CURRENTS IN BIOMEDICAL SIGNALS PROCESSING - METHODS AND APPLICATIONS

EDITED BY: Aleksandra Dagmara Kawala-Sterniuk, Mariusz Pelc, Radek Martinek and Grzegorz Marcin Wójcik

PUBLISHED IN: Frontiers in Neuroscience, Frontiers in Human Neuroscience and Frontiers in Bioengineering and Biotechnology

frontiers Research Topics

frontiers

ISSN 1664-8714
ISBN 978-2-88976-901-8
DOI 10.3389/978-2-88976-901-8

About Frontiers

Frontiers is more than just an open-access publisher of scholarly articles: it is a pioneering approach to the world of academia, radically improving the way scholarly research is managed. The grand vision of Frontiers is a world where all people have an equal opportunity to seek, share and generate knowledge. Frontiers provides immediate and permanent online open access to all its publications, but this alone is not enough to realize our grand goals.

Frontiers Journal Series

The Frontiers Journal Series is a multi-tier and interdisciplinary set of open-access, online journals, promising a paradigm shift from the current review, selection and dissemination processes in academic publishing. All Frontiers journals are driven by researchers for researchers; therefore, they constitute a service to the scholarly community. At the same time, the Frontiers Journal Series operates on a revolutionary invention, the tiered publishing system, initially addressing specific communities of scholars, and gradually climbing up to broader public understanding, thus serving the interests of the lay society, too.

Dedication to Quality

Each Frontiers article is a landmark of the highest quality, thanks to genuinely collaborative interactions between authors and review editors, who include some of the world's best academicians. Research must be certified by peers before entering a stream of knowledge that may eventually reach the public - and shape society; therefore, Frontiers only applies the most rigorous and unbiased reviews.
Frontiers revolutionizes research publishing by freely delivering the most outstanding research, evaluated with no bias from both the academic and social point of view. By applying the most advanced information technologies, Frontiers is catapulting scholarly publishing into a new generation.

What are Frontiers Research Topics?

Frontiers Research Topics are very popular trademarks of the Frontiers Journals Series: they are collections of at least ten articles, all centered on a particular subject. With their unique mix of varied contributions from Original Research to Review Articles, Frontiers Research Topics unify the most influential researchers, the latest key findings and historical advances in a hot research area! Find out more on how to host your own Frontiers Research Topic or contribute to one as an author by contacting the Frontiers Editorial Office: frontiersin.org/about/contact

CURRENTS IN BIOMEDICAL SIGNALS PROCESSING - METHODS AND APPLICATIONS

Topic Editors:
Aleksandra Dagmara Kawala-Sterniuk, Opole University of Technology, Poland
Mariusz Pelc, University of Greenwich, United Kingdom
Radek Martinek, VSB-Technical University of Ostrava, Czechia
Grzegorz Marcin Wójcik, Marie Curie-Sklodowska University, Poland

Citation: Kawala-Sterniuk, A. D., Pelc, M., Martinek, R., Wójcik, G. M., eds. (2022). Currents in Biomedical Signals Processing - Methods and Applications. Lausanne: Frontiers Media SA. doi: 10.3389/978-2-88976-901-8

Table of Contents

frontiers | Frontiers in Neuroscience

TYPE Editorial
PUBLISHED 29 July 2022
DOI 10.3389/fnins.2022.989400

OPEN ACCESS

EDITED AND REVIEWED BY
Michele Giugliano,
International School for Advanced Studies (SISSA), Italy

*CORRESPONDENCE
Aleksandra Kawala-Sterniuk
kawala84@gmail.com

SPECIALTY SECTION
This article was submitted to Neural Technology, a section of the journal Frontiers in Neuroscience

RECEIVED 08 July 2022
ACCEPTED 18 July 2022
PUBLISHED 29 July 2022

CITATION
Kawala-Sterniuk A, Pelc M, Martinek R and Wójcik GM (2022) Editorial: Currents in biomedical signals processing—methods and applications.
Front. Neurosci. 16:989400.
doi: 10.3389/fnins.2022.989400

Editorial: Currents in biomedical signals processing—methods and applications

Aleksandra Kawala-Sterniuk[1]*, Mariusz Pelc[1,2], Radek Martinek[3] and Grzegorz Marcin Wójcik[4,5]

[1]Faculty of Electrical Engineering, Automatic Control and Informatics, Opole University of Technology, Opole, Poland, [2]School of Computing and Mathematical Sciences, University of Greenwich, London, United Kingdom, [3]Department of Cybernetics and Biomedical Engineering, Faculty of Electrical Engineering and Computer Science, VSB-Technical University of Ostrava, Ostrava, Czechia, [4]Department of Neuroinformatics, Faculty of Mathematics, Physics and Computer Science, Institute of Computer Science, Maria Curie-Sklodowska University in Lublin, Lublin, Poland, [5]Department of Neuroinformatics and Biomedical Engineering, Institute of Computer Science, Maria Curie-Sklodowska University, Lublin, Poland

KEYWORDS

biomedical signals, signal processing, cognitive sciences, neuroscience, biomedical engineering

Editorial on the Research Topic
Currents in biomedical signals processing—methods and applications

1. Introduction

Biosignals as measurement of the human body's functions provide useful information regarding human condition. Thus, the analysis of biomedical signals has become one of the most important methods for both interpretations and visualization in numerous research areas such as inter alia biology or medicine. They also play a very important role in health monitoring, diagnosis, but also as a source of data for the control purposes (in Human-Machine Interfaces). It has also led to development of numerous modern instruments designed for their detection, storage, transmission, and analysis. As the biological signals appear to be random (stochastic), it is impossible to predict their value in any time instant and therefore only statistical measures may be used in order to determine their features.

Recent advances in signal computational methods have enabled the biomedical signals in an efficient way in order to extract appropriate features and has been an interesting subject for numerous research groups for over 40 years. More and more sophisticated methods are being developed and applied including various classifiers and filters. This Research Topic should focus on the most current trends in analysis of the most popular biosignals such as EEG, ECG, EMG, which would be processed for various applications such as inter alia clinical (rehabilitation, diagnostic purposes) or control (human-machine interaction).

In this Research Topic titled "Currents in Biomedical Signals Processing—Methods and Applications" a collection of contributions showing new advancements and applications of advanced processing of biosignals for various applications was presented.

The paper Binkowska et al. was focused on association between long-term cannabis use and memory impairments. Based on authors' thorough literature review—very few studies have examined this relationship, in particular in a poly-drug context, when cannabis was combined with other addictive substances. For this purpose the authors used event-related potentials for the examination of the recognition process in a visual episodic memory task in both cannabis users (CU) and cannabis poly-drug users (PU). The hypothesis stated that both groups (CU and PU) will have their behavioral and phychophysiological-indicators of memory processes affected with the use of cannabis alone or with other substances, with the PU patients expressing stronger changes. The obtained results proved alterations in recognition memory processing in both groups compared to the healthy participants from the control group (CG), and these were at most visible in cannabis poly-drug users (PU), while there was no significant effect for patients using cannabis alone. The paper proves that the use of cannabis with other substances results with the biggest disturbance in the brain function. The authors applied analysis of electroencephalography (EEG) signals for this study, which is non-invasive.

Another study involving analysis of EEG signals was presented in Kwasniewicz et al., where the authors used brain signals for the purpose of message credibility evaluation. No similar studies could be found in the literature and the efficiency of the obtained results was over 0.7. The authors found out appropriate brain areas active depending on positive or negative message reliability.

Also in Jakubowska et al., the authors used EEG signals, where they assessed visual working memory (VWM) capacity. The authors tired to assess the efficiency of training conducted on a Real-Time Strategy (RTS) video game (StarCraft II) for the purpose of VWM capacity improvement. The study involved 62 participants, who took part in two EEG sessions (before and after the StarCraft II-based training).

Analysis of the EEG signals, which is a very challenging task mostly due to the nature of these signals, was also presented in paper Xu et al.; where the authors applied topological data analysis (TDA) in order to analyse and understand the brain signals. This paper is a short review and shows the opportunities in analysis of EEG signals brought by the TDA, which differs from traditional signal processing methods and has a very wise potential usage in among the others Brain-Computer Interfaces (BCIs).

As mentioned above—analysis of the EEG signals is difficult among the others due to these signals' non-stationarity. Also the electroencephalography signals are prone to various external and internal artifacts occurrence, where the internal artifacts are frequently related with physiological activity, such as the one related with eye movement and blinking. The eye blinking (EB) artifacts strongly affect the signals' quality. In Jurczak et al., the authors compared the most popular methods applied for the electrooculography (EOG) artifacts removal—independent component analysis (ICA) and regression with the convolutional neural network (CNN) in order to eliminate the eye blinking artifacts. The obtained results showed much higher performance of the CNN method than the one given by regression or ICA.

In Barry et al. the authors focused on using more invasive brain signals—electrocorticography (ECoG). The main aim of this work was to examine ECoG spectrogram images for training proper and reliable cross-patient seizure classifiers. In this paper, the data collected from 113 patients was converted into RGB spectrogram images. For classification purposes five different convolutional neural networks (CNN) were applied, which gave the cross-patient classification (seizure/non-seizure) accuracy of 87.9%, while the appropriate trained ResnEt50-based models provided efficiency of 95.7%, which is very high.

Another study using deep learning algorithms such as convolutional neural networks, in this case the osteogenic convolutional neural network (OCNN), was presented in Lan et al., where the authors decided to use them for the rat bone marrow mesenchymal stem cells (rBMSCs) osteogenic differentiation quantitative measurement. The study showed that the OCNN enabled successful distinguishing of differentiated cells at a very early stage and gave better prediction performance compared with the single morphological parameters. Based on the conducted experiments it is possible apply the OCNN-based online learning models for further rBMSCs osteogenic differentiation recognition.

The authors of Haraguchi et al. described the use of less common signals, namely magnetoencephalography (MEG), where the effect of menstrual cycle on brain activity has been analyzed. In this work, the authors focused on objective quantitative MEG parameters in order to investigate the effects of the menstrual cycle on spontaneous neural oscillations. The study involved 25 healthy female participant, with normal, regular menstrual cycle. The testing was conducted twice: during menstrual period (MP) and outside period (OP). The authors of this paper showed that it is possible to use the menstrual cycle as an accurate interpretation of functional neuroimaging in clinical practice.

In Bartosik et al. statistical analysis was applied for the purpose of selecting the most important personality features based on attractiveness assessment. Based on authors thorough literature study and on their research experience they claimed that trust is based on facial appearance appraisal, which is made based on facial morphological characteristics, such as among the others color, complexion, shape, etc. In order to select appropriate features the authors modeled a backward step-wise logistic regression; they also analyzed the results of the psychological tests together with the attractiveness and trust survey.

In paper Boschen et al., the authors presented Fast Scan Cyclic Voltammetry (FSCV), which has been used for many years in animal models, but has not been applied on

humans. The main aim of this work was to bring interest into using the FSCV in human clinical studies. This article showed also some technical challenges, which may limit its clinical implementation, but these can be overcome while properly addressed.

As mentioned above—analysis of biomedical data is a very challenging task, but this makes is very interesting. We hope that our Research Topic will be found interesting to readers and researchers in fields of medicine, biomedical engineering or neuroscience.

Author contributions

All authors reviewed and accepted the manuscript for submission.

Conflict of interest

The 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

frontiers in Human Neuroscience

ORIGINAL RESEARCH
published: 11 June 2021
doi: 10.3389/fnhum.2021.677793

Not Just a Pot: Visual Episodic Memory in Cannabis Users and Polydrug Cannabis Users: ROC and ERP Preliminary Investigation

Alicja Anna Binkowska[1], Natalia Jakubowska[1,2], Maciej Gaca[3], Natalia Galant[4], Agnieszka Piotrowska-Cyplik[5] and Aneta Brzezicka[1,6]*

[1] SWPS University of Social Sciences and Humanities, Warsaw, Poland, [2] Polish-Japanese Academy of Information Technology, Warsaw, Poland, [3] Nencki Institute of Experimental Biology PAS, Warsaw, Poland, [4] Institute of Forensic Genetics, Bydgoszcz, Poland, [5] Institute of Food Technology of Plant Origin, Poznań University of Life Sciences, Poznań, Poland, [6] Department of Neurosurgery, Cedars-Sinai Medical Center, Los Angeles, CA, United States

Background: While research has consistently identified an association between long-term cannabis use and memory impairments, few studies have examined this relationship in a polydrug context (i.e., when combining cannabis with other substances).

Aims: In this preliminary study, we used event-related potentials to examine the recognition process in a visual episodic memory task in cannabis users (CU) and cannabis polydrug users (PU). We hypothesized that CU and PU will have both–behavioral and psychophysiological–indicators of memory processes affected, compared to matched non-using controls with the PU expressing more severe changes.

Methods: 29 non-using controls (CG), 24 CU and 27 PU were enrolled into the study. All participants completed a visual learning recognition task while brain electrical activity was recorded. Event-related potentials were calculated for familiar (old) and new images from a signal recorded during a subsequent recognition test. We used receiver operating characteristic curves for behavioral data analysis.

Results: The groups did not differ in memory performance based on receiver operating characteristic method in accuracy and discriminability indicators nor mean reaction times for old/new images. The frontal old/new effect expected from prior research was observed for all participants, while a parietal old/new effect was not observed. While, the significant differences in the late parietal component (LPC) amplitude was observed between CG and PU but not between CG and CU nor CU and PU. Linear regression analysis was used to examine the mean amplitude of the LPC component as a predictor of memory performance accuracy indicator. LPC amplitude predicts recognition accuracy only in the CG.

Conclusion: The results showed alterations in recognition memory processing in CU and PU groups compared to CG, which were not manifested on the behavioral level, and were the most prominent in cannabis polydrug users. We interpret it as a manifestation of the cumulative effect of multiple drug usage in the PU group.

Keywords: cannabis (marijuana), polydrug use, recognition memory, EEG, late parietal component, SDT, ROC

OPEN ACCESS

Edited by:
Aleksandra Dagmara Kawala-Sterniuk, Opole University of Technology, Poland

Reviewed by:
Edward Jacek Gorzelańczyk, Nicolaus Copernicus University, Poland
Santiago Galdo-Alvarez, University of Santiago de Compostela, Spain
Grzegorz Miêkisiak, Opole University, Poland

****Correspondence:***
Alicja Anna Binkowska
abinkowska@swps.edu.pl

Specialty section:
This article was submitted to Brain Imaging and Stimulation, a section of the journal Frontiers in Human Neuroscience

Received: *08 March 2021*
Accepted: *14 May 2021*
Published: *11 June 2021*

Citation:
Binkowska AA, Jakubowska N, Gaca M, Galant N, Piotrowska-Cyplik A and Brzezicka A (2021) Not Just a Pot: Visual Episodic Memory in Cannabis Users and Polydrug Cannabis Users: ROC and ERP Preliminary Investigation. Front. Hum. Neurosci. 15:677793. doi: 10.3389/fnhum.2021.677793

INTRODUCTION

Cannabis is the most commonly used illicit substance in the world (United Nations Office on Drugs & Crime (UNODC), 2019) and approximately 26% of adult Europeans have a history of using cannabis (European Monitoring Centre for Drugs and Drug Addiction (EMCDDA), 2018). And it is not without influence on their cognitive functioning. The most consistent and prominent acute effects of cannabis are impairments in the verbal learning and memory and working memory as shown in recent meta-analysis, where reported effect sizes were medium, between 0.5 and 0.7 (Zhornitsky et al., 2021). Moreover, impairments in various cognitive functions have been associated with the chronic cannabis use (lasting beyond intoxication phase), with the most consistent evidence in impairment of verbal episodic and working memory (Broyd et al., 2016; Curran et al., 2016; Figueiredo et al., 2020). Meta-analyses concentrated on the potential long-term effects of cannabis use on cognitive domains, has shown the significant deficits in executive functions, learning, working memory, attention, processing speed, and the estimated effect sizes were small (Grant et al., 2003; Schreiner and Dunn, 2012; Scott et al., 2018; Figueiredo et al., 2020; Kroon et al., 2020a).

Less is known about the long-term impact of cannabis on visual memory and learning. Studies suggest that THC (*tetrahydrocannabinol*–the main psychoactive compound of cannabis) may interfere with visual information processing, which can lead to impairment of the visual memory of objects and increase the chances of false recognition (Böcker et al., 2010; Winton-Brown et al., 2011). Recently conducted study has shown that people directly under the influence of cannabis exhibit impairment in visual episodic memory, including learning process (Doss et al., 2020). The question arises whether this effect persists in regular cannabis users when they are no longer directly intoxicated and if it is accompanied by altered brain function. The formation and retrieval in episodic memory involve mostly medial temporal lobe (MTL) structures such as the hippocampus and parahippocampal gyrus (Brewer et al., 1998; Desgranges et al., 1998). It is important to note that the highest density of cannabinoid receptors (CB1) is in the hippocampus and prefrontal cortex (Herkenham et al., 1990; Ameri, 1999).

There is strong evidence coming from studies on the relation between brain function and episodic memory, that event-related potential (ERP) components are helpful in understanding physiological correlates of episodic memory (Rugg and Curran, 2007; Hoppstädter et al., 2015). The majority of studies on learning and recognition tasks show that ERPs evoked by previously experienced stimuli ("old" or "familiar") are more positive than ERPs evoked by new stimuli ("new")–this effect is known in the literature as the so called "Old/New effect" (Rugg and Curran, 2007; Hoppstädter et al., 2015). There are two main components of the Old/New effect–the frontal negative (FN) deflection peaking between 300 and 500 ms, labeled the FN400, which is related to the familiarity process (Curran, 2000; Paller et al., 2007) and late positive component (LPC), which is a posterior positive deflection, peaking between 400 and 800 ms, related to the recollection process (Curran and Cleary, 2003; Rugg and Curran, 2007).

While there is still a discussion about the cognitive mechanisms involved in LPC generation, the research suggests that it is related to categorical response, decision accuracy, maintenance of visual working memory representations and memory match, despite recognition memory (Danker et al., 2008; Schendan and Maher, 2009; Gunseli et al., 2014). Study conducted by Finnigan et al. (2002) has shown that the LPC amplitude was significantly more positive for correct than incorrect recognition responses not only when old but also when new items were considered. Another research based on drift-diffusion modeling of behavioral data showed that the LPC amplitude predicts participants' accuracy of recognition-memory decisions on a trial-by-trial basis (Ratcliff et al., 2016).

To the best of our knowledge, there is only one study investigating ERPs in non-intoxicated regular cannabis users and heavy alcohol drinkers, which used the verbal memory task (the Rey Auditory Verbal Learning Test, RAVLT) conducted by Smith et al. (2017). The results did not show any behavioral impairments, but indicated alterations in recognition memory processing manifested by a larger LPC in heavy drinkers compared to a control group. Cannabis-related differences were related to a smaller FN400 component and a lack of the Old/New effect, usually observed in a LPC response.

Detailed analysis of our collected data on self-reported substance use as well as the hair sample analysis on drug use (results were delivered after study accomplishment) revealed that the majority of recruited cannabis users actually use other illicit psychoactive substances as well.

Cannabis is the most commonly used drug within a polydrug context (Mitchell and Plunkett, 2000; Carlson et al., 2005; Lynskey et al., 2006; World Health Organisation, 2016). The majority of cannabis users do not restrict substance use to cannabis, but use other illicit or licit psychoactive substances (mainly alcohol). As far as we know, research including polydrug users (cannabis use only vs. combining cannabis with other illicit substances) is extremely rare. Such comparison offers a deeper insight into neurocognitive functioning of cannabis users and polydrug cannabis users. That is why we decided to include two cannabis users groups in our study–cannabis users and polydrug users (cannabis $+ 1 \leq$ other illicit drugs). Alcohol and tobacco consumption level were controlled among groups as a potential confounding variable. As we were interested in the residual impact of cannabis, we included users who were not acutely intoxicated during the time of study and who have had at least 12 h abstinence from cannabis.

The objective of the current study was to investigate if visual episodic memory impairment lasts beyond the intoxication phase in regular cannabis users (residual effect) and in those with cannabis polydrug use, and if there are differences between users and the controls in electrical brain activity related to memory recognition. We decided to use sensitive analyses methods of memory performance based on signal detection theory (SDT) as they provide reliable information on memory qualities and are considered as an indicator of MTL-dependent declarative memories successful formation (Wais et al., 2006).

The primary aim of our study was to answer the following research questions: First, is performance on a visual episodic memory task affected in regular cannabis users compared to non-using controls? Second, are there differences in ERPs components related to recognition between these users and the controls? If this is the case, is there any relationship between electrophysiological and behavioral indicators of memory performance? Additionally, in line with previous research, we expected to observe the frontal (FN400) and parietal (LPC) Old/New effect. We also expected to find more positive LPC amplitude for correct than incorrect answers, which would predict memory accuracy.

MATERIALS AND METHODS

Participants

Eighty-two participants provided informed consent to take part in this study, 79 were included in study (reasons for exclusion of three participants are described below), while 64 were included in EEG data analysis (15 participants were excluded because of bad EEG signal). The research protocol was approved by the SWPS University Research Ethics Committee. We included in study 29 non-using controls (CG) who used cannabis on fewer than two occasions a year, and had not used in the preceding 90 days and 50 cannabis users, that were further divided in two subgroups which consist of 24 cannabis users (CU) using cannabis at least once a month (regular use) for at least 2 years (long-term use), and 26 cannabis polydrug users (PU) defined as using cannabis (at least once a month for at least 2 years) and using at least one other illicit drug in the last 3 months. Inclusion criteria for all participants were as follows: 21–42 years of age; Polish as a first language; normal or corrected to normal vision. Participants were excluded if they self-reported a history of brain injury, diagnosis of neurological disease, psychotropic medications usage. Additional criteria for cannabis users included: using cannabis at least once a month (regular use) for at least 2 years (long-term use) and negative results in screening test for cannabis use disorder (measured as $\leq$12 points at The Cannabis Use Disorder Identification Test–Revised (CUDIT-R; Adamson et al., 2010) delivered in an online recruitment questionnaire before study. CUDIT-R is often used in academic settings and is consider as having good psychometric properties and time-efficient measurement (Adamson et al., 2010; Kroon et al., 2020b). It is important to note that we invited to our study participants who declared cannabis use only (and no other illicit drugs) while recruiting to study, however, analyses of collected data in lab settings and hair sample analyses revealed polydrug use patterns in half of cannabis users. That is why we decided to include them in a study as a separate group which constitute a representative sample of cannabis users (Mitchell and Plunkett, 2000; Carlson et al., 2005; Lynskey et al., 2006; World Health Organisation, 2016).

Participants were screened for diagnosed psychiatric disorders based on self-declaration of the presence of a diagnosis by a mental health specialist, eight participants reported depression or anxiety (4 CG, 1CU, and 3 PU), all other participants reported no psychiatric disorders. **Table 1** shows demographic and substance use characteristics for 79 participants included in behavioral analyses, while **Table 2** shows the same characteristics for 64 participants included in EEG data analyses.

Cannabis users were included to CU group if they reported in self-assessment regular and long-term cannabis use (described above) and hair sample analysis reflected no other drug metabolites detected in their hair [from analysis of hair samples reflecting past 3-months exposure: THC+ (n = 12); no cannabinoid metabolites detected (n = 12)]. Cannabis polydrug users were included in PU group, if they reported in self-assessment regular and long-term cannabis use and hair sample analysis reflected other drug metabolites [from analysis of hair samples reflecting past 3-months exposure: THC+ (n = 19); 1 $\leq$ other illicit drug metabolites detected (n = 26)]. The most popular drug used in PU was MDMA (n = 17). Non-drug using controls (CG) reported cannabis use on fewer than two occasions a year, no use in the preceding 90 days (and no other illicit drug use) in self-assessment and had no drug metabolites detected in hair samples. Hair samples were not collected from eight participants from the non-drug using control group.

As mentioned before three participants were excluded from all analyses (1 in CG and 1 in PU because of current use of psychotropic medication and another one in PU because of deviant results – performance at random level indicating no engagement in task, described in "Behavioral Performance" section and psychotropic medication use). 15 participants were excluded from EEG data analyses due to technical problems with recording. Four participant reported shorter that 12 h abstinence since last cannabis use (2 PU and 2 CU), while it was highly possible that they used cannabis at night preceding experimental sessions, we decided to include them in behavioral and EEG analyses (2 CU and 1 PU, one of them was excluded because of bad EEG signal), controlling whether the results would change in case of their absence.

Participants were recruited via advertisements and social media and received the description of their IQ test score and a sample of their brains' electrical activity for their participation.

Substance Use Assessment

Substance use was assessed by the self-reported drug history questionnaire, which included the age of when cannabis use first started, years of usage, days per month of usage, dose in grams per week and time since last use. Questions about other drug use were included. Additionally, illicit substance use over the last 3 months was examined by 3 cm-hair samples. The average concentration of each hair segment was calculated and used for the final analyses. Hair samples were analyzed for 512 drugs and their metabolites by an extremely sensitive and specific analytical technique–Liquid Chromatography Mass Spectrometry (LC-MS/MS). It is important to note that one participant in CG and two in PU did not self-report psychotropic medication which were detected in hair sample analysis, these participants were excluded from the study.

What is more 23.1% of PU (n = 6) did not report use of any other illicit psychoactive substance than cannabis, but we detected them in their hair sample analyses (see **Supplementary Table 1**).

TABLE 1 | Demographic, neuropsychological and substance use characteristics for participants included in behavioral analyses.

Group (n = 79)	Controls (n = 29)	Cannabis users (n = 24)	Polydrug users (n = 26)	Three group comparisons	
				$F_{2, 76}$	P
Male/female, %	48.3/51.7	54.2/45.8	65.4/34.6		
Age, years, mean (SD)	28 (5.13)	29.29 (5.09)	27.67 (4.6)	0.611[a]	0.545
Highest level of education (Years)	16.8 (1.85)	17 (2.23)	16.3 (2.04)	0.737[a]	0.482
IQ test scores					
[1pt] WAIS scores Vocabulary	13.7 (2.06)	13.1 (1.72)	13.2 (1.7)	1.086[a]	0.343
WAIS scores Similarities	13 (1.92)	13.3 (2.22)	12.9 (1.81)	0.259[a]	0.773
WAIS scores Block design	12.9 (2.74)	13.4 (2.65)	13.6 (2.84)	0.441[a]	0.645
WAIS scores Matrix reasoning	12.9 (2.3)	13.2 (2.54)	13.2 (2.22)	0.196[a]	0.822
WAIS scores Digit span	12.1 (3.08)	13.1 (2.83)	13.2 (3.11)	1.086[a]	0.343
Diagnosed psychiatric disorders (% in group)	13.8	4.2	11.5		
Alcohol, standard drinks per week, %					
0	3.4	0	0		
<1	44.8	50	46.2		
1–3	37.9	45.8	38.5		
4–6	13.8	4.2	11.5		
7–14	0	0	3.9		
14≤	0	0	0		
Tobacco, %					
No	75.9	45.8	65.4		
Occasionally	24.1	45.8	26.9		
Regularly	0	8.3	7.7		
Cannabis use pattern					
Onset age, years	–	19.96 (5.7)	19.7 (3.43)		
Duration, years	–	9.04 (7.09)	8.61 (4.48)		
Frequency of cannabis use (% of subjects) Lifetime					
0	82.8	0	0		
Less than twice a year	17.2	0	0		
2–3 times per month	0	0	7.7		
1–3 times per week	0	33.3	23.1		
3–6 times per week	0	33.3	23.1		
Daily	0	33.3	38.7		
No answer	0	0	7.6		
Frequency of cannabis use within past 30 days					
0	100	4.2	7.7		
2–3 times per month	0	8.3	3.8		
1–3 times per week	0	33.3	26.9		
3–6 times per week	0	29.2	30.8		
Daily	0	25	30.8		
Dose in grams per week (%)					
Less than 1 g	0	33.3	30.8		
1–2 grams	0	41.7	23.1		
3–5 grams	0	25	30.8		
>5 grams	0	0	11.5		
No answer	0	0	3.8		
Dose in puffs per one use	–	7.5 (3)	7 (2.7)		
Time since last cannabis use (%)					
<12 h	0	8.3	7.7		
12–24 h	0	41.7	50		
1–3 days	0	16.7	26.9		
3–7 days	0	20.8	7.7		
7–14 days	0	4.2	0		
>14 days ago	0	8.3	3.8		
Other illicit drug use in last 30 days (% of subjects)					
0	100	100	61.5		
1 time per month	0	0	34.6		
2 < per month	0	0	3.8		
Hair sample pos (number)					
THC	–	12	19		
MDMA	–	–	17		
LSD	–	–	1		
Amphetamine	–	–	4		
Methcathinone	–	–	2		
Cocaine	–	–	8		
Cathine	–	–	1		

Note: [a]One-way ANOVAs; There were no significant differences between cannabis users (CU), polydrug users (PU), and control group (CG) in tobacco H(2) = 5.441, p = 0.066 and alcohol use patterns H(2) = 0.4, p = 0.819 as Kruskal–Wallis H tests revealed, gender X(2) = 1.659, p = 0.436, diagnosed psychiatric disorders X(2) = 3.256, p = 0.516 as χ^2-tests shown. Comparison between CU and PU on cannabis use pattern shown no significant differences in: onset age t(44) = 0.19, p = 0.852, duration t(44) = 0.249, p = 0.805, dose in puffs per one use t(43) = 0.587, p = 0.561 (series of t-tests has shown), lifetime frequency of use Z = −0.326, p = 0.745, frequency of use in last 30 days Z = −0.403, p = 0.687, dose in grams per week Z = −1.050, p = 0.294 and time since last cannabis use Z = −1.044, p = 0.279 (series of Mann–Whitney U tests). It is important to note that the ordinal data on substance use are presented in table in percentages, however, during analyses (while performing series of Mann–Whitney U or Kruskal–Wallis H tests) they were coded numerically. Hair sample analyses time-window was 3 months preceding study participation.

Procedure

Researchers collecting data were blind to the group status and had no knowledge of the cannabis/illicit drug use by the participants. Participants were asked to refrain from cannabis and other psychoactive substance use 12 h before attending the assessment session to ensure that examination would occur while they

TABLE 2 | Demographic, neuropsychological and substance use characteristics for participants included in event-related potential (ERPs) analyses.

Group (n = 64)	Controls (n = 24)	Cannabis users (n = 19)	Polydrug users (n = 21)	Three group comparisons	
				$F_{2, 61}$	P
Male/female, %	50/50	52.6/47.4	66.7/33.3		
Age, years, mean (SD)	28.3 (5.34)	28.9 (5)	27.9 (4.34)	0.246[a]	0.783
Highest level of education (Years)	16.8 (1.86)	16.8 (2.26)	16.3 (2.1)	0.42[a]	0.659
IQ test scores					
WAIS scores Vocabulary	13.8 (2.1)	13 (1.73)	13.4 (1.77)	0.933[a]	0.399
WAIS scores Similarities	13.1 (2.03)	13.2 (2.46)	13 (1.87)	0.032[a]	0.969
WAIS scores Block design	12.8 (2.95)	13.3 (2.68)	13.4 (2.71)	0.317[a]	0.73
WAIS scores Matrix reasoning	12.5 (2.28)	13.4 (2.36)	13 (1.94)	0.865[a]	0.426
WAIS scores Digit span	12.6 (2.64)	12.8 (3.08)	12.6 (3.12)	0.046[a]	0.955
Diagnosed psychiatric disorders (% in group)	16.7	5.7	4.8		
Alcohol, standard drinks per week, %					
0	4.2	0	0		
<1	41.7	47.4	47.6		
1–3	41.7	52.6	38.1		
4–6	12.5	0	14.3		
7–14	0	0	0		
14≤	0	0	0		
Tobacco, %					
No	75	47.4	61.9		
Occasionally	25	42.1	28.6		
Regularly	0	10.5	9.5		
Cannabis use pattern					
Onset age, years	–	19.47 (5.95)	20.06 (3.56)		
Duration, years	–	9.47 (7.73)	8.28 (3.97)		
Frequency of cannabis use (% of subjects) Lifetime					
0	87.5	0	0		
Less than twice a year	12.5	0	0		
2–3 times per month	0	0	9.5		
1–3 times per week	0	36.8	19		
3–6 times per week	0	26.3	28.6		
Daily	0	36.8	33.3		
No answer	0	0	9.6		
Frequency of cannabis use within past 30 days					
0	100	0	9.5		
2–3 times per month	0	10.5	4.8		
1–3 times per week	0	36.8	23.8		
3–6 times per week	0	26.3	38.1		
Daily	0	26.4	23.8		
Dose in grams per week (%)					
Less than 1 g	0	31.6	33.3		
1–2 grams	0	36.8	23.8		
3–5 grams	0	31.6	33.3		
>5 grams	0	0	4.8		
No answer	0	0	4.8		
Dose in puffs per one use	–	8.2 (2.6)	6.6 (2.7)		
Time since last cannabis use (%)					
<12 h	0	10.5	4.8		
12–24 h	0	42.1	52.4		
1–3 days	0	15.8	23.8		
3–7 days	0	21.1	10		
7–14 days	0	5.3	0		
>14 days ago	0	5.3	5		
Other illicit drug use in last 30 days (% of subjects)					
0	100	100	57.1		
1 time per month	0	0	42.9		
2 < per month	0	0	0		
Hair sample pos (number)					
THC	–	10	15		
MDMA	–	–	15		
LSD	–	–	1		
Amphetamine	–	–	4		
Methcathinone	–	–	1		
Cocaine	–	–	7		
Cathine	–	–	0		

Note: [a]One-way ANOVAs; There were no significant differences between CU, PU, and CG in tobacco H(2) = 3.459, p = 0.117 and alcohol use patterns H(2) = 0.837, p = 0.242 as Kruskal–Wallis H tests revealed, gender X(2) = 1.408, p = 0.495, diagnosed psychiatric disorders X(2) = 2.406, p = 0.3 as χ^2-tests shown. Comparison between CU and PU on cannabis use pattern shown no significant differences in: onset age t(35) = −0.359, p = 0.722, duration t(35) = 0.587, p = 0.561, dose in puffs per one use t(34) = 1.832, p = 0.076 (series of t-tests has shown), lifetime frequency of use Z = −0.566, p = 0.571, frequency of use in last 30 days Z = −0.028, p = 0.978, dose in grams per week Z = −0.282, p = 0.788 and time since last cannabis use Z = −0.511, p = 0.609 (series of Mann–Whitney U tests). It is important to note that the ordinal data on substance use are presented in table in percentages, however, during analyses (while performing series of Mann–Whitney U or Kruskal–Wallis H tests) they were coded numerically. Hair sample analyses time-window was 3 months preceding study participation.

were not intoxicated. The abstinence was verified via the self-reported time and date of last use, and no observable signs of intoxication.

All participants were tested individually in one session. The experimenter showed the participant the lab and recording equipment and described the experimental protocol before written informed consent was obtained. Participants then completed a short demographics questionnaire and answered

questions about their drug use in a separate room to protect their privacy. After that, subjects performed a shortened version of Wechsler Adult Intelligence Scale–Revised including five subtests (Wechsler, 1981; Brzeziński and Hornowska, 1998). Afterward, a 3 cm hair sample was taken from each participant (from the scalp).

Prior to the beginning of the experimental task, participants were verbally instructed as to what they would be experiencing and were shown what the procedure of EEG electrode mounting entails. Then participants were brought into a laboratory setting and seated in front of a 24 inch BenQ XL2411Z computer monitor (1,920 × 1,080 resolution, 100 Hz refresh rate) at a distance of 60 cm. Electrodes were then mounted and participants were briefly shown the EEG signal and explained how it is affected by eye blinks and muscular movements, which was a part of the procedure aimed at minimizing the amount of artifacts in the signal. The procedure was then started, and upon its completion subjects were provided with a place to wash their hair. The entire procedure lasted no more than 3 h.

Experimental Task

The episodic memory task consisted of two sessions: learning session and a recognition memory session at least 15 min later (**Figure 1**). During the learning session, participants viewed a set of 100 previously unseen, randomly selected images (each shown only once) from five categories, in random order (cars, people, animals, landscapes, food; from each category the same number of images was presented). Each image was presented for 1 s. After the image presentation, the subject assessed whether it showed an animal or not (the aim of that question was to maintain concentration on the task). The participants were informed that a test based on these images will be carried out later. Between the learning and recognition session there was at least a 15 min delay, during which a distraction task was administered (Sternberg) to prevent active rehearsal.

During the recognition session, participants viewed a set of 100 randomly selected images, 50 of which were new and 50 of which had been presented during the learning session (the same number of images were presented for each category). By using novel and complex pictures, we wanted to decrease chances of subjects to use verbalization as a memory strategy. If subjects had difficulty attaching a verbal label to the complex stimuli, then the probability of using this technique diminished. What is more, for the long delay period between learning and recognition phase (as in this study) information maintenance using a verbal strategy is not as efficient as in the case of immediate testing phase (Ellmore and Reichert, 2017). Subjects indicated whether they had seen the image before (response "old") or not (response "new") on a six-point confidence scale, along with the degree of confidence in their answer.

One hundred and fifty pictures were selected to be used in the current study from Wikimedia Commons under a Creative Commons license (stimuli from the **Figure 1** have not been used in the study and come from the author's private archive; examples of stimuli used in the study are in **Supplementary Figure 1**). The created database contains high quality photographs. All images were shown at the center of the screen of a computer placed in front of the participant. Participants responded by pressing marked buttons on a keyboard. There were no time restrictions for an answer. Stimulus presentation and recording of responses were attained using PsychoPy (v1.85.6; Peirce, 2007).

Signal Detection Theory

Performance during the memory task was analyzed with the SDT approach. An image correctly identified as old is a hit (remembered), an image incorrectly identified as old is a false alarm, an image incorrectly identified as novel is a miss (forgotten), and an image correctly identified as novel is a correct rejection.

The confidence levels in the ROC are cumulative and are calculated according to standard SDT analysis methods (Macmillan and Creelman, 2004). Each ROC curve was generated by plotting the hit rate against the false alarm rate for each confidence level of the 6-point confidence rating scale, from the most strict criterion (the proportion of hits and false alarms at the highest level of confidence) to the most liberal criterion, ending at (1,1). Recognition accuracy – the proportion of recognition (Pr) was calculated by subtracting false alarm rates from hit rates (Pr = H − FA) for each subject and was used as a measure of overall memory performance (Snodgrass and Corwin, 1988). Higher Pr values indicate better discriminability between old and new stimuli, which means better recognition accuracy.

The area under the ROC curve (AUC) is a measure of the discriminability, where 50% is chance and 100% is perfect discrimination. It is a reliable measure of performance accuracy as it is not influenced by response bias (C). *C* value ($C = -0.5[(z(\mathrm{H}) + z(\mathrm{FA})]$) is a decision bias parameter. Positive C values indicate conservative response bias, while negative indicate liberal response bias. The other measure of discriminability or sensitivity is *d*-prime [$d' = z(\mathrm{H}) - z(\mathrm{FA})$] which usually refers to the degree to which latent memory or perceptual signals from two classes of repeatedly presented stimuli–target stimuli or lure stimuli–overlap in the brain. Statistic *d'* is the standardized distance between the means of two underlying strength distributions that are assumed to be Gaussian in form and to have equal variance and take values between 0 and 4 SD. If those two distributions overlap completely, then discriminability is equal to zero. Detection capability/sensitivity (*d'*) increases as the number of hits increases and as the number of false alarms decreases.

EEG Recordings and Analysis

A 64-channel SynAmps RT Neuroscan EEG amplifier and BrainProducts actiCap Ag/AG–Cl active electrode set were used to record brain activity during task performance. All channels were recorded at 1000 Hz sampling rate. Impedances were held below 15 kΩ. All data was preprocessed offline using MATLAB environment, EEGlab (Delorme and Makeig, 2004), and ERPlab (Lopez-Calderon and Luck, 2014) software packages. The signal was initially re-referenced to a common average and then down-sampled to 250 Hz, followed by a high-pass filter (cut-off = 0.1 Hz). Signals from bad electrodes were interpolated. Movement artifacts were manually removed from the data, after

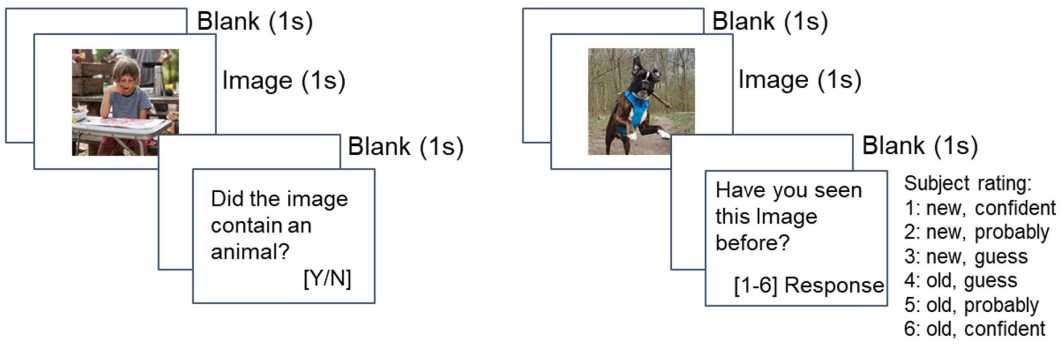

FIGURE 1 | Trial structure at learning and recognition.

which an independent component analysis (ICA) was applied for an eyeblink artifact rejection. Data epochs between −200 and 996 ms (with zero being the image presentation) were extracted. The epochs were visually inspected for remaining eye-blinks/movements and excessive muscle activity. Then the segments were averaged from trials with correct responses, according to the condition [(i) New image, (ii) Old image]. Correct old trials (Old image) were calculated for old items rated 5 or 6, while correct new trials (New image) were calculated for new items rated 1 or 2. Incorrect responses were not included in analyses as there were not enough of them. Baseline correction of ERP amplitudes was performed for the interval from -200 to 0 ms.

The FN component was measured from the 350–550 ms time window, from the F3, Fz, and F4 electrodes. LPC time window was specified to 450–750 ms, and was obtained from electrodes: C2, C4, C6, CP2, CP4, CP6, P2, P4, and P6 (**Figure 2B**). ERPs' time windows and electrodes were defined according to the literature and corresponding topographical maps (Rugg and Curran, 2007; Danker et al., 2008; Hoppstädter et al., 2015; Tsivilis et al., 2015; Kamp et al., 2016).

All analyses were conducted using IBM Corp. Released 2017. IBM SPSS Statistics for Windows, Version 25.0. Armonk, NY: IBM Corp. and MATLAB custom scripts. An alpha level of 0.05 was used for all statistical tests.

RESULTS

Group Characteristics

Group comparisons for demographic, general functioning, intelligence quotient and substance use were conducted with a series of *t*-tests or ANOVAs for continuous variables, Mann–Whitney *U* or Kruskal–Wallis *H* tests for most substance use measures (on ordinal scales) and a χ^2-test for categorical variables. There were no significant differences between groups in age, sex, education years, self-reported diagnosed psychiatric disorders, verbal and fluid intelligence, alcohol and tobacco use patterns nor for groups included in behavioral analyses nor in ERPs analyses as shown in **Tables 1**, **2**, respectively. Fluid intelligence was assessed using Matrix Reasoning and Block Design from WAIS-R. Verbal intelligence was assessed using Vocabulary, Similarities and Digit Span from WAIS-R.

There were no differences in age of cannabis use onset, frequency of cannabis use, dose in grams per week, frequency of cannabis use within the past 30 days and dose in puffs per one use between CU and PU (see **Tables 1**, **2**).

Behavioral Performance

Memory performance estimates were calculated using the ROC toolbox for MATLAB (Koen et al., 2017). **Table 3** lists memory performance indices for three groups. One-way ANOVA was computed for each of the behavioral indices to test for significant differences. There were no differences in *H* [$F(2, 76) = 0.472$, $p = 0.626$, $\eta^2 = 0.12$], FA [$F(2, 76) = 1.510$, $p = 0.228$, $\eta^2 = 0.038$], Pr [$F(2, 76) = 0.159$, $p = 0.854$, $\eta^2 = 0.004$], *d'* [$F(2, 73) = 0.238$, $p = 0.789$, $\eta^2 = 0.006$], AUC [$F(2, 76) = 0.648$, $p = 0.526$, $\eta^2 = 0.017$], and c [$F(2, 73) = 1.574$, $p = 0.214$, $\eta^2 = 0.041$] between groups. The statistically insignificant differences were preserved when we excluded four participants who reported last use less than 12 h before testing from analyses.

Participants successfully indicated the presence or absence of an animal in 94% of the trials, which means they were focused on the task. In general, subjects performed well. As shown in **Table 3** the average sensitivity (*d'*) and recognition accuracy (Pr) were high. Participant's reports of confidence fit well to their performance (**Figures 3A,C**) and the ROC were assymetrical (**Figures 3B,D**) with high average AUC values. Only one participant AUC was 0.58, while 0.5 indicated performance at a random level, this subject was excluded from the study due to deviant results and lack of engagement in the task.

We also compared reaction times between correct old and correct new trials. Correct old trials were calculated for old items rated 5 or 6, while correct new trials were calculated for new items rated 1 or 2, the same as in EEG analyses. Reaction times were individually standardized and the observations above the three standard deviations (for each participant separately) were excluded from data analysis. RT for each group separately is reported in **Table 3**. Repeated-measures ANOVAs, with Old/New

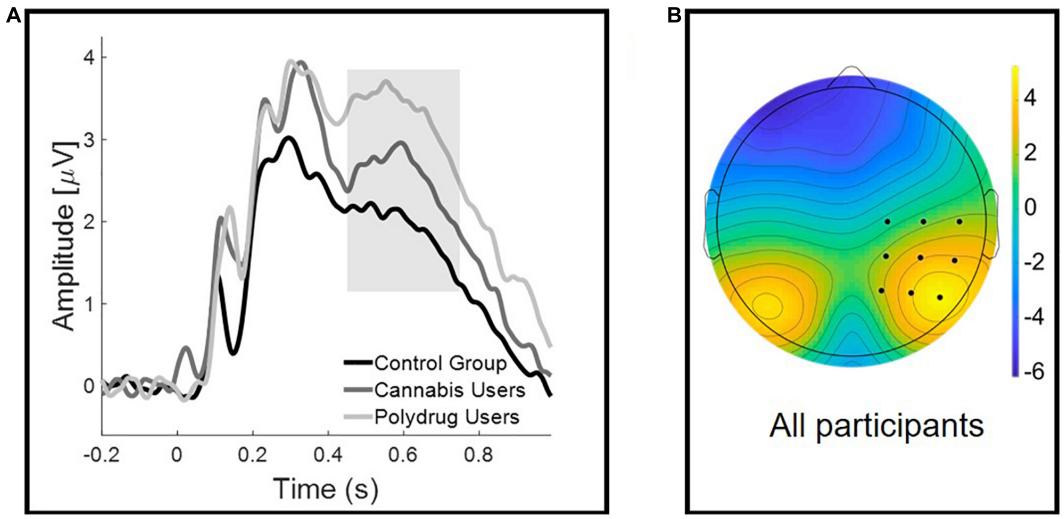

FIGURE 2 | Group differences in the Late Positive Component (LPC). **(A)** Grand-averaged waveforms at a representative right centro-occipital cluster for correctly recognized images. The shaded area represents the late (450–750 ms) time window used for the analyses. **(B)** Scalp topographies of the mean activity in 450–750 ms time window.

(two levels: Old Image vs. New Image) as a within-participant factor, and Group (three levels: CG vs. CU vs. PU) as a between-participant factor, were performed on the reaction times. ANOVA analysis revealed that there was neither a main effect of Old/New [$F(1, 76) = 0.002$, $p = 0.962$, $\eta^2 = 0.00$] nor the Old/New × Group interaction [$F(2, 76) = 0.078$, $p = 0.925$, $\eta^2 = 0.002$]. All effects for RT remain statistically insignificant when we excluded four participants who reported last use less than 12 h before testing from analyses.

Electrophysiological Data

Frontal Negativity

Figure 4A illustrates the grand average ERPs for correctly recognized old and new images. Repeated-measures ANOVA, with Old/New (two levels: Old Image vs. New Image) as a within-participant factor, and Group (three levels: CG vs. CU vs. PU) as a between-participant factor, was performed on the mean FN amplitudes. ANOVA analysis revealed that there was a significant Old/New effect [$F(1, 60) = 5.211$, $p = 0.026$, $\eta^2 = 0.08$] but no Group effect [$F(2, 60) = 0.35$, $p = 0.706$, $\eta^2 = 0.012$], nor Old/New × Group interaction [$F(2, 60) = 0.019$, $p = 0.981$, $\eta^2 = 0.001$].

Similarly, analysis, which excluded participants with shorter than 12 h abstinence since last cannabis use, showed significant Old/New effect [$F(1, 57) = 5.882$, $p = 0.018$, $\eta^2 = 0.094$] and no Group effect [$F(2, 57) = 0.305$, $p = 0.738$, $\eta^2 = 0.011$], nor Old/New × Group interaction [$F(2, 57) = 0.032$, $p = 0.969$, $\eta^2 = 0.001$]. Specific values of the FN are presented in **Table 4**.

TABLE 3 | Recognition memory performance – mean (SD).

	Control group	Cannabis users	Polydrug users
Hits (*H*)	0.78 (0.11)	0.77 (0.12)	0.75 (0.11)
False alarms (FAs)	0.12 (0.08)	0.1 (0.06)	0.09 (0.06)
Discrimination indices			
Pr	0.66 (0.13)	0.68 (0.13)	0.66 (0.12)
d'	2.08 (0.58)	2.19 (0.51)	2.12 (0.5)
AUC	0.89 (0.59)	0.89 (0.57)	0.88 (0.62)
Response bias indices			
C	0.2 (0.31)	0.25 (0.27)	0.35 (0.29)
Reaction time (RT)			
Old correct RT [s]	0.672 (0.3)	0.741 (0.21)	0.639 (0.27)
New correct RT [s]	0.681 (0.34)	0.728 (0.21)	0.646 (0.44)

Note Reaction times reported in seconds [s].

Late Positive Component

Figure 2A illustrates the mean amplitudes of correctly recognized images for each group. As mentioned before, we did not include data for incorrectly recognized images in the analysis.

Repeated-measures ANOVA, with Old/New (two levels: Old Image vs. New Image) as a within-participant factor, and Group (three levels: CG vs. CU vs. PU) as a between-participant factor, was performed on the mean LPC amplitudes. ANOVA analysis revealed that there was a significant Group effect [$F(2, 60) = 20.478$, $p = 0.017$, $\eta^2 = 0.127$] but no main effect of Old/New [$F(1, 60) = 0.706$, $p = 0.404$, $\eta^2 = 0.012$], nor Old/New × Group interaction [$F(2, 60) = 1.336$, $p = 0.271$, $\eta^2 = 0.043$].

Bonferroni-corrected multiple comparisons revealed significant difference between CG and PU ($p = 0.013$) but not between CG and CU ($p = 0.593$) or CU and PU ($p = 0.425$).

Analysis, which excluded participants with shorter than 12 h abstinence since last cannabis use, showed significant Group effect [$F(2, 57) = 3.961$, $p = 0.025$, $\eta^2 = 0.122$] and almost significant Old/New × Group interaction [$F(2, 57) = 2.693$, $p = 0.076$, $\eta^2 = 0.086$], but no main Old/New effect [$F(1, 57) = 0.918$, $p = 0.342$, $\eta^2 = 0.016$]. Bonferroni-corrected multiple

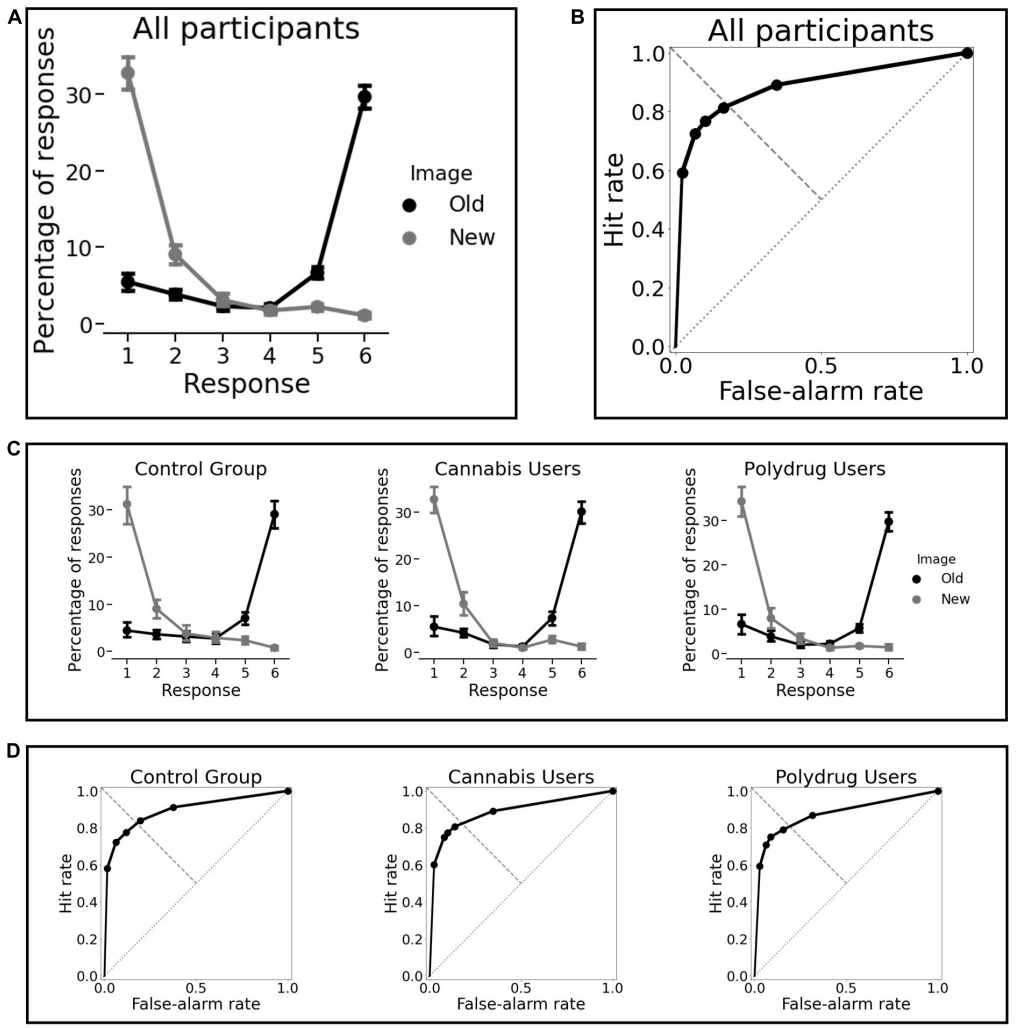

FIGURE 3 | Behavioral results. **(A)** Percentage of responses "old" and "new" as a function of confidence for all participants As *X*-axis corresponds to participants' rating (1 = new, confident; 6 = old, confident) the highest probabilities for responses Old and New were 6 and 1. **(B)** The average receiver operating characteristic of all subjects. **(C)** Percentage of responses "old" and "new" as a function of confidence for each group separately. **(D)** The average receiver operating characteristic for each group separately. There were no significant differences between groups for any memory performance indicator and we can observe obvious similarities in memory performance across groups.

comparisons revealed significant difference between CG and PU ($p = 0.02$) but not between CG and CU ($p = 0.571$) or CU and PU ($p = 0.636$).

As previous studies have reported, LPC is usually lateralized on the left centro-occipital sphere (Curran and Cleary, 2003; Rugg and Curran, 2007), we decided to duplicate presented analyses on the left-lateralized electrode cluster (electrodes: C1, C3, C5, CP1, CP3, CP5, P1, P3, and P5). Then it was revealed that neither Group difference [$F(2, 60) = 1.174$, $p = 0.317$, $\eta^2 = 0.038$], Old/New effect [$F(1, 60) = 1.815$, $p = 0.183$, $\eta^2 = 0.029$], nor Old/New × Group interaction [$F(2, 60) = 0.121$ $p = 0.886$, $\eta^2 = 0.004$] were significant on the left-lateralized cluster.

Specific values of the LPC are presented in **Table 5**.

Regression Models

We did not observe any significant differences in any of the memory performance indices between groups, which was also observed in some previous studies (Smith et al., 2017; Sagar and Gruber, 2019). While the Old/New effect was not observed for the LPC component, there were significant differences in mean amplitude between groups in this component for correct answers. As LPC reflects attention, motivation, higher cognitive function, decision accuracy, and memory judgments (MacNamara et al., 2011; Wiens and Syrjänen, 2013; Ratcliff et al., 2016; Yang et al., 2019), we decided to look for its influence on the measure of overall memory performance – recognition accuracy (Pr). To test whether the mean amplitude of the LPC component predicts participants' Pr, a linear regression was calculated.

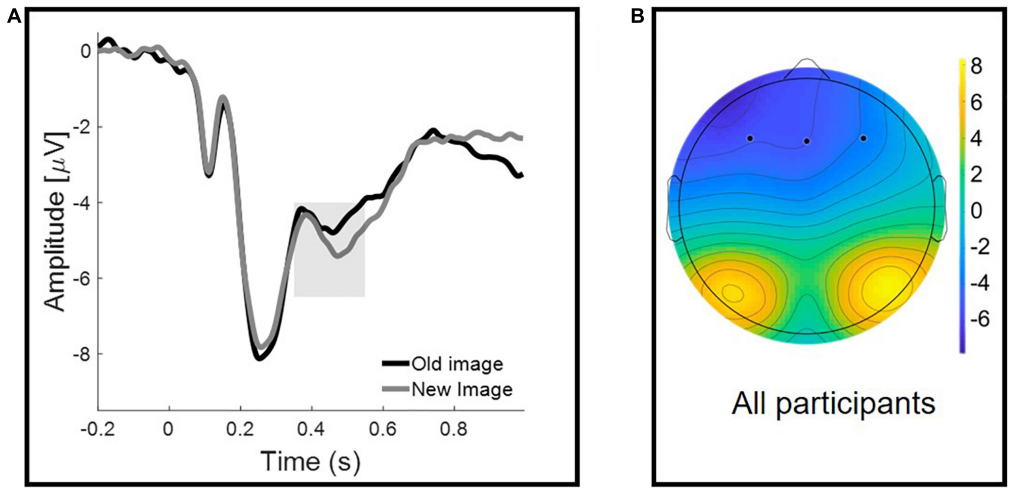

FIGURE 4 | Frontal Negativity for Old/New effect reflecting familiarity. **(A)** Grand-averaged waveforms at a representative frontal cluster for correctly recognized old and new images. The shaded area represents the early (350–550 ms) time window used for the analyses. **(B)** Scalp topographies of the mean activity in 350–550 ms time window.

TABLE 4 | Means and standard deviations of the mean amplitudes (μV) of the Frontal Negativities for all groups and conditions.

	All participants			Excluding participants with shorter than 12 h abstinence since last cannabis use	
	Control group	**Cannabis users**	**Polydrug users**	**Cannabis users**	**Polydrug users**
Old image	-3.92 (3.60)	-4.80 (4.38)	-4.54 (3.26)	-4.61 (4.62)	-4.65 (3.31)
New image	-4.45 (3.26)	-5.28 (3.69)	-4.98 (2.95)	-5.23 (3.90)	-5.12 (2.95)
Mean	-4.16 (3.35)	-5.04 (3.97)	-4.74 (3.00)	-4.90 (4.20)	-4.88 (3.01)

TABLE 5 | Means and standard deviations of the mean amplitudes (μV) of the Late Positive Components for all groups and conditions.

	All participants			Excluding participants with shorter than 12 h abstinence since last cannabis use	
	Control group	**Cannabis users**	**Polydrug users**	**Cannabis users**	**Polydrug users**
Old image	2.01 (1.75)	2.71 (1.82)	3.22 (1.45)	2.81 (1.91)	3.16 (1.45)
New image	1.88 (1.63)	2.42 (1.45)	3.37 (1.36)	2.41 (1.45)	3.38 (1.40)
Mean	1.95 (1.67)	2.57 (1.55)	3.30 (1.33)	2.61 (1.61)	3.27 (1.36)

While the initial model, containing three groups, turned out to be statistically insignificant [$F(3, 59) = 1.902$, $p = 0.139$] with an $R^2 = 0.088$, it contained significant influence of the mean LPC amplitude ($b = 0.027$, $p = 0.043$), interaction between the amplitude and the Group ($b = -0.023$, $p = 0.045$), and almost significant influence of the Group ($b = 0.07$, $p = 0.05$). As the insignificance of the initial model may be a result of an actual relation existing in only one group, we decided to look at each group independently (see **Figure 5**). Then it was revealed that while the influence of the LPC was significant in CG ($b = 0.034$, $p = 0.02$), both CU ($b = -0.02$, $p = 0.248$) and PU ($b = -0.002$, $p = 0.903$) models were statistically insignificant.

Similarly, a model computed with extension of participants with abstinence shorter than 12 h since last cannabis use was statistically insignificant [$F(3, 56) = 2.104$, $p = 0.11$, R2 = 0.101], but contained significant influence of the mean LPC amplitude ($b = 0.028$, $p = 0.036$), Group ($b = 0.077$, $p = 0.037$) and interaction between those ($b = -0.025$, $p = 0.028$). Auxiliary models, created for each of groups independently, presented similar results (CG: $b = 0.034$, $p = 0.02$; CU: $b = -0.023$, $p = 0.156$; PU: $b = -0.004$, $p = 0.844$).

DISCUSSION

We believe this is the first EEG investigation of prolonged THC exposure on memory processes in regular cannabis users and regular cannabis polydrug users. The results showed alterations in recognition memory processing, which were not manifested on a behavioral level, but were prominent on a physiological level and expressed the most in cannabis polydrug users.

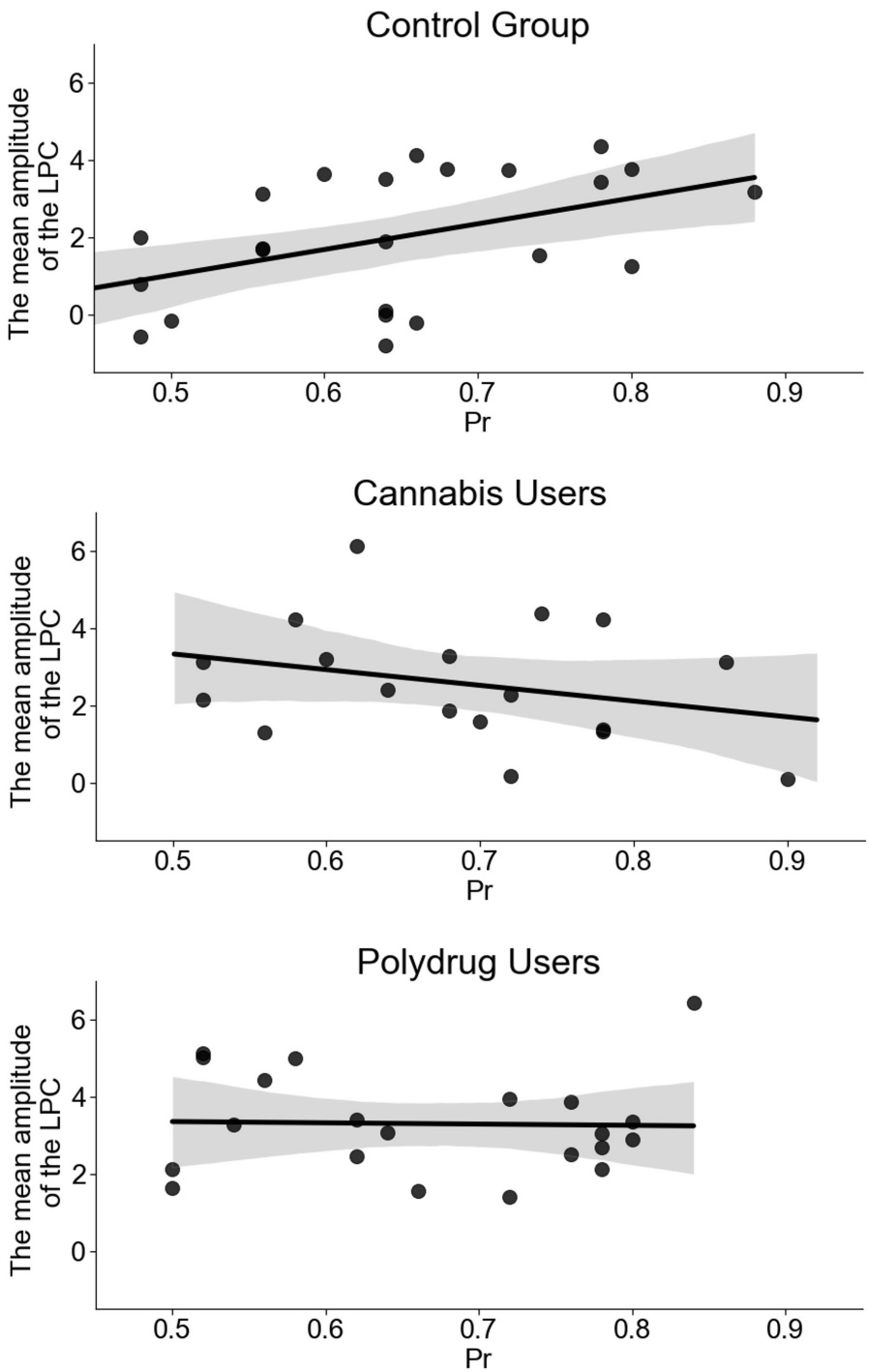

FIGURE 5 | Regression models for recognition accuracy indicator (proportion of recognition, Pr = Hits - False Alarms) as a function of the mean amplitude of the LPC for each group separately. Scatter plot presents a significant relation between the mean amplitude of the LPC and Pr (r = 0.472, p = 0.02) in control group (CG) and statistically insignificant relations in cannabis users (CU) (r = −0.287, p = 0.248) and polydrug users (PU) (r = −0.028, p = 0.903).

Based on the experimental procedure used in our study and calculated behavioral indicators, we were able to precisely analyse participants' performance. Subjects' confidence rating charts and receiver operating characteristics, averaged for each group separately (**Figure 3**) indicate they had an overall good quality of their memories. It is also a signature that MTL-dependent declarative memories were formed successfully (Wais et al., 2006).

The results have shown significant differences in mean LPC amplitudes for correct answers between CG and PU. As mentioned before the LPC is not unique for episodic memory recollection, but is engaged in decision making during memory judgments and related to various cognitive functions (Brezis et al., 2016; Ratcliff et al., 2016). Increased brain activity during cognitive task could be interpreted as a compensational, stronger "neurophysiological" effort to overcome drug-induced brain dysfunction and maintain normal behavioral performance (Zeineh et al., 2003; Bhattacharyya et al., 2009; Roberts et al., 2009), possibly higher engagement of attentional and motivational resources in this case.

It is quite surprising, however, that we have not found significant differences in LPC between CU and CG or PU. The possible explanation is that use of other illicit drugs among cannabis users altered brain function much more than cannabis use only, and the differences observed between the control group and cannabis users in previous studies were enhanced by co-occurring use of other illicit drugs. The most popular other illicit psychoactive substance among PU in our study was MDMA. As previous studies have shown MDMA is the most commonly used drug in polydrug context with cannabis (Scholey et al., 2004; Daumann et al., 2004; European Monitoring Centre for Drugs and Drug Addiction (EMCDDA), 2018). MDMA acts via serotonergic receptors and may cause disturbances of serotonergic pathways (Benningfield and Cowan, 2013) and electrophysiological changes that reflect recruitment of additional resources to perform cognitive tasks (Roberts et al., 2018). The other psychoactive substances used among PU, such as cocaine, amphetamine, and others may interact in a complex way and cumulative effect may impact the further increased brain activity (Zilverstand et al., 2018; Khajehpour et al., 2019). Still, little is known about the interactive impact of various illicit drugs on the human brain (Gouzoulis-Mayfrank and Daumann, 2006; Van Dam et al., 2008; Shevlin et al., 2017), regardless of the observed high frequency of polydrug use pattern.

What is more, the age of cannabis use onset in our participants (∼20) could play an important role in shaping our results (no difference between CU and CG), as usually in cannabis research the age of onset is lower (during adolescence). That is why, our results may be not possible to generalize on cannabis users that started regular cannabis use much earlier. At the same time, our participants were middle-aged and have shown long lifetime cannabis use (∼10 years), giving us the opportunity to investigate higher cumulative effects of cannabis use. It is worthy to note that Smith et al. (2017) observed larger LPC (indexing recollection) in heavy drinkers compared to a control group, but not in cannabis users, without any concomitant behavioral impairments.

We have also examined the direct relationship between the ERP components' amplitude and memory performance. We checked whether the value-modulated LPC predicted behavioral measures of a participant's recognition accuracy. The initial regression model containing LPC × Group interaction was statistically insignificant, but further exploration revealed that LPC amplitude predicts recognition accuracy in the control group only. This result suggests that the higher the involvement of the processes reflected in the LPC, the better non-drug using participants are in discriminating between old and new stimuli. Previous study has shown a positive correlation between performance level on recognition memory task and LPC amplitude only in healthy older people, but it was not observed in participants suffering from mild cognitive impairment (Waninger et al., 2018).

Due to a small CU sample size in our study, this explanation should be treated with caution and should be further explored in a larger sample. It is important to note that LPC is located predominantly over posterior sites, often but not always exhibiting a left-sided maximum, while in our study it showed a right-sided maximum. It is not common but was observed before (Tanguay et al., 2018; Hebscher et al., 2020).

Consistent with previous studies, we have observed a frontal old/new effect thought to index familiarity, ERPs elicited by previously seen stimuli (Old Images) were significantly more positive in all groups (Curran, 2000; Paller et al., 2007). However, we did not observe a parietal old/new effect (LPC). The absence of LPC mean amplitude changes, depending on whether stimuli was correctly identified as old or new, has been reported in the literature before (Danker et al., 2008; Wolk et al., 2009; Addante et al., 2012) and could be interpreted as reduction in recollective processes following successful familiarity (Tibon and Henson, 2015; Kamp et al., 2016). The FN400 and LPC are usually interpreted in the context of the dual-process model of recognition memory, where the LPC is the index of recollection of episodic details about the prior stimulus encounter, and the FN400 marks item-based, context free, familiarity assessment (Yonelinas, 2002). It should be mentioned that there is a still ongoing debate on the interpretation of these two ERP components of recognition memory and the processes that underlie them (Voss and Federmeier, 2011; Bridger et al., 2012).

However, what was a little bit surprising for us, we did not see any visible differences in behavioral indices of visual episodic memory between groups. In light of our results, it should be noted that studies employing neuroimaging and electrophysiological techniques proved to be more sensitive to detect drug effects (especially in non-intoxicated users) than traditional, behavioral measures (Jager et al., 2007; Nestor et al., 2008; Cousijn et al., 2014). Altered brain activation patterns in cannabis users relative to non-users were observed across numerous brain regions even when no differences in task performance were detected (Smith et al., 2017; Sagar and Gruber, 2019).

In line with these results, it is interesting to note that LPC amplitude in general was higher in CU and PU (**Figure 2A**), but it did not predict memory accuracy. It may reflect subtle alterations in neural circuits engaged in memory processing in both CU and PU. Possibly, LPC rather reflects compensatory mechanisms mentioned before, and higher attentional and/or motivational engagement in PU and CU, as it does not correspond to a recognition accuracy indicator in this case.

Strengths and Limitations

The use of hair sample analysis to define groups exposed or not to cannabis and other drugs is limited in that it only provides

information on exposure over the prior 3 months (when 3 cm of hair is analysed). Unfortunately, hair sample analysis was not available for eight participants from the control group. However, it is important to note that biological measures of drug use (illicit substances or psychotropic medication) is rarely used in non-using controls in the research field on cannabis (or other illicit drugs) and self-reported measures dominate for CU (Smith et al., 2017; Khajehpour et al., 2019; Rangel-Pacheco et al., 2020) or the urine/hair sample analyses detect cannabis use only (no other illicit drugs nor psychotropic medication) (Yücel et al., 2016; Prashad et al., 2018). That is why we would consider it as an additional value of our study and exclusion of this participants form analyses would make CG too small to obtain reliable results. What is more, the obtained memory performance results and ERPs are in line with previous studies and do not show any artifacts (which possible drug use in this group could generate). The reason why hair samples were not collected in this group was that participants did not want to lose a big amount of hair (diameter of a pencil) required for hair sample analyses, because of esthetic reasons. As the observed tendency among participants was rather to under-report drug use in cannabis users (mainly in case of other illicit drug use) we assumed their self-reports were reliable. In general, the best practice is to use hair analysis with complementary tests for urine and blood analysis as THC (similar to other drugs) takes about 2 weeks to reach the hair shaft (Shah et al., 2019). Unfortunately, we did not use these measurements in our study, because of funding limitations and ethical concerns. At the same time, combined self-reported and objective hair sample analyses were also the strength of our study. While there is growing popularity of using a combination of self-report and objective drug use assessment in research on cannabis and neurocognitive functioning, the most popular are drug urine tests, which despite all advantages have a relatively short time-frame for various illicit drug detection. The hair sample analysis provides the opportunity to detect many drugs metabolites among much longer time-frames (restricted by subjects' hair length and research funding limits), which makes it a suitable tool for long-term drug use assessment. This kind of analysis has been used in research on cannabis and neurocognition, but it has been restricted to cannabinoid metabolites detection only. While hair sample analyses did not prove THC presence in all cannabis users, it allowed to exclude other illicit drug use in CU. Previous research showed that the sensitivity of THC detection in hair is almost 80% in heavy cannabis smokers compared to light and non- cannabis users, but fell to 55% in any cannabis users compared to non-cannabis users (Taylor et al., 2017).

The highly heterogeneous PU group is an important limitation in our study (**Supplementary Table 1**), because we did not include it preliminary in our research plan (as described in "Participants" section). However, we decided to include them in a study as a separate group (PU) as constituting a representative sample of cannabis users (Mitchell and Plunkett, 2000; Carlson et al., 2005; Lynskey et al., 2006; World Health Organisation, 2016). That is why we did not collect detailed information about the polydrug use pattern (e.g., frequency of use, lifetime use, whether different substances were used concurrently or sequentially, substance dependence) as we intended to recruit cannabis users only. This information would be beneficial to understand better the issues of polydrug use and select a more homogeneous group in future studies.

Moreover, our measure of psychiatric symptoms is based exclusively on self-report (self-declaration of the presence of a diagnosis by a mental health specialist), rather than clinical evaluation or a structured interview. It would be important to engage medical professionals in future studies for psychiatric diagnosis as it is an important factor in polydrug use context. While we used CUDIT-R to screen the severity of use-related problems and recruit only participants with negative results in screening for cannabis use disorder, we did not include in our study screening for the severity of other illicit drug use-related problems.

The control group in our study did not include only individuals who had never used cannabis, but some of them reported minimal use in their lifetime (<50 occasions). It is considered acceptable and attenuates a potential cumulative effect of cannabis use (Sagar and Gruber, 2019). We are aware of the modest group sizes in our study, however, most neuroimaging investigations in cannabis research have similar sample sizes (Sagar and Gruber, 2019). While these samples appear to be large enough to detect between-group differences in brain activation patterns, it should be noted that our study has preliminary character and further research is needed.

In summary, the findings of the present study indicate that, when patterns of cannabis and polydrug use are examined in greater detail, the unique effect of cannabis consumption seems to be greatly attenuated. There was no significant effect for cannabis alone, but only for cannabis in conjunction with other illicit drugs, which most likely produce the biggest disturbance in brain function.

DATA AVAILABILITY STATEMENT

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

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the SWPS University Research Ethics Committee. The patients/participants provided their written informed consent to participate in this study. Written informed consent was obtained from the minor(s)' legal guardian/next of kin for the publication of any potentially identifiable images or data included in this article.

AUTHOR CONTRIBUTIONS

AAB, and AB: study the design. AAB, NJ, and MG: data processing and data analysis. NG and AP-C: data analysis. AAB, AB, NJ, MG, NG, and AP-C: writing the manuscript. All authors contributed to the article and approved the submitted version.

FUNDING

This research and open access of this article was financed by the Ministry of Science and Higher Education in Poland under the 2019–2022 program "Regional Initiative of Excellence", project number 012/RID/2018/19.

ACKNOWLEDGMENTS

We cordially thank Monika Myśliwiec, Klaudia Krystecka, Sandra Łuczak, Martyna Bryłka, Klaudia Wojtal, Paulina Lewandowska, and Agnieszka Ozimek for help with data collection, and Wiktoria Podolewska for linguistic proofreading. We would also like to express gratitude to Free Cannabis Society for support in the rectrutation process.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fnhum.2021.677793/full#supplementary-material

REFERENCES

Adamson, S. J., Kay-Lambkin, F. J., Baker, A. L., Lewin, T. J., Thornton, L., Kelly, B. J., et al. (2010). An improved brief measure of cannabis misuse: the cannabis use disorders identification test-revised (CUDIT-R). *Drug Alcohol Depend.* 110, 137–143. doi: 10.1016/j.drugalcdep.2010.02.017

Addante, R. J., Ranganath, C., and Yonelinas, A. P. (2012). Examining ERP correlates of recognition memory: evidence of accurate source recognition without recollection. *Neuroimage* 62, 439–450. doi: 10.1016/j.neuroimage.2012.04.031

Ameri, A. (1999). The effects of cannabinoids on the brain. *Prog. Neurobiol.* 58, 315–348.

Benningfield, M. M., and Cowan, R. L. (2013). Brain serotonin function in MDMA (ecstasy) users: evidence for persisting neurotoxicity. *Neuropsychopharmacology* 38:253. doi: 10.1038/npp.2012.178

Bhattacharyya, S., Fusar-Poli, P., Borgwardt, S., Martin-Santos, R., Nosarti, C., O'Carroll, C., et al. (2009). Modulation of mediotemporal and ventrostriatal function in humans by Δ9-tetrahydrocannabinol. *Arch. Gen. Psychiatry* 66, 442–457. doi: 10.1001/archgenpsychiatry.2009.17

Böcker, K. B. E., Gerritsen, J., Hunault, C. C., Kruidenier, M., Mensinga, T. T., and Kenemans, J. L. (2010). Cannabis with high Δ9-THC contents affects perception and visual selective attention acutely: an event-related potential study. *Pharmacol. Biochem. Behav.* 96, 67–74. doi: 10.1016/j.pbb.2010.04.008

Brewer, J. B., Zhao, Z., Desmond, J. E., Glover, G. H., and Gabrieli, J. D. E. (1998). Making memories: brain activity that predicts how well visual experience will be remembered. *Science* 281, 1185–1187. doi: 10.1126/science.281.5380.1185

Brezis, N., Bronfman, Z. Z., Yovel, G., and Goshen-Gottstein, Y. (2016). The electrophysiological signature of remember–know is confounded with memory strength and cannot be interpreted as evidence for dual-process theory of recognition. *J. Cogn. Neurosci.* 29, 322–336. doi: 10.1162/jocn_a_01053

Bridger, E. K., Bader, R., Kriukova, O., Unger, K., and Mecklinger, A. (2012). The FN400 is functionally distinct from the N400. *Neuroimage* 63, 1334–1342. doi: 10.1016/j.neuroimage.2012.07.047

Broyd, S. J., Van Hell, H. H., Beale, C., Yücel, M., and Solowij, N. (2016). Acute and chronic effects of cannabinoids on human cognition – A systematic review. *Biol. Psychiatry* 79, 557–567. doi: 10.1016/j.biopsych.2015.12.002

Brzeziński, J., and Hornowska, E. (1998). *Skala Inteligencji Wechslera WAIS-R. Polska adaptacja, Standaryzacja, Normalizacja i Wykorzystanie w Diagnostyce Psychologicznej* (wyd. 2). Warsaw: Wydawnictwo Naukowe PWN.

Carlson, R. G., Wang, J., Falck, R. S., and Siegal, H. A. (2005). Drug use practices among MDMA/ecstasy users in Ohio: a latent class analysis. *Drug Alcohol Depend.* 79, 167–179. doi: 10.1016/j.drugalcdep.2005.01.011

Cousijn, J., Vingerhoets, W. A. M., Koenders, L., de Haan, L., van den Brink, W., Wiers, R. W., et al. (2014). Relationship between working-memory network function and substance use: a 3-year longitudinal fMRI study in heavy cannabis users and controls. *Addict. Biol.* 19, 282–293. doi: 10.1111/adb.12111

Curran, H. V., Freeman, T. P., Mokrysz, C., Lewis, D. A., Morgan, C. J. A., and Parsons, L. H. (2016). Keep off the grass? Cannabis, cognition and addiction. *Nat. Rev. Neurosci.* 17, 293–306. doi: 10.1038/nrn.2016.28

Curran, T. (2000). Brain potentials of recollection and familiarity. *Mem. Cogn.* 28, 923–938. doi: 10.3758/bf03209340

Curran, T., and Cleary, A. M. (2003). Using ERPs to dissociate recollection from familiarity in picture recognition. *Cogn. Brain Res.* 15, 191–205. doi: 10.1016/s0926-6410(02)00192-1

Danker, J. F., Hwang, G. M., Gauthier, L., Geller, A., Kahana, M. J., and Sekuler, R. (2008). Characterizing the ERP Old-New effect in a short-term memory task. *Psychophysiology* 45, 784–793. doi: 10.1111/j.1469-8986.2008.00672.x

Daumann, J., Hensen, G., Thimm, B., Rezk, M., Till, B., and Gouzoulis-Mayfrank, E. (2004). Self-reported psychopathological symptoms in recreational ecstasy (MDMA) users are mainly associated with regular cannabis use: further evidence from a combined cross-sectional/longitudinal investigation. *Psychopharmacology* 173, 398–404. doi: 10.1007/s00213-003-1719-0

Delorme, A., and Makeig, S. (2004). EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *J. Neurosci. Methods* 134, 9–21. doi: 10.1016/j.jneumeth.2003.10.009

Desgranges, B., Baron, J. C., and Eustache, F. (1998). the functional neuroanatomy of episodic memory: the role of the frontal lobes, the hippocampal formation, and other areas. *Neuroimage* 8, 198–213. doi: 10.1006/nimg.1998.0359

Doss, M. K., Weafer, J., Gallo, D. A., and de Wit, H. (2020). Δ9-Tetrahydrocannabinol during encoding impairs perceptual details yet spares context effects on episodic memory. *Biol. Psychiatry Cogn. Neurosci. Neuroimaging* 5, 110–118. doi: 10.1016/j.bpsc.2019.08.007

Ellmore, T. M., and Reichert, C. P. (2017). Early and late compo- nents of EEG delay activity correlate differently with scene working memory performance. *PLoS One* 12:e0186072. doi: 10.1371/journal.pone.0186072

European Monitoring Centre for Drugs and Drug Addiction (EMCDDA) (2018). *European Drug Report 2018: Trends and Developments*. Avaiable online at: https://www.emcdda.europa.eu/system/files/publications/8585/20181816_TDAT18001ENN_PDF.pdf (accessed September 22, 2020).

