.* (1987) 13:261–76. doi: 10.1093/schbul/13.2.261
59. Vogel EK, McCollough AW, Machizawa MG. Neural measures reveal individual differences in controlling access to working memory. *Nature.* (2005) 438:500–3. doi: 10.1038/nature04171
60. Peirce J, Gray JR, Simpson S, MacAskill M, Höchenberger R, Sogo H, et al. PsychoPy2: Experiments in behavior made easy. *Behav Res Methods.* (2019) 51:195–203. doi: 10.3758/s13428-018-01193-y
61. Roudier JN, Morey RD, Morey CC, Cowan N. How to measure working memory capacity in the change detection paradigm. *Psych Bull Rev.* (2011) 18:324–30. doi: 10.3758/s13423-011-0055-3
62. Delorme A, Makeig S. EEGLAB: An open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *J Neurosci Methods.* (2004) 134:9–21. doi: 10.1016/j.jneumeth.2003.10.009
63. Mullen T. *NITRC: CleanLine: Tool/Resource Info.* New Delhi: NITRC (2012).
64. Hsu SS. *Adaptive Mixture ICA (AMICA): Theory & Practicum.* Lincoln: AMICA (2018).
65. Kingma DP, Ba JL. Adam: A method for stochastic optimization. *Proceedings of the 3rd International Conference on Learning Representations, ICLR 2015 - Conference Track Proceedings.* Vienna (2015).
66. Paszke A, Gross S, Massa F, Lerer A, Bradbury J, Chanan G, et al. PyTorch: An Imperative Style, High-Performance Deep Learning Library. In: Wallach H, Larochelle H, Beygelzimer A, Alché-Buc F, Fox E, Garnett R editors. *Advances in Neural Information Processing Systems.* New York, NY: Curran Associates, Inc (2019).
67. R Core Team. *R: A language and environment for statistical computing.* Vienna: R Foundation for Statistical Computing (2017).
68. Forbes NE, Carrick LA, McIntosh AM, Lawrie SM. Working memory in schizophrenia: A meta-analysis. *Psychol Med.* (2009) 39:889–905. doi: 10.1017/S0033291708004558
69. Gold JM, Luck SJ. Working Memory in People with Schizophrenia. *Curr Topics Behav Neurosci.* (2023) 63:137–52. doi: 10.1007/7854\_2022\_381
70. Gold JM, Robinson B, Leonard CJ, Hahn B, Chen S, McMahon RP, et al. Selective attention, working memory, and executive function as potential independent sources of cognitive dysfunction in schizophrenia. *Schizop Bull.* (2018) 44:1227–34. doi: 10.1093/schbul/sbx155
71. Zhao YJ, Ma T, Zhang L, Ran X, Zhang RY, Ku Y. Atypically larger variability of resource allocation accounts for visual working memory deficits in schizophrenia. *PLoS Comput Biol.* (2021) 17:1009544. doi: 10.1371/journal.pcbi.1009544
72. Heinrichs RW, Pinnock F, Muharib E, Hartman L, Goldberg J, McDermid Vaz S. Neurocognitive normality in schizophrenia revisited. *Schizop Res.* (2015) 2:227–32. doi: 10.1016/j.sch.2015.09.001
73. Light GA, Williams LE, Minow F, Sprock J, Rissling A, Sharp R, et al. Electroencephalography (EEG) and event-related potentials (ERPs) with human participants. *Curr Protoc Neurosci.* (2010) 52:1–24. doi: 10.1002/0471142301.n5062552
74. So RP, Kegeles LS, Mao X, Shungu DC, Stanford AD, Chen CMA. Long-range gamma phase synchronization as a compensatory strategy during working memory in high-performing patients with schizophrenia. *J Clin Exp Neuropsychol.* (2018) 40:663–81. doi: 10.1080/13803395.2017.1420142
75. Modesti MN, Arena JF, Palermo N, Del Casale A. A Systematic Review on Add-On Psychotherapy in Schizophrenia Spectrum Disorders. *J Clin Med.* (2023) 12:1021. doi: 10.3390/jcm12031021
76. Von Stein A, Chiang C, König P. Top-down processing mediated by interareal synchronization. *Proc Natl Acad Sci U.S.A.* (2000) 97:14748–53. doi: 10.1073/pnas.97.26.14748
77. Womelsdorf T, Schoffelen JM, Oostenveld R, Singer W, Desimone R, Engel AK, et al. Modulation of neuronal interactions through neuronal synchronization. *Science.* (2007) 316:1609–12. doi: 10.1126/science.1139597
78. Thut G, Nietzel A, Brandt SA, Pascual-Leone A.  $\alpha$ -Band electroencephalographic activity over occipital cortex indexes visuospatial attention bias and predicts visual target detection. *J Neurosci.* (2006) 26:9494–502. doi: 10.1523/JNEUROSCI.0875-06.2006
79. Klimesch W, Sauseng P, Hanslmayr S. EEG alpha oscillations: The inhibition-timing hypothesis. *Brain Res Rev.* (2007) 53:63–88. doi: 10.1016/j.brainresrev.2006.06.003
80. Lewis DA, Hashimoto T, Volk DW. Cortical inhibitory neurons and schizophrenia. *Nat Rev Neurosci.* (2005) 6:312–24. doi: 10.1038/nrn1648
81. Avberšek LK, Repovš G. Deep learning in neuroimaging data analysis: Applications, challenges, and solutions. *Front Neuroimag.* (2022) 1:981642. doi: 10.3389/fnimg.2022.981642
82. Cordonnier J-B, Loukas A, Jaggi M. On the relationship between self-attention and convolutional layers. *arXiv [Preprint].* (2019) doi: 10.48550/arXiv.1911.03584
83. Wolfers T, Doan NT, Kaufmann T, Alnaes D, Moberget T, Agartz I, et al. Mapping the heterogeneous phenotype of schizophrenia and bipolar disorder using normative models. *JAMA Psychiatry.* (2018) 75:1146–55. doi: 10.1001/jamapsychiatry.2018.2467
84. Unsworth N, Robison MK. The importance of arousal for variation in working memory capacity and attention control: A latent variable pupillometry study. *J Exp Psychol.* (2017) 43:1962–87. doi: 10.1037/xlm0000421
85. Takahashi S, Ukai S, Kose A, Hashimoto T, Iwatani J, Okumura M, et al. Reduction of cortical GABAergic inhibition correlates with working memory impairment in recent onset schizophrenia. *Schizop Res.* (2013) 146:238–43. doi: 10.1016/j.schres.2013.02.033