Figueiredo, P. R., Tolomeo, S., Steele, J. D., and Baldacchino, A. (2020). Neurocognitive consequences of chronic cannabis use: a systematic review and meta-analysis. *Neurosci. Biobehav. Rev.* 108, 358–369. doi: 10.1016/j.neubiorev.2019.10.014

Finnigan, S., Humphreys, M. S., Dennis, S., and Geffen, G. (2002). ERP 'old/new'effects: memory strength and decisional factor (s). *Neuropsychologia* 40, 2288–2304. doi: 10.1016/s0028-3932(02)00113-6

Gouzoulis-Mayfrank, E., and Daumann, J. (2006). The confounding problem of polydrug use in recreational ecstasy/MDMA users: a brief overview. *J. Psychopharmacol.* 20, 188–193. doi: 10.1177/0269881106059939

Grant, I., Gonzalez, R., Carey, C. L., Natarajan, L., and Wolfson, T. (2003). Non-acute (residual) neurocognitive effects of cannabis use: a meta-analytic study. *J. Int. Neuropsychol. Soc.* 9, 679–689. doi: 10.1017/s1355617703950016

Gunseli, E., Meeter, M., and Olivers, C. N. (2014). Is a search template an ordinary working memory? Comparing electrophysiological markers of working memory maintenance for visual search and recognition. *Neuropsychologia* 60, 29–38. doi: 10.1016/j.neuropsychologia.2014.05.012

Hebscher, M., Ibrahim, C., and Gilboa, A. (2020). Precuneus stimulation alters the neural dynamics of autobiographical memory retrieval. *Neuroimage* 210:116575. doi: 10.1016/j.neuroimage.2020.116575

Herkenham, M., Lynn, A. B., Little, M. D., Johnson, M. R., Melvin, L. S., de Costa, B. R., et al. (1990). Cannabinoid receptor localization in brain. *Proc. Natl. Acad. Sci. U.S.A.* 87, 1932–1936.

Hoppstädter, M., Baeuchl, C., Diener, C., Flor, H., and Meyer, P. (2015). Simultaneous EEG-fMRI reveals brain networks underlying recognition memory ERP old/new effects. *Neuroimage* 116, 112–122. doi: 10.1016/j.neuroimage.2015.05.026

Jager, G., Van Hell, H. H., De Win, M. M. L., Kahn, R. S., Van Den Brink, W., Van Ree, J. M., et al. (2007). Effects of frequent cannabis use on hippocampal activity during an associative memory task. *Eur. Neuropsychopharmacol.* 17, 289–297. doi: 10.1016/j.euroneuro.2006.10.003

Kamp, S. M., Bader, R., and Mecklinger, A. (2016). The effect of unitizing word pairs on recollection versus familiarity-based retrieval-further evidence from ERPs. *Adv. Cogn. Psychol.* 12, 168–177. doi: 10.5709/acp-0196-2

Khajehpour, H., Makkiabadi, B., Ekhtiari, H., Bakht, S., Noroozi, A., and Mohagheghian, F. (2019). Disrupted resting-state brain functional network in methamphetamine abusers: a brain source space study by EEG. *PLoS One* 14:e0226249. doi: 10.1371/journal.pone.0226249

Koen, J. D., Barrett, F. S., Harlow, I. M., and Yonelinas, A. P. (2017). The ROC Toolbox: a toolbox for analyzing receiver-operating characteristics derived from confidence ratings. *Behav. Res. Methods* 49, 1399–1406. doi: 10.3758/s13428-016-0796-z

Kroon, E., Kuhns, L., and Cousijn, J. (2020a). The short-term and long-term effects of cannabis on cognition: recent advances in the field. *Curr. Opin. Psychol.* 38, 49–55. doi: 10.1016/j.copsyc.2020.07.005

Kroon, E., Kuhns, L., Hoch, E., and Cousijn, J. (2020b). Heavy cannabis use, dependence and the brain: a clinical perspective. *Addiction* 115, 559–572. doi: 10.1111/add.14776

Lopez-Calderon, J., and Luck, S. J. (2014). ERPLAB: an open-source toolbox for the analysis of event-related potentials. *Front. Hum. Neurosci.* 8, 213. doi: 10.3389/fnhum.2014.00213

Lynskey, M. T., Agrawal, A., Bucholz, K. K., Nelson, E. C., Madden, P. A., Todorov, A. A., et al. (2006). Subtypes of illicit drug users: a latent class analysis of data from an Australian twin sample. *Twin Res. Hum. Genet.* 9, 523–530. doi: 10.1375/twin.9.4.523

Macmillan, N. A., and Creelman, D. C. (2004). *Detection Theory: A User's Guide*, 2nd Edn. New York, NY: Psychology Press.

MacNamara, A., Ferri, J., and Hajcak, G. (2011). Working memory load reduces the late positive potential and this effect is attenuated with increasing anxiety. *Cogn. Affect. Behav. Neurosci.* 11, 321–331. doi: 10.3758/s13415-011-0036-z

Mitchell, C. M., and Plunkett, M. (2000). The latent structure of substance use among American Indian adolescents: an example using categorical variables. *Am. J. Commun. Psychol.* 28, 105–125. doi: 10.1023/a:1005146530634

Nestor, L., Roberts, G., Garavan, H., and Hester, R. (2008). Deficits in learning and memory: parahippocampal hyperactivity and frontocortical hypoactivity in cannabis users. *Neuroimage* 40, 1328–1339. doi: 10.1016/j.neuroimage.2007.12.059

Paller, K. A., Voss, J. L., and Boehm, S. G. (2007). Validating neural correlates of familiarity. *Trends Cogn. Sci.* 11, 243–250. doi: 10.1016/j.tics.2007.04.002

Peirce, J. W. (2007). PsychoPy—psychophysics software in Python. *J. Neurosci. Methods* 162, 8–13. doi: 10.1016/j.jneumeth.2006.11.017

Prashad, S., Dedrick, E. S., and Filbey, F. M. (2018). Cannabis users exhibit increased cortical activation during resting state compared to non-users. *Neuroimage* 179, 176–186. doi: 10.1016/j.neuroimage.2018.06.031

Rangel-Pacheco, A., Lew, B. J., Schantell, M. D., Frenzel, M. R., Eastman, J. A., Wiesman, A. I., et al. (2020). Altered fronto-occipital connectivity during visual selective attention in regular cannabis users. *Psychopharmacology* 238, 1351–1361. doi: 10.1007/s00213-020-05717-3

Ratcliff, R., Sederberg, P. B., Smith, T. A., and Childers, R. (2016). A single trial analysis of EEG in recognition memory: tracking the neural correlates of memory strength. *Neuropsychologia* 93, 128–141. doi: 10.1016/j.neuropsychologia.2016.09.026

Roberts, C. A., Quednow, B. B., Montgomery, C., and Parrott, A. C. (2018). MDMA and brain activity during neurocognitive performance: an overview of neuroimaging studies with abstinent 'Ecstasy'users. *Neurosci. Biobehav. Rev.* 84, 470–482. doi: 10.1016/j.neubiorev.2017.07.015

Roberts, G. M., Nestor, L., and Garavan, H. (2009). Learning and memory deficits in ecstasy users and their neural correlates during a face-learning task. *Brain Res.* 1292, 71–81. doi: 10.1016/j.brainres.2009.07.040

Rugg, M. D., and Curran, T. (2007). Event-related potentials and recognition memory. *Trends Cogn. Sci.* 11, 251–257. doi: 10.1016/j.tics.2007.04.004

Sagar, K. A., and Gruber, S. A. (2019). Interactions between recreational cannabis use and cognitive function: lessons from functional magnetic resonance imaging. *Ann. N. Y. Acad. Sci.* 1451, 42–70. doi: 10.1111/nyas.13990

Schendan, H. E., and Maher, S. M. (2009). Object knowledge during entry-level categorization is activated and modified by implicit memory after 200 ms. *Neuroimage* 44, 1423–1438. doi: 10.1016/j.neuroimage.2008.09.061

Scholey, A. B., Parrott, A. C., Buchanan, T., Heffernan, T. M., Ling, J., and Rodgers, J. (2004). Increased intensity of ecstasy and polydrug usage in the more experienced recreational Ecstasy/MDMA users: a WWW study. *Addict. Behav.* 29, 743–752. doi: 10.1016/j.addbeh.2004.02.022

Schreiner, A. M., and Dunn, M. E. (2012). Residual effects of cannabis use on neurocognitive performance after prolonged abstinence: a meta-analysis. *Exp. Clin. Psychopharmacol.* 20, 420–429. doi: 10.1037/a0029117

Scott, J. C., Slomiak, S. T., Jones, J. D., Rosen, A. F., Moore, T. M., and Gur, R. C. (2018). Association of cannabis with cognitive functioning in adolescents and young adults: a systematic review and meta-analysis. *JAMA Psychiatry* 75, 585–595. doi: 10.1001/jamapsychiatry.2018.0335

Shah, I., Al-Dabbagh, B., Salem, A. E., Hamid, S. A. A., Muhamma, N., and Naughton, D. P. (2019). A review of bioanalytical techniques for evaluation of cannabis (Marijuana, weed, Hashish) in human hair. *BMC Chem.* 13:106. doi: 10.1186/s13065-019-0627-2

Shevlin, M., McElroy, E., Murphy, J., Hyland, P., Vallières, F., Elklit, A., et al. (2017). Cannabis and psychosis: the impact of polydrug use. *Drugs Alcohol Today* 17, 186–194. doi: 10.1108/dat-03-2017-0014

Smith, J. L., De Blasio, F. M., Iredale, J. M., Matthews, A. J., Bruno, R., Dwyer, M., et al. (2017). Verbal learning and memory in cannabis and alcohol users: an event-related potential investigation. *Front. Psychol.* 8:2129. doi: 10.3389/fpsyg.2017.02129

Snodgrass, J. G., and Corwin, J. (1988). Pragmatics of measuring recognition memory: applications to dementia and amnesia. *J. Exp. Psychol. Gen.* 117, 34–50. doi: 10.1037/0096-3445.117.1.34

Tanguay, A. N., Benton, L., Romio, L., Sievers, C., Davidson, P. S., and Renoult, L. (2018). The ERP correlates of self-knowledge: are assessments of one's past, present, and future traits closer to semantic or episodic memory? *Neuropsychologia* 110, 65–83. doi: 10.1016/j.neuropsychologia.2017.10.024

Taylor, M., Lees, R., Henderson, G., Lingford-Hughes, A., Macleod, J., Sullivan, J., et al. (2017). Comparison of cannabinoids in hair with self-reported cannabis consumption in heavy, light and non-cannabis users. *Drug Alcohol Rev.* 36, 220–226. doi: 10.1111/dar.12412

Tibon, R., and Henson, R. (2015). Commentary on: recollection reduces unitised familiarity effect. *Front. Psychol.* 6:757. doi: 10.3389/fpsyg.2015.00757

Tsivilis, D., Allan, K., Roberts, J., Williams, N., Downes, J. J., and El-Deredy, W. (2015). Old-new ERP effects and remote memories: the late parietal effect is absent as recollection fails whereas the early mid-frontal effect persists as familiarity is retained. *Front. Hum. Neurosci.* 9:532. doi: 10.3389/fnhum.2015.00532

United Nations Office on Drugs & Crime (UNODC) (2019). *World Drug Report 2019*. Avaiable online at: https://wdr.unodc.org/wdr2019/ (accessed September 20, 2020).

Van Dam, N. T., Earleywine, M., and DiCiacomo, G. (2008). Polydrug use, cannabis, and psychosis-like symptoms. *Hum. Psychopharmacol.* 23, 475–485. doi: 10.1002/hup.950

Voss, J. L., and Federmeier, K. D. (2011). FN400 potentials are functionally identical to N400 potentials and reflect semantic processing during recognition testing. *Psychophysiology* 48, 532–546. doi: 10.1111/j.1469-8986.2010.01085.x

Wais, P. E., Wixted, J. T., Hopkins, R. O., and Squire, L. R. (2006). The hippocampus supports both the recollection and the familiarity components of recognition memory. *Neuron* 49, 459–466. doi: 10.1016/j.neuron.2005.12.020

Waninger, S., Berka, C., Meghdadi, A., Karic, M. S., Stevens, K., Aguero, C., et al. (2018). Event-related potentials during sustained attention and memory tasks: utility as biomarkers for mild cognitive impairment. *Alzheimers Dement.* 10, 452–460. doi: 10.1016/j.dadm.2018.05.007

Wechsler, D. (1981). *Wechsler Adult Intelligence Scale-Revised*. Test manual. New York, NY: Psychological Corporation.ää

Wiens, S., and Syrjänen, E. (2013). Directed attention reduces processing of emotional distracters irrespective of valence and arousal level. *Biol. Psychol.* 94, 44–54. doi: 10.1016/j.biopsycho.2013.05.001

Winton-Brown, T. T., Allen, P., Bhattacharrya, S., Borgwardt, S. J., Fusar-Poli, P., Crippa, J. A., et al. (2011). Modulation of auditory and visual processing by delta-9- tetrahydrocannabinol and cannabidiol: an fMRI study. *Neuropsychopharmacology* 36, 1340–1348. doi: 10.1038/npp.2011.17

Wolk, D. A., Sen, N. M., Chong, H., Riis, J. L., McGinnis, S. M., Holcomb, P. J., et al. (2009). ERP correlates of item recognition memory: effects of age and performance. *Brain Res.* 1250, 218–231. doi: 10.1016/j.brainres.2008.11.014

World Health Organisation (WHO) (2016). *The Health and Social Effects of Nonmedical Cannabis Use*. Available online at: http://www.who.int/substance_abuse/publications/cannabis_report/en/index5.html (accessed September 22, 2020).

Yang, H., Laforge, G., Stojanoski, B., Nichols, E. S., McRae, K., and Köhler, S. (2019). Late positive complex in event-related potentials tracks memory signals when they are decision relevant. *Sci. Rep.* 9:9469.

Yonelinas, A. P. (2002). The nature of recollection and familiarity: a review of 30 years of research. *J. Mem. Language* 46, 441–517. doi: 10.1006/jmla.2002.2864

Yücel, M., Lorenzetti, V., Suo, C., Zalesky, A., Fornito, A., Takagi, M. J., et al. (2016). Hippocampal harms, protection and recovery following regular cannabis use. *Transl. Psychiatry* 6:e710. doi: 10.1038/tp.2015.201

Zeineh, M. M., Engel, S. A., Thompson, P. M., and Bookheimer, S. Y. (2003). Dynamics of the hippocampus during encoding and retrieval of face-name pairs. *Science* 299, 577–580. doi: 10.1126/science.1077775

Zhornitsky, S., Pelletier, J., Assaf, R., Giroux, S., Li, C., Shan, R., et al. (2021). Acute effects of partial CB1 receptor agonists on cognition – A meta-analysis of human studies. *Prog. Neuropsychopharmacol. Biol. Psychiatry* 104:110063. doi: 10.1016/j.pnpbp.2020.110063

Zilverstand, A., Huang, A. S., Alia-Klein, N., and Goldstein, R. Z. (2018). Neuroimaging impaired response inhibition and salience attribution in human drug addiction: a systematic review. *Neuron* 98, 886–903. doi: 10.1016/j.neuron.2018.03.048

Conflict of Interest: The 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.

frontiers in Neuroscience

ORIGINAL RESEARCH
published: 28 June 2021
doi: 10.3389/fnins.2021.667373

A High Accuracy Electrographic Seizure Classifier Trained Using Semi-Supervised Labeling Applied to a Large Spectrogram Dataset

Wade Barry[1†], Sharanya Arcot Desai[1†], Thomas K. Tcheng[1] and Martha J. Morrell[1,2]*

[1] NeuroPace, Inc., Mountain View, CA, United States, [2] Department of Neurology, Stanford University, Stanford, CA, United States

The objective of this study was to explore using ECoG spectrogram images for training reliable cross-patient electrographic seizure classifiers, and to characterize the classifiers' test accuracy as a function of amount of training data. ECoG channels in ∼138,000 time-series ECoG records from 113 patients were converted to RGB spectrogram images. Using an unsupervised spectrogram image clustering technique, manual labeling of 138,000 ECoG records (each with up to 4 ECoG channels) was completed in 320 h, which is an estimated 5 times faster than manual labeling without ECoG clustering. For training supervised classifier models, five random folds of data were created; with each fold containing 72, 18, and 23 patients' data for model training, validation and testing respectively. Five convolutional neural network (CNN) architectures, including two with residual connections, were trained. Cross-patient classification accuracies and F_1 scores improved with model complexity, with the shallowest 6-layer model (with ∼1.5 million trainable parameters) producing a class-balanced seizure/non-seizure classification accuracy of 87.9% on ECoG channels and the deepest ResNet50-based model (with ∼23.5 million trainable parameters) producing a classification accuracy of 95.7%. The trained ResNet50-based model additionally had 93.5% agreement in scores with an independent expert labeller. Visual inspection of gradient-based saliency maps confirmed that the models' classifications were based on relevant portions of the spectrogram images. Further, by repeating training experiments with data from varying number of patients, it was found that ECoG spectrogram images from just 10 patients were sufficient to train ResNet50-based models with 88% cross-patient accuracy, while at least 30 patients' data was required to produce cross-patient classification accuracies of >90%.

Keywords: semi-supervised labeling, ECoG labeling, big data, electrographic seizure classifier, epilepsy

OPEN ACCESS

Edited by:
Grzegorz Marcin Wójcik, Marie Curie-Skłodowska University, Poland

Reviewed by:
Xin Liu, University of California, San Diego, United States
Danny Eytan, Technion Israel Institute of Technology, Israel

****Correspondence:***
Sharanya Arcot Desai
sdesai@neuropace.com

[†] These authors have contributed equally to this work and share first authorship

Specialty section:
This article was submitted to Neural Technology, a section of the journal Frontiers in Neuroscience

Received: *12 February 2021*
Accepted: *17 May 2021*
Published: *28 June 2021*

Citation:
Barry W, Arcot Desai S, Tcheng TK and Morrell MJ (2021) A High Accuracy Electrographic Seizure Classifier Trained Using Semi-Supervised Labeling Applied to a Large Spectrogram Dataset. Front. Neurosci. 15:667373. doi: 10.3389/fnins.2021.667373

INTRODUCTION

One of the major challenges in epilepsy treatment is the ability to reliably assess patient outcomes (Engel, 2011; Engel et al., 2013). Patient reports of seizures can be unreliable and incomplete because patients may be amnestic for seizures, may not document their seizures, or because seizures often occur during sleep (Bazil et al., 2004; Kerling et al., 2006; Hoppe et al., 2007).

A potential solution to this problem is to automatically detect and count electrographic seizures from electrocorticographic/physiological data captured using devices such as implanted neuromodulation devices (Ryapolova-Webb et al., 2014; Skarpaas et al., 2019), or health-monitoring wearables that record long-term ambulatory patient data (Bruno et al., 2018; Regalia et al., 2019).

Machine and deep learning models previously developed for detecting electrographic seizures from physiological data have demonstrated excellent performance (Thodoroff et al., 2016; Acharya et al., 2018; O'Shea et al., 2018; Ansari et al., 2019; Roy et al., 2019). However, most of these models are either patient-specific or trained on small subsets of patients, which limits their applicability to new patients or patients in whom only small datasets are available. Additionally, previous models were mostly trained on electroencephalographic (EEG) data captured during intracranial EEG diagnostic monitoring, which may vary substantially from data captured in a long-term ambulatory setting and hence may not translate effectively to data captured outside the clinical setting (Ung et al., 2017; Baumgartner and Koren, 2018; Sun et al., 2018).

Large multi-patient ambulatory electrocorticographic (ECoG) datasets obtained during clinical trials of the NeuroPace® RNS® System may facilitate the development of machine learning (ML) algorithms for seizure detection (Morrell, 2011; Bergey et al., 2015; Skarpaas et al., 2019). However, training of a supervised machine/deep learning electrographic seizure classifier (ESC) requires ECoG datasets that contain training labels to mark both electrographic seizure and non-seizure portions in the datasets. In computer vision classification tasks involving classifying everyday objects such as cats and dogs, the task of labeling large datasets is often crowdsourced to non-specialist workers around the world (example: Amazon Mechanical Turk[1], FigureEight[2], Samasource[3]). A similar crowdsourcing technique may not be suitable for labeling ECoG datasets given their complex nature. Individuals specifically trained on labeling ECoG records are needed, and even then, interlabeler agreement is not guaranteed (Halford et al., 2011; Halford et al., 2015). The process of creating training labels for large ECoG datasets can thus be laborious and in many cases may even deter the development of powerful ML models.

Electrographic seizure patterns are often stereotypical within individual patients in time windows spanning a few months to many years (Manford et al., 1996). Hence, if a representative electrographic seizure pattern in a patient is manually labeled, a method to automatically pre-label other similar electrographic seizure patterns in the patient would substantially expedite the process of manual review and labeling of ECoG data. Further, if this method is generic and can be readily applied to individual patient's ECoGs without the need for patient-level customizations, the solution could easily scale to large multi-patient datasets. 'One-shot' and 'few-shot' learning techniques for classifying objects with just one or a few training examples of each class are currently being explored in computer vision with promising results (Vinyals et al., 2016; Snell et al., 2017). This paper describes a semi-supervised labeling technique, based on unsupervised features extracted by a pre-trained deep convolution neural network, for expediting the process of labeling the large multi-patient ECoG dataset captured with the RNS System.

Electrographic seizure patterns can vary substantially between patients due to differences in seizure onset zones, disease etiologies, and location and orientation of recording electrodes (Haas et al., 2007). Given the heterogeneous nature of electrographic seizures, development of ESCs that generalize in new patients can be a complex problem, one on which the rules- or feature-based algorithms may fare poorly. Deep learning models can learn rules and features directly from data and are particularly well-suited for problems that are too complicated to craft rules (LeCun et al., 2015). Many previous papers have demostrated the superiority of deep learning based EEG classification models over traditional machine learning models (Arora et al., 2018; Zeng et al., 2018). In this paper several different deep learning models of varying architectures and depths have been trained to classify ECoG records as electrographic seizures or non-seizures.

Performance of deep learning models generally improves with depth. However, the majority of the previous work on training deep learning models for seizure detection have used convolutional neural networks (CNNs) that are <10 layers deep (Roy et al., 2019),which is relatively shallow compared to the more recently developed contest-winning models (Russakovsky et al., 2015; He et al., 2016). This is presumably due to the limited amount of EEG training data available in the previous studies. In the current study, training and validation experiments were performed on the large multi-patient ambulatory ECoG dataset (137,985 ECoG records from 113 patients) captured with the RNS System. Because of this abundance of data, the deeper 18-layer ResNet18 and 50-layer ResNet50 (He et al., 2016) architectures were experimented with, in addition to shallower CNNs. ResNet models contain residual blocks which have been shown to alleviate the problem of exploding and vanishing gradients which can arise in deep neural network architectures and were chosen for this reason (He et al., 2016).

Despite the tremendous progress made by CNNs in classifiying 2 dimensional (2D) image data (LeCun et al., 2015), most of the previous work in training CNNs for ECoG/EEG classification has focused on using raw time-series signals as input to one dimensional CNNs (Ullah et al., 2018; Roy et al., 2019; Yildirim et al., 2020). Very few studies have explored converting ECoG time-series signals to spectrogram images for training 2 dimensional CNNs, and even the ones which do have demostrated their methods on relatively small EEG datasets (Kuanar et al., 2018; Vrbancic and Podgorelec, 2018). Validating 2D CNN training on large multi-patient ECoG spectrogram image datasets will certainly add confidence in this technique, and will encourage similar EEG/ECoG classification studies to leverage the latest developments in

[1]https://www.mturk.com/

[2]https://www.figure-eight.com/

[3]https://www.samasource.com/

2D CNN image processing, potentially accelerating the rate of neuroscience discoveries.

Although it is widely accepted that large ECoG training datasets will lead to better model generalizability to new patients (LeCun et al., 2015), to the best of our knowledge, characterizing electrographic seizure classification accuracy in held-out patients as a function of the number of patients whose data was used for training, has not been performed. In this paper, CNN models were trained with ECoG records from 10 to 80 patients, in increments of 10, to determine the generalizability of models trained with ECoG records from varying numbers of patients. All trained models were tested on expert-labeled ECoG records from 80 additional held-out patients (i.e., in addition to the 113 patients mentioned above) not used for training or validation.

The work in this paper significantly adds to the existing body of literature on labeling and training electrographic seizure classifiers in several ways. First, it introduces a semi-supervised labeling method for rapid manual-labeling of large ECoG datasets. Second, it validates the use of ECoG spectrogram images as inputs for training convolutional neural networks by producing trained classification models with very high (>95%) cross-patient classification accuracies. Third, it establishes a new benchmark cross-patient electrographic seizure classification accuracy level for ambulatory ECoG records. Fourth, it characterizes classification accuracy as a function of the amount of training data, thereby guiding the neuroscience community on data collection requirements for solving similar ECoG/EEG classification problems.

METHODS

The present study's dataset comes from clinical trials of 256 patients treated with the NeuroPace® RNS® System (Bergey et al., 2015). One hundred and ninety three randomly selected patients were used for the analyses in this study.

All study protocols were approved by the US FDA and the institutional review boards of the participating investigation sites. All participants gave written informed consent. The RNS System Feasibility, Pivotal and LTT studies are registered on clinicaltrials.gov (NCT00079781, NCT00264810, and NCT00572195).

The RNS System

The NeuroPace® RNS® System is an FDA approved adjunctive treatment for patients with medically intractable partial onset epilepsy having 1-2 seizure foci. Details about the RNS System and the types of data it captures can be found in several previous publications (Morrell, 2011; Bergey et al., 2015; Desai et al., 2019b; Skarpaas et al., 2019). Briefly, the RNS System (**Figure 1A**) consists of a closed-loop responsive neurostimulator device that is placed in the skull. One or two quadripolar depth or strip leads are connected to the device and implanted at the seizure foci. The RNS System continuously senses brain activity and sends electrical stimulation when patient-specific abnormal patterns, as defined by the physician, are detected.

ECoG Acquisition and Patient Selection

ECoG records captured with the RNS System have a sampling rate of 250 Hz per channel and are typically 90 s in duration, however length may vary to a maximum of 180 s. An ECoG record typically contains four channels of ECoG activity. A variety of recording triggers save ECoG records including, time(s) of day (scheduled ECoG records), detection of long abnormal patient-specific patterns (long episode ECoG records), and ECoG activity that saturates the recording amplifiers (saturation ECoG records). About half of all ECoG records captured with the RNS System are long episode (LE) ECoG records. LE ECoG records can contain varying degrees of abnormal epileptiform events on one or more ECoG channels and are the only type included in this study. **Figures 1B,C** shows four example ECoG records captured in one RNS System patient.

In the remainder of this paper, the term 'ECoG record' refers to an ECoG data file with up to 4 channels of ECoG data, and the term 'ECoG channel' refers to each channel of ECoG activity within an ECoG record. The terms 'electrographic seizures,' 'seizures,' and 'sz' are used interchangeably to refer to electrographic seizures; and 'electrographic non-seizures,' 'non-seizures,' and 'nsz' are used to refer to electrographic non-seizures.

Of the 256 patients enrolled in the RNS System clinical trials, 193 were randomly selected for inclusion in this study. Data from all 256 patients could not be processed due to limited human labeler time resources. Data from 113 patients were used to train, test and validate the ESC, and data from the remaining 80 patients were used only for testing by comparison of the ESC's classification scores with those of a board certified epileptologist. **Figure 2** outlines the data split from the 193 patients, and **Figure 3** shows example spectral image of ECoG channels labeled as non-seizures (top) and seizures by a human labeler.

Patient-Specific 2D Embedding and Clustering of ECoG Records

All analyses were performed using Python 3.5. Patient-specific 2D embeddings of all LE ECoG records from 193 patients were created using unsupervised feature extraction and dimensionality reduction techniques. Data preparation involved removing any stimulation artifacts present in the ECoG records (Desai et al., 2019a). In brief, blanked portions of ECoGs (**Figures 1B,C**) were identified and marked along with 10 samples preceding and 30 samples following the detected artifact. ECoG data flanking either side of the marked ECoG data were concatenated to delete stimulation and any amplifier recovery artifact. Spectrograms of the time-series ECoG data were computed using Tensorflow's built-in function tensorflow.contrib.signal.stft. Since high frequency (>90 Hz) seizure and interictal activity is often observed in ECoG data captured with the RNS System, frequencies from 0 Hz to the Nyquist frequency (i.e., 125 Hz; sampling rate = 250 Hz) were included in the spectrograms. The resulting grayscale spectrograms were resized to (299 × 299) using Tensorflow's built-in function tensorflow.image.resize_nearest_neighbor, and

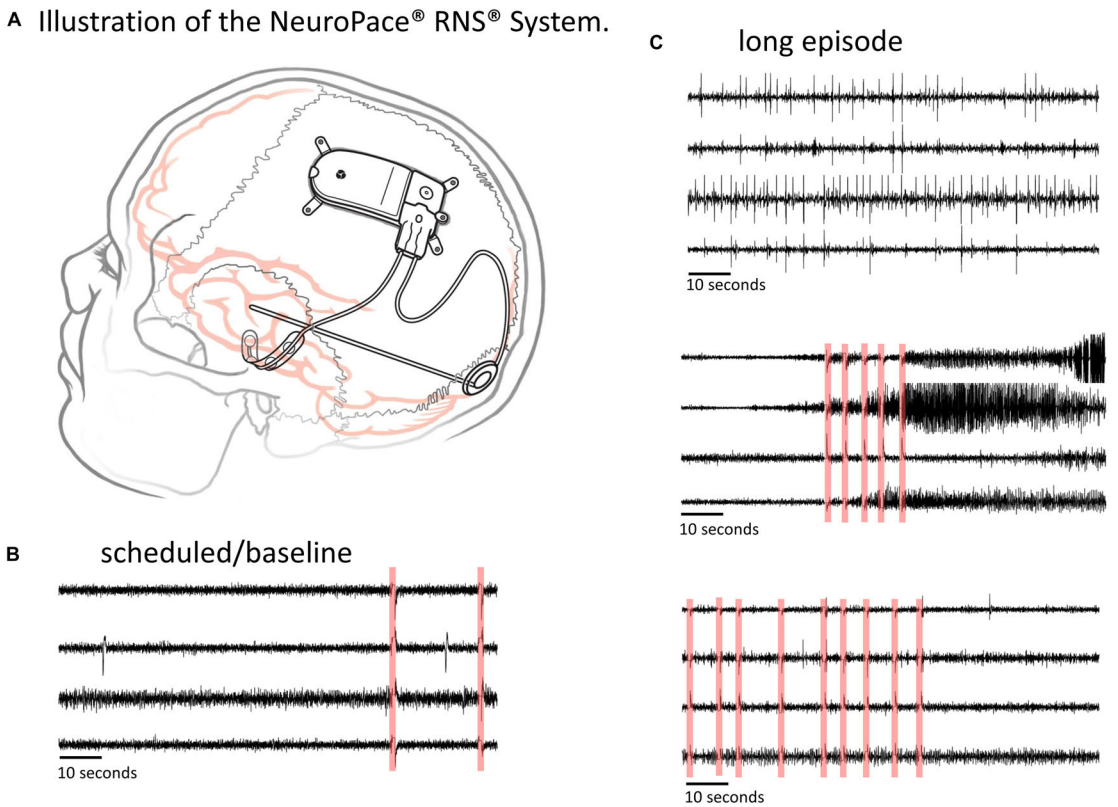

FIGURE 1 | The RNS System and types of ECoG data captured. **(A)** Illustration of the NeuroPace RNS System. The neurostimulator is placed in the skull and connected to up to two leads, each of which contains four electrode contacts. The leads can be cortical strip leads, depth leads or a combination. **(B,C)** Examples of 4-channel electrocorticographic (ECoG) records for a single patient captured with the RNS System. **(B)** Scheduled/baseline ECoG record. **(C)** 'Long Episode' (LE) ECoG records (i.e., ECoG records with detection of long abnormal patient-specific patterns). LE ECoG records can contain a variety of abnormal ECoG activity on one to several ECoG channels. Stimulation artifacts are highlighted in red in panels **(B,C)**. Stimulation artifact consists of short blanks in data followed by a brief amplifier recovery artifact.

expanded to 3 identical color channels using Tensorflow's built-in function tensorflow.image.grayscale_to_rgb (**Figure 4** step [1]). Expansion of spectrograms to 3 channels was performed because the pre-trained CNN (GoogLeNet Inception-V3) used for feature extraction requires the input data to have shape (299 × 299 × 3). The resulting 3 color channel spectrograms were passed through the pre-trained GoogLeNet Inception-V3 model for feature extraction (**Figure 4** step [2]). Tensorflow code for converting the time-series ECoG data to spectrograms and extracting features using the pre-trained GoogLeNet Inception-V3 model is provided in the **Supplementary Material**. A similar technique for embedding time-series ECoG data in 2 dimenional surfaces for differentiating ECoGs by patient outcomes was previously published by NeuroPace and are described in Desai et al. (2019a).

The extracted features (dimensions 8 × 8 × 2048) were flattened resulting in a vector of 131,072 floating point numbers for each channel of ECoG data (**Figure 4** step [3]). Features vectors from the 4 channels in ECoG records were concatenated to produce 'ECoG-record-vectors' that had 524,288 features. If less than 4 channels were present in ECoG records (which happened in a small fraction of ECoG records), zero-filled vectors were used in the missing channel's place. Principal Component Analysis (PCA) was applied to randomly selected 1,024 ECoG-record-vectors to derive a mapping function between the 524,288 features and the top 50 principal components. The mapping function was then applied to all ECoG-record-vectors in a patient to produce a reduced feature vector with 50 principal components for each ECoG record. The resulting 50 components were then passed to the t-SNE (t-distributed stochastic gradient descent) algorithm to represent all ECoGs in a patient-specific further reduced 2-dimensional embedding space.

Bayesian Gaussian Mixture (BGM; python function: sklearn.mixture.BayesianGaussianMixture) was used for automatically clustering ECoG records represented in the patient-specific 2D embedding spaces. The BGM clustering technique was chosen over other popular clustering methods such as k-means, spectral, and dbscan (density based spatial clustering) because of its ability to infer the number of clusters from the data and because it produced more sensible cluster identifications compared to the other clustering methods that were evaluated. The sklearn.mixture.BayesianGaussianMixture function takes as input the maximum number of components/clusters

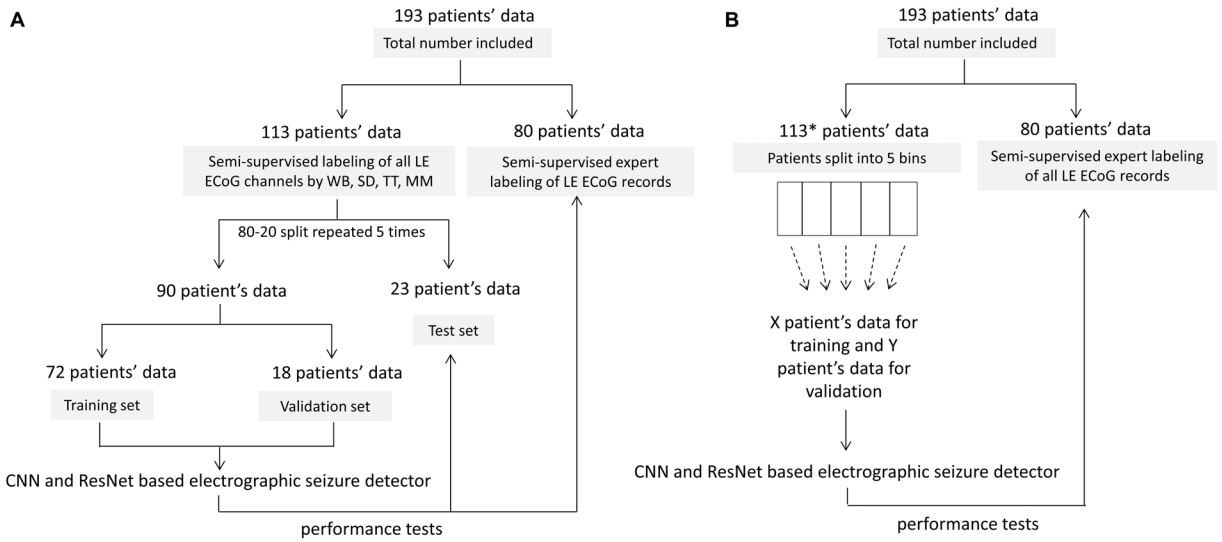

FIGURE 2 | Patient splits for labeling, training, validating and testing electrographic seizure classifier models. **(A)** Patient splits for training and testing electrocorticographic (ECoG) channel-level and ECoG record-level classification models with 80% patients' data used for training and validation, and 20% patients' data used for testing. The trained models were tested on individual ECoG channels within ECoG records from the 23 held-out patients in each training fold, to generate ECoG channel-level classification test metrics. The trained models were additionally tested on an independent expert-labeled dataset from 80 patients with seizure or non-seizure ECoG record-level labels. In this case, model predictions on individual ECoG channels within ECoG records were combined with the OR operator applied to the seizure classification to produce ECoG record-level predictions. These were compared with expert labels to generate ECoG record-level classification test metrics. **(B)** 10-80 patient splits for characterizing model performance as a function of amount of training data. *8 out of 113 patients did not have ECoG channels with the 'seizure' label and were not included in this analysis. This was done to avoid creating training datasets (especially the ones with data from small number of patients) with highly skewed number of examples of seizure and non-seizure classifications. The remaining 105 patients were split into 5 bins (with 21 patients in each bin) based on the number of labeled ECoG records available in the patient. Patients were uniformly and randomly selected from each of the 5 bins to create the 10-80 patient training sets. The 20 patient training set, for example, contained four patients randomly selected from each of the 5 bins. The trained models were tested on ECoG records from 80 patients independently labeled by an expert. Note that unlike in panel **(A)**, the only test dataset used in this case was the expert labeled dataset from 80 patients.

(n_components). Depending on the data, the BGM model can choose not to use all the components. Therefore, the number of effective components/clusters can be smaller than the number specified in n_components. For clustering ECoG records in each patient using BGM, the n_components attribute was set as maximum of [number of ECoG records in patient/100] and 10. For example, for a patient with 5,000 ECoGs, n_components would be 50, and for a patient with 200 ECoGs, n_components would be 10.

Similar to observations in Macosko et al. (2015) where authors clustered gene expression data, preliminary studies by NeuroPace also showed superior clustering in the 2 dimensional t-SNE output dataset compared to clustering in the original high dimensional dataset. This is presumably because of a frequently observed phenomenon called 'curse of dimensionality' in which clustering algorithms lose effectiveness in high dimensional spaces (Verleysen and François, 2005).

Labeling of ECoG Records

All LE ECoG records in 193 patients were manually labeled and verified using a 2-step method. In 113 out of 193 patients, ECoG records were given one of six labels: 'ictal,' 'interictal,' 'baseline,' 'noise,' 'low voltage fast only,' or 'unsure,' and channels in the ECoG record with the designated activity were selected by author WB. WB was specifically trained in labeling activity in ECoG records captured with the NeuroPace RNS System over a 2 month period by authors SD, TT and MM. Both time-series waveforms and spectrogram views of data were used to guide ECoG labeling. In case of labeling ambiguity, multiple reviewers (authors SA, TT, and MM) provided inputs to ensure accurate labeling. The 'ictal' label was selected if there was clear evolution of baseline activity into electrographic seizure that lasted at least 10 s. The 'unsure' label was only selected when reviewers could not reach a consensus on the activity type and ECoG records with the 'unsure' label were not used for training, validation or testing. ECoG channels with 'interictal,' 'baseline,' 'noise,' and 'low voltage fast only' labels formed the 'electrographic non-seizure' classification while ECoG channels with the 'ictal' label were used as the 'electrographic seizure' classification. When cluster centroid ECoG records were manually labeled, the remaining ECoG records within the cluster were automatically pre-assigned the same label.

The next step involved manual verification of the pre-assigned labels. Thumbnails of ECoG records within each pre-labeled cluster were displayed for label verification in a sorted order based on their Euclidean distance to the centroid. If a pre-assigned channel label did not match the activity observed on that ECoG channel, the label was manually corrected by author WB after consulting with additional reviewers as necessary. Thumbnails of 15 ECoG records were displayed on each page

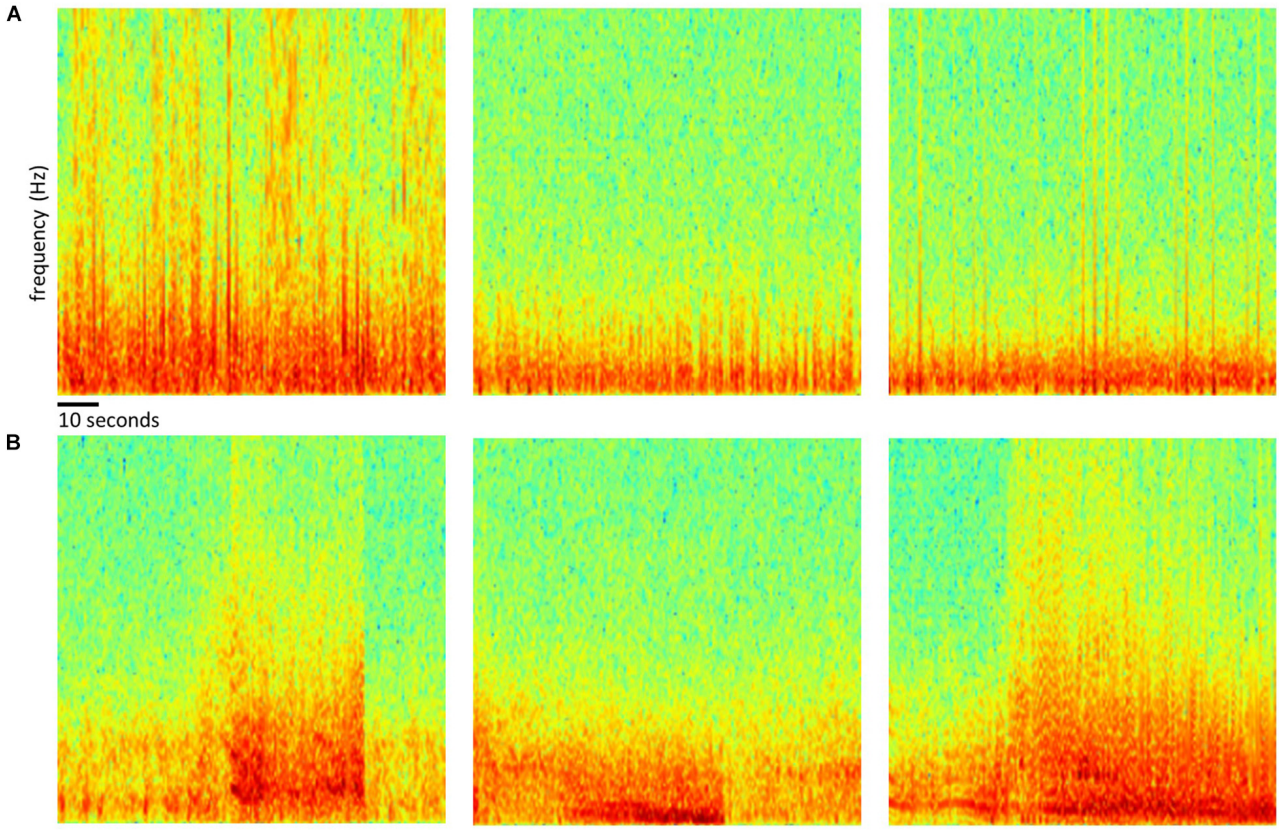

FIGURE 3 | Examples of spectrogram images of ECoG channels. Top row **(A)** shows examples of spectrogram images computed from electrocorticographic (ECoG) channels manually labeled as 'electrographic non-seizure.' Bottom row **(B)** shows examples of spectrogram images computed from ECoG channels labeled as 'electrographic seizure.'

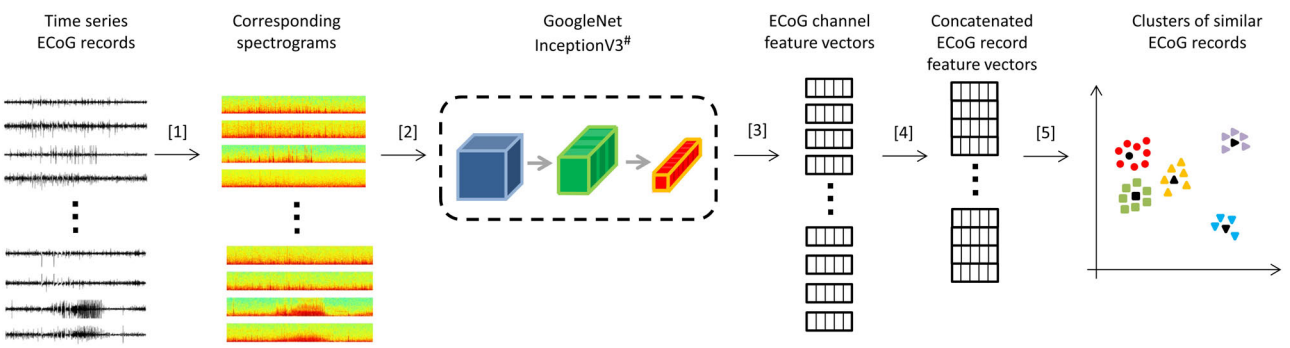

FIGURE 4 | Steps for creating 2D embeddings of patient-specific ECoG records. [1] Each channel in the time series electrocorticographic (ECoG) record was converted to a spectrogram image after removal of stimulation artifacts. [2] Spectrogram image (size 224 × 224 × 3) of each ECoG channel was passed to a pre-trained GoogLeNet Inception-V3 model for feature extraction. [3] The convolutional neural networks (CNN) extracted a feature matrix of dimensions 8 × 8 × 2048 for each spectrogram image. The feature matrix was flattened to create a feature vector of length 131,072. [4] Four feature vectors corresponding to the four channels in an ECoG record were concatenated to create a feature vector of length 524,288 for each ECoG record. Vectors of zeros were substituted for missing channels. [5] Dimensionality reduction with PCA and t-SNE was performed on the concatenated feature vectors to represent ECoG records in 2-dimensional patient-specific embedding spaces. Unsupervised clustering of the resulting 2D data was performed. ECoG records closest to cluster centers (shown in black) were identified and presented to the human labeler.# Cartoon representation of the GoogleNet Inception-V3 CNN model. Please refer to https://cloud.google.com/tpu/docs/inception-v3-advanced for details about the GoogleNet Inception-V3 model.

for label verification. Therefore if an ECoG cluster contained 90 ECoG records, 6 pages of ECoG records were displayed with page 1 containing ECoG records that were closest to the centroid and likely requiring no corrections to pre-assigned labels. Whereas page 6 contained ECoG records that were farthest from the centroids and hence was more likely to contain ECoG records that required pre-assigned label corrections compared to previous pages. When a closer look at activity in a ECoG was desired, the thumbnail could be expanded to display high resolution versions of the ECoG records in time series and spectrogram view.

To compare the performance of the trained models with those of an independent ECoG rater, an additional labeled test ECoG dataset was created. In the remaining 80 patients, a board certified epileptologist independently labeled centroid ECoG records as either 'electrographic seizure' or 'electrographic non-seizure.' A total of 262 and 333 ECoG records were labeled as either seizure and non-seizure, respectively, by the epileptologist. Additionally, the expert only provided labels at the ECoG record level, without identifying the channels in the ECoG record to which the label applied. That is, the seizure label was assigned to an ECoG record if any of the channels in the ECoG record contained an electrographic seizure, and the non-seizure label was assigned otherwise. To compute labeler agreement percentage, author WB additionally labeled all 595 expert-labeled ECoG records.

ECoG Preprocessing for Training CNN

ECoG records that were less than 80 s and greater than 100 s were repeated or cropped to 90 s in length respectively. ECoG records shorter than 80 s had a portion equal to the disparity duplicated from the beginning of the record and concatenated onto the start. Experiments were repeated with zero padding applied to short ECoG records. In this case, instead of concatenating duplicate portions of the ECoG, a vector of zeros was concatenated onto the start to create the 90 s ECoG. ECoG records greater than 100 s had 60 s before and 30 s after the storage trigger (detection of LE) selected, the remaining portions of the ECoG record were discarded.

Stimulation artifact rejection was performed on all ECoG records as described in section "Patient-Specific 2D Embedding and Clustering of ECoG Records." Model training and testing experiments were repeated without stimulation artifact rejection to assess the impact of this step on the model's classification performance.

To facilitate additional ESC model training and testing, spectrogram images of ECoGs were saved in folders organized by class labels and patient IDs. This made it convenient to perform error analyses for the different model architecture and hyperparameter selections. Further, this type of data organization made it stratightforward to apply several of Keras's built-in functions (such as ImageDataGenerator.flow_from_directory function) for reading large datasets in batches for model training and testing. The code used for converting ECoG data to spectral images and saving them in.PNG format is provided in the **Supplementary Material** section. Briefly, Matplotlib's built-in function matplotlib.pyplot.specgram with window size 256 and step size 128 was used for creating the spectrograms (spanning 0-125 Hz on the frequency axis), and saved as RBG images using the 'jet' colorsmap. Pixel values in the RGB images were scaled between −1 and + 1 which is a standard preprocessing step for training CNN models. ECoG classification experiments were repeated with spectrograms images saved using the 'grayscale' colormap in which case spectrogram images were saved with the 3 color channels having the same value.

Model Training, Validation, and Testing

Experiments to Test the ECoG-Channel Level and ECoG-Record Level Classification Performance of Trained ESC Models

The 113 patients with ECoG channel level labeling were randomly divided into three groups: 72 patients for training, 18 patients for validation and 23 patients for testing. This was repeated five times for creating 5 folds of data for training, validation and testing (**Figure 2A**). In the training dataset, the majority classification (non-seizure class in all training folds) was randomly downsampled to match the number of training examples in the seizure and non-seizure classes. CNN models were trained to classifiy each ECoG channel (note that each ECoG record can contain up to 4 ECoG channels, see section "ECoG Acquisition and Patient Selection" for details) as electrographic seizure or non-seizure. The trained models in each of the 5 folds were tested on ECoG channels in the 23 patients held-out in that fold.

Model performance was also tested on ECoG records from 80 patients independently labeled by an epileptologist. Model predictions for ECoG records were derived by applying the OR operator to the seizure classification. Consequently, if any of the ECoG channels in the ECoG record were predicted as a seizure by the trained model, the ECoG record was labeled a seizure. A non-seizure label was applied only if all ECoG channels in the ECoG record were predicted as non-seizures by the trained model. **Table 1A** shows the number of patients, ECoG records and ECoG channels in each of 5 folds.

Gradient-Based Saliency Maps

To gain some understanding of features learned by the trained CNNs, and to ensure that classification is based on relevant portions of the spectrograms, saliency maps (Simonyan et al., 2013) of the trained classification model were created using the built-in keras API, vizualize_saliency. Saliency maps are computed as the gradient of the output with respect to the input, and highlight the input regions in the datasets that contribute most toward the output classification.

Trained Model's Generalizability to Other Epilepsy Datasets of Time-Series Brain Recordings

To test the trained ESC models' generalizability to EEG datasets captured with devices other than the RNS System, the models were evaluated on the TUH EEG Seizure Corpus (v.5.1.0)[4] (Shah et al., 2018), the largest publicly available EEG dataset. Therein, 2,915 annoted seizures had an onset at >45 s into the EDF data files and were used for testing. 90 s of EEG data

[4]https://www.isip.piconepress.com/projects/tuh_eeg/

TABLE 1A | Number of patients, seizure and non-seizure ECoG channels in the training, validation and test datasets in each fold (and average across the 5 folds) for testing ECoG channel-level and ECoG record-level classification performances.

Fold	Training			Validation			ECoG channel-level testing			ECoG record-level testing		
	# of pts	# of SZ ECoG channels	#of NSZ ECoG channels*	# of pts	# of SZ ECoG channels	# of NSZ ECoG channels	# of pts	# of SZ ECoG channels	# of NSZ ECoG channels	# of pts	# of SZ ECoG records	# of NSZ ECoG records
1	72	81,573	188,858	18	21,692	40,239	23	36,918	42,477	80	262	333
2	72	78,218	174,949	18	33,398	35,751	23	28,567	60,874	80	262	333
3	72	103,572	172,136	18	8,381	36,832	23	28,230	62,606	80	262	333
4	72	91,214	150,498	18	24,957	56,364	23	24,012	64,712	80	262	333
5	72	100,983	205,036	18	15,917	25,066	23	23,283	41,472	80	262	333
Avg	**72**	**91,112**	**178,295**	**18**	**20,869**	**38,850**	**23**	**28,202**	**54,428**	**80**	262	333

*ECoG = electrocorticographic; NSZ = non-seizure; pts = patients; SZ = seizure. *Shown here is the total number of non-seizure ECoG channels available in the training dataset. Note that the non-seizure ECoG channels were randomly downsampled to match the number of seizure ECoG channels to create balanced 50/50 class splits for training. Bold values are average of 5 folds.*

TABLE 1B | Average number of ECoG channels (seizures and non-seizures) in training and validation datasets (and number of ECoG records in the test dataset) used in experiments for characterizing seizure classification accuracy as a function of number of patients' data used for training.

Training			Validation			ECoG record-level testing		
# of pts	Average # of SZ ECoG channels across 5 folds	Average # of NSZ ECoG channels across 5 folds*	# of pts	Average # of SZ ECoG channels across 5 folds	Average # of NSZ ECoG channels across 5 folds	# of pts	# of SZ ECoG records in the 5 folds	# of NSZ ECoG records in the 5 folds
10	14,469	23,263	5	5,229	7,717	80	262	333
20	26,064	57,695	5	7,037	15,701	80	262	333
30	39,356	87,258	6	8,441	19,504	80	262	333
40	51,730	113,262	8	10,873	13,212	80	262	333
50	67,034	133,792	10	13,733	25,686	80	262	333
60	80,767	159,479	12	15,352	25,598	80	262	333
70	94,769	181,877	14	16,938	39,042	80	262	333
80	108,277	214,530	16	20,656	37,438	80	262	333

The averages are computed across the 5 folds for each patient-level split. Numbers of training and validation ECoG channels in each fold for each patient-level split is shown in Supplementary Table 2. ECoG = electrocorticographic; NSZ = non-seizure; pts = patients; SZ = seizure. Shown here is the total number of non-seizure ECoG channels available in the training dataset. Note that the non-seizure ECoG channels were randomly downsampled to match the number of seizure ECoG channels to create balanced 50/50 class splits for training.

spanning 45 s respectively on either side of annotated seizure start times were converted to RGB spectrogram images using the methods described in section "ECoG Preprocessing for Training CNN." The only differences in the data processing steps were (1) skipping the stimulation artifact rejection step since it is irrelevant to EEG data and (2) applying a 60 Hz denoising notch filter to the raw timeseries data before the spectrograms were created since 60 Hz noise is commonly present in EEG data. Each EEG file resulted in 21 channels of data with channel referencing performed as described in the.lbl file associated with each.edf file. An equal number (2,915) of randomly selected 90-s EEG samples with background activity were selected for testing.

Experiments to Characterize ESC Classification Performance as a Function of the Amount of Training Data

Multiple training sets were created by selecting ECoG records from 10 to 80 patients in increments of 10 patients. Each training dataset had an equal distribution of patients with few to many ECoG channels to approximate the availability of data in the real-world. Eight of the 113 patients had no ECoG channels with seizure labels and were not included in these experiments. This was done to avoid creating training datasets (especially the ones with data from small number of patients) with highly skewed number of examples of seizure and non-seizure classifications. The remaining 105 patients with ECoG channel labeling were divided into 5 bins based on the number of labeled ECoG channels available for the patient. Each bin contained 21 patients, where bin 1 had patients with the fewest labeled ECoG channels and bin 5 had the patients with the most labeled ECoG channels. Patients for each training set were equally (and randomly) selected from all 5 bins. For example, the training dataset with 10 patients had 2 patients from each of the 5 bins, and the training dataset with 80 patients had 16 patients from each bin. To create incrementally larger training datasets, the same patients from the prior smaller datasets were retained and new patients were added.

The validation dataset for each training dataset was created by randomly sampling from the remaining (non-training) patients in each of the 5 bins equally. Each validation set contained the greater of five patients or 20% of the number of patients in the training dataset. For example, there were 5 patients in the validation datasets for the 10 and 20 patient training datasets, and 6 and 8 patients in the validation datasets for the 30 and 40 patient training datasets, respectively. **Table 1B** and **Supplementary Table 2** shows the number of patients and ECoG channels in each of the 5 folds for each level of patient split.

All trained models were tested on the same set of 262 seizure and 333 non-seizure ECoG records from the 80 patients independently labeled by an epileptologist.

Model Architectures, Training Hyperparameters, and Training Hardware

The five deep learning models evaluated and are summarized graphically in **Figure 5**. The original ResNet50 and ResNet18 (available for download from keras.applications) were modified by replacing the final 196 neuron dense layer with a 2 neuron dense layer. This was done to adapt the 18 and 50 layer ResNet models for the 2 class (seizure and non-seizure) classification task described in this paper. Training was performed for a maximum of 70 epochs with a learning rate of 10^{-6} for the 6, 7, and 12 layer CNNs or 10^{-7} for ResNets (18 and 50 layers) with a learning rate decay factor of 0. The choice of learning rate and learning rate decay factor was made after experimenting with a range of values in preliminary experiments. Learning rates higher than those listed above resulted in drastic fluctuations in training and validation performance indicating undesirable divergent behavior in the loss function, and those below the above listed values resulted in very slow, suboptimal training. Initial experiments with fine-tuning the training parameters of only the final few layers of the ResNet models with the initial layers retaining pretrained weights and biases from the imagenet dataset demonstrated substantially worse classification performance on the test dataset, compared to training the parameters in all layers. Hence, the choice was made to fine-tune the weights and biases of all layers. A training, validation and test batch size of 32 was used with all models, and models were trained with the Adam and Nadam optimizers. Training was stopped earlier than 70 epochs, if $<0.1\%$ improvement in validation accuracy was observed over 10 consecutive training epochs. The trained model at the epoch number which produced the highest validation accuracy was selected for testing. Keras v2.2.2 with Tensorflow v1.10.1 backend on an on-premise Ubuntu 16.04 machine with two NVIDIA GeForce GTX 1080 Ti GPUs was used for running the experiments described previously in Methods section "Experiments to Test the ECoG-Channel Level and ECoG-Record Level Classification Performance of Trained ESC Models." Keras v2.3.1 with Tensorflow v2.1.0 on Compute Engine Virtual Machines (machine type: n1-standard-4) with NVIDIA Tesla K80 GPUs on the Google Cloud Platform was used for running the experiments described in 2.6.4. Training was enabled on all layers of all model architectures used in this study.

Trained Model and Code Availabilty

The trained deep learning models, and ECoG pre-processing code described in this publication may be made available to researchers for academic use. Requests sent to the research-requests@neuropace.com will be reviewed in accordance with NeuroPace data sharing policy and guidelines for requesting support from NeuroPace for research. Key lines of python code for pre-processing ECoGs are made available in the **Supplementary Material**.

RESULTS

Semi-Supervised Labeling of ECoG Records

Different patterns, sizes and/or numbers of LE ECoG record clusters were obtained for each patient by the 2D embedding and clustering process (**Figure 6A**). The median number of clusters obtained with the BGM method was 9 (range: 2-45) with a median of 161 (range: 1-1624) ECoG records in each cluster. While in most cases, sensible cluster identification was obtained with the patient-specific 2D embeddings, in some cases

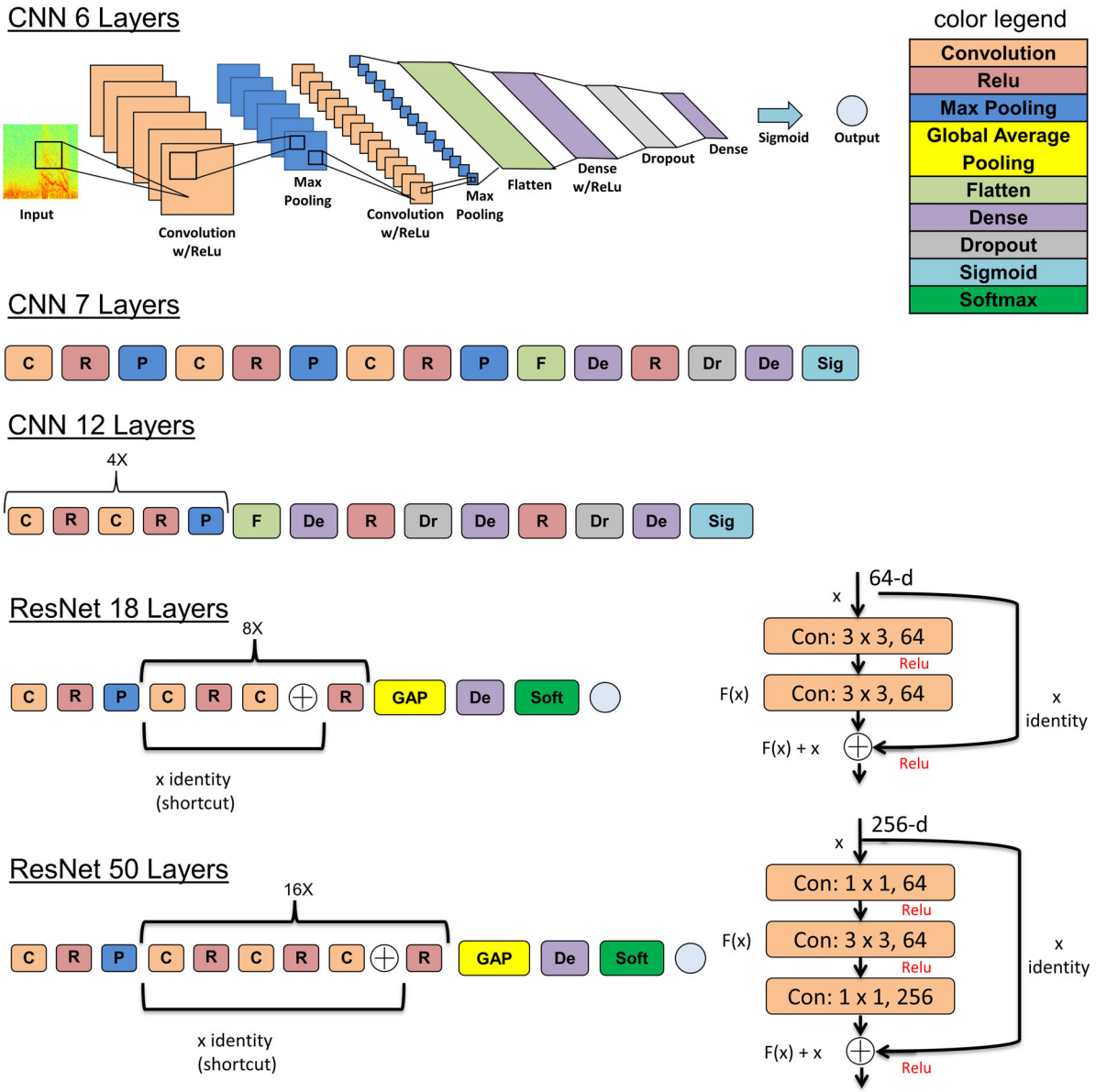

FIGURE 5 | The 5 model architectures used in this study. 6, 7, and 12 layer convolutional neural networks (CNN) models were trained and tested in this study along with 18 and 50 layer ResNet models. Inspiration for the CNN model architectures was derived from a wide range of sources including https://tinyurl.com/y76stqkc, while the ResNet model architectures are available on the Keras webpage: https://keras.io/api/applications/.

more dispersed LE ECoG record clusters were observed (example orange and purple clusters in **Figure 6A** bottom-right panel).

Based on *ad hoc* analyses and visual inspection, two primary factors appeared to be responsible for LE ECoG record clustering. One factor was the type of electrographic activity in the ECoG record and the other factor was the ECoG record length. Electrographic activity indicative of seizures tended to produce discrete clusters. Within a patient, multiple seizure clusters were often apparent and distinguishable by seizure morphology and the presence of seizures on different channels. Typically, ECoG records <70 s in length tended to form separate clusters. When electrographic seizures were captured in these short ECoG records, they often clustered separately but in the vicinity of other clusters containing short ECoG records. Since only a minority of LE ECoG records captured with the RNS System were short in length, clusters with short ECoG records were an uncommon occurrence and did not substantially lengthen the manual labeling effort. Similarly, in some patients, ECoG records longer than 90 s (typically 180 s) were captured and in those cases, ECoG records also tended to form separate clusters similar to the short ECoG records. With the clustering tool described above, channel-level manual labeling of 137,985 ECoG

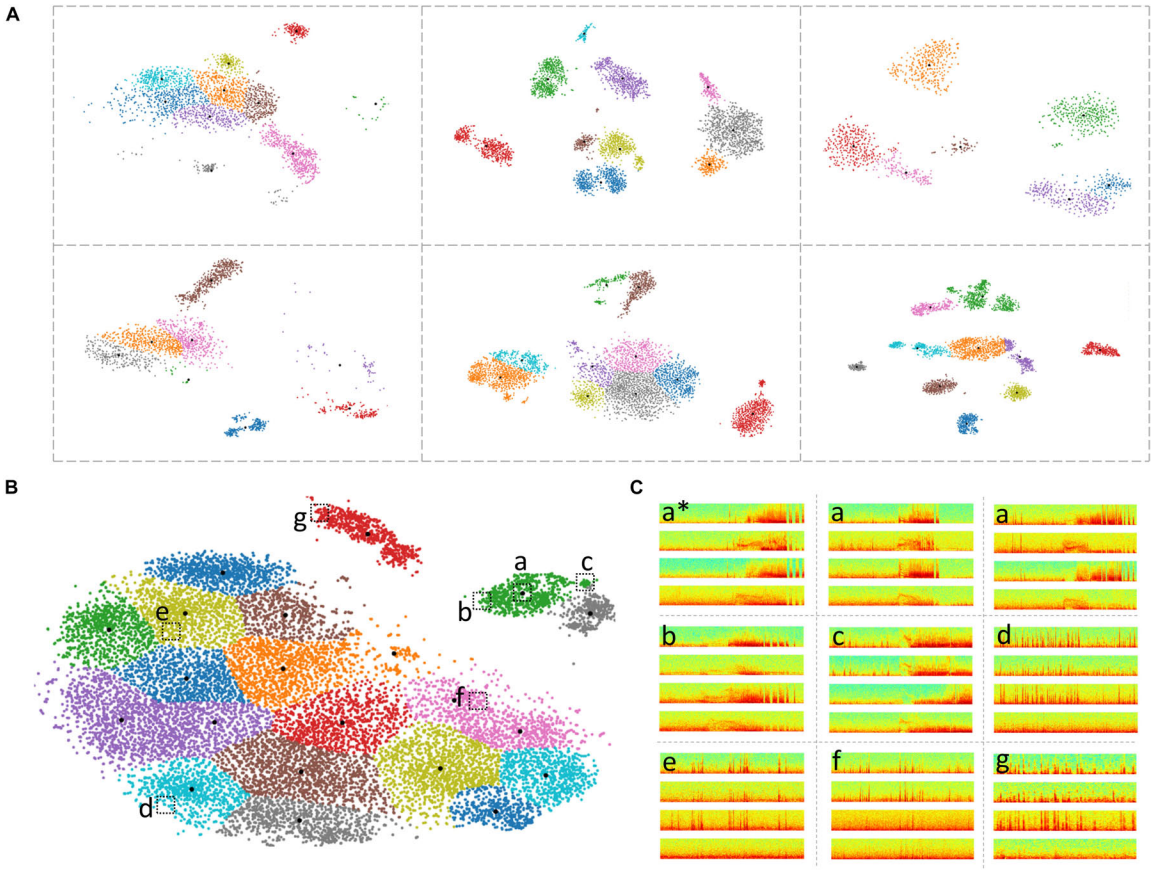

FIGURE 6 | Clustering of ECoG records in patient-specific 2D embedding spaces. **(A)** 2-dimensional embeddings of long episode (LE) electrocorticographic (ECoG) records from six example patients included in this study. Different numbers and patterns of clusters were observed in the patient-specific 2D embedding spaces. **(B)** 2D embeddings of 17,422 LE ECoG records from the patient with the most ECoG records included in this study. Only the green and gray clusters in the top right of the embedding space contained electrographic seizures. **(C)** Example ECoG records from different locations in the 2D embedding space shown in panel **(B)**. ECoG record denoted with 'a*' is the centroid ECoG record within the green cluster in panel **(B)**. ECoG records denoted with 'a' are two ECoG records that were closest to the centroid ECoG (a*). ECoGs records b and c are two example ECoG records further away from the centroid in the same green cluster. Three example ECoG records (d-f) from other areas of the embedding space contained interictal activity. ECoG record (g) is from a cluster (red cluster in panel **B**) that exclusively contained short (<70 s) interictal ECoG records.

records from 113 patients took approximately 320 h. The time required to label all of a patient's ECoG records depended on the number of records and variability within clusters. The median labeling time was 1.5 h, while for some patients it took as little as <15 min or as long as 3 days. These time estimates include the time spent manually examining all cluster centroid ECoG records in time and spectral domains, assigning each ECoG channel to one of six labels ('ictal,' interictal,' 'baseline,' 'noise,' 'low voltage fast only,' and 'unsure') and manually reviewing the pre-assigned (based on the centroid ECoG labels) channel labels for every member of each cluster. As described in the Methods section, if the pre-assigned channel labels did not apply to a given ECoG channel, it was manually corrected after consulting with additional reviewers as needed. The ML-assisted labeling and verification process was faster in patients where most clusters contained similar ECoG records compared to other patients with more ECoG record variability within clusters. In patients with highly similar ECoG records within clusters, usually associated with stereotypical seizure and interictal activity, the entire labeling and verification process took 15 mins or less (for ∼1,200 ECoG records on average). However, it was more common for 10-20% of the ECoG channels within a cluster to require label correction. ML-assisted labeling of 17,422 long episode ECoG records, the most from any patient (**Figure 6B**) took only 2 h. In this patient, only two out of the 21 clusters contained electrographic seizures. **Figure 6C** shows nine example ECoG records embedded in the 2D space in panel **(B)**. ECoG records (a) closer to a cluster centroid ECoG (a*) looked more similar to the centroid ECoG compared to ECoG records (b,c) further away from the cluster centroid. This trend was generally seen in all clusters in all patients. Further, ECoG records from other clusters (d-f) looked very different from ECoG records in the green cluster (a-c).

Agreement between an independent expert labeller and author WB on 595 ECoG records from 80 held-out patients (see Methods section "Labeling of ECoG Records") was at 98.3%. Eight and two

ECoG records respectively labeled as seizures and non-seizures by the expert did not agree with the labels provided by author WB.

ECoG Channel-Level and ECoG Record-Level Classification Performance of Trained ESC Models

Model Performance on ECoG Channels From 20% Held-Out Patients

Test performances improved with increased depth of models used for training, with the deepest trained model (ResNet50) producing the highest precision and recall values among the five model architectures trained. **Figure 7** shows the precision recall curves and confusion matrices for the five CNN models for a randomly chosen data fold. F_1 scores and test accuracies were generally higher with the Nadam optimizer compared to the Adam optimizer and are shown in **Table 2**. The corresponding values with the Adam optimizer are shown in **Supplementary Table 1**. While the 'sz' (seizure) class classification accuracy of individual ECoG channels drastically increased with depth of training models, the increase in accuracy in the 'nsz' (non-seizure) class was moderate (**Figure 7B** and **Table 2**). Increases in overall accuracies of 3.8% and 3.26% were observed when the depth of the training model was increased from 6 to 7 layers, and 7 to 12 layers respectively. However, accuracy increases of only 0.51% and 0.28% were observed when model depth was further increased from 12 layers to 18 layers, and 18 layers to 50 layers respectively. Similar trends were observed with F_1 scores, indicating that the point of diminishing returns relative to model depth is around 12 layers. Among the five different model architectures studied, the ResNet50 model had the highest F_1 score (94.26%) and class-balanced accuracy (95.72%), with 97.17% non-seizure class accuracy and 94.26% seizure class accuracy. In addition to producing higher test performances, deeper ($\geq$12 layers) models also resulted in less fold to fold variation in F_1 scores and accuracies compared to the shallower models (2.9%, 1.5%, 0.6%, 0.4% and 0.4% standard deviation in overall accuracy values for the 6 layer, 7 layer, 12 layer CNN, ResNet18 and ResNet50 respectively), indicating that the shallower models not only underfit the training data, but were also less robust compared to the deeper models.

ECoG length affected classification accuracy when comparing short (<80 s), regular (90 s) and long (>100 s) records. With the ResNet50 model, the average classification accuracies on short (92.7%) and long (94.1%) ECoG channels were significantly lower ($p < 0.05$, Wilcoxon rank sum test) compared to classification accuracy on 90-s (96.4%) ECoG channels. Similar trends in classification performance with short, long and regular-length ECoG channels were observed with all model architectures trained. The average classification performance of the ResNet50 models on short ECoG channels was slighty better when preprocessing was performed by duplicating and concatenating the starting portion of the ECoG (92.7%), when compared to zero-padding (91.8%), although this difference was not statistically significant. The average classification performance was slightly better (statistically non-significant) when the stimulation artifact rejection step was applied (95.7%), compared to when it was not applied (95.5%). Additionally, the classification accuracy was significantly better ($p < 0.05$, Wilcoxon rank sum test) when model training was done with the spectrogram images saved in the 'jet' colormap (95.7%), compared to spectrogram images saved in the 'grayscale' colormap (94.9%).

Model Performance on ECoG Records From 80 Expert-Labeled Held-Out Patients

As described in the Methods section, an ECoG record was classified as a seizure if any of the 4 channels were classified as a seizure by the trained ESC models. A non-seizure label was assigned only if all 4 channels were classified as non-seizures. A random binary classifier (with 50% chance of classifying an ECoG channel as a seizure) would classify a 4-channel ECoG record as seizure 93.75% of the time, and it would classify a 4-channel ECoG record as non-seizure only 6.25% of the time. Hence, a random classifier would produce extremely skewed class-specific accuracies guessing seizure ECoG records correctly most of the time, while having a very poor non-seizure class accuracy. In comparison, all trained models produced more or less balanced seizure and non-seizure class accuracies (as shown in **Table 2B**), demonstrating successful training, with the test performance generally improving with depth. A significant ($p < 0.01$, Wilcoxon rank sum test) increase of 9.72% in overall accuracy and an 11.32% increase in F_1 score was observed between the 6 layer CNN and the 12 layer CNN models, and a non-significant difference in test performance (< 1.41%, $p > 0.05$, Wilcoxon rank sum test) was observed between the 12 layer CNN and ResNet50-based model. Similar to observations with ECoG channel level test accuracies, substantial improvements in seizure class accuracies were observed with increased CNN model depths, while the non-seizure class accuracies remained relatively constant.

Training and validation curves for each type of model are shown in **Supplementary Figure 1**. Training and validation accuracies with the deeper ResNet50 and ResNet18 models increased rapidly over the first few epochs with the condition for early stopping (< 0.1% improvement in validation accuracy observed over 10 consecutive training epochs) applying substantially earlier than with shallower models.

Error Analysis

Type 1 (False negative) and Type 2 (False positive) error rates for the representative data fold shown in **Figure 7** are 17% and 6% for the 6 layer CNN model, 12% and 6% for the 7 layer CNN model, 7% and 4% for the 12 layer CNN model, 6% and 5% for ResNet18 model, and 6% and 3% for the ResNet50 model respectively. In all data folds, the deepest ResNet50 model produced the least percentage of type 1 and type 2 errors. Examination of errors showed that between 10-15% of the misclassified ECoG channels could be attributed to labeler error and were not model performance errors. The remaining errors were due to model performance. A few example ResNet50 classification errors are shown in **Figure 8**. Seizure spectrograms with only faint frequency bands, lower amplitude changes, or short durations (around 10 s) were sometimes

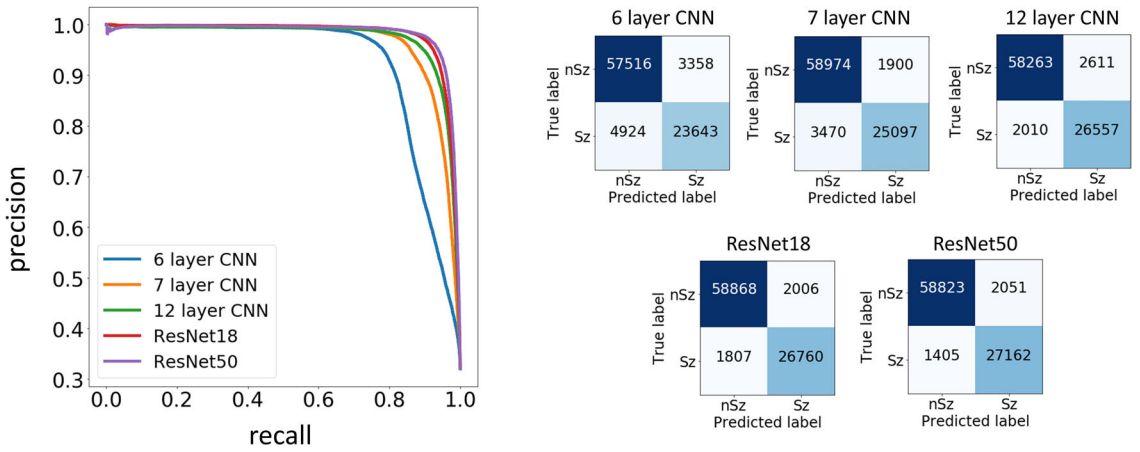

FIGURE 7 | Precision Recall curves and confusion matrices for one example data fold. The trained models assigned a seizure or non-seizure label to individual electrocorticographic (ECoG) channels, along with class prediction probabilities. The precision recall curves **(left)** show the precision or positive predictive value on the *y*-axis, and recall or model sensitivity on the *x*-axis for different thresholds of prediction probabilities, for each of the five model architectures trained in this study. The confusion matrices **(right)** show the number of ECoG channels of each classification (seizure and non-seizure) correctly and incorrectly predicted by the trained models.

misclassified (**Figure 8A**). Also, non-seizure spectrograms were sometimes misclassified when they contained cropped interictal ECoG channels, resulting in short bursts of high frequency or amplitude activity (typically found in electrographic seizures) being repeated for longer than 10 s (**Figure 8B**). Finally, if ECoGs longer than 100 s contained multiple long episode triggers, the last one would be used for creating the spectrogram image. This infrequently resulted in the main portion of the seizure being cropped out of the resulting image, resulting in the ECoG channel getting erroneously labeled as non-seizure.

Saliency Maps

Gradient-based saliency maps associated with the final fully connected layer of the trained ResNet50 model are shown in **Figure 9**. It appears that in seizure class, horizontal and diagonal spectral power bands within the seizure activity are generally highlighted, confirming that the neural network's seizure classifications are indeed based on seizure activity, and is not a consequence of irrelevant feature learning. In comparison, non-seizure class predictions had scattered background activation patterns often associated with interictal spiking activity.

Model Generalizability to EEG Datasets Not Captured With the RNS System

On the TUH EEG Seizure Corpus, the ResNet50 ESC models had an average overall classification accuracy of 70.1% with an average F_1 score of 69.5%. The average seizure class classification accuracy was 68.2%, while the non-seizure class classification accuracy was 72%. In other words, 926 out of 2,915 seizure spectrograms were missclassified as non-seizures, while 1,989 were correctly classified. Similarly, 817 out of 2,915 non-seizure spectrograms were missclassified as seizures, while 2,089 were correctly classified.

Error analysis revealed that the most common explanation for the seizure and non-seizure misclassification in the TUH EEG dataset was the presence of various types of noise in the raw EEG data, with the classification performance generally degrading with increased levels of noise in the EEG spectrograms. A few examples of correctly and incorrectly classified EEG spectrograms from the TUH EEG Seizure Corpus are shown in **Supplementary Figure 2**.

ESC Classification Performance as a Function of Amount of Training Data

Trends in F_1 scores and class-balanced accuracies vs amount of training data are shown in **Figure 10** and **Table 3**. Models were tested on expert labeled ECoG records from 80 patients. The results show that training data from a minimum of 30 patients is required to achieve > 90% generalizability in new patients. With the 6 layer CNN model, none of the 8 patient splits achieved > 90% accuracy or F_1 score. Training data from 50 patients were required to achieve F_1 scores and accuracies of over 90% with the 12 layer CNN models. On the other hand, the deeper ResNet18 and ResNet50 models performed better, requiring training data from fewer patients i.e., 40 and 30 patients respectively, to achieve ECoG record-level classification accuracies of >90% in new patients. Deeper training models and larger training datasets produced lower fold to fold variation in performance metrics (see standard deviation values in **Table 3** and **Supplementary Table 3**), compared to shallow models and smaller training datasets.

DISCUSSION AND CONCLUSION

The work in this paper is novel and significant for several reasons. First, it describes a semi-supervised technique for labeling large ECoG datasets. This can be an important step for building supervised machine learning models on large datasets, which is becoming increasingly common in the heathcare domain.

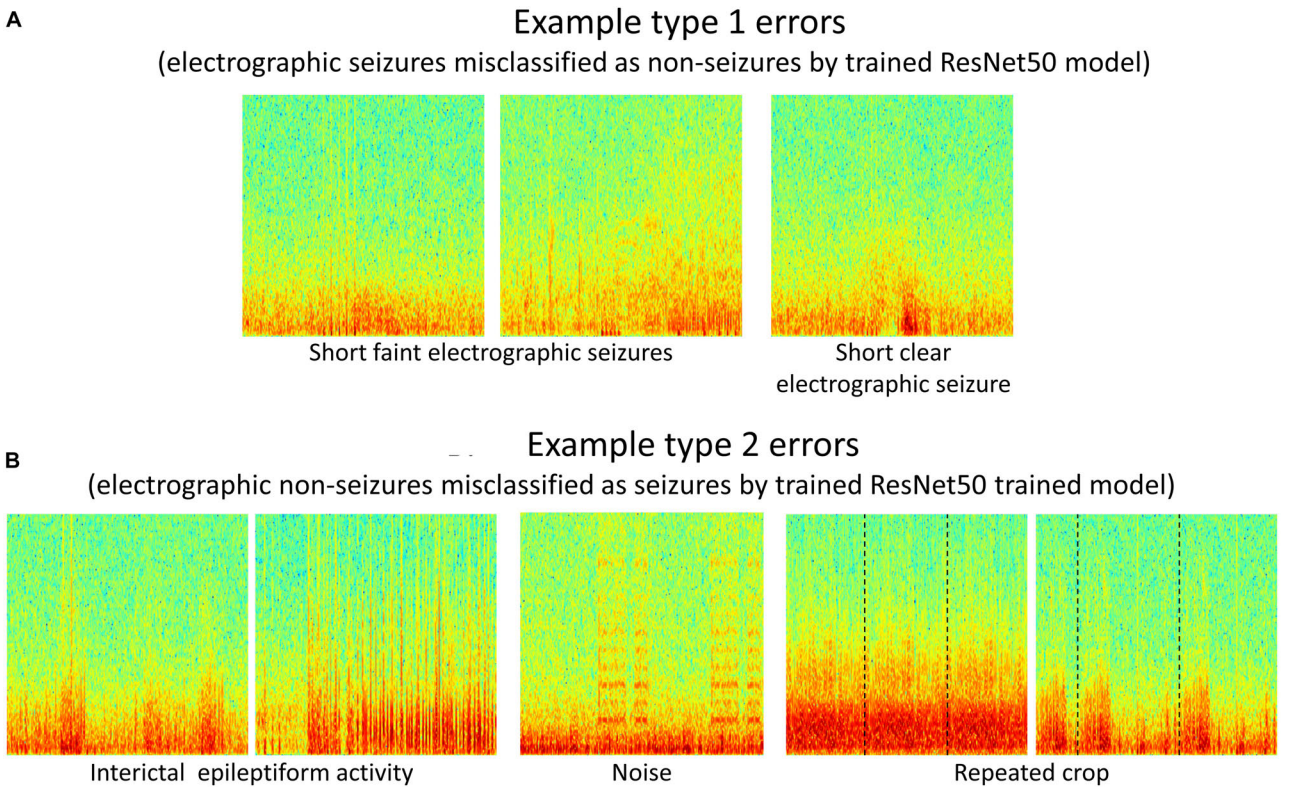

FIGURE 8 | (A,B) Example type 1 and type 2 errors by a trained ResNet50 model. Type 1 errors typically included short faint electrographic seizures, while type 2 errors typically included long trains of high amplitude interictal events, noise and repeated crops created during pre-processing electrocorticographic (ECoG) channels in cases where length was shorter than 80 s.

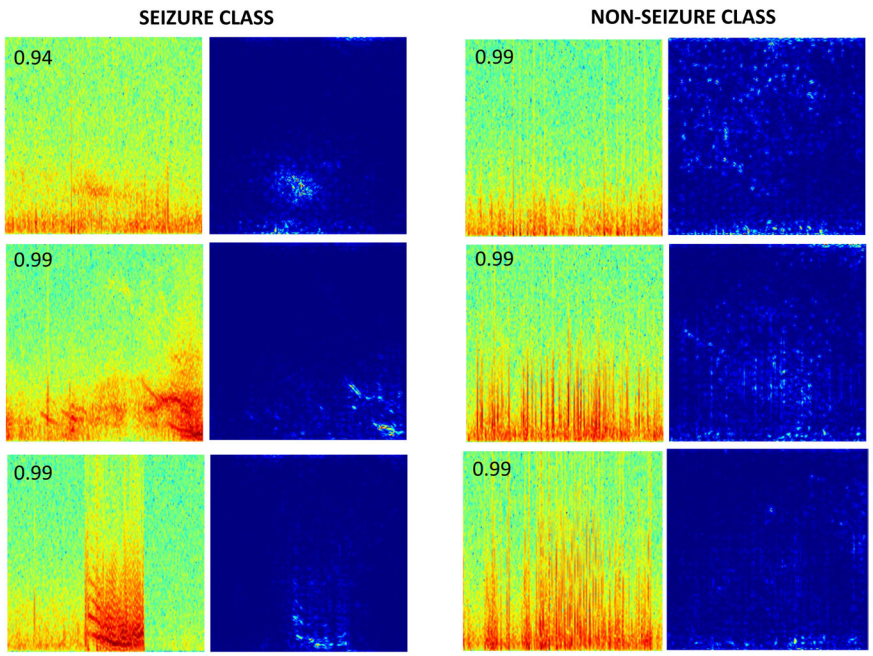

FIGURE 9 | Saliency maps of a few correctly classified seizure and non-seizure class ECoG channels by the ResNet50 ESC model. The number on the top left corner in each spectrogram image is the prediction probability of the model for the respective class.

TABLE 2 | Test performance on held-out ECoG channels and expert labeled ECoG records.

Model	Fold	Performance on ECoG channels from 20% held-out patients				Performance on ECoG records from 80 expert labeled held-out patients			
		Test accuracy %			F_1 score	Test accuracy %			F_1 score
		Overall	NSZ Class	SZ Class		Overall	NSZ Class	SZ Class	
6 layer LR: 10^{-6} Opt: Nadam	1	90.16	93.59	86.74	89.37	86.32	92.49	80.15	84.51
	2	88.62	94.48	82.76	85.10	84.21	93.99	74.43	81.76
	3	83.03	93.91	72.14	77.72	81.00	89.49	72.52	78.03
	4	89.79	92.62	86.97	84.08	83.68	97.90	69.47	80.71
	5	87.80	89.93	85.66	84.15	89.85	94.59	85.11	88.67
	Avg	**87.88**	**92.90**	**82.86**	**84.08**	**85.01**	**93.69**	**76.34**	**82.73**
7 layer LR: 10^{-6} Opt: Nadam	1	91.84	95.20	88.48	91.22	90.35	95.20	85.50	89.24
	2	92.37	96.88	87.85	90.34	90.60	87.69	93.51	89.42
	3	89.65	97.55	81.75	87.34	90.15	95.20	85.11	89.02
	4	93.61	95.24	91.98	89.83	91.97	97.30	86.64	91.16
	5	90.89	93.21	88.58	88.28	92.64	95.20	90.08	91.83
	Avg	**91.67**	**95.62**	**87.73**	**89.40**	**91.14**	**94.11**	**88.17**	**90.13**
12 Layer LR: 10^{-6} Opt: Nadam	1	95.15	96.51	93.79	94.84	95.68	96.70	94.66	95.20
	2	94.34	95.71	92.96	92.00	93.48	94.59	92.37	92.72
	3	94.28	97.12	91.46	92.44	94.19	92.19	96.18	93.33
	4	95.62	97.29	93.95	93.37	95.13	93.69	96.56	94.40
	5	95.24	98.27	92.20	94.43	95.19	96.10	94.27	94.64
	Avg	**94.93**	**96.98**	**92.87**	**93.41**	**94.73**	**94.65**	**94.81**	**94.06**
RN18 LR: 10^{-7} Opt: Nadam	1	96.05	97.03	95.07	95.79	93.82	90.69	96.95	92.87
	2	95.19	96.71	93.68	93.35	91.88	85.29	98.47	90.69
	3	95.13	97.98	92.27	93.79	93.97	90.99	96.95	93.04
	4	95.07	95.67	94.47	91.66	94.99	91.89	98.09	94.14
	5	95.76	97.22	94.30	94.65	95.52	92.19	98.85	94.70
	Avg	**95.44**	**96.92**	**93.96**	**93.85**	**94.04**	**90.21**	**97.86**	**93.09**
RN50 LR: 10^{-7} Opt: Nadam	1	96.07	97.09	95.04	95.81	94.57	92.19	96.95	93.73
	2	95.86	96.63	95.08	94.02	91.91	90.69	93.13	90.88
	3	95.14	98.04	92.24	93.84	92.89	91.89	93.89	91.96
	4	95.57	96.32	94.82	92.63	95.00	95.50	94.66	94.48
	5	95.94	97.74	94.14	95.01	93.26	89.19	97.33	92.22
	Avg	**95.72**	**97.17**	**94.26**	**94.26**	**93.54**	**91.89**	**95.19**	**92.65**

ECoG = electrocorticographic; NSZ = non-seizure; SZ = seizure. Bold values are average of 5 folds.

Second, it shows that despite the heterogeneous nature of electrographic seizures, robust electrographic seizure detection models can be built that classify ambulatory ECoG channels in new patients with over 95% classification accuracy. Third, it validates the method of converting time-series ECoG data to spectrogram images for the purpose of CNN-based ECoG classification models. Finally, it shows that with good CNN architecture selection, data from as few as 10 patients can produce cross-patient electrographic seizure classification accuracies of 88% (F_1 score 87%), while a minimum of 30 patient's labeled ECoG records may be required for achieving a classification accuracy of over 90%.

The prohibitive task of labeling a large ECoG dataset was made manageable with the aid of an ECoG record clustering tool. This tool enabled the manual labeling of ∼138,000 ECoG records in 320 h. Using only 2 labels instead of 6 (as was the case in this study) would have resulted in even faster labeling. It should be emphasized that the unsupervised ECoG record custering step was followed by a manual label validation step in which pre-assigned labels given to every ECoG record were manually verified, and corrections were made as necessary. However, without the use of such an ECoG clustering tool for pre-labeling ECoG records, a conservative estimate for labeling ∼138,000 4-channel ECoG records is around 1533 h (at ∼10 s on average for labeling each ECoG channel), not accounting for delays caused by labeler-fatigue. Thus, use of the ECoG record clustering tool to speed-up labeling of a large ECoG dataset was key to the success of this project.

A goal of this study was to manually label the large NeuroPace ECoG dataset for the purpose of building an electrographic seizure classifier. A simple, 'out-of-the-box' technique that produced reasonably good within-patient ECoG clustering was desired, and hence the pre-trained GoogLeNet Inception-V3 model which can be applied to 3-color channel image data

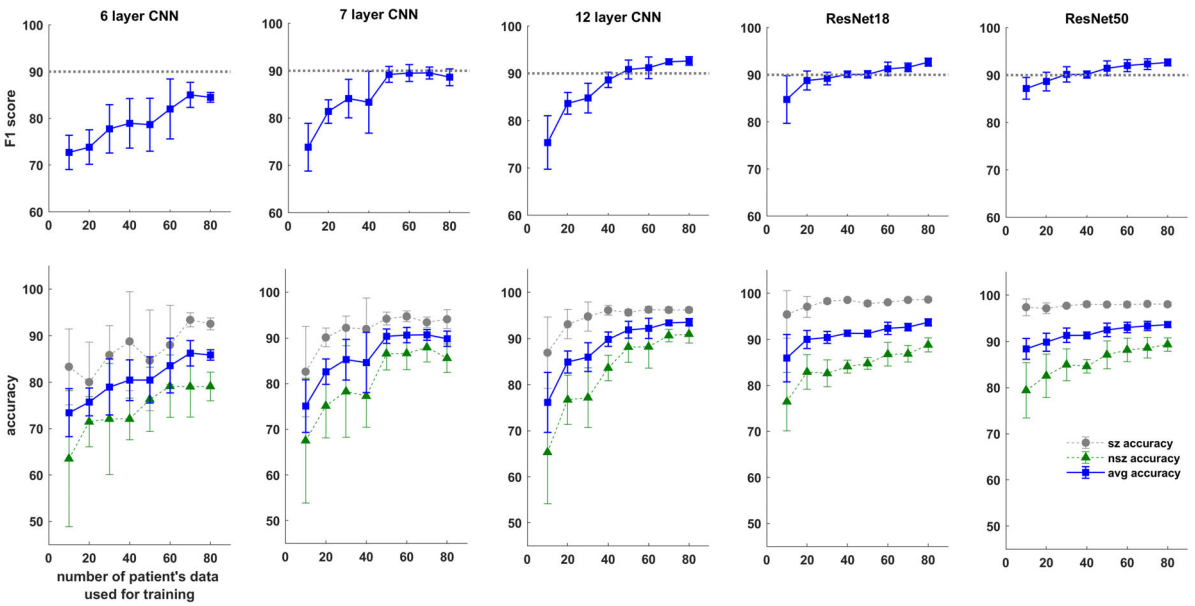

FIGURE 10 | F_1 score and test accuracy of CNN and ResNet models trained with data from increasing numbers of patients. Convolutional neural networks (CNN) and ResNet models were tested on expert labeled electrocorticographic (ECoG) records from 80 patients. A dotted line is drawn at F_1 score value of 90% for reference.

TABLE 3 | F_1 scores (mean ± standard deviation) as a function of number of patient's data used for training.

Model	Number of patients' data used for training mean (standard deviation)							
	10	20	30	40	50	60	70	80
6 layer CNN	72.7 (3.7)	73.9 (3.7)	77.8 (5.2)	78.9 (5.3)	78.7 (5.7)	82.0 (6.4)	85.0 (2.7)	84.5 (1.1)
7 layer CNN	73.8 (5.0)	81.4 (2.5)	84.1 (4.1)	83.3 (6.5)	89.2 (1.7)	89.5 (1.8)	89.5 (1.3)	88.6 (1.8)
12 layer CNN	75.4 (5.7)	83.7 (2.3)	84.8 (3.1)	88.6 (1.6)	90.9 (2.0)	91.2 (2.3)	92.5 (0.6)	92.7 (0.9)
ResNet18	84.8 (5.1)	88.8 (2.0)	89.2 (1.3)	90.1 (0.7)	90.1 (0.8)	91.3 (1.4)	91.6 (0.9)	92.7 (0.8)
ResNet50	87.3 (2.3)	88.7 (2.0)	90.2 (1.6)	90.1 (0.7)	91.4 (1.6)	92.0 (1.3)	92.3 (1.1)	92.7 (0.7)

CNN = convolutional neural networks. Cells with F_1 scores ≥ 90% are filled with green color.

was chosen as feature extractor, after experimenting with a few different pre-trained CNN models. Dimensionality reduction was then performed with PCA and t-SNE. Even though t-SNE is a technique not known for preserving intercluster distances, in our case, where ECoG labeling was done on a per-cluster basis, preserving intercluster distances was not a priority and did not negatively impact the manual labeling process. In applications where preserving inter-cluster distances is of importance, other more recently developed techniques such as UMAP could be used instead (McInnes et al., 2018).In future studies, models specifically trained on ECoG records, such as autoencoders trained on unlabeled ECoG records (Tsinalis et al., 2016), or CNN models trained on auxiliary tasks such ECoG classification (i.e., similar to the ESC models trained in this study) may be used as feature extractors to improve ECoG record clustering performance. Alternatively, deep ranking models trained using the triplet loss function with the goal of learning optimal embedding functions, may be used (Wang et al., 2014). Nevertheless, as shown in this present study, the use of pre-trained CNN models for feature extraction followed by dimensionally reduction proved to be adequately useful, and as far as we know this is the first study to demonstrate the usefulness of transferring pre-trained CNN weights to spectrogram images of brain recordings for clustering brain data. Such a transfer learning technique is likely to translate effectively to other types of physiological data (Salem et al., 2018).

Long episode ECoG records make up over 50% of all ECoG records captured by the NeuroPace RNS System and contain a mix of baseline, interictal and ictal activity. Out of the 414,933 manually labeled channels from 137,985 long episode ECoG records, 140,183 were given the seizure label, and 274,750 were given one of five non-seizure labels. Since a sufficient number of ECoG channels with the non-seizure label was available from only labeling the long episode ECoG records, in the interest of minimizing manual labeling time, it was deemed unnecessary to label other types of ECoG records captured by the RNS System (such as scheduled records and records triggered by a patient applied magnet) for the purpose of training ESC

models. The other types of ECoG records contain a much lower proportion of seizures and many more non-seizure records. Since balanced numbers of seizure and non-seizure ECoGs were used for training, it is unlikely that including these ECoGs in the training sets would have improved performance.

Five types of model architectures were trained and tested. Three of them involved convolutional layers, max pooling, and dense layers, connected linearly, and two of them involved residual connections, as shown in **Figure 5**. Increasing the depth of CNNs from 6 to 7 layers, and from 6 to 12 layers resulted in improvements in ECoG channel-level accuracies of 3.8% and 7.1%, respectively, with an accompanying decrease in fold-to-fold performance variability. Since issues such as vanishing and exploding gradients can manifest from pushing a neural network too deep, the ResNet architectures, which have residual connections to mitigate the above mentioned problems, were used when expanding beyond 12 layers. Modest improvements in accuracy were observed between the 12 layer CNN and the 18 layer ResNet-based (0.51% improvement), and between the 12 layer CNN and 50 layer ResNet-based models (0.79% improvement) suggesting that the point of diminishing returns with respect to model depth/complexity may be around 12 layers. When selecting a model for incorporation of an ESC in embedded systems, the 12 layer CNN may be a better choice compared to the ResNet-based models because of its substantially smaller computational demands while having only slightly worse classification performance. It should be noted, however, that since the models described here were optimized for classifying 90-s ECoG records, pipeline modifications and model revisions may be necessary to identify seizures in a continuous data stream. Additionally, the trained models may need to be converted to compressed, lighter formats (the Tensorflow Lite format, for example) for real-time model inference in embedded systems with limited compute capability.

The classification performance of the ESC trained models were only slightly worse on short (ResNet50 model accuracy: 92.7%, 12 layer CNN model accuracy: 91.1%) and long (94.1%, 93.3%) ECoG channels compared to the classification performance on 90-s ECoG channels (96.4%, 95.4%). This small discrepancy in performance was expected given that ∼67% of the data used in this study were the regular-length (i.e., 90-s) ECoG channels. Even though the trained ESC models performed best on regular-length EcoG channels, the classification performances on short and long EcoG channels were still very impressive making the trained models applicable to EcoG channels of all durations captured with the RNS System (maximum EcoG storage length is 240 s, minimum is 30 s).

A few different hyperparameters such as the initial learning rate, the learning rate decay factor, the optimizer type were tuned empirically. Preliminary tests showed that initial learning rates of 10^{-6} for the shallower (6,7, and 12 layer) CNNs, and 10^{-7} for the deeper ResNet-based architectures, with a learning rate decay of 0, produced the best training and validation accuracies among the parameter values tested, so these values were used for all experiments reported in this paper. The choice of weight initialization method (Xavier vs ImageNet weights) did not make a difference on the model's final training performance, so ImageNet weight initialization was used where available (i.e., for the ResNet models), and Xavier initialization was used in other cases. Additionally, the choice of gradient descent optimization algorithm, i.e., Adam vs Nadam, did not make a substantial difference to the models' performance, although the Nadam (which is Adam with Nesterov Momentum) optimizer produced slightly better results. Overall, model training and testing performance seemed to be robust to small variations in hyperparameter selections.

Gradient-based saliency maps were examined to confirm that the features learned by the trained ESC models are relevant to the classification task. Saliency maps associated with seizure classifications, generally had pixels associated with horizontal and diagonal seizure power bands highlighted, while saliency maps associated with non-seizure classifications had pixels associated with any interictal spiking activity occasionally highlighted. Finally, tests were run to confirm that the choice of training platform (virtual machines on the Google Cloud Platform vs on-premise machines) did not affect model training and testing performances.

Although model architectures and hyperparameters such as the learning rate, optimizers and spectrogram colormaps (which are often considered to be among the most important hyperparameters) were experimented with, the hyperparameter search space is enormous. In future experiments, neural architecture search methods such as AutoML on Google Cloud Platform will be used to further improve classification accuracies. Additionally, data from only 113 patients were used for training, validation and testing in this study. Manually labeling data from all 256 RNS study patients and using data from additional patients for training may lead to further improvements in model performance.

The trained ESC models performed substantially better than chance (70.1%; chance level is 50%) in classifying spectrogram images of EEG data in the TUH EEG Seizure Corpus (an EEG dataset not captured with the RNS System). Error analysis revealed that the most common explanation for misclassification was various types of noise present in the raw EEG data, with the classification performance generally being much better on EEG datasets with little/no noise, compared to nosier datasets. This was expected given that the RNS System ECoG data used to train the CNN models were practically noise-free. The only type of denoising applied by us to the TUH EEG data was a 60 Hz notch filter. The addition of other denoising steps, while outside the scope of this paper, could potentially lead to substantially better classification outcomes. The fact that the ESC models trained on ECoG data captured with the RNS System could classify seizures in EEG datasets captured with markedly different lead types, electrode configurations and recording electronics demonstrates the transferability of preprocessing methods and trained models across different types of time-series brain recording datasets. Fine-tuning weights and biases in final layers of neural networks trained on large ECoG/EEG datasets (such as the RNS System data used in this study) with smaller labeled time-series brain datasets from a different source could be a potential strategy for applying deep learning classification/regression models to datasets with limited training

data. Such a technique is frequently used for solving computer vision problems (Shin et al., 2016).

As expected, the performance of trained models continued to improve with the amount of training data used. In all patient-level splits the deepest model (ResNet50 model) produced the best results while shallowest model (CNN 6 layer) produced the worst. Class-balanced accuracies were >90% when the ResNet50 model was trained with labeled ECoG records from 30 patients. Surprisingly, training ResNet50-based models with labeled spectrogram images from just 10 patients produced a mean ECoG record classification accuracy of ∼88% in new patients suggesting that the deeper ResNet50 models can effectively learn electrographic seizure signatures even with limited training examples, and can successfully generalize the learnings to ECoG records from new patients. The shallow 6 layer CNN, on the other hand, only produced a classification accuracy of 73% with 10 patient's training data. These findings suggest that for EEG/ECoG classification tasks, it may be beneficial to train deep CNN architectures (such as the ResNet50 architecture) even in cases where data from only a limited number of subjects are available, and that training experiments should not be limited to shallow CNN models as has mostly been the case in the past (Roy et al., 2019).

Future experiments are needed to study the relationship between the distinct number of electrographic seizure patterns used for training and cross-patient classification accuracy. In our experiments to characterize model performance as a function of number of patient's data used for training, a substantial number (∼14,500 per class or ∼29,000 in total) of seizure and non-seizure spectrogram images were available from just 10 patients. Since some patients have multiple seizure foci and seizure waveforms can change over time, each patient likely contributed a few different seizure patterns for training. This is because seizures may originate in different brain areas with different seizure waveforms, and because ECoG records used in this study cover an average of 7.5 years per patient. This variety could explain why models built with just 10 patients' data generalized to new patients with > 88% accuracy. Repeat analyses with data captured from small numbers of patients over shorter periods will provide additional insight into the minimum number of seizure examples required for good performance.

About 34% of labeled ECoG channels used in this paper belonged to the seizure class, while the remaining 66% of ECoG channels belonged to the non-seizure class. Class balancing of seizure and non-seizure class ECoG channels was performed only in the training datasets and not in the test datasets. This was done to compute the test performance on realistic distributions of ECoG records captured with the RNS System. Hence, the precision (positive predictive value), recall (sensitivity) and F_1 scores (harmonic mean of precision and recall) reported in this paper were computed on imbalanced datasets with about twice the number of non-seizure examples as seizure examples (see **Table 1**). However, these metrics do not reflect model performance on continuous ECoG records. The RNS System captures data intermittently with data capture biased toward abnormal ECoG activity. In this paper, a training approach that produced high cross-patient test classification performance on such intermittently captured ECoG records was chosen. Although training was not performed with the goal of applying the models to continuous ECoG records, extrapolating the results (97.17% non-seizure accuracy and 94.26% seizure accuracy on 90 s ECoG records; see **Table 2**) to continuous ECoG records would result in a false positive rate of ∼1.1/h per ECoG channel (with the sensitivity still being high at 94.26%). This is assuming that the continuous ECoG records are cut into 90 s non-overlapping segments and fed into the model for inference. If a patient only has a few seizures per month, this could produce a large number of false positives relative to true positives. Nevertheless, such a generic cross-patient model could still be used in an offline setting to filter out ∼97% of data and present the remaining 3% for further human review. A better strategy, however, might be to alter the training approach to optimize model classification performance on continuous ECoG records. For example, data augmentation techniques could be used to mimic continuous data capture by enriching scheduled ECoGs in a training set of all ECoG record types, and custom loss functions with additional penalties could be used for minority class misclassifications. Such experiments are outside the scope of this paper but certainty warrant further studies.

Error analysis revealed some shortcomings in model performance. In general, seizure ECoG channels with very subtle seizure signals in spectrogram images were sometimes misclassified as non-seizures. Conversely, non-seizure spectrogram images with noise artifact or long trains of high amplitude interictal activity were occasionally misclassified as seizures. Several approaches may be taken to further improve the classification performance of the electrographic seizure detection models. Training ECoG channel classification models with greater spatial context such as including spectrogram images of ECoG activity from adjacent ECoG channels could lead to higher classification accuracy for the primary channel. This would mimic the behavior of a human labeler evaluating faint seizure activity by using activity on adjacent ECoG channels. Additionally, separate CNN or ResNet models could be trained on shorter and longer ECoGs to improve the classification performance of ECoG channels that are outside the typical length range of 90 s. Since relatively few ECoG channels fall outside the typical length range, one way to achieve this could be to fine-tune the final layer(s) parameters of a base CNN or ResNet model trained with 90-s ECoGs for the shorter and longer ECoGs respectively. Alternatively, recurrent neural networks architectures (RNN) or combined CNN and RNN architectures which do not require the input data to be of fixed dimensions may be trained.

In summary, this study demonstrates that converting time-series ECoG records into spectrogram images and using them as input to CNN models can be used to effectively train robust cross-patient seizure classification models. Healthcare tools built using these models may facilitate the physician's review of EEG data for epilepsy patients, and have the potential to improve clinical outcomes due to improved diagnostic assessments. Additionally, by characterizing the performance of various CNN models as a function of amount of training data, this research provides

ECoG/EEG data collection guidance for researchers interested in solving similar ECoG classification problems.

DATA AVAILABILITY STATEMENT

The data analyzed in this study is subject to the following licenses/restrictions: The trained deep learning models, and ECoG pre-processing code described in this publication may be made available to researchers for academic use. Requests sent to the research-requests@neuropace.com will be reviewed in accordance with NeuroPace data sharing policy and guidelines for requesting support from NeuroPace for research. Key lines of python code for pre-processing ECoGs are made available in the **Supplementary Material**.

ETHICS STATEMENT

All study protocols were approved by the FDA and the institutional review boards of the participating investigation sites. All participants gave written informed consent. The RNS System Feasibility, Pivotal, and LTT studies are registered on clinicaltrials.gov (NCT00079781, NCT00264810, and NCT00572195). The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

SA designed the study and performed the ECoG clustering experiments, ESC model training and testing experiments, and primarily wrote the manuscript. WB manually labeled ECoG records from 113 patients, performed ESC model training and testing experiments, and error analyses. TT and MM were involved in study and manuscript development. All authors contributed to the article and approved the submitted version.

ACKNOWLEDGMENTS

We thank Dr. Vikram R. Rao for labeling ECoG records from 80 held-out patients included in this study. We are grateful to all the patients and their families for their participation in the RNS System pivotal clinical study. We also thank the investigators and study coordinators who contributed to the RNS System clinical study research at the participating epilepsy centers.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fnins.2021.667373/full#supplementary-material

REFERENCES

Acharya, U. R., Oh, S. L., Hagiwara, Y., Tan, J. H., and Adeli, H. (2018). Deep convolutional neural network for the automated detection and diagnosis of seizure using EEG signals. *Comput. Biol. Med.* 100, 270–278. doi: 10.1016/j.compbiomed.2017.09.017

Ansari, A. H., Cherian, P. J., Caicedo, A., Naulaers, G., De Vos, M., Van Huffel, S., et al. (2019). Neonatal seizure detection using deep convolutional neural networks. *Int. J. Neural Syst.* 29:1850011. doi: 10.1142/S0129065718500119

Arora, A., Lin, J. J., Gasperian, A., Maldjian, J., Stein, J., Kahana, M., et al. (2018). Comparison of logistic regression, support vector machines, and deep learning classifiers for predicting memory encoding success using human intracranial EEG recordings. *J. Neural Eng.* 15:066028. doi: 10.1088/1741-2552/aae131

Baumgartner, C., and Koren, J. P. (2018). Seizure detection using scalp-EEG. *Epilepsia* 59, 14–22. doi: 10.1111/epi.14052

Bazil, C. W., Pitkänen, A., Loeb, J. A., Dudek, F. E., Bertram, E. H. III, Cole, A. J., et al. (2004). Sleep, sleep apnea, and epilepsy. *Curr. Treat. Options Neurol.* 6, 339–345. doi: 10.1007/s11940-004-0033-4

Bergey, G. K., Morrell, M. J., Mizrahi, E. M., Goldman, A., King-Stephens, D., Nair, D., et al. (2015). Long-term treatment with responsive brain stimulation in adults with refractory partial seizures. *Neurology* 84, 810–817. doi: 10.1212/WNL.0000000000001280

Bruno, E., Simblett, S., Lang, A., Biondi, A., Odoi, C., Schulze-Bonhage, A., et al. (2018). Wearable technology in epilepsy: the views of patients, caregivers, and healthcare professionals. *Epilepsy Behav.* 85, 141–149. doi: 10.1016/j.yebeh.2018.05.044

Desai, S. A., Tcheng, T., and Morrell, M. (2019a). "Transfer-learning for differentiating epileptic patients who respond to treatment based on chronic ambulatory ECoG data," in *Proceedings of the 2019 9th International IEEE/EMBS Conference on Neural Engineering (NER)*, (New York, NY: IEEE). doi: 10.1109/NER.2019.8717007

Desai, S. A., Tcheng, T. K., and Morrell, M. J. (2019b). Quantitative electrocorticographic biomarkers of clinical outcomes in mesial temporal lobe epileptic patients treated with the RNS§system. *Clin. Neurophysiol.* 130, 1364–1374. doi: 10.1016/j.clinph.2019.05.017

Engel, J. (2011). Biomarkers in epilepsy: introduction. *Biomark. Med.* 5, 537–544. doi: 10.2217/bmm.11.62

Engel, J. Jr., Pitkänen, A., Loeb, J. A., Dudek, F. E., Bertram, E. H. III, Cole, A. J., et al. (2013). Epilepsy biomarkers. *Epilepsia* 54, 61–69. doi: 10.1111/epi.12299

Haas, S. M., Frei, M. G., and Osorio, I. (2007). Strategies for adapting automated seizure detection algorithms. *Med. Eng. Phys.* 29, 895–909. doi: 10.1016/j.medengphy.2006.10.003

Halford, J. J., Pressly, W. B., Benbadis, S. R., Tatum, W. O. IV, Turner, R. P., Arain, A., et al. (2011). Web-based collection of expert opinion on routine scalp EEG: software development and interrater reliability. *J. Clin. Neurophysiol.* 28, 178–184. doi: 10.1097/WNP.0b013e31821215e3

Halford, J., Shiau, D., Desrochers, J. A., Kolls, B. J., Dean, B. C., Waters, C. G., et al. (2015). Inter-rater agreement on identification of electrographic seizures and periodic discharges in ICU EEG recordings. *Clin. Neurophysiol.* 126, 1661–1669. doi: 10.1016/j.clinph.2014.11.008

He, K., Zhang, X., Ren, S., and Sun, J. (2016). "Deep residual learning for image recognition," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, (Las Vegas, NV). doi: 10.1109/CVPR.2016.90

Hoppe, C., Poepel, A., and Elger, C. E. (2007). Epilepsy: accuracy of patient seizure counts. *Arch. Neurol.* 64, 1595–1599. doi: 10.1001/archneur.64.11.1595

Kerling, F., Mueller, S., Pauli, E., and Stefan, H. (2006). When do patients forget their seizures? An electroclinical study. *Epilepsy Behav.* 9, 281–285. doi: 10.1016/j.yebeh.2006.05.010

Kuanar, S., Athitsos, V., Pradhan, N., Mishra, A., Rao, K. R., et al. (2018). "Cognitive analysis of working memory load from EEG, by a deep recurrent neural network," in *Proceedings of the 2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, (New York, NY: IEEE). doi: 10.1109/ICASSP.2018.8462243

LeCun, Y., Bengio, Y., and Hinton, G. (2015). Deep learning. *Nature* 521, 436–444. doi: 10.1038/nature14539

Macosko, E. Z., Basu, A., Satija, R., Nemesh, J., Shekhar, K., Goldman, M., et al. (2015). Highly parallel genome-wide expression profiling of individual cells using nanoliter droplets. *Cell* 161, 1202–1214. doi: 10.1016/j.cell.2015.05.002

Manford, M., Fish, D., and Shorvon, S. (1996). An analysis of clinical seizure patterns and their localizing value in frontal and temporal lobe epilepsies. *Brain* 119, 17–40. doi: 10.1093/brain/119.1.17

McInnes, L., Healy, J., and Melville, J. (2018). Umap: uniform manifold approximation and projection for dimension reduction. *arXiv [Preprint]*. doi: 10.21105/joss.00861

Morrell, M. J. (2011). Responsive cortical stimulation for the treatment of medically intractable partial epilepsy. *Neurology* 77, 1295–1304. doi: 10.1212/WNL.0b013e3182302056

O'Shea, A., Lightbody, G., Boylan, G., and Temko, A. (2018). "Investigating the impact of CNN depth on neonatal seizure detection performance," in *Proceedings of the 2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, (New York, NY: IEEE). doi: 10.1109/EMBC.2018.8513617

Regalia, G., Onorati, F., Lai, M., Caborni, C., and Picard, R. W. (2019). Multimodal wrist-worn devices for seizure detection and advancing research: Focus on the Empatica wristbands. *Epilepsy Res.* 153, 79–83. doi: 10.1016/j.eplepsyres.2019.02.007

Roy, Y., Banville, H., Albuquerque, I., Gramfort, A., Falk, T. H., Faubert, J., et al. (2019). Deep learning-based electroencephalography analysis: a systematic review. *J. Neural Eng.* 16: 051001. doi: 10.1088/1741-2552/ab260c

Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., et al. (2015). Imagenet large scale visual recognition challenge. *Int. J. Comput. Vis.* 115, 211–252. doi: 10.1007/s11263-015-0816-y

Ryapolova-Webb, E., Afshar, P., Stanslaski, S., Denison, T., de Hemptinne, C., Bankiewicz, K., et al. (2014). Chronic cortical and electromyographic recordings from a fully implantable device: preclinical experience in a nonhuman primate. *J. Neural Eng.* 11:016009. doi: 10.1088/1741-2560/11/1/016009

Salem, M., Taheri, S., and Yuan, J. S. (2018). "ECG arrhythmia classification using transfer learning from 2-dimensional deep CNN features," in *Proceedings of the 2018 IEEE Biomedical Circuits and Systems Conference (BioCAS)*, (New York, NY: IEEE). doi: 10.1109/BIOCAS.2018.8584808

Shah, V., von Weltin, E., Lopez, S., McHugh, J. R., Veloso, L., Golmohammadi, M., et al. (2018). The temple university hospital seizure detection corpus. *Front. Neuroinform.* 12:83. doi: 10.3389/fninf.2018.00083

Shin, H.-C., Roth, H. R., Gao, M., Lu, L., Xu, Z., Nogues, I., et al. (2016). Deep convolutional neural networks for computer-aided detection: CNN architectures, dataset characteristics and transfer learning. *IEEE Trans. Med. Imaging* 35, 1285–1298. doi: 10.1109/TMI.2016.2528162

Simonyan, K., Vedaldi, A., and Zisserman, A. (2013). Deep inside convolutional networks: visualising image classification models and saliency maps. *arXiv [Preprint]*.

Skarpaas, T. L., Jarosiewicz, B., and Morrell, M. J. (2019). Brain-responsive neurostimulation for epilepsy (RNS§System). *Epilepsy Res.* 153, 68–70. doi: 10.1016/j.eplepsyres.2019.02.003

Snell, J., Swersky, K., and Zemel, R. (2017). Prototypical networks for few-shot learning. *Adv. Neural Inf. Process. Syst. arXiv*. arXiv:1703.05175

Sun, F. T., Arcot Desai, S., Tcheng, T. K., and Morrell, M. J. (2018). Changes in the electrocorticogram after implantation of intracranial electrodes in humans: the implant effect. *Clin. Neurophysiol.* 129, 676–686. doi: 10.1016/j.clinph.2017.10.036

Thodoroff, P., Pineau, J., and Lim, A. (2016). "Learning robust features using deep learning for automatic seizure detection," in *Paper Presented at the Machine Learning for Healthcare Conference*. Los Angeles, CA.

Tsinalis, O., Matthews, P. M., and Guo, Y. (2016). Automatic sleep stage scoring using time-frequency analysis and stacked sparse autoencoders. *Ann. Biomed. Eng.* 44, 1587–1597. doi: 10.1007/s10439-015-1444-y

Ullah, I., Hussain, M., and Aboalsamh, H. (2018). An automated system for epilepsy detection using EEG brain signals based on deep learning approach. *Exp. Syst. Appl.* 107, 61–71. doi: 10.1016/j.eswa.2018.04.021

Ung, H., Baldassano, S. N., Bink, H., Krieger, A. M., Williams, S., Vitale, F., et al. (2017). Intracranial EEG fluctuates over months after implanting electrodes in human brain. *J. Neural Eng.* 14:056011. doi: 10.1088/1741-2552/aa7f40

Verleysen, M., and François, D. (2005). "The curse of dimensionality in data mining and time series prediction," in *Proceedings of the International Work-Conference on Artificial Neural Networks*, (Berlin: Springer). doi: 10.1007/11494669_93

Vinyals, O., Blundell, C., Lillicrap, T., Kavukcuoglu, K., Wierstra, D., et al. (2016). Matching networks for one shot learning. *Adv. Neural Inf. Process. Syst. arxiv*. arXiv:1606.04080

Vrbancic, G., and Podgorelec, V. (2018). Automatic classification of motor impairment neural disorders from EEG signals using deep convolutional neural networks. *Elektronika ir Elektrotechnika* 24, 3–7. doi: 10.5755/j01.eie.24.4.21469

Wang, J., Song, Y., Leung, T., Rosenberg, C., Wang, J., Philbin, J., et al. (2014). "Learning fine-grained image similarity with deep ranking," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, (New York, NY: IEEE). doi: 10.1109/CVPR.2014.180

Yildirim, Ö, Baloglu, U. B., and Acharya, U. R. (2020). A deep convolutional neural network model for automated identification of abnormal EEG signals. *Neural Comput. Appl.* 32, 15857–15868. doi: 10.1007/s00521-018-3889-z

Zeng, H., Yang, C., Dai, G., Qin, F., Zhang, J., Kong, W., et al. (2018). EEG classification of driver mental states by deep learning. *Cogn. Neurodyn.* 12, 597–606. doi: 10.1007/s11571-018-9496-y

Conflict of Interest: WB, TT, and SA have equity ownership/stock options with NeuroPace and was current employee of NeuroPace. MM has equity ownership/stock options with NeuroPace and is currently employed as Chief Medical Officer of NeuroPace.

frontiers in Human Neuroscience

ORIGINAL RESEARCH
published: 26 July 2021
doi: 10.3389/fnhum.2021.652789

The Menstrual Cycle Alters Resting-State Cortical Activity: A Magnetoencephalography Study

*Rika Haraguchi[1†], Hideyuki Hoshi[2†], Sayuri Ichikawa[1†], Mayuko Hanyu[3], Kohei Nakamura[3,4], Keisuke Fukasawa[1], Jesús Poza[5,6,7], Víctor Rodríguez-González[5], Carlos Gómez[5,6] and Yoshihito Shigihara[2,8]**

[1] Clinical Laboratory, Kumagaya General Hospital, Kumagaya, Japan, [2] Precision Medicine Centre, Hokuto Hospital, Obihiro, Japan, [3] Department of Gynaecology, Kumagaya General Hospital, Kumagaya, Japan, [4] Genomics Unit, Keio Cancer Centre, Keio University School of Medicine, Minato, Japan, [5] Biomedical Engineering Group, Higher Technical School of Telecommunications Engineering, University of Valladolid, Valladolid, Spain, [6] Centro de Investigación Biomédica en Red en Bioingeniería, Biomateriales y Nanomedicina (CIBER-BBN), Valladolid, Spain, [7] Instituto de Investigación en Matemáticas (IMUVA), University of Valladolid, Valladolid, Spain, [8] Precision Medicine Centre, Kumagaya General Hospital, Kumagaya, Japan

OPEN ACCESS

Edited by:
Grzegorz Marcin Wójcik, Marie Curie-Skłodowska University, Poland

Reviewed by:
Görsev Yener, Ýzmir University of Economics, Turkey
Waldemar Bauer, AGH University of Science and Technology, Poland

***Correspondence:**
Yoshihito Shigihara
y-shigihara@hokuto7.or.jp

† These authors have contributed equally to this work and share first authorship

Specialty section:
This article was submitted to Brain Imaging and Stimulation, a section of the journal Frontiers in Human Neuroscience

Received: *20 January 2021*
Accepted: *28 June 2021*
Published: *26 July 2021*

Citation:
Haraguchi R, Hoshi H, Ichikawa S, Hanyu M, Nakamura K, Fukasawa K, Poza J, Rodríguez-González V, Gómez C and Shigihara Y (2021) The Menstrual Cycle Alters Resting-State Cortical Activity: A Magnetoencephalography Study. Front. Hum. Neurosci. 15:652789. doi: 10.3389/fnhum.2021.652789

Resting-state neural oscillations are used as biomarkers for functional diseases such as dementia, epilepsy, and stroke. However, accurate interpretation of clinical outcomes requires the identification and minimisation of potential confounding factors. While several studies have indicated that the menstrual cycle also alters brain activity, most of these studies were based on visual inspection rather than objective quantitative measures. In the present study, we aimed to clarify the effect of the menstrual cycle on spontaneous neural oscillations based on quantitative magnetoencephalography (MEG) parameters. Resting-state MEG activity was recorded from 25 healthy women with normal menstrual cycles. For each woman, resting-state brain activity was acquired twice using MEG: once during their menstrual period (MP) and once outside of this period (OP). Our results indicated that the median frequency and peak alpha frequency of the power spectrum were low, whereas Shannon spectral entropy was high, during the MP. Theta intensity within the right temporal cortex and right limbic system was significantly lower during the MP than during the OP. High gamma intensity in the left parietal cortex was also significantly lower during the MP than during the OP. Similar differences were also observed in the parietal and occipital regions between the proliferative (the late part of the follicular phase) and secretory phases (luteal phase). Our findings suggest that the menstrual cycle should be considered to ensure accurate interpretation of functional neuroimaging in clinical practice.

Keywords: menstrual cycle, quantitative spectral parameters, regional oscillatory intensity, representative value, spontaneous neural oscillation

INTRODUCTION

Advancements in neuroimaging techniques have led to their widespread adoption in the clinical examination of functional brain diseases (Brammer, 2009; Khanna et al., 2015; Hoshi and Shigihara, 2020; Tanoue et al., 2021). In functional neuroimaging, brain function is primarily assessed based on changes in metabolism (i.e., position emission tomography), blood flow (i.e., functional

magnetic resonance imaging, fMRI), or electrophysiology [i.e., magnetoencephalography (MEG) and electroencephalography (EEG)]. Both MEG and EEG record brain activity in terms of "spontaneous neural oscillations," which are altered by diverse brain disorders such as epilepsy and dementia (Fernández et al., 2013; Ahmed and Rutka, 2016). However, physiological factors such as age and sex can also affect the frequency and regional patterns of spontaneous neural oscillations, representing potential confounding factors in relevant studies (Dustman et al., 1993; Vysata et al., 2012; Barry and De Blasio, 2017; Hoshi and Shigihara, 2020).

Several studies have indicated that the menstrual cycle also alters brain activity (Lindsley and Rubenstein, 1937; de Barenne and Gibbs, 1942; Creutzfeldt et al., 1976; Becker et al., 1982; Bazanova et al., 2014). The menstrual cycle is produced by a network between the hypothalamus, pituitary gland, and ovaries, which interact via sex hormones such as gonadotropin-releasing hormone (GnRH), follicle-stimulating hormone (FSH), luteinising hormone (LH), oestradiol, and progesterone (Creutzfeldt et al., 1976; Franz, 1988; Hawkins and Matzuk, 2008; Brötzner et al., 2014; see **Figure 1**). The cycle causes changes at the functional, molecular, and structural levels of the brain and affects both emotion and cognition (Sacher et al., 2013). Functional changes appear as a slowing down or attenuation of spontaneous neural oscillations during the menstrual cycle, and are accompanied by molecular-related changes in sex hormone levels such as oestradiol and progesterone (Creutzfeldt et al., 1976; Becker et al., 1982; Bazanova et al., 2014; Brötzner et al., 2014). These two alterations have recently attracted attention because they are often observed in patients with cognitive dysfunction, including those with dementia (Ferna and Hornero, 2006; Poza et al., 2007; López et al., 2014; Shigihara et al., 2020a). In the case of cognitive dysfunction, damage to the nucleus basalis of Meynert in the basal forebrain leads to decreased cholinergic input to the cortices, thereby leading to changes in spontaneous neural oscillations (Gratwicke et al., 2013; **Figure 1**). Although the hormones primarily responsible for these changes differ between the menstrual cycle and cognitive dysfunction, evidence suggests a shared neural basis given the interactions between the cholinergic system and oestradiol (Newhouse and Dumas, 2015). Despite the available evidence, most previous studies were limited because they were based on visual inspection (i.e., a subjective analysis). Given that neural signal processing methods have dramatically improved over the last two decades, further studies are required to explore the influence of the menstrual cycle on neural patterns.

During the last few decades, researchers have proposed several parameters for characterising the spectral features of neural oscillations, including median frequency (MF), individual alpha frequency (IAF), and Shannon entropy (SE) (Poza et al., 2007). These parameters can be calculated mechanically from the power spectral density (PSD) of the MEG signals. Thus, they are objective and independent of the skill or experience of the examiner. This is of paramount importance in clinical practice, as it improves the reproducibility, replicability, and reliability of the research findings. These parameters enable intuitive quantification of the diverse properties of brain activity. Furthermore, they are sensitive to changes in cognitive performance and have been highlighted as potential biomarkers for cognitive disorders such as dementia (Poza et al., 2007; Hughes et al., 2019). Indeed, our hospital group utilises these MEG parameters (MF, IAF, and SE) as clinical tools for evaluating hundreds of patients with dementia, stroke, and epilepsy each year. However, no such analyses have revealed how the menstrual cycle affects spontaneous neural oscillations.

Advances in signal processing are not limited to representative parameters of neural oscillations (MF, IAF, and SE). Improvements in processing speed have allowed us to calculate the regional oscillatory intensity of brain activity. Each brain region is associated with specific functions (i.e., "functional specialisation" Mahon and Cantlon, 2011), and there is often a direct link between the damaged cortex and clinical symptoms. Consequently, regional oscillatory intensity gives us information relevant to neurological diseases (Fernández et al., 2013; Sakamoto et al., 2016; Pratt et al., 2017; Shigihara et al., 2020b). Previous studies have revealed that the menstrual cycle affects regional brain activity at both the functional (Bayer et al., 2014; Albert et al., 2015; Pletzer et al., 2019; Weis et al., 2019) and structural levels (Hagemann et al., 2011; Pletzer et al., 2018). fMRI studies have indicated that the menstrual cycle modifies brain activity in various regions, such as the prefrontal cortex, limbic system, hippocampus, amygdala, and striatum (Albert et al., 2015; Pletzer et al., 2019; Weis et al., 2019). Anatomical MRI studies have also reported that grey matter volume changes along the menstrual cycle, reaching its maximum value at the time of ovulation (Hagemann et al., 2011). In contrast, the right parahippocampal/fusiform gyrus reaches its maximum volume during the early follicular phase of the cycle (Pletzer et al., 2018). Although these findings suggest that regional oscillatory intensity should also vary along the menstrual cycle, this hypothesis remains to be verified.

Therefore, in the present study, we aimed to clarify the effect of the menstrual cycle on spontaneous neural oscillations based on quantitative parameters rather than visual inspection, with the goal of providing personalised clinical assessments (i.e., "precision medicine"). More specifically, we aimed to (i) update previous findings related to the "slowing down and attenuation of oscillatory intensity" using quantitative parameters (Study Goal 1) and (ii) identify the role of the regions responsible for the observed changes in resting-state brain activity (Study Goal 2). To achieve these goals, we measured resting-state spontaneous neural oscillations in 25 healthy women using MEG once during the menstrual period (MP) (i.e., "menses") and once outside of the menstrual period (OP) (see **Figure 2**). The representative values of the spontaneous neural oscillations were calculated using established algorithms and compared along the cycle to standardise changes in these values.

MATERIALS AND METHODS

Participants and Ethics

Twenty-five healthy women (mean age ± standard deviation: 28.4 ± 8.0 years; age range: 22–48 years) were enrolled

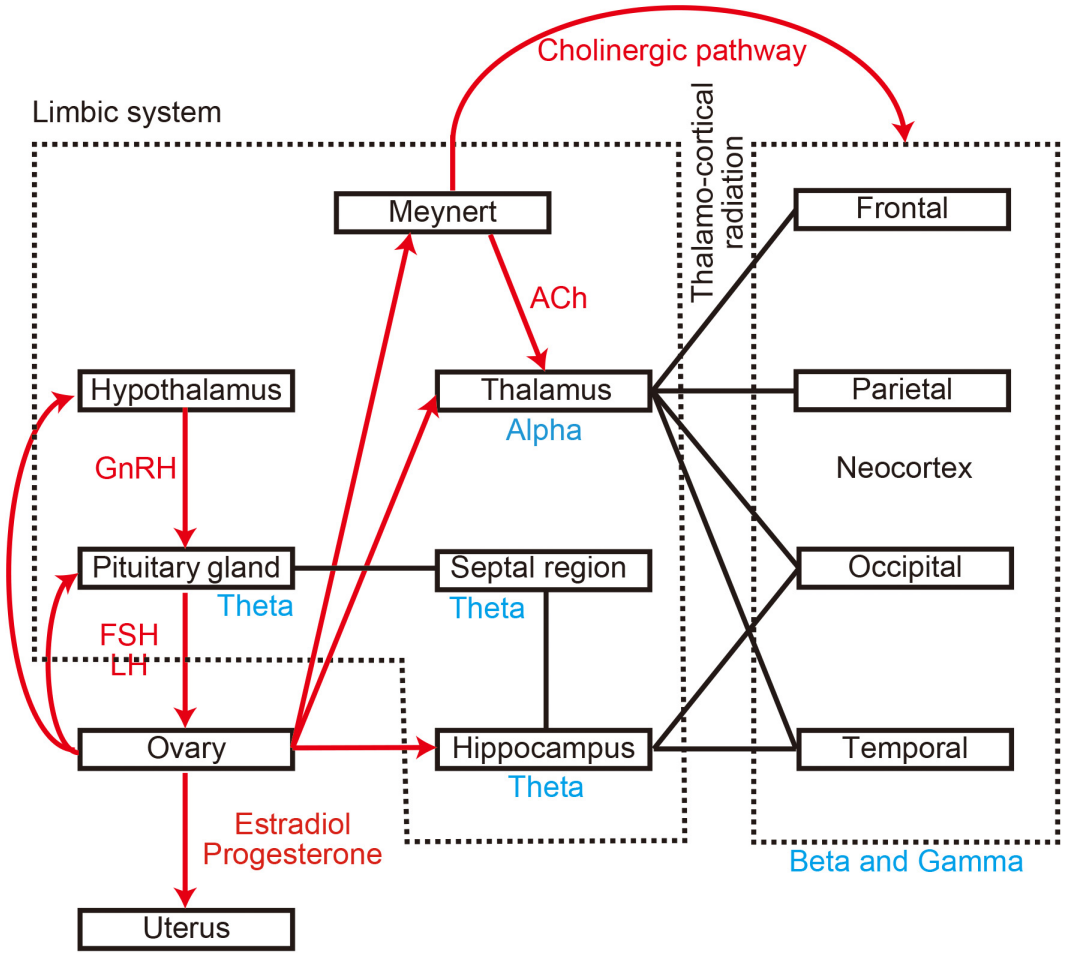

FIGURE 1 | Anatomical and molecular relationship between brain regions. Black lines represent anatomical connections. Red arrows represent molecular connections. Blue words indicate generators of oscillatory activity. GnRH, gonadotropin-releasing hormone; FSH, follicle-stimulating hormone; LH, luteinising hormone; Ach, acetylcholine.

in the present study. The length of their menstrual cycle ranged between 25 and 36 days (29.6 ± 2.7 days), and none of them used oral contraceptives. All participants were staff members at Kumagaya General Hospital, and their health condition was checked annually in accordance with Japanese law (Industrial Safety and Health Act). The present study was conducted in accordance with the ethical principles of the Declaration of Helsinki and was approved by the Ethics Committee of Kumagaya General Hospital (approval number: 14). Written informed consent was obtained from each participant during enrolment.

Procedure

All participants visited the MEG room twice: once during the MP and once during the OP (see **Figure 2**). For eight participants, MEG activity was acquired during the MP first. For the remaining 17 participants, MEG activity was acquired during the OP first. The interval between the two MEG scans ranged from 6 to 133 days (mean ± standard deviation: 41.3 ± 43.4 days). For the OP, participants were asked to report the beginning of the last MP and the average length of their cycle to identify the phase of the cycle (proliferative phase or secretory phase) on the day of MEG recording. One participant failed to report the last MP, and her phase was not identified.

The cycle is generally defined in two ways: ovarian cycle and uterine cycle (**Figure 2**). These two cycles start with the first day (beginning) of the MP, with ovulation occurring in the middle of the cycle. In the ovarian cycle, the period from the first day of menstruation to the day of ovulation is known as the follicular phase, which lasts 16.9 days on average (Bull et al., 2019). The interval from the day of ovulation to the next MP is known as the luteal phase and lasts 12.4 days on average. Thus, we assumed that the menstrual cycle can be divided into follicular and luteal phases using the ratio 16.9:12.4, regardless of the length of the menstrual cycle (Bull et al., 2019). We used this ratio to estimate the phase of the cycle (follicular or luteal) on the recording day. In the uterine cycle, the MP and subsequent proliferative phase (i.e., "pre-ovulatory phase") occupy the follicular phase of the ovarian cycle. The secretory phase of the uterine cycle corresponds to the luteal phase of the ovarian cycle. Thus, the

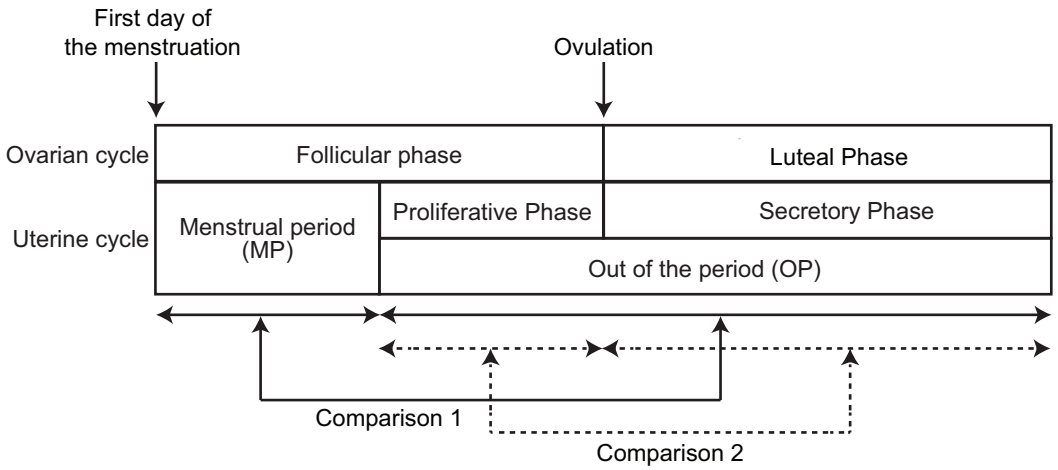

FIGURE 2 | Periods and phases of the menstrual cycle. The first row defines the ovarian cycle, whereas the second row defines the uterine cycle. One cycle includes two periods: the menstrual period (MP) and the period outside of menstruation (OP). The latter is further broken down into two phases: the proliferative and secretory phases. The follicular phase comprises the MP and proliferative phase, whereas the secretory phase is equivalent to the luteal phase. In the present study, neural activity was compared in two ways: (i) between the MP and OP (Comparison 1, solid arrows) and (ii) between the proliferative and secretory phases (Comparison 2, broken arrows).

OP consists of the proliferative phase and secretory phase of the uterine cycle.

In the present study, we performed two different comparisons to assess changes in the spectral content of neural oscillatory MEG activity along the menstrual cycle. First, we compared spontaneous neural oscillations between the MP and OP (Comparison 1 in **Figure 2**, within-participant design). Second, we compared spontaneous neural oscillations between the proliferative phase (the later part of the follicular phase) and secretory phase (luteal phase) within the OP (Comparison 2 in **Figure 2**, between-participant design). Eleven participants visited the MEG room during the proliferative phase, while 13 participants visited during the secretory phase. Data from one participant was used for Comparison 1 only and was excluded from Comparison 2 because the participant failed to report the last MP, and the phase could not be identified.

MEG Scanning Details

Spontaneous neural oscillations were recorded for 5 min using a 160-channel whole-head type MEG system (RICOH160-1; RICOH Co., Ltd., Tokyo, Japan) in a magnetically shielded room. During the scan, participants were asked to remain calm in the supine position with their eyes closed. The scanning conditions were controlled to be as consistent and comfortable for participants as possible. The sensors and reference coils were gradiometers, with diameters of 15.5 and 50 mm at the baseline, respectively. Each pair of sensor coils was separated by a distance of 23 mm. The sampling frequency was 2,000 Hz, and a 500-Hz low-pass filter was used during recording. To co-register MEG source images with structural brain images acquired using canonical MRI, three fiducial magnetic marker coils were placed on each participant's face (5 mm above the nasion and bilaterally 10 mm in front of the tragus) during the MEG scan.

MEG Data Analysis

MEG data were pre-processed offline using the software package SPM-12 (Wellcome Trust Centre for Neuroimaging, London, United Kingdom[1]) and the MEAW system[2]. Two types of standard MEG analyses were applied: sensor-level and source-level analyses. Each method is associated with advantages over the other (Shigihara et al., 2020a). Sensor-level analysis produces mathematically reliable results and is less time-consuming than source-level analysis (i.e., taking only a few minutes), making it valuable for clinical practice. Sensor-level analysis is also sensitive to global changes in brain activity, allowing researchers to examine individual differences in spatial distribution. In the present study, sensor-level analysis was used to obtain representative values of the PSD and to replicate the previous finding that spontaneous neural oscillations slow during the MP (Lindsley and Rubenstein, 1937; de Barenne and Gibbs, 1942; Creutzfeldt et al., 1976; Becker et al., 1982; Brötzner et al., 2014) (Study Goal 1). In contrast, source-level analysis provides information regarding regional brain activity, although it is time-consuming (i.e., taking 30–60 min for a single patient). Source-level analyses were adopted to assess differences in regional changes between two conditions (MP vs. OP, or proliferative phase vs. secretory phase) (Study Goal 2).

Sensor-Level Data Processing

Sensor-level analyses were performed in accordance with the protocol described in our previous study (Shigihara et al., 2020a). If necessary, prominent artefacts were manually removed via principal component analysis (PCA) using the MEG analysis software developed by the manufacturer because spectral parameters are sensitive to artefacts. The software is authorised

[1]https://www.fil.ion.ucl.ac.uk/spm/

[2]https://www.hokuto7.or.jp/hospital/lang/english-home/meaw/

for clinical use by the Ministry of Health, Labour, and Welfare of Japan (equivalent to FDA approval). Most artefacts are caused by silver tooth fillings and eye movements; hence, their frequency is usually low (delta to theta range), which can influence the computation of spectral parameters. In addition, artefacts from outside the brain present as characteristic patterns on contour maps and can lead to unnatural time courses for the PCA components (Gross et al., 2013). Experienced clinicians and technicians can distinguish these artefacts from brain signals based on visual inspection. A 50-Hz band-stop filter was also applied to remove power line noise. Thereafter, three spectral parameters were calculated to summarise different properties of spontaneous neural oscillations: MF, IAF, and SE (Poza et al., 2007). The spectral parameters were computed from the PSD, which was estimated using the Blackman–Tukey method considering non-overlapping 10-s epochs. Afterward, the PSD was normalised between 1 and 70 Hz (PSDn) (Gómez et al., 2013). The first parameter, MF, refers to the median of the distribution represented by the PSDn (i.e., the frequency that splits the PSDn into two halves of equal power) (Poza et al., 2007). The second parameter, IAF, refers to the frequency corresponding to the peak of the PSDn in the extended alpha band (4–15 Hz) (i.e., the dominant alpha activity), which usually appears in human adults in the eyes-closed resting condition (Poza et al., 2007). IAF is useful for describing the loss of neural oscillations in the alpha band (i.e., the "*shift-to-the-left*" of the alpha peak). Finally, SE is an irregularity measure closely related to the concept of order in information theory that quantifies the distribution of the oscillatory components of the PSDn (Poza et al., 2007). These three parameters were calculated for each epoch and MEG sensor, following which they were averaged across epochs and sensor position (left and right). Of note, the side of the sensors was not completely matched with each brain hemisphere because our MEG system was equipped with axial gradiometers. As such, left sensors received some signals from the right hemisphere. However, it should be noted that a previous study demonstrated that there is little signal contamination across regions, and that sensor-level information (i.e., that from left and right sensors) is nearly consistent with source-level information (i.e., that from the left and right hemispheres) (Rodríguez-González et al., 2020). Source-level analysis was performed to accurately determine the source of the signals (see section "Source-Level Analysis"). Finally, the averaged parameters were statistically analysed as described in section "Statistical Analysis".

Source-Level Analysis

The source-level analysis for individual participants (first-level analysis) followed the pipeline used in a previous study (Shigihara et al., 2020b). Continuous MEG signals were divided into non-overlapping 10-s epochs. Because the experimental environment generated a utility frequency, a 50-Hz band-stop filter was applied to the epoched data. These filtered data were then directly used for source-level analyses. To identify the brain regions producing the resting-state-induced components, a source inversion procedure was applied to the delta (0–3 Hz), theta (4–7 Hz), alpha (8–12 Hz), beta (13–25 Hz), and gamma (low-gamma, 26–40 Hz; high gamma, 41–80 Hz) oscillatory components separately, using a maximal smoothness algorithm with a spatially coherent sources model (i.e., the COH algorithm implemented in SPM-12) (Friston et al., 2008). This source localisation algorithm is comparable to those used in standardised low-resolution brain electromagnetic tomography (Pascual-Marqui, 2002). The COH algorithm is a popular source inversion algorithm that is often used in clinical environments (Terakawa et al., 2008; Ray and Bowyer, 2010; Shigihara et al., 2020b). Forward modelling was performed for the whole brain using a single-shell model with canonical MRIs provided by SPM-12. The source inversion and estimation steps were performed by applying filters corresponding to each frequency band (from delta to high gamma). No source priors were used for source estimation. The estimated oscillatory intensity at each frequency band and in each brain region (i.e., regional oscillatory intensity) was saved as a source image file in the NIfTI format and used for the second (group)-level analysis.

To increase the sensitivity to changes in regional neural oscillatory intensity between the two conditions in two ways (Comparisons 1 and 2 in **Figure 2**), we analysed the data in different regions of interest (ROIs). Five sets of ROI mask images (NIfTI format) were created using the WFU pick atlas[3] for each hemisphere (left and right) in order to cover the whole cortex: frontal, temporal, parietal, occipital, and limbic system. Oscillatory intensities were averaged within each mask using the SPM function "spm_summarise" for each condition in each participant. Averaged oscillatory intensities in the ROIs (ROI values) were analysed as described in section "Statistical Analysis".

Statistical Analysis

We first performed Comparison 1 within each participant. As condition order was not counterbalanced (eight participants were scanned during the MP first while 17 were scanned in reverse order), we examined the effect of order. The three sensor-level spectral parameters were compared between first and second MEG scans using a bootstrapping method, irrespective of the MP or phases. For each parameter, the average difference between the first and second scans was computed via resampling with replacement data across all participants 20,000 times, and the percentage of the resampled average difference larger or smaller than 0 (the smaller value) was taken as the significance level (p-value). As we observed statistically significant effects of order, the effects were taken into account in the following statistical analysis using a weighted bootstrapping approach. For this task, the weights (i.e., proportion for resampling) were controlled so that half of the resampled data were derived from the first MEG scan, while the other half were derived from the second scan. This approach resulted in pseudo-counterbalancing after resampling, which allowed us to examine the effects of interest (e.g., difference between phases) while controlling for the effects of order.

Second, sensor-level spectral parameters were compared between the MP and OP (Comparison 1), and between the proliferative and secretory phases (Comparison 2) using a weighted bootstrapping method. For Comparison 1, the average

[3]https://www.nitrc.org/projects/wfu_pickatlas/

difference between MP and OP was computed via (weighted) resampling with replacement data across all participants 20,000 times, and the percentage of the resampled average difference larger or smaller than 0 (the smaller value) was taken as the significance level (p-value). For Comparison 2, the group (i.e., proliferative and secretory phases) average was computed via (weighted) resampling with replacement data across all participants 20,000 times, and the group difference was stored for each iteration. The percentage of the group differences larger or smaller than 0 (the smaller value) was taken as the significance level (p-value).

We also examined the relationships between participant age, length of the menstruation cycle, and sensor-level spectral parameters (during the MP, OP, and the difference between the two) using a weighted bootstrapping approach. For each pair of variables, Pearson's coefficients were calculated via (weighted) resampling with replacement data across all participants 20,000 times. The percentage of the resampled coefficients larger or smaller than 0 (the smaller value) was taken as the significance level (p-value).

Finally, we performed Comparisons 1 and 2 for source-level ROI values using the same method used for sensor-level spectral data. The comparisons were made for each ROI and frequency band.

For all statistical analyses, we report the grand mean of the statistical values (e.g., mean value of each condition, standard error of resampled cases, group-difference: d, and average correlation coefficient: r) across bootstrap iterations, as well as p-values. Throughout all statistical examinations, statistically significant p-values were determined after controlling the false detection rate (FDR) using the Benjamini and Hochberg method (Benjamini and Hochberg, 1995).

RESULTS

Sensor-Level Analysis (Spectral Parameters)

Effect of Order

We observed significant effects of order on MF for sensors on the right side (rtMF) and on IAF for sensors on both sides (rtMF: $d = -0.283$, $p = 0.025$; IAF in the left, ltIAF: $d = -0.199$, $p = 0.027$; IAF in the right, rtIAF: $d = -0.214$, $p = 0.020$), indicating that values measured first were higher than those measured second.

Comparison 1: MP vs. OP

For sensors on both sides, rtMF and IAF were lower during MP than during OP (rtMF: $d = -0.288$, $p = 0.011$; ltIAF: $d = -0.238$, $p = 0.004$; rtIAF: $d = -0.253$, $p = 0.002$; **Figure 3**). Although MF values for the left side (ltMF) were lower during MP than during OP, the difference did not reach significance after FDR correction ($d = -0.236$, $p = 0.040$). SE values on both sides were higher during MP than during OP (right side, rtSE: $d = 0.008$, $p = 0.001$; left side, ltSE: $d = 0.009$, $p = 0.002$).

For sensors on both sides, the difference between MF values and the difference between IAF values were positively correlated with MF and IAF values during the MP, respectively (ltMF: $r = 0.593$, $p = 0.006$; rtMF: $r = 0.623$, $p = 0.002$; ltIAF: $r = 0.587$, $p = 0.001$; rtIAF: $r = 0.614$, $p < 0.001$), but not during OP. Differences in SE on both sides were negatively correlated with SE during OP (ltSE: $r = -0.572$, $p < 0.001$; rtSE: $r = -0.635$, $p < 0.001$), but not during MP.

Comparison 2: Proliferative Phase (Later Part of the Follicular Phase) vs. Secretory Phase (Luteal Phase)

Although MF and IAF values were higher, while SE values were lower, on both sides during the secretory phase than during the proliferative phase (**Figure 4**), these differences did not reach significance after FDR correction (ltMF, $p = 0.190$; rtMF, $p = 0.240$; ltIAF, $p = 0.202$; rtIAF, $p = 0.270$; ltSE, $p = 0.314$; rtSE = 0.329).

Age and Cycle Length

For sensors on both sides, age was positively correlated with the MF value during the MP (ltMF: $r = 0.416$, $p = 0.012$; rtMF: $r = 0.398$, $p = 0.018$) (**Table 1**). In addition, age was positively correlated with the change in IAF between MP and OP on the left side ($r = 0.206$, $p = 0.040$). The length of the menstrual cycle was negatively correlated with the IAF value on both sides during the OP (ltIAF: $r = -0.379$, $p = 0.006$; rtIAF: $r = -0.407$, $p = 0.002$). The correlation between participant age and cycle length was not statistically significant ($p = 0.214$). No statistically significant correlations were observed for the other pairs between age/cycle length and spectral parameters (**Table 1**).

Source-Level Analysis (Regional Oscillatory Intensity)

Comparison 1: MP vs. OP

Theta intensity within the right temporal cortex and right limbic system was significantly lower during the MP than during the OP (right temporal: $d = -0.007$, $p = 0.010$; right limbic system: $d = -0.002$, $p = 0.003$) (**Table 2**, **Figure 5**, and **Supplementary Figure 1**). High gamma intensity in the left parietal cortex was also significantly lower during the MP than during the OP ($d = -0.017$, $p = 0.037$). No other differences in regional oscillatory intensity were observed between the two conditions.

Comparison 2: Proliferative Phase vs. Secretory Phase

Significant changes in regional oscillatory intensity were observed on both sides for the parietal and occipital regions. Theta oscillatory intensity was higher on both sides of the parietal region during the proliferative phase than during the secretory phase (**Table 3** and **Figure 6**), while that in the occipital region was lower during the proliferative phase than during the secretory phase. In addition, delta oscillatory intensity was higher while low gamma oscillatory intensity was lower during the proliferative phase than during the secretory phase on both hemispheres. No other differences were observed for any region or frequency band.

DISCUSSION

The present study investigated the effect of the menstrual cycle on spontaneous neural oscillations based on quantitative MEG

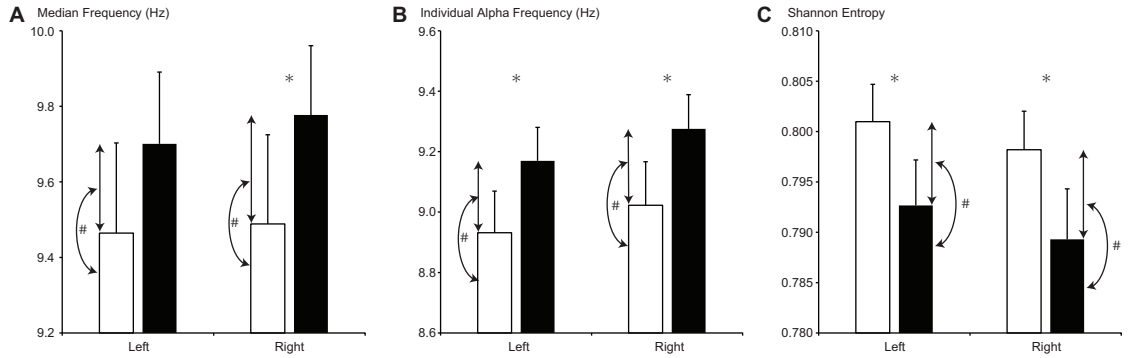

FIGURE 3 | Changes in **(A)** median frequency, **(B)** individual alpha frequency, and **(C)** Shannon entropy between the menstrual period (MP, open columns) and the period outside of menstruation (OP, filled columns). Error bars represent the grand mean of standard errors over bootstrapped iterations. Asterisks (*) indicate statistically significant differences (after false-discovery rate (FDR) correction) between the two periods. Hash tags (#) indicate statistically significant correlations (after FDR correction) between the value of the parameter in each period and its change.

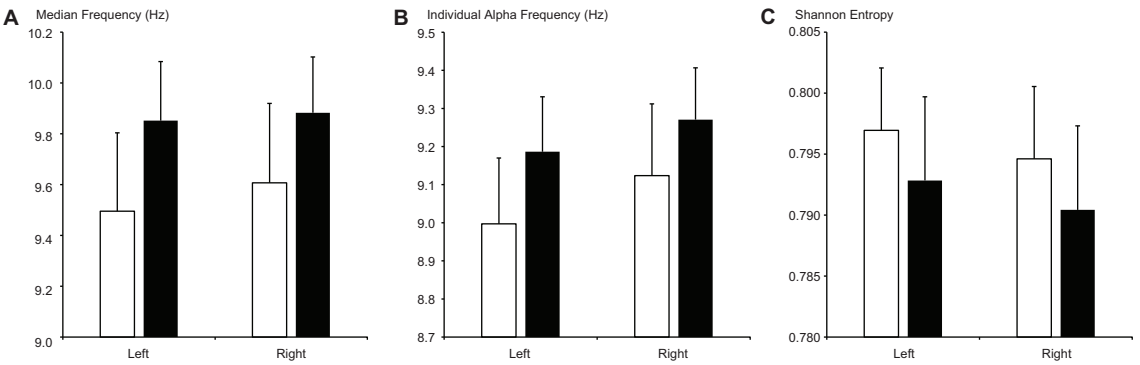

FIGURE 4 | Changes in **(A)** median frequency, **(B)** individual alpha frequency, and **(C)** Shannon entropy between the proliferative phase (open columns) and secretory phase (filled columns). Error bars represent the grand mean of standard errors over bootstrapped iterations.

TABLE 1 | Correlation between spectral parameters and participant age/length of the menstrual cycle (Comparison 1 in sensor-level analysis).

		MF				**IAF**				**SE**			
		Left		**Right**		**Left**		**Right**		**Left**		**Right**	
		r	***p***	***r***	***p***	***r***	***p***	***r***	***p***	***r***	***p***	***r***	***p***
Menstrual period (MP)	Age	0.416	0.012*	0.398	0.018*	0.222	0.086	0.248	0.071	0.104	0.322	0.117	0.281
	Cycle	−0.171	0.150	−0.124	0.247	−0.219	0.083	−0.169	0.155	−0.273	0.100	−0.232	0.107
Outside of the menstrual period (OP)	Age	0.381	0.050	0.310	0.070	0.103	0.207	0.132	0.167	0.008	0.498	−0.032	0.432
	Cycle	−0.120	0.250	−0.213	0.115	−0.379	0.006*	−0.407	0.002*	−0.120	0.260	−0.157	0.185
Change	Age	0.163	0.236	0.275	0.059	0.206	0.040*	0.214	0.049	0.121	0.257	0.180	0.170
	Cycle	−0.116	0.219	0.082	0.296	0.117	0.187	0.216	0.087	−0.134	0.161	−0.015	0.446

*r, average correlation coefficient across bootstrap iterations; p, p-values of weighted bootstrapping statistics, * indicates statistically significant correlation after FDR correction.*

parameters. Our analysis revealed four major findings: (i) MF and IAF were lower during MP than during OP, mainly on the right side of the MEG dewar; (ii) SE was higher during MP than during OP; (iii) theta oscillatory intensity was lower in the right temporal cortex and limbic system during MP than during OP; and (iv) theta oscillatory intensity in the parietal and occipital regions differed between the proliferative and secretory phases.

The menstrual cycle is a fundamental body rhythm that is regulated by five basic hormones: GnRH, FSH, LH, oestradiol, and progesterone (Creutzfeldt et al., 1976; Franz, 1988; Hawkins and Matzuk, 2008; Brötzner et al., 2014; **Figure 1**). These hormones interact with each other and affect various brain functions in the sensory, cognitive, and emotional domains (Farage et al., 2008; Poromaa and Gingnell, 2014). As neural

TABLE 2 | Average change in oscillatory intensity within ROIs and their significance between the menstrual period (MP) and the period outside of menstruation (OP) (Comparison 1 in source-level analysis).

	Frontal				Temporal				Parietal				Occipital				Limbic system			
	Left		Right		Left		Right		Left		Right		Left		Right		Left		Right	
Frequency band	*d*	*p*	*d*	*p*	*d*	*p*	*d*	*p*	*d*	*p*	*d*	*p*	*d*	*p*	*d*	*p*	*d*	*p*	*d*	*p*
Delta	0.000	0.474	0.000	0.350	0.000	0.453	0.000	0.407	−0.002	0.155	−0.001	0.369	−0.002	0.266	−0.001	0.210	−0.001	0.293	−0.001	0.237
Theta	−0.001	0.265	−0.002	0.177	0.002	0.330	−0.007	0.010*	0.006	0.238	0.003	0.254	0.009	0.097	−0.001	0.459	0.001	0.296	−0.002	0.003*
Alpha	0.001	0.171	0.001	0.231	0.000	0.498	0.000	0.491	0.002	0.353	0.005	0.162	−0.002	0.343	0.000	0.482	0.000	0.434	−0.001	0.131
Beta	0.001	0.323	0.000	0.465	0.004	0.129	0.001	0.429	0.007	0.144	0.008	0.086	0.011	0.075	0.006	0.098	0.001	0.270	−0.001	0.249
Low gamma	0.000	0.484	0.001	0.353	0.000	0.474	−0.001	0.428	−0.002	0.391	0.008	0.115	0.006	0.238	0.000	0.461	−0.002	0.079	−0.001	0.213
High gamma	−0.002	0.234	0.004	0.110	−0.004	0.202	0.004	0.180	−0.017	0.037*	0.003	0.327	−0.005	0.204	−0.002	0.355	−0.004	0.136	0.001	0.378

*d, average difference across bootstrap iterations; p, p-values of weighted bootstrapping statistics, * indicates statistically significant difference after FDR correction. ROI, region of interest.*

oscillations are sensitive to changes in brain function (BaŞar, 2013; Mazaheri et al., 2018; Ohki and Takei, 2018), previous studies have already reported that the EEG power spectrum changes across the menstrual cycle (Lindsley and Rubenstein, 1937; de Barenne and Gibbs, 1942; Creutzfeldt et al., 1976; Becker et al., 1982; Bazanova et al., 2014). Most of these previous studies focussed on changes in alpha oscillations, which can easily be identified and evaluated via visual inspection without sophisticated techniques such as signal processing methods (Lewine and Orrison, 1995).

Mean alpha frequency is higher during the luteal phase (secretory phase) than during menstruation, and the change is coupled with an increase in progesterone levels (Creutzfeldt et al., 1976; Becker et al., 1982; Bazanova et al., 2014; Brötzner et al., 2014). Other studies have reported that increased alpha frequency during the luteal phase is associated with changes in cognitive performance (Grandy et al., 2013; Bazanova et al., 2014). Although it is difficult to analyse other neural oscillations (i.e., delta, theta, beta, and gamma) without sophisticated techniques, some studies have also reported that alterations in these rhythms occur during the menstrual cycle. Nonetheless, changes in theta oscillations remain controversial. On one hand, Creutzfeldt et al. (1976) reported that mean theta power was lower during the luteal phase than during the follicular phase. On the other hand, Becker et al. (1982) reported that mean theta power was lower during the periovulatory period and higher during the perimenstrual period. This discrepancy may be explained by methodological and statistical limitations in the earlier study, which was conducted approximately 40–50 years ago. Further evidence indicated that beta intensity was significantly larger during the follicular phase than during the luteal phase (Becker et al., 1982).

Despite these interesting findings, most previous studies regarding the relationship between the menstrual cycle and oscillatory patterns were based on visual inspection rather than signal processing methods. In the present study, we updated these findings by computing three different spectral parameters (MF, IAF, and SE) (Study Goal 1) and regional oscillatory intensities (Study Goal 2). We primally focused on differences between MP and OP (Comparison 1) because they are practically/clinically useful. Clinical practice is based on a good relationship between the patient and clinician. Details regarding menstruation can be difficult to discuss during a patient's first visit, especially when the clinician is male and when the patient's chief complaint does not seem to be related. Given its relevance to many conditions, the classification of the MP and OP is essential in clinical practice. In addition, we performed Comparison 2 (proliferative vs. secretory phase) within the OP because the endocrine environment differs between these phases. Although information from Comparison 2 is applicable in limited gynaecological situations, such information can help us to evaluate MEG data more appropriately. Another potential area of interest may be the relationship between hormone levels and dementia, as mentioned in the Introduction. Some research indicates that hormone replacement therapy can attenuate the progression of dementia in women, for example (Moradi et al., 2018; Zhou et al., 2020). MEG may therefore aid in determining the efficacy of such therapy.

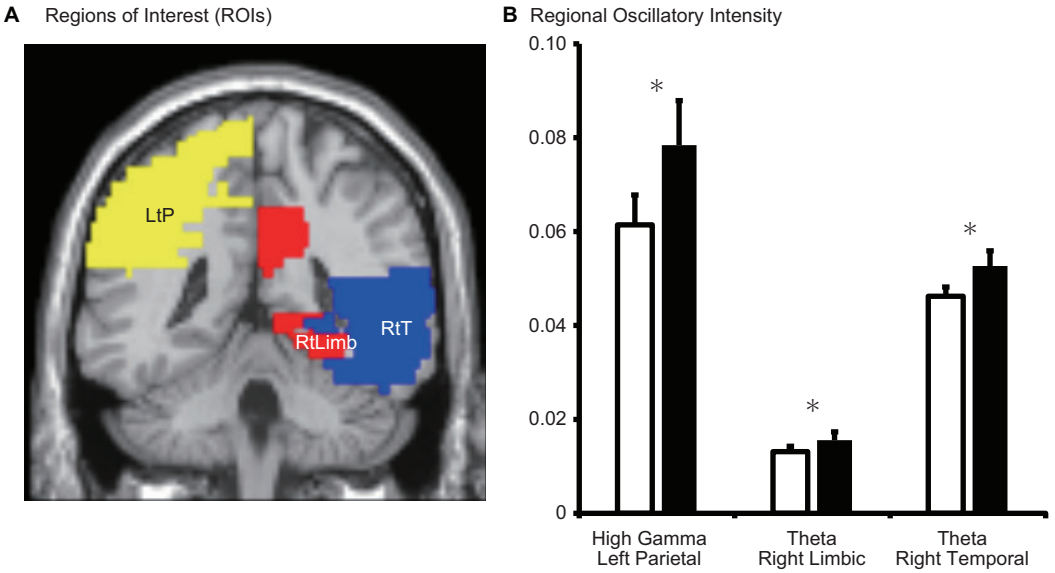

FIGURE 5 | (A) Schematic representation of regions of interest (ROIs) and **(B)** regional oscillatory intensity changes during the menstrual period (MP, open columns) and the period outside of menstruation (OP, filled columns). Three parameters with significant changes (after false-discovery rate (FDR) correction) were displayed. Error bars represent the grand mean of standard errors over bootstrapped iterations. LtP, left parietal; RtLimb, right limbic system; RtT, right temporal. Asterisks (*) indicate statistically significant difference after FDR correction.

Spectral Parameters (Sensor-Level Analysis)

In the sensor-level analysis, three spectral parameters (MF, IAF, and SE) were used to quantify global changes in neural oscillations related to the menstrual cycle. In this analysis, we aimed to replicate and expand upon the previous visually based finding that alpha frequency is low during the MP and high during the OP, especially during the secretory phase (i.e., luteal phase) (Study Goal 1). The IAF reflects the peak frequency of alpha oscillations. IAF values were lower during the MP than during the OP (Comparison 1). Although they were also lower during the proliferative phase than during the secretory phase, this difference was not significant (Comparison 2). These findings are largely consistent with those of previous studies (Lindsley and Rubenstein, 1937; de Barenne and Gibbs, 1942; Creutzfeldt et al., 1976; Becker et al., 1982).

The MF reflects the oscillatory power balance between high and low frequencies. Although both IAF and MF are affected by changes in peak alpha frequency, MF is also affected by changes in the theta, beta, and gamma bands. As observed for IAF, MF was lower during the MP than during the OP (Comparison 1). Similarly, although MF values were lower during the proliferative phase (the later part of the follicular phase) than during the secretory phase (luteal phase), this difference was not significant (Comparison 2). There are three possible interpretations regarding the change in MF: (i) changes in alpha (peak) frequency, (ii) changes in power balance between high and low frequencies, or (iii) both. These changes are noteworthy because low MF and IAF values are often observed in patients with cognitive dysfunction and dementia (Poza et al., 2007). Previous studies have indicated that some cognitive functions are impaired during the MP, including mental rotation, visuospatial ability, verbal memory, and verbal fluency, although these findings remain controversial (Sommer, 1992; Schöning et al., 2007; Hatta and Nagaya, 2009; Poromaa and Gingnell, 2014; Farrar et al., 2015; Sundström-Poromaa, 2018). Low values of MF and IAF during the MP may reflect impairments in cognitive function due to the menstrual cycle.

The SE is another spectral parameter that represents the irregularity of the distribution of the neural oscillatory components in the PSDn. Our findings indicated that SE was higher during the MP than during the OP (Comparison 1). Although SE was also higher during the proliferative phase than during the secretory phase, this difference was not significant (Comparison 2). Patients with dementia tend to exhibit lower SE values, in addition to low MF and IAF values (Poza et al., 2007). However, our participants exhibited lower MF/IAF values and higher SE values during the MP than during the OP, suggesting that the distribution of neural oscillatory components during MP differs from that related to dementia.

Changes in MF and IAF values between the MP and OP (Comparison 1, marked with # in **Figure 3**) were positively correlated with their values during the MP, but not during the OP. This result suggests that participants with low values of MF or IAF during the MP exhibited larger changes in these parameters between the MP and OP. Furthermore, despite being a relatively short period within the menstrual cycle, the MP appears to regulate changes in MF or IAF. In addition, the changes in SE were negatively correlated with the value of SE during the OP, but not during the MP, suggesting that participants with low SE

TABLE 3 | Average oscillatory intensity within ROIs and the corresponding statistical differences between the proliferative and secretory phases (Comparison 2 in source-level analysis).

		Left			Right		
	Freq	**Proliferative *M***	**Secretory *M***	***p***	**Proliferative *M***	**Secretory *M***	***p***
Frontal	Delta	0.035	0.033	0.299	0.030	0.027	0.168
	Theta	0.026	0.024	0.325	0.025	0.023	0.334
	Alpha	0.009	0.008	0.326	0.010	0.009	0.212
	Beta	0.021	0.016	0.079	0.018	0.016	0.221
	Low gamma	0.032	0.028	0.172	0.027	0.023	0.151
	High gamma	0.028	0.025	0.252	0.020	0.018	0.242
Temporal	Delta	0.039	0.037	0.233	0.030	0.028	0.235
	Theta	0.058	0.066	0.096	0.053	0.054	0.450
	Alpha	0.047	0.048	0.419	0.047	0.044	0.330
	Beta	0.051	0.045	0.203	0.038	0.033	0.185
	Low gamma	0.049	0.053	0.288	0.032	0.036	0.235
	High gamma	0.044	0.042	0.441	0.029	0.031	0.410
Parietal	Delta	0.016	0.009	0.076	0.015	0.010	0.139
	Theta	0.106	0.078	0.033*	0.083	0.058	0.013*
	Alpha	0.079	0.064	0.141	0.062	0.049	0.125
	Beta	0.111	0.080	0.092	0.072	0.051	0.062
	Low gamma	0.124	0.116	0.331	0.087	0.074	0.145
	High gamma	0.077	0.071	0.370	0.048	0.042	0.272
Occipital	Delta	0.022	0.013	0.036*	0.008	0.004	0.006*
	Theta	0.098	0.121	0.038*	0.068	0.090	0.015*
	Alpha	0.121	0.133	0.144	0.089	0.091	0.435
	Beta	0.100	0.102	0.446	0.064	0.071	0.219
	Low gamma	0.083	0.115	0.007*	0.054	0.074	0.008*
	High gamma	0.111	0.109	0.402	0.065	0.066	0.458
Limb	Delta	0.043	0.038	0.099	0.030	0.028	0.151
	Theta	0.016	0.021	0.124	0.015	0.017	0.306
	Alpha	0.012	0.013	0.329	0.011	0.011	0.442
	Beta	0.021	0.022	0.390	0.015	0.014	0.427
	Low gamma	0.022	0.024	0.293	0.015	0.016	0.370
	High gamma	0.023	0.028	0.205	0.015	0.018	0.231

M: average oscillatory intensity within ROI at given frequency in given phase; p, p-values of weighted bootstrapping statistics, asterisks (*) indicate statistically significant difference after FDR correction. ROI, region of interest.

values during OP exhibited larger changes in SE between the MP and OP. These results support the notion that MF/IAF and SE reflect different properties of the PSDn.

In accordance with previous findings, MF and IAF exhibited a partial positive correlation with age (Gómez et al., 2013; Hoshi and Shigihara, 2020). However, we observed no significant correlation between participant age and cycle length. We speculate that this is due to bias in participant age: Most participants were in their 20 s, only one was in her 30 s, and four were in their 40 s. Research has indicated that the length of the mensural cycle progressively decreases with age, and that IAF progressively increases with age (Gómez et al., 2013; Bull et al., 2019; Hoshi and Shigihara, 2020), consistent with our finding that the length of the menstrual cycle was negatively correlated with IAF.

Overall, the sensor level analysis successfully replicated the previous findings and provided two additional results: (i) It is plausible that the menstrual cycle affects frequencies other than alpha rhythms; (ii) oscillatory brain patterns during the MP (i.e., increased SE) differ from those associated with dementia.

Regional Oscillatory Intensity (Source-Level Analysis)

Whereas the sensor-level analysis investigated global changes in neural oscillations in terms of quantitative spectral parameters (MF, IAF, and SE), the source-level analysis provided deeper insight into the brain regions responsible for these changes (Study Goal 2). Interestingly, we observed significant changes in regional oscillatory intensity in the theta and gamma bands rather than the alpha band (Comparison 1 in **Figure 5** and **Table 2**, Comparison 2 in **Figure 6** and **Table 3**). Theta oscillations are generated by a neural network between the hypothalamus, septal region, and hippocampus (Colom, 2006; **Figure 1**). The hippocampus is a major structure of the limbic system. Given that the hippocampus is rich in oestrogen receptors, it is not surprising that it exhibits

changes in function during the menstrual cycle (Woolley et al., 1990; Shughrue and Merchenthaler, 2000; Lisofsky et al., 2015). The hippocampus plays a key role in short-term memory: While the right hippocampus is important for spatial navigation, the left hippocampus is important for verbal memory (Hamid, 2014; Ezzati et al., 2016; Kang et al., 2016). Performance on mental rotation tasks, which are used to investigate spatial navigation, fluctuates with oestradiol and progesterone levels. Previous studies have indicated that task performance is high during the MP and low during the midluteal phase (Hampson et al., 2014; Barel et al., 2019). The hippocampus is located in the medial part of the temporal lobe and is connected with the other parts of the temporal lobe (Rolls, 1989; Wilson et al., 1990; Maller et al., 2019). Consequently, it is reasonable to suggest that theta intensity in the limbic system and temporal region changes between the MP and OP (Comparison 1 in **Figures 2**, **5**). Comparison 1 also revealed that high gamma intensity in the left parietal region differed between the MP and OP. The left parietal lobe is a critical region for verbal processing (Coslett and Schwartz, 2018), which is also modified during the menstrual cycle (Šimič and Santini, 2012; Barel et al., 2019): Women exhibit their best performance on verbal fluency tasks during the menstrual and midluteal phases and worst performance during the early follicular phase (Šimič and Santini, 2012; Barel et al., 2019). Furthermore, gamma oscillatory activity is affected by grey matter volume (Schwarzkopf et al., 2012), and research has indicated that the volume of the left partial lobe changes during the menstrual cycle (Protopopescu et al., 2008). It is plausible that changes in grey matter volume during the menstrual cycle are linked with changes in verbal fluency task performance and gamma oscillatory intensity between the MP and OP (Comparison 1).

Comparison 2 (proliferative phase vs. secretory phase) revealed that the menstrual cycle is associated with changes in oscillatory intensity in the parietal and occipital regions. Changes in theta oscillatory intensity were observed in both regions. These two regions are important for visual perception, which is also affected by the menstrual cycle (Ward et al., 1978). The amplitude of the typical visual-evoked response, which is modulated by the top-down attention system, is larger during the mid-luteal phase (Lusk et al., 2015) than during other phases. Notably, changes in these regions were observed in the theta band but not in the gamma band, which is typically associated with the bottom-up components of visual perception (Tallon-Baudry, 2009; Bastos et al., 2012; Sauseng et al., 2015; Bartoli et al., 2020). Theta oscillations are mainly generated in the hippocampus (Colom, 2006). However, it is unlikely that hippocampal activity directly modifies theta oscillations in these regions. The strongest target of hippocampal connectivity is the temporal lobe, followed by the limbic system and subcortical structures such as the thalamus (Maller et al., 2019). The occipital lobe is the third target, while the parietal lobe exhibits weak direct connectivity with the hippocampus. Thus, it is not plausible that the hippocampus modifies theta oscillations in the parietal and occipital regions without eliciting significant changes in the temporal region.

The hippocampus is connected with the thalamus, which contains receptors for oestradiol, and several studies have demonstrated that the menstrual cycle influences thalamic activity (Sacher et al., 2013; Brinton et al., 2015; Hidalgo-Lopez et al., 2020). In addition, the thalamus is connected with almost all cortices via the thalamocortical radiations (Cunningham et al., 2017; Caspers and Zilles, 2018; George and Das, 2019). These radiations are categorised into four distinct parts based on anatomical position: anterior thalamic radiations, posterior thalamic radiations, superior thalamic radiations, and inferior thalamic radiations (George and Das, 2019). Hence, we assume that parietal and occipital theta oscillatory activity is modulated by the hippocampus via the posterior thalamic radiations, whose target regions are the parietal and occipital regions. This assumption may explain the mechanism underlying the effect of the menstrual cycle on visual perception. The posterior thalamic radiation conveys top-down signals to these two regions from other neural networks such as the attention system using theta oscillations, which serve to provide top-down control (Bastos et al., 2012; Sauseng et al., 2015). Attention exhibits a close relationship with visual perception (Luck and Ford, 1998), and previous studies have indicated that the attention system is affected by the menstrual cycle via progesterone levels (Pletzer et al., 2017). Other studies have indicated that attention is associated with theta oscillations (Hermens et al., 2005; Keller et al., 2017). We speculate that the attention system or relevant networks modulate visual perception via theta oscillations.

In the occipital region, delta oscillatory intensity was higher during the proliferative phase than during the secretory phase. Enhanced delta oscillations in the caudal part of the brain are a typical feature of cognitive dysfunction and dementia (Fernández et al., 2013; Hughes et al., 2019). Although research has indicated that the menstrual cycle alters cognition (Sacher et al., 2013), these alterations are not the same as those observed in patients with dementia. Delta oscillations are associated with network activity rather than local activity (Bastos et al., 2012; Sauseng et al., 2015), and enhanced delta oscillation is associated with loss of cholinergic input from the nucleus basalis of Meynert (Lopes da Silva, 1991; Gratwicke et al., 2013; **Figure 1**), which contains oestrogen receptors (Ishunina and Swaab, 2001). Changes in the activity of the nucleus basalis of Meynert modulate delta intensity in the occipital region via cholinergic pathways.

Low gamma oscillatory intensity in the occipital region also differed between these two phases. Interestingly, cross-frequency coupling studies suggest that gamma oscillations are modulated by low-frequency oscillations, such as delta and theta (Florin and Baillet, 2015). Thus, alterations in gamma activity may occur secondary to changes in the delta and theta bands. Overall, the source-level analysis revealed that the menstrual cycle most strongly influences theta oscillatory intensity, and that this change originates in the limbic system and deep structures such as the thalamus.

In addition, the source-level analysis revealed a trend in which the rostral (frontal/parietal) and caudal (temporal/occipital) regions exhibited reversal changes in oscillatory intensity between the proliferative and secretory phases (**Figure 6**). In most of the frequency bands, oscillatory intensities in rostral regions during the proliferative phase were higher than those during the secretory phase, while the inverse relationship was observed in caudal regions. Similar results were observed in our

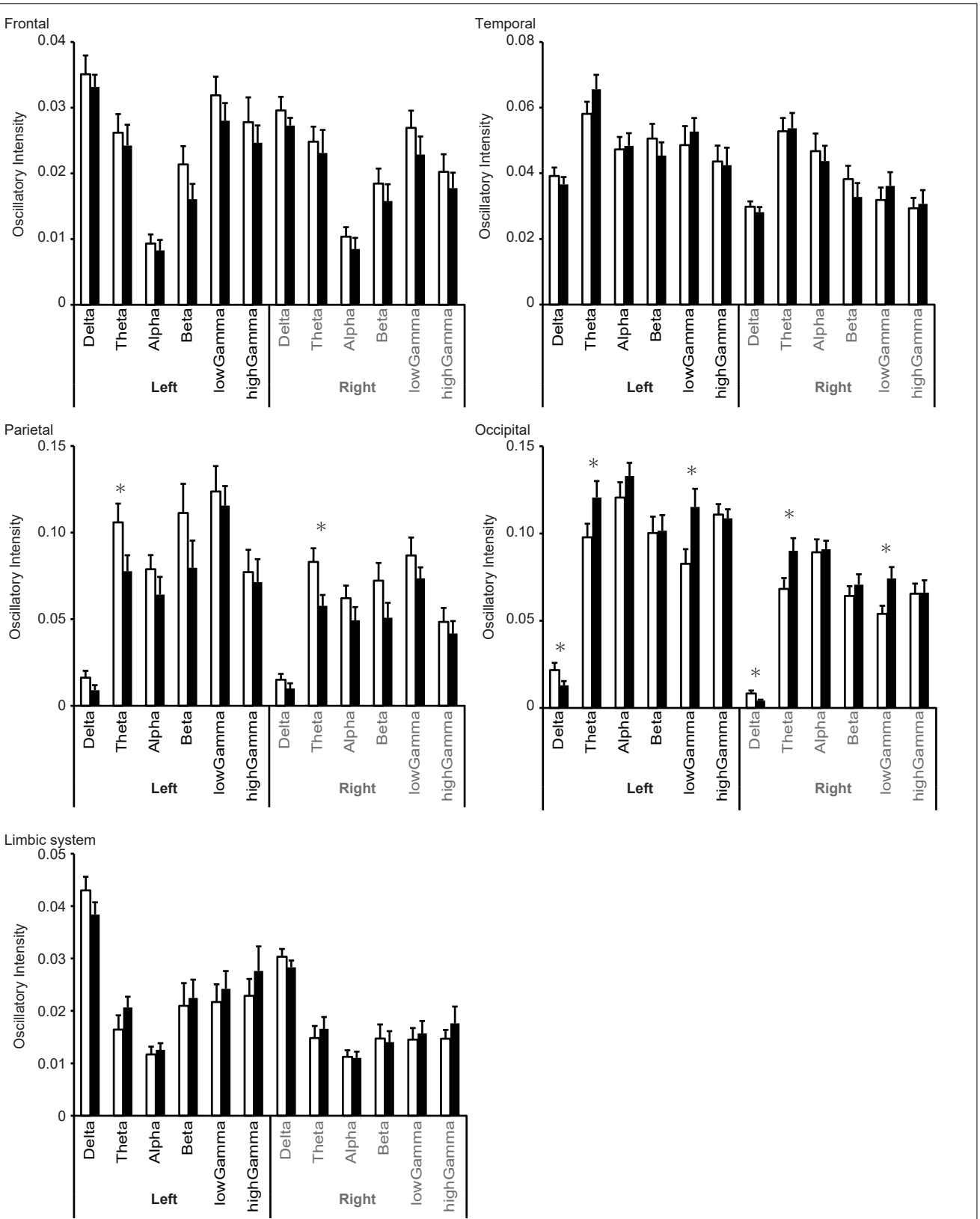

FIGURE 6 | Changes in regional oscillatory intensity during the proliferative (open columns) and secretory (filled columns) phases. Error bars represent the grand mean of standard errors over bootstrapped iterations, whereas asterisks (*) indicate statistically significant differences (after false-discovery rate (FDR) correction).

previous study regarding healthy ageing (Hoshi and Shigihara, 2020). In the previous study, we speculated that this inverse relationship may be explained by differences in the distribution of the dopaminergic system, which dominantly controls the rostral part of the brain (Alcaro et al., 2007). This may also explain the present result, as the dopaminergic system is affected by the menstrual cycle as well (Hidalgo-Lopez and Pletzer, 2017).

Limitations

The study has five main limitations. First, we did not measure basal body temperature or levels of hormones such as oestrogen and progesterone due to ethical reasons. These factors are important and essential for identifying the phase of the menstrual cycle. However, these measurements place a substantial burden on participants, and many of them hesitated to mention the details of their menstrual cycle even in the clinic. Although the relationship between resting-state neural activity and the menstrual cycle has been investigated in previous studies, most of the results were based on subjective observations of EEG data. Therefore, we computed quantitative measures of resting-state neural activity and conducted the present study with a minimum setting to simulate the same conditions as clinical practice: recording neural oscillations non-invasively for 5 min and asking the participants to report the first day of the last MP. Despite this limited information, our results suggest that the influence of the menstrual cycle should be taken into account when functional neuroimaging is required in general practice settings. Second, the order of measurements between the MP and OP was not well balanced. MEG data from eight of 25 participants were acquired during the MP first, while data were acquired during the OP first in the remaining participants. To remove this potential confounding factor, the two conditions should have included an equal number of participants. In addition, the number of participants in the proliferative and secretory phases was not well matched due to the same ethical reasons. To overcome the potential bias, we applied a sophisticated statistical method, although we recognise that this is not the ideal approach. Nevertheless, our results were significant, supporting the notion that the menstrual cycle should be included as a control variable in future studies. Third, participants were recruited from a limited population: clinical staff in Kumagaya General Hospital. Thus, some variables (e.g., educational levels) may not have been the same as those in the general population. However, this conservative approach allowed us to easily confirm that all participants were healthy (Hoshi and Shigihara, 2020), which is not always possible when participants are sampled from the general population. Fourth, the study design was not identical between Comparison 1 (i.e., within-participant design) and Comparison 2 (i.e., between-participant design). The present research was primarily designed to investigate differences in neural oscillatory activity between the MP and OP (Comparison 1), as discrimination between the two phases is frequently necessary in clinical practice. We undertook an additional analysis (Comparison 2) to explore the mechanisms underlying these alterations, which thereby required a different study design. Fifth, to minimise the burden on research participants, we did not assess cognitive or emotional function using neuropsychological tests. Such tests may have been helpful in interpreting alterations in resting-state cortical activity. Although these analyses were beyond the scope of the present study, it may be interesting to include these variables along with basal body temperature and hormone levels in future studies.

CONCLUSION

Resting-state neural oscillations are used as biomarkers for functional diseases, such as dementia (Poza et al., 2007; Hughes et al., 2019), epilepsy (Worrell et al., 2004; Lundstrom et al., 2019; Tanoue et al., 2021), and stroke (Sakamoto et al., 2016). However, accurate interpretation of clinical outcomes requires the identification and minimisation of potential confounding factors. Our findings suggest that spontaneous neural oscillations are affected by the menstrual cycle at both global and regional levels, as well as by age and sex. The changes elicited by the cycle are not limited to alpha rhythms, but also affect high- and low-frequency oscillations, particularly theta rhythms. The limbic system and parietal lobe play important roles in these changes in spontaneous neural oscillations. In conclusion, our findings suggest that the menstrual cycle should be considered to ensure accurate interpretation of functional neuroimaging in clinical practice, especially in the treatment of epilepsy and dementia. Future studies may wish to focus on the association between the frequency of epileptic seizures and the menstrual cycle (i.e., catamenial epilepsy) (Maguire and Nevitt, 2019) and on the interactions between the healthy menstrual cycle and dementia via hormonal systems (Moradi et al., 2018; Zhou et al., 2020).

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found below: Shigihara, Yoshihito, 2021, "Replication Data for: Healthy menstruation", https://doi.org/10.7910/DVN/H4O0RE, Harvard Dataverse, V1.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the Ethics Committee of Kumagaya General Hospital. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

YS managed and designed the research project. RH and SI recruited participants. RH, SI, and KF scanned and managed MEG data. HH and YS analysed the data. HH, JP, VR-G, CG, and YS wrote the manuscript. KN and MH supervised the project as gynaecologists. All authors contributed to the article and approved the submitted version.

FUNDING

This study was partially sponsored by RICOH Co., Ltd., the "Ministerio de Ciencia, Innovación y Universidades—Agencia Estatal de Investigación," and "European Regional Development Fund (FEDER)" under project PGC2018-098214-A-I00; and by "CIBER en Bioingeniería, Biomateriales y Nanomedicina (CIBER-BBN)" through "Instituto de Salud Carlos III" co-funded with FEDER funds. The sponsors had no role in the study concept, design, methods, data analysis, or preparation of the manuscript. Their financial support was used for English language editing service fee. VR-G was in receipt of a PIF-UVa grant from the University of Valladolid.

ACKNOWLEDGMENTS

We would like to thank Editage (www.editage.com) for providing English language editing services. Lastly, we thank our participants and affirm our genuine respect for their contributions to continued progress in the medical sciences.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fnhum.2021.652789/full#supplementary-material

REFERENCES

Ahmed, R., and Rutka, J. T. (2016). The role of MEG in pre-surgical evaluation of epilepsy: current use and future directions. *Exp. Rev. Neurother.* 16, 795–801. doi: 10.1080/14737175.2016.1181544

Albert, K., Pruessner, J., and Newhouse, P. (2015). Estradiol levels modulate brain activity and negative responses to psychosocial stress across the menstrual cycle. *Psychoneuroendocrinology* 59, 14–24. doi: 10.1016/j.psyneuen.2015.04.022

Alcaro, A., Huber, R., and Panksepp, J. (2007). Behavioral functions of the mesolimbic dopaminergic system: An affective neuroethological perspective. *Brain Res. Rev.* 56, 283–321. doi: 10.1016/j.brainresrev.2007.07.014

Barel, E., Krispil, M., and Yaari, I. (2019). Cognitive performance across the menstrual cycle. *J. Psychol. Cogn.* 4:47. doi: 10.35841/PSYCHOLOGY-COGNITION.4.3.41-47

Barry, R. J., and De Blasio, F. M. (2017). EEG differences between eyes-closed and eyes-open resting remain in healthy ageing. *Biol. Psychol.* 129, 293–304. doi: 10.1016/J.BIOPSYCHO.2017.09.010

Bartoli, E., Bosking, W., and Foster, B. L. (2020). Seeing Visual Gamma Oscillations in a New Light. *Trends Cogn. Sci.* 24, 501–503. doi: 10.1016/j.tics.2020.03.009

BaŞar, E. (2013). Brain oscillations and neuropsychiatry. *Dialogues Clin. Neurosci.* 15, 291–300.

Bastos, A., Userey, M., Adams, R., Mangun, G., Fries, P., and Friston, K. J. (2012). Canonical microcircuits for predictive coding. *Neuron* 76, 695–711. doi: 10.1016/j.neuron.2012.10.038.Canonical

Bayer, J., Schultz, H., Gamer, M., and Sommer, T. (2014). Menstrual-cycle dependent fluctuations in ovarian hormones affect emotional memory. *Neurobiol. Learn. Mem.* 110, 55–63. doi: 10.1016/j.nlm.2014.01.017

Bazanova, O. M., Kondratenko, A. V., Kuz'minova, O. I., Muravleva, K. B., and Petrova, S. E. (2014). EEG alpha indices depending on the menstrual cycle phase and salivary progesterone level. *Fiziol Cheloveka* 40, 31–40. doi: 10.1111/j.1365-2265.1986.tb03251.x

Becker, D., Creutzfeldt, O. D., Schwibbe, M., and Wuttke, W. (1982). Changes in physiological, eeg and psychological parameters in women during the spontaneous menstrual cycle and following oral contraceptives. *Psychoneuroendocrinology* 7, 75–90. doi: 10.1016/0306-4530(82)90057-9

Benjamini, Y., and Hochberg, Y. (1995). Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *J. R. Stat. Soc. Ser. B* 57, 289–300. doi: 10.2307/2346101

Brammer, M. (2009). The role of neuroimaging in diagnosis and personalized medicine-current position and likely future directions. *Dialog. Clin. Neurosci.* 11, 389–396. doi: 10.31887/dcns.2009.11.4/mbrammer

Brinton, R. D., Yao, J., Yin, F., Mack, W. J., and Cadenas, E. (2015). Perimenopause as a neurological transition state. *Nat. Rev. Endocrinol.* 11, 393–405. doi: 10.1038/nrendo.2015.82

Brötzner, C. P., Klimesch, W., Doppelmayr, M., Zauner, A., and Kerschbaum, H. H. (2014). Resting state alpha frequency is associated with menstrual cycle phase, estradiol and use of oral contraceptives. *Brain Res.* 1577, 36–44. doi: 10.1016/j.brainres.2014.06.034

Bull, J. R., Rowland, S. P., Scherwitzl, E. B., Scherwitzl, R., Danielsson, K. G., and Harper, J. (2019). Real-world menstrual cycle characteristics of more than 600,000 menstrual cycles. *npj Digit. Med.* 2:83. doi: 10.1038/s41746-019-0152-7

Caspers, S., and Zilles, K. (2018). Microarchitecture and connectivity of the parietal lobe. *Handb. Clin. Neurol.* 151, 53–72. doi: 10.1016/B978-0-444-63622-5.00003-6

Colom, L. V. (2006). Septal networks: relevance to theta rhythm, epilepsy and Alzheimer's disease. *J. Neurochem.* 96, 609–623. doi: 10.1111/j.1471-4159.2005.03630.x

Coslett, H. B., and Schwartz, M. F. (2018). The parietal lobe and language. *Handbook Clin. Neurol.* 2018, 365–375. doi: 10.1016/B978-0-444-63622-5.00018-8

Creutzfeldt, O. D., Arnold, P. M., Becker, D., Langenstein, S., Tirsch, W., Wilhelm, H., et al. (1976). EEG changes during spontaneous and controlled menstrual cycles and their correlation with psychological performance. *Electroencephalogr. Clin. Neurophysiol.* 40, 113–131. doi: 10.1016/0013-4694(76)90157-7

Cunningham, S. I., Tomasi, D., and Volkow, N. D. (2017). Structural and functional connectivity of the precuneus and thalamus to the default mode network. *Hum. Brain Mapp.* 38, 938–956. doi: 10.1002/hbm.23429 doi: 10.1002/hbm.23429

de Barenne, D. D., and Gibbs, F. A. (1942). Variations in the electroencephalogram during the menstrual cycle. *Am. J. Obstet. Gynecol.* 44, 687–690. doi: 10.1016/S0002-9378(42)90438-1

Dustman, R. E., Shearer, D. E., and Emmerson, R. Y. (1993). EEG and event-related potentials in normal aging. *Prog. Neurobiol.* 41, 369–401. doi: 10.1016/0301-0082(93)90005-D

Ezzati, A., Katz, M. J., Zammit, A. R., Lipton, M. L., Zimmerman, M. E., Sliwinski, M. J., et al. (2016). Differential association of left and right hippocampal volumes with verbal episodic and spatial memory in older adults. *Neuropsychologia* 93, 380–385. doi: 10.1016/j.neuropsychologia.2016.08.016

Farage, M. A., Osborn, T. W., and MacLean, A. B. (2008). Cognitive, sensory, and emotional changes associated with the menstrual cycle: A review. *Arch. Gynecol. Obstet.* 278, 299–307. doi: 10.1007/s00404-008-0708-2

Farrar, D., Neill, J., Scally, A., Tuffnell, D., and Marshall, K. (2015). Is objective and accurate cognitive assessment across the menstrual cycle possible? A feasibility study. *SAGE Open Med.* 3:205031211456519. doi: 10.1177/2050312114565198

Ferna, A., and Hornero, R. (2006). Quantitative Magnetoencephalography of Spontaneous Brain Activity in Alzheimer Disease. *Alzheimer Dis. Assoc. Disord.* 20, 153–159. doi: 10.1097/00002093-200607000-00006

Fernández, A., Turrero, A., Zuluaga, P., Gil-Gregorio, P., Del Pozo, F., Maestu, F., et al. (2013). MEG delta mapping along the healthy aging-Alzheimer's disease continuum: Diagnostic implications. *J. Alzheimer's Dis.* 35, 495–507. doi: 10.3233/JAD-121912

Florin, E., and Baillet, S. (2015). The brain's resting-state activity is shaped by synchronized cross-frequency coupling of neural oscillations. *Neuroimage* 111, 26–35. doi: 10.1016/j.neuroimage.2015.01.054

Franz, W. B. (1988). Basic review: endocrinology of the normal menstrual cycle. *Prim. Care-Clin. Off. Pract.* 15, 607–616. doi: 10.1016/s0095-4543(21)01279-3

Friston, K., Harrison, L., Daunizeau, J., Kiebel, S., Phillips, C., Trujillo-Barreto, N., et al. (2008). Multiple sparse priors for the M/EEG inverse problem. *Neuroimage* 39, 1104–1120. doi: 10.1016/j.neuroimage.2007.09.048

George, K., and Das, M. J. (2019). *Neuroanatomy, Thalamocortical Radiations*. Treasure Island, FL: StatPearls Publishing.

Gómez, C., Pérez-Macías, M., Poza, J., Fernández, A., and Hornero, R. (2013). Spectral changes in spontaneous MEG activity across the lifespan. *J. Neural. Eng.* 10:066006. doi: 10.1088/1741-2560/10/6/066006

Grandy, T. H., Werkle-Bergner, M., Chicherio, C., Lövdén, M., Schmiedek, F., and Lindenberger, U. (2013). Individual alpha peak frequency is related to latent factors of general cognitive abilities. *Neuroimage* 79, 10–18. doi: 10.1016/j.neuroimage.2013.04.059

Gratwicke, J., Kahan, J., Zrinzo, L., Hariz, M., Limousin, P., Foltynie, T., et al. (2013). The nucleus basalis of Meynert: a new target for deep brain stimulation in dementia? *Neurosci. Biobehav. Rev.* 37, 2676–2688. doi: 10.1016/j.neubiorev.2013.09.003

Gross, J., Baillet, S., Barnes, G. R., Henson, R. N., Hillebrand, A., Jensen, O., et al. (2013). Good practice for conducting and reporting MEG research. *Neuroimage* 65, 349–363. doi: 10.1016/j.neuroimage.2012.10.001

Hagemann, G., Ugur, T., Schleussner, E., Mentzel, H. J., Fitzek, C., Witte, O. W., et al. (2011). Changes in brain size during the menstrual cycle. *PLoS One* 6:e14655. doi: 10.1371/journal.pone.0014655

Hamid, H. (2014). *Networks in Mood and Anxiety Disorders in Neuronal Networks in Brain Function, CNS Disorders, and Therapeutics*. Cambridge, MA: Academic Press, 327–334. doi: 10.1016/B978-0-12-415804-7.00024-1

Hampson, E., Levy-Cooperman, N., and Korman, J. M. (2014). Estradiol and mental rotation: relation to dimensionality, difficulty, or angular disparity? *Horm. Behav.* 65, 238–248. doi: 10.1016/j.yhbeh.2013.12.016

Hatta, T., and Nagaya, K. (2009). Menstrual Cycle Phase Effects on Memory and Stroop Task Performance. *Arch. Sex. Behav.* 38, 821–827. doi: 10.1007/s10508-008-9445-7

Hawkins, S. M., and Matzuk, M. M. (2008). The menstrual cycle: Basic biology. *Ann. N. Y. Acad. Sci.* 1135, 10–18. doi: 10.1196/annals.1429.018

Hermens, D. F., Soei, E. X. C., Clarke, S. D., Kohn, M. R., Gordon, E., and Williams, L. M. (2005). Resting EEG theta activity predicts cognitive performance in attention-deficit hyperactivity disorder. *Pediatr. Neurol.* 32, 248–256. doi: 10.1016/j.pediatrneurol.2004.11.009

Hidalgo-Lopez, E., and Pletzer, B. (2017). Interactive effects of dopamine baseline levels and cycle phase on executive functions: The role of Progesterone. *Front. Neurosci.* 11:403. doi: 10.3389/fnins.2017.00403

Hidalgo-Lopez, E., Mueller, K., Harris, T., Aichhorn, M., Sacher, J., and Pletzer, B. (2020). Human menstrual cycle variation in subcortical functional brain connectivity: a multimodal analysis approach. *Brain Struct. Funct.* 225, 591–605. doi: 10.1007/s00429-019-02019-z

Hoshi, H., and Shigihara, Y. (2020). Age- and gender-specific characteristics of the resting-state brain activity: a magnetoencephalography study. *Aging* 12, 21613–21637. doi: 10.18632/aging.103956

Hughes, L. E., Henson, R. N., Pereda, E., Bruña, R., López-Sanz, D., Quinn, A. J., et al. (2019). Biomagnetic biomarkers for dementia: A pilot multicentre study with a recommended methodological framework for magnetoencephalography. *Alzheimer's Dement. Diag. Assess. Dis. Monit.* 11, 450–462. doi: 10.1016/j.dadm.2019.04.009

Ishunina, T. A., and Swaab, D. F. (2001). Increased expression of estrogen receptor α and β in the nucleus basalis of Meynert in Alzheimer's disease. *Neurobiol. Aging* 22, 417–426. doi: 10.1016/S0197-4580(00)00255-4

Kang, J. K., Lee, E. M., Kim, D., Hye Lee, S., Kim, T., Park, Y. M., et al. (2016). Asymmetric behaviours of brain oscillations in the human hippocampus during spatial navigation tasks. *Neuroreport* 27, 192–196. doi: 10.1097/WNR.0000000000000523

Keller, A. S., Payne, L., and Sekuler, R. (2017). Characterizing the roles of alpha and theta oscillations in multisensory attention. *Neuropsychologia* 99, 48–63. doi: 10.1016/j.neuropsychologia.2017.02.021

Khanna, N., Altmeyer, W., Zhuo, J., and Steven, A. (2015). Functional Neuroimaging: Fundamental Principles and Clinical Applications. *Neuroradiol. J.* 28, 87–96. doi: 10.1177/1971400915576311

Lewine, J. D., and Orrison, W. W. (1995). *Clinical Electroencephalography and Event-Related Potentialsin Functional Brain Imaging*. Maryland Heights: Mosby, 327–368. doi: 10.1016/b978-0-8151-6509-5.50012-6

Lindsley, D. B., and Rubenstein, B. B. (1937). Relationship Between Brain Potentials and Some Other Physiological Variables. *Proc. Soc. Exp. Biol. Med.* 35, 558–563. doi: 10.3181/00379727-35-9053

Lisofsky, N., Lindenberger, U., and Kühn, S. (2015). Amygdala/hippocampal activation during the menstrual cycle: Evidence for lateralization of effects across different tasks. *Neuropsychologia* 67, 55–62. doi: 10.1016/j.neuropsychologia.2014.12.005

Lopes da Silva, F. (1991). Neural mechanisms underlying brain waves: from neural membranes to networks. *Electroencephalogr. Clin. Neurophysiol.* 79, 81–93. doi: 10.1016/0013-4694(91)90044-5

López, M. E., Cuesta, P., Garcés, P., Castellanos, P. N., Aurtenetxe, S., Bajo, R., et al. (2014). MEG spectral analysis in subtypes of mild cognitive impairment. *Age* 36, 1095–1112. doi: 10.1007/s11357-014-9624-5

Luck, S. J., and Ford, M. A. (1998). On the role of selective attention in visual perception. *Proc. Natl. Acad. Sci. U. S. A.* 95, 825–830. doi: 10.1073/pnas.95.3.825

Lundstrom, B. N., Boly, M., Duckrow, R., Zaveri, H. P., and Blumenfeld, H. (2019). Slowing less than 1 Hz is decreased near the seizure onset zone. *Sci. Rep.* 9:6218. doi: 10.1038/s41598-019-42347-y

Lusk, B. R., Carr, A. R., Ranson, V. A., Bryant, R. A., and Felmingham, K. L. (2015). Early visual processing is enhanced in the midluteal phase of the menstrual cycle. *Psychoneuroendocrinology* 62, 343–351. doi: 10.1016/j.psyneuen.2015.08.022

Maguire, M. J., and Nevitt, S. J. (2019). Treatments for seizures in catamenial (Menstrual-related) epilepsy. *Cochrane Database Syst. Rev.* 2019, 13225. doi: 10.1002/14651858.CD013225.pub2

Mahon, B. Z., and Cantlon, J. F. (2011). The specialization of function: Cognitive and neural perspectives. *Cogn. Neuropsychol.* 28, 147–155. doi: 10.1080/02643294.2011.633504

Maller, J. J., Welton, T., Middione, M., Callaghan, F. M., Rosenfeld, J. V., and Grieve, S. M. (2019). Revealing the Hippocampal Connectome through Super-Resolution 1150-Direction Diffusion MRI. *Sci. Rep.* 9:2418. doi: 10.1038/s41598-018-37905-9

Mazaheri, A., Slagter, H. A., Thut, G., and Foxe, J. J. (2018). Orchestration of brain oscillations: principles and functions. *Eur. J. Neurosci.* 48, 2385–2388. doi: 10.1111/ejn.14189

Moradi, F., Sadatmahalleh, S. J., and Ziaei, S. (2018). The effect of hormone replacement therapy on cognitive function in postmenopausal women: An RCT. *Int. J. Reprod. Biomed.* 16, 767–774. doi: 10.18502/ijrm.v16i12.3682

Newhouse, P., and Dumas, J. (2015). Estrogen-cholinergic interactions: Implications for cognitive aging. *Horm. Behav.* 74, 173–185. doi: 10.1016/j.yhbeh.2015.06.022

Ohki, T., and Takei, Y. (2018). Neural mechanisms of mental schema: a triplet of delta, low beta/spindle and ripple oscillations. *Eur. J. Neurosci.* 48, 2416–2430. doi: 10.1111/ejn.13844

Pascual-Marqui, R. D. (2002). Standardized low-resolution brain electromagnetic tomography (sLORETA): technical details. *Methods Find. Exp. Clin. Pharmacol.* 24(Suppl. D), 5–12.

Pletzer, B., Harris, T. A., and Hidalgo-Lopez, E. (2018). Subcortical structural changes along the menstrual cycle: beyond the hippocampus. *Sci. Rep.* 8:16042. doi: 10.1038/s41598-018-34247-4

Pletzer, B., Harris, T. A., and Ortner, T. (2017). Sex and menstrual cycle influences on three aspects of attention. *Physiol. Behav.* 179, 384–390. doi: 10.1016/j.physbeh.2017.07.012

Pletzer, B., Harris, T. A., Scheuringer, A., and Hidalgo-Lopez, E. (2019). The cycling brain: menstrual cycle related fluctuations in hippocampal and fronto-striatal activation and connectivity during cognitive tasks. *Neuropsychopharmacology* 44, 1867–1875. doi: 10.1038/s41386-019-0435-3

Poromaa, I. S., and Gingnell, M. (2014). Menstrual cycle influence on cognitive function and emotion processing from a reproductive perspective. *Front. Neurosci.* 8:380. doi: 10.3389/fnins.2014.00380

Poza, J., Hornero, R., Abásolo, D., Fernández, A., and García, M. (2007). Extraction of spectral based measures from MEG background oscillations in Alzheimer's disease. *Med. Eng. Phys.* 29, 1073–1083. doi: 10.1016/j.medengphy.2006.11.006

Pratt, M., Goldstein, A., Levy, J., and Feldman, R. (2017). Maternal Depression Across the First Years of Life Impacts the Neural Basis of Empathy in Preadolescence. *J. Am. Acad. Child Adolesc. Psychiatry* 56, 20.e–29.e. doi: 10.1016/j.jaac.2016.10.012

Protopopescu, X., Butler, T., Pan, H., Root, J., Altemus, M., Polanecsky, M., et al. (2008). Hippocampal structural changes across the menstrual cycle. *Hippocampus* 18, 985–988. doi: 10.1002/hipo.20468

Ray, A., and Bowyer, S. (2010). Clinical applications of magnetoencephalography in epilepsy. *Ann. Indian Acad. Neurol.* 13, 14–22. doi: 10.4103/0972-2327.61271

Rodríguez-González, V., Gómez, C., Shigihara, Y., Hoshi, H., Revilla-Vallejo, M., Hornero, R., et al. (2020). Consistency of local activation parameters at sensor- And source-level in neural signals. *J. Neural Eng.* 17, 56020. doi: 10.1088/1741-2552/abb582

Rolls, E. T. (1989). *Functions of Neuronal Networks in the Hippocampus and Neocortex in Memory in Neural Models of Plasticity*. Cambridge, MA: Academic Press, 240–265. doi: 10.1016/b978-0-12-148955-7.50017-5

Sacher, J., Okon-Singer, H., and Villringer, A. (2013). Evidence from neuroimaging for the role of the menstrual cycle in the interplay of emotion and cognition. *Front. Hum. Neurosci.* 2013:374. doi: 10.3389/fnhum.2013.00374

Sakamoto, S., Ikeda, H., Tsuyuguchi, N., Uda, T., Okumura, E., Asakawa, T., et al. (2016). MEG Frequency Analysis Depicts the Impaired Neurophysiological Condition of Ischemic Brain. *PLoS One* 11:e0168588. doi: 10.1371/journal.pone.0168588

Sauseng, P., Conci, M., Wild, B., and Geyer, T. (2015). Predictive coding in visual search as revealed by cross-frequency EEG phase synchronization. *Front. Psychol.* 6:1655. doi: 10.3389/fpsyg.2015.01655

Schöning, S., Engelien, A., Kugel, H., Schäfer, S., Schiffbauer, H., Zwitserlood, P., et al. (2007). Functional anatomy of visuo-spatial working memory during mental rotation is influenced by sex, menstrual cycle, and sex steroid hormones. *Neuropsychologia* 45, 3203–3214. doi: 10.1016/j.neuropsychologia.2007.06.011

Schwarzkopf, D. S., Robertson, D. J., Song, C., Barnes, G. R., and Rees, G. (2012). The frequency of visually induced γ-band oscillations depends on the size of early human visual cortex. *J. Neurosci.* 32, 1507–1512. doi: 10.1523/JNEUROSCI.4771-11.2012

Shigihara, Y., Hoshi, H., Poza, J., Rodríguez-González, V., Gómez, C., Kanzawa, T., et al. (2020a). Predicting the outcome of non-pharmacological treatment for patients with dementia-related mild cognitive impairment. *Aging* 12, 24101–24116. doi: 10.18632/aging.202270

Shigihara, Y., Hoshi, H., Shinada, K., Okada, T., and Kamada, H. (2020b). Non-pharmacological Treatment Changes Brain Activity in Patients With Dementia. *Sci. Rep.* 10:6744. doi: 10.1038/S41598-020-63881-0

Shughrue, P. J., and Merchenthaler, I. (2000). Estrogen is more than just a "sex hormone": novel sites for estrogen action in the hippocampus and cerebral cortex. *Front. Neuroendocrinol.* 21, 95–101. doi: 10.1006/frne.1999.0190

Šimič, N., and Santini, M. (2012). Verbal and spatial functions during different phases of the menstrual cycle. *Psychiatr. Danub.* 24, 73–79.

Sommer, B. (1992). Cognitive Performance and the Menstrual Cycle. *J. Psychol. Cogn.* 1992, 39–66. doi: 10.1007/978-1-4613-9148-7_2

Sundström-Poromaa, I. (2018). The Menstrual Cycle Influences Emotion but Has Limited Effect on Cognitive Function in Vitamins and Hormones. *Vitam Horm* 107, 349–376. doi: 10.1016/bs.vh.2018.01.016

Tallon-Baudry, C. (2009). The roles of gamma-band oscillatory synchrony in human visual cognition. *Front. Biosci.* 14, 321–332. doi: 10.2741/3246

Tanoue, Y., Uda, T., Hoshi, H., Shigihara, Y., Kawashima, T., Nakajo, K., et al. (2021). Specific Oscillatory Power Changes and Their Efficacy for Determining Laterality in Mesial Temporal Lobe Epilepsy: A Magnetoencephalographic Study. *Front. Neurol.* 12:617291. doi: 10.3389/fneur.2021.617291

Terakawa, Y., Tsuyuguchi, N., Tanaka, H., Shigihara, Y., Sakamoto, S., Takami, T., et al. (2008). Quantitative analysis of MEG using modified sLORETA for clinical application. *Clin. Neurophysiol.* 119, 1917–1922. doi: 10.1016/j.clinph.2008.04.008

Vysata, O., Kukal, J., Prochazka, A., Pazdera, L., and Valis, M. (2012). Age-Related Changes in the Energy and Spectral Composition of EEG. *Neurophysiology* 44, 63–67. doi: 10.1007/s11062-012-9268-y

Ward, M. M., Stone, S. C., and Sandman, C. A. (1978). Visual perception in women during the menstrual cycle. *Physiol. Behav.* 20, 239–243. doi: 10.1016/0031-9384(78)90215-9

Weis, S., Hodgetts, S., and Hausmann, M. (2019). Sex differences and menstrual cycle effects in cognitive and sensory resting state networks. *Brain Cogn.* 131, 66–73. doi: 10.1016/J.BANDC.2017.09.003

Wilson, C. L., Isokawa, M., Babb, T. L., and Crandall, P. H. (1990). Functional connections in the human temporal lobe. *Exp. Brain Res.* 82, 279–292. doi: 10.1007/BF00231248

Woolley, C. S., Gould, E., Frankfurt, M., and McEwen, B. S. (1990). Naturally occurring fluctuation in dendritic spine density on adult hippocampal pyramidal neurons. *J. Neurosci.* 10, 4035–4039. doi: 10.1523/JNEUROSCI.10-12-04035.1990

Worrell, G. A., Parish, L., Cranstoun, S. D., Jonas, R., Baltuch, G., and Litt, B. (2004). High-frequency oscillations and seizure generation in neocortical epilepsy. *Brain* 127, 1496–1506. doi: 10.1093/brain/awh149

Zhou, C. C., Wu, Q., Wang, Z., Wang, Q., Liang, Y., and Liu, S. (2020). The Effect of Hormone Replacement Therapy on Cognitive Function in Female Patients With Alzheimer's Disease: A Meta-Analysis. *Am. J. Alzheimers. Dis. Other Demen.* 35:153331752093858. doi: 10.1177/1533317520938585

Conflict of Interest: YS leads the joint research projects that were supported by RICOH Co., Ltd. HH was employed by RICOH Co., Ltd. VR-G received a PIF-UVa grant from the University of Valladolid.

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.

frontiers in Human Neuroscience

ORIGINAL RESEARCH
published: 26 July 2021
doi: 10.3389/fnhum.2021.685530

Are You Able to Trust Me? Analysis of the Relationships Between Personality Traits and the Assessment of Attractiveness and Trust

Bernadetta Bartosik[1], *Grzegorz M. Wojcik*[1]*, *Aneta Brzezicka*[2] and *Andrzej Kawiak*[1,3]

[1] Chair of Neuroinformatics and Biomedical Engineering, Institute of Computer Science, Maria Curie-Sklodowska University, Lublin, Poland, [2] Neurocognitive Research Center, SWPS University of Social Sciences and Humanities, Warsaw, Poland, [3] Polish-Japanese Academy of Information Technology, Warsaw, Poland

Behavioral and neuroimaging studies show that people trust and collaborate with others based on a quick assessment of the facial appearance. Based on the morphological characteristics of the face, i.e., features, shape, or color, it is possible to determine health, attractiveness, trust, and some personality traits. The study attempts to indicate the features influencing the perception of attractiveness and trust. In order to select individual factors, a model of backward stepwise logistic regression was used, analyzing the results of the psychological tests and the attractiveness and trust survey. Statistical analysis made it possible to select the most important personality traits related to attractiveness and trust assessments.

Keywords: trust and distrust, trust and reputation management, credibility, regress algorithm, machine learning

OPEN ACCESS

Edited by:
Filippo Brighina, University of Palermo, Italy

Reviewed by:
Stefano I. Di Domenico, Australian Catholic University, Australia
Davide Valeriani, Neurable Inc., United States

***Correspondence:**
Grzegorz M. Wojcik
gmwojcik@live.umcs.edu.pl

Specialty section:
This article was submitted to Brain Imaging and Stimulation, a section of the journal Frontiers in Human Neuroscience

Received: *25 March 2021*
Accepted: *16 June 2021*
Published: *26 July 2021*

Citation:
Bartosik B, Wojcik GM, Brzezicka A and Kawiak A (2021) Are You Able to Trust Me? Analysis of the Relationships Between Personality Traits and the Assessment of Attractiveness and Trust. Front. Hum. Neurosci. 15:685530. doi: 10.3389/fnhum.2021.685530

1. INTRODUCTION

The face is like a book that allows you to obtain information that is an important element of social communication. Watching the faces of strangers, people make social (attractiveness, credibility, intelligence, dominance) and personality (gender, age, emotions) assessments based on facial readings (Kościński, 2007; Oosterhof and Todorov, 2008, 2009; Zebrowitz and Montepare, 2008). The information obtained is particularly important in the process of analyzing the degree of credibility (Wierzbicki, 2008). It has been noted that attractive people are more often seen as trustworthy (Shinners, 2009). In addition, credibility assessment is coupled with expressed emotions. Happy and smiling faces are more credible in contrast to sad or angry faces (Sutherland et al., 2017). Scientists have shown that women and people with children's facial features have higher trust (Buchan et al., 2008; Zebrowitz et al., 2015).

It is believed that the ability to recognize faces develops automatically from an early age and is improved with development (De Heering et al., 2012; Jessen and Grossmann, 2019; Mondloch et al., 2019). Depending on the circumstances, the human brain can detect faces in just over 100 ms (Crouzet et al., 2010; Martin et al., 2018). This is important in the situations requiring an immediate decision. It is worth adding that the first impressions regarding the seen face may appear about 33 ms after the stimulus exposure (Bar et al., 2006; Freeman et al., 2014). Recently, scientists have presented views on the stage processing of the face. Visual features related to facial recognition (among others, gender and age) are recognized the fastest, followed by identity identification (di Oleggio Castello and Gobbini, 2015; Dobs et al., 2019).

Facial credibility has been proven to be essential to the trust that influences cooperation (Zebrowitz and Montepare, 2008). Murderers whose faces are trustworthy are more likely to get milder punishments (Wilson and Rule, 2015; Ancāns and Austers, 2018). Based on the assessment of the credibility of the face, scientists are able to predict the results of political choices (Ballew and Todorov, 2007). Face perception also has a big impact on online sales, when buyers are more likely to choose the offer of a seller with a more reliable face, regardless of the issued reviews (Ert et al., 2016). Similarly in the "trust games," people are more likely to spend money on a trustworthy partner, and the amount of the stake depends on the level of trust (Van't Wout and Sanfey, 2008; Chang et al., 2010).

Perception of socially important stimuli relies on the temporal cortical areas of the temporal lobe, and their connection with emotions and motivation is provided, among others, by the amygdala, orbitofrontal cortex, or cingulate cortex (Adolphs, 2001). Numerous brain tests have shown that the amygdala is one of the most important regions during the credibility assessment process (Engell et al., 2007; Todorov et al., 2008). The amygdala activity is associated with the processing of lower level emotional stimuli. It increases its activity during social assessment based on the emotional state and intentions of others (Costafreda et al., 2008). Depending on the degree of credibility of the face, the amygdala is more or less activated, with reliably looking faces reducing activity, and a decrease in credibility causing an increase in its activity (Haas et al., 2015).

Appearance is one of numerous characteristics that can influence the initiation or continuation of a relationship with another person. It can also affect the level of trust in the other person. As a matter of fact, there are so many canons of beauty and criteria for choosing in people. The aim of the research was to show the relationship between the personality traits and the assessment of trust and attractiveness toward the faces shown in the photos. For this purpose, there were carried out a survey to examine whether the faces of the people in the photos inspire trust and whether they are attractive as well as two psychological tests: the IVE Impulsivity Questionnaire and the NEO PI-R Personality Inventory. Both tests are used to diagnose personality traits. Using IVE you can define three features (impulsiveness, risk-aversion, empathy), and through NEO PI-R five (neuroticism, extrovertness, openness to experience, agreeableness, conscientiousness), each of these features has six more elements. After applying the stepwise backward logistic regression model, only those features that had the most significant impact on the dependent variables, i.e., attractiveness and trust, were selected. To our knowledge, this is one of the first EEG protocols planned on this subject. The current study is a pilot for further research using EEG.

2. TOOLS

Personality traits were examined using the NEO PI-R personality questionnaire by the authors of McCrae and Costa, in the Polish translation of Siuta. The NEO PI-R questionnaire is a test modeled on the five-factor personality model (Big Five), which takes into account five main dimensions (Costa and McCrae, 1992):

- neuroticism—a dimension defined by fear, guilt, dissatisfaction, anger (high neuroticism). Susceptibility to negative feelings causes weaker control over emotions, increases stress, and leads to illogical behavior. Low neuroticism characterizes emotionally stable, calm, and composed people.
- extroversion—it defines people prone to social interactions, able to feel positive emotions, active, and energetic. Extroverts are friendly to others, talkative, focused on searching for new stimuli. The opposite and at the same time the opposite end of the scale of this dimension is introversion, which characterizes less daring, more secretive people and avoiding such active social contacts.
- openness to experience—expresses a tendency to look for new life experiences. People with high openness are curious about the world, more tolerant and easily take on new tasks.
- agreeableness—defined as an attitude toward other people. A high level of agreeableness is characterized by trust in others, honesty, and a disinterested willingness to help. Agreeableness in a pejorative version reflects egocentrism, aggression, and dry relationships with other people.
- conscientiousness—described by such values as conscientiousness, punctuality, and diligence. High conscientiousness is primarily duty, goal-oriented action, high motivation, but also perfectionism or excessive dedication to work. On the other side of the conscientiousness scale, there are no defined life goals, low motivation to act and spontaneity as well as impulse decision-making.

The first personality inventory consisted of three factors, each of which had six subscales. On the basis of numerous observations, the model was extended with two further features and only modified in the following years. The current test consists of the following trait factors: neuroticism, extroversion, openness, agreeableness, and conscientiousness. Each of them is divided into six subscales. The worksheet consists of 240 questions. The respondent's task is to answer questions on a scale of 0–4 depending on how true the question is for the participant.

The Impulsiveness Questionnaire (IVE) by Hans J. Eysenck and Sybil B. G. Eysenck is a test used to diagnose personality traits in adults and high school students. It consists of three scales:

- the scale of impulsivity characteristic of people who make decisions without thinking about their effect,
- the scale of empathy, describing people sensitive to other people's emotions along with an adequate action,
- the scale of propensity to risk typical of people willing to take on new challenges.

The sheet contains 54 questions to which the respondent answers yes by marking "YES" or marking "NO" in the negative. The result for individual scales is the sum of points scored for the answers belonging to them (Caci et al., 2003).

From the available databases, two relatively large and generally available resources were selected, which are intended for the

development of science. Both sets are characterized by good resolution and accuracy of the taken photos. Both databases are designed to provide users with standardized and multi-aspect-tested photo sets that have been used in other facial processing studies (Silver et al., 2020; Assem et al., 2021). The first is the Development Emotional Faces Stimulus Set (DEFSS) which contains the total of 404 face pictures showing different emotions such as sadness, happiness, fear, anger, and neutral facial expressions. The models were people aged 8–30, mostly white. An additional advantage is the verification of photos made by the creators, during which independent respondents and photographed people assessed the presented emotions (Meuwissen et al., 2017). The second is the Multi-Racial Mega-Resolution (MR2) which shows the photos of 74 people between the ages of 18 and 25. Contrary to the previous database, it presents the photos of various races people (European, African, and East Asian), with only a neutral facial expression without makeup (Strohminger et al., 2016).

One hundred photos (50 women and 50 men) were selected from the two sets. Only photographs showing the face from the front without emotions (with a neutral expression on the face) were taken into account during the selection. Overall, 49 photos show the people of European descent, 31 photos the people of African descent, and 20 photos those of East Asian descent. A questionnaire was made for the obtained base, in which three questions were displayed for each photo. The first question concerned the gender of the person in the photo and the respondent chose a woman and a man from the answers. The second and third questions were about attractiveness and trust, respectively. Using a five-point scale, the participants determined to what extent the person in the photo is attractive and to what extent they can trust the person in the photo, where in both questions 1 meant not at all, and 5 very much. Eighty-five students of computer science and cognitive science at Maria Curie-Skłodowska University were invited to participate in the survey, and their answers were statistically analyzed. The photos are divided into four groups: attractive and trustworthy, unattractive and untrustworthy, attractive and untrustworthy, unattractive and trustworthy. Within these groups, based on the division by sex and origin of the people in the photos, six photos were selected for each group (**Figure 1**), rated the highest by the respondents. As a result, 24 photos were selected for future study.

3. PROCEDURE

As part of additional activity in the classroom, 85 students of Maria Curie-Skłodowska University enrolled in the research. The participants were mainly first-year students of cognitive science and first-year computer science, aged 18–24. Taking care of the confidentiality of personal data and the comfort of the participants, the laboratory employees generated random logins and passwords, which were used by the students to identify themselves while completing the questionnaires ensuring their anonymity.

The study consists of two psychological tests and a face survey. The first test in each case was the NEO PI-R. The test consisted of 240 questions, the participant assigned an answer to each question from 0 to 4 depending on how much he agreed with the statement. The next test was IVE consisting of 54 questions with the possibility of answering "yes" or "no." The last one was a questionnaire containing 100 photos of faces. For each photo, three questions were displayed in turn: "What is the gender of the person in the photo," "How attractive is the person in the photo," "To what extent are you able to trust the person in the photo." The participant chose the answer to the first question, a woman or a man, and chose the next two on a scale from 1 to 5 where 1 meant not at all and 5 very much.

4. DESCRIPTION OF THE STATISTICAL METHOD

The statistical analysis was performed using the logistic regression model in the SPSS program. It is a type of nonlinear analysis that allows you to describe the direction and strength of the relationship between individual explanatory variables in quantitative or qualitative form and dichotomous dependent variables that assume values of 0 or 1. The model was composed of 34 explanatory variables which include of personality traits tested with the NEO PI-R and IVE psychological tests. Due to the characteristics of logistic regression and the desire to extract the variables as well as possible,the dependent variables of the model are attractiveness and trust represented by a dichotomous variable where 1 means that the respondent assessed the face from the photo as attractive/trustworthy, 0 means that the respondent assessed the face in the photo negatively. The data set introduced into the model did not include the division into training and test sets. In addition, it is worth mentioning that the presented study is a pilot for further research in which ERP will be tested using EEG in the context of trust/not trust.

The designed backward stepwise logistic regression model took into account additionally the Wald criterion to optimize the number of variables influencing the dependent variables. The higher Wald's coefficient, the more influencing on attractiveness and trust the variable is. A stepwise regression model was used, which means that the model gradually changed. In this case, along with the next step, one variable with the lowest value of Wald's criterion was rejected from the model and further statistical activities were carried out on the remaining data set (**Table 1**). The number of explanatory variables decreased to 5 in the attractiveness model which obtained satisfactory results in 30 steps (Nagellerke's $R2 = 0.261$) and to 9 in the confidence model with the number of steps equal to 26 (Nagellerke's $R2 = 0.434$).

5. CORRECTNESS OF THE LOGISTIC MODEL

Principal Component Analysis (PCA) is one of the statistical methods that examines data without supervision. The main goal is to reduce the number of input variables describing the phenomenon under study. By design, this method is used to explain the variability of complex data using new principal components that are linear combinations of the observed

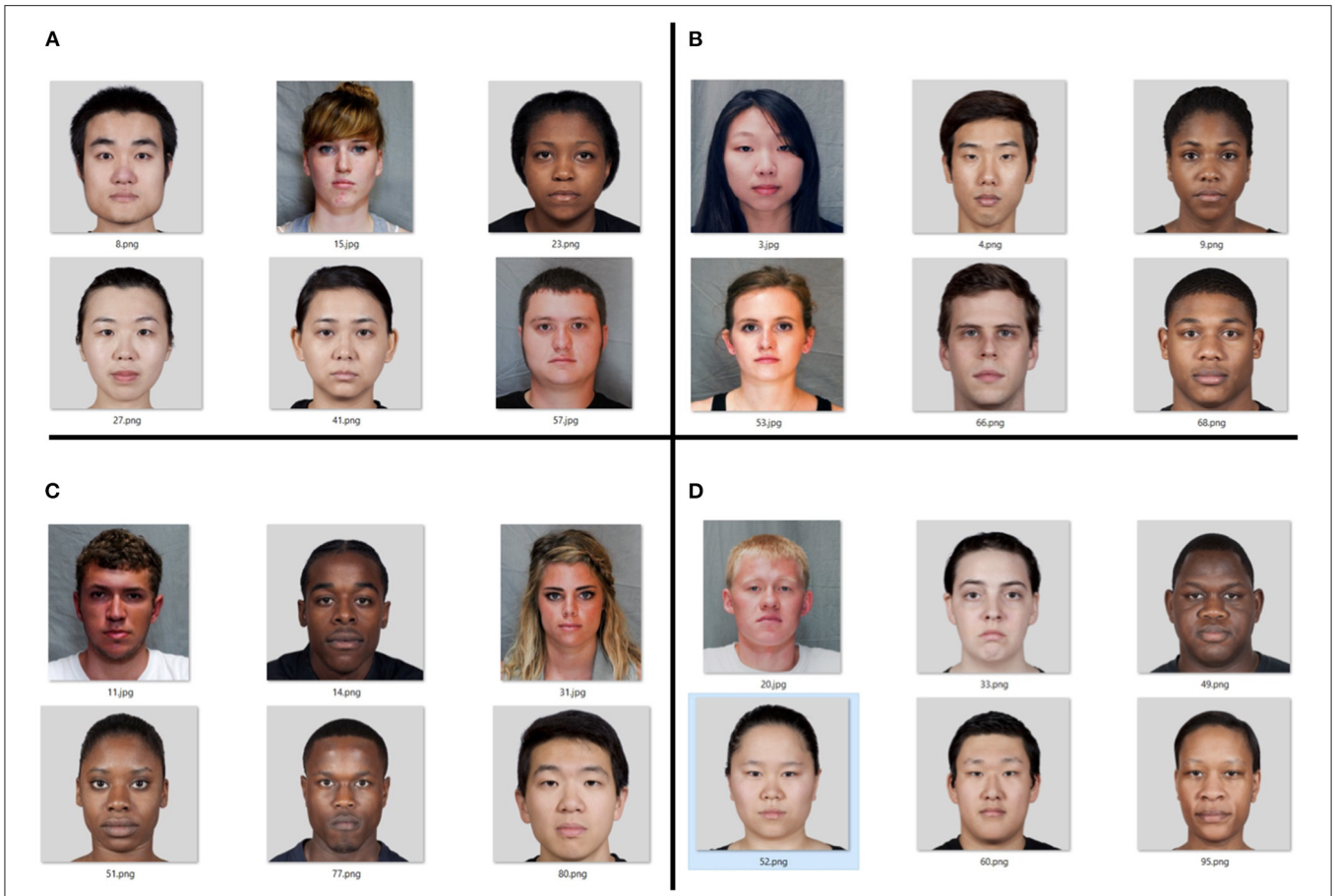

FIGURE 1 | A set of photos selected for future research, divided into groups: **(A)** low attractiveness and high confidence, **(B)** high attractiveness and high confidence, **(C)** high attractiveness and low confidence, **(D)** low attractiveness and low trust. Photos taken from the following databases: DEFSS (Meuwissen et al., 2017) and MR2 (Strohminger et al., 2016).

variables. The new set is characterized basing on most of the data from the original set with a reduced number of variables. The PCA analysis can be an introduction before proceeding to further statistical methods, such as cluster or discrimination analysis.

In the study, the backward stepwise logistic regression eliminated, based on the Wald coefficient, the least significant explanatory variables for the model. According to the purpose of PCA, calculations were performed and the number of variables necessary to describe the phenomenon was determined and compared with the number of variables determined in the backward stepwise logistic regression model. For the calculations, a correlation matrix containing 34 explanatory variables and a constant was used. The operations were performed with the use of functions available in the programming language "R." The PCA analysis was used to check that the logistic regression model uses a sufficient number of describing variables. **Tables 2**, **3** provide the PCA statistics for attractiveness and trust.

On the basis of the Kaiser criterion, which recommends distinguishing only those factors whose eigenvalues are >1, PCA selected five components for attractiveness and seven for trust. The components for the variable "attractiveness" explain more than 90% of the variability of the input data (first component: 43.6%, second component: 21.6%, third component: 9.9%, fourth component: 6.6%, fifth component: 3%) while the components for the variable "trust" in about 87% (first component: 34%, second component: 22%, third component: 10%, fourth component: 9%, fifth component: 5%, sixth component: 4%, seventh component: 3%). The above statistical data indicate a sufficient number of components that can be used to describe the studied phenomenon. The logistic regression model, as a result of the stepwise elimination of the least statistically significant features, left five predictors for attractiveness and nine predictors for confidence. Comparing the number of predictors necessary to describe the phenomenon determined by the PCA with the number of predictors left as a result of the logistic regression model calculations, it is concluded that the model left a minimum and sufficient number of predictors to describe the attractiveness while for the description of confidence it distinguished two additional predictors above the minimum necessary.

The quality of the model was assessed using the Hosmer-Lemeshow test (**Table 4**). In both cases the obtained results

TABLE 1 | Summary of the most important explanatory variables (marked with a "+" in the table) for the variables attractiveness and trust, which remained after elimination by backward logistic stepwise regression based on the Wald coefficient.

Variable	Attractiveness	Trust
N1- Anxiety	–	–
N2- Angry hostility	+	–
N3- Depression	–	–
N4- Self-consciousness	+	–
N5- Impulsiveness	+	–
N6- Vulnerability	–	–
E1- Warmth	–	–
E2- Gregariousness	–	–
E3- Assertiveness	–	–
E4- Activity	+	–
E5- Excitement seeking	–	–
E6- Positive emotions	–	–
O1- Fantasy	–	–
O2- Aesthetics	–	–
O3- Feelings	–	–
O4- Actions	+	–
O5- Ideas	–	+
O6- Values	–	–
U1- Trust	+	–
U2- Straightforwardness	–	–
U3- Altruism	+	+
U4- Compliance	–	–
U5- Modesty	–	–
U6- Tendermindedness	–	–
S1- Competence	–	–
S2- Order	–	+
S3- Dutifulness	–	–
S4- Achievement striving	–	–
S5- Self-discipline	–	–
S6- Deliberation	–	–
Empathy	+	+
sdr-Tendency to take risks	–	+
Impulsiveness	+	–
constant	+	+

TABLE 2 | Principal component analysis of a logistic regression model based on the correlation matrix for the attractiveness predictor.

	Standard deviation	Proportion of variance	Cumulative proportion
Comp1	4.167	0.496	0.496
Comp2	2.752	0.216	0.712
Comp3	1.858	0.099	0.811
Comp4	1.515	0.066	0.877
Comp5	1.027	0.030	0.907
Comp6	0.968	0.027	0.933
Comp7	0.874	0.022	0.955
Comp8	0.744	0.016	0.971
Comp9	0.470	0.006	0.977
Comp10	0.454	0.006	0.983

TABLE 3 | Principal component analysis of a logistic regression model based on the correlation matrix for the trust predictor.

	Standard deviation	Proportion of variance	Cumulative proportion
Comp1	3.466	0.343	0.343
Comp2	2.783	0.222	0.565
Comp3	1.914	0.105	0.670
Comp4	1.765	0.088	0.758
Comp5	1.298	0.048	0.806
Comp6	1.171	0.039	0.845
Comp7	1.050	0.031	0.877
Comp8	0.897	0.023	0.900
Comp9	0.886	0.022	0.922
Comp10	0.740	0.016	0.938

TABLE 4 | Goodness of fit test results from Hosmer and Lemeshow.

	Step	Chi-square	df	Relevance
Attractiveness	1	7.827	8	0.451
	30	9.075	8	0.336
Trust	1	9.612	7	0.212
	26	7.411	7	0.387

showed no significance which proves the similarity of the observed and expected values and a good fit of the model.

6. RESULTS

As a result of the analysis of the obtained data, the impact of individual descriptive variables on dependent variables was assessed. For a better understanding of the results of the regression model, a table of correlation between the dimensions of the NEO-Pi-R questionnaire and the subscales of these dimensions is presented below. **Table 5** shows only high correlations between dimensions and their subscales. The data show that the subscales do not correlate significantly with each other.

The model uses a default cutoff of 0.5. The case classification derived from the logistic regression uses the predicted probability. The case with the predicted probability greater than the cutoff value is classified as positive (1), and less, as negative (0). Based on **Tables 6**, **7**, which compare the classification of the tested values with respect to the dependent variable with the classification resulting from the use of the model, it is concluded that the model in which the dependent variable is "attractiveness" classifies correctly 72.6% of the data. In the model where the dependent variable is "trust," the total of 78.8% of correctly predicted responses was recorded.

TABLE 5 | Correlation between the dimensions of the NEO-Pi-R questionnaire and their subscales.

	Neuroticism	Extroversion	Openness	Agreeableness	Conscientiousness
N1	0.780				
E1		0.748			
S1					0.711
E2		0.877			
O2			0.742		
U2				0.739	
S2					0.711
N3	0.772				
E3		0.761			
U3				0.704	
S3					0.801
N4	0.721				
E4		0.815			
U4				0.835	
E5		0.740			
S5					0.715
N6	0.759				

TABLE 6 | Percentage of correct classifications in the training data for the dependent variable "attractiveness".

Observed			Predicted		
			Attractiveness		Percentage of correct classifications
			0	1	
Step 1	Attractiveness	0	48	4	92.3
		1	5	27	84.4
	Total percentage				89.3
Step 30	Attractiveness	0	45	7	86.5
		1	16	16	50
	Total percentage				72.6

TABLE 7 | Percentage of correct classifications in the training data for the dependent variable "trust."

Observed			Predicted		
			Trust		Percentage of correct classifications
			0	1	
Step 1	Trust	0	33	7	82.5
		1	5	40	88.9
	Total percentage				85.9
Step 26	Trust	0	30	10	75
		1	8	37	82.2
	Total percentage				78.8

In order to check how the model behaves in the event of data reduction, the dimensions of the NEO-PI-R questionnaire were omitted, leaving only the subscales of dimensions in the model.

TABLE 8 | Percentage of correct classifications in the training data for the dependent variable "trust."

Observed		Predicted		
		Trust		Percentage of correct classifications
		0	1	
Trust	0	30	10	75.0
	1	8	37	82.2
Total percentage				78.8

TABLE 9 | Percentage of correct classifications in the training data for the dependent variable "attractiveness".

Observed		Predicted		
		Attractiveness		Percentage of correct classifications
		0	1	
Attractiveness	0	46	6	88.5
	1	9	23	71.9
Total percentage				82.1

The above tables (**Tables 8**, **9**) show that the model limitation contributed to the improvement of the model quality in the case of the dependent variable "attractiveness," thus giving 82.1% of correct classifications. The changes in the model did not improve the classification for the dependent variable "trust." The interpretation of the classification accuracy of 70% + may be misleading, although only the model from which the individual data is predicted was presented in the studies. The study is only a pilot and will be used for further analysis in EEG studies, the

current indicators will be auxiliary indicators, while the main indicators will be indicators taken from the EEG.

The values of the parameters of the regression model are presented in **Table 10**. As a result of the reduction in the number of variables increasing with each successive regression step, several most influential components of the model were obtained. For better illustration of the influence of individual features on the assessment of attractiveness and trust, the data are presented in **Table 11**. Additionally, the features influencing significantly both predictors simultaneously were marked.

Among the variables for the assessment of attractiveness, the most important is altruism (U3), which is the agreeableness subscale in the NEO PI-R test. According to the data, as the variable increases, the probability that the face in the photo will be assessed as attractive increases over 2.5 times. This variable is also the most important predictor when trust is the dependent variable in the model. In this case the probability of trust grows also more than 2.5 times with an increase in the value of the predictor. Another factor that influences the assessed attractiveness significantly is order (S2, component of conscientiousness in the NEO PI-R test). The increase in the variable enhances the chance of a positive visual assessment of the person in the photo by about 50%. In the case of empathy, which is a variable derived from the IVE questionnaire, it follows that more emphatic people are prone to negative assessment of attractiveness. The most similar results were obtained for the variables O5 (ideas, openness subscale) and sdr (propensity to risk from the IVE questionnaire). In both cases, an increase in the value of the variable leads to an increase in the chances that the respondent will evaluate the person in the photo as attractive.

As mentioned before for the dependent variable trust, a particularly strong predictor is the variable U3 or altruism. In addition to this variable, the N2 component, in other words, angry hostility from the NEO PI-R test, described as a tendency to irritation or anger, has a large impact. As follows from the data the stronger the personality trait, the greater the probability of trusting the person in the photo. Excessive self-consciousness (N4) is another variable that plays a significant role in the model. According to the presented statistics, the chances of trust by people manifesting social anxiety or low self-esteem drop by half. The group describing the dependent variable also included a feature of the same name, i.e., trust (U1) from the NEO PI-R test. When the variable changes, the probability that the face in the photo is judged trustworthy doubles. Very similar results were obtained for the variable O4 (actions from the NEO PI-R test) and impulsiveness from the IVE test. In both cases, the probability is doubled. The variable E4 (activity) has a negative component which means that its increase results in a reduction of the probability of trust by ∼52%. The same relationship is found empathy from the IVE test and for the E5 variable (excitement seeking). The decrease in odds is ∼44 and ∼45%, respectively.

7. DISCUSSION

The subject of personality and its influence on human behavior has been of interest for researchers for many years. Müller and Schwieren (2020) examined the significance of individual personality types on the behavior of participants during a trust game. They showed that personality influences human behavior based on trust with higher correlations with ambiguous decisions than with risky decisions. They indicated that neuroticists spend lower stakes during the game while people characterized by agreeableness are inclined to donate higher amounts. This statement can be translated into trust, i.e., agreeable people have a higher level of trust toward another person than in the case of neurotics. Similar conclusions were obtained by Ben-Ner and Halldorsson (2010) in the studies based also on the trust game. They found that the personality type positively influenced trust when the participant was characterized by high agreeableness, extraversion, or low neuroticism.

In the above paper, the impact of individual personality traits on the assessment of trust in people in the photos and on their attractiveness was examined. Personality traits are listed on the basis of the Neo Pi-R and IVE psychological tests and compared with the questionnaires examining the attractiveness and trust toward people in the photos. The faces obtained from generally available databases with a neutral expression were used in the research to minimize the impact of facial expression on the assessment. The analysis was performed based on the logistic stepwise backward regression. The included Wald coefficient allowed for the elimination of the least significant descriptive variables from the model with each successive step. The model was developed in the SPSS program and the learning and training of the model was carried out according to the procedures available in the program. Based on the obtained data and the performed analyses, it was shown that altruism has the greatest impact on the perceived attractiveness and trust. In both cases this trait has a positive effect on the dependent variable which indicates that altruists are more likely to judge others positively. Overall, it was noted that trust is largely influenced by the components of agreeableness and neuroticism.

After collecting the data from the respondents, apart from determining the personality traits influencing the decisions, a set of photos was extracted to be used in the electroencephalographic (EEG) examinations. The set of all 100 photos was divided into four groups: attractive and trustworthy, unattractive and untrustworthy, attractive and untrustworthy, unattractive and trustworthy. These are different combinations of dependent variables. Each of these groups has a different number of photos. The most numerous groups are great attractiveness with great confidence and small attractiveness with small confidence. The reason for this allocation of photos is the relationship between the attraction and the trust. Attractive people are believed to be more trustworthy, and less attractive people are likewise unreliable (Oosterhof and Todorov, 2008; Sutherland et al., 2017). Among the photos from the group of high attractiveness and low trust, there are only faces of people of African and European descent with a slight predominance of people of African descent. On the other hand, in the group where trust is great and attractiveness small, there are usually faces of people of Asian and European origin, with a slight predominance of the former. Based on the highest average ratings for each photo and the gender of each group, six photos were selected to be used in the EEG tests.

TABLE 10 | Estimating the values of the parameters of the logistic regression model for the dependent variable—attractiveness and trust.

		Trust					Attractiveness				
		B	**Standard error**	**Wald**	**Relevance**	**Exp(B)**	**B**	**Standard error**	**Wald**	**Relevance**	**Exp(B)**
Step 1	N1	−0.061	0.753	0.006	0.936	0.941	0.77	1.102	0.488	0.485	2.16
	E1	0.073	0.758	0.009	0.923	1.076	−2.149	0.967	4.941	0.026	0.117
	O1	−0.015	0.67	0	0.983	0.985	−0.11	0.828	0.018	0.894	0.896
	U1	1.347	0.811	2.759	0.097	3.848	0.719	0.748	0.925	0.336	2.053
	S1	−0.826	0.701	1.387	0.239	0.438	−2.232	1.363	2.683	0.101	0.107
	N2	2.499	0.498	6.948	0.008	12.173	0.073	0.955	0.006	0.939	1.076
	E2	−0.381	0.737	0.268	0.605	6.683	2.351	1.331	3.119	0.077	10.493
	O2	−0.566	0.754	0.564	0.453	0.568	0.261	0.918	0.081	0.776	1.298
	U2	−0.979	0.69	2.014	0.156	0.376	−1.705	1.137	2.248	0.134	0.182
	S2	0.07	0.533	0.017	0.896	1.072	−2.252	0.905	6.19	0.352	0.487
	N3	−0.246	0.603	0.166	0.684	0.782	−0.72	0.773	0.866	0.352	0.487
	E3	0.133	0.781	0.029	0.865	1.142	−1.92	1.182	2.636	0.104	0.147
	O3	−0.307	0.68	0.204	0.651	0.735	−1.339	0.936	2.048	0.152	0.262
	U3	2.266	0.817	7.687	0.006	9.643	5.185	1.753	8.746	0.003	178.51
	S3	−1.206	0.79	2.33	0.127	0.299	0.043	1.129	0.001	0.969	1.044
	N4	−1.111	0.714	2.419	0.12	0.329	−1.757	1.041	2.849	0.091	0.173
	E4	−1.168	0.755	2.391	0.122	0.311	−1.684	1.053	2.557	0.11	0.186
	O4	1.204	0.785	2.351	0.125	3.335	1.183	0.827	2.049	0.152	3.265
	U4	0.27	0.683	0.157	0.692	1.311	0.048	0.982	0.002	0.961	1.05
	S4	−0.154	0.646	0.057	0.811	0.857	2.137	1.166	3.361	0.067	8.475
	N5	−1.282	0.754	2.892	0.089	0.278	−0.188	0.835	0.051	0.821	0.828
	E5	−0.322	0.718	0.201	0.654	0.725	−2.679	1.338	4.009	0.045	0.069
	O5	−0.585	0.775	0.57	0.45	0.557	1.33	1.042	1.629	0.202	3.782
	U5	−0.72	0.661	1.187	0.276	0.487	−2.238	1.099	4.146	0.042	0.107
	S5	0.156	0.599	0.068	0.795	1.168	−1.189	0.794	2.239	0.135	0.305
	N6	−0.135	0.795	0.029	0.865	0.874	−0.358	1.094	0.107	0.744	0.699
	E6	0.417	0.669	0.389	0.533	1.517	1.075	0.891	1.456	0.227	2.93
	O6	1.176	0.726	2.622	0.105	3.24	1.496	0.822	3.307	0.069	4.462
	U6	−0.111	0.485	0.053	0.819	0.895	1.213	0.758	2.56	0.11	3.362
	S6	1.317	0.847	2.419	0.12	3.733	0.575	0.865	0.442	0.506	1.777
	impulsiveness	0.602	0.538	1.253	0.263	1.825	−1.418	0.824	2.958	0.085	0.242
	sdr	0.346	0.57	0.369	0.544	1.414	1.505	0.937	2.581	0.108	4.505
	empathy	−1.243	0.621	4.002	0.045	0.288	−1.339	0.818	2.676	0.102	0.262
	tmzS	−1.016	0.652	2.43	0.119	0.362	−1.385	0.98	1.998	0.157	0.25
	constant	0.04	0.358	0.013	0.91	1.041	−1.948	0.701	7.721	0.005	0.143
Step 26	U1	0.709	0.331	4.599	0.032	2.032	–	–	–	–	–
	N2	1.153	0.442	6.8	0.009	3.169	–	–	–	–	–
	U3	1.041	0.371	7.859	0.005	2.833	–	–	–	–	–
	N4	−0.721	0.333	4.703	0.03	0.486	–	–	–	–	–
	E4	−0.753	0.398	3.579	0.059	0.471	–	–	–	–	–
	O4	0.694	0.357	3.776	0.052	2.001	–	–	–	–	–
	N5	−0.605	0.374	2.609	0.106	0.546	–	–	–	–	–
	Impulsiveness	−0.717	0.368	3.794	0.051	2.047	–	–	–	–	–
	Empathy	−0.592	0.351	2.846	0.092	0.553	–	–	–	–	–
	Constant	0.096	0.273	0.124	0.725	1.1	–	–	–	–	–
Step 30	S2	–	–	–	–	–	−0.741	0.303	6	0.014	0.477
	U3	–	–	–	–	–	0.961	0.32	9	0.003	2.615
	O5	–	–	–	–	–	0.501	0.298	2.82	0.093	1.65
	sdr	–	–	–	–	–	0.535	0.299	3.208	0.073	1.707
	Empathy	–	–	–	–	–	−0.471	0.292	2.608	0.106	0.624
	Constant	–	–	–	–	–	−0.678	0.263	6.675	0.01	0.507

The table shows the statistics for each of the predictors in the model. As a result of the gradual elimination of the predictors in the light parts of the table (step 26 for the confidence dependent variable, step 30 for the attractiveness dependent variable), the variables furthest away from the dependent variable are presented.

TABLE 11 | List of personality traits with the most important ones for attractiveness (marked light yellow—the positive impact on the dependent variable, dark yellow—the negative impact on the dependent variable), trust (marked light green—the positive impact on the dependent variable, dark green—the negative impact on the dependent variable) and attractiveness and trust at the same time (marked light blue—the positive impact on the dependent variables, dark blue—the negative impact on the dependent variables).

Type of test	Personality factors	Component factors
Neo-Pi-R	Neuroticism	N1- Anxiety
		N2- Angry hostility
		N3- Depression
		N4- Self-consciousness
		N5- Impulsiveness
		N6- Vulnerability
	Extroversion	E1- Warmth
		E2- Gregariousness
		E3- Assertiveness
		E4- Activity
		E5- Excitement seeking
		E6- Positive emotions
	Openness	O1- Fantasy
		O2- Aesthetics
		O3- Feelings
		O4- Actions
		O5- Ideas
		O6- Values
	Agreeableness	U1- Trust
		U2- Straightforwardness
		U3- Altruism
		U4- Compliance
		U5- Modesty
		U6- Tendermindedness
	Conscientiousness	S1- Competence
		S2- Order
		S3- Dutifulness
		S4- Achievement striving
		S5- Self-discipline
		S6- Deliberation
IVE	Empathy	
	sdr - Tendency to take risks	
	Impulsiveness	

In the presented study, the focus was primarily on finding the relationship between the choices made toward people in the photos and personality trait. It is an introduction to further research using electroencephalography. EEG research will be developed on the above-mentioned set of faces and will be conducted on a group of 60 students.

Stimulating the brain with various types of stimuli results in the arousal coming from the centers responsible for reading and processing them. The resulting neural processes are analyzed in terms of time dynamics which are estimated using the event-related potential (ERP). Component N170 is assumed to be a component characteristic of facial perceptual processing and appears ∼170 ms after the stimulus has occurred. N170 can also appear as a result of multiple face presentations and is called the adaptive effect of N170 (Eimer et al., 2010). ERP components for trust and attractiveness appear both in the initial stages of facial processing (e.g., P100) and later (e.g., late positive potential, LPP). The study (Marzi et al., 2014) analyzed, inter alia, the differences between the ERP components of trustworthy and untrustworthy faces. The stimulus responses included P100 (110–130 ms), EPN (200–350 ms), and LPP (300–500 ms). Additionally, components for trustworthy faces showed lower amplitudes than untrustworthy faces. Both early and late facial processings were recorded in the study (Yang et al., 2011) where the reliability of pre-categorized faces was analyzed and the most significant responses were obtained for the C1 (40–90 ms) and LPC (400–600 ms). Early ERP modulations occur in response to attractive/unattractive faces (Marzi and Viggiano, 2010; Hahn et al., 2016). Early registration of stimulus-induced activity is the result of facial perceptual processing. The course of facial identity processing results in the registration of the signal within the N250 limits (Werheid et al., 2007) while the signals resulting from cognitive processing that occur within 300–600 ms from the stimulus occurrence are represented the latest (Calvo et al., 2018).

Taking into account the information already available on the processing of visual stimuli by the brain in the context of attractiveness and trust, and the data obtained in the present study, an ERP analysis is planned in the next works to find the correlation between the "class" of the face and the activity of selected areas of the brain using photos of women and men different facial features and nationalities, and then, using appropriate algorithms, to accurately indicate the areas of the cerebral cortex that showed the highest activity during the experiment in individual people. As shown in the above study, the personality traits of the evaluators influence largely the decisions. In addition to the above-mentioned studies, the signal analysis is planned in terms of individual differences of respondents. The collected data in combination with EEG data will constitute a sufficient set to build a classifier that will be able to predict consensus based on personality traits.

8. CONCLUSIONS

On the basis of the estimated models there were demonstrated the personality traits that are most important concerning the behavior toward others, and more precisely they affect the assessment of attractiveness and trust in people from the photos significantly. Among all the features for both dependent variables, altruism (U3) is the most important. The growing probability of giving the grade "attractive" and "trustworthy" can be justified by the character traits of these people, i.e., sensitivity to the fate of another person and selfless help. It has been noticed that the components belonging to the groups of agreeableness and neuroticism have a particularly large impact on trust while altruism (U3), trust (U1), and angry hostility (N2) increase the probability of trust twice or three times and excessive self-consciousness(N4) reduces the probability by about 50%.

Future study will assess the course of ERP induced during social assessment based on the first impressions and facial appearance and locate the most active areas of the brain, detailing the Brodmann's area (BA). For this purpose an electroencephalographic test will be carried out to assess the credibility and attractiveness of the presented faces. EEG data are expected to help find a correlation between the face "class" and the activity in the selected areas of the cerebral cortex.

DATA AVAILABILITY STATEMENT

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

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Maria Curie-Sklodowska University in Lublin Bioethical Commission. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

BB: project idea, manuscript writing, and surveying. GW: project idea and concept and model design. AB: psychological tests selection. AK: statistical analysis. All authors contributed to the article and approved the submitted version.

REFERENCES

Adolphs, R. (2001). The neurobiology of social cognition. *Curr. Opin. Neurobiol.* 11, 231–239. doi: 10.1016/S0959-4388(00)00202-6

Ancāns, K., and Austers, I. (2018). The influence of face trustworthiness on judgments in forensic context. *Baltic J. Psychol.* 19, 100–111.

Assem, M., Shashidhara, S., Glasser, M. F., and Duncan, J. (2021). Precise topology of adjacent domain-general and sensory-biased regions in the human brain. *bioRxiv [Preprint].* doi: 10.1101/2021.02.21.431622

Ballew, C. C., and Todorov, A. (2007). Predicting political elections from rapid and unreflective face judgments. *Proc. Natl. Acad. Sci. U.S.A.* 104, 17948–17953. doi: 10.1073/pnas.0705435104

Bar, M., Neta, M., and Linz, H. (2006). Very first impressions. *Emotion* 6:269.

Ben-Ner, A., and Halldorsson, F. (2010). Trusting and trustworthiness: what are they, how to measure them, and what affects them. *J. Econ. Psychol.* 31, 64–79. doi: 10.1016/j.joep.2009.10.001

Buchan, N. R., Croson, R. T., and Solnick, S. (2008). Trust and gender: an examination of behavior and beliefs in the Investment Game. *J. Econ. Behav. Organ.* 68, 466–476. doi: 10.1016/j.jebo.2007.10.006

Caci, H., Nadalet, L., Baylé, F. J., Robert, P., and Boyer, P. (2003). Cross-cultural study of the impulsiveness-venturesomeness-empathy questionnaire (IVE-7). *Comprehens. Psychiatry* 44, 381–387. doi: 10.1016/S0010-440X(03)00105-6

Calvo, M. G., Gutiérrez-García, A., and Beltrán, D. (2018). Neural time course and brain sources of facial attractiveness vs. trustworthiness judgment. *Cogn. Affect. Behav. Neurosci.* 18, 1233–1247. doi: 10.3758/s13415-018-0634-0

Chang, L. J., Doll, B. B., van't Wout, M., Frank, M. J., and Sanfey, A. G. (2010). Seeing is believing: trustworthiness as a dynamic belief. *Cogn. Psychol.* 61, 87–105. doi: 10.1016/j.cogpsych.2010.03.001

Costa, P. T., and McCrae, R. R. (1992). Normal personality assessment in clinical practice: the NEO personality inventory. *Psychol. Assess.* 4:5. doi: 10.1037/1040-3590.4.1.5

Costafreda, S. G., Brammer, M. J., David, A. S., and Fu, C. H. (2008). Predictors of amygdala activation during the processing of emotional stimuli: a meta-analysis of 385 PET and fMRI studies. *Brain Res. Rev.* 58, 57–70. doi: 10.1016/j.brainresrev.2007.10.012

Crouzet, S. M., Kirchner, H., and Thorpe, S. J. (2010). Fast saccades toward faces: face detection in just 100 ms. *J. Vis.* 10:16. doi: 10.1167/10.4.16

De Heering, A., Rossion, B., and Maurer, D. (2012). Developmental changes in face recognition during childhood: evidence from upright and inverted faces. *Cogn. Dev.* 27, 17–27. doi: 10.1016/j.cogdev.2011.07.001

di Oleggio Castello, M. V., and Gobbini, M. I. (2015). Familiar face detection in 180 ms. *PLoS ONE* 10:e0136548. doi: 10.1371/journal.pone.0136548

Dobs, K., Isik, L., Pantazis, D., and Kanwisher, N. (2019). How face perception unfolds over time. *Nat. Commun.* 10, 1–10. doi: 10.1038/s41467-019-09239-1

Eimer, M., Kiss, M., and Nicholas, S. (2010). Response profile of the face-sensitive N170 component: a rapid adaptation study. *Cereb. Cortex* 20, 2442–2452. doi: 10.1093/cercor/bhp312

Engell, A. D., Haxby, J. V., and Todorov, A. (2007). Implicit trustworthiness decisions: automatic coding of face properties in the human amygdala. *J. Cogn. Neurosci.* 19, 1508–1519. doi: 10.1162/jocn.2007.19.9.1508

Ert, E., Fleischer, A., and Magen, N. (2016). Trust and reputation in the sharing economy: the role of personal photos in Airbnb. *Tour. Manage.* 55, 62–73. doi: 10.1016/j.tourman.2016.01.013

Freeman, J. B., Stolier, R. M., Ingbretsen, Z. A., and Hehman, E. A. (2014). Amygdala responsivity to high-level social information from unseen faces. *J. Neurosci.* 34, 10573–10581. doi: 10.1523/JNEUROSCI.5063-13.2014

Haas, B. W., Ishak, A., Anderson, I. W., and Filkowski, M. M. (2015). The tendency to trust is reflected in human brain structure. *Neuroimage* 107, 175–181. doi: 10.1016/j.neuroimage.2014.11.060

Hahn, A. C., Symons, L. A., Kredel, T., Hanson, K., Hodgson, L., Schiavone, L., et al. (2016). Early and late event-related potentials are modulated by infant and adult faces of high and low attractiveness. *Soc. Neurosci.* 11, 207–220. doi: 10.1080/17470919.2015.1059361

Jessen, S., and Grossmann, T. (2019). Neural evidence for the subliminal processing of facial trustworthiness in infancy. *Neuropsychologia* 126, 46–53. doi: 10.1016/j.neuropsychologia.2017.04.025

Kościński, K. (2007). Facial attractiveness: general patterns of facial preferences. *Anthropol. Rev.* 70, 45–79. doi: 10.2478/v10044-008-0001-9

Martin, J. G., Davis, C. E., Riesenhuber, M., and Thorpe, S. J. (2018). Zapping 500 faces in less than 100 seconds: evidence for extremely fast and sustained continuous visual search. *Sci. Rep.* 8, 1–12. doi: 10.1038/s41598-018-30245-8

Marzi, T., Righi, S., Ottonello, S., Cincotta, M., and Viggiano, M. P. (2014). Trust at first sight: evidence from ERPs. *Soc. Cogn. Affect. Neurosci.* 9, 63–72. doi: 10.1093/scan/nss102

Marzi, T., and Viggiano, M. P. (2010). When memory meets beauty: insights from event-related potentials. *Biol. Psychol.* 84, 192–205. doi: 10.1016/j.biopsycho.2010.01.013

Meuwissen, A. S., Anderson, J. E., and Zelazo, P. D. (2017). The creation and validation of the developmental emotional faces stimulus set. *Behav. Res. Methods* 49, 960–966. doi: 10.3758/s13428-016-0756-7

Mondloch, C. J., Gerada, A., Proietti, V., and Nelson, N. L. (2019). The influence of subtle facial expressions on children's first impressions of trustworthiness and dominance is not adult-like. *J. Exp. Child Psychol.* 180, 19–38. doi: 10.1016/j.jecp.2018.12.002

Müller, J., and Schwieren, C. (2020). Big five personality factors in the trust game. *J. Bus. Econ.* 90, 37–55. doi: 10.1007/s11573-019-00928-3

Oosterhof, N. N., and Todorov, A. (2008). The functional basis of face evaluation. *Proc. Natl. Acad. Sci. U.S.A.* 105, 11087–11092. doi: 10.1073/pnas.0805664105

Oosterhof, N. N., and Todorov, A. (2009). Shared perceptual basis of emotional expressions and trustworthiness impressions from faces. *Emotion* 9:128. doi: 10.1037/a0014520

Shinners, E. (2009). Effects of the "what is beautiful is good" stereotype on perceived trustworthiness. *UW-L J. Undergrad. Res.* 12, 1–5.

Silver, B. M., Conte, M. M., Victor, J. D., and Jones, R. M. (2020). Visual search for circumscribed interests in autism is similar to that of neurotypical individuals. *Front. Psychol.* 11:2656. doi: 10.3389/fpsyg.2020.582074

Strohminger, N., Gray, K., Chituc, V., Heffner, J., Schein, C., and Heagins, T. B. (2016). The MR2: A multi-racial, mega-resolution database of facial stimuli. *Behav. Res. Methods* 48, 1197–1204. doi: 10.3758/s13428-015-0641-9

Sutherland, C. A., Young, A. W., and Rhodes, G. (2017). Facial first impressions from another angle: how social judgements are influenced by changeable and invariant facial properties. *Br. J. Psychol.* 108, 397–415. doi: 10.1111/bjop.12206

Todorov, A., Baron, S. G., and Oosterhof, N. N. (2008). Evaluating face trustworthiness: a model based approach. *Soc. Cogn. Affect. Neurosci.* 3, 119–127. doi: 10.1093/scan/nsn009

Van't Wout, M., and Sanfey, A. G. (2008). Friend or foe: the effect of implicit trustworthiness judgments in social decision-making. *Cognition*, 108, 796–803. doi: 10.1016/j.cognition.2008.07.002

Werheid, K., Schacht, A., and Sommer, W. (2007). Facial attractiveness modulates early and late event-related brain potentials. *Biol. Psychol.* 76, 100–108. doi: 10.1016/j.biopsycho.2007.06.008

Wierzbicki, A. (2008). The case for fairness of trust management. *Electron. Notes Theor. Comput. Sci.* 197, 73–89. doi: 10.1016/j.entcs.2007.12.018

Wilson, J. P., and Rule, N. O. (2015). Facial trustworthiness predicts extreme criminal-sentencing outcomes. *Psychol. Sci.* 26, 1325–1331. doi: 10.1177/0956797615590992

Yang, D., Qi, S., Ding, C., and Song, Y. (2011). An ERP study on the time course of facial trustworthiness appraisal. *Neurosci. Lett.* 496, 147–151. doi: 10.1016/j.neulet.2011.03.066

Zebrowitz, L. A., Franklin, R. G. Jr., and Boshyan, J. (2015). Face shape and behavior: implications of similarities in infants and adults. *Pers. Individ. Differ.* 86, 312–317. doi: 10.1016/j.paid.2015.06.036

Zebrowitz, L. A., and Montepare, J. M. (2008). Social psychological face perception: why appearance matters. *Soc. Pers. Psychol. Compass* 2, 1497–1517. doi: 10.1111/j.1751-9004.2008.00109.x

Conflict of Interest: The 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.

frontiers in Human Neuroscience

ORIGINAL RESEARCH
published: 17 September 2021
doi: 10.3389/fnhum.2021.659243

What to Believe? Impact of Knowledge and Message Length on Neural Activity in Message Credibility Evaluation

Lukasz Kwasniewicz[1], *Grzegorz M. Wojcik*[1]*, *Piotr Schneider*[1], *Andrzej Kawiak*[1] and *Adam Wierzbicki*[2]

[1] Chair of Neuroinformatics and Biomedical Engineering, Institute of Computer Science, Maria Curie-Sklodowska University in Lublin, Lublin, Poland, [2] Polish-Japanese Academy of Information Technology, Warsaw, Poland

Understanding how humans evaluate credibility is an important scientific question in the era of fake news. Message credibility is among crucial aspects of credibility evaluations. One of the most direct ways to understand message credibility is to use measurements of brain activity of humans performing credibility evaluations. Nevertheless, message credibility has never been investigated using such a method before. This article reports the results of an experiment during which we have measured brain activity during message credibility evaluation, using EEG. The experiment allowed for identification of brain areas that were active when participant made positive or negative message credibility evaluations. Based on experimental data, we modeled and predicted human message credibility evaluations using EEG brain activity measurements with F1 score exceeding 0.7.

Keywords: EEG, credibility, source localization, LORETA, classifiers

OPEN ACCESS

Edited by:
Changming Wang,
Capital Medical University, China

Reviewed by:
Marco Viviani,
University of Milano-Bicocca, Italy
Jack Adam Noah,
Yale University, United States

***Correspondence:**
Grzegorz M. Wojcik
gmwojcik@live.umcs.edu.pl

Specialty section:
This article was submitted to Brain Imaging and Stimulation, a section of the journal Frontiers in Human Neuroscience

Received: *31 March 2021*
Accepted: *28 July 2021*
Published: *17 September 2021*

Citation:
Kwasniewicz L, Wojcik GM, Schneider P, Kawiak A and Wierzbicki A (2021) What to Believe? Impact of Knowledge and Message Length on Neural Activity in Message Credibility Evaluation. Front. Hum. Neurosci. 15:659243. doi: 10.3389/fnhum.2021.659243

1. INTRODUCTION

The World Wide Web has been designed for low barriers of entry, enabling fast, and cheap publication of content. At the same time, the prevalent business model of the web provides high incentives for producing Web content that impacts opinions and beliefs of Web users. These commercial incentives are caused by the popularity of Web-based marketing and advertising. However, Web content affects not just our shopping decisions, but also decisions regarding our health, or politics. In this technical and economic environment, the spread of fake news has become an increasingly significant social problem (Sharma et al., 2019). Fake news disseminate through social media, Web-based newspapers, blogs, and regular Web pages.

Although combating fake news has been the focus of policy and scientific research since 2016, to date, little is known about why people believe in fake news. While factors that contribute to belief in fake news have been studied by social psychology (Rutjens and Brandt, 2018; Forgas and Baumeister, 2019), these results have been obtained from the declarative studies. Simply asking Web users whether they believe fake news, or indirectly inferring this conclusion from their behavior, cannot reveal the real reasons for such a decision. The response to a question about the believability of fake news also cannot be a basis for a certain conclusion that fake news was indeed credible, because of possible biases in the response.

Surprisingly, almost no previous research has attempted to directly measure brain activity to study basic processes occurring in brain during credibility evaluation. Previous research using EEG or fMRI has been devoted to lie detection (Wang et al., 2016; Meijer and Verschuere, 2017). This approach is based on the investigation of the brain activity of the author, and not the receiver of the message.

The focus of our research is the brain activity during evaluation of message credibility. It is a fundamental aspect of credibility evaluation that focuses on the content, and not on the source of a message. In many online scenarios, Web users must evaluate the credibility of content without knowing the content's author or source. Our goal is to identify brain areas and periods of brain activity that are most active or most important in the process of textual message credibility evaluation. This process relies on the individual, subjective perception of a Web user and can therefore be studied experimentally using EEG.

This basic question leads us to a more applicable goal: creation of a method for EEG-based message credibility evaluation based only on the observed brain activity. In the future, we envisage the use of EEG for either testing the credibility of information in the form of fake news, or (to the contrary) correcting information designed to counteract fake news. Similarly to the use of EEG in online marketing (Deitz et al., 2016; Guixeres et al., 2017), researchers could evaluate in such a setting the credibility of information using a panel of information consumers.

The goal of this article is to address the following research questions:

- What brain areas are active while a receiver is evaluating message credibility?
- Does brain activity during credibility evaluation depend on message design?
- Can we model and predict human message credibility evaluations using EEG brain activity measurements?

One of the difficulties in addressing these questions lies in the fact that message credibility evaluation can be affected by two competing factors: message design and prior knowledge of message recipients. An experiment for studying message credibility must control prior knowledge of experiment participants about the message, as well as other factors that may influence message credibility evaluation. In this article, we describe an experiment that enables the study of message credibility evaluation without prior knowledge, and with perfect knowledge. In the former case, the message credibility evaluation of experiment participants can be influenced by irrelevant factors of message design. This situation reflects the reality of many Web users who encounter fake news on various subjects.

2. RELATED WORK

2.1. Basic Concepts: Credibility and Truth

The concept of credibility is grounded in common sense and in used in scientific research. Modern research on credibility is active especially in the field of psychology, media science and informatics (Viviani and Pasi, 2017). In research, credibility is usually understood as a perceived quality of individuals. The earliest theoretical work on credibility from the 1950s is due to Hovland and Weiss (1951), who distinguished between *source, message, and media credibility*.

Out of these three, source credibility and message credibility are a good starting point for a top-down study of the complex concept of credibility. These two concepts are closely related to the dictionary definition of the term "credibility" (Oxford Advanced Learner's Dictionary): "the quality that somebody/something has that makes people believe or trust them." A part of this definition focuses on a person ("somebody") and is close to the concept of source credibility. Another part is about "something"—the message itself. That part defines message credibility that frequently needs to be evaluated on the Web without knowing the source of the message.

Information scientists have studied credibility evaluations aiming at designing systems that could evaluate Web content credibility automatically or support human experts in making credibility evaluations (Wawer et al., 2014; Liu et al., 2015; Kakol et al., 2017). However, human credibility evaluations are often subjective, biased or otherwise unreliable (Kakol et al., 2013; Rafalak et al., 2014), making it necessary to search for new methods of credibility evaluation, such as the EEG-based methods proposed in this article.

The concept of truth is even more complex than the concept of credibility. Without going into details (the reader is referred to Wierzbicki, 2018 for a detailed discussion), there exist several conflicting definitions of truth, such as scientific truth or post-structuralist truth. Truth may be also undecidable, or truth evaluation may be impossible in practice. However, the purpose of processes of disinformation verification and debunking is to discover information that is untrue and correct it by pointing out the truth. In this article, we shall assume a definition of truth as an objectively verifiable information that is the basis for disinformation checking and debunking.

Having said that, the relationship between credibility and truth is not simple. Non-expert web users may evaluate information that is not true as credible. We would expect that experts would evaluate only true information as credible. However, experts are human too, and can make mistakes. On the other hand, in many areas we have no choice but to rely on expert opinion, and to accept experts' credibility evaluations as truth.

2.2. Message Credibility

A search for the term "message credibility" on Google Scholar returns over 1,000 results (for an overview of recent publications, especially on the subject of Web content credibility, see Viviani and Pasi, 2017; Wierzbicki, 2018). Media science researchers have attempted to create scale for declarative measurements of message credibility (Appelman and Sundar, 2016). Message credibility has been investigated in the area of healthcare (Borah and Xiao, 2018).

Message credibility has been defined as a scientific concept by Hovland and Weiss (1951) as the aspect of credibility that depends only on the communicated message, instead of the message's source or communication medium. On the Web, the message is a webpage that includes an article (or a shorter text in case of social media). Message credibility depends on the

textual content, on images or videos (as well as advertisements) embedded in the webpage, and on webpage design or style.

What follows is that message credibility is affected by many factors (features of the message). Even if we consider only the textual content, message credibility can be affected by semantic or pragmatic aspects of the message (its meaning and style, persuasiveness, sentiment etc.) This complexity is especially important because message credibility is usually evaluated rapidly on the Web.

Tseng and Fogg (1999) introduced the two concepts of "surface credibility" and "earned credibility". Surface credibility is based on a fast and superficial examination of the message (similar to System I reasoning, as introduced by Kahneman, 2011). Earned credibility, on the other hand, is the result of a more deliberate and time-consuming evaluation of the message, like System II reasoning. Research (Wierzbicki, 2018) has established that most users evaluate webpage credibility quickly, usually requiring several minutes (3 min are enough for most Web page credibility evaluations). Earned credibility evaluation requires much more time and usually involves a debunking or verification process. These observations are relevant for our experiment design. In this article, we focus on surface credibility evaluations based on the contents of the message. In order to begin understanding brain activity during message credibility evaluation, we shall design messages that differ by a single aspect that can be evaluated quickly.

Message credibility evaluation of Web content is often difficult for ordinary Web users. This problem has led to numerous attempts of designing automated or semi-automated IT systems that support Web content credibility evaluation (Viviani and Pasi, 2017; Wierzbicki, 2018; Sharma et al., 2019). In this article, we focus on how humans make message credibility evaluations without computer support.

2.3. Experimental fMRI and EEG Findings

Research on brain signaling in decision-making focuses on neuroimaging (fMRI) and the activity of specific parts of the brain in the situation when participants solve various tasks.

Many of these studies concern confidence in the person, as in facial plausibility studies in which Amygdala activity has been demonstrated (Rule et al., 2013), also as Precuneus-the medial part of Brodmann Area (BA) 7, Inferior Frontal Gyrus—BA44, BA45, BA47, Medial Prefrontal Cortex—BA12, BA25, BA32, BA33, BA24 (Filkowski et al., 2016). The dynamic role of the Paracingulate Cortex(BA9/32) and Septal Area in supporting conditional and unconditional trust strategies in Trust Games was investigated (Krueger et al., 2007). Also difference in neural activation (BA7, BA8, BA40) between prosocials and proselfs people during decision making and interaction effect between dispositional trust and social value orientation (BA9, BA31, BA39) was shown in Emonds et al. (2014).

Few of these studies are concerned with trusting the message itself. In the Processing of Online Trust Signals study, online shopping activity of Rolandic Operculum was reported with the most trustworthy signal (BA 44 is part of it), calcarine—is where the Primary Visual Cortex is concentrated (BA17), Angular Gyrus (BA39) and Superior Motor Area, pre-SMA (BA 8) (Casado-Aranda et al., 2019).

Anterior cingulate cortex (BA24, 32, 33) activity was observed, but only within high effort condition such as the quickest possible pressing of the button task (Mulert et al., 2008) in the decision-making study on the pitch of the tone.

Heekeren et. al. suggest that Posterior Superior Temporal Sulcus and Ventromedial Prefrontal Cortex (BA10) are involved in decision-making regarding scenarios devoid of violence and direct bodily harm (Heekeren et al., 2003).

In terms of EEG, the N1 and P300 signals combined with fMRI data in value-based decision-making were examined (Mulert et al., 2008; Larsen and O'Doherty, 2014). In this study, decision making involved activation of Dorsomedial Prefrontal Cortex (BA8, BA9, BA10, BA24, and BA32) and Ventromedial Prefrontal Cortex (BA10).

In a study by Douglas et al. (2013) with the help of ICA, the authors managed to create satisfactory models using EEG signal (power envelopes derived from spectral bands as features) in classification of belief/disbelief decision-making.

Areas such as BA.08, BA.09 have been proposed to participate in a general mechanism for perceptual decision-making in the human brain (Heekeren et al., 2004).

3. EXPERIMENT DESIGN

3.1. Motivation for Experiment Design

As a practical reference situation that motivates our experiment design, consider a receiver of a message on social media. The message could be true, or it could be disinformation. The receiver evaluates the surface credibility of the message. This means that she will quickly (within a matter of minutes or even seconds) decide whether the message is credible or not, and act accordingly (by forwarding, retweeting, or liking the message). This kind of situation is so common that experts on media literacy have coined the slogan: "think before you like" (Harrison, 2017). Fast, superficial credibility evaluation is not only common, but can lead to innumerable social harm, especially in the context of health-related Web content.

For the sake of our reasoning, consider that the message contains information related to health or medicine (about 60% of Americans and Europeans go online looking for health information; Viviani and Pasi, 2017). It could be a simple statement like "Low doses of aspirin can be safely consumed in the second trimester of pregnancy" to increasingly complex statements, for example: "Coenzyme Q10 supplements may help prevent statin side effects in some people, though more studies are needed to determine any benefits of taking it." For the record, the first statement is generally true, especially if there is risk of miscarriage. The second statement is mostly false and contains a hedging part to make it more credible; it is designed to sell coenzyme Q10 supplements, but can also discourage the use of statins by people who need them because of high cholesterol and arteriosclerosis.

Please note that the first factor that impacts a receiver's credibility evaluation of such a statement is the receiver's knowledge and experience. Controlling this factor is therefore

a crucial element of our experiment design. However, it is hard to evaluate degrees of knowledge and control their impact on credibility evaluation. For this reason, we designed two experimental conditions: full knowledge and lack of knowledge. Lack of knowledge is particularly characteristic of online situations, such as when a non-expert social media user is evaluating a message concerning health or medicine. The full knowledge scenario is more applicable to credibility evaluations by experts.

If the message is just a simple textual statement, the second factor that can influence credibility evaluation is the message's persuasiveness. This factor is even harder to understand and control. Contemporary research on Natural Language Processing for detecting persuasive disinformation uses complex language models, and achieve accuracy of 70–80% (Wawer et al., 2014). For experiment design, this is insufficient. Moreover, persuasive disinformation can influence many different cognitive biases or heuristics—they can appeal to positive or negative emotions, use hyperbolization, forgery, selective presentation of information, and many other techniques. In a single experiment, it would be impossible to consider all of them, and focusing on a single, yet complex technique is contrary to our research goal of achieving a generalizable understanding of how the brain processes message credibility evaluation.

For this reason, in our experiment we decided to use a simple factor that differentiates between messages—the message's complexity. Research has found that this factor has a significant impact on message persuasiveness (Wawer et al., 2014; Kerz et al., 2021). We use short messages that consist of one or a few words, and long messages that give a detailed description of the translated kanji characters. Results of our pilot experiment (Kwaśniewicz et al., 2020) (and of the main experiment itself—see section 4.1) indicate that receivers are influenced by message complexity and tend to positively evaluate message credibility of long messages more frequently than of short messages. Message credibility evaluation without knowledge, affected only by message length, is not random, as evidenced by statistical tests.

3.2. Controlling Participant Knowledge

The goal of the experiment was to observe electrical activity and the most active areas of the participant's brain cortex during tasks involving message credibility evaluation, as well as the influence of the message design and content on this process. In order to ensure that participants could only rely on message design during the experiment, it was designed so that the participants would not be familiar with the topic of the messages. Selected topic of the messages concerned the meaning of Japanese kanji signs.

The experiment was designed to create a situation in which the participants assess truthfulness or falsehood with practically no prior knowledge of the message subject. Knowledge of participants about the correct meaning of Kanji signs used in experiment was controlled. The initial condition ensured that participants had no knowledge of Kanji signs[1]. However, the participants were taught the meaning of three Kanji signs. When they were shown a Kanji sign unfamiliar to them, experimental setting resembled the case when a person who has no knowledge of the subject receives fake news. In order to study the effect of knowledge, we also showed them the previously taught Kanji signs.

Note that instead of Kanji signs, we could have used other images (for example, USG scans of different types of tissue). Our choice of Kanji signs was motivated by the fact that this type of image has been extensively studied using EEG (Sakurai et al., 2000; Ardila et al., 2015; Higashino and Wakamiya, 2021), and we knew from literature what brain activity to expect from participants who examined Kanji signs.

3.3. Participants and Ethical Commission's Permission

The participants to the experiment were right-handed male students without any knowledge of Japanese. A total of 107 participants took part in the experiment. EEG signal from 105 participants was collected. The experiment was carried upon the permission of the University's Bioethical Commission (MCSU Bioethical Commission permission 13.06.2019).

3.4. Message Credibility Evaluation Task

In the first part of the experiment, participants were requested to learn three Japanese Kanji signs by heart: eye,mouth and mountain. These characters were chosen because of their simplicity and a lack of similarity with other characters, making it simple for participants to correctly identify them. In fact, during the entire experiment, no participant made even a single mistake during the identification of these characters.

In the next step, participants saw a single message on the screen that contained a translation of a Kanji sign into their native language. They had to answer whether they considered the translation of the Kanji sign to be true or not. Participants had to answer in a maximum of 3.5 s from the appearance of the translation.

The decision task was formulated in the form of a question: "Is this Japanese translation true." The participants could answer the question by selecting "Yes" or "No." The choice of the answer "yes" was equivalent to a positive evaluation of message credibility. An example screen is shown on **Figure 1**. There were 260 such questions in total.

3.5. Controlling the Language Complexity of the Message

The proposed translations shown to the participants differed in the length of the explanation of the Kanji sign. The meaning of 160 translations was explained in a single word, while the meaning of the remaining 80 translations was explained in a full sentence in the participants' native language. Longer explanations included were designed to give additional detail or to logically explain the relationship between the shape and meaning of the Kanji sign. For examples, see **Figure 2**. For brevity, we refer to

[1]All experiment participants had no knowledge of Japanese. Japanese language is not popular in Poland, it is not taught neither in schools nor at our University. Before the participation, students filled up the questionnaire in which they were asked to tick the appropriate level of knowledge of foreign languages as well as their age and others.

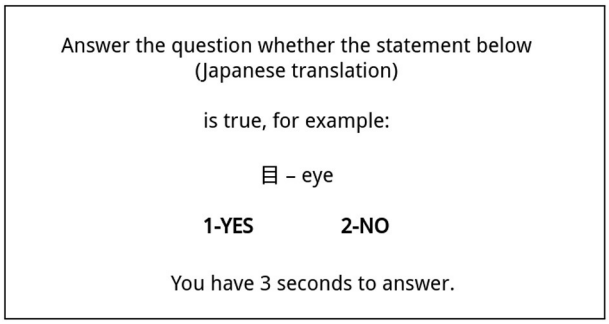

FIGURE 1 | Typical screen shown to participant during the experiment. In this case participant was to decide whether this Kanji sign means Eye, having perfect knowledge about it.

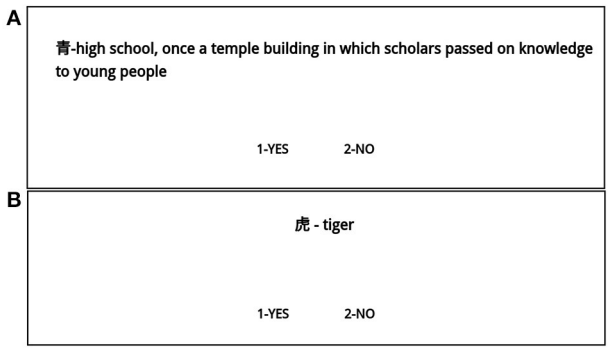

FIGURE 2 | Examples of screens with a long **(A)** and short **(B)** notes used in experiment for translation of Kanji sign.

the single-word Kanji sign translations as "short note," and to the longer translations as "long note."

3.6. Experimental Cases and Data

There were 240 screens shown to each participant during the experiments.

The first group of 80 screens contained one of the three Kanji signs that the participants have learned before, described by a short note. Therefore, the participants had perfect knowledge of them. This first group of 80 screens consisted of 40 screens with correct translations and 40 screens with incorrect translations.

In the second group, there were 80 screens with Kanji signs that were completely unknown to the participants, described by a short note. In the third group, there were 80 screens also with completely unknown Kanji signs, described by a long note.

It should be noted that there was no group of screens with perfectly-known Kanji signs described by a long note. The reason for this is that for the perfect knowledge cases, we did not wish to confuse users with other factors that could have an impact on their message credibility evaluations.

Such a setup allowed us to register electroencephalographic activity in the following six cases of choice:

1. TT: true translation of a known Kanji sign was evaluated as credible
2. FF: false translation of a known Kanji sign was evaluated as not credible
3. ST: short translation of an unknown Kanji sign was evaluated as credible
4. SF: short translation of an unknown Kanji sign was evaluated as not credible
5. LT: long translation of an unknown Kanji sign was evaluated as credible
6. LF: long translation of an unknown Kanji sign was evaluated as not credible

In this experiment, participants were not mistaken in questions about known Kanji signs, and no signal has been registered for the following hypothetical cases:

- true translation of a known Kanji sign was evaluated as not credible (TF)
- false translation of a known Kanji sign was evaluated as credible (FT)

This means that in our experiment, we can consider the cases when participants knew the Kanji signs as a model of message credibility evaluation with perfect knowledge.

3.7. Hypotheses

We formulated the following hypotheses:

1. Length of the note has a significant positive influence on the participant's decision about message credibility.
2. Decision of participants about message credibility in all three cases (short note with or without previous knowledge, and long note without previous knowledge) can be predicted based on measurements of mean electric charges in participant's brains.
3. Length of the note has a significant influence on brain activity during making decisions process concerning message credibility, and there are some significant differences in the models predicting decisions of participants who had seen long note compared to participants who had seen short notes.
4. Previously learned knowledge of the shown Kanji sign had a significant influence on brain activity during making decisions process concerning message credibility and there are some significant differences in the models predicting decisions of participants who had previously knowledge as compared to participants who did not have this knowledge.
5. There are some significant differences in the models predicting decisions of participants who frequently choose long note as compared to models of other participants.
6. There is a satisfactory model to predict the participants' decision based on the left Brodmann Areas 08,09 described in the literature as related to the decision-making process.

Hypothesis 1 is not directly related to participants' brain activities. It is rather a test of our experiment's internal validity. Positive validation of hypothesis 1 would confirm that there is a relationship between one of the main independent variables of

our experiment and the participant's decision. Such a relationship would partially confirm the internal validity of our experiment.

Hypothesis 2 can be validated by constructing classifiers that predict the (binary) decision of participants with sufficiently high accuracy. However, the validation of hypothesis 2 requires training of three classifiers: first based on the set of participants who evaluated message credibility based on their knowledge (this classifier would have two classes: TT and FF), second classifier based on message credibility evaluations made without knowledge (classes: ST and SF), and third one based on the set of participants who made decision under the impact of message design (LT and LF).

Hypothesis 3 is related to the first and second research question. It is focused on differences in the brain processes during credibility evaluation processes under the impact of message design (length). To validate this hypothesis, we need to compare two models based on situation when short note is evaluated (classes ST and SF) and when long note is evaluated (classes LT and LF). To enable this comparison, the classifiers trained for validating hypothesis 2 should be interpretable (based on logistic regression, decision trees or a similar method).

Hypothesis 4 concerns the effect of other main independent variable: previous learned knowledge of the Kanji sign. The experiment design allowed us to control this variable: our participants had perfect knowledge or no knowledge. We can therefore study the effect of knowledge on message credibility evaluation process. A validation of this hypothesis requires a comparison between two different classification models: one for classes TT and FF, and the other for ST and SF.

Similarly, the validation of hypothesis 5 requires training of two classifiers, one based on the set of participants who tend to evaluate long messages as credible, and another one based on the remaining set of participants. The comparison of these two classifiers is only possible if the two of them are explainable, which excludes the use of black-box classifiers such as neural networks.

Hypotheses 6 can be validated by constructing classifiers that will use signal from left Brodmann Areas 08,09 and predict the (binary) decision of participants with sufficiently high accuracy.

3.8. EEG Measurements

Our empirical experiments involved top EEG devices. We were equipped with a dense array amplifier recording the cortical activity with up to 500 Hz frequency through 256 channels HydroCel GSN 130 Geodesic Sensor Nets provided by EGI[2]. In addition, in the EEG Laboratory the Geodesic Photogrammetry System (GPS) was used.

Estimating ERP for each of the 256 electrodes is not necessary for ERP observation, as in general standards there are just a few electrodes (in our case 26) playing an important role in cognitive tasks[3]. However, for the sLORETA source localization analyses (used for verification of the next hypotheses) the ERP for all 256 electrodes had to be in fact calculated on the fly.

Having the ERP signal estimated for each electrode out of 256, it was possible to calculate the mean electric charge (MEC) flowing through the BA situated under these electrodes on the brain cortex in cognitive processing time interval (CPTI) as described in Wojcik et al. (2018) and Kawiak et al. (2020). Moreover, it was also possible to conduct the full source localization analysis of the signal originating from all 256 electrodes using sLORETA algorithm (GeoSourse parameters set as follow: Dipole Set: 2 mm Atlas Man, Dense: 2,447 dipoles Source Montages: BAs). Mean electric current flowing through each BA and varying in time was given as an output. Having those values calculated, it was possible to integrate that current in time and then get the MEC. The mean electric charge calculated for each electrode using source localization techniques could, as we intended, indicate the hyperactivity of some BAs that are not necessary precisely situated under the cognitive electrodes. For all calculations of MEC, the CPTI was divided into 5 ms time intervals. The procedure of calculating MEC has been described detail in Wojcik et al. (2018).

4. EXPERIMENT RESULTS

4.1. Impact of Note Length on Message Credibility Evaluations Without Prior Knowledge

As described in section 3.3, we have collected sufficient data to measure ERP and execute source localization from 105 male, right-handed participants. Signal was collected from 105 people in cases TT∪FF, 104 in cases ST∪SF, and 95 in cases LT∪LF. One person finished experiment after the first part (TT∪FF), while nine others did not answer the question about long notes (LT∪LF).

The impact of note length on message credibility evaluations can be established by comparing four experimental cases: LT, LF, ST, and SF. In particular, hypothesis 1 states that we expect that note length has a significant positive influence on the participants' decision about message credibility. Thus, we expect that

$$\frac{|LT|}{|LT| + |LF|} > \frac{|ST|}{|ST| + |SF|} \tag{1}$$

To check whether this condition is satisfied, we only needed to count the number of experimental results in the four cases. There were 4,079 results when participants evaluated long translation of an unknown Kanji sign as true (LT case) and 3,526 results when participants evaluated long translation of an unknown Kanji sign as false (LF case). This gives a proportion of

$$\frac{|LT|}{|LT| + |LF|} = 53.6\% \tag{2}$$

positive message credibility evaluations of long notes. Correspondingly, there were 3,455 experiment results when

[2]Electrical Geodesic Systems, Inc., 500 East 4th Ave. Suite 200, Eugene, OR 97401, USA.

[3]The electrodes are described in EGI 256-channel cap specification as best for cognitive ERP observations, covering the scalp regularly, and numbered as follows: E98, E99, E100, E101, E108, E109, E110, E116, E117, E118, E119, E124, E125, E126, E127, E128, E129, E137, E138, E139, E140, E141, E149, E150, E151, E152. Those electrodes are automatically chosen for observing P-300 ERP signal by NetStation software.

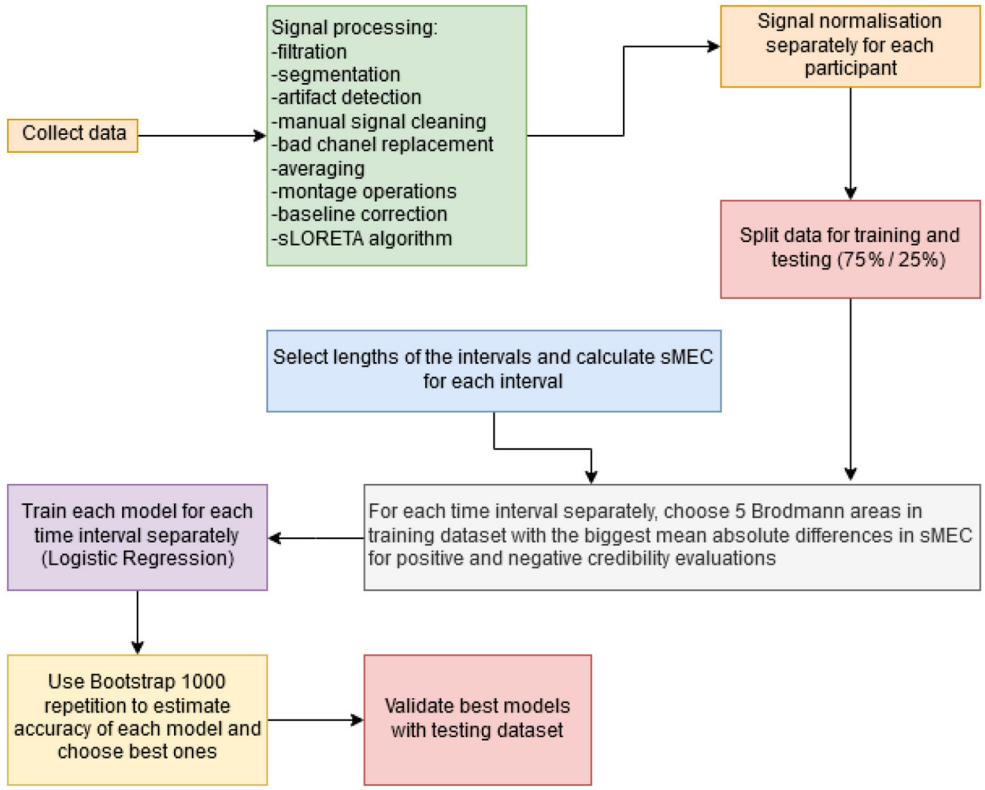

FIGURE 3 | Data processing steps in analysis of experimental results.

participants evaluated short translation of an unknown Kanji sign as true (ST case) and 4,977 experiment results when participants evaluated short translation of an unknown Kanji sign as false (SF case). The resulting proportion of positive message credibility evaluations of short notes is

$$\frac{|ST|}{|ST| + |SF|} = 40.9\% \tag{3}$$

which is a significantly lower number than for long notes. This observation positively verifies Hypothesis 1.

We also tested whether or not the message credibility evaluation was random in the cases without knowledge. Participants could evaluate the long or short note as credible. If the choice would be random, the choices of the long or short note should form a binomial distribution with probability 0.5. We used the binomial test and calculated the p-value, which was <0.000001 (we observed 3,455 choices of the short note out of 7,534 message credibility evaluations). Therefore, we concluded that we could reject the possibility that the choices of short or long notes in the experiment cases without knowledge were binomially random.

4.2. Method for Selecting Independent Model Variables

In order to verify hypotheses 2, 3, 4, 5, 6 we needed to build machine learning models of message credibility evaluations in our experiment.

First step toward the creation of machine classification models consisted in the selection of independent variables. While it would be possible to use the MEC from all Brodman Areas in all time intervals to define independent variables, such a model would most likely be overfitted, and would also have a minimal capacity for interpretation. Therefore, we followed a special method for selecting a smaller subset of independent variables of the model based on the MEC from various Brodman Areas.

All steps of the method for selecting independent variables are shown in **Figure 3**. First step is state-of-the-art EEG signal processing and the use of the sLORETA algorithm for source location. In contrast to our earlier research (Wojcik et al., 2018; Kwaśniewicz et al., 2020) that used the mean electric charge (MEC) flowing through each BA, in our new approach, the MEC signal was normalized to the range from 0 to 1 for each BA of every participant. Brain cortex is covered with the mantle of meninges, bones of the skull, skin and hair which results in a different SNR. Therefore, the electrophysiological activity of particular participants may be different in its measured power. To avoid the impact of such individual differences, we propose to normalize the

MEC values. The normalized (scaled) MEC will be referred to as sMEC.

In the next step, the dataset was divided into training and validating datasets. There were 78 participants in the training dataset and 27 participants in the validating one for TT and FF classification, 78 (training) and 26 (validating) for ST and SF classification, and 74 (training) and 21 (validating) for LT and LF classification. This split was performed to avoid any possible data leakage during the testing of the models. The remaining steps of independent variable selection were done on the training dataset. This means that the model was designed and trained using only training data, and finally tested on a separate testing dataset.

An important step within the method was the selection of time intervals for MEC calculation. Time intervals were selected taking into account all possible lengths and offsets from stimulus, with a resolution of 5 ms. This means that a considered time interval could be [0, 25 ms] from stimulus, or [0, 990 ms] (the entire duration of the experiment)—as well as all possible combinations of the beginning and ending time, with a resolution of 5 ms. The choice of 5 ms intervals was then found to be optimal for classifier efficiency in such a temporal resolution turned out to be the best for achieving satisfactory classification results. Choosing higher time resolution would result in averaging the MEC changes and the smaller resolution would be too tiny for observation statistically significant differences in activity.

It was assumed that the difference in decision making would be reflected in the signal of brain activity. It was not known how long it took the participants to make a decision or when exactly it was made before the response pad was clicked. The 5 ms interval turned out to be insufficient to follow the decision. The comparison of sMEC from different time intervals made it possible to select those intervals with the largest mean value differences for the selected classes. The selection of the time intervals for which the sMEC was calculated was made only from the training data set, preventing data leakage. The validation of the classifiers on the testing set with good accuracy results, positively verified the correctness of selecting such explanatory variables.

For every selected time interval, we calculated sMEC for all Brodmann Areas, for all responses. Next, we chose up to 5 Brodmann Areas that had the largest mean absolute difference in sMEC for positive and negative credibility evaluations (for the two classified cases). We limited the maximum number of explanatory variables to 5, because in our previous studies, logistic regression models using five Brodmann's Areas had been as accurate as those using more areas (Kwaśniewicz et al., 2020).

For 5 Brodmann Areas, there could be 39,175,752 models (all combinations without repetitions of 5 Brodmann Areas out of 88) for a single time period. The number of time intervals to be checked depends on the selected length, e.g., for 25 ms we had 194 intervals to be checked for timeline 0–900 ms and the smallest possible interval as 5 ms. From the above selected lengths, we obtained 2,261 time intervals, and 391,758,752 combinations. This means that there were 88,576,375,272 models to check. Since this number was too large, we limited the number of combinations of Brodmann Areas by only considering the 5 Brodmann Areas with the largest mean absolute difference in sMEC for a given time interval between the two classes as in the following equation:

$$N = I \sum_{i=1}^{Q} \binom{Q}{i} \tag{4}$$

where Q = 5 and I = 2,261.

This limiting approach left us with possible 70,091 models. Out of these, we selected the model with the best accuracy that was estimated with the use of the bootstrap validation method (1,000 repetitions) on the training data. Finally, the accuracy of the chosen model was evaluated on the validating dataset. It should be note that the models could have up to five explanatory variables that are sMEC values from the chosen Brodmann Areas in the selected time interval.

4.3. Machine Classification Models of Message Credibility Evaluations

The logistic regression classifier was implemented in R language using the stats v3.6.3 library.

Results are shown in **Table 1**. The best models achieved an accuracy of at least 0.7 on the training and validating datasets, which confirms hypothesis 2. Overall, the best results were achieved for the classification of the ST and SF cases (short notes without knowledge). However, the best models for the different classification problems differ significantly with respect to the time interval and the Brodmann Areas selected to determine the independent model variables.

In particular, different time intervals and different Brodmann Areas were used in the best model for long messages, and best model for the short ones, which supports the Hypothesis 3. Similarly, Brodmann Areas and time intervals differ for the best models for known and unknown messages, which confirms the Hypothesis 4.

The validation results of individual classifiers are shown in the form of confusion matrices, ROC curves and Area Under the Curve on **Figures 4–6**.

All areas used in the classifiers are shown in **Figure 7**. In the case when participants evaluated message credibility without prior knowledge, the best model used a signal in the range of 105-330 ms. The model used Brodmann Areas such as:

- BA.46 which is the area in which Fleck et al. (2006) observed increased activity in decision making under uncertainty
- BA.36—activity of this area was observed when listening to a foreign language (Perani et al., 1996)
- BA.02, BA.04—motor sensor areas, changes in these areas have been observed associated with error processing in the context of visual feedback (Wilson et al., 2019)
- Another area is BA.33—a part of Anterior cingulate cortex. Basic theory states that the Anterior cingulate cortex is involved in error detection (Bush et al., 2000). Evidence for this conclusion has been derived from studies involving a Stroop task (Posner and DiGirolamo, 1998).

When the participants had knowledge about the meaning of Japanese characters, the model used the signal from a later time interval: 330–530 ms and Brodmann Areas such as:

TABLE 1 | Best models' scores for each case used for verification of hypothesis 2.

Message case	Intervals (ms)	Brodmann areas	Bootstrap accuracy	Validation accuracy	Validation precision	Validation recall	Validation f1
Short note UNKNOWN (ST and SF)	105–330	R.BA.46, R.BA.0402 R.BA.36,L.BA.33 L.BA.	0.70	0.79	0.78	0.81	0.79
Short note KNOWN (TT and FF)	330–530	R.BA.39, L.BA.370.71 L.BA.25,R.BA.31	0.71	0.70	0.69	0.74	0.71
Long note UNKNOWN (LT and LF)	830–855	L.BA.31, L.BA.44 R.Hippocampus L.BA.09, R.BA.08	0.70	0.74	0.67	0.78	0.72

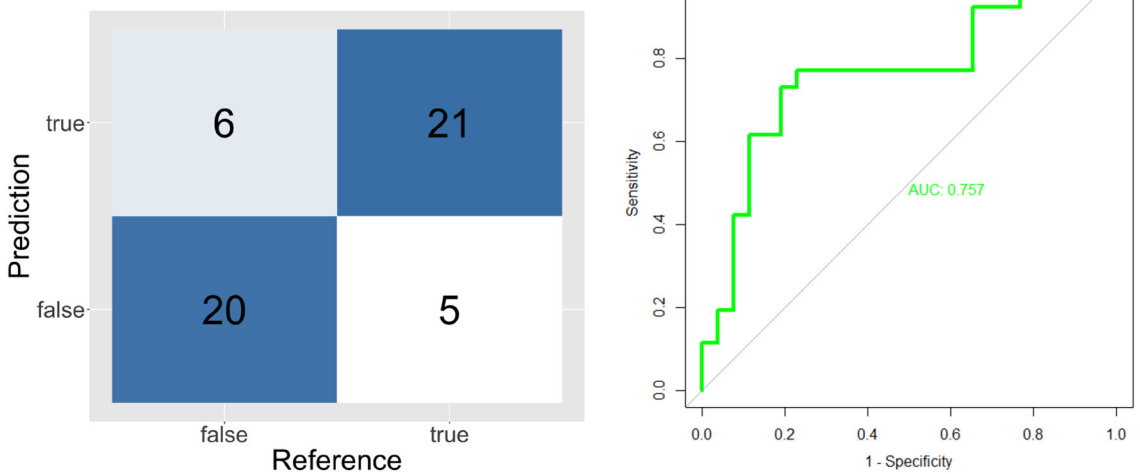

FIGURE 4 | Confusion matrix and ROC curve for short unknown message's model (ST and SF).

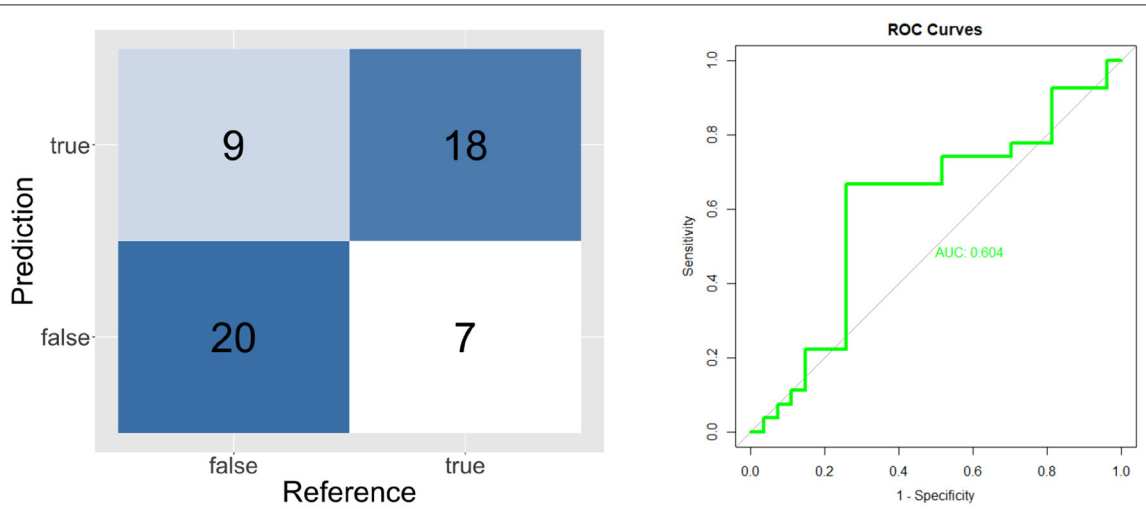

FIGURE 5 | Confusion matrix and ROC curve for short known message's model (TT and FF).

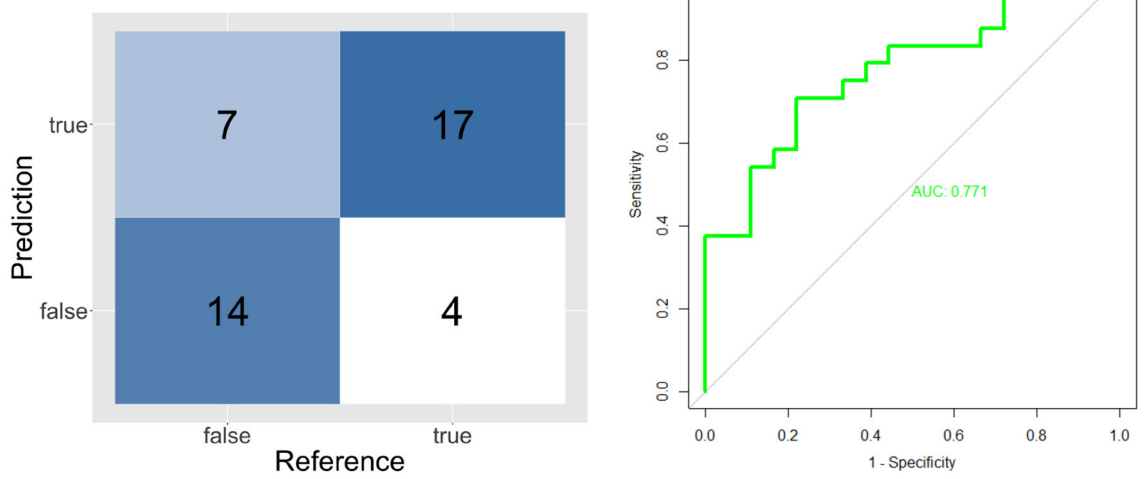

FIGURE 6 | Confusion matrix and ROC curve for long unknown message's model (LT and LF).

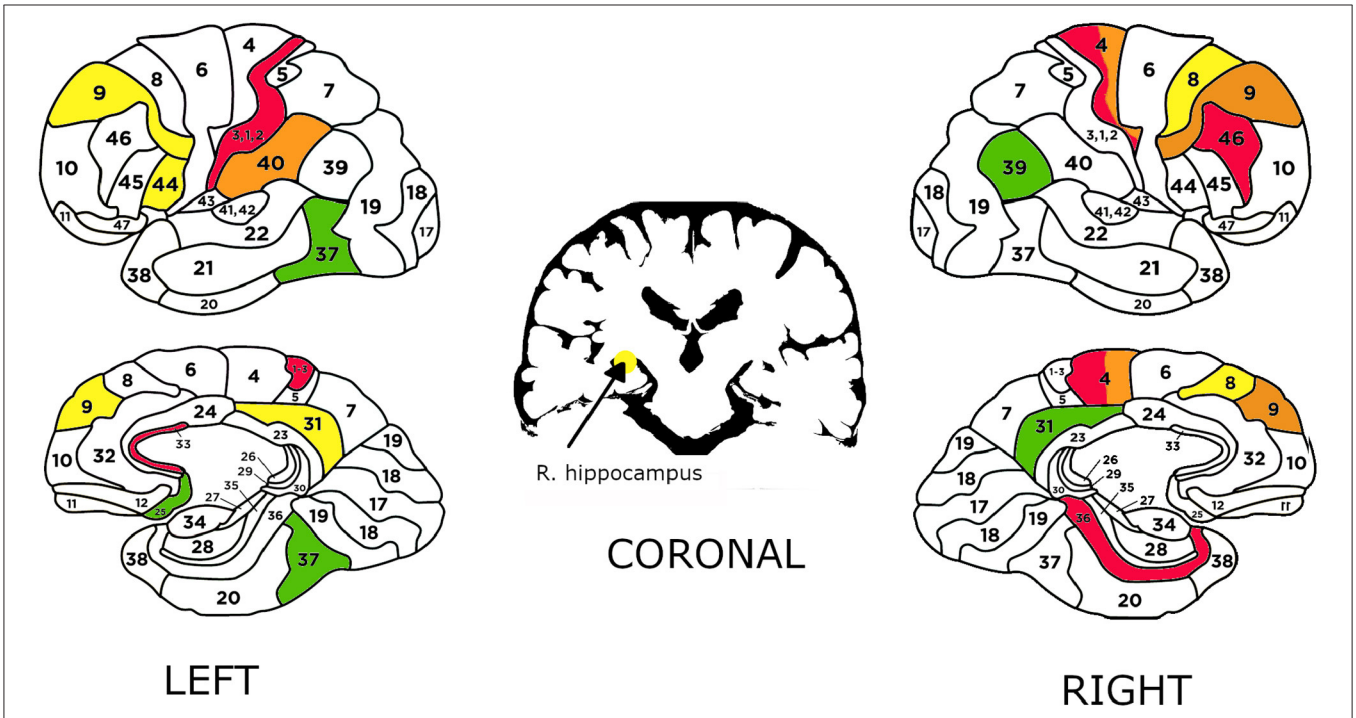

FIGURE 7 | Brodmann areas (left and right view) and hippocampus (coronal view) used in the best classifiers. The colors: green, yellow, orange, and red correspond to the classifiers in the following order: TT and FF, LT and LF (entire population), LT and LF (long note prefered), ST and SF.

- BA37—a common node of two distinct networks-visual recognition (perception) and semantic language functions (Ardila et al., 2015)
- BA.39—involved in language reception and understanding (Ardila et al., 2016)
- BA.31—studied in case of decision-making in perceptual decisions (Heekeren et al., 2004)
- BA.25—considered a governor for a vast network involving areas like hypothalamus, brain stem, amygdala, and hippocampus (Ressler and Mayberg, 2007)
- BA.44—involved in processing of different types of linguistic information (Heim et al., 2009).

Also, areas like BA.08, BA.09, BA.31, Hippocampus have been linked to decision making

TABLE 2 | Models for classification of message credibility evaluations of long notes without prior knowledge (LT and LF cases) for participants who preferred long notes and for the entire population.

Case and group	Intervals (ms)	Brodmann areas	Bootstrap accuracy	Validation accuracy	Val. precision	Val. recall	Val. f1
LT and LF long note preferred	800–875	L.BA.40, R.BA.04 R.BA.09	0.71	0.73	0.7	0.7	0.7
LT and LF entire population	830–855	L.BA.31, L.BA.44 R.Hippocampus L.BA.09, R.BA.08	0.70	0.74	0.67	0.78	0.72

Boostrap and validation accuracies are presented in columns 4 and 5.

TABLE 3 | Coefficients in logistic regression for best models.

ST and SF		LT and LF		TT and FF	
Variables	**Coefficients**	**Variables**	**Coefficients**	**Variables**	**Coefficients**
Intercept	−3.23286	Intercept	0.4133	Intercept	−2.82279
R.BA.46	0.08884	L.BA.31	0.4360	R.BA.39	0.09693
R.BA.04	0.05689	L.BA.44	−0.3214	L.BA.37	0.03943
R.BA.36	0.06321	R.Hippocampus	−0.2790	L.BA.25	0.06532
L.BA.33	0.02682	L.BA.09	0.4657	R.BA.31	−0.02018
L.BA.02	0.07420	R.BA.08	−0.4305		

(Maddock et al., 2003; Heekeren et al., 2004; Deppe et al., 2005; Volz et al., 2005; O'Neil et al., 2015).

Areas such as BA.08, BA.09 have been proposed to participate in a general mechanism for perceptual decision-making in the human brain (Heekeren et al., 2004).

4.3.1. Cognitive Bias for Long Notes

In our experiment, message credibility evaluation could be affected by an irrelevant factor: the length of the message. Hypothesis 5 concerns the existence of differences in the models obtained for participants who frequently evaluated long note as credible, as compared to other participants. To verify the hypothesis, we needed to identify participants who evaluated long notes as credible more frequently than short notes. In order to do this, we calculated for each participant how many times that participant evaluated long notes as credible (LC) and how many times the same participant evaluated short notes as credible in the absence of prior knowledge (SC). Next, for each participant we calculated the ratio *LC*/*SC*. This ratio indicates how much more frequently the participant evaluated long notes as credible, as compared to short notes.

Forty-seven participants had a *LC*/*SC* ratio above the mean value of all participants. We have followed the same procedures as described above to build a model that classified the cases LT and LF for participants in this group.

The best model for the classification of *LT* and *LF* for participants who prefer long notes is different from the model for all participants, as shown in **Table 2** and its best coefficients in **Table 3**. The model for participants who prefer long notes does not use the areas L.BA.31, L.BA.44, R.Hippocampus, R.BA.08. Both models also use other language networks.

BA40 is part of a language reception/understanding system and it is involved in language associations (associating words with other information) (Ardila et al., 2016). BA44—language production system. BA 44 supports modality-independent lexical decision making (Heim et al., 2007). Hippocampus is associated with declarative memory, including the memory of facts (Squire, 1992). Overall findings showed that Hippocampus may be part of a larger cortical and subcortical network seen to be important in decision making in uncertain conditions (O'Neil et al., 2015), some studies show that activity in regions BA 9, BA 31, increases with increasing trust as well (Emonds et al., 2014). Research also indicates activation of BA 8 reflects that we are uncertain (Volz et al., 2005).

These results may indicate smaller uncertainty and distrust associated with a negative message credibility evaluation of a long note by participants who prefer long notes. Further work should consist in creating a model using the EEG signal to classify the confidence level of the participants who evaluate long notes as credible.

Changes in areas BA.9, BA.25, BA.31, BA.44 (Altshuler et al., 2008; Klempan et al., 2009; Goodman et al., 2011; Alexander et al., 2019), may also be related to mental illnesses such as depression. Kim et al. (2012) indicates that low interpersonal trust appears to be an independent risk factor for new-onset and long-term depression.

4.4. Discussion and Limitations

For each case—short known message, short unknown message, and long unknown message, we received different best models that used sMEC from differing time intervals and Brodmann Areas.

The Brodmann Areas, used by our models, confirm previous studies in which the activity of these areas was observed in decision-making processes under uncertainty, or it was related to language processing.

Our method of model design relies on choosing time intervals that had the largest differences in brain activity for the classified cases. Optimal time intervals chosen for classifying pairs of cases: ST and SF, TT and TF, LT and LF were: 105–330, 330–530, and 830–855 ms, respectively. Significant differences in

these time intervals correspond to the nature of the task for each pair of cases. First pair, ST and SF, is message credibility evaluation of short notes without prior knowledge. In this case, models used the earliest time interval, which indicates that brain activity crucial for making decisions occurred quite quickly after the stimulus. On the other hand, second pair of cases, TT and TF, involved message credibility evaluation with prior knowledge. In this case participants needed more time, likely for comparing messages to memorized facts and making decisions based on comparison results. Last pair of cases based on message credibility evaluation of long notes without prior knowledge. Late time interval was most likely caused by the subjects having to read a longer text.

Using the same time intervals as in our models for the same cases, we created models consisting of the signal from the left BA08 and BA09 given as areas involved in the decision making process. The results we obtained were less effective than the results of the models presented in **Table 1**: short known: accuracy = 0.56, precision = 0.56, recall = 0.52, f1 = 0.54 short unknown: accuracy = 0.58,precision = 0.57, recall = 0.62, f1 = 0.59 long unknown: accuracy = 0.64, precision = 0.59, recall = 0.55, f1 = 0.57. The models obtained from our research have a much higher ability to predict message credibility evaluations, which seems to be a specific brain process that cannot be explained as general decision making. This means that Hypothesis 6 is not supported.

We had designed the experiment of message credibility evaluation with consideration for internal validity. The task of Japanese language (Kanji signs) translation was chosen in order to exclude confounding variables, such as participant prior knowledge, experience or opinions on the subject of the evaluated message. We were also able to verify internal validity by evaluating the impact of one of the main independent variables (length of the note) on the participant's decisions (as stated by hypothesis 1). The experiment also had low attrition rate, as over 90% of participants completed the experiment.

The comparison of the models that predicted decisions in the cases with full knowledge (TT and FF) to the models that predicted cases with no knowledge, but under the influence of message design, shows great differences in the Brodmann areas selected by the machine learning algorithm as most significant for the prediction. This means that our results would allow to determine whether a person that evaluates message credibility bases this evaluation on what he knows (or believes that he knows). Note that in our experiment, participants accepted our translation of the three Kanji signs used in the full knowledge cases without verifying them).

In order to evaluate the external validity of our experiment, we can compare the results to results of a pilot experiment (Kawiak et al., 2020). The pilot experiment involved a different and smaller set of participants (57). It had a similar setting, but the long note and the short note were presented on a single screen. Hypotheses in the pilot experiment were similar to the ones described in this article, and their verification results are the same.

The Brodmann Areas used in the best classifiers for both experiments were different, because only one large time interval was taken into account in the pilot study, while in the main experiment we tested different smaller time intervals.

In the pilot study, short messages and long messages were displayed on the screen at the same time, making it impossible to determine whether the participant's response was connected to a positive message credibility evaluation of the first note or a negative evaluation of the second note (and vice versa). We have redesigned the main experiment to overcome this shortcoming by displaying short and long messages separately. This difference makes it difficult to directly compare models and results obtained in the pilot study and the main experiment. The results reported here regarding differences in brain activity during message credibility evaluation with and without prior knowledge are also missing in the pilot study (Kawiak et al., 2020).

Nevertheless, in the pilot study message credibility evaluation models used areas related to language processing and word comprehension such as BA38, BA39, BA40. In both experiments, there were classifiers that used areas like BA39—language reception and understanding and BA46—decision making under uncertainty.

Overall, we consider that the preliminary results in the pilot experiment confirmed the external validity of our experiment, because the results of hypothesis verification based on both experiments were the same.

Our experiment had several limitations. First, only right-handed, young men who were university students of a technical subject were included in our sample.

Second, our experiment controlled and limited the factors that could influence credibility evaluation. Only message credibility was available to experiment participants, who did not know the source of the message. While this setting resembled a situation in which a Web user evaluates credibility of content by an unknown author, the experimental setting was still very limiting. Other factors, such as the message look, persuasiveness, or emotional content, could influence message credibility. A limitation of the new method for selecting independent variables was the use of the same time interval for all Brodmann Areas for the sMEC calculations. The next step should be to search for different time intervals for different Brodmann Areas in a single model.

5. CONCLUSION AND FUTURE WORK

Our results indicate that by using source localization algorithm (sLORETA) and an easy-to-interpret logistic regression algorithm, we can demonstrate and make use of the difference in brain activity during the decision making process to classify message credibility evaluations. This is an important first step toward a deeper understanding of human credibility evaluations. For instance, consider the credibility evaluation of debunking information designed to counteract fake news. The findings from this study can be used to guide the design of future experiments with a panel of judges who would evaluate the credibility of fake news or debunking information. Our results allow to determine, by observing the brain activity of such a judge, whether he made a credibility evaluation based on his knowledge (or what he believes to know), or not.

The next step is to investigate other cognitive biases that can affect message credibility evaluations, and to learn how to detect

them using brain activity measurements. Our results show that this is a promising research direction.

In future work, we will also study the activity of parts of the brain in different time intervals and different frequency bands for each part of the brain, and to build even better classifiers of message credibility evaluation using more advanced models.

The study by Douglas et al. (2013) proposed models using various spectral bands of the collected signal, characterized by good accuracy. These models were not based on source location but on wavelets of EEG signals. In further work, we can also use different bands to increase the number of explanatory variables in the model.

The results of machine learning algorithms other than logistic regression may turn out to be better than the results of models presented in this article, at the expense of reduced interpretability. This is another direction of our future research.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this manuscript will be made available by the authors, without undue reservation, to any qualified researcher.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the Maria Curie-Sklodowska University's Bioethical Commission (MCSU Bioethical Commission permission 13.06.2019). The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

All authors listed have made a substantial, direct and intellectual contribution to the work, and approved it for publication.

ACKNOWLEDGMENTS

The content of this manuscript has been presented in part at the International Conference in Computational Science, (Kwaśniewicz et al., 2020). The experiment presented herein is different than presented during the conference and research conducted on different cohort. Finally, results presented herein are then different from the results presented during the conference as they refer to new experiment.

REFERENCES

Alexander, L., Clarke, H. F., and Roberts, A. C. (2019). A focus on the functions of area 25. *Brain Sci*. 9:129. doi: 10.3390/brainsci9060129

Altshuler, L., Bookheimer, S., Townsend, J., Proenza, M. A., Sabb, F., Mintz, J., et al. (2008). Regional brain changes in bipolar i depression: a functional magnetic resonance imaging study. *Bipolar Disord*. 10, 708–717. doi: 10.1111/j.1399-5618.2008.00617.x

Appelman, A., and Sundar, S. S. (2016). Measuring message credibility: construction and validation of an exclusive scale. *J. Mass Comm. Quart*. 93, 59–79. doi: 10.1177/1077699015606057

Ardila, A., Bernal, B., and Rosselli, M. (2015). Language and visual perception associations: meta-analytic connectivity modeling of brodmann area 37. *Behav. Neurol*. 2015:565871. doi: 10.1155/2015/565871

Ardila, A., Bernal, B., and Rosselli, M. (2016). How localized are language brain areas? A review of brodmann areas involvement in oral language. *Arch. Clin. Neuropsychol*. 31, 112–122. doi: 10.1093/arclin/acv081

Borah, P., and Xiao, X. (2018). The importance of "likes": The interplay of message framing, source, and social endorsement on credibility perceptions of health information on Facebook. *J. Health Commun*. 23, 399–411. doi: 10.1080/10810730.2018.1455770

Bush, G., Luu, P., and Posner, M. I. (2000). Cognitive and emotional influences in anterior cingulate cortex. *Trends Cogn. Sci*. 4, 215–222. doi: 10.1016/S1364-6613(00)01483-2

Casado-Aranda, L.-A., Dimoka, A., and Sánchez-Fernández, J. (2019). Consumer processing of online trust signals: a neuroimaging study. *J. Interact. Market*. 47, 159–180. doi: 10.1016/j.intmar.2019.02.006

Deitz, G. D., Royne, M. B., Peasley, M. C., and Coleman, J. T. (2016). EEG-based measures versus panel ratings: predicting social media-based behavioral response to super bowl ads. *J. Advert. Res*. 56, 217–227. doi: 10.2501/JAR-2016-030

Deppe, M., Schwindt, W., Kugel, H., Plassmann, H., and Kenning, P. (2005). Nonlinear responses within the medial prefrontal cortex reveal when specific implicit information influences economic decision making. *J. Neuroimag*. 15, 171–182. doi: 10.1111/j.1552-6569.2005.tb00303.x

Douglas, P., Lau, E., Anderson, A., Kerr, W., Head, A., Wollner, M. A., et al. (2013). Single trial decoding of belief decision making from EEG and fMRI data using independent components features. *Front. Hum. Neurosci*. 7:392. doi: 10.3389/fnhum.2013.00392

Emonds, G., Declerck, C. H., Boone, C., Seurinck, R., and Achten, R. (2014). Establishing cooperation in a mixed-motive social dilemma. An fMRI study investigating the role of social value orientation and dispositional trust. *Soc. Neurosci*. 9, 10–22. doi: 10.1080/17470919.2013.858080

Filkowski, M. M., Anderson, I. W., and Haas, B. W. (2016). Trying to trust: brain activity during interpersonal social attitude change. *Cogn. Affect. Behav. Neurosci*. 16, 325–338. doi: 10.3758/s13415-015-0393-0

Fleck, M. S., Daselaar, S. M., Dobbins, I. G., and Cabeza, R. (2006). Role of prefrontal and anterior cingulate regions in decision-making processes shared by memory and nonmemory tasks. *Cereb. Cortex* 16, 1623–1630. doi: 10.1093/cercor/bhj097

Forgas, J. P., and Baumeister, R. (2019). *The Social Psychology of Gullibility: Conspiracy Theories, Fake News and Irrational Beliefs*. New York, NY: Routledge. doi: 10.4324/9780429203787

Goodman, M., Hazlett, E. A., Avedon, J. B., Siever, D. R., Chu, K.-W., and New, A. S. (2011). Anterior cingulate volume reduction in adolescents with borderline personality disorder and co-morbid major depression. *J. Psychiatr. Res*. 45, 803–807. doi: 10.1016/j.jpsychires.2010.11.011

Guixeres, J., Bigné, E., Ausín Azofra, J. M., Alcañiz Raya, M., Colomer Granero, A., Fuentes Hurtado, F., et al. (2017). Consumer neuroscience-based metrics predict recall, liking and viewing rates in online advertising. *Front. Psychol*. 8:1808. doi: 10.3389/fpsyg.2017.01808

Harrison, G. P. (2017). *Think Before You Like: Social Media's Effect on the Brain and the Tools You Need to Navigate Your Newsfeed*. Buffalo: Prometheus Books.

Heekeren, H. R., Marrett, S., Bandettini, P. A., and Ungerleider, L. G. (2004). A general mechanism for perceptual decision-making in the human brain. *Nature* 431, 859–862. doi: 10.1038/nature02966

Heekeren, H. R., Wartenburger, I., Schmidt, H., Schwintowski, H.-P., and Villringer, A. (2003). An fMRI study of simple ethical decision-making. *Neuroreport* 14, 1215–1219. doi: 10.1097/00001756-200307010-00005

Heim, S., Eickhoff, S. B., Ischebeck, A. K., Friederici, A. D., Stephan, K. E., and Amunts, K. (2009). Effective connectivity of the left BA 44, BA 45, and inferior temporal gyrus during lexical and phonological decisions identified with DCM. *Hum. Brain Mapp*. 30, 392–402. doi: 10.1002/hbm.20512

Heim, S., Eickhoff, S. B., Ischebeck, A. K., Supp, G., and Amunts, K. (2007). Modality-independent involvement of the left BA 44 during lexical decision making. *Brain Struct. Funct*. 212, 95–106. doi: 10.1007/s00429-007-0140-6

Higashino, T., and Wakamiya, N. (2021). Verification and proposal of information processing model using eeg-based brain activity monitoring. *Int. J. Psychol. Behav. Sci*. 15, 10–17. Available online at: https://publications.waset.org/pdf/10011731

Hovland, C. I., and Weiss, W. (1951). The influence of source credibility on communication effectiveness. *Publ. Opin. Quart*. 15, 635–650. doi: 10.1086/266350

Kahneman, D. (2011). *Thinking, Fast and Slow*. Macmillan.

Kakol, M., Jankowski-Lorek, M., Abramczuk, K., Wierzbicki, A., and Catasta, M. (2013). "On the subjectivity and bias of web content credibility evaluations," in *Proceedings of the 22nd International Conference on World Wide Web* (New York, NY: ACM), 1131–1136. doi: 10.1145/2487788.2488133

Kakol, M., Nielek, R., and Wierzbicki, A. (2017). Understanding and predicting web content credibility using the content credibility corpus. *Inform. Process. Manage*. 53, 1043–1061. doi: 10.1016/j.ipm.2017.04.003

Kawiak, A., Wójcik, G. M., Kwasniewicz, L., Schneider, P., and Wierzbicki, A. (2020). "Look who's talking: modeling decision making based on source credibility," in *International Conference on Computational Science* (Cham: Springer), 327–341. doi: 10.1007/978-3-030-50371-0_24

Kerz, E., Qiao, Y., and Wiechmann, D. (2021). "Language that captivates the audience: predicting affective ratings of ted talks in a multi-label classification task," in *Proceedings of the Eleventh Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis*, 13–24.

Kim, S.-S., Chung, Y., Perry, M. J., Kawachi, I., and Subramanian, S. (2012). Association between interpersonal trust, reciprocity, and depression in South Korea: a prospective analysis. *PLoS ONE* 7:e30602. doi: 10.1371/journal.pone.0030602

Klempan, T. A., Sequeira, A., Canetti, L., Lalovic, A., Ernst, C., Turecki, G., et al. (2009). Altered expression of genes involved in ATP biosynthesis and gabaergic neurotransmission in the ventral prefrontal cortex of suicides with and without major depression. *Mol. Psychiatry* 14, 175–189. doi: 10.1038/sj.mp.4002110

Krueger, F., McCabe, K., Moll, J., Kriegeskorte, N., Zahn, R., Strenziok, M., et al. (2007). Neural correlates of trust. *Proc. Natl. Acad. Sci. U.S.A*. 104, 20084–20089. doi: 10.1073/pnas.0710103104

Kwaśniewicz, Ł., Wójcik, G. M., Kawiak, A., Schneider, P., and Wierzbicki, A. (2020). "How you say or what you say? Neural activity in message credibility evaluation," in *International Conference on Computational Science* (Cham: Springer), 312–326. doi: 10.1007/978-3-030-50371-0_23

Larsen, T., and O'Doherty, J. P. (2014). Uncovering the spatio-temporal dynamics of value-based decision-making in the human brain: a combined fMRI-EEG study. *Philos. Trans. R. Soc. B Biol. Sci*. 369:20130473. doi: 10.1098/rstb.2013.0473

Liu, X., Nielek, R., Adamska, P., Wierzbicki, A., and Aberer, K. (2015). Towards a highly effective and robust web credibility evaluation system. *Decis. Support Syst*. 79, 99–108. doi: 10.1016/j.dss.2015.07.010

Maddock, R. J., Garrett, A. S., and Buonocore, M. H. (2003). Posterior cingulate cortex activation by emotional words: fMRI evidence from a valence decision task. *Hum. Brain Mapp*. 18, 30–41. doi: 10.1002/hbm.10075

Meijer, E. H., and Verschuere, B. (2017). Deception detection based on neuroimaging: better than the polygraph? *J. Forens. Radiol. Imag*. 8, 17–21. doi: 10.1016/j.jofri.2017.03.003

Mulert, C., Seifert, C., Leicht, G., Kirsch, V., Ertl, M., Karch, S., et al. (2008). Single-trial coupling of EEG and fMRI reveals the involvement of early anterior cingulate cortex activation in effortful decision making. *Neuroimage* 42, 158–168. doi: 10.1016/j.neuroimage.2008.04.236

O'Neil, E. B., Newsome, R. N., Li, I. H., Thavabalasingam, S., Ito, R., and Lee, A. C. (2015). Examining the role of the human hippocampus in approach-avoidance decision making using a novel conflict paradigm and multivariate functional magnetic resonance imaging. *J. Neurosci*. 35, 15039–15049. doi: 10.1523/JNEUROSCI.1915-15.2015

Perani, D., Dehaene, S., Grassi, F., Cohen, L., Cappa, S. F., Dupoux, E., et al. (1996). Brain processing of native and foreign languages. *NeuroRep. Int. J. Rapid Commun. Res. Neurosci*. 7, 2439–2444. doi: 10.1097/00001756-199611040-00007

Posner, M. I., and DiGirolamo, G. J. (1998). Executive attention: conflict, target detection, and cognitive control. *Variet. Attent*. Available online at: https://www.researchgate.net/profile/Michael-Posner-2/publication/232582595_Executive_attention_Conflict_target_detection_and_cognitive_control/links/565ddb7a08aeafc2aac8a40f/Executive-attention-Conflict-target-detection-and-cognitive-control.pdf

Rafalak, M., Abramczuk, K., and Wierzbicki, A. (2014). "Incredible: is (almost) all web content trustworthy? Analysis of psychological factors related to website credibility evaluation," in *Proceedings of the 23rd International Conference on World Wide Web* (New York, NY), 1117–1122. doi: 10.1145/2567948.2578997

Ressler, K. J., and Mayberg, H. S. (2007). Targeting abnormal neural circuits in mood and anxiety disorders: from the laboratory to the clinic. *Nat. Neurosci*. 10, 1116–1124. doi: 10.1038/nn1944

Rule, N. O., Krendl, A. C., Ivcevic, Z., and Ambady, N. (2013). Accuracy and consensus in judgments of trustworthiness from faces: behavioral and neural correlates. *J. Pers. Soc. Psychol*. 104:409. doi: 10.1037/a0031050

Rutjens, B. D., and Brandt, M. D. (2018). *Belief Systems and the Perception of Reality*. London: Routledge. doi: 10.4324/9781315114903

Sakurai, Y., Momose, T., Iwata, M., Sudo, Y., Ohtomo, K., and Kanazawa, I. (2000). Different cortical activity in reading of kanji words, kana words and kana nonwords. *Cogn. Brain Res*. 9, 111–115. doi: 10.1016/S0926-6410(99)00052-X

Sharma, K., Qian, F., Jiang, H., Ruchansky, N., Zhang, M., and Liu, Y. (2019). Combating fake news: a survey on identification and mitigation techniques. *ACM Trans. Intell. Syst. Technol*. 10, 1–42. doi: 10.1145/3305260

Squire, L. R. (1992). Memory and the hippocampus: a synthesis from findings with rats, monkeys, and humans. *Psychol. Rev*. 99:195. doi: 10.1037/0033-295X.99.2.195

Tseng, S., and Fogg, B. (1999). Credibility and computing technology. *Commun. ACM* 42, 39–44. doi: 10.1145/301353.301402

Viviani, M., and Pasi, G. (2017). Credibility in social media: opinions, news, and health information–a survey. *Wiley Interdisc. Rev*. 7:e1209. doi: 10.1002/widm.1209

Volz, K. G., Schubotz, R. I., and von Cramon, D. Y. (2005). Variants of uncertainty in decision-making and their neural correlates. *Brain Res. Bull*. 67, 403–412. doi: 10.1016/j.brainresbull.2005.06.011

Wang, H., Chang, W., and Zhang, C. (2016). Functional brain network and multichannel analysis for the p300-based brain computer interface system of lying detection. *Expert Syst. Appl*. 53, 117–128. doi: 10.1016/j.eswa.2016.01.024

Wawer, A., Nielek, R., and Wierzbicki, A. (2014). "Predicting webpage credibility using linguistic features," in *Proceedings of the 23rd International Conference on World Wide Web* (New York, NY), 1135–1140. doi: 10.1145/2567948.2579000

Wierzbicki, A. (2018). *Web Content Credibility*. Cham: Springer. doi: 10.1007/978-3-319-77794-8

Wilson, N. R., Sarma, D., Wander, J. D., Weaver, K. E., Ojemann, J. G., and Rao, R. P. (2019). Cortical topography of error-related high-frequency potentials during erroneous control in a continuous control brain-computer interface. *Front. Neurosci*. 13:502. doi: 10.3389/fnins.2019.00502

Wojcik, G. M., Masiak, J., Kawiak, A., Schneider, P., Kwasniewicz, L., Polak, N., et al. (2018). New protocol for quantitative analysis of brain cortex electroencephalographic activity in patients with psychiatric disorders. *Front. Neuroinformatics* 12:27. doi: 10.3389/fninf.2018.00027

Conflict of Interest: The 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.

frontiers in Human Neuroscience

ORIGINAL RESEARCH
published: 28 October 2021
doi: 10.3389/fnhum.2021.763821

Psychophysiological, but Not Behavioral, Indicator of Working Memory Capacity Predicts Video Game Proficiency

Natalia Jakubowska[1,2]*, *Paweł Dobrowolski*[3], *Alicja Anna Binkowska*[1], *Ibrahim V. Arslan*[1], *Monika Myśliwiec*[1] *and Aneta Brzezicka*[1]*

[1]Department of Psychology, SWPS University of Social Sciences and Humanities, Warsaw, Poland, [2]Polish-Japanese Academy of Information Technology, Warsaw, Poland, [3]Institute of Psychology, Polish Academy of Sciences, Warsaw, Poland

Visual working memory (VWM) is the ability to actively maintain visual information over short periods of time and is strongly related to global fluid intelligence and overall cognitive ability. In our study, we used two indices of visual working memory capacity: the behavioral estimate of capacity (K) and contralateral delay activity (CDA) in order to check whether training in a Real-Time Strategy (RTS) video game StarCraft II can influence the VWM capacity measured by the change detection task. We also asked a question whether individual differences in behavioral and psychophysiological indices of VWM can predict the effectiveness of video game training. Sixty-two participants (non-players) were recruited to the experiment. Participants were randomly assigned to either experimental (Variable environment), active control (Fixed environment), and passive control groups. Experimental and active control groups differed in the type of training received. Training consisted of 30 h of playing the StarCraft II game. Participants took part in two EEG sessions (pre- and post-training) during which they performed the VWM task. Our results showed that working memory capacity (K calculated according to Pashler's formula) increases after training in both experimental groups, but not in a control group. We have also found a correlation between average visual working memory capacity (calculated as K) and mean CDA amplitude no matter which group we are looking at. And, last but not least, we have found that we can predict the amount of improvement in the RTS video game by looking at the psychophysiological indices (CDA amplitude) recorded at baseline (before training), but only in the experimental group. We think that the strength of the psychophysiological indicator of VWM capacity might be a marker of the future success in video game acquisition.

Keywords: action video games, visual working memory, trainings, ERPs, EEG

OPEN ACCESS

Edited by:
Aleksandra Dagmara Kawala-Sterniuk, Opole University of Technology, Poland

Reviewed by:
Santiago Galdo-Alvarez, University of Santiago de Compostela, Spain
Sergio Iglesias-Parro, University of Jaén, Spain

***Correspondence:**
Natalia Jakubowska
njakubowska@swps.edu.pl
Aneta Brzezicka
abrzezi2@swps.edu.pl

Specialty section:
This article was submitted to Brain Imaging and Stimulation, a section of the journal Frontiers in Human Neuroscience

Received: *24 August 2021*
Accepted: *12 October 2021*
Published: *28 October 2021*

Citation:
Jakubowska N, Dobrowolski P, Binkowska AA, Arslan IV, Myśliwiec M and Brzezicka A (2021) Psychophysiological, but Not Behavioral, Indicator of Working Memory Capacity Predicts Video Game Proficiency. Front. Hum. Neurosci. 15:763821. doi: 10.3389/fnhum.2021.763821

INTRODUCTION

Visual working memory (VWM) allows us to maintain visual information over short periods of time for manipulation or later access (Baddeley, 2003; D'Esposito and Postle, 2015). VWM is an important cognitive function in our daily life and is essential for many higher-level cognitive processes, like problem-solving, learning by observation, or reading (Fukuda et al., 2010;

Shipstead et al., 2012). The capacity of VWM relates to the amount of visual information, which can be maintained in memory simultaneously and accessible if needed (Luck and Vogel, 2013). Previous research (including neuroimaging studies) has shown that VWM capacity is highly limited (Luck and Vogel, 1997; Todd and Marois, 2004), differs across individuals (Rouder et al., 2008), and predicts fluid intelligence in adults (Fukuda et al., 2010; Unsworth et al., 2014). Studies on VWM have relied on a well-established paradigm that measures VWM capacity—the *change detection task* (Luck and Vogel, 1997, 2013), where participant maintains a visual image in memory over a short delay interval and answers if any item (or items) in a later probe image have changed compared to the sample image. The number of items presented (memory load) is manipulated, and performance (working memory capacity, an estimate of the number of items stored in WM measured by K calculated according to Pashler's formula in our study) is compared between trials of different loads. Change detection accuracy mirrors a participant's limitation of VWM capacity and is usually limited to 3–4 items (Vogel and Awh, 2008). It is suggested that the limitation of VWM capacity is associated with visual search and multiple-object tracking performance (Drew et al., 2011; Luria and Vogel, 2011). Previous research has shown that participants with higher VWM capacity are more effective in ignoring unnecessary items during task performance (Vogel et al., 2005). In neurophysiological studies of lateralized VWM, stimuli are presented peripherally, and the subject's task is to attend and maintain in VWM only the items presented in a cued visual hemifield. This generates a lateralized representation, which is larger contralateral compared to ipsilateral of the memorized hemifield, in posterior cortical areas over the retention period that results in a contralateral delay activity (CDA). CDA is a negative slow-wave evoked component that amplitude relates to the number of objects maintained in VWM, so it could be interpreted as a neural index of WM load (Vogel and Machizawa, 2004; Luria et al., 2016). Previous research has shown that CDA amplitude is correlated with memory capacity (Vogel and Machizawa, 2004; Ikkai et al., 2010) and can be modified as a result of WM training (Li et al., 2017). In this study, we used video games as a specific kind of cognitive training having the potential for VWM improvement.

The growing body of research suggests that playing video games enhances the performance on tasks measuring visual and attentional abilities (Green and Bavelier, 2007; Jakubowska et al., 2021). Potential cognitive benefits are possible even with relatively short periods of engagement in playing activity (Green and Bavelier, 2007; Wilms et al., 2013), which makes video games an attractive training option for restoring cognitive functions following brain impairments and in preventive cognitive interventions (Achtman et al., 2008). As there are different kinds of video games, the particular category called action video gaming (AVG) is thought to have a substantial impact on human cognitive functioning. AVG requires players to scan many different complex visual stimuli at the same time and react to multiple stimuli or situations under time pressure (Green and Bavelier, 2003, 2012). AVG is cognitively demanding because of engaging many cognitive functions like working memory, visual attention, and inhibitory control (Green and Bavelier, 2003, 2012). Previous research has shown that long experience in AVG was associated with VWM improvement measured with a change detection task (Boot et al., 2008; Blacker et al., 2014; Li et al., 2015) as well as other tasks (Colzato et al., 2010; Sungur and Boduroglu, 2012; Waris et al., 2019). These results suggest that AVG training may lead to the enhancement of VWM. At the same time, VWM is a key cognitive function in effective video gaming, because it allows players to keep task-relevant visual stimuli over short periods of time for manipulation or later access (Logie, 2011; Blacker et al., 2014). Noteworthy, some studies suggest that cognitive enhancement connected to video game playing does not show far transfer's characteristics (like general improvement in cognitive functioning or learning), but seems to be limited to functions being involved in a given type of video game (Oei and Patterson, 2014).

It is important to note that there are studies that have not found a cognitive improvement after gaming training (Seçer and Satyen, 2014; Dominiak and Wiemeyer, 2016). The possible explanation of these divergent results could be connected to different kinds of games being considered as AVG is actually a broad category with wide inclusion criteria. The study conducted by Dobrowolski et al. (2015) has shown that the achievement of expertise in two different game genres, while both included in AVG category called real-time strategy (RTS) and first-person shooter (FPS), impacts differently cognitive functioning of players. The higher performance in task engaging visual attention and task-switching ability were observed only in RTS (but not in FPS) players as compared to non-players (Dobrowolski et al., 2015). Similarly, RTS experts seem to have higher accuracy and larger VWM capacity than non-experts (Yao et al., 2020). The possible interpretation of these results is that video gaming-related cognitive benefits may depend on the type of actions performed within the game (Dobrowolski et al., 2015). As RTS gaming requires extensive interaction with the complex visual environments, we assume it is highly possible to improve VWM through training with this type of video game. While previous investigations indicate that AVG experts have larger visual attentional capacities, greater capacity of working memory, and higher visual acuity as compared to non-gamers (Green and Bavelier, 2003, 2012; Oei and Patterson, 2013), and that specific AVG can positively affect the level of a given function (Bejjanki et al., 2014; Choi et al., 2020), the impact of the initial level of cognitive functioning on player performance remains largely unexplored.

That is why we decided to use the (RTS) video game StarCraft II with two different types of environments requiring diverse cognitive workloads. Our training types were based on either variable or fixed game environments. The opponent's faction and strategy varied in the variable environment group only (and it was connected to the higher level of difficulty). Our participants were randomly assigned to either a Variable environment, a Fixed environment or the control group. Then it's important to mention that differences between variable and fixed training models were investigated in previous studies, which proved

that variable training enhances learning rates and retention, and induce transfer to untrained tasks more, effectively than fixed training (Kramer et al., 1999; Bherer et al., 2008; Erickson et al., 2010). Moreover, training based on a variable environment seems to have more in common—than training with a fixed environment—with people's gaming experiences in everyday life.

The objective of the current study was to investigate the impact of RTS video game StarCraft II training on VWM capacity by comparing training groups' and control group's behavioral (k estimate of WM capacity) and ERP (contralateral delay activity) data in a change detection task. Furthermore we were also interested in whether initial, individual differences in behavioral and psychophysiological indices of VWM can predict the effectiveness of video game training, which could extend our knowledge of the relationship between VWM and in-game performance.

MATERIALS AND METHODS

Participants

A total of 104 participants were recruited online *via* a covert questionnaire (Sobczyk et al., 2015). As a result of: (1) resignation (n = 13); (2) wrong hardware configuration (n = 7); (3) failure to meet all training objectives (n = 6); (4) bad quality of data (n = 7); and (5) lost data (n = 9) only 62 of participants were included in analyses reported here. Participants were randomly assigned to two training groups: with Variable Environment training (VEG; n = 22; 12 males; M_{age} = 25.05, SD_{age} = 2.97), with Fixed Environment training model (FEG; n = 21; 8 males; M_{age} = 25.33, SD_{age} = 3.01), and to two control groups: passive control (PC) group (n = 8; 5 males, M_{age} = 24.63, SD_{age} = 2.97), that did not receive any training and active control (AC) group (n = 11; males = 8; M_{age} = 25.55, SD_{age} = 4.41). The participants played Heart Stone for 30 h (8 h in the laboratory and 22 h at home). As the size of the control groups was inappropriate to analyze them individually, and neither 4 × 2 × 2 repeated measures ANOVAs with Load and Session as the within-subjects factors and Group as the between subject factor, nor One-way ANOVAs with Group as a factor showed any between group differences on behavioral or neurophysiological levels, we decided to merge the groups into one Control group (CG; n = 19; 13 males; M_{age} = 25.16, SD_{age} = 3.80). Then it is important to mention that dropout, which largely contributed to the reduction of the size of the control groups, is a common problem in longitudinal training studies (e.g., Moore et al., 2017). Furthermore, our study employed restrictive recruitment criteria, especially in terms of experience in video game playing, which finally resulted in an inability to re-complete the control groups. All participants reported normal or corrected-to-normal visual acuity, normal color vision and normal hearing. They were right-handed and reported not being on any medications, no history of neurological or psychiatric disorders and injuries, including no previous head trauma, no previous head or neck surgery, and no brain tumors. All participants declared less than 5 h of video games played per week over the past 6 months and no experience with Real Time Strategy or First Person Shooter games. Informed consent was obtained from each participant before the start of the experimental procedure.

Procedure

The study design and the informed consent form were approved by the Ethics Committee of the SWPS University of Social Sciences and Humanities. The research consisted of three steps: (1) Pre-training measurement of cognitive function *via* change detection task (Visual Working Memory task; VWM); (2) Training sessions applied to active groups; and (3) Post-training measurement (**Figure 1**). Experimenters were present during all meetings. Measurement and training sessions took place in the laboratories of the SWPS University in Warsaw.

Experimental Procedure

Prior to the beginning of the experiment, participants were verbally instructed as to what they would be experiencing and were shown what the procedure of EEG electrode mounting entails. Then, after signing a consent form, participants were brought into a laboratory setting and seated in front of a 24 inch BenQ XL2411Z computer monitor (1,920 × 1,080 resolution, 100 Hz refresh rate) at a distance of 60 cm. Electrodes were then mounted and participants were briefly shown the EEG signal and explained how it is affected by eye blinks and muscular movements, which was a part of the procedure aimed at minimizing the number of artifacts in the signal. The procedure was then started, and upon its completion subjects were provided with a place to wash their hair. The entire procedure lasted no more than 2 h and was identical during both measurements. All subjects, who fulfilled training requirements and participated in both measurement sessions, were compensated for their participation with approx. 184 USD after post-training measurement.

Experimental Task—Change Detection Task Paradigm

The experimental task was based on the procedure outlined by Vogel and Machizawa (2004). An initial fixation cross was followed by an arrow, pointing which side of the screen needs to be attended (whether right or left hemifield), after which a pattern (memory array) of two to five colored squares appeared in each hemifield of the screen. The same array appeared again (test array) after a brief retention interval, with a 50% chance that one of the squares in the cued hemifield changed its color. Participants were tasked with detecting changes between the memory and test array by responding with one of the keys (same or different). Square colors were chosen at random from seven possibilities (red, blue, violet, green, yellow, black, white), with the constraint that one color appeared no more than twice in a given test array. Squares (0.65 × 0.65 visual degrees) were randomly positioned at the start of each trial in two 4 deg. × 7.4 deg. hemifields (centered 3 deg. to the left and right of a central fixation, light gray background), with a minimum 2 deg. (center to center) distance between squares. All participants completed 576 trials (144 per load) of the task along with 16 initial practice trials.

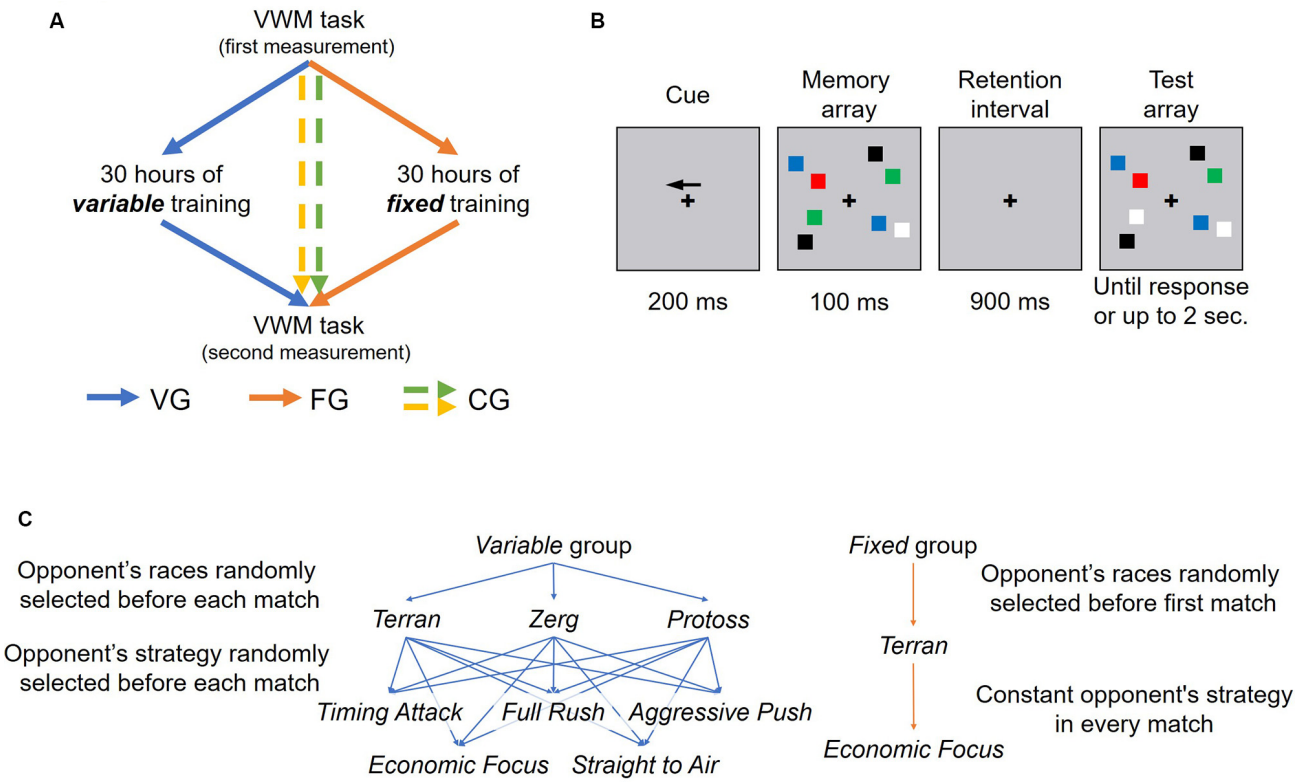

FIGURE 1 | (A) Study design: two measurement sessions were carried out during the study (pre-training and post-training). Training included 30 h of playing in the real-time strategy game (*StarCraft II*), spread over 4 weeks. Training varied depending on the group. The Control group (CG) merges participants from Passive Control, who participated only in the measurement sessions and Active Control, who additionally played Heart Stone for 30 h (8 h in the laboratory and 22 h at home). **(B)** The visual-working memory task. Participants directed their attention to a cued hemifield (left of right, guided by an arrow at the beginning of each trial) and compared two arrays of colored squares (memory and test arrays) separated by a retention interval. The test array was either identical to the memory array (no-change condition) or differed by one color (change condition). Participants answered whether the two arrays were identical or not. **(C)** While all of the participants from training groups played as a Terran faction during training, the opponent's race and strategy varied according to the training group type. Participants from the Variable Environment Group (VEG) could match three factions, from each could use one of five strategies. The faction and the strategy were randomly selected before each match for the Variable group. In the case of a Fixed Environment Group (FEG), participants always played against the Terran faction, which used an economic strategy.

Training

StarCraft II Training

The StarCraft II (SC2) training consisted of 30 h of training time over a 4-week period. Training consisted of playing matches (approx. 20 min each) against SC2's artificial intelligence (AI), and all matches were played at our laboratory. Training objectives required the participants to train a minimum of 10 h per week, but no more than 5 h per day. This was done to avoid excessive skew in the distribution of training hours across the training period. There were also two possible training types: Fixed and Variable. The exact differences between the types of training are described below and were presented in **Figure 1**.

Participants had to access an online platform before each match in order to receive configuration parameters; the parameters consisted of the difficulty setting, the opponent's faction, the opponent's strategy, and the game map. Participants from both groups played all of their matches as a Terran faction. While the map was randomly selected from 14 maps before each match in both—Fixed and Variable—training versions, the opponent's faction, and strategy only varied in the Variable group. The Fixed group always faced the same faction (Terran), and their opponent always applied a more passive "Economic Focus" strategy. The Variable group could face any of the three factions (each with their own unique units and abilities) and also any of five opponent strategies: Full Rush, Timing Attack, Aggressive Push, Economic Focus, Straight to Air. The game difficulty was set adaptively for both training types spanning across eight levels (1. Very Easy; 2. Easy; 3. Medium; 4. Hard; 5. Harder; 6. Very Hard; 7. Elite; 8. Cheater) The online platform software recorded the number of wins (+1) and losses (−1) and each time the total passed the multiple of four threshold, the difficulty was increased by one. The difficulty decreased whenever the total dropped below the multiple of four threshold. The training was preceded by an introduction phase designed to familiarize participants with the core concepts of the game and basic gameplay mechanics (see "StarCraft II Introduction" section).

Starcraft II Introduction

The introductory phase consisted of eight parts: (1) a short text describing the goals of the meeting; (2) a text and video-based description of the overall game; (3) a video introduction to the Terran faction, its units and buildings; (4) a text-based description of the fundamental game concepts and in-game interface; (5) an AI guided tutorial that introduces the gameplay in real time, allowing participants to experience the game for the first time; (6) a quiz requiring that the correct labels be attached to each of the five basic unit and building types that are available to the Terran faction, which was intended to check if participants were attentive to the training materials; (7) two films (25 min. each) describing basic strategies and explaining the various stages that each match progresses through; and (8) a three-match series in which the game progressively increased its difficulty, speed, and available units, with no specific guiding instructions. The entire introduction lasted approx. 2.5 h, and did not count into the required 30 h of training. It was also automated and self-paced, with experimenters only providing assistance when needed and also during part 8 of the introduction where assistance was provided to keep up the pace and direction of each training game. Upon completion of this introduction, participants were free to begin training on the following day.

EEG Recording and Analysis

A 64-channel SynAmps RT Neuroscan EEG amplifier and BrainProducts actiCap Ag/AG-Cl active electrode set were used to record brain activity during task performance. All channels were recorded at 1,000 Hz sampling rate. Impedances were held below 5 kΩ. All data were preprocessed offline using MATLAB environment and EEGlab (Delorme and Makeig, 2004), and ERPlab (Lopez-Calderon and Luck, 2014) software packages. The signal was initially re-referenced to a common average and then down-sampled to 250 Hz, followed by a band-pass filter between 0.1 and 40 Hz. Data epochs between −0.2 and 0.996 s were extracted, and all epochs with incorrect behavioral responses were rejected. The remaining epochs were manually filtered for eye-blinks/movements and excessive muscle activity and then averaged.

Data Reduction and Analysis

All analyses were conducted using R Statistical Software (Foundation for Statistical Computing, Vienna, Austria), IBM Corp. Released 2017. IBM SPSS Statistics for Windows, Version 25.0. Armonk, NY: IBM Corp, python and MATLAB custom scripts.

Mixed ANOVAs (3 × 4 × 2) were used to analyze the behavioral and neurophysiological data including the between group variables of group (three levels: CG *vs.* FEG *vs.* VEG) and the within group variables of load (four levels: 2 *vs.* 3 *vs.* 4 *vs.* 5) and session (two levels: pre-training measurement *vs.* post-training measurement). Group comparisons for telemetric data were conducted by a series of t-tests (two-group comparisons). *Post hoc* pairwise t-tests were also performed in case of significant main effects or interactions, with Bonferroni correction for multiple comparisons.

Telemetric data were collected from a total of 5,494 games. While SC2 replays allow obtaining dozens of different variables, participants' expertise or game results do not depend on any particular one. Nevertheless, we selected basic predictor variables that relate to cognitive-motor abilities and game proficiency. We focused on (1) the number of matches played by each player; (2) first army unit creation latency; and (3) first supply collection latency. As better SC2 players play shorter matches, the first of mentioned variables should reflect general players' proficiency. It should be emphasized that the number of played matches positively correlated with the number of won matches ($r = 0.964$, $p < 0.001$) and matches played on more difficult levels (Harder: $r = 0.458$, $p = 0.002$; Very hard: $r = 0.634$, $p < 0.001$, Elite: $r = 0.605$, $p < 0.001$; Cheater: $r = 0.595$, $p < 0.001$), but not easier ones (Very Easy: $r = -0.145$, $p = 0.354$; Easy: $r = -0.146$; $p = 0.351$; Medium: r -0.239, $p = 0.122$; Hard: $r = -0.019$; $p = 0.904$). Then it can be assumed that a higher number of played matches is due to players' higher skills rather than multiple lost matches. Latencies of first army unit creation and first supply collection relate to two key moments in the game environment, which faster execution should result in better performance in the game. We also calculated the overall time each player spent in the game environment which allowed us to confirm the fulfillment of training assumptions. All mentioned telemetric variables were tested for between-group differences by a series of t-tests.

For behavioral data, the capacity of visual working memory, which is measured by the K value, was calculated using the formula proposed by Pashler (1988),

where P(hit) = hits/(hits + misses), and P(FA) = false alarms/(false alarms + correct rejections). In addition to the K values of each set size, we also computed the average K value (K_{mean}) for each participant's visual working memory capacity.

For neurophysiological data, mean amplitudes of CDA (lateralized waveforms; contra—ipsi), averaged across P7/P8 electrodes, from 400 to 900 ms time window were outcome variables (**Figures 4A–C**).

To examine the relationship between behavioral, psychophysiological, and telemetric data, linear regression analyses were conducted.

RESULTS

Telemetric Data

We started by calculating the total time spent in the game and the mean number of played matches for each player. Although there were no significant difference between groups in time spent playing SC2 ($p = 0.513$), participants from Variable group were able to play significantly more matches in that time period (VEG: Mean = 99.68, $SD = 24.782$; FEG: Mean = 85.05, $SD = 14.5$); $t_{(34.152)} = 2.376$, $p = 0.023$ (**Figure 2**). Then we calculated the mean latencies of first army unit creation and first supply collection. Analysis revealed that participants from the Variable group created their army units significantly faster (VEG: Mean = 170.685, $SD = 40.505$; FEG: Mean = 219.243, $SD = 66.683$); $t_{(41)} = -2.901$, $p = 0.006$, but there were no differences in the latency of the first supply collection ($p = 0.696$).

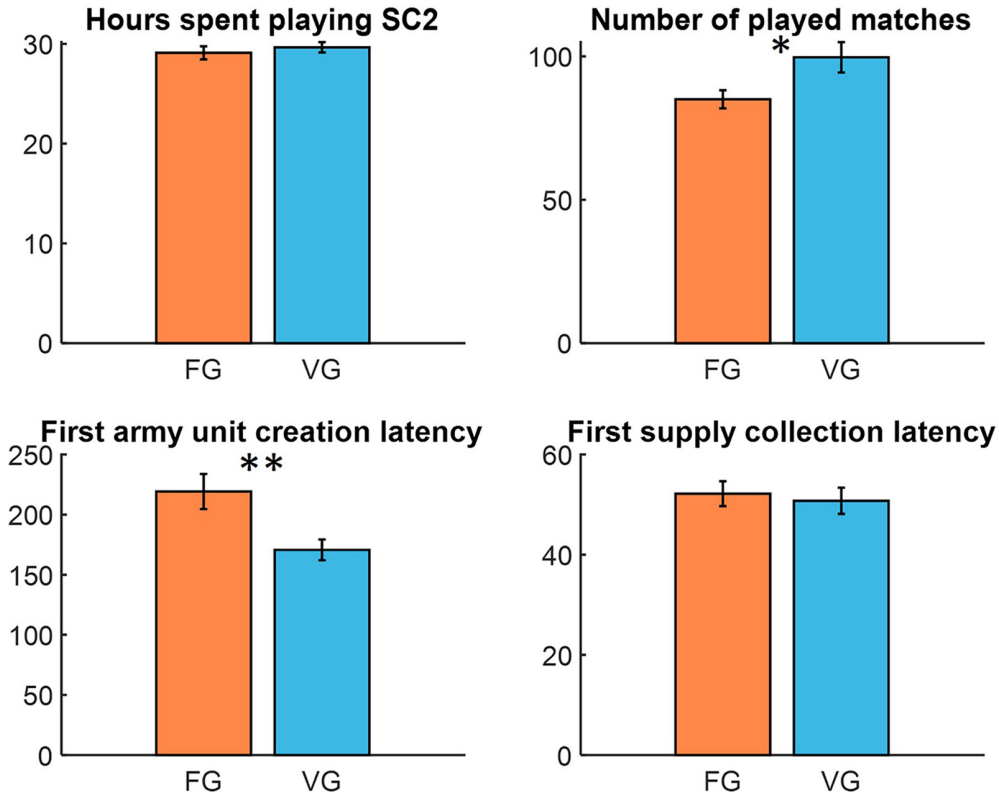

FIGURE 2 | Telemetric variables obtained from the StarCraft II (SC2) environment. While hours spent playing SC2 (upper left) allows us to confirm that both groups (VG: Variable Group; FG: Fixed Group) fulfill training assumptions, other variables indicate players' proficiency. Barplots presenting first army unit creation (lower left) and first supply collection (lower right) give latencies in seconds. Asterisks indicate statistical significance: $^{*}p < 0.05$, $^{**}p < 0.01$.

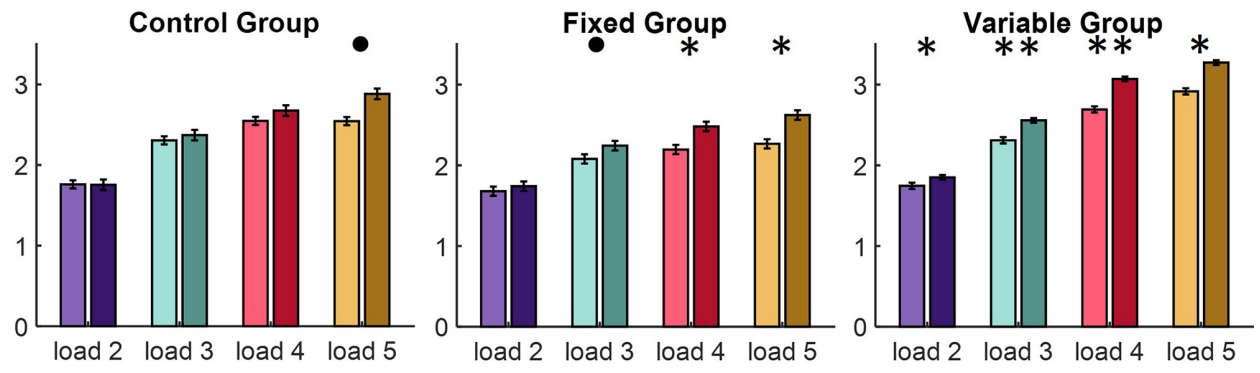

FIGURE 3 | The average K values for each set size in the two tests presented separately for each group. Lighter colors in the pair correspond to the pre-training measurement and darker to the post-training measurement. Asterisks indicate statistical significance: $^{\bullet}p < 0.08$, $^{*}p < 0.05$, $^{**}p < 0.01$.

Behavioral Data

The capacity of visual working memory, measured by the K values, were analyzed using a 4 (Load: load 2 *vs.* load 3 *vs.* load 4 *vs.* load 5) × 2 (Sessions: pre-training *vs.* post-training) × 3 (Group: Control *vs.* Fixed *vs.* Variable) repeated-measures ANOVA, with Load and Session as the within-subjects factors and Group as the between subject factor (**Figure 3**).

Analysis revealed the main effects of Group [$F_{(2, 59)} = 3.209$, $p = 0.048$, $\eta^2 = 0.1$], Load [$F_{(3, 57)} = 134.515$, $p < 0.001$, $\eta^2 = 0.49$], Session [$F_{(1, 59)} = 30.22$, $p < 0.001$, $\eta^2 = 0.07$], Load * Session interaction [$F_{(3, 57)} = 2.808$, $p = 0.039$, $\eta^2 = 0.02$] and Group * Load interaction [$F_{(6, 116)} = 3.992$, $p < 0.001$, $\eta^2 = 0.05$] but no Load * Session * Group interaction [$F_{(6, 116)} = 1.806$, $p = 0.104$, $\eta^2 = 0.085$] or Session * Group interaction [$F_{(2, 59)} = 0.541$, $p = 0.585$, $\eta^2 = 0.018$].

Additional analyses revealed that, while Control group wasn't able to significantly increase its capacity of visual working

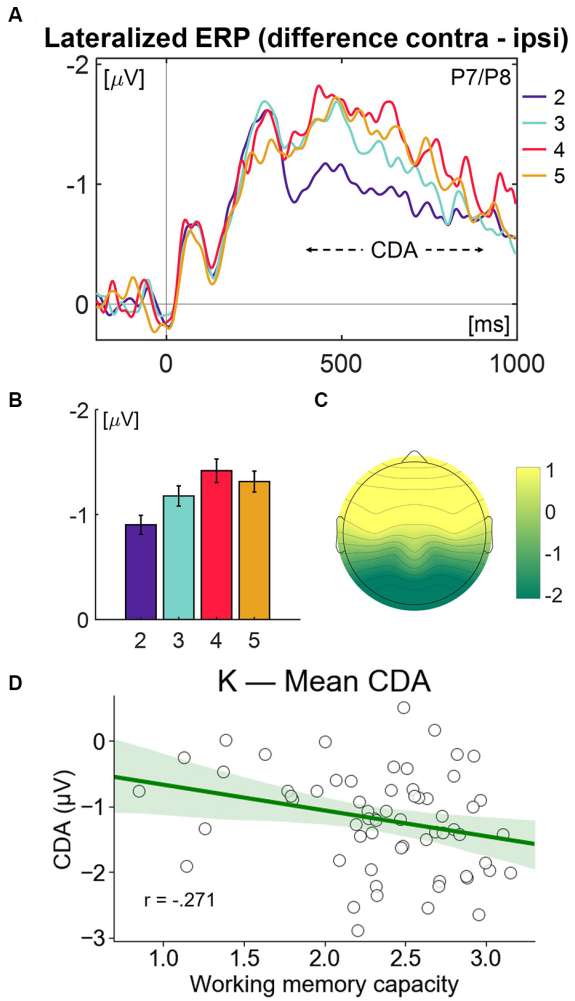

FIGURE 4 | (A) Grand average lateralized waveforms (contra—ipsi), averaged across P7/P8 electrodes, separately for all lateralized target distributions. For statistical analyses of CDA, the mean amplitude from 400 to 900 ms was used. **(B)** Mean CDA amplitude from 400 to 900 ms, separately for each load. Error bars denote standard errors of the mean, corrected with within-subjects comparisons. **(C)** Topography of the average activity at each electrode site from 400 to 900 ms. As values were averaged across paired electrodes, the topography is perfectly symmetrical. **(D)** Scatterplot of working memory capacity (K) averaged across loads and contralateral delay activity (CDA) averaged across loads.

memory at any of used loads, Fixed group increased it at the load 4 ($p = 0.025$) and load 5 ($p = 0.049$) and Variable group was able to significantly increase it at every load (load 2, $p = 0.029$; load 3, $p = 0.008$; load 4, $p = 0.003$, load 5, $p = 0.044$).

Psychophysiological Data

Contralateral delay activity was analyzed using a 4 (Load: load 2 vs. load 3 vs. load 4 vs. load 5) × 2 (Sessions: pre-training vs. post-training) × 3 (Group: Control vs. Fixed vs. Variable) repeated-measures ANOVA, with Load and Session as the within-subjects factors and Group as the between subject factor.

Analysis revealed that the only significant effect was the Load effect [$F_{(3, 57)} = 89$, $p < 0.001$, $\eta^2 = 0.288$], but no Session

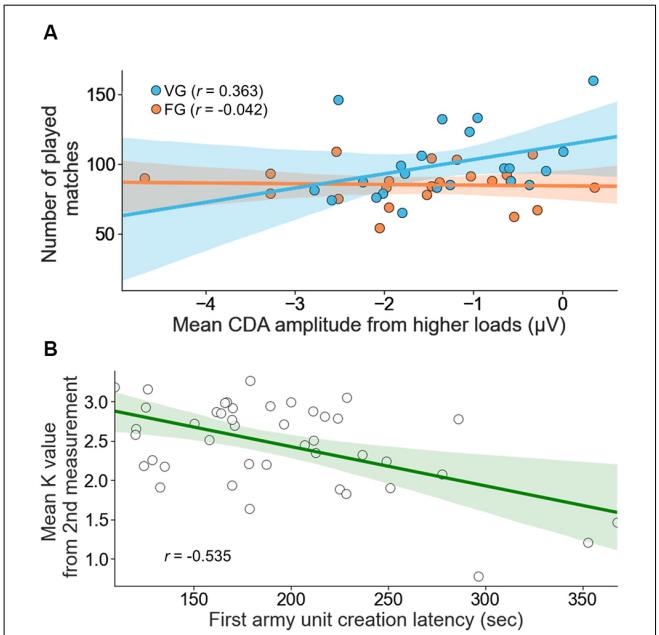

FIGURE 5 | In-game behaviors relationship to psychophysiological and behavioral variables. **(A)** Initial (from the 1st measurement point) contralateral delay activity (CDA) averaged across loads 4 and 5 predicts the overall number of played matches in the Variable Environment Group (VEG). Observed effect turned out to be insignificant in the Fixed Environment Group (FEG) only. **(B)** The averaged across training latency of first army unit creation predicts participants' mean K value obtained in post-training measurement. The effect was significant in both groups.

[$F_{(1, 59)} = 0.087$, $p = 0.769$, $\eta^2 = 0.002$], Group [$F_{(2, 59)} = 2.212$, $p = 0.118$, $\eta^2 = 0.02$], Load * Session interaction [$F_{(3, 57)} = 1.336$, $p = 0.272$, $\eta^2 = 0.066$], Load * Group interaction [$F_{(6, 116)} = 0.412$, $p = 0.87$, $\eta^2 = 0.021$] or Session * Group [$F_{(2, 59)} = 0.667$, $p = 0.517$, $\eta^2 = 0.022$].

Psychophysiological, Telemetric, and Behavioral Data Relations

In the next step, we created a model containing a mean contralateral delay activity (CDA) averaged across loads 4 and 5 obtained from pre-training measurement as a predictor, Group as a moderator variable and mean number of played matches as a dependent variable. Created model turned out to be significant [$F_{(3, 39)} = 3.387$, $p = 0.028$, $R^2 = 0.207$] and contained significant influence of the Group [$b = 29.077$, $t_{(39)} = 2.68$, $p = 0.011$] and tendency of interaction between CDA and Group [$b = 10.736$, $t_{(39)} = 1.734$, $p = 0.079$]. Next, it was revealed that while there was no relationship between CDA and number of played matches in the Fixed Group ($p = 0.891$), there was a significant negative relationship in the Variable Group: the smaller initial CDA amplitude averaged from loads 4 and 5, the more matches participants played [one unit decrease in the average CDA component's amplitude resulted in an increase of 10.219 matches played ($t_{(39)} = 2.077$, $p = 0.044$); **Figure 5A**].

In the final analysis, we created a model containing a mean latency of first army unit creation as a predictor, Group as a moderator variable, and mean K obtained from post-training

measurement as a dependent variable. Created model turned out to be significant [$F_{(3, 39)} = 8.384$, $p < 0.001$, $R^2 = 0.392$] and contained significant influence of the predictor [$b = -0.009$, $t_{(39)} = -2.499$, $p = 0.017$]. Group influence and interaction turned out to be insignificant ($p = 0.349$; $p = 0.12$; **Figure 5B**).

DISCUSSION

The study presented here examined the relationship between the RTS video game proficiency acquired during the training and the improvement of the VWM capacity indexed with behavioral and ERP measures. To properly inspect players' game proficiency, telemetric data from the game environment were used. EEG and behavioral data were collected from non-gamers, who were assigned to one of three groups (Control Group, Fixed Group, and Variable Group).

Participants completed a change detection task, which is the typical experimental paradigm used to examine the VWM capacity, twice during their participation (pre-training and post-training in active groups or over a period of 4 weeks in the passive control group).

The obtained results suggest that VWM capacity improvement was the most significant in the group of participants with the Variable training model. This finding stands in agreement with our initial hypothesis, which assumes that video game influence may vary depending on the training model.

Most importantly, our results show that we can successfully explain game performance by looking at the initial values of the psychophysiological index of VWM and also the behavioral index of VWM (mean K value) at the post-training measurement can be predicted from in-game behavior.

We believe that natural predispositions are an important aspect of achieving success in training, but a good training environment is no less crucial. Therefore, potential players can reach their full potential only under the right conditions. The combination of aspects of natural predispositions and different training models allows for a better understanding of differences in the obtained results, but above all—it shows how important it is to control game environment conditions, which can diversify the gameplay in an enormous number of ways.

VEG Participants Were Able to Achieve the Biggest Improvement of Their VWM Capacity During the Study

The participants from the group with the variable environment training model were able to significantly improve their VWM capacity (measured by Pashler's formula of K value) on each of the tested loads (from load 2 to load 5). This after-training improvement in accuracy stands in agreement with studies, which show that AVG experience is related to VWM abilities (Green and Bavelier, 2003; Boot et al., 2008; Colzato et al., 2010; Clark et al., 2011; Blacker and Curby, 2013; Oei and Patterson, 2013; Li et al., 2015). Still, the Fixed Environment Group had only a significant increase on load 4 and load 5. Then it is important to emphasize that AVG influence corresponds to applied game mechanisms: SC2 matches require players to rapidly switch between multiple sources of action and information in general, but the training's demands were different depending on the training's model. A similar effect was not observed in the control group. Presented results argue that variable training strategies can be more beneficial and allow not only to achieve bigger improvement in specific task but also the occurrence of the far transfer. The fact that VEG players were able to achieve the biggest improvement of their VWM capacity after their training is consistent with this interpretation. In contrast, FEG players were not encouraged to thoroughly explore the game environment, learn different strategies and maximize their various skills, but rather, were trained to repeat one gameplay model in a non-engaging way.

VEG Participants Were Able to Achieve the Biggest Game Proficiency

As mentioned above, three in-game indicators were chosen to measure game proficiency. (1) The number of played matches by each player; (2) latency of creating the first army unit; and (3) first supply collection latency. Telemetric data analysis shows us, even though there were no significant differences in groups about time spent on games, VEG players were able to play significantly more games in that period of time.

In comparison with FEG, VEG participants were significantly faster in creating their first army unit. However, there were no associations between the collection of first supply latency and group types.

These taken into account, we see that comparing with FEG, VEG settings allowed non-gamer participants to be greatly proficient in SC II.

CDA Component, K Value, and Game-Related Factor Analysis

Neurophysiological output was closely analyzed with all parameters using repeated-measure ANOVA. Analyses did not pinpoint significant association either for group type or session. Yet, the load variable had a significant effect on mean CDA amplitudes. This means we observed different CDA amplitudes on different loads. Our data support the notion that CDA is a VWM indicator (**Figure 4D**).

Additionally, the K value had a correlation with CDA. Therefore we understand that low-valued CDA components are significantly associated with both increased VWM capacity and increased input on VWM.

Game Proficiency Indicator Predicts VWM Capacity (K Value)

Two predictive models give us key insights about the relation between game performance, CDA, and K value obtained from the measurements. Model A holds a predictive value about the number of played SC II matches and the mean CDA amplitude on loads 4 and 5 (collected from pre-training session). Participants who have lower initial mean CDA amplitude are less likely to play a higher number of matches, which implies greater natural predispositions to succeed in the game environment. Then it needs to be

highlighted that this model was only found to be significant for VEG. It shows that players' natural predispositions can result in better in-game development only in a favorable environment.

Model B enables us to obtain information about participants' level of VWM (K-value obtained from post-training measurement) just by looking at the latency of creating the first army unit. Such a model could help us (in the future) not only to create a rule of thumb for measuring VWM in a specific setting but also to determine players' level of specific cognitive skills in a more natural environment.

Although performed analyzes did not reveal a significant model of moderated mediation, two independent regression models, it's important to interpret obtained results in a broader, common context. As a complex game environment can be reflected by dozens of telemetric variables, which only together make up the full picture of the match and players' skills, it may not be possible to create a simple and efficient model with only one telemetry variable.

Furthermore, initial VWN capacity, measured by K-value, didn't determine in-game performance regardless of the analyzed indicator. Then behavioral results obtained from pre-training measurement cannot be clearly associated with participants' natural predispositions. It should be noted then, that AVG requires more than one cognitive function, so the result of any single behavioral variable may turn out to be insufficient to fully reflect players' in-game proficiency or predispositions.

Presented models, taken together, hold promising results for both: RTS gaming's impact on VWM, and the role of neurophysiological indicators in recognizing the natural predispositions of AVG players. In conclusion, this study confirms that playing RTS games increases VWM capacity. As these improvements were majorly observed in VEG participants (yet still, FEG showed higher results in comparison with the control group), it can be assumed that the intensity of AVG influence depends on the adopted training model. What is more, in the presented study we propose a neurophysiological indicator, which may allow us to identify AVG players with higher predispositions to become better gamers. Last but not least: telemetric data sheds light on game performance, and combining it with other variables *via* regression models holds promising information as such, predicting the capacity of VWM (K-value, scored) from just one game proficiency indicator.

All these findings combined and experimental settings may hold a guiding reference for future research opportunities and commercial usage. Therefore it's important to mention that future investigations should examine a wider range of carefully selected tasks, which can contribute to create a more complete spectrum of cognitive functions and changes that they undergo through VG training.

DATA AVAILABILITY STATEMENT

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

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Komisja ds. Etyki Badań Naukowych Wydziału Psychologii w Warszawie [Ethics committee of Department of Psychology at University of Social Sciences and Humanities]. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

NJ: analyzed data, prepared figures, and wrote the first version of the manuscript. PD: designed the study and prepared paradigms' code. IA: helped with EEG data preparation and analysis. MM: collected data. ABr: study conceptualization, data interpretation, manuscript correction, and final approval. All authors contributed to the article and approved the submitted version.

FUNDING

This study was supported by the Polish National Science Centre, NCN (Narodowe Centrum Nauki) Grants 2016/23/B/HS6/03843, 2013/10/E/HS6/00186. Open access of this article was financed by the Ministry of Science and Higher Education in Poland under the 2019–2022 program "Regional Initiative of Excellence", project number 012/RID/2018/19.

REFERENCES

Achtman, R. L., Green, C. S., and Bavelier, D. (2008). Video games as a tool to train visual skills. *Restor. Neurol. Neurosci.* 26, 435–446.

Baddeley, A. (2003). Working memory: looking back and looking forward. *Nat. Rev. Neurosci.* 4, 829–839. doi: 10.1038/nrn1201

Bejjanki, V. R., Zhang, R., Li, R., Pouget, A., Green, C. S., Lu, Z., et al. (2014). Action video game play facilitates the development of better perceptual templates. *Proc. Natl. Acad. Sci. U S A* 111, 16961–16966. doi: 10.1073/pnas.1417056111

Bherer, L., Kramer, A. F., Peterson, M., Colcombe, S. J., Erickson, K. I., and Becic, E. (2008). Transfer effects in task-set cost and dual-task cost after dualtask training in older and younger adults: further evidence for cognitive plasticity in attentional control in late adulthood. *Exp. Aging Res.* 34, 188–219. doi: 10.1080/03610730802070068

Blacker, K. J., and Curby, K. M. (2013). Enhanced visual short-term memory in action video game players. *Attn. Percept. Psychophys.* 75, 1128–1136. doi: 10.3758/s13414-013-0487-0

Blacker, K. J., Curby, K. M., Klobusicky, E., and Chein, J. M. (2014). Effects of action video game training on visual working memory. *J. Exp. Psychol. Hum. Percept. Perform.* 40, 1992–2004. doi: 10.1037/a0037556

Boot, W. R., Kramer, A. F., Simons, D. J., Fabiani, M., and Gratton, G. (2008). The effects of video game playing on attention, memory and executive control. *Acta Psychol. (Amst)* 129, 387–398. doi: 10.1016/j.actpsy.2008.09.005

Choi, E., Shin, S.-H., Ryu, J.-K., Jung, K.-I., Kim, S.-Y., and Park, M.-H. (2020). Commercial video games and cognitive functions: video game genres and modulating factors of cognitive enhancement. *Behav. Brain Funct.* 16:2. doi: 10.1186/s12993-020-0165-z

Clark, K., Fleck, M. S., and Mitroff, S. R. (2011). Enhanced change detection performance reveals improved strategy use in avid action video

game players. *Acta Psychol. (Amst)* 136, 67–72. doi: 10.1016/j.actpsy.2010.10.003

Colzato, L. S., van Leeuwen, P. J. A., van den Wildenberg, W. P. M., and Hommel, B. (2010). DOOM'd to switch: superior cognitive flexibility in players of first person shooter games. *Front. Psychol.* 1:8. doi: 10.3389/fpsyg.2010.00008

D'Esposito, M., and Postle, B. R. (2015). The cognitive neuroscience of working memory. *Annu. Rev. Psychol.* 66, 115–142. doi: 10.1146/annurev-psych-010814-015031

Delorme, A., and Makeig, S. (2004). EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *J. Neurosci. Methods* 134, 9–21. doi: 10.1016/j.jneumeth.2003.10.009

Dobrowolski, P., Hanusz, K., Sobczyk, B., Skorko, M., and Wiatrow, A. (2015). Cognitive enhancement in video game players: the role of video game genre. *Comp. Hum. Behav.* 44, 59–63. doi: 10.1016/j.chb.2014.11.051

Dominiak, A., and Wiemeyer, J. (2016). "Training of spatial competencies by means of gesture-controlled sports games," in *Proceedings of the 10th International Symposium on Computer Science in Sports (ISCSS). Advances in Intelligent Systems and Computing*, (Cham: Springer), 392, 243–249. doi: 10.1007/978-3-319-24560-7_31

Drew, T., Horowitz, T. S., Wolfe, J. M., and Vogel, E. K. (2011). Delineating the neural signatures of tracking spatial position and working memory during attentive tracking. *J. Neurosci.* 31, 659–668. doi: 10.1523/JNEUROSCI.1339-10.2011

Erickson, K. I., Boot, W. R., Basak, C., Neider, M. B., Prakash, R. S., Voss, M. W., et al. (2010). Striatal volume predicts level of video game skill acquisition. *Cereb. Cortex* 20, 2522–2530. doi: 10.1093/cercor/bhp293

Fukuda, K., Vogel, E., Mayr, U., and Awh, E. (2010). Quantity, not quality: the relationship between fluid intelligence and working memory capacity. *Psychon. Bull. Rev.* 17, 673–679. doi: 10.3758/17.5.673

Green, C. S., and Bavelier, D. (2003). Action video game modifies visual selective attention. *Nature* 423, 534–537. doi: 10.1038/nature01647

Green, C. S., and Bavelier, D. (2007). Action-video-game experience alters the spatial resolution of vision: research article. *Psychol. Sci.* 18, 88–94. doi: 10.1111/j.1467-9280.2007.01853.x

Green, C. S., and Bavelier, D. (2012). Learning, attentional control and action video games. *Curr. Biol.* 22, R197–R206. doi: 10.1016/j.cub.2012.02.012

Ikkai, A., McCollough, A. W., and Vogel, E. K. (2010). Contralateral delay activity provides a neural measure of the number of representations in visual working memory. *J. Neurophysiol.* 103, 1963–1968. doi: 10.1152/jn.00978.2009

Jakubowska, N., Dobrowolski, P., Rutkowska, N., Skorko, M., Myśliwiec, M., Michalak, J., et al. (2021). The role of individual differences in attentional blink phenomenon and real-time-strategy game proficiency. *Heliyon* 7:e06724. doi: 10.1016/j.heliyon.2021.e06724

Kramer, A. F., Larish, J. F., and Strayer, D. L. (1999). Training for executive control: task coordination strategies and aging. *Atten. Perform.* 17, 616–652.

Li, X., Cheng, X., Li, J., Pan, Y., Hu, Y., and Ku, Y. (2015). Examination of mechanisms underlying enhanced memory performance in action video game players: a pilot study. *Front. Psychol.* 6:843. doi: 10.3389/fpsyg.2015.00843

Li, C. H., He, X., Wang, Y. J., Hu, Z., and Guo, C. Y. (2017). Visual working memory capacity can be increased by training on distractor filtering efficiency. *Front. Psychol.* 8:196. doi: 10.3389/fpsyg.2017.00196

Logie, R. H. (2011). The functional organization and capacity limits of working memory. *Curr. Dir. Psychol. Sci.* 20, 240–245. doi: 10.1177/0963721411415340

Lopez-Calderon, J., and Luck, S. J. (2014). ERPLAB: an open-source toolbox for the analysis of event-related potentials. *Front. Hum. Neurosci.* 8:213. doi: 10.3389/fnhum.2014.00213

Luck, S. J., and Vogel, E. K. (1997). The capacity of visual working memory for features and conjunctions. *Nature* 390, 279–281. doi: 10.1038/36846

Luck, S. J., and Vogel, E. K. (2013). Visual working memory capacity: from psychophysics and neurobiology to individual differences. *Trends Cogn. Sci.* 17, 391–400. doi: 10.1016/j.tics.2013.06.006

Luria, R., Balaban, H., Awh, E., and Vogel, E. K. (2016). The contralateral delay activity as a neural measure of visual working memory. *Neurosci. Biobehav. Rev.* 62, 100–108. doi: 10.1016/j.neubiorev.2016.01.003

Luria, R., and Vogel, E. K. (2011). Visual search demands dictate reliance on working memory storage. *J. Neurosci.* 31, 6199–6207. doi: 10.1523/JNEUROSCI.6453-10.2011

Moore, C. M., MaWhinney, S., Forster, J. E., Carlson, N. E., Allshouse, A., Wang, X., et al. (2017). Accounting for dropout reason in longitudinal studies with nonignorable dropout. *Stat. Methods Med. Res.* 26, 1854–1866. doi: 10.1177/0962280215590432

Oei, A. C., and Patterson, M. D. (2013). Enhancing cognition with video games: a multiple game training study. *PLoS One* 8:e58546. doi: 10.1371/journal.pone.0058546

Oei, A. C., and Patterson, M. D. (2014). Are videogame training gains specific or general? *Front. Syst. Neurosci.* 8:54. doi: 10.3389/fnsys.2014.00054

Pashler, H. (1988). Familiarity and visual change detection. *Percept. Psychophys.* 44, 369–478. doi: 10.3758/bf03210419

Rouder, J. N., Morey, R. D., Cowan, N., Zwilling, C. E., Morey, C. C., and Pratte, M. S. (2008). An assessment of fixed-capacity models of visual working memory. *Proc. Natl. Acad. Sci. U S A* 105, 5975–5979. doi: 10.1073/pnas.0711295105

Seçer, I., and Satyen, L. (2014). Video game training and reaction time skills among older adults. *Activ. Adaptation Aging* 38, 220–236. doi: 10.1080/01924788.2014.935908

Shipstead, Z., Redick, T. S., and Engle, R. W. (2012). Is working memory training effective? *Psychol. Bull.* 138, 628–654. doi: 10.1037/a0027473

Sobczyk, B., Dobrowolski, P., Skorko, M., Michalak, J., and Brzezicka, A. (2015). Issues and advances in research methods on video games and cognitive abilities. *Front. Psychol.* 6:1451. doi: 10.3389/fpsyg.2015.01451

Sungur, H., and Boduroglu, A. (2012). Action video game players form more detailed representation of objects. *Acta Psychol. (Amst)* 139, 327–334. doi: 10.1016/j.actpsy.2011.12.002

Todd, J. J., and Marois, R. (2004). Capacity limit of visual short-term memory in human posterior parietal cortex. *Nature* 428, 751–754. doi: 10.1038/nature02466

Unsworth, N., Fukuda, K., Awh, E., and Vogel, E. K. (2014). Working memory and fluid intelligence: capacity, attention control and secondary memory retrieval. *Cogn. Psychol.* 71, 1–26. doi: 10.1016/j.cogpsych.2014.01.003

Vogel, E. K., and Awh, E. (2008). How to exploit diversity for scientific gain: Using individual differences to constrain cognitive theory. *Curr. Dir. Psychol. Sci.* 17, 171–176. doi: 10.1111/j.1467-8721.2008.00569.x

Vogel, E. K., and Machizawa, M. G. (2004). Neural activity predicts individual differences in visual working memory capacity. *Nature* 428, 748–751. doi: 10.1038/nature02447

Vogel, E. K., McCollough, A. W., and Machizawa, M. G. (2005). Neural measures reveal individual differences in controlling access to working memory. *Nature* 438, 500–503. doi: 10.1038/nature04171

Waris, O., Jaeggi, S. M., Seitz, A. R., Lehtonen, M., Soveri, A., Lukasik, K. M., et al. (2019). Video gaming and working memory: a large-scale cross-sectional correlative study. *Comput. Hum. Behav.* 97, 94–103. doi: 10.1016/j.chb.2019.03.005

Wilms, I. L., Petersen, A., and Vangkilde, S. (2013). Intensive video gaming improves encoding speed to visual short-term memory in young male adults. *Acta Psychol. (Amst)* 142, 108–118. doi: 10.1016/j.actpsy.2012.11.003

Yao, Y., Cui, R., Li, Y., Zeng, L., Jiang, J., Qiu, N., et al. (2020). Action real-time strategy gaming experience related to enhanced capacity of visual working memory. *Front. Hum. Neurosci.* 14:333. doi: 10.3389/fnhum.2020.00333

Conflict of Interest: The 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.

frontiers in Neuroscience

MINI REVIEW
published: 03 November 2021
doi: 10.3389/fnins.2021.761703

Topological Data Analysis as a New Tool for EEG Processing

Xiaoqi Xu[1,2]*, *Nicolas Drougard*[1,2] and *Raphaëlle N. Roy*[1,2]

[1] ISAE-SUPAERO, Université de Toulouse, Toulouse, France, [2] ANITI–Artificial and Natural Intelligence Toulouse Institute, Université de Toulouse, Toulouse, France

Electroencephalography (EEG) is a widely used cerebral activity measuring device for both clinical and everyday life applications. In addition to denoising and potential classification, a crucial step in EEG processing is to extract relevant features. Topological data analysis (TDA) as an emerging tool enables to analyse and understand data from a different angle than traditionally used methods. As a higher dimensional analogy of graph analysis, TDA can model rich interactions beyond pairwise relations. It also distinguishes different dynamics of EEG time series. TDA remains largely unknown to the EEG processing community while it fits well the heterogeneous nature of EEG signals. This short review aims to give a quick introduction to TDA and how it can be applied to EEG analysis in various applications including brain-computer interfaces (BCIs). After introducing the objective of the article, the main concepts and ideas of TDA are explained. Next, how to implement it for EEG processing is detailed, and lastly the article discusses the benefits and limitations of the method.

Keywords: topological data analysis (TDA), electroencephalography (EEG), persistent homology, brain-computer interface (BCI), machine learning

OPEN ACCESS

Edited by:
Grzegorz Marcin Wójcik, Marie Curie-Sklodowska University, Poland

Reviewed by:
Emanuele Olivetti, Bruno Kessler Foundation (FBK), Italy
Claire Cury, Inria Rennes-Bretagne Atlantique Research Centre, France

***Correspondence:**
Xiaoqi Xu
xiaoqi.xu@isae-supaero.fr

Specialty section:
This article was submitted to Brain Imaging Methods, a section of the journal Frontiers in Neuroscience

Received: *20 August 2021*
Accepted: *11 October 2021*
Published: *03 November 2021*

Citation:
Xu X, Drougard N and Roy RN (2021) Topological Data Analysis as a New Tool for EEG Processing. Front. Neurosci. 15:761703. doi: 10.3389/fnins.2021.761703

1. INTRODUCTION

Electroencephalography (EEG) records brain electrical activity in a non-invasive way and contains rich information about the underlying brain state and function. It is intensively used in diagnosis and analysis of various neurological disorders such as epilepsy, schizophrenia and autism spectrum disorder (ASD) (van der Stelt and Belger, 2007; Billeci et al., 2013; Acharya et al., 2015) as well as for non-clinical applications such as sport and sleep monitoring (Borbély et al., 1981; Thompson et al., 2008).

A key step in EEG processing is to extract relevant features or markers for the considered application. Many techniques have been developed, ranging from traditional spectral analysis, to non-linear analysis, as well as to recent deep learning techniques (Muthuswamy and Thakor, 1998; Müller et al., 2008; Murugappan et al., 2008; Subha et al., 2010; Craik et al., 2019). Since the brain is a huge network of neurons wired together and its function is based on the synchronization of neurons, it is natural to study EEG signals using functional connectivity metrics as features (Sporns, 2013). Most of the current work use graph theory as tools to extract features e.g., small-worldness, global clustering coefficient and characteristic path length (Ismail and Karwowski, 2020). While being powerful tools, graph models oversimplify the interactions between neurons by reducing them to nodes and edges, thus capturing only low-dimensional information (0 and 1). In contrast, topological data analysis (TDA) allows to explore higher-dimensional information by using higher dimensional representations called simplicial complexes (a set of points, segments, triangles and their higher dimensional analogs, cf. section 2.1 for a formal definition).

Although it is not novel to classify time series by extracting topological features using TDA (Seversky et al., 2016; Umeda, 2017), it is only recently that this technique has attracted attention in the domain of EEG processing, especially for clinical applications. Some pioneering work has shown that topological features extracted from EEG signals reveal relevant information for various neurological disorders (Wang et al., 2018; Ibáñez-Marcelo et al., 2019; Yamanashi et al., 2021).

In this mini review, we aim to identify where and how TDA can be used for EEG processing by reviewing the current literature, and point out potential directions for future work. The organization is as follows: section 2 is a short presentation of the main principle of TDA; section 3 summarizes various ways to apply TDA in EEG processing based on the current literature; section 4 discusses the advantages and limitations of TDA for EEG analysis and the gaps in current research.

2. TOPOLOGICAL DATA ANALYSIS

Topological data analysis is a young but rapidly growing domain at the intersection of algebraic topology and data science. There already exist some good tutorials for non-mathematicians like data scientists or neuroscientists (Sizemore et al., 2019; Chazal and Michel, 2021). For a more mathematical introduction to algebraic topology, the books of Hatcher (2000) and Ghrist (2014) could be a good starting point. In this section we briefly describe the main idea of TDA and refer the interested readers to the above references. All the definitions are gathered in section 2.1 to facilitate the reading.

2.1. Definitions

DEFINITION 1. *Two functions $f, g : X \to Y$ are said to be **homotopic** if there exist a continuous function $H : X \times [0,1] \to Y$ such that $H(\cdot, 0) = f$ and $H(\cdot, 1) = g$. Two topological spaces X and Y are **homotopic** if there exist continuous functions $f : X \to Y$ and $g : Y \to X$ such that $f \circ g$ and $g \circ f$ are homotopic to id_Y and id_X, respectively.*

DEFINITION 2. *A **k-simplex**, noted as $\Delta_k = [v_0, \ldots, v_k]$, is the convex hull of a set of $k+1$ linearly independent points.*

DEFINITION 3. *A **simplicial complex** is a collection of simplices satisfying following conditions: every subset and their intersections are also simplices in the collection.*

DEFINITION 4. *The **boundary** of a k-simplex $\Delta_k = [v_0, \ldots, v_k]$ is defined as an alternating formal sum of $(k-1)$-simplices, $\partial_k \Delta_k = \sum_i (-1)^i [v_0, \ldots, \hat{v_i}, \ldots, v_k]$ where $\hat{v_i}$ means omitting v_i.*

EXAMPLE 1. *Take a 2-simplex $\Delta_2 = [v_0, v_1, v_2]$ for example: $\partial_2 \Delta_2 = [v_1, v_2] - [v_0, v_2] + [v_0, v_1] = [v_1, v_2] + [v_2, v_0] + [v_0, v_1]$. Its boundary is in fact the loop formed by its edges.*

REMARK 1. *A simple yet important property is that the boundary of a boundary is always zero: $\partial_k \circ \partial_{k+1} = 0, \forall k \geq 0$. In other words, $\text{im}\, \partial_{k+1} \subseteq \ker \partial_k$. Homology is defined based on this property.*

EXAMPLE 2. *To have an intuition of what goes on here, we could take the same example above. $\partial_1 \circ \partial_2\, \Delta_2 = \partial_1([v_1, v_2] - [v_0, v_2] + [v_0, v_1]) = v_2 - v_1 - v_2 + v_0 + v_1 - v_0 = 0$.*

DEFINITION 5. *The k-dimensional **homology** of a simplicial complex C is defined as the quotient space $H_k(C) := \ker \partial_k / \text{im}\, \partial_{k+1}$, in which two elements that differ by a k-boundary are considered as the same element. The dimension of $H_k(C)$ is called the k^{th} **Betti number**.*

2.2. Topology and Homology in a Nutshell

Topology is the mathematical branch which studies the properties that are preserved by continuous deformation, such as scaling, twisting but not tearing. It could be viewed as "geometry without metric." Distance is of no importance in topology, instead the whole theory is based on the notion of "closeness." More precisely, what we call topological properties are those invariant under homeomorphism (continuous map whose inverse is also continuous). However, homeomorphism is often too strict. A looser but useful notion is homotopy (see Definition 1). It is not only useful in the theory, but also in practice since a lot of noise in the real world data can also be viewed in a homotopic way, such as an blurred edge is homotopic (but not homeomorphic) to the ideal edge in a digital image.

Based on the notion of homotopy, homotopy groups can be defined and distinguish different topological spaces. But they are difficult to compute in general. An alternative notion is homology (see Definition 5), which is based on simplicial complex (see Definition 3) and can be computed effectively using the "divide and conquer" strategy.

Intuitively, the k-dimensional homology catches k-dimensional "holes," i.e., independent cycles that are not filled. Homology is a homotopy invariant and the dimensions of the homology group, i.e., Betti numbers, are among the first topological statistics applied in real world applications (Giusti et al., 2015).

2.3. Persistent Homology

What makes TDA powerful and particularly suitable to capture hierarchical features is persistent homology, a method that extracts persistent topological features across scales. The key notion is filtration, which is a nested family of subcomplexes indexed by a parameter. The homology of these subcomplexes evolves as the parameter grows, giving rise to the barcode or persistence diagram as a description of persistence of connected components (described with components' birth and death) by 0-dimensional homology, and of multidimensional holes by higher dimensional homology.

Figure 1 illustrates two kinds of filtration that are frequently used. Let f be a map from $\mathbb{R}$ to $\mathbb{R}$ as shown in **Figure 1A**. Then the *sublevel filtration* consists of the family of subcomplexes $S_l = \{\Delta \in \mathbb{R} : f(\Delta) < l\}$. For this example the only non trivial topological information is carried by a 0-dimensional homology since we have only segments (as shown at the bottom of **Figure 1A**) and no loops. We see as the parameter l increases, there are new components emerge, marked as birth, and also some component merges with the other one, leading to its death (by convention the elder rule is applied that keeps the older one). The parameters corresponding to birth and death form a family of intervals. Chazal et al. (2012) demonstrated that this family of intervals is unique up to reordering and can be used as a

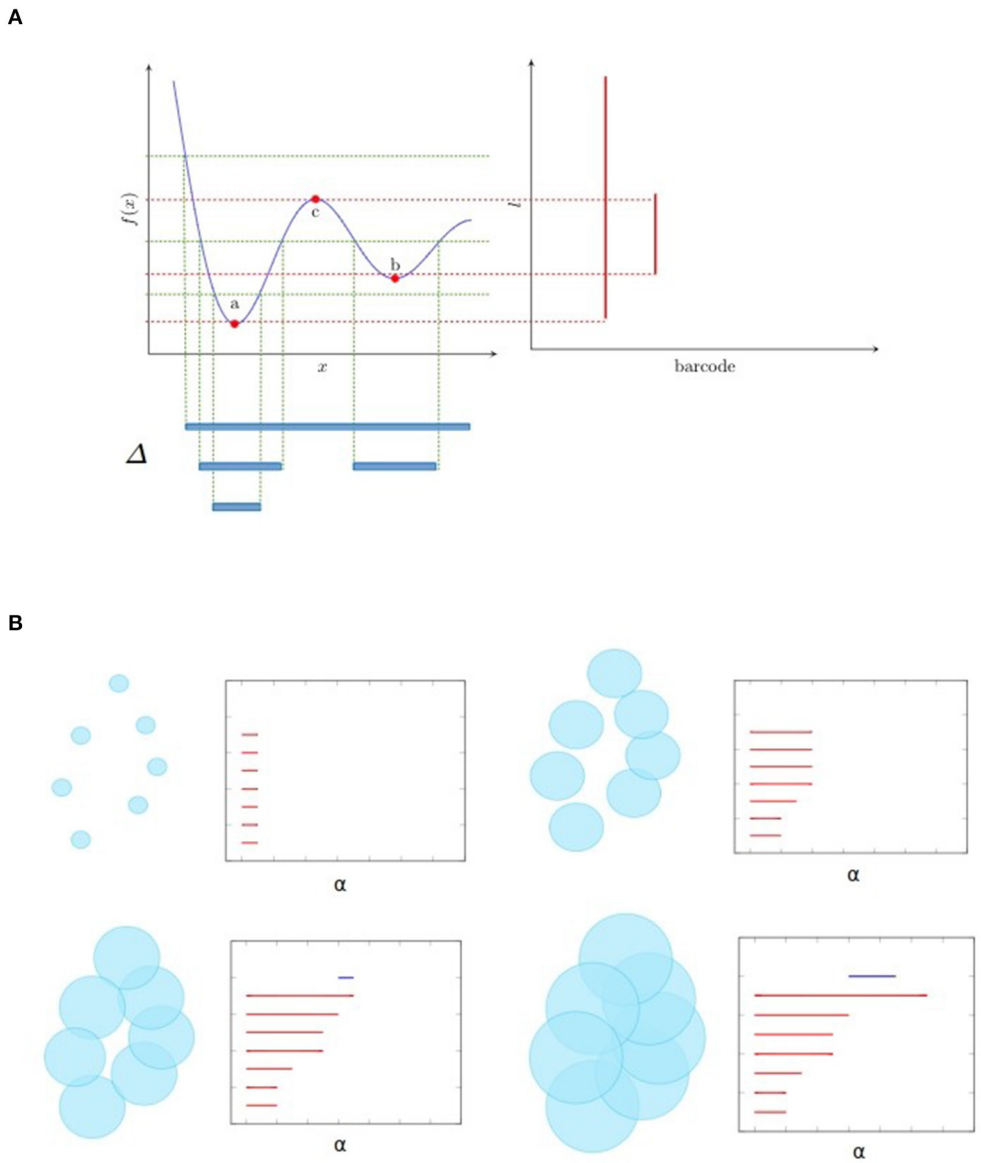

FIGURE 1 | Examples of filtration and the associated barcode. **(A)** shows the sublevel filtration of function $f(x)$. The first and second connected component appear at point *a* and *b* respectively and merge together at point *c*, giving rise to two bars: one from *a* till infinity and the other from *b* to *c*. **(B)** shows several stages of Rips filtration. At first, all points are isolated and we have the same number of bars and points. As balls grow larger, some balls start to merge together, giving death to certain bars. Then one loop appears, marked as a 1-dimensional feature corresponding to the blue line in the barcode. Finally the loop is filled so the blue bar stops and every point merges into one component living forever.

topological feature named *barcode*. We could also mark on the plan $\mathbb{R}^2$ the point at which the x and y coordinates correspond to birth and death of a feature, respectively. These points, together with the diagonal, are called *persistence diagram*.

Figure 1B is an example of Rips filtration. Given a set of points X and a positive number α, the *Rips complex* $\text{Rips}_\alpha(X)$ is the simplicial complex including all simplices in which the distance between any two of their vertices is smaller than α. *Rips filtration* consists of the family of Rips complex $\{\text{Rips}_\alpha(X)\}_\alpha$ indexed by α. In the example in $\mathbb{R}^2$, we could imagine balls with increasing radius around the initial point cloud. As the balls become bigger, they merge with each other, leading to the death of certain connected components but also the birth of some loops (only one for the example in **Figure 1B**). The larger the scale of the loop is, the more persistent it is in the barcode. In most cases, long bars correspond to significant features while short bars correspond to noise. As α goes from 0 to ∞, the topology of each Rips complex goes from one trivial case, disjoint unions of points, to another

TABLE 1 | Summary of applications of TDA to EEG analysis (ordered by publication date).

References	Domain	Transformation	Method	Features	Classifier	Dataset
Altındiş et al. (2021)	MI-BCI	-	Time delay embedding	Persistence diagram	-	Graz dataset
Bischof and Bunch (2021)	Eyes-open/eyes-closed classification	-	Time delay embedding	Betti-numbers	CNN	Bonn dataset
Yamanashi et al. (2021)	Delirium	-	Time delay embedding	Area of the 1-dimensional Betti curve	-	Private dataset
Majumder et al. (2020)	Autism Spectrum Disorder (ASD)	-	Sublevel filtration	Persistent entropy	SVM	Private dataset
Wang et al. (2020b)	Aphasia	ICA	Sublevel filtration	Persistence landscape	-	Private dataset
Wang et al. (2020a)	Aphasia	ICA	Gradient filtration	Persistence landscape	-	Private dataset
Ibáñez-Marcelo et al. (2019)	Hypnotizability	ICA	Connectivity	Homological scaffold	-	Private dataset
Nasrin et al. (2019)	Brain state classification	-	Sublevel filtration	Persistence diagram	Bayesian learning	US Army Aberdeen Proving Ground (APG) simulation dataset
Wang et al. (2019)	Seizure localization	Fourier transform	Sublevel filtration	Persistence landscape	-	Private dataset
Altindis et al. (2018)	MI-BCI	-	Time delay embedding	Betti numbers	kNN	Graz dataset
Piangerelli et al. (2018)	Epileptic seizure	-	Sublevel filtration	Persistent entropy	Linear classifier	Physionet
Wang et al. (2018)	Epilepsy	Fourier transform	Sublevel filtration	Persistence landscape	-	Private dataset

trivial case, all points merged to one connected component. The persistence homology records the evolution between the two extreme cases, in the middle of which interesting features come to light. So, unlike hard thresholding methods usually used in the construction of graphs, TDA preserves more information.

3. TDA APPLIED ON EEG DATA

In order to retrieve and evaluate in a comprehensive manner all research works related to TDA and EEG processing, a systematic search was conducted as described here after. Hence, using (topological data analysis, TDA) and (electroencephalography, EEG) as search terms, we found 70 publications on PubMed and 85 publications on Web of Science. Then we examined the title and abstract of each non-redundant publication and excluded those concerning recording methods other than EEG (e.g., MEG, fMRI) or using graph theory analysis instead of TDA as tools. The remaining publications are summarized in **Table 1**.

We note that besides standard preprocessing steps, e.g., band-pass filtering, downsampling, and artifacts removal, few authors have chosen to transform data into another space. Independent component analysis (ICA) is the one mostly used with the purpose of getting source components, then followed by Fourier transform for the purpose of denoising data.

From the papers that were collected, three methods of employing TDA emerge: the first one applies it directly onto the EEG signals; the second one applies it onto the connectivity network; and the third one onto the phase space. For the first group, sublevel filtration illustrated in **Figure 1A** is applied directly on the EEG time series of each channel which is as the function f in the example. For the second group, Rips filtration, as illustrated in **Figure 1B**, is applied on the point cloud in which each point represents a channel or a source component, and the distance is measured by connectivity measures such as Pearson correlation. For the third group, even though it is possible to embed data spatially, all authors have followed time-delay embedding well backed by Takens' theorem (Takens, 1981) which gives the minimum dimension of embedding in order to reconstruct attractors (a set of points in the phase space that is invariant under dynamics and "attracts" neighboring points) up to diffeomorphism. Time-delay embedding of a time series $\{x(t)\}_t$ is formed by keeping k observations before current time $(x(t-k+1), x(t-k+2), \ldots, x(t-1), x(t))$. In the phase space each point represents the state at a certain time. Takens' theorem guarantees that the differential, so topological, properties of the attractors are preserved by time delay embedding, which permits to distinguish different EEG time series based on the topology of the attractors in the phase space.

A barcode or persistence diagram is constructed along with the filtration process as shown in section 2.3. A metric structure is still needed in order to measure differences between barcodes or persistence diagrams. The Bottleneck distance, or Wasserstein distance, measures the distance between persistence diagrams by pairing points in the two diagrams.

There are many other ways to extract topological features from persistence diagrams other than measuring distance in the original space. The simplest way might be extracting Betti numbers. Another simple method is persistence entropy (Chintakunta et al., 2015) which gives a scalar description of the barcode. Additionally, persistence landscapes introduced by Bubenik (2015) smartly embed persistence diagrams into a Hilbert space where most machine learning algorithms can be applied, which makes them a popular choice.

Finally, after the feature extraction step, a classifier might be applied or not depending on the application at hand (e.g., diagnostic, active brain-computer interface, or mental state estimation via a passive brain-computer interface). The most used ones are linear classifiers due to the small size of clinical datasets. However, when the amount of data allows, deep learning methods show promising performance (Bischof and Bunch, 2021). TDA can also be combined with other frameworks such as Bayesian networks (Nasrin et al., 2019).

4. DISCUSSION

TDA has many advantages. Firstly, topological features are by nature robust and invariant to transformations such as translation, amplitude and frequency scaling (Wang et al., 2018). Secondly, TDA is well suited for neuroscience, especially analyses involving connectivity networks. Neurons far apart can communicate with each other since some axons extend up to one meter or more, so it's how they are connected, i.e., the topological structure of the network, and not the distance that counts. Thirdly, TDA can capture global and higher dimensional features where other methods such as graph theory fails.

The main limitation of TDA stems also from its strength. Since it neglects all metric related information, this harms its ability to distinguish data of different categories. Considering the pros and cons of TDA, it is recommended to combine TDA with other methods to use it to its full potential. One promising direction is the combination of TDA with deep learning techniques. There have been some pioneering work e.g., the work of Carriere et al. (2020), Kim et al. (2020), and Royer et al. (2021). Much more still remains to be explored, especially leveraging the particular structure of EEG signals. Another direction is to associate TDA with statistics. The topological features could be seen not as deterministic but random variables. The notions such as convergence rate, consistency, confidence region etc. of the extracted topological features could be studied.

Although there are theoretical results showing the robustness of persistence diagrams under perturbations (Cohen-Steiner et al., 2007; Bubenik, 2015), in practice the technical details of applying TDA for EEG signals need to be further investigated. Altındiş et al. (2021) started in this direction by trying to find the optimal embedding dimension, time delay and time window size using false nearest neighbor (FNN) test. The paired t-test showed that the significance level of extracted topological features was very sensitive to the choice of embedding parameters and hence it was important to use the optimal parameters.

The current domain of application is still quite restricted to clinical studies to improve the diagnostic of neurological diseases. However, TDA is also quite suitable for other non-clinical EEG related areas, e.g., EEG based brain-computer interface (BCI; Wolpaw et al., 2000) which allows the explicit control of machines or implicit mental state estimation using only EEG signals. Further, the datasets used in current publications are mostly private datasets, which if possible should be replaced by publicly available ones for increased reproducibility and comparison with other work.

TDA is going through rapid development, both in theory and in application. With better theoretical foundation and more open source software and code published online[1], we believe that it will become part of the arsenal of tools for a broader scientific community. Hopefully we will see more publications on EEG processing using TDA in the future.

AUTHOR CONTRIBUTIONS

XX: original idea and drafting of the article. ND and RR: writing supervision and critical revisions. All authors contributed to the article and approved the submitted version.

FUNDING

The work was funded by ANITI (Artificial and Natural Intelligence Toulouse Institute), Toulouse, France.

[1]For example see https://github.com/FatemehTarashi/awesome-tda for a list of various softwares and resources.

REFERENCES

Acharya, U. R., Fujita, H., Sudarshan, V. K., Bhat, S., and Koh, J. E. (2015). Application of entropies for automated diagnosis of epilepsy using eeg signals: a review. *Knowl. Based Syst*. 88, 85–96. doi: 10.1016/j.knosys.2015.08.004

Altındiş, F., Yılmaz, B., Borisenok, S., and İçöz, K. (2021). Parameter investigation of topological data analysis for eeg signals. *Biomed. Signal. Process. Control*. 63:102196. doi: 10.1016/j.bspc.2020.102196

Altindis, F., Yilmaz, B., Borisenok, S., and Icoz, K. (2018). "Use of topological data analysis in motor intention based brain-computer interfaces," in *2018 26th European Signal Processing Conference (EUSIPCO)* (Rome), 1695–1699.

Billeci, L., Sicca, F., Maharatna, K., Apicella, F., Narzisi, A., Campatelli, G., et al. (2013). On the application of quantitative eeg for characterizing autistic brain: a systematic review. *Front. Hum. Neurosci.* 7:442. doi: 10.3389/fnhum.2013.00442

Bischof, B., and Bunch, E. (2021). Geometric feature performance under downsampling for eeg classification tasks. *arXiv[Preprint].arXiv:2102.07669*.

Borbély, A. A., Baumann, F., Brandeis, D., Strauch, I., and Lehmann, D. (1981). Sleep deprivation: Effect on sleep stages and eeg power density in man. *Electroencephalogr. Clin. Neurophysiol.* 51, 483–493. doi: 10.1016/0013-4694(81)90225-X

Bubenik, P. (2015). Statistical topological data analysis using persistence landscapes. *J. Mach. Learn. Res.* 16, 77–102. doi: 10.5555/2789272.2789275

Carriere, M., Chazal, F., Ike, Y., Lacombe, T., Royer, M., and Umeda, Y. (2020). "Perslay: a neural network layer for persistence diagrams and new graph topological signatures," in *Proceedings of the Twenty Third International Conference on Artificial Intelligence and Statistics, Vol. 108 of Proceedings of Machine Learning Research*, eds S. Chiappa and R. Calandra (PMLR), 2786–2796

Chazal, F., De Silva, V., Glisse, M., and Oudot, S. (2012). *The Structure and Stability of Persistence Modules*. New York, NY: Springer International Publishing.

Chazal, F., and Michel, B. (2021). An introduction to topological data analysis: fundamental and practical aspects for data scientists. *arXiv math.ST 1710.04019*.

Chintakunta, H., Gentimis, T., Gonzalez-Diaz, R., Jimenez, M.-J., and Krim, H. (2015). An entropy-based persistence barcode. *Pattern Recognit*. 48, 391–401. doi: 10.1016/j.patcog.2014.06.023

Cohen-Steiner, D., Edelsbrunner, H., and Harer, J. (2007). Stability of persistence diagrams. *Discrete Comput. Geometry* 37, 103–120. doi: 10.1007/s00454-006-1276-5

Craik, A., He, Y., and Contreras-Vidal, J. L. (2019). Deep learning for electroencephalogram (EEG) classification tasks: a review. *J. Neural Eng*. 16, 031001. doi: 10.1088/1741-2552/ab0ab5

Ghrist, R. (2014). *Elementary Applied Topology*. Scotts Valley, CA: CreateSpace Independent Publishing Platform.

Giusti, C., Pastalkova, E., Curto, C., and Itskov, V. (2015). Clique topology reveals intrinsic geometric structure in neural correlations. *Proc. Natl. Acad. Sci. U.S.A*. 112, 13455–13460. doi: 10.1073/pnas.1506407112

Hatcher, A. (2000). *Algebraic Topology*. Cambridge: Cambridge Univ. Press.

Ibáñez-Marcelo, E., Campioni, L., Phinyomark, A., Petri, G., and Santarcangelo, E. L. (2019). Topology highlights mesoscopic functional equivalence between imagery and perception: the case of hypnotizability. *Neuroimage* 200, 437–449. doi: 10.1016/j.neuroimage.2019.06.044

Ismail, L. E., and Karwowski, W. (2020). A graph theory-based modeling of functional brain connectivity based on eeg: a systematic review in the context of neuroergonomics. *IEEE Access*. 8, 155103–155135. doi: 10.1109/ACCESS.2020.3018995

Kim, K., Kim, J., Zaheer, M., Kim, J., Chazal, F., and Wasserman, L. (2020). "Pllay: efficient topological layer based on persistent landscapes," in *Advances in Neural Information Processing Systems, Vol. 33*, eds H. Larochelle, M. Ranzato, R. Hadsell, M. F. Balcan, and H. Lin (Curran Associates, Inc.), 15965–15977.

Majumder, S., Apicella, F., Muratori, F., and Das, K. (2020). "Detecting autism spectrum disorder using topological data analysis," in *ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* (Barcelona: IEEE), 1210–1214.

Müller, K.-R., Tangermann, M., Dornhege, G., Krauledat, M., Curio, G., and Blankertz, B. (2008). Machine learning for real-time single-trial eeg-analysis: from brain–computer interfacing to mental state monitoring. *J. Neurosci. Methods* 167, 82–90. doi: 10.1016/j.jneumeth.2007.09.022

Murugappan, M., Rizon, M., Nagarajan, R., Yaacob, S., Hazry, D., and Zunaidi, I. (2008). "Time-frequency analysis of eeg signals for human emotion detection," in *4th Kuala Lumpur International Conference on Biomedical Engineering 2008*, eds N. A. Abu Osman, F. Ibrahim, W. A. B. Wan Abas, H. S. Abdul Rahman, and H.-N. Ting (Berlin; Heidelberg: Springer), 262–265.

Muthuswamy, J., and Thakor, N. V. (1998). Spectral analysis methods for neurological signals. *J. Neurosci. Methods* 83, 1–14. doi: 10.1016/S0165-0270(98)00065-X

Nasrin, F., Oballe, C., Boothe, D., and Maroulas, V. (2019). "Bayesian topological learning for brain state classification," in *2019 18th IEEE International Conference On Machine Learning And Applications (ICMLA)* (Boca Raton, FL: IEEE), 1247–1252.

Piangerelli, M., Rucco, M., Tesei, L., and Merelli, E. (2018). Topological classifier for detecting the emergence of epileptic seizures. *BMC Res. Notes* 11:392. doi: 10.1186/s13104-018-3482-7

Royer, M., Chazal, F., Levrard, C., Umeda, Y., and Ike, Y. (2021). "Atol: measure vectorization for automatic topologically-oriented learning," in *Proceedings of The 24th International Conference on Artificial Intelligence and Statistics, Vol. 130 of Proceedings of Machine Learning Research*, eds A. Banerjee and K. Fukumizu (PMLR), 1000–1008.

Seversky, L. M., Davis, S., and Berger, M. (2016). "On time-series topological data analysis: New data and opportunities," in *2016 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)* (Las Vegas, NV: IEEE), 1014–1022.

Sizemore, A. E., Phillips-Cremins, J. E., Ghrist, R., and Bassett, D. S. (2019). The importance of the whole: topological data analysis for the network neuroscientist. *Netw. Neurosci*. 3, 656–673. doi: 10.1162/netn_a_00073

Sporns, O. (2013). Structure and function of complex brain networks. *Dialogues Clin. Neurosci*. 15, 247–262. doi: 10.31887/DCNS.2013.15.3/osporns

Subha, D. P., Joseph, P. K., Acharya, U., R., and Lim, C. M. (2010). Eeg signal analysis: a survey. *J. Med. Syst*. 34, 195–212. doi: 10.1007/s10916-008-9231-z

Takens, F. (1981). "Detecting strange attractors in turbulence," in *Dynamical Systems and Turbulence, Warwick 1980*, eds D. Rand and L.-S. Young (Berlin; Heidelberg: Springer), 366–381.

Thompson, T., Steffert, T., Ros, T., Leach, J., and Gruzelier, J. (2008). Eeg applications for sport and performance. *Methods* 45, 279–288. doi: 10.1016/j.ymeth.2008.07.006

Umeda, Y. (2017). Time series classification via topological data analysis. *Trans. Jpn. Soc. Artif. Intell*. 32, 1–12. doi: 10.1527/tjsai.D-G72

van der Stelt, O., and Belger, A. (2007). Application of electroencephalography to the study of cognitive and brain functions in schizophrenia. *Schizophr. Bull*. 33, 955–970. doi: 10.1093/schbul/sbm016

Wang, Y., Behroozmand, R., Johnson, L. P., Bonilha, L., and Fridriksson, J. (2020a). "Topological signal processing in neuroimaging studies," in 2020 *IEEE 17th International Symposium on Biomedical Imaging Workshops (ISBI Workshops)* (Iowa City, IA: IEEE), 1–4.

Wang, Y., Behroozmand, R., Johnson, L. P., and Fridriksson, J. (2020b). "Topology highlights neural deficits of post-stroke aphasia patients," in *2020 IEEE 17th International Symposium on Biomedical Imaging (ISBI)* (Iowa City, IA: IEEE), 754–757.

Wang, Y., Ombao, H., and Chung, M. K. (2018). Topological data analysis of single-trial electroencephalographic signals. *Ann. Appl. Stat*. 12, 1506–1534. doi: 10.1214/17-AOAS1119

Wang, Y., Ombao, H., and Chung, M. K. (2019). "Statistical persistent homology of brain signals," in *ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* (Brighton: IEEE), 1125–1129.

Wolpaw, J., Birbaumer, N., Heetderks, W., McFarland, D., Peckham, P., Schalk, G., et al. (2000). Brain-computer interface technology: a review of the first international meeting. *IEEE Trans. Rehabil. Eng*. 8, 164–173. doi: 10.1109/TRE.2000.847807

Yamanashi, T., Kajitani, M., Iwata, M., Crutchley, K. J., Marra, P., Malicoat, J. R., et al. (2021). Topological data analysis (tda) enhances bispectral eeg (bseeg) algorithm for detection of delirium. *Sci. Rep*. 11:304. doi: 10.1038/s41598-020-79391-y

Conflict of Interest: The 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.

frontiers in Neuroscience

REVIEW
published: 12 November 2021
doi: 10.3389/fnins.2021.728092

Defining a Path Toward the Use of Fast-Scan Cyclic Voltammetry in Human Studies

Suelen Lucio Boschen[1], James Trevathan[2], Seth A. Hara[3], Anders Asp[1,4] and J. Luis Lujan[1,5]*

[1] Applied Computational Neurophysiology and Neuromodulation Laboratory, Department of Neurologic Surgery, Mayo Clinic, Rochester, MN, United States, [2] Department of Biomedical Engineering, University of Wisconsin-Madison, Madison, WI, United States, [3] Division of Engineering, Mayo Clinic, Rochester, MN, United States, [4] Mayo Clinic Graduate School of Biomedical Sciences, Mayo Clinic, Rochester, MN, United States, [5] Department of Physiology and Biomedical Engineering, Mayo Clinic, Rochester, MN, United States

Fast Scan Cyclic Voltammetry (FSCV) has been used for decades as a neurochemical tool for *in vivo* detection of phasic changes in electroactive neurotransmitters in animal models. Recently, multiple research groups have initiated human neurochemical studies using FSCV or demonstrated interest in bringing FSCV into clinical use. However, there remain technical challenges that limit clinical implementation of FSCV by creating barriers to appropriate scientific rigor and patient safety. In order to progress with clinical FSCV, these limitations must be first addressed through (1) appropriate pre-clinical studies to ensure accurate measurement of neurotransmitters and (2) the application of a risk management framework to assess patient safety. The intent of this work is to bring awareness of the current issues associated with FSCV to the scientific, engineering, and clinical communities and encourage them to seek solutions or alternatives that ensure data accuracy, rigor and reproducibility, and patient safety.

Keywords: clinical neurochemistry, deep brain stimulation, fast scan cyclic voltammetry, intraoperative, neurochemical signaling, neurophysiology

OPEN ACCESS

Edited by:
Grzegorz Marcin Wójcik, Marie Curie-Sklodowska University, Poland

Reviewed by:
Jit Muthuswamy, Arizona State University, United States
Vassiliy Tsytsarev, University of Maryland, College Park, United States
Parastoo Hashemi, University of South Carolina, United States

****Correspondence:***
Suelen Lucio Boschen
Souza.suelen@mayo.edu

Specialty section:
This article was submitted to Neural Technology, a section of the journal Frontiers in Neuroscience

Received: *20 June 2021*
Accepted: *25 October 2021*
Published: *12 November 2021*

Citation:
Lucio Boschen S, Trevathan J, Hara SA, Asp A and Lujan JL (2021) Defining a Path Toward the Use of Fast-Scan Cyclic Voltammetry in Human Studies. Front. Neurosci. 15:728092. doi: 10.3389/fnins.2021.728092

INTRODUCTION

Fast scan cyclic voltammetry (FSCV) is an electrochemistry technique used for over 30 years to study rapid neurotransmission in the brain of anesthetized and awake and behaving animals (Ganesana et al., 2017; Rodeberg et al., 2017). FSCV detects electroactive neurotransmitters by varying the electric potential between a small working electrode (on the order of a few micrometers in diameter) and a larger reference electrode (on the order of millimeters in diameter). This process oxidizes and/or reduces neurotransmitters at specific potentials and results in electrical currents with amplitude proportional to the concentration of the neurotransmitter in the extracellular space (Clark et al., 2010; Bucher and Wightman, 2015; **Figure 1**). Under the right conditions, this enables the estimation of changes in neurotransmitter levels with high spatial and temporal resolution, sensitivity, and chemical selectivity (Borland and Michael, 2007; Johnson et al., 2017).

However, *in vivo* neurotransmitter measurement via FSCV can be confounded by several factors including interferant molecules with similar oxidation/reduction potentials to the neurotransmitter of interest, electrode biofouling, shifts in pH and ionic concentrations, or increased oxygenated blood flow in the electrode microenvironment. Additionally, electrical and motion artifacts

can disrupt the electrochemical interface creating artificial signals that appear similar to neurotransmitter release (Ariansen et al., 2012; Lee et al., 2017). In small animal studies, a set of guidelines known as the "Five Golden Rules" is typically used to validate and increase reliability of *in vivo* neurotransmitter measurements (Clark et al., 2010). These Golden Rules include (i) identification of neurotransmitter-specific electrochemical signatures, (ii) additional confirmation of the chemical identity of each recorded neurotransmitter (e.g., through microdialysis at the FSCV site), (iii) anatomical validation of the recording location, (iv) kinetic validation of spontaneous or evoked changes in neurotransmitter concentration, and (v) pharmacological validation of recorded neurotransmitters (Clark et al., 2010; Meunier et al., 2018). These guidelines are instrumental toward ensuring the validity of pre-clinical FSCV measurements.

In the last 10 years, there has been a trend toward the use of FSCV in patients undergoing neurosurgical procedures to study human neurophysiology (Bucher and Wightman, 2015). Undoubtedly, advancing the knowledge of underlying mechanisms and biological processes associated with the complex chemistry of the human brain is critical to the development of new and improved therapeutic interventions. For example, the clinical use of FSCV may allow characterization of changes in neurochemical signaling evoked by deep brain stimulation (DBS) and could potentially advance treatment of neurologic diseases (Lee et al., 2017). However, clinical questions that can be currently answered with FSCV are limited by technical barriers that must be addressed to ensure that clinical studies have appropriate scientific rigor and mitigate risks to patient safety. To address potential confounds during data acquisition in humans, extensive pre-clinical testing of new technologies designed to enable human use must be performed. Additionally, there is a critical need for appropriate pre-clinical studies mimicking the clinical environment. Risks should be weighed against actual benefits of the clinical work to ensure patient safety and high-quality data. The objective of this manuscript is to initiate a discussion between the scientific, engineering, and clinical communities regarding these issues and encourage collaborative solutions that are safe and suitable for neurochemical measurements in clinical studies.

INVESTIGATIONAL USE OF FAST SCAN CYCLIC VOLTAMMETRY

While a clinical study may not directly benefit the patients participating in the study, studies are likely to improve future patient care. Notably, treatment improvements and benefits to patients can only be accomplished by leveraging appropriate pre-clinical studies and scientific rigor.

Large Animal Studies

Anatomical and physiological similarities, as well as behavioral correlates to human conditions often make non-human primate and swine models a better alternative for translational research (Lind et al., 2007; Phillips et al., 2014). Unfortunately, only a small number of FSCV studies have investigated neurotransmitter release *in vivo* using large animal models (**Table 1**).

Dopamine release in large animals was first measured via FSCV in 1984 in a non-human primate model (Cheney-Thamm et al., 1984). In the study, the investigators detected increased current peaks in the oxidation range for dopamine in the caudate in response to pleasant stimuli and amphetamine administration. However, these measurements were not properly validated accordingly to the Five Golden Rules, i.e., use of rudimentary electrochemical parameters and working electrode material, which reduces reliability on the data collected.

It was not until 10 years later that FSCV was used again in non-human primates to demonstrate reduced dopaminergic release in response to electrical stimulation of the medial forebrain bundle in an anesthetized 1-methyl-4-phenyl-1,2,3,6-tetrahydropyridine marmoset model of Parkinson's Disease (Earl et al., 1998). In this study, investigators validated dopamine measurements by confirming dopamine-specific electrochemical signature, and by demonstrating anatomically, electrically, and pharmacologically evoked changes in the kinetics of dopamine concentration.

Since then, other studies using FSCV in swine and non-human primate models have utilized FSCV recording electrodes (including both working and reference electrodes), recording systems, and analysis techniques that had been well validated in small animal models (Adams, 1976; Stamford et al., 1984; Millar et al., 1985; Suaud-Chagny et al., 1992), ensuring rigor and data reproducibility. Consistently, they have reported increased dopamine release in the caudate-putamen as a function of reward (Yoshimi et al., 2015), and electrical stimulation of the subthalamic nucleus (Nakajima et al., 2017), fornix, and ventral tegmental area (Lee et al., 2017). In addition, FSCV has been used to detect cortical adenosine during seizure termination in swine (Van Gompel et al., 2014), and to serve as the basis of a neurochemical-based closed-loop DBS system in rodent, swine, and non-human primate (Ariansen et al., 2012). More recently, studies in non-human primates proposed new strategies to overcome challenges regarding chronic FSCV recordings in large animals, such as target optimization and spatial resolution (Schluter et al., 2014; Schwerdt et al., 2017b). However, they could not resolve other limitations such as tissue damage, glial encapsulation, electrode biofouling, and material deterioration that caused continuous signal loss (Kozai et al., 2015).

Overall, these pre-clinical studies show the feasibility of recording of neurotransmitters via FSCV in acute and long-term animal models. However, they do not directly address the significant barriers to clinical use of FSCV, especially in the operating room setting.

Scientific Rigor

It is not possible to apply the Five Golden Rules for neurotransmitter measurement in the human operating room. However, many of the potential confounds for FSCV measurement in clinical studies can be anticipated. Data analysis, including calibration and signal extraction algorithms designed to distinguish neurochemical signals from interferents and noise, depend mainly on the electrochemical properties of the

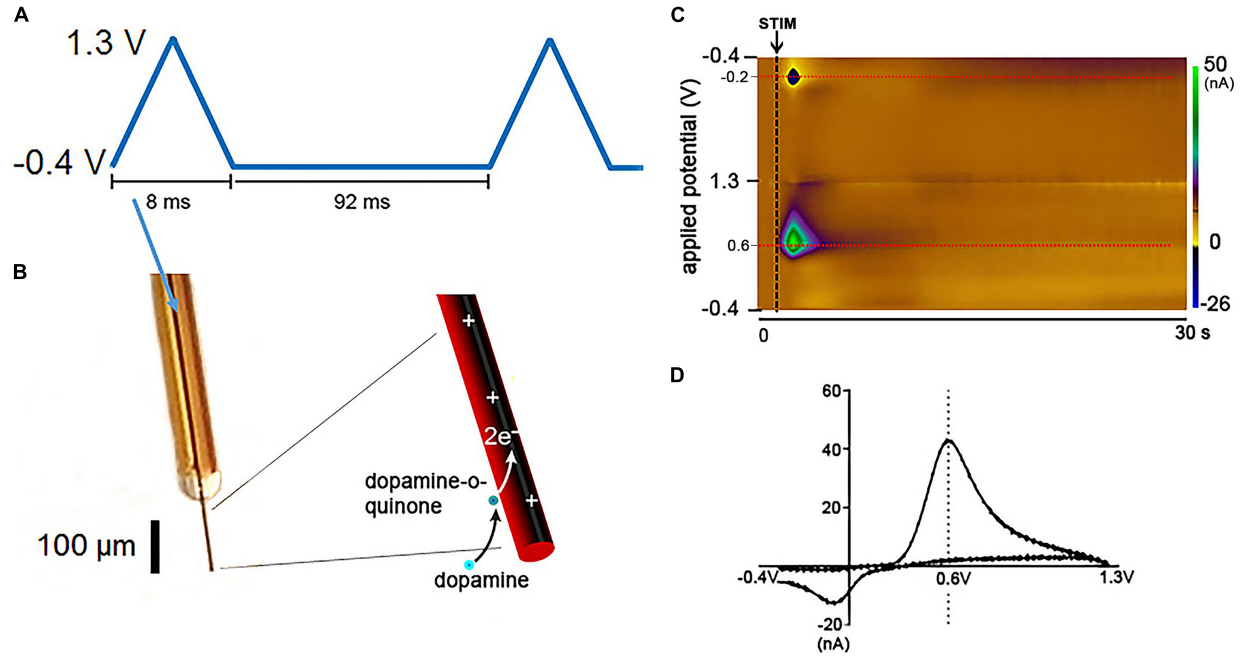

FIGURE 1 | Fast Scan Cyclic Voltammetry in dopamine detection. Schematic of FSCV dopamine detection at CFM. **(A)** Triangular waveform typically used for catecholamine detection. Each waveform ranging from –0.4 to 1.3 V takes 8.0 ms to be delivered and is followed by 92 ms of a constant potential held at –0.4 V. Therefore, each waveform cycle takes 100 ms to be completed. **(B)** Illustration of a silica tubing-insulated CFM with the oxidation/reduction reactions of dopamine at the carbon fiber exposed tip. **(C)** Pseudocolor-plot shows the changes in oxidation (0.6 V) and reduction (–0.2 V) currents (red dashed lines) corresponding to dopamine release evoked by medial forebrain bundle electrical stimulation (black arrow and dashed line) in an anesthetized rat. The oxidation and reduction currents (in nA) are represented in the color scale on the right side of the pseudocolor-plot. The pseudocolor-plot shows 300 triangular waveform cycles (–0.4, 1.3, and –0.4 V) applied from 0 to 30 s at 10 Hz. Additional time after neurotransmitter release is typically showed in pseudocolor-plots to demonstrate that the background current returns to baseline levels. **(D)** Dopamine voltammogram associated with the oxidation and reduction currents represented in the pseudocolor-plot. Adapted with permission from The Schwerdt lab (http://schwerdt.pitt.edu/resources.html).

FSCV recording system. For example, the material composition, configuration, and surface properties of FSCV electrodes define the electrochemical properties of the recording system and can affect the recorded signals (Bucher and Wightman, 2015; Ganesana et al., 2017; Johnson et al., 2017; Meunier et al., 2018; Puthongkham and Venton, 2019). Additionally, electrode materials and recording system commonly used in small animal models must be modified to ensure biocompatibility and the ability to reach deep brain targets. Furthermore, confirmation of measured signals accuracy via pharmacologic validation or alternate neurochemical technique in humans is largely not possible in the operating room due to increased hemorrhage risk with additional alternate neurochemical recordings (Sansur et al., 2007; Park et al., 2011), and risk of pharmacological manipulations interfering with patient's wellbeing during the procedure. Moreover, the time required for recording system stabilization and signal detection is large, but the total allowed operating room time for safe experimental procedures is limited. For example, in acute animal experiments using a single Carbon fiber microelectrode (CFM), identifying an optimal area for recording evoked neurochemical release within a given target region can take several minutes to a few hours. This process requires slowly moving the electrode to different locations within the target and allowing enough time for the signal drift to settle after each movement. In contrast, Institutional Review Boards (IRBs) typically require that the research use of FSCV adds no longer than 15–60 min to the overall procedure, which varies across institutions with an average duration of 4 h. The compressed timeframe of the operating room environment combined with patient safety concerns require the use of one single track, with very little or no time available for electrode repositioning for recording optimization. The operating room environment is also full of sources of noise and interference from electronic equipment that can affect FSCV recordings in unexpected ways. Although not exhaustive, the next sections highlight the need for appropriate pre-clinical studies addressing changes to the recording methods, experimental paradigm, and data analysis techniques beyond what is described in the FSCV literature before human neurophysiology studies can be performed.

Novel Working Electrodes Designs

The working electrode is typically composed of carbon-based materials (**Table 2**) to improve biocompatibility and chemical inertness, as well as to reduce background currents that interfere with the detection of neurochemical signatures associated with specific neurotransmitters (McDermott and McCreery, 1994; Patel et al., 2013; Rodeberg et al., 2017; Roberts and Sombers, 2018). CFMs have become the standard working electrode for pre-clinical FSCV recordings. The popularity of CFMs can

TABLE 1 | Fast scan cyclic voltammetry studies in large animals.

Study	Species (common name), *n*	Goal	FSCV parameters	FSCV electrodes	Major findings
Cheney-Thamm et al., 1984	*Macaca nemestrina* (Pigtail macaque) 2 females, 2 males	1 – Determine if D-amphetamine increased current oxidation in the caudate of macaques, as previously reported in rodents. 2 – Determine if altered behavioral states induced by D-amphetamine affected oxidation peaks.	Triangular waveform from –0.2 to +0.6 V at 10 mV/s	*Working electrode*: Teflon coated 250 μm diameter SS wire ∼ 1 mm exposed tip *Reference electrode:* Ag/AgCl wires	D-Amphetamine treatment enhances dopaminergic signaling during presentation of both pleasant and unpleasant stimuli.
Cheney-Thamm et al., 1987	*Macaca nemestrina* (Pigtail macaque) 2 females, 1 male	Demonstrate the feasibility of using acetaminophen as *in vivo* internal standard for electrode calibration.	Triangular waveform from –0.2 to +0.6 V at 10 mV/s	*Working electrode:* Teflon coated 250 μm diameter SS wire *Reference electrode:* Ag/AgCl wire	Acetaminophen peak signal was successfully detected by FSCV in the non-human primate brain.
Earl et al., 1998	*Callithrix jacchus* (marmoset) 8 (unspecified number of males and females)	Evaluate electrically-evoked dopaminergic efflux in the striatum of normal and MPTP-treated marmosets.	Triphasic triangular waveform from –1.0 to +1.4 V and resting potential at 0 V at 480 V/s at 2 Hz (15 ms/scan)	*Working electrode:* Glass insulated 7 μm diameter CFM ∼ 20–50 μm exposed tip *Reference electrode:* Ag/AgCl wire	Dopamine signaling in the striatum remains responsive to MFB electrical stimulation after dopaminergic lesion induced by MPTP.
Shon et al., 2010	*Sus scrofa domesticus* (Swine) 4 males	1 – Investigate if electrical stimulation of the STN evokes striatal dopamine release in a large animal model. 2 – Demonstrate feasibility of performing FSCV recordings in an environment similar to human OR.	Triangular waveform from –0.4 to +1.5 V at 400 V/s at 10 Hz	*Working electrode:* glass insulated CFM (characteristics unspecified) *Reference electrode:* Ag/AgCl wire	STN electrical stimulation evoked intensity and frequency dependent striatal dopamine release.
Lee et al., 2011	*Sus scrofa domesticus* (Swine) Unknown number of animals – males	Use a wireless instantaneous neurotransmitter concentration measurement system to monitor electrochemical signaling in the brain.	Triangular waveform from –0.4 to +1.5 V at 400 V/s at 10 Hz	*Working electrode:* CFM (characteristics unspecified) *Reference electrode:* Ag/AgCl wire	Dopamine signaling responded in sigmoidal-like fashion to pulse intensity and pulse-width STN electrical stimulation.
Ariansen et al., 2012	*Macaca mulatta* (Rhesus macaque) 3 (unspecified number of males and females)	Determine if changes in pH and oxygen are associated with reward and/or reward prediction.	Dopamine – Triangular waveform from –0.4 or –0.6 V to +1.0 or +1.4 V at 400 V/s at 10 Hz. Oxygen – from 0.0 to +0.8 V, a reversal to –1.4 V, and then returned to 0.0 V, at 10 Hz.	*Working electrode:* glass insulated 12 μm diameter CFM with 250 μm exposed tip and coated with Nafion and 4-sulfobenzene; 33 μm coated with Nafion. *Reference electrode:* Ag/AgCl wire	Oxygen and pH changes were associated with the reward and cues that predicted reward. Dopamine responses evoked by reward and cues were overshadowed by pH changes detected by FSCV.
Schluter et al., 2014	*Macaca mulatta* (Rhesus macaque) 2 males	Demonstrate the feasibility of using FSCV to measure real-time changes of dopamine levels in the striatum of macaques.	Triangular waveform from –0.4 to +1.3 V at 400 V/s at 10 Hz	*Working electrode:* fused silica capillary CFM – 125–150 μm exposed tip and 7 μm diameter. *Reference electrode:* Ag/AgCl wire	Demonstrated striatal dopamine responses evoked by VTA/SNc and striatum electrical stimulation, and by unexpected rewards in awake monkeys.
Van Gompel et al., 2014	*Sus scrofa domesticus* (Swine) 3 males	Determine if extracellular adenosine concentration increases during seizure termination (in swine and humans).*	Triangular waveform from –0.4 to +1.5 V at 900 V/s at 10 Hz	*Working electrode:* CFM – 100 μm exposed tip and 7 μm diameter *Reference electrode:* Ag/AgCl wire	Increased adenosine levels were observed just prior to seizure termination. FSCV recordings were also performed in human patients.*
Yoshimi et al., 2015	*Macaca fuscata* (Japanese macaques) 3 females	Measure changes in dopamine levels associated with reward response.	Triangular waveform from –0.4 to +1.3 V at 400 V/s at 10 Hz	*Working electrode:* fused silica CFM and glass capillary CFM – 7 μm diameter and 300 μm exposed tip. *Reference electrode:* Ag/AgCl wire	Dopamine release induced by electrical stimulation and reward signals was detected in the macaque striatum by FSCV on carbon fibers.
Min et al., 2016	*Macaca mulatta* (Rhesus macaque) 3 males	Characterize striatal dopamine release evoked by STN DBS as a function of stimulating and recording electrode location.	Triangular waveform from –0.4 to +1.5 V at 400 V/s at 10 Hz	*Working electrode:* CFM – 7 μm diameter and 100 μm exposed tip. *Reference electrode:* Ag/AgCl wire	Evoked dopamine responses were higher at the stimulation of the dorsolateral posterior border of the STN.

(Continued)

TABLE 1 | (Continued)

Study	Species (common name), *n*	Goal	FSCV parameters	FSCV electrodes	Major findings
Ross et al., 2016	*Sus scrofa domesticus* (Swine) 17 (unspecified number of males and females)	Determine the functional connectivity between the medial limbic and corticolimbic circuits following fornix DBS via evoked-dopamine release in the NAc.	Triangular waveform from –0.4 to +1.5 V at 400 V/s at 10 Hz	*Working electrode:* CFM – 7 μm diameter and 100 μm exposed tip. *Reference electrode:* Ag/AgCl wire	Electrical stimulation of the fornix induced dopamine release in the NAc and increased BOLD activity in structures along the medial-corticolimbic circuitry.
Lee et al., 2017	*Rattus norvegicus* (Sprague-Dawley rats) 40 males *Sus scrofa domesticus* (Swine) 12 males *Macaca mulatta* (Rhesus macaque) 3 males	Demonstrate proof-of-principle for wireless measurement, characterization, and control of neurotransmitter release.	For dopamine and adenosine – triangular waveform from –0.4 to +1.5 V; for serotonin – N-shaped waveform: –0.4, +1.0, –0.4, +1.4 V	*Working electrode:* CFM – 7 μm diameter and 100 μm exposed tip. *Reference electrode:* Ag/AgCl wire	Demonstrated successful *in vivo*, wireless, single or multi-channel detection of dopamine, adenosine, and serotonin, with integrated sensing and stimulation feedback capabilities.
Nakajima et al., 2017	*Macaca fuscata* (Japanese macaques) 3 females	Evaluate the effects of clinically relevant STN and GPi DBS in the modulation of the activity of tonically active striatal cholinergic interneurons.	Triangular waveform from –0.4 to +1.5 V at 408.6 V/s at 10 kHz for 60 s, and resting potential at 0 V.	*Working electrode:* SS insulated CFM – 250–300 μm exposed tip. *Reference electrode:* Ag/AgCl wire	STN DBS, but not GPi DBS, induced striatal dopamine release that was correlated to increased activity of tonically active cholinergic striatum interneurons.
Schwerdt et al., 2017b	*Macaca mulatta* (Rhesus macaque) 3 females	Demonstrate the feasibility of using an integrated neurochemical modular platform for monitoring dopamine release from sensors chronically implanted in the brain of non-human primates during behavior and stimulation-evoked dopamine release.	Triangular waveform from –0.4 to +1.3 V at 400 V/s at 10 Hz	*Working electrode:* Array of chronic and acute fused silica insulated CFMs – 7 μm diameter and 150–300 μm exposed tip *Reference electrode:* Ag/AgCl wire or SS electrode	Modular platform allowed measurements of dopamine release from multiple sites in the striatum while electrically stimulating the SNc/VTA for up to 170 days.
Settell et al., 2017	*Sus scrofa domesticus* (Swine) 4 (unspecified number of males and females)	Determine the neuromodulatory effects of VTA DBS on dopamine release in the NAc.	Triangular waveform from –0.4 to +1.5 V; or N-shaped waveform: –0.4, +1.0, –0.4, +1.4 V	*Working electrode:* CFM – 7 μm diameter and 100 μm exposed tip. *Reference electrode:* not specified	VTA DBS resulted in increased dopamine release in the NAc and increased BOLD activity in the striatum, cortical and limbic structures.
Trevathan et al., 2017	*Rattus norvegicus* (Sprague-Dawley rats) 12 females *Sus scrofa domesticus* (Swine) 4 males *Macaca mulatta* (Rhesus macaque) 1 male	Characterize subject-specific kinetics of stimulation-evoked dopamine release using computational modeling.	Triangular waveform from –0.4 to +1.5 V at 400 V/s at 10 Hz	*Working electrode:* CFM – 7 μm diameter and 100 μm exposed tip. *Reference electrode:* not specified	Dopamine dynamics in response to electrical stimulation was modeled and characterized based on non-linear increased responses of dopamine to increasing SNc/VTA stimulus intensity.

SS, stainless steel; FSCV, fast scan cyclic voltammetry; MPTP, 1-methyl-4-phenyl-1,2,3,6-tetrahydropyridine CFM, carbon fiber microelectrode; MFB, medial forebrain bundle; STN, subthalamic nucleus; OR, operating room; VTA, ventral tegmental area; SNc, substatia nigra pars compacta; DBS, deep brain stimulation; NAc, nucleus accumbens; BOLD, blood-oxygen-level-dependent imaging; and GPi, globus pallidus internum.
See **Table 3.*

be attributed mainly to their small size, which is critical for minimizing damage to neural tissue and reducing the impact on neurochemical transmission near the electrode (see **Figure 2** for CFM and DBS electrode leads size comparison). Similarly, CFM have shown good performance in maintaining normal neurotransmitter kinetics at the measurement site and high adsorption of neurotransmitters of interest (Borland et al., 2005; Wang and Michael, 2012; Jaquins-Gerstl and Michael, 2015). However, despite their desirable neurochemical detection properties, CFMs are not optimal for clinical applications due to their high manufacturing variability (Roberts and Sombers, 2018) and fragility, which increase the risk of breakage during implantation into brain tissue (Schluter et al., 2014). Similarly, surface modified CFM made by coating with polymer films

TABLE 2 | Characteristics of commonly used fast scan cyclic voltammetry working electrodes in *in vivo* studies.

Electrode material	Type of electrode modification	Insulation	Advantages*	Disadvantages*	References
Carbon fiber	N/A	Borosilicate glass	Ease of fabrication, low cost, biocompatible, recommended for acute recordings.	Fragile, high risk of breakage, poor fabrication uniformity, susceptible to fouling.	Patel et al., 2013; Bucher and Wightman, 2015; Rodeberg et al., 2017; Roberts and Sombers, 2018
Carbon fiber	N/A	Fused-silica capillaries sealed with epoxy	Good insulating properties, increased flexibility, low cost, lower risk of breakage, small diameter, biocompatible, recommended for chronic recordings.	Susceptible to fouling, poor fabrication uniformity, torsion applied to the carbon fiber not well translated to the fused-silica insulation.	Clark et al., 2010; Rodeberg et al., 2017; Roberts and Sombers, 2018
Carbon nanotube	Surface-modified CFM	Borosilicate glass or Fused-silica capillaries sealed with epoxy	High mechanical strength, improved electron-transfer kinetics, high sensitivity to adsorbed dopamine	Poor fabrication uniformity and reproducibility.	Bucher and Wightman, 2015; Ganesana et al., 2017; Yang et al., 2017
Gold-nanoparticles	Surface-modified CFM	Borosilicate glass or Fused-silica capillaries sealed with epoxy	High electroactive surface area, fast electron-transfer kinetics, highly sensitive	Increased fouling	Zhao et al., 2012; Mohanaraj et al., 2019
Carbon fiber	Nafion-coated CFM	Borosilicate glass or Fused-silica capillaries sealed with epoxy	Improved selectivity and sensitivity for dopamine, fast response time	Poor uniformity and reproducibility in creating Nafion film around the electrode, fouling susceptibilibility is similar to bare CFM.	Brazell et al., 1987; Pihel et al., 1996; Roberts and Sombers, 2018
Carbon fiber	PEDOT:Nafion-coated CFM	Borosilicate glass or Fused-silica capillaries sealed with epoxy	Improved selectivity and sensitivity for dopamine, fast response time, low fouling	PEDOT is not approved for human use	Vreeland et al., 2015; Ganesana et al., 2017; Boehler et al., 2020
Diamond	Boron doped conductive diamond	Borosilicate glass sealed with resin or Stainless tube sealed with resin or Parylene	Wide potential window, low and stable background current, low fouling, improved longevity and resistance	Low sensitivity, slow electron-transfer kinetics, costly fabrication, poor doping may lead to impurities on the surface	Yoshimi et al., 2011; Patel et al., 2013; Bennet et al., 2016; Ganesana et al., 2017; Chang et al., 2019
Platinum	Metal	Unknown	Less susceptible to breakage, versatile for surface modification and assembly into arrays for multi-neurotransmitter monitoring	Susceptible to corrosion, passivation, increased fouling, low sensitivity.	Jackowska and Krysinski, 2013; Roberts and Sombers, 2018
Gold	Metal	Unknown	Less susceptible to breakage, versatile for surface modification and assembly into arrays or miniature gold electrodes, good sensitivity to catecholamines,	Susceptible to corrosion, passivation, increased fouling, low sensitivity	Zachek et al., 2008; Jackowska and Krysinski, 2013; Roberts and Sombers, 2018

CFM, Carbon Fiber Microelectrode; PEDOT, poly(3,4-ethylenedioxythiophene).
**Relative to borosilicate glass CFM.*

such as Nafion and PEDOT provides enhanced sensitivity, selectivity, and kinetic properties for dopamine detection. However, electrode reproducibility is still poor due to the lack of coating uniformity (Brazell et al., 1987; Pihel et al., 1996; Vreeland et al., 2015; Ganesana et al., 2017; Roberts and Sombers, 2018). In addition, PEDOT is not currently approved for clinical use due to the large range of PEDOT variations determined by multiple factors such as the fabrication technique, the dopant, and the functionalization of the polymer (Boehler et al., 2020).

As an alternative to CFM and surface modified CFM, several groups have attempted to develop Boron-doped diamond (BDD) electrodes, which offer increased strength and longevity compared to CFM (Martin et al., 1998; Suzuki et al., 2007; Yoshimi et al., 2015; Bennet et al., 2016). However, BDD has distinct electrochemical properties and significant characterization studies must be performed to better understand and optimize BDD for neurochemical sensing in the clinical setting. To date, studies have shown that BDD electrodes are not suitable for chronic *in vivo* neurochemical detections due

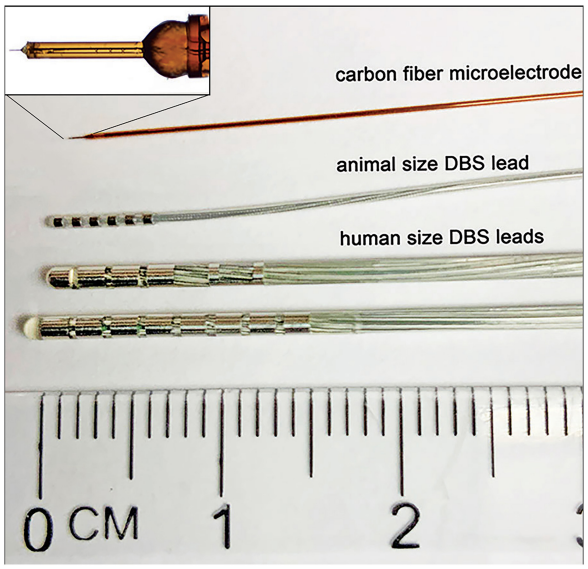

FIGURE 2 | Comparison between typically used FSCV working electrode and DBS electrode leads. From top to bottom: In-house made epoxy and silica tubing-insulated CFM for detection of neurotransmitters via FSCV; Animal size 6-contact DBS electrode lead – NuMED, Inc. (2.0 Fr shaft,0.5 mm spacing, and 42 cm length); Human size DBS electrode leads – Boston Scientific, Inc. (Vercise DBS Directional Lead, 6.0 mm^2 dome tip contact surface area, 1.5 mm^2 segmented contacts surface area, 6.0 mm^2 ring contacts surface area, 1.5 mm contact length, 2.0 mm spacing, 1.3 mm diameter, 30–45 cm length; and Vercise DBS lead, 1.5 mm contact length, 2.0 mm spacing, 1.3 mm diameter, 30–45 cm length).

to poor sensitivity and neurochemical adsorption rate, increased fouling, and distorted signals (Yoshimi et al., 2011; Jackowska and Krysinski, 2013; Chang et al., 2019).

Metal electrodes such as Gold and Platinum electrodes provide increased mechanical strength being more resistant to breakage during implantation. Additionally, these materials allow for novel electrode arrays and the capability of multiple neurotransmitter simultaneous detection. However, metal electrodes are hardly used for clinical applications because of the risk of corrosion that can lead to tissue damage, passivation and increased biofouling that interfere with neurotransmitter detection (Zachek et al., 2008; Jackowska and Krysinski, 2013; Roberts and Sombers, 2018).

Although novel electrode materials or configurations, such as Platinum, Gold, and Nafion-coated CFM described above, provide numerous benefits, the representation of signal and artifacts *in vitro* and *in vivo* can vary significantly to those measured with CFMs (Bucher and Wightman, 2015). For this reason, it is important to fully validate novel working electrode designs via large animal experiments with similar signal levels and experimental paradigms to anticipated human studies.

Reference Electrode Biocompatibility and Stability

Reference electrodes are pivotal to ensure stable, consistent, and neurotransmitter-specific measurements. Reference electrodes should not be polarized (i.e., pulled away from their open circuit potential) during FSCV to avoid sensor drift, low signal-to-noise ratio, and distorted neurochemical signals (Rodeberg et al., 2017). Thus, investigators need to ensure that polarization is avoided by increasing the surface area of the reference electrodes relative to the active CFM, as well as by using materials that have a stable *in vivo* reference potential and low polarizability properties. Reference electrodes for *in vivo* neurotransmitter measurements in animal models are typically made of Ag/AgCl, which are highly cytotoxic making them unsafe for long-term (in the order of hours/days) human use (Hashemi et al., 2011; Kolli et al., 2015; Seaton et al., 2020). To avoid the safety concerns associated with Ag/AgCl electrodes, stainless steel (SS) reference electrodes have been used in humans (Kishida et al., 2011; Bucher and Wightman, 2015). Unfortunately, SS is highly polarizable and has an inhomogeneous grain structure that decreases the stability of its open circuit potential by changing the protein adsorption to the electrode surface, which can shift its open circuit potential (Rajendrachari, 2018). The use of SS reference electrodes for clinical FSCV (see **Table 3**) is an unresolved confound that can significantly affect measurement sensitivity and specificity. Thus, there is a critical need for the pre-clinical investigation and validation of novel reference electrodes that are not only biocompatible, but also that provide stable reference potentials prior to human use.

Tissue Damage and Signal Integrity

Implantation of both working and reference electrodes disrupts the blood-brain barrier, triggering a cascade of complex molecular and cellular responses such as activation of glial cells, loss of perfusion, secondary metabolic injury, neuronal degeneration, and introduction of mechanical strain, which can affect neurochemical sensing (Woolley et al., 2013; Kozai et al., 2015). For example, studies have shown that brain hemorrhage associated with electrode implantation in the human brain can result in 1–2% of symptomatic hemorrhage (i.e., speech arrest, hemiplegia, agitation, and confusion) and 0.5–0.9% of permanent neurological deficit (Sansur et al., 2007; Park et al., 2011). Therefore, any clinical FSCV work should minimize the risks associated with mechanical tissue damage through rigorous pre-clinical studies aimed at understanding and reducing mechanical damage to neural tissue (ISO10993-6:2016, 2016).

Reduction/oxidation (redox) reactions can also disrupt neurochemical diffusion to the region of the sensing electrode, changing neurochemical concentrations near the electrode and affecting detection of the neurochemical species of interest (Wang and Michael, 2012; Jaquins-Gerstl and Michael, 2015).

Carbon fiber microelectrodes used in pre-clinical studies have traditionally been fabricated from silica glass capillaries or tubing with a small implantation profile (Bucher and Wightman, 2015; Rodeberg et al., 2017). However, clinical FSCV studies have mimicked setups used clinically to implant electrophysiological sensing electrodes and relied on stainless steel cannulas to implant either CFM or diamond electrodes, e.g., FHC#66-IT(AO6) 165 mm SS insertion tube for placing depth electrodes along a trajectory. The effect of the relatively large tissue insult

TABLE 3 | FSCV studies in humans.

Study	Goal	FSCV parameters	FSCV electrodes	Major findings
Kishida et al., 2011	Demonstrate feasibility of sub-second dopamine detection in humans during brain surgery.	Triangular waveform from –0.425 V to +1.380 at 400 V/s at 10 Hz.	*Working electrode*: Glass insulated 7 μm CFM *Reference electrode*: Stainless steel	First demonstration of real-time human dopamine release measured during a behavioral investment task.
Chang et al., 2012	Quantify adenosine concentrations in human subjects with essential tremor during VIM DBS.	Triangular waveform from –0.4 V to +1.5 at 400 V/s at 10 Hz.	*Working electrode*: Glass and polyamide insulated 7 μm CFM with 50 μm exposed tip or silicone insulated 30 μm CFM *Reference electrode*: Stainless steel	Demonstrated changes in extracellular concentrations of adenosine in the VIM of patients undergoing VIM DBS.
Van Gompel et al., 2014	Determine changes in extracellular adenosine concentration during epileptic activity (in swine and humans).*	Triangular waveform from –0.4 V to +1.5 at 400 V/s at 10 Hz.	*Working electrode*: Glass insulated 7 μm CFM *Reference electrode*: Stainless steel, 24 gage	Extracellular adenosine concentration increases prior to seizure termination.
Bennet et al., 2016	Demonstrate sensitivity of diamond-based electrode to detect and quantify neurotransmitter release in human patients undergoing DBS.	Triangular waveform from –0.4 V to +1.5 at 400 V/s at 10 Hz.	*Working electrode*: Parylene-C coated boron-doped diamond, conical shape, 50 μm at the base and 100 μm long. *Reference electrode*: Stainless steel	Dopamine and adenosine release were detected using boron-doped diamond electrode in human patients undergoing DBS for tremor.
Kishida et al., 2016	Determine whether subsecond dopamine fluctuations in the human striatum encode reward prediction errors during a sequential choice task.	Triangular waveform from –0.6 to +1.4 V at 400 V/s at 10 Hz.	*Working electrode*: 7 μm CFM insulated with a polyimide-coated fused-silica capillary *Reference electrode*: Stainless steel	Demonstrated striatal dopamine changes encoding experience-dependent reward prediction error and counterfactual prediction error.
Lohrenz et al., 2016	Compare BOLD activity and dopamine responses during a sequential choice task in humans.	Triangular waveform from –0.6 to +1.4 V at 400 V/s at 10 Hz.	*Working electrode*: 7 μm CFM insulated with a polyimide-coated fused-silica capillary *Reference electrode*: Stainless steel	Demonstrated non-linear relationship between BOLD activity and dopamine release of subjects performing an investment task.
Moran et al., 2018	Demonstrate changes in the serotonergic signaling during a decision making task.	Triangular waveform from –0.6 to +1.4 V at 400 V/s at 10 Hz.	*Working electrode*: 7 μm CFM insulated with a polyimide-coated fused-silica capillary *Reference electrode*: Stainless steel	Striatal serotonin correlates with decisions in a sequential investment game and may encode a strategy that modulates choice selection to mitigate risk.
Montague and Kishida, 2018	Demonstrate a new approach for simultaneous multi-neurotransmitter detection and quantification.	Triangular waveform from –0.6 to +1.4 V at 400 V/s at 10 Hz.	*Working electrode*: 7 μm CFM insulated with a polyimide-coated fused-silica capillary *Reference electrode*: Stainless steel	The elastic net approach extracted a concentration-prediction model for multiple analytes that included dopamine, serotonin, and norepinephrine.

FSCV, fast scan cyclic voltammetry; CFM, carbon fiber microelectrode; VIM, ventral intermediate nucleus of thalamus; DBS, deep brain stimulation; and BOLD, blood-oxygen-level-dependent imaging.
See **Table 1.*

resulting from cannula insertion on measured FSCV signals is not well understood and should be explored in large animal models.

Spatial Resolution and Heterogeneity of Neurotransmitter Release

The high spatial resolution of FSCV, determined by the size of the working electrode relative to the volume of neurochemical release, is one of the driving motivators for use of FSCV to detect neurochemical signaling. Working electrodes are typically made from 7 to 30 μm diameter carbon fibers cut to 50–500 μm lengths (see **Figure 2**). It is the small size of the CFM that allows for rapid scan rates (>100 V/s) which enables neurotransmitter measurements on a subsecond timescale without producing large charging currents potentially harmful to the brain tissue (Zachek et al., 2009). The spatial resolution of the CFM is on the order of tens to hundreds of microns (Rodeberg et al., 2017), much higher than microdialysis (Jaquins-Gerstl and Michael, 2015) and functional imaging (Glover, 2011). FSCV high spatial resolution allows for neurotransmitter measurement in discrete brain regions, which could potentially serve as guidance for improved targeting of DBS electrode implantation.

It must be considered that studies have shown inconsistent dopamine kinetics across different microdomains in small and large animal models (Wightman et al., 2007; Moquin and Michael, 2009; Taylor et al., 2015), which can lead to data misinterpretation and demonstrate that FSCV recordings from a single CFM are not adequate for testing specific hypotheses of systemic neurotransmission. Thus, measurements obtained at these high resolutions may not accurately represent activity within the target brain nuclei due to sparse synaptic organization, neuronal heterogeneity, and differences in neurochemical release kinetics throughout the structures of interest. These inconstancies in neurotransmitter release across a target of interest, coupled with limited recording time in the operating room, can make finding a suitable recording site difficult for clinical studies. Sparsity and heterogeneity can be addressed by developing FSCV electrode arrays capable of simultaneously

sampling multiple brain regions (Zachek et al., 2010; Patel et al., 2016; Schwerdt et al., 2017a). This, however, comes at the expense of increased risk of hemorrhage and inflammation, which must be weighed carefully against the expected benefit of the study.

Effects of Sterilization of Fast Scan Cyclic Voltammetry Electrodes on Recorded Signals

Sterilization of implantable devices is paramount to reduce the risk of infection. However, this creates a challenge for FSCV measurements as sterilization procedures can alter the geometrical and chemical surface of the CFM, compromise electrode insulation, affect electrical connections, and ultimately, hinder biocompatibility (Nair, 1995; Godara et al., 2007; Zheng et al., 2011; Chong et al., 2015). For example, CFM working electrodes and Ag/AgCl reference electrodes are composed of heat-sensitive materials, which makes steam and dry heat sterilization inappropriate because these techniques can cause polymer degradation and changes in physical or mechanical properties (Nair, 1995; Mendes et al., 2007). Thus, sterilization procedures should not only be evaluated for sterility prior to clinical use, but also for the effect of surface modification on signal specificity and accuracy.

Effects of the Operating Room Environment on Recorded Signals

Continuous electrochemical recording is associated with baseline instability and sensor drift as a function of time, ultimately reducing measurement accuracy and reproducibility (DeWaele et al., 2017; Mitchell et al., 2017). This drift can be caused by a myriad of external factors such as temperature fluctuations, changes in the chemical environment, or chemical reactions at the sensing surface (Roberts and Sombers, 2018). The chemical environment can be affected by biofouling caused by extracellular proteins, redox reaction debris, cellular encapsulation around the surface of the electrode, and many other factors (Kachoosangi and Compton, 2007; Patel et al., 2013; Nicolai et al., 2017; Seaton et al., 2020). To minimize these effects, several strategies such as sensor cleaning, recalibration, and even replacement have been used (Hermans et al., 2008; DeWaele et al., 2017; Mitchell et al., 2017). However, these are not feasible in an intraoperative setting, where available time is limited. Additionally, the relatively long period of stabilization following implantation of an FSCV electrode (Ramsson et al., 2015; Rodeberg et al., 2017; Castagnola et al., 2020) limits the time available for data collection in the operating room. Furthermore, the operating room environment has numerous sources of noise. These are known to affect electrophysiological recordings, but their effects have not been characterized for electrochemical measurements using FSCV.

Challenges for Data Analysis

One of the most challenging aspects of FSCV is the complex and variable nature of the data (Olivieri, 2008; Johnson et al., 2016). Signals obtained *in vivo* often have contributions from multiple analytes that require resolution prior to positive identification and quantification. Consistency and accuracy can be improved with automated multivariate statistical data analyses, such as principal component regression, partial least squares regression, and statistical models. Typically, information across the scan-potential window can be used to separate overlapping signals by using training sets (i.e., signals obtained from electrodes, recording sessions, and/or subjects other than those used for experimental data collection) as calibration models (Johnson et al., 2016, 2017; Kishida et al., 2016; Lohrenz et al., 2016; Rodeberg et al., 2017; Meunier et al., 2018). The effectiveness of these analysis techniques depends on the existence of well-characterized relationships between the potential at the working electrode and the measured redox currents for the neurochemical species of interest. For clinical FSCV, analysis algorithms should be validated *in vivo* using the exact recording setup that will be used clinically, at signal levels comparable to those normally observed in pre-clinical settings.

Recent studies adopted a data analysis algorithm built from a novel statistical model trained on *in vitro* data recorded against Ag/AgCl reference electrodes. However, the algorithm was applied to data recorded using SS reference electrodes in patients during a task performed in the operating room without previous pre-clinical validation (Kishida et al., 2011; Montague and Kishida, 2018). Pre-clinical validation using the Five Golden Rules afore mentioned should be performed prior to clinical use of any analysis algorithms. Thus, algorithms such as the one described above remains largely unvalidated until it is tested both *in vitro* and *in vivo* in a pre-clinical model using the same type of working and reference electrodes as those that will be used clinically.

OTHER CONSIDERATIONS AND TECHNICAL CHALLENGES TO FAST SCAN CYCLIC VOLTAMMETRY IN HUMANS

Patient Safety

Patient safety is paramount whenever clinical studies are performed. As with all clinical studies, the onus of determining safety risk lies with the investigator and local IRB, regardless of whether the study involves the use of a device subjected to an Investigational Device Exemption or not (Code of Federal Regulations, 2021a). While not all risk can be eliminated, it is the responsibility of the investigator and local IRB to create a risk management process in order to identify, isolate, and mitigate as much risk as possible (ISO14971:2019, 2019). In this way, not only is the well-being of the patient protected, but also is the integrity, rigor, and reproducibility of the collected data and resulting conclusions.

Material Biocompatibility

Any material or device that comes into contact with tissue, especially neural tissue, in a clinical study should be evaluated for biocompatibility. Although extensive discussion of biocompatibility testing is outside of the scope of this work, guidance for this biocompatibility testing is provided by the

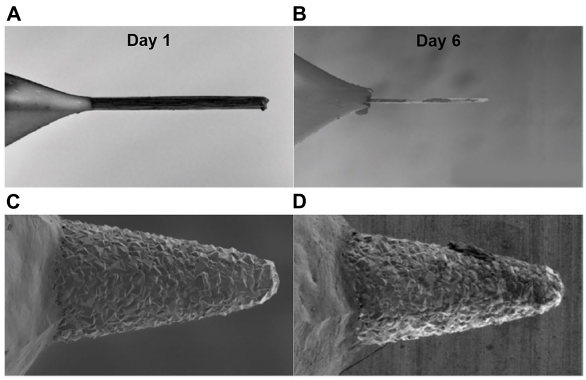

FIGURE 3 | Impact of FSCV waveform on CFM and BDD electrodes longevity. Scanning electron microscope images demonstrate the surface erosion of CFMs **(A,B)** and BDDs **(C,D)** on day 1 **(A,C)** and day 6 **(B,D)** of continuous FSCV. CFMs were significantly eroded after 6 days of continuous FSCV while BDD electrodes show no noticeable changes. A FSCV triangular waveform from -0.4 to 1.5 V and back to -0.4 V at 10 Hz at 400 V/s was continuously applied to CFMs and BDDs immersed in Tris buffer over an extended period of time. Adapted from Bennet et al. (2016).

ISO 10993 standard (ISO10993-1:2018, 2018). It is important to note that even if the material itself is biocompatible, there is a risk of harmful substances to be present on the material as byproducts of manufacturing or other processing steps. In commercially marketed medical devices, these risks are mitigated by manufacturing under Good Manufacturing Practice protocols (Code of Federal Regulations, 2021b), which help control any deviation between manufacturing batches. Currently, no medical device suppliers provide electrodes for FSCV, so it falls on the investigator to ensure consistency between materials tested for biocompatibility and those that are used for clinical FSCV work.

Electrochemical Reactions Due to Waveform Selection

Different electrode materials exhibit unique electrochemical properties that may lead to changes in the waveforms that are used for neurochemical detection. These waveforms, in turn, affect the various safety considerations listed above. Thus, testing of the electrical FSCV potential waveforms is required to determine the range of voltages that can be safely applied to the tissue without inducing excitotoxicity, electrolysis, and reduced electrode longevity. For example, harmful reactive oxygen species are generated when oxygen is reduced (negative potentials) or when water is oxidized (positive potentials; Roberts and Sombers, 2018). However, the potential at which these redox reactions occur varies depending on the water window for the specific electrode material. For voltage-controlled applications such as FSCV, the water window informs whether a particular waveform will result in electrolysis, which is damaging to tissue (Cogan et al., 2004; Cogan, 2008). Therefore, all FSCV waveforms used should be evaluated for tissue damage and safety for each electrode material and configuration should be stablished.

It is also vital to ensure that the waveform does not damage the electrode itself, as material dissolution is detrimental to the longevity of the electrode (**Figure 3**) and the released particulate could be harmful to tissue. To prevent electrode damage, the largest potentials of a waveform should be applied to the electrode in solution in a closed-system benchtop setup over an extended period of time and optionally, at a higher frequency than planned. The beaker solution should be sampled at regular intervals and can be tested with mass spectrometry to determine whether the electrode material is dissolved into solution (Boehler et al., 2020). Subsequent biological tests should be conducted to determine whether the intended waveforms would damage cells due to excitotoxicity by conducting safety tests of electrical stimulation in neural tissue

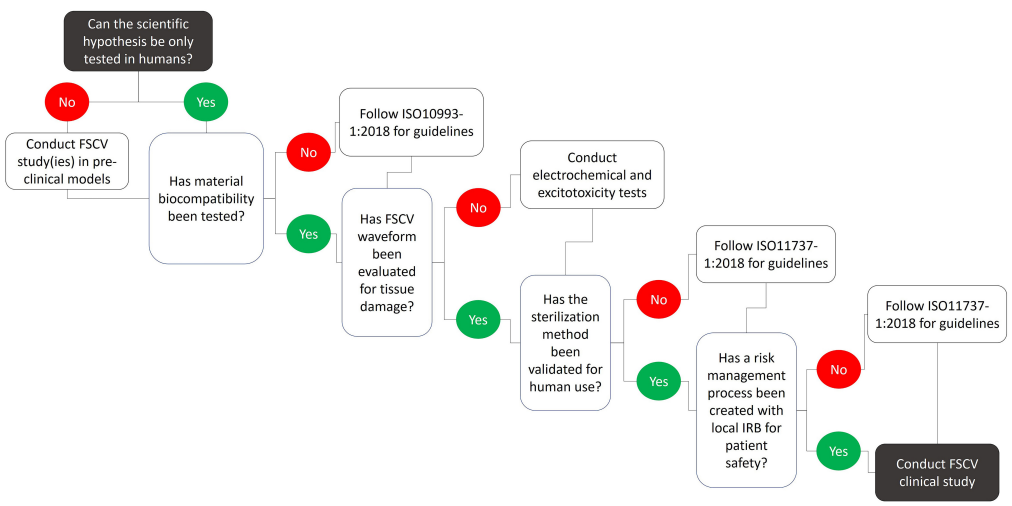

FIGURE 4 | Technical considerations and recommendations for conducting FSCV studies in humans. Numerous limitations associated with FSCV limit the type and number of scientific questions that can be answered by clinical FSCV studies. These limitations pose safety risks to the patient and need to be addressed prior to the clinical use of FSCV. Extensive pre-clinical testing including material biocompatibility, waveform safety, tissue damage, and sterilization methods must be conducted following specific guidelines in order to develop a risk management process in cooperation with local IRBs to establish patient safety.

(McCreery et al., 1988, 1990, Robblee and Rose, 1990; Shannon, 1992). Finally, pre-clinical studies with thorough histological evaluation of tissue adjacent to and in the vicinity of the implanted electrode should be conducted prior to human studies (Cogan et al., 2004, 2016).

Sterilization Validation

As would be expected for any invasive medical device, sterilization validation is essential to patient safety (ISO11737-1:2018, 2018). Balancing this need with that of ensuring electrode integrity may require exploration of novel sterilization methods. As previously discussed, the heat-sensitive nature of some of the components commonly used for FSCV electrodes, steam or dry heat sterilization methods are not viable (Nair, 1995; Mendes et al., 2007). However, sterilization methods that do not rely on heat pose other challenges to patient safety. For example, γ radiation generates free radicals that can cause tissue damage (Ries et al., 1996; Gorna and Gogolewski, 2003; Mendes et al., 2007). Similarly, ethylene oxide and its derivatives have been reported to accumulate in some materials, potentially increasing teratogenicity risk (Lucas et al., 2003; Mendes et al., 2007). As previously mentioned, the investigator should characterize and validate each sterilization technique for each FSCV sensor type, composition, and geometry to ensure patient safety, in addition to signal accuracy.

DISCUSSION

Many of the safety and efficacy concerns associated with FSCV recordings in humans are due to interrelated technical challenges, which makes these challenges difficult to address. However, once the scientific and clinical benefits of the study are clearly defined and the risks to patient safety minimized, residual risks can be weighed against anticipated benefits. This risk-benefit analysis is crucial to the success of the clinical work.

Investigators should perform a compelling risk/benefit analysis to justify the clinical use of FSCV, ensure scientific rigor, and articulate the scientific and clinical benefit of FSCV studies. Addressing and mitigating risk to patient safety will help ensure patients are not put at undue risk. Of course, not all risk can be eliminated, and it is this remaining risk that must be weighed against the proposed clinical benefit. Furthermore, it is essential that both risks and benefits are appropriately and timely communicated to patients through consent forms so patients can make an informed decision about their participation.

Risk-benefit analyses are complex, so the authors propose that the FSCV community develop the necessary guidelines through partnership with the FDA to engender high-quality scientific discoveries that FSCV can offer.

Clinical studies need not demonstrate that FSCV provides direct patient benefit. However, evidence from pre-clinical studies should be used to guide approval for use in human subjects. Application of FSCV in clinical studies should be carefully evaluated to ensure that only hypothesis-driven studies that cannot be appropriately and methodically performed in animal models are conducted in human subjects. Thus, translational FSCV research should demonstrate scientific rationale, feasibility, efficacy, and safety prior to attempting clinical application.

While progress has been made in the clinical use of FSCV (**Table 3**), it must be stressed that electrode safety and data robustness have yet to be characterized. To date, clinical FSCV studies have employed working and reference electrodes and signal processing techniques that have not been validated by translational research. The Five Golden Rules are not directly applicable in the clinical setting. Thus, the operating room environment necessitates the reconsideration of human FSCV methodologies. Specifically, the following processes need to be performed: acute and chronic target assessment and confirmation, standardization of electrode manufacturing processes, signal verification in human recordings, and characterization of sources of electrical noise in the intra-operative environment. Addressing these issues will facilitate application of FSCV in the clinical setting, particularly if FDA-approved working and reference electrodes become commercially available.

The use of FSCV in humans must be understood as a clinical intervention, and as such, should follow the same protocols and guidelines required by the FDA to establish device safety (**Figure 4**). Additionally, there are many gaps in knowledge that complicate the risk/benefit analysis for the clinical use of FSCV during existing surgical procedures. Until these gaps are addressed, the likelihood of obtaining useful scientifically rigorous data is low due to all the unknowns and the quality of the recorded signal in the operating room environment. Thus, clinical FSCV should only be performed when there is a clear hypothesis that can further neuroscientific understanding in the pursuit of improved therapeutic interventions, and only after all risk factors discussed in this manuscript have been mitigated.

AUTHOR CONTRIBUTIONS

SLB, JT, SH, and JLL contributed to the conceptualization and design of this work. SLB, JT, and SH wrote the first draft. AA wrote sections of this work. SLB was responsible for the project administration and supervision. JLL supervised the project. All authors contributed to manuscript revision, read, and approved the submitted version review and editing.

FUNDING

This work was supported by the National Institutes of Health, NINDS (R01 NS084975 award to JLL) and The Grainger Foundation.

ACKNOWLEDGMENTS

We thank Kai Miller for providing human size deep brain stimulation electrode lead (Boston Scientific, Inc., Vercise DBS electrode lead) used in **Figure 2**.

REFERENCES

Adams, R. N. (1976). Probing brain chemistry with electroanalytical techniques. *Anal. Chem.* 48, 1126A–1138A. doi: 10.1021/ac50008a001

Ariansen, J., Heien, M. L. A. V., Hermans, A., Phillips, P. E. M., Hernadi, I., Bermudez, M., et al. (2012). Monitoring extracellular pH, oxygen, and dopamine during reward delivery in the striatum of primates. *Front. Behav. Neurosci.* 6:36. doi: 10.3389/fnbeh.2012.00036

Bennet, K. E., Tomshine, J. R., Min, H.-K., Manciu, F. S., Marsh, M. P., Paek, S. B., et al. (2016). A diamond-based electrode for detection of neurochemicals in the human brain. *Front. Hum. Neurosci.* 10:102. doi: 10.3389/fnhum.2016.00102

Boehler, C., Carli, S., Fadiga, L., Stieglitz, T., and Asplund, M. (2020). Tutorial: guidelines for standardized performance tests for electrodes intended for neural interfaces and bioelectronics. *Nat. Protocols* 15, 3557–3578. doi: 10.1038/s41596-020-0389-382

Borland, L. M., and Michael, A. C. (2007). "An introduction to electrochemical methods," in *Neuroscience. Electrochemical Methods for Neuroscience*, eds A. C. Michael and L. M. Borland (Boca Raton, FL: Taylor & Francis Group, LLC.).

Borland, L. M., Shi, G., Yang, H., and Michael, A. C. (2005). Voltammetric study of extracellular dopamine near microdialysis probes acutely implanted in the striatum of the anesthetized rat. *J. Neurosci. Methods* 146, 149–158. doi: 10.1016/j.jneumeth.2005.02.002

Brazell, M. P., Kasser, R. J., Renner, K. J., Feng, J., Moghaddam, B., and Adams, R. N. (1987). Electrocoating carbon fiber microelectrodes with Nafion improves selectivity for electroactive neurotransmitters. *J. Neurosci. Methods* 22, 167–172. doi: 10.1016/0165-0270(87)90011-90012

Bucher, E. S., and Wightman, R. M. (2015). Electrochemical analysis of neurotransmitters. *Ann. Rev. Anal. Chem.* 8, 239–261. doi: 10.1146/annurev-anchem-071114-40426

Castagnola, E., Robbins, E. M., Woeppel, K. M., McGuier, M., Golabchi, A., Taylor, I. M., et al. (2020). Real-Time fast scan cyclic voltammetry detection and quantification of exogenously administered melatonin in mice brain. *Front. Bioeng. Biotechnol.* 8:602216. doi: 10.3389/fbioe.2020.602216

Chang, S.-Y., Jay, T., Muñoz, J., Kim, I., and Lee, K. H. (2012). Wireless fast-scan cyclic voltammetry measurement of histamine using WINCS – a proof-of-principle study. *Analyst* 137, 2158–2165. doi: 10.1039/c2an16038b

Chang, A.-Y., Dutta, G., Siddiqui, S., and Arumugam, P. U. (2019). Surface fouling of ultrananocrystalline diamond microelectrodes during dopamine detection: improving lifetime via electrochemical cycling. *ACS Chem. Neurosci.* 10, 313–322. doi: 10.1021/acschemneuro.8b00257

Cheney-Thamm, J., Reite, M., Alianiello, E. A., Yamamoto, B. K., Capitanio, J. P., and Freed, C. R. (1984). Caudate electrochemical response following amphetamine administration in pigtail monkeys. *Life Sci.* 35, 1453–1460. doi: 10.1016/0024-3205(84)90161-90169

Cheney-Thamm, J., Alianello, E. A., Freed, C. R., and Reite, M. (1987). In vivo electrochemical recording of acetaminophen in non human primate brain. *Life Sci.* 40, 375–379.

Chong, A., Fischer, G., Dart, B. R., and Wooley, P. H. (2015). Long-term elastic durability of polymer matrix composite materials after repeated steam sterilization. *Am. J. Orthoped.* 44, E427–E433.

Clark, J. J., Sandberg, S. G., Wanat, M. J., Gan, J. O., Horne, E. A., Hart, A. S., et al. (2010). Chronic microsensors for longitudinal, subsecond dopamine detection in behaving animals. *Nat. Methods* 7, 126–129. doi: 10.1038/nmeth.1412

Code of Federal Regulations (2021a). *Part 812 – Investigational Device Exemptions. Title 21 Chapter I – Subchapter H.* e-CFR.

Code of Federal Regulations (2021b). *Part 820 – Quality System Regulation. Title 21 – Chapter I – Subchapter H.* e-CFR.

Cogan, S. F. (2008). Neural stimulation and recording electrodes. *Annu. Rev. Biomed. Eng.* 10, 275–309. doi: 10.1146/annurev.bioeng.10.061807.160518

Cogan, S. F., Guzelian, A. A., Agnew, W. F., Yuen, T. G. H., and McCreery, D. B. (2004). Over-pulsing degrades activated iridium oxide films used for intracortical neural stimulation. *J. Neurosci. Methods* 137, 141–150. doi: 10.1016/j.jneumeth.2004.02.019

Cogan, S. F., Ludwig, K. A., Welle, C. G., and Takmakov, P. (2016). Tissue damage thresholds during therapeutic electrical stimulation. *J. Neural Eng.* 13:021001. doi: 10.1088/1741-2560/13/2/021001

DeWaele, M., Oh, Y., Park, C., Kang, Y. M., Shin, H., Blaha, C. D., et al. (2017). A baseline drift detrending technique for fast scan cyclic voltammetry. *Analyst* 142, 4317–4321. doi: 10.1039/c7an01465a

Earl, C. D., Sautter, J., Xie, J., Kruk, Z. L., Kupsch, A., and Oertel, W. H. (1998). Pharmacological characterisation of dopamine overflow in the striatum of the normal and MPTP-treated common marmoset, studied in vivo using fast cyclic voltammetry, nomifensine and sulpiride. *J. Neurosci. Methods* 85, 201–209. doi: 10.1016/S0165-0270(98)00134-134

Ganesana, M., Lee, S. T., Wang, Y., and Venton, B. J. (2017). Analytical techniques in neuroscience: recent advances in imaging, separation, and electrochemical methods. *Anal. Chem.* 89, 314–341. doi: 10.1021/acs.analchem.6b04278

Glover, G. H. (2011). Overview of functional magnetic resonance imaging. *Neurosurg. Clin. N. Am.* 22, 133–139. doi: 10.1016/j.nec.2010.11.001

Godara, A., Raabe, D., and Green, S. (2007). The influence of sterilization processes on the micromechanical properties of carbon fiber-reinforced PEEK composites for bone implant applications. *Acta Biomaterialia* 3, 209–220.

Gorna, K., and Gogolewski, S. (2003). The effect of gamma radiation on molecular stability and mechanical properties of biodegradable polyurethanes for medical applications. *Polymer Degradation Stabil.* 79, 465–474.

Hashemi, P., Walsh, P. L., Guillot, T. S., Gras-Najjar, J., Takmakov, P., Crews, F. T., et al. (2011). Chronically implanted, nafion-coated Ag/AgCl reference electrodes for neurochemical applications. *ACS Chem. Neurosci.* 2, 658–666. doi: 10.1021/cn2000684

Hermans, A., Keithley, R. B., Kita, J. M., Sombers, L. A., and Wightman, R. M. (2008). Dopamine detection with fast-scan cyclic voltammetry used with analog background subtraction. *Anal. Chem.* 80, 4040–4048. doi: 10.1021/ac800108j

ISO10993-1:2018 (2018). Biological Evaluation of Medical Devices - Part 1: Evaluation and Testing within a Risk Management Process. Geneva: International Organization for Standardization.

ISO10993-6:2016 (2016). *Biological Evaluation of Medical Devices. Part 6: Tests for Local Effects After Implantation*. Geneva: International Organization for Standardization.

ISO11737-1:2018 (2018). *Sterilization of Health Care Products*. Geneva: International Organization for Standardization.

ISO14971:2019 (2019). *Medical Devices - Application of Risk Management to Medical Devices*. Geneva: International Organization for Standardization.

Jackowska, K., and Krysinski, P. (2013). New trends in the electrochemical sensing of dopamine. *Anal. Bioanal. Chem.* 405, 3753–3771. doi: 10.1007/s00216-012-6578-6572

Jaquins-Gerstl, A., and Michael, A. C. (2015). A review of the effects of FSCV and microdialysis measurements on dopamine release in the surrounding tissue. *Analyst* 140, 3696–3708. doi: 10.1039/c4an02065k

Johnson, J. A., Gray, J. H., Rodeberg, N. T., and Wightman, R. M. (2017). Multivariate curve resolution for signal isolation from fast-scan cyclic voltammetric data. *Anal. Chem.* 89, 10547–10555. doi: 10.1021/acs.analchem.7b02771

Johnson, J. A., Rodeberg, N. T., and Wightman, R. M. (2016). Failure of standard training sets in the analysis of fast-scan cyclic voltammetry data. *ACS Chem. Neurosci.* 7, 349–359. doi: 10.1021/acschemneuro.5b00302

Kachoosangi, R. T., and Compton, R. G. (2007). A simple electroanalytical methodology for the simultaneous determination of dopamine, serotonin and ascorbic acid using an unmodified edge plane pyrolytic graphite electrode. *Anal. Bioanal. Chem.* 387, 2793–2800. doi: 10.1007/s00216-007-1129-y

Kishida, K. T., Saez, I., Lohrenz, T., Witcher, M. R., Laxton, A. W., Tatter, S. B., et al. (2016). Subsecond dopamine fluctuations in human striatum encode superposed error signals about actual and counterfactual reward. *Proc. Natl. Acad. Sci. U S A.* 113, 200–205.

Kishida, K. T., Sandberg, S. G., Lohrenz, T., Comair, Y. G., Sáez, I., Phillips, P. E., et al. (2011). Sub-second dopamine detection in human striatum. *PLoS One* 6:e23291. doi: 10.1371/journal.pone.0023291

Kolli, R., Kaivosoja, E., and Levon, K. (2015). Choice of reference electrode is critical for potentiometric whole cell-based sensor. *Electroanalysis* 27, 1636–1641. doi: 10.1002/elan.201400710

Kozai, T. D. Y., Jaquins-Gerstl, A. S., Vazquez, A. L., Michael, A. C., and Cui, X. T. (2015). Brain tissue responses to neural implants impact signal sensitivity and intervention strategies. *ACS Chem. Neurosci.* 6, 48–67. doi: 10.1021/cn500256e

Lee, K. H., Chang, S. Y., Jang, D. P., Kim, I., Goerss, S., Van Gompel, J., et al. (2011). Emerging techniques for elucidating mechanism of action of deep brain stimulation. *Conf. Proc. IEEE Eng. Med. Biol. Soc.* 2011, 677–680. doi: 10.1109/IEMBS.2011.6090152

Lee, K. H., Lujan, J. L., Trevathan, J. K., Ross, E. K., Bartoletta, J. J., Park, H. O., et al. (2017). WINCS harmoni: closed-loop dynamic neurochemical

control of therapeutic interventions. *Sci. Rep.* 7:46675. doi: 10.1038/srep46675

Lind, N. M., Moustgaard, A., Jelsing, J., Vajta, G., Cumming, P., and Hansen, A. K. (2007). The use of pigs in neuroscience: modeling brain disorders. *Neurosci. Biobehav. Rev.* 31, 728–751. doi: 10.1016/j.neubiorev.2007.02.003

Lohrenz, T., Kishida, K. T., and Montague, P. R. (2016). BOLD and its connection to dopamine release in human striatum: a cross-cohort comparison. *Philos. Trans. R. Soc. B: Biol. Sci.* 371:20150352.

Lucas, A. D., Merritt, K., Hitchins, V. M., Woods, T. O., McNamee, S. G., Lyle, D. B., et al. (2003). Residual ethylene oxide in medical devices and device material. *J. Biomed. Mater. Res. Part B: Appl. Biomater.* 66, 548–552.

Martin, H. B., Smith, B. A., Angus, J. C., Landau, U., and Anderson, A. B. (1998). Boron-Doped diamond films for electrochemical applications. *MRS Proc.* 555:217. doi: 10.1557/PROC-555-217

McCreery, D. B., Agnew, W. F., Yuen, T. G., and Bullara, L. (1990). Charge density and charge per phase as cofactors in neural injury induced by electrical stimulation. *IEEE Trans. Biomed. Eng.* 37, 996–1001. doi: 10.1109/10.102812

McCreery, D. B., Agnew, W. F., Yuen, T. G., and Bullara, L. A. (1988). Comparison of neural damage induced by electrical stimulation with faradaic and capacitor electrodes. *Ann. Biomed. Eng.* 16, 463–481. doi: 10.1007/bf02368010

McDermott, M. T., and McCreery, R. L. (1994). Scanning tunneling microscopy of ordered graphite and glassy carbon surfaces: electronic control of quinone adsorption. *Langmuir* 10, 4307–4314.

Mendes, G. C., Brandao, T. R., and Silva, C. L. (2007). Ethylene oxide sterilization of medical devices: a review. *Am. J. Infect. Control* 35, 574–581.

Meunier, C. J., Mitchell, E. C., Roberts, J. G., Toups, J. V., McCarty, G. S., and Sombers, L. A. (2018). Electrochemical selectivity achieved using a double voltammetric waveform and partial least squares regression: differentiating endogenous hydrogen peroxide fluctuations from shifts in pH. *Anal. Chem.* 90, 1767–1776. doi: 10.1021/acs.analchem.7b03717

Millar, J., Stamford, J. A., Kruk, Z. L., and Wightman, R. M. (1985). Electrochemical, pharmacological and electrophysiological evidence of rapid dopamine release and removal in the rat caudate nucleus following electrical stimulation of the median forebrain bundle. *Eur. J. Pharmacol.* 109, 341–348. doi: 10.1016/0014-2999(85)90394-90392

Min, H. K., Ross, E. K., Jo, H. J., Cho, S., Settell, M. L., Jeong, J. H., et al. (2016). Dopamine release in the nonhuman primate caudate and putamen depends upon site of stimulation in the subthalamic nucleus. *J. Neurosci.* 36, 6022–6029. doi: 10.1523/JNEUROSCI.0403-16.2016

Mitchell, E. C., Dunaway, L. E., McCarty, G. S., and Sombers, L. A. (2017). Spectroelectrochemical characterization of the dynamic carbon-fiber surface in response to electrochemical conditioning. *Langmuir* 33, 7838–7846.

Mohanaraj, S., Wonnenberg, P., Cohen, B., Zhao, H., Hartings, M. R., Zou, S., et al. (2019). Gold nanoparticle modified carbon fiber microelectrodes for enhanced neurochemical detection. *J. Vis. Exp.*

Montague, P. R., and Kishida, K. T. (2018). Computational underpinnings of neuromodulation in humans. *Cold. Spring Harb. Symp. Quant. Biol.* 83, 71–82. doi: 10.1101/sqb.2018.83.038166

Moquin, K. F., and Michael, A. C. (2009). Tonic autoinhibition contributes to the heterogeneity of evoked dopamine release in the rat striatum. *J. Neurochem.* 110, 1491–1501. doi: 10.1111/j.1471-4159.2009.06254.x

Moran, R. J., Kishida, K. T., Lohrenz, T., Saez, I., Laxton, A. W., Witcher, M. R., et al. (2018). The protective action encoding of serotonin transients in the human brain. *Neuropsychopharmacology* 43, 1425–1435. doi: 10.1038/npp.2017.304

Nair, P. D. (1995). Currently practised sterilization methods-some inadvertent consequences. *J. Biomater. Appl.* 10, 121–135.

Nakajima, A., Shimo, Y., Uka, T., and Hattori, N. (2017). Subthalamic nucleus and globus pallidus interna influence firing of tonically active neurons in the primate striatum through different mechanisms. *Eur. J. Neurosci.* 46, 2662–2673.

Nicolai, E. N., Trevathan, J. K., Ross, E. K., Lujan, J. L., Blaha, C. D., Bennet, K. E., et al. (2017). Detection of norepinephrine in whole blood via fast scan cyclic voltammetry. *IEEE Int. Symp. Med. Meas. Appl.* 2017, 111–116. doi: 10.1109/MeMeA.2017.7985859

Olivieri, A. C. (2008). Analytical advantages of multivariate data processing. one, two, three, infinity? *Anal. Chem.* 80, 5713–5720. doi: 10.1021/ac800692c

Park, J. H., Chung, S. J., Lee, C. S., and Jeon, S. R. (2011). Analysis of hemorrhagic risk factors during deep brain stimulation surgery for movement disorders: comparison of the circumferential paired and multiple electrode insertion methods. *Acta Neurochir (Wien)* 153, 1573–1578. doi: 10.1007/s00701-011-0997-992

Patel, A. N., Tan, S.-Y., Miller, T. S., Macpherson, J. V., and Unwin, P. R. (2013). Comparison and reappraisal of carbon electrodes for the voltammetric detection of dopamine. *Anal. Chem.* 85, 11755–11764.

Patel, P. R., Zhang, H., Robbins, M. T., Nofar, J. B., Marshall, S. P., Kobylarek, M. J., et al. (2016). Chronic in vivo stability assessment of carbon fiber microelectrode arrays. *J. Neural Eng.* 13:066002. doi: 10.1088/1741-2560/13/6/066002

Phillips, K. A., Bales, K. L., Capitanio, J. P., Conley, A., Czoty, P. W., 't Hart, B. A., et al. (2014). Why primate models matter. *Am. J. Primatol.* 76, 801–827. doi: 10.1002/ajp.22281

Pihel, K., Walker, Q. D., and Wightman, R. M. (1996). Overoxidized polypyrrole-coated carbon fiber microelectrodes for dopamine measurements with fast-scan cyclic voltammetry. *Anal. Chem.* 68, 2084–2089. doi: 10.1021/ac960153y

Puthongkham, P., and Venton, B. J. (2019). Nanodiamond coating improves the sensitivity and antifouling properties of carbon fiber microelectrodes. *ACS Sensors* 4, 2403–2411.

Rajendrachari, S. (2018). *Investigation of Electrochemical Pitting Corrosion by Linear Sweep Voltammetry: A Fast and Robust Approach. Voltammetry*. London: IntechOpen.

Ramsson, E. S., Cholger, D., Dionise, A., Poirier, N., Andrus, A., and Curtiss, R. (2015). Characterization of fast-scan cyclic voltammetric electrodes using paraffin as an effective sealant with in vitro and in vivo applications. *PLoS One* 10:e0141340. doi: 10.1371/journal.pone.0141340

Ries, M. D., Weaver, K., Rose, R. M., Gunther, J., Sauer, W., and Seals, N. (1996). Fatigue strength of polyethylene after sterilization by gamma irradiation or ethylene oxide. *Clin. Orthopaed. Related Res.* 333, 87–95.

Robblee, L. S., and Rose, T. L. (1990). "The electrochemistry of electrical stimulation," in *Proceedings of the 20th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, (Piscataway, NJ: IEEE), 1479–1480.

Roberts, J. G., and Sombers, L. A. (2018). Fast-Scan cyclic voltammetry: chemical sensing in the brain and beyond. *Anal. Chem.* 90, 490–504. doi: 10.1021/acs.analchem.7b04732

Rodeberg, N. T., Sandberg, S. G., Johnson, J. A., Phillips, P. E. M., and Wightman, R. M. (2017). Hitchhiker's guide to voltammetry: acute and chronic electrodes for in vivo fast-scan cyclic voltammetry. *ACS Chem. Neurosci.* 8, 221–234. doi: 10.1021/acschemneuro.6b00393

Ross, E. K., Kim, J. P., Settell, M. L., Han, S. R., Blaha, C. D., Min, H.-K., et al. (2016). Fornix deep brain stimulation circuit effect is dependent on major excitatory transmission via the nucleus accumbens. *Neuroimage* 128, 138–148. doi: 10.1016/j.neuroimage.2015.12.056

Sansur, C. A., Frysinger, R. C., Pouratian, N., Fu, K. M., Bittl, M., Oskouian, R. J., et al. (2007). Incidence of symptomatic hemorrhage after stereotactic electrode placement. *J. Neurosurg.* 107, 998–1003. doi: 10.3171/jns-07/11/0998

Schluter, E. W., Mitz, A. R., Cheer, J. F., and Averbeck, B. B. (2014). Real-time dopamine measurement in awake monkeys. *PLoS One* 9:e98692. doi: 10.1371/journal.pone.0098692

Schwerdt, H. N., Shimazu, H., Amemori, K.-I., Amemori, S., Tierney, P. L., Gibson, D. J., et al. (2017b). Long-term dopamine neurochemical monitoring in primates. *Proc. Natl. Acad. Sci. U S A.* 114, 13260–13265. doi: 10.1073/pnas.1713756114

Schwerdt, H. N., Kim, M. J., Amemori, S., Homma, D., Yoshida, T., Shimazu, H., et al. (2017a). Subcellular probes for neurochemical recording from multiple brain sites. *Lab. Chip* 17, 1104–1115. doi: 10.1039/c6lc01398h

Seaton, B. T., Hill, D. F., Cowen, S. L., and Heien, M. L. (2020). Mitigating the effects of electrode biofouling-induced impedance for improved long-term electrochemical measurements in vivo. *Anal. Chem.* 92, 6334–6340. doi: 10.1021/acs.analchem.9b05194

Settell, M. L., Testini, P., Cho, S., Lee, J. H., Blaha, C. D., Jo, H. J., et al. (2017). Functional circuitry effect of ventral tegmental area deep brain stimulation: imaging and neurochemical evidence of mesocortical and mesolimbic pathway modulation. *Front. Neurosci.* 11:104. doi: 10.3389/fnins.2017.00104

Shannon, R. V. (1992). A model of safe levels for electrical stimulation. *IEEE Trans. Biomed. Eng.* 39, 424–426. doi: 10.1109/10.126616

Shon, Y. M., Lee, K. H., Goerss, S. J., Kim, I. Y., Kimble, C., Van Gompel, J. J., et al. (2010). High frequency stimulation of the subthalamic nucleus evokes

striatal dopamine release in a large animal model of human DBS neurosurgery. *Neurosci. Lett.* 475, 136–140. doi: 10.1016/j.neulet.2010.03.060

Stamford, J. A., Kruk, Z. L., Millar, J., and Wightman, R. M. (1984). Striatal dopamine uptake in the rat: in vivo analysis by fast cyclic voltammetry. *Neurosci. Lett.* 51, 133–138. doi: 10.1016/0304-3940(84)90274-x

Suaud-Chagny, M. F., Chergui, K., Chouvet, G., and Gonon, F. (1992). Relationship between dopamine release in the rat nucleus accumbens and the discharge activity of dopaminergic neurons during local in vivo application of amino acids in the ventral tegmental area. *Neuroscience* 49, 63–72. doi: 10.1016/0306-4522(92)90076-E

Suzuki, A., Ivandini, T. A., Yoshimi, K., Fujishima, A., Oyama, G., Nakazato, T., et al. (2007). Fabrication, characterization, and application of boron-doped diamond microelectrodes for in vivo dopamine detection. *Anal. Chem.* 79, 8608–8615. doi: 10.1021/ac071519h

Taylor, I. M., Nesbitt, K. M., Walters, S. H., Varner, E. L., Shu, Z., Bartlow, K. M., et al. (2015). Kinetic diversity of dopamine transmission in the dorsal striatum. *J. Neurochem.* 133, 522–531. doi: 10.1111/jnc.13059

Trevathan, J. K., Yousefi, A., Park, H. O., Bartoletta, J. J., Ludwig, K. A., Lee, K. H., et al. (2017). Computational modeling of neurotransmitter release evoked by electrical stimulation: nonlinear approaches to predicting stimulation-evoked dopamine release. *ACS Chem. Neurosci.* 8, 394–410.

Van Gompel, J. J., Bower, M. R., Worrell, G. A., Stead, M., Chang, S.-Y., Goerss, S. J., et al. (2014). Increased cortical extracellular adenosine correlates with seizure termination. *Epilepsia* 55, 233–244. doi: 10.1111/epi.12511

Vreeland, R. F., Atcherley, C. W., Russell, W. S., Xie, J. Y., Lu, D., Laude, N. D., et al. (2015). Biocompatible PEDOT:Nafion composite electrode coatings for selective detection of neurotransmitters in vivo. *Anal. Chem.* 87, 2600–2607. doi: 10.1021/ac502165f

Wang, Y., and Michael, A. C. (2012). Microdialysis probes alter presynaptic regulation of dopamine terminals in rat striatum. *J. Neurosci. Methods* 208, 34–39. doi: 10.1016/j.jneumeth.2012.04.009

Wightman, R. M., Heien, M. L., Wassum, K. M., Sombers, L. A., Aragona, B. J., Khan, A. S., et al. (2007). Dopamine release is heterogeneous within microenvironments of the rat nucleus accumbens. *Eur. J. Neurosci.* 26, 2046–2054. doi: 10.1111/j.1460-9568.2007.05772.x

Woolley, A. J., Desai, H. A., and Otto, K. J. (2013). Chronic intracortical microelectrode arrays induce non-uniform, depth-related tissue responses. *J. Neural Eng.* 10:026007. doi: 10.1088/1741-2560/10/2/026007

Yang, C., Trikantzopoulos, E., Jacobs, C. B., and Venton, B. J. (2017). Evaluation of carbon nanotube fiber microelectrodes for neurotransmitter detection: correlation of electrochemical performance and surface properties. *Anal. Chim. Acta* 965, 1–8. doi: 10.1016/j.aca.2017.01.039

Yoshimi, K., Kumada, S., Weitemier, A., Jo, T., and Inoue, M. (2015). Reward-Induced phasic dopamine release in the monkey ventral striatum and putamen. *PLoS One* 10:e0130443. doi: 10.1371/journal.pone.0130443

Yoshimi, K., Naya, Y., Mitani, N., Kato, T., Inoue, M., Natori, S., et al. (2011). Phasic reward responses in the monkey striatum as detected by voltammetry with diamond microelectrodes. *Neurosci. Res.* 71, 49–62.

Zachek, M. K., Hermans, A., Wightman, R. M., and McCarty, G. S. (2008). Electrochemical dopamine detection: comparing gold and carbon fiber microelectrodes using background subtracted fast scan cyclic voltammetry. *J. Electroanal Chem. (Lausanne)* 614, 113–120. doi: 10.1016/j.jelechem.2007.11.007

Zachek, M. K., Park, J., Takmakov, P., Wightman, R. M., and McCarty, G. S. (2010). Microfabricated FSCV-compatible microelectrode array for real-time monitoring of heterogeneous dopamine release. *Analyst* 135, 1556–1563. doi: 10.1039/c0an00114g

Zachek, M. K., Takmakov, P., Moody, B., Wightman, R. M., and McCarty, G. S. (2009). Simultaneous decoupled detection of dopamine and oxygen using pyrolyzed carbon microarrays and fast-scan cyclic voltammetry. *Anal. Chem.* 81, 6258–6265. doi: 10.1021/ac900790m

Zhao, Y., Li, S.-H., Chu, J., Chen, Y.-P., Li, W.-W., Yu, H.-Q., et al. (2012). A nano-sized Au electrode fabricated using lithographic technology for electrochemical detection of dopamine. *Biosens. Bioelectron.* 35, 115–122. doi: 10.1016/j.bios.2012.02.030

Zheng, J., Clogston, J. D., Patri, A. K., Dobrovolskaia, M. A., and McNeil, S. E. (2011). Sterilization of silver nanoparticles using standard gamma irradiation procedure affects particle integrity and biocompatibility. *J. Nanomed. Nanotechnol.* 2011(Suppl. 5):001.

Conflict of Interest: We declare that the following investigators Kenneth T. Kishida, P. Read Montague, Charles D. Blaha, Michael Heien, Kendall H. Lee, Terry Lohrenz may present conflict of interest with the ideas presented in this review because of their past or current use of fast scan cyclic voltammetry in humans.

The 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.

frontiers in Bioengineering and Biotechnology

ORIGINAL RESEARCH
published: 27 January 2022
doi: 10.3389/fbioe.2021.802794

Morphology-Based Deep Learning Approach for Predicting Osteogenic Differentiation

Yiqing Lan[1,2,3], *Nannan Huang*[1,2,3], *Yiru Fu*[1,2,3], *Kehao Liu*[1,2,3], *He Zhang*[1,2,3], *Yuzhou Li*[1,2,3]* *and Sheng Yang*[1,2,3]*

[1]*Stomatological Hospital of Chongqing Medical University, Chongqing, China,* [2]*Chongqing Key Laboratory of Oral Diseases and Biomedical Sciences, Chongqing, China,* [3]*Chongqing Municipal Key Laboratory of Oral Biomedical Engineering of Higher Education, Chongqing, China*

Early, high-throughput, and accurate recognition of osteogenic differentiation of stem cells is urgently required in stem cell therapy, tissue engineering, and regenerative medicine. In this study, we established an automatic deep learning algorithm, i.e., osteogenic convolutional neural network (OCNN), to quantitatively measure the osteogenic differentiation of rat bone marrow mesenchymal stem cells (rBMSCs). rBMSCs stained with F-actin and DAPI during early differentiation (day 0, 1, 4, and 7) were captured using laser confocal scanning microscopy to train OCNN. As a result, OCNN successfully distinguished differentiated cells at a very early stage (24 h) with a high area under the curve (AUC) (0.94 ± 0.04) and correlated with conventional biochemical markers. Meanwhile, OCNN exhibited better prediction performance compared with the single morphological parameters and support vector machine. Furthermore, OCNN successfully predicted the dose-dependent effects of small-molecule osteogenic drugs and a cytokine. OCNN-based online learning models can further recognize the osteogenic differentiation of rBMSCs cultured on several material surfaces. Hence, this study initially demonstrated the foreground of OCNN in osteogenic drug and biomaterial screening for next-generation tissue engineering and stem cell research.

Keywords: deep learning, convolutional neural network, osteogenic differentiation, drug screening, online learning

OPEN ACCESS

Edited by:
Grzegorz Marcin Wójcik, Marie Curie-Sklodowska University, Poland

Reviewed by:
Debanjan Sarkar, University at Buffalo, United States
Zaozao Chen, Southeast University, China

***Correspondence:**
Yuzhou Li
501182@cqmu.edu.cn
Sheng Yang
ysdentist@hospital.cqmu.edu.cn

Specialty section:
This article was submitted to Biomaterials, a section of the journal Frontiers in Bioengineering and Biotechnology

Received: *27 October 2021*
Accepted: *30 December 2021*
Published: *27 January 2022*

Citation:
Lan Y, Huang N, Fu Y, Liu K, Zhang H, Li Y and Yang S (2022) Morphology-Based Deep Learning Approach for Predicting Osteogenic Differentiation. Front. Bioeng. Biotechnol. 9:802794. doi: 10.3389/fbioe.2021.802794

INTRODUCTION

BMSCs are the most frequently used subtype of stem cells with a vigorous proliferative and differential capacity, making them a promising tool in tissue engineering, biomedicine, biomaterials, and many other fields (Mauney et al., 2005; Guan et al., 2012; Chiu et al., 2014; Yang et al., 2017; Farokhi et al., 2018; Qi et al., 2020). Assessing the osteogenic differentiation of BMSCs is of great importance for these applications, but is challenging because of the time-consuming process and low temporal-spatial resolution of conventional methods. For example, polymerase chain reaction and western blot only assess the bulk expression level, while histochemical staining, such as Alkaline phosphatase staining (ALP) and Alizarin red staining (ARS), often requires 14 days, 28 days, or even longer to induce observable biochemical changes (Waisman et al., 2019), which hinders the high-throughput screening of small molecules, cytokines, and biomaterials. Hence, an accurate, early-stage, and single-cell resolution method is urgently required to assess the osteogenic differentiation of BMSCs for next-generation biomedical applications.

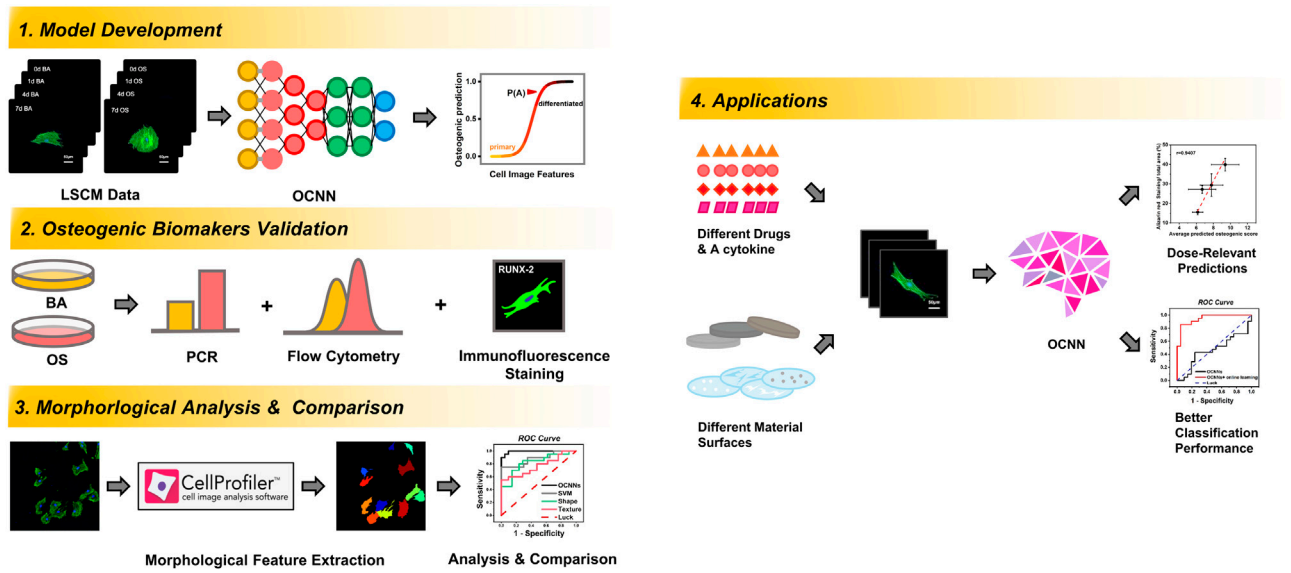

FIGURE 1 | Schematic flow chart of the training, validation, comparison, and application procedures and outcomes of OCNN in this study.

During osteogenic differentiation, BMSCs tend to change from a spindle-like shape to a polygon-like shape and are enlarged *in vitro* (Fan et al., 2012), while the arrangement or texture of the cytoskeleton also manifests distinct alterations (Treiser et al., 2010; Oei et al., 2019). On the one hand, the osteogenesis process is accompanied by augmented cell volume and programmed cytoskeleton remodeling. On the other hand, the direct modulation of cellular adhesive areas and cytoskeletal texture substantially influence osteogenic differentiation (McBeath et al., 2004; Engler et al., 2006; Zhang et al., 2015). For instance, our previous studies demonstrated that the regulation of the cytoskeleton by surface topography directly mediates cell differentiation and is associated with the activation of multiple adhesion and morphological proteins, such as FAK, RhoA, and YAP (Zhang et al., 2015; Zhang et al., 2016; Li et al., 2019). Therefore, the cellular morphology of BMSCs provides invaluable information for osteogenic differentiation prediction (Thomas et al., 2002; Marklein et al., 2016).

The cellular morphology data comprise a large number of high-dimensional image features and are challenging for prediction methods. Previously, machine learning models, including the Bayesian linear regression (Cutiongco et al., 2020) and support vector machine (SVM) (Chen et al., 2016; Yelin et al., 2019; Chen et al., 2021), were successfully applied to predict early-stage cell osteogenic differentiation based on the cytoskeletal morphology of biomaterials with different micro-environmental cues. Nonetheless, the data processing and parameter optimization of conventional machine learning require a high degree of specialized knowledge and considerable human efforts. Fortunately, novel deep learning methods can avoid these manual processes and achieve a high performance (Moen et al., 2019), which has been successfully applied to predict cellular senescence (Kusumoto et al., 2021), neural stem cell differentiation (Zhu et al., 2021), and screening drugs (X. Yang et al., 2019).

Herein, as illustrated in **Figure 1**, we established and trained a deep learning model, called osteogenic convolutional neural network (OCNN), on the high-content laser scanning confocal microscope (LSCM) images of the rBMSCs during the early stages of osteogenic differentiation. The predicted osteogenic scores (POS) obtained from OCNN were verified using traditional biomarkers. Subsequently, we compared the performance of OCNN with single-cell morphological parameters and support vector machine models. Lastly, we evaluated the performance of the OCNN and its modification on the osteogenic differentiation predictions of the rBMSCs cultured with different soluble drugs and on different biomaterial substrates. We hope that this study could preliminarily demonstrate the promising potential of deep learning in stem cell research, drug screening, and novel biomaterial development.

MATERIALS AND METHODS

Fabrication of Materials and Characterization of Surface Topography

Titanium plates with different nano-topographies were fabricated via sandblasting and acid etching, as previously described in our work (Li et al., 2019). Pure titanium plates (diameter of 14 mm and thickness of 1 mm, Chongqing University, Chongqing, China) were polished to 600 grit. Smooth surfaces were treated with 30 wt% HNO3 for 5 min. Micro surfaces were created by blasting with 100 l m aluminum oxide particles and incubated in 5 N HCl for 12 h and 30 wt% HNO3 for 5 min. Nano surfaces were manufactured *via* treatment with a 50/50 v/v% solution of 30% H_2O_2 and 2 N H_2SO_4 for 2 h. Titanium plates

were ultrasonically washed with dd water and sterilized with 75 wt% ethanol and ultraviolet light.

The substrates were coated with hyaluronic acid (HA), collagen I (Col-I), poly-dopamine (DOPA), and amyloid fibrils as previously described briefly. The glass coverslips were cleaned ultrasonically in ethanol and then rinsed with deionized water. After air-drying and UV irradiation for 30 min, 1 ml aqueous solution of HA (1 mg/ml), Col-I (0.1 mg/ml, PBS), DOPA (2 mg/ml, 10 mM Tris-HCl, pH 8.5), and lysozyme (1 mg/ml, 50 mM TCEP) were poured onto the substrate and allowed to react for 3 days, 3, 16, and 2 h, respectively. The topography of the prepared titanium plates and glass coverslips with different surface coatings was observed using high-resolution scanning electron microscopy (SEM, Hitachi S-4700).

Cell Culture

The rBMSCs were collected from the bone marrow of 4–6-week-old female rats as previously described (Ren et al., 2021) and identified by flow cytometry and differentiation phenotypes (**Supplementary Figure S1**). The cells were resuspended in complete medium and transferred to a Petri dish, cultured at 37°C with 5% CO_2. The medium was changed after 1 day and 2 days respectively. After achieving confluency of 80–90%, the rBMSCs were digested, counted, and seeded on substrates with an initial cell density of 5,000–10,000 cells/cm^2. In addition, they were cultured with Dulbecco's Modified Eagle's medium-low glucose containing 10% FBS and 1% penicillin/streptomycin and supplemented with PBS (basic medium, BA) or 10 nM dexamethasone, 10 mM β-glycerophosphate, and 50 μM ascorbic acid (osteogenic medium, OS). For the glass coverslips, the rBMSCs were cultured for 0, 1, 4, and 7 days. Whereas for the drug screening assays, they were cultured for 1 and 14 days, and for the biomaterial assays, they were cultured for 1 day. All animal operations were performed in accordance with the guidelines of the Animal Care and Use Committee of China and were approved by the ethics committee of Chongqing Medical University Affiliated Hospital of Stomatology (Ethic No. 2021033).

Data Acquisition and Pre-processing

The rBMSCs were washed with PBS three times, fixed with 4% formaldehyde in Dulbecco's phosphate-buffered saline (D-PBS), and washed repeatedly. Subsequently, they were permeabilized in 0.2% Triton-X100 (PBS) for 5 min and washed with PBS three times. Next, they were soaked in staining buffer containing 0.33 mM Alexa Fluor 488 phalloidin (Yisheng, China) and 10 mg/ml bovine serum albumin (PBS) for 1 h. After washing three times with PBS, nuclei were stained with 0.3 mM 40, 6-diamidino-2-phenylindole (DAPI, Beyotime, China) for 3 min and then rinsed with PBS. Samples were observed and imaged with LSCM at a magnification of ×200 (Leica TCS SP8, Germany), and each 2-D image was taken at the maximum projection of the z-stack and saved as a 1,024 × 1024-pixel RGB image in the Tiff format.

For model development, single-cell images were cropped from the original LSCM images (5–10 cells per image) to obtain 2,916 single-cell images, approximately 500–600 images in each group (0, 1, 4, and 7 days with BA/OS). Notably, cells adjunct to other cells were abandoned because of the potential influence of cell-to-cell contact (Chen et al., 2016). The fluorescence intensity was globally normalized to correct the batch effect from different biological repeats.

rBMSC Osteogenic Differentiation Assays

An mRNA was extracted by lysing the cells with TRIzol (Takara, Japan) and incubated with chloroform for 10 min, followed by centrifugation at 12,000 rpm for 15 min at 4°C. Then, it was purified using ethanol, resuspended in DEPC water (Biosharp, Japan), and quantified using a Nanodrop spectrophotometer (Thermo Scientific, Waltham, United States). Next, it was reverse transcribed to cDNA by using the RNAiso Plus reagent kit (Takara) and amplified using the ProFlex PCR system (Thermo Scientific). Finally, the genes expression levels were quantified using the Power SYBR Green PCR master mix (Takara) in a real-time PCR machine (Applied Biosystems 7500, Life Technologies, Waltham, United States). The primer sequences used in this study are listed in **Supplementary Table S1**.

For flow cytometry, the rBMSCs were digested, diluted with PBS, and stained with FITC Mouse Anti-Rat CD90 (Biolegend, San Diego, United States) and FITC Mouse Anti-Rat CD44 (Biolegend).

For immunofluorescence staining, they were fixed with 4% formaldehyde, permeabilized with 0.2% Triton-X100, and washed repeatedly. Then, they were blocked using 3% donkey serum for 1 h and incubated overnight with a Runx-2 primary antibody (1:200, Abcam, United Kingdom). After washing with PBS three times, cells were incubated with the fluorescent secondary antibody (1:350, Alexa 647, Jackson ImmunoResearch, West Grove, United States) for 1 h. Finally, the cytoskeleton and nuclei were stained according to the above-mentioned procedure.

For drug screening, the rBMSCs were treated with an OS medium accompanied by incremental concentrations of alendronate sodium (1, 5, or 10 μM), simvastatin (0.1 μg/ml, 0.5 μg/ml or 1 μg/ml), 1α, 25-dihydroxyvitamin D3 (1, 10 or 100 nM) as well as BMP-2 (50 ng/ml, 100 ng/ml and 200 ng/ml). After cultivation for 14 days, the cells were fixed with 4% paraformaldehyde for 10 min and washed with PBS. Then, they were stained with 1% Alizarin Red S (Solarbio) for 10 min and washed twice with PBS. The stained cells were captured using a stereomicroscope (Zeiss SteREO Discovery.V12, Germany) and quantified using ImageJ (NIH Image, Bethesda, MD).

Development of Convolutional Neural Networks

To develop OCNN, three classic deep learning models with pre-trained weights (ImageNet), including VGG 16 (Simonyan and Zisserman, 2014), Inception V3 (Szegedy et al., 2016), and ResNet50 (He et al., 2016), were screened to select the convolutional core. The transfer learning method was used to extract the rBMSC features, and a binary classifier with softmax

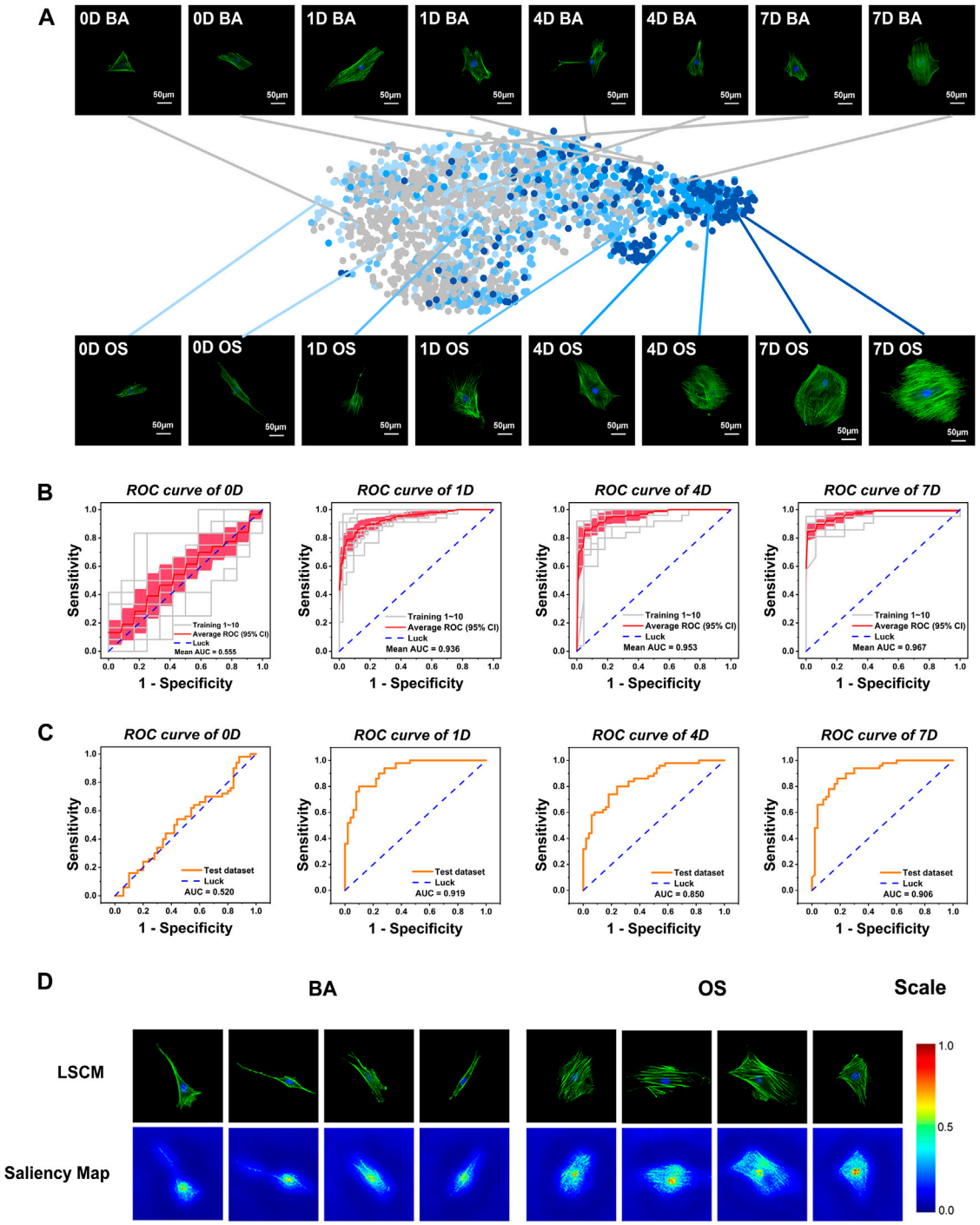

FIGURE 2 | Development and validation of a conventional, static, and baseline model (OCNN) for sensitive high-throughput and automatic osteogenesis prediction on flat coverslips. **(A)** Visualization of diverse single-cell datasets. The styles from the network for all images in the cell dataset were embedded using t-SNE. Each point represents a different LSCM image. Grey: basic (BA) group; light to deep blue: osteogenic (OS) group on days 0,1,4, and 7. Each photo: green, F-actin; blue, nuclei. Scale bar, 50 μm. **(B)** 10-fold cross-validated ROC curves of Inception V3 at four-time points (validation dataset). The gray line represents each independent validation; the red line represents the average ROC; the light red area represents the 95% confidence interval. **(C)** A randomly selected new test dataset was used to evaluate the classification performance of OCNN (50 images of BA/OS group each). ROC curves at four-time points. The orange line represents the independent validation. **(D)** The saliency map showed key identification regions for the prediction of BA or OS cells.

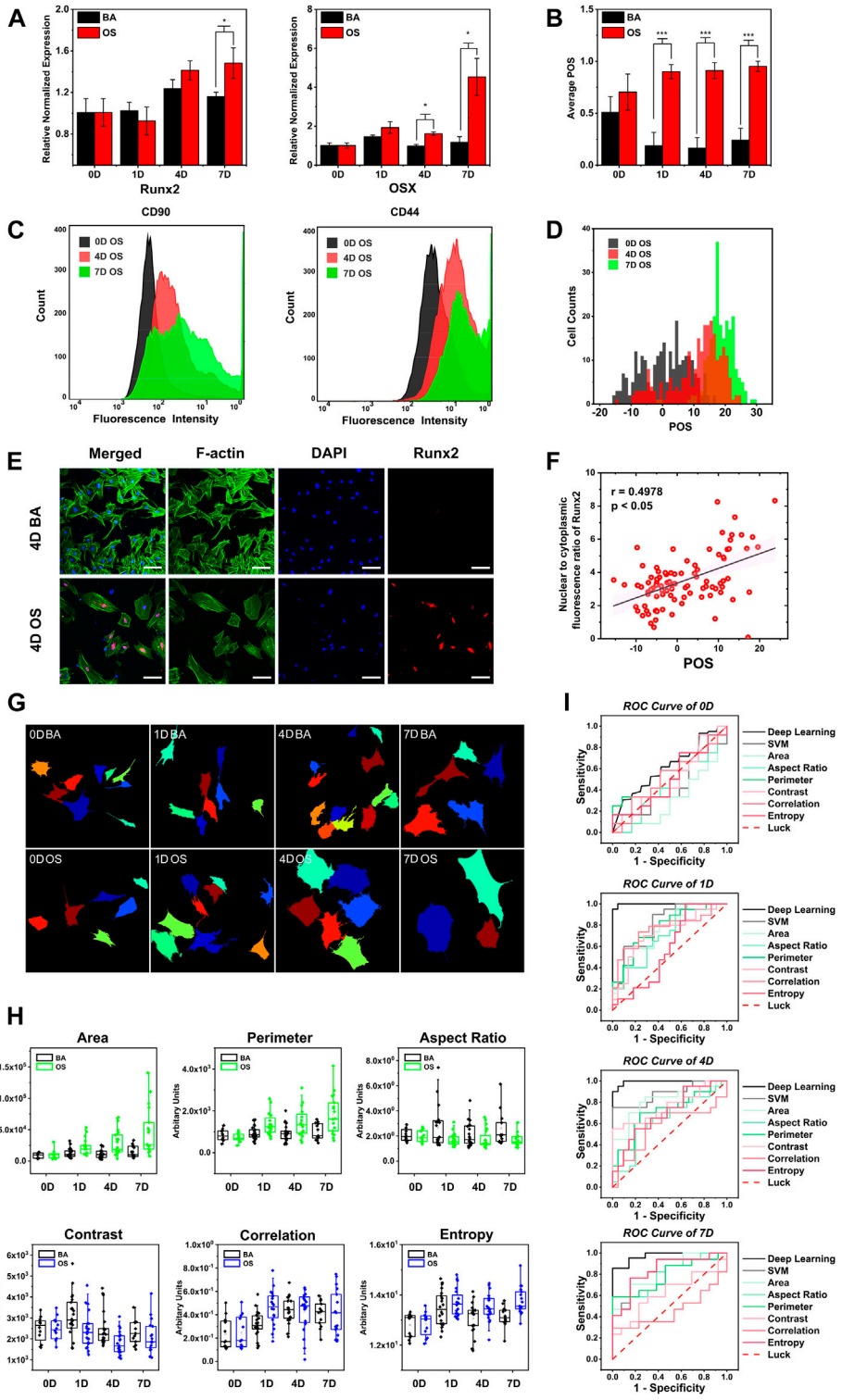

FIGURE 3 | OCNN highly correlated with conventional biochemical markers and performed better than single morphological parameter and support vector machine in osteogenic differentiation prediction. **(A)** Real-time PCR gene expression levels of *Osx*, *Runx2* relative to *GAPDH* in rBMSCs cultured in BA/OS for 0, 1, 4, and 7 days **(B)** BA/OS Images of 0, 1, 4, and 7 days were classified and scored according to OCNN. **(C)** Flow cytometry determination of the MSC-specific surface markers (CD44, CD90) in OS groups at different induction times (0, 4, 7 days). **(D)** Images in the OS group of 0, 4, and 7 days were classified and scored according to OCNN. **(E)** rBMSCs' immunofluorescence staining of Runx2 protein cultured in both types of the medium on day 4. Green: F-actin; Blue: Nuclei; Red: Runx2 protein. Scale bar: 100 μm. **(F)** Using the Nuclear/cytoplasm intensity ratio of Runx2 protein to define the extent of single-cell osteogenic differentiation and compared with OCNN

(*Continued*)

FIGURE 3 | predicting score on day 4. Data are shown as mean. *p* values by two-sided Student's t-test. *: $p < 0.05$; ***: $p < 0.001$. **(G)** Schematic diagram of Cellprofiler software for cell localization and cell morphology capture. **(H)** Cell morphology was measured on the cell images of the validation dataset of 0, 1, 4, and 7 days using CellProfiler software (United States), and six representative morphological parameters were selected: area, perimeter, aspect ratio (shape); contrast, correlation, and entropy (texture). Box plots were used to observe the differences between the BA/OS groups on the six parameters. **(I)** ROC curves of deep learning, support vector machines (SVM), and six parameters, comparing the classification performance of four-time points.

activation was added to the top of the models for predictions. The training process was conducted on a computer with an Intel Core i7-9700F, 32 GB RAM, and an NVIDIA GeForce RTX 2080Ti and implemented with Keras v.2.2.4 (http://github.com/fchollet/keras). Detailed information about the training of convolutional neural networks can be found in **Supplementary Table S2**.

During training, single-cell images were randomly distributed into training and validation sets at a ratio of 9:1. Conventional data augmentation was performed to reduce overfitting (**Supplementary Figure S2**). The Adam optimizer was adopted, and hyper-parameters (dropout ratio, learning rate, and batch size) were optimized to improve the performance. Ten-fold cross-validation was used to evaluate the prediction performance of the three models. For online learning, OCNN was further supplemented with small training samples (50–100) from different biomaterials.

To quantitatively measure the osteogenic differentiation in a linear space, predicted osteogenic score (POS) was proposed by a logit transformation of the final model output *p* for data scaling, similarly to a previously described method to quantify the cell senescence (Kusumoto et al., 2021):

$$POS = ln\frac{p}{1-p}$$

Single-Cell Morphological Parameters and Support Vector Machines

Cell morphology was measured on single-cell images using an open-source software Cellprofiler (the Broad Institute of Harvard and MIT, United States) (Lamprecht et al., 2007), yielding 25 single-cell morphological parameters. Six representative morphological parameters were selected by correlation analysis, including three shape parameters (area, perimeter, aspect ratio) and three texture parameters (contrast, correlation, and entropy).

Support vector machines were constructed using the Sci-kit package in Python. Feature dimensions were reduced to three from the morphological parameters using the linear kernel support vector machine technique. Hyperparameter optimization was conducted using the wrapping algorithm.

Evaluation of Model Performance

To evaluate the model performance, the true positive (TP), true negative (TN), false positive (FP), and false negative (FN) values were counted. Then, six measurements, including accuracy, sensitivity, specificity, precision, recall, and F1-score, were calculated as follows:

$$Sensitivity = \frac{TP}{TP+FN}$$

$$Specificity = \frac{TN}{TN+FP}$$

$$Accuracy = \frac{TP+TN}{TP+FP+TN+FN}$$

$$Precision = \frac{TP}{TP+FP}$$

$$F1-score = \frac{2*Precision*Recall}{Precision+Recall}$$

The ROC curve was plotted based on the sensitivity and 1—Specificity scores, and the AUC value was computed.

Statistical Analysis

The statistical charts were created using Origin 2018 (OriginLab, Northampton, United States), and the data were presented as mean ± standard deviations. For comparison between two groups, the unpaired Student's t-test was used, and a value of <0.05 was considered statistically significant. For correlation analysis, a value of <0.05 was considered statistically significant.

RESULTS

Morphological Characteristics of rBMSCs Under LSCM

To investigate the morphological changes of the rBMSCs during the early stages of osteogenic differentiation, we collected the rBMSCs cultured in osteogenic (OS) and basal (BA) mediums and then took LSCM images on days 0, 1, 4, and 7. As shown in **Figure 2A**, *via* the naked eye observation, the cellular shape changed from spindle-like to more extensive, and the cytoskeleton arrangement seemed to become more complex and crossed after induction for 4 days. However, by depositing 25 objective morphological parameters into two-dimensional t-SNE (Van der Maaten et al., 2008), we found a clear left-to-right shift of group OS starting from day 1, while group BA exhibited a more randomized distribution, which partially overlapped with the OS group.

To recognize the distinct but overlapping underlying pattern of the cellular morphology, we next developed OCNN models via the transfer learning of single-cell images using three classical deep learning models, VGG-16, Inception V3, and ResNet-50, as shown in **Supplementary Figure S3** and **Supplementary Table S3**. Based on the general performance, we selected the pre-trained Inception V3 as the convolutional core of the OCNN to perform our follow-up studies. As shown in **Figure 2B**, the OCNN showed average AUCs of 0.936, 0.953, and 0.967 on days 1, 4, and 7, respectively, in 10-fold internal cross-validation. To validate the model generalization ability, OCNN was further tested on an independent, external dataset from biological repeats at

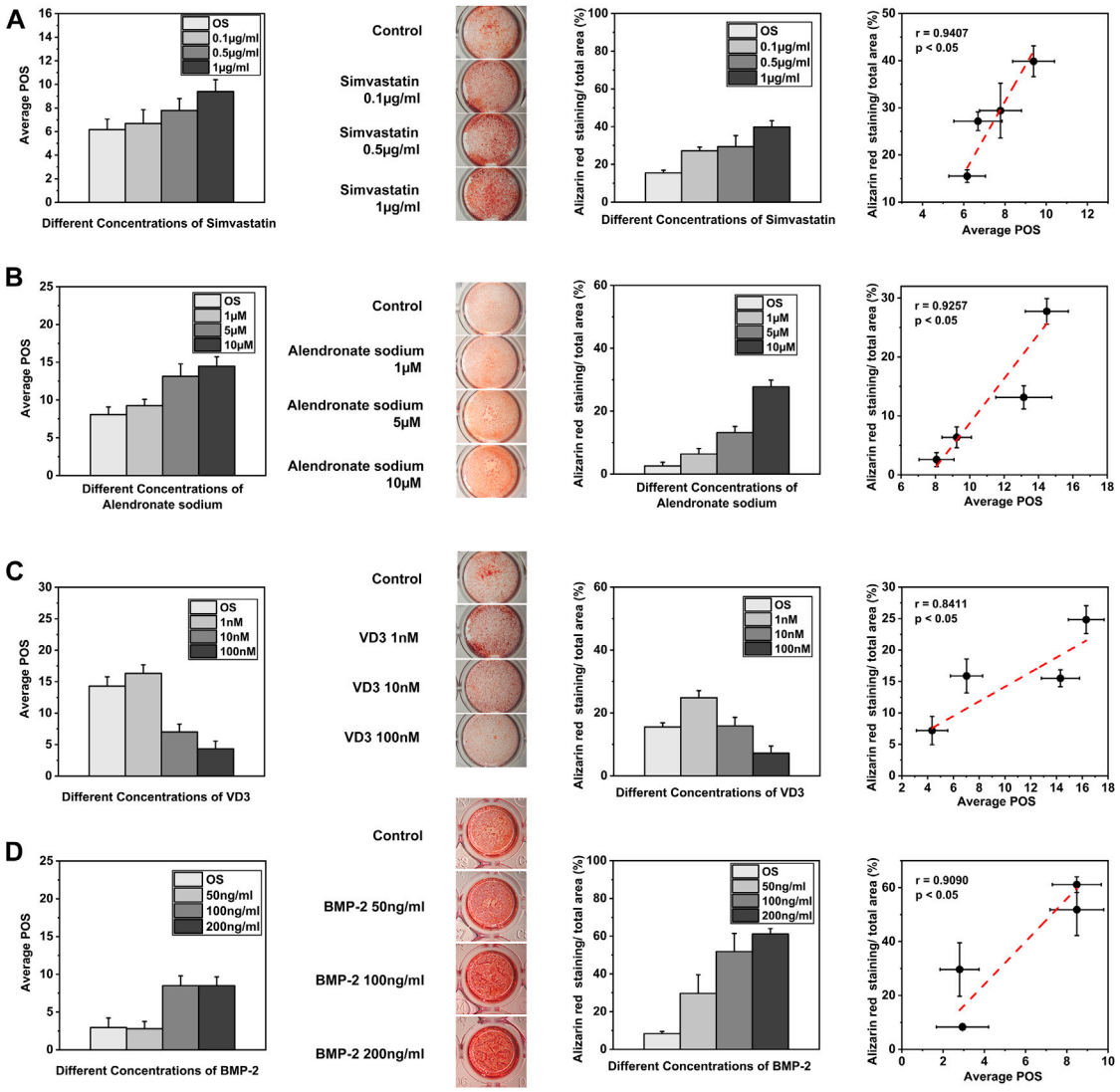

FIGURE 4 | OCNN for drug screening. **(A)** The predicted osteogenic score (POS) of the rBMSCs cultured with different concentrations of simvastatin on day 1 and Alizarin red staining on day 14 was quantified by ImageJ. Linear fits were performed using the POS and Alizarin red staining areas to calculate the correlation, $p < 0.05$. **(B)** The predicted osteogenic score (POS) of the rBMSCs cultured with different concentrations of alendronate sodium on day 1 and Alizarin red staining on day 14 was quantified by ImageJ. Linear fits were performed using the POS and Alizarin red staining areas to calculate the correlation, $p < 0.05$. **(C)** The predicted osteogenic score (POS) of the rBMSCs cultured with different concentrations of VD3 on day 1 and Alizarin red staining on day 14 was quantified by ImageJ. Linear fits were performed using the POS and Alizarin red staining areas to calculate the correlation, $p < 0.05$. **(D)** The predicted osteogenic score (POS) of the rBMSCs cultured with different concentrations of BMP-2 on day 1 and Alizarin red staining on day 14 was quantified by ImageJ. Linear fits were performed using the POS and Alizarin red staining areas to calculate the correlation, $p < 0.05$.

four-time points, which also showed satisfactory AUCs of 0.919, 0.850, and 0.906 on days 1, 4, and 7, respectively (**Figure 2C**). The AUCs on day 0 were 0.555 (internal validation) and 0.520 (external validation), which is expected, since day 0 implies that cells had not received any treatment.

To better understand the morphological differences recognized by OCNN, we visualized the important regions that are relevant to the predictions via a Saliency Map (Simonyan, Vedaldi, et al., 2014), which reflected the activation of specific pixels upon specific predictions. Interestingly, as shown in **Figure 2D**, the nucleus and peri-nucleus cytoskeleton were mostly activated for the BA predictions, while more cytoplasmic cytoskeletons were activated in the OS prediction, probably reflecting the importance of the cytoskeleton in stemness or osteogenic differentiation. Nuclear morphology is an important indicator of cell function and is correlated with osteogenic differentiation via lamin A/C under external forces, nano-topography, and chemical coatings (Werner et al., 2017). On the other hand, the strong stress fibers of the cytoskeleton drew considerable concern in the OS group, as revealed in previous findings (Engler et al., 2006) that the intensity and arrangement of F-actin and

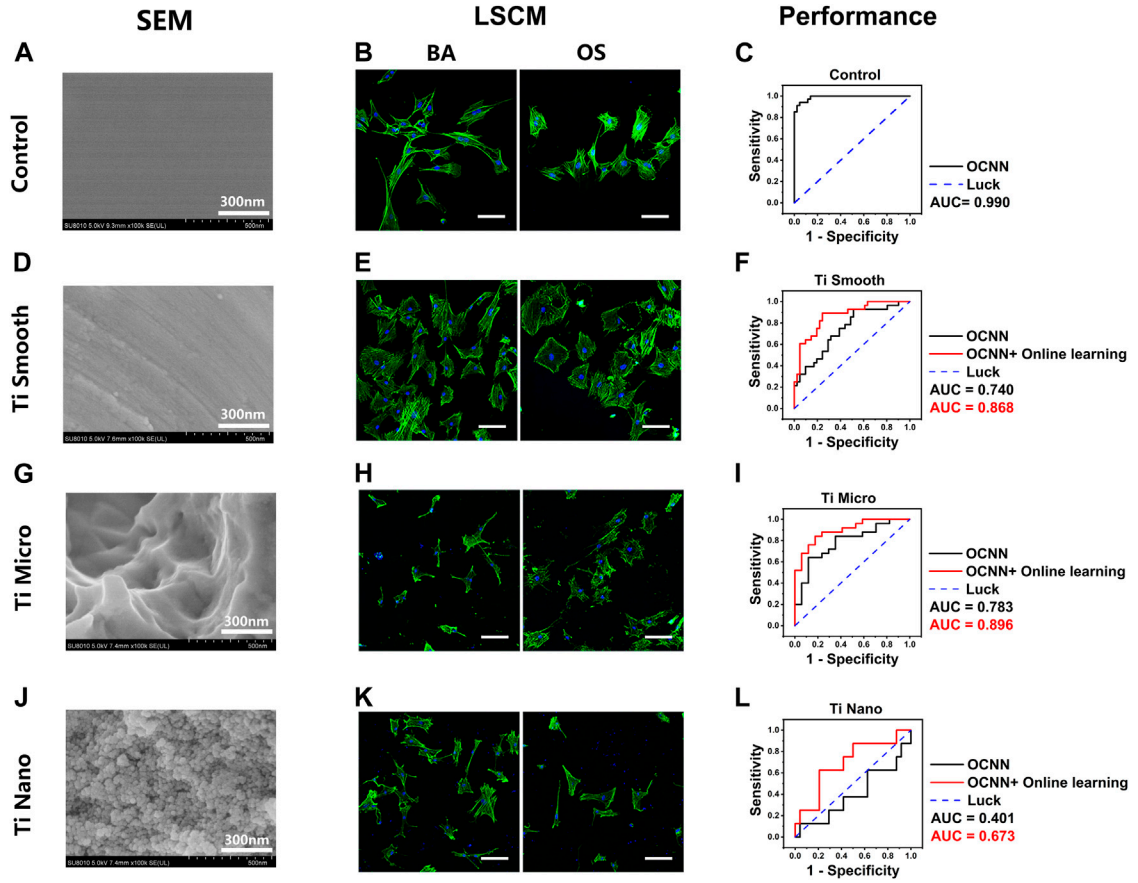

FIGURE 5 | OCNN prediction and OCNN-based online learning for rBMSCs on titanium surfaces. **(A) (D) (G) (J)** SEM images of control and three different titanium surfaces: smooth, micro, and nano, scale bar = 300 nm. **(B) (E) (H) (K)** LSCM images of the rBMSCs on control and three different titanium surfaces: smooth, micro, and nano. Basal medium (BA), osteogenic supplement medium (OS). Induction time: 24 h; green: F-actin; blue: nuclei. Scale bar, 100 μm. **(C) (F) (I) (L)** Comparison of OCNN prediction and OCNN-based online learning results using ROC curves. Black line: OCNN prediction; red line: OCNN accompanied with online learning.

non-muscle myosin-II plays an important role in MSC osteogenic differentiation. Considered together, these data showed the plausibility of the binary prediction models based on deep learning trained via nucleus and cytoskeleton morphology images.

It was worth noting that a small proportion of the images were incorrectly classified when using the OCNN, indicating that there may be some degree of differences in the level of single-cell differentiation within the stem cell population under nonosteogenic and osteogenic induction conditions, which cannot be classified via deep learning (**Supplementary Figure S4**).

Biochemical Changes and Morphology-Based Predictions During Osteogenic Differentiation Compared with OCNN

Next, we analyzed the consistency between OCNN predictions and conventional biochemical measurements, including qRT-PCR, flow cytometry, and immunofluorescence staining. To compare in a one-dimensional linear space, we came up with the concept of the predicted osteogenic score (POS), a logit transformation value from the OCNN output for each single-cell image, as a modified method from previous studies.

As illustrated in **Figure 3A**, we first compared the POS with the mRNA expression level of two osteogenic markers, osterix (*Osx*) and runt-related transcription factor 2 (*Runx2*), on days 0, 1, 4, and 7. From the result, there were no significant differences in both *Runx2* and *Osx* gene expression from the BA/OS groups on days 0 and 1. The Osx gene expression in the OS group was upgraded on day 4 ($n = 3$, $p < 0.05$) and day 7 ($n = 3$, $p < 0.05$), and the Runx2 gene expression in the OS group was upgraded on day 7 ($n = 3$, $p < 0.05$). In comparison, the POS was significantly upregulated from day 1 ($n = 50$ cells from three biological repeats, $p < 0.001$), indicative of the early sensitivity of OCNN (**Figure 3B**).

Second, to scrutinize the POS at the single-cell level, we compared it with the flow cytometric analysis of the cell surface markers in the OS group on days 0, 4, and 7. On day 0, the rBMSCs had a high expression of mesenchymal stem cell marker CD90, which decreased with time; however,

there were still partial overlaps with the results of day 0 (Dhaliwal et al., 2016) (**Figure 3C**). This decreasing trend was more evident for CD44 than CD90. Similarly, the POS also showed a right-shift tendency, similar to flow cytometry (**Figure 3D**).

Then, to visually inspect the POS, we compared it with the immunofluorescence staining of Runx2 at the protein level on day 4 (**Figure 3E**), since the spatial distribution of this nuclear transcription factor could reflect the osteogenic differentiation (Z. Chen et al., 2019). Compared with group BA, the red staining of Runx2 in the OS group was more concentrated in the nucleus, which is consistent with previous findings. Subsequently, we performed a correlation analysis of the Runx2 nuclear/cytoplasm ratio and POS (**Figure 3F**) and found a significant correlation ($p < 0.05$) with a moderate relationship (r = 0.4978). Based on these results, the POS is consistent with conventional biochemical markers; thus, laying the biological basis for further applications.

Single-cell morphological parameters have been widely suggested to be associated with cell phenotypes (Bakal et al., 2007; Prasad and Alizadeh, 2019; Wu et al., 2020), and some machine learning methods based on these features have been recently developed to predict osteogenesis. We extracted 25 morphological parameters (**Figure 3G**; **Supplementary Figure S5**) and selected six typical features that were significantly correlated with the cell phenotypes (**Figure 3H**; **Supplementary Figure S6**), suggesting the upregulation of the cell area, perimeter, cytoskeleton correlation, and entropy, as well as the aspect ratio and cytoskeleton contrast during osteogenesis. Support vector machine models were then developed based on these parameters.

We then compared the prediction performance between single-cell morphological parameters, support vector machine, and OCNN in a biologically independent dataset ($n = 20$) (**Figure 3I**). As expected, no model could distinguish cells in the two groups on day 0. Impressively, OCNN achieved a fantastic AUC of 0.998 on day 1, which was higher than that of the support vector machine (AUC = 0.807) and single-cell morphological parameters (AUCs ranged from 0.579 to 0.775). The advantage of the OCNN was maintained at day 4 (AUC = 0.993) and day 7 (AUC = 0.972) compared to those of other methods. It is worth noting that three highly correlated morphological parameters during every support vector machine training process were displayed: area, perimeter, and contrast on day 0; area, aspect ratio, and perimeter on day 1 and day 4; and area, perimeter, and entropy on day 7. From the results, it can be concluded that the area and perimeter are significant indicators for distinguishing cellular morphology. In addition, these results further support the idea of using deep-learning-based models in cell phenotype analysis.

Screening of Osteogenic Small Molecule Drugs

We examined the application potential of OCNN in predicting the osteoinduction ability of small molecule drugs or cytokines at an early stage (**Figure 4**). First, we examined the effects of simvastatin (**Figure 4A**), whose osteogenic induction ability was reported by enhancing the Rho/actin/cell rigidity pathway as well as increasing the actin filament organization and cell rigidity (Tai et al., 2015). We calculated the POS from the rBMSCs supplied with 0.1, 0.5, and 1 μg/ml simvastatin on day 1 ($n \geq 50$), and alizarin red staining was carried out on day 14. The Pearson analysis revealed a significant correlation ($p < 0.05$) with a very strong relationship (r = 0.9407) between the OCNN prediction at the early stage (day 1) and the final osteogenesis *in vitro*.

Second, we examined the dose-dependent osteogenic induction ability of the two small molecules via other biochemical mechanisms (**Figures 4B,C**). Alendronate sodium affected the osteogenic differentiation by activating ERK and JNK (Fu et al., 2008), while 1α, 25-dihydroxyvitamin D3 (VD3) activated the nuclear vitamin D receptor (VDR) and promoted the osteogenic differentiation (Lou et al., 2017) (He et al., 2020). The Pearson correlation coefficient was 0.9257 ($p < 0.05$) for alendronate sodium and 0.8411 ($p < 0.05$) for 1α, 25-dihydroxyvitamin D3. These results suggested that the OCNN was able to distinguish osteogenic phenotypes modulated by different biochemical signals.

Lastly, we examined the predictive ability of the OCNN under the influence of a classic osteogenic growth factor, i.e., bone morphogenetic protein 2 (BMP-2, **Figure 4D**). Bone morphogenetic proteins (BMPs) are effective regulators of osteoblast proliferation and differentiation, and among them, BMP-2 has been the most studied cell growth factor in the bone tissue regeneration field (Salazar et al., 2016). The Pearson analysis revealed a significant correlation ($p < 0.05$) with a strong relationship (r = 0.9090) between the OCNN prediction and BMP-2 induced osteogenesis. Collectively, these results demonstrate the feasibility and reliability of OCNN for drug screening.

OCNN Prediction and OCNN-Based Online Learning for Cells on Titanium Surfaces and Chemical Coatings

The substrate characteristics of biomaterials are of great importance and have a significant impact on cell morphology. Therefore, we hypothesized that the baseline OCNN may not be suitable for predicting osteogenic differentiation in this scenario. This dilemma may be tackled by the idea of online learning, which implies supplying small additional samples to the baseline model, due to the migratory nature of deep learning (Lobo et al., 2018).

To examine this hypothesis, we fabricated titanium surfaces with different nanotopographies: smooth, micro, and nano, as described in our previous work. Glass coverslips were chosen as the control and were flat with no extra features under SEM (**Figure 5A**). On day 1, the cell morphology in the OS group exhibited no visible changes to the naked eye (**Figure 5B**). However, the baseline OCNN model still captured certain underlying patterns; thus, achieving a satisfactory AUC of 0.990 (**Figure 5C**).

For titanium surfaces, the smooth group (**Figure 5D**) is flat with no obvious ridges or nanoscale features, while the micro

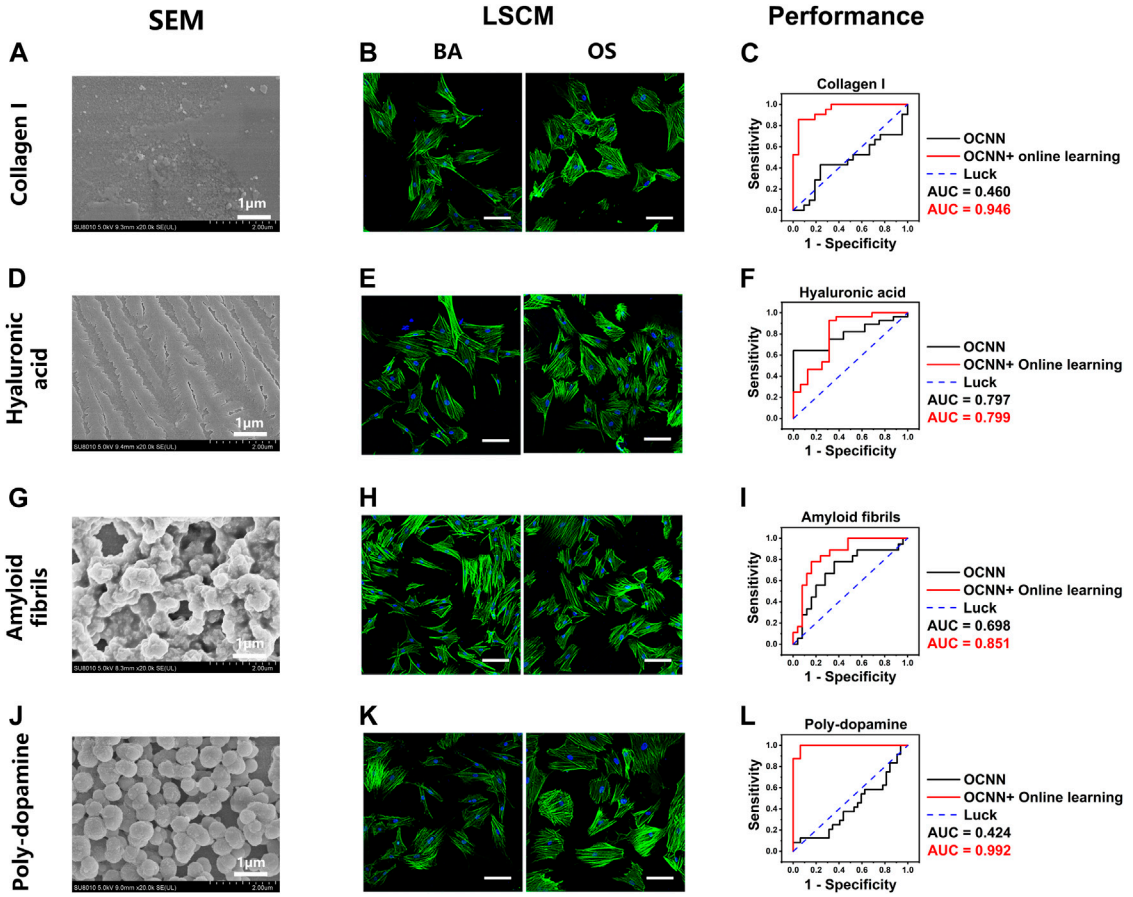

FIGURE 6 | OCNN prediction and OCNN-based online learning for rBMSCs on chemical coatings. **(A) (D) (G) (J)** SEM images of four different chemical coatings: collagen I, hyaluronic acid, amyloid fibrils, and poly-dopamine, scale bar = 1 μm. **(B) (E) (H) (K)** LSCM images of the rBMSCs on four different chemical coatings: collagen I, hyaluronic acid, amyloid fibrils, and poly-dopamine. Basal medium (BA), osteogenic supplement medium (OS). Induction time: 24 h; green: F-actin; blue: nuclei. Scale bar, 100 μm. **(C) (F) (I) (L)** Comparison of the OCNN prediction and OCNN-based online learning results using ROC curves. Black line: OCNN prediction and red line: OCNN accompanied with online learning.

group (**Figure 5G**) showed numerous ridges and grooves, and the nano group (**Figure 5J**) showed dense nano-features. On day 1, the cell morphology on these substrates showed extreme inconsistency and diversity; thus, no patterns in group BA/OS could be observed with the naked eyes (**Figures 5E,H,K**). Interestingly, the baseline OCNN also performed poorly in predictions on titanium surfaces, with AUCs of 0.740 (smooth, **Figure 5F**), 0.783 (micro, **Figure 5I**), and 0.401 (nano, **Figure 5L**). Fortunately, after supplying small data (n = 50–100 and 1/16–1/8 of baseline training data) to the baseline OCNN, the AUC increased to 0.868 (smooth), 0.896 (micro), and 0.673 (nano).

Next, we examined the performance of the baseline and online OCNNs on several common chemical coatings, including collagen I (Col-I), hyaluronic acid (HA), amyloid fibrils, and poly-dopamine (DOPA). On Col-I (**Figure 6A**) and HA (**Figure 6D**) coatings, no micro or nanostructures were observed under SEM, similar to the glass coverslips. Under LSCM, the cell morphology in group BA/OS could not be distinguished by the naked eye at day 1 (**Figures 6B,E**). However, unlike on the glass coverslips, the baseline OCNN could not predict the phenotypes associated with the OS on Col-I (AUC = 0.460) but can moderately predict on HA (AUC = 0.797), suggesting that the chemical components of substrates have an essential influence on cell phenotypes (Arora et al., 2020). After supplying online data, the AUC increased to 0.946 (Col-I, **Figure 6C**) and 0.799 (HA, **Figure 6F**). This result demonstrates the ability of online OCNN to distinguish cells cultured in microenvironments with different chemical cues.

Amyloid fibrils and DOPA coatings exhibited microspheres with diverse shapes and continuity (**Figures 6G,J**); thus, representing a mixture of different nanotopographies and chemical components. Under LSCM, the cell morphology in group BA/OS could not be distinguished with the naked eyes on day 1 (**Figures 6H,K**). The predictions of the baseline OCNN were acceptable in the amyloid fibrils (AUC = 0.698, **Figure 6I**) but performed poorly in DOPA (AUC = 0.424, **Figure 6L**). These results further suggest that the material properties of the

substrates, including the nanotopography and chemical components, can substantially reform the cell morphology, which may explain the controversial association between certain cell phenotypes and final cell fate (Arora et al., 2020). After supplying online data, the AUC increased drastically to 0.851 (amyloid fibrils) and 0.992 (DOPA).

DISCUSSION

In this paper, we show that the CNNs can be trained with images taken by a laser confocal microscope and then classify images with slight morphological variations correctly. We applied three classical deep learning CNNs VGG-16, Inception V3, and ResNet-50 for transfer learning and achieved accuracy higher than 80% on the validation set except for day 0. Then the selected OCNN models from above also achieved excellent results on the independent test set. We believe that several conditions allowed us to achieve such high accuracy with the trained neural networks. First, we set the cell seeding at a suitable density for the study. Second, the starting size of the single-cell images is a uniform 640*640 pixels, which ensured the size and uniformity of the original input data. Third, by certain image pre-processing methods, for example, rotation, mirror flip, and scale, the number of images provided to CNN was increased and the training effect gained better than that of images without image processing.

We validated the prediction results of OCNN by conventional biochemical markers to verify the osteogenic differentiation phenotypes of rBMSCs at the corresponding training time points and there was a good fit between them, which laid the foundation for the subsequent application of OCNNs. In addition, the feature extraction of single-cell images was performed by Cellprofiler, and the classical shape and material parameters were selected to distinguish the single-cell phenotypes at uniform time points. Also, the SVM models that integrated these parameters were applied simultaneously. Neither single-cell parameters nor SVM models could reach the classification accuracy of OCNNs, reflecting the superiority of OCNNs.

For further application of OCNNs, we conducted dose-relevant predictions of osteogenic drugs and online learning-based predictions for cells cultured on different material surfaces, both of which yielded well results. In the drug screening part, we examined the ability of OCNN in predicting the osteoinduction ability of small molecule drugs or cytokines at an early stage and the results were in high consistency with 14-days alizarin red staining results. On OCNN-based online learning for cells on material surfaces, it can be seen that by online tuning the base model with a small amount of data, better prediction effects can be obtained, overcoming the phenotypical alterations on substrates with different nanotopographies and chemical components.

Compared with previous studies, this study presents several advances. First, previous relative studies adopted some machine learning techniques like linear discriminant analysis (Marklein et al., 2016), unsupervised clustering (Khayal et al., 2018), and bayesian linear regression (Cutiongco et al., 2020), to trace the osteogenic differentiation of stem cells, requiring considerable human efforts to select features and construct models, which are avoided by applying deep learning methods. Second, when we focus on image-based deep learning in the biomedical area, many other deep learning methods like YOLO (Li et al., 2021), GANs (Rubin et al., 2019; Sirinukunwattana et al., 2021) were applied and achieved good results. Compared with those methods, we conducted the evaluations of more parameters (Accuracy, Sensitivity, Specificity, Precision, F1-score, AUC) with satisfactory results. Furthermore, the strong correlation between OCNN's predictions and osteogenic biomarkers was reached, and applications in osteogenic drug and biomaterial screening were implemented by OCNN and OCNN-based online learning. Third, cell morphology showed distinct changes and extensive versatility on substrates with different topography and chemical components. Moreover, conventional approaches usually need to train different models to associate cell phenotypes with microenvironment cues, which demands a large amount of data. By applying online learning techniques to OCNN, deep learning models can drastically improve their performance with a small amount of additional data.

Nonetheless, this study has several limitations. First, only the 2D cellular morphological characteristics were studied. Some 3D characteristics, such as cell volume and cell sphericity, are tightly correlated with 3D microenvironment cues and may have a substantial impact on the fate of cells (Bao et al., 2019; Remuzzi et al., 2020). Second, despite the cytoskeleton, more features such as nuclear skeleton, nuclear transcription factors, and chromatin morphology are also related to osteogenic differentiation. Multicolor immunofluorescent staining can capture more of this high-dimensional information simultaneously; thus, improving the prediction accuracy and generalizability. Prospectively, for future OCNN development on 3D and high-throughput data, advanced network structures and cloud-based techniques need to be constructed to tackle the increased computational complexity and the consumption of computing power. Last, in the original data acquisition process, automatic single-cell identification and segmentation procedures were lacking. In future studies, an automatic technique shall be carried out to accelerate the whole process by some algorithms like the watershed method (Ng et al., 2006), YOLO (Redmon et al., 2016), or U-net (Ronneberger et al., 2015).

In conclusion, to predict the osteogenic differentiation of rBMSCs, a deep learning model, OCNN, was successfully developed based on single-cell LSCM images. The output of the OCNN and POSs correlated well with conventional biomarkers. OCNN showed better predictions than single morphological parameters and support vector machines. It successfully predicted the dose-dependent osteogenic effects of three small molecule drugs (simvastatin, alendronate sodium, and 1α, 25-dihydroxyvitamin D3) and the osteogenic cytokine BMP-2. Moreover, OCNN with online learning successfully predicted the phenotypes associated with osteogenic differentiation on different biomaterial substrates. Therefore, this study preliminarily proved the application value and promising prospect of deep learning-based techniques in osteogenic drug screening, biomaterial development for bone tissue engineering, and cell-matrix interaction research.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/**Supplementary Material**, further inquiries can be directed to the corresponding authors.

ETHICS STATEMENT

The animal study was reviewed and approved by the ethics committee of Chongqing Medical University Affiliated Hospital of Stomatology.

AUTHOR CONTRIBUTIONS

YLI and SY designed the study. YLA did the *in vitro* experiments, developed the model training process, performed the analysis and wrote the manuscript. YF and HZ helped the data acquisition in vitro experiments. NH, and KL assisted in the statistical analysis and image processing. YL and SY provided funding support, supervised the study, and critically revised the manuscript.

FUNDING

This work was supported by the National Natural Science Foundation of China (Grant No. 82171010, 82170936, 81500894, 81901057, 82001103); the Natural Science Foundation of Chongqing, China (Grant No. cstc2021jcyj-jqX0028, cstc2019jcyj-msxmX0366, cstc2019jcyj-bshX0005); Chongqing Yuzhong District Science and Technology Project (Grant No.20190103).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fbioe.2021.802794/full#supplementary-material

REFERENCES

Arora, S., Lin, S., Cheung, C., Yim, E. K. F., and Toh, Y.-C. (2020). Topography Elicits Distinct Phenotypes and Functions in Human Primary and Stem Cell Derived Endothelial Cells. *Biomaterials* 234, 119747. doi:10.1016/j.biomaterials.2019.119747

Bakal, C., Aach, J., Church, G., and Perrimon, N. (2007). Quantitative Morphological Signatures Define Local Signaling Networks Regulating Cell Morphology. *Science* 316, 1753–1756. doi:10.1126/science.1140324

Bao, M., Xie, J., Katoele, N., Hu, X., Wang, B., Piruska, A., et al. (2019). Cellular Volume and Matrix Stiffness Direct Stem Cell Behavior in a 3D Microniche. *ACS Appl. Mater. Inter.* 11, 1754–1759. doi:10.1021/acsami.8b19396

Chen, D., Sarkar, S., Candia, J., Florczyk, S. J., Bodhak, S., Driscoll, M. K., et al. (2016). Machine Learning Based Methodology to Identify Cell Shape Phenotypes Associated with Microenvironmental Cues. *Biomaterials* 104, 104–118. doi:10.1016/j.biomaterials.2016.06.040

Chen, Z., Zhang, Z., Ma, X., Duan, Z., Hui, J., Zhu, C., et al. (2019). Newly Designed Human-Like Collagen to Maximize Sensitive Release of BMP-2 for Remarkable Repairing of Bone Defects. *Biomolecules* 9, 450. doi:10.3390/biom9090450

Chen, D., Dunkers, J. P., Losert, W., and Sarkar, S. (2021). Early Time-Point Cell Morphology Classifiers Successfully Predict Human Bone Marrow Stromal Cell Differentiation Modulated by Fiber Density in Nanofiber Scaffolds. *Biomaterials* 274, 120812. doi:10.1016/j.biomaterials.2021.120812

Chiu, L.-H., Lai, W.-F. T., Chang, S.-F., Wong, C.-C., Fan, C.-Y., Fang, C.-L., et al. (2014). The Effect of Type II Collagen on MSC Osteogenic Differentiation and Bone Defect Repair. *Biomaterials* 35, 2680–2691. doi:10.1016/j.biomaterials.2013.12.005

Cutiongco, M. F. A., Jensen, B. S., Reynolds, P. M., and Gadegaard, N. (2020). Predicting Gene Expression Using Morphological Cell Responses to Nanotopography. *Nat. Commun.* 11, 1384. doi:10.1038/s41467-020-15114-1

Dhaliwal, A., Brenner, M., Wolujewicz, P., Zhang, Z., Mao, Y., Batish, M., et al. (2016). Profiling Stem Cell States in Three-Dimensional Biomaterial Niches Using High Content Image Informatics. *Acta Biomater.* 45, 98–109. doi:10.1016/j.actbio.2016.08.052

Engler, A. J., Sen, S., Sweeney, H. L., and Discher, D. E. (2006). Matrix Elasticity Directs Stem Cell Lineage Specification. *Cell* 126, 677–689. doi:10.1016/j.cell.2006.06.044

Fan, Z.-X., Lu, Y., Deng, L., Li, X.-Q., Zhi, W., Li-Ling, J., et al. (2012). Placenta- Versus Bone-Marrow-Derived Mesenchymal Cells for the Repair of Segmental Bone Defects in a Rabbit Model. *FEBS J.* 279, 2455–2465. doi:10.1111/j.1742-4658.2012.08625.x

Farokhi, M., Mottaghitalab, F., Samani, S., Shokrgozar, M. A., Kundu, S. C., Reis, R. L., et al. (2018). Silk Fibroin/Hydroxyapatite Composites for Bone Tissue Engineering. *Biotechnol. Adv.* 36, 68–91. doi:10.1016/j.biotechadv.2017.10.001

Fu, L., Tang, T., Miao, Y., Zhang, S., Qu, Z., and Dai, K. (2008). Stimulation of Osteogenic Differentiation and Inhibition of Adipogenic Differentiation in Bone Marrow Stromal Cells by Alendronate via ERK and JNK Activation. *Bone* 43, 40–47. doi:10.1016/j.bone.2008.03.008

Guan, M., Yao, W., Liu, R., Lam, K. S., Nolta, J., Jia, J., et al. (2012). Directing Mesenchymal Stem Cells to Bone to Augment Bone Formation and Increase Bone Mass. *Nat. Med.* 18, 456–462. doi:10.1038/nm.2665

He, K., Zhang, X., Ren, S., and Sun, J. (2016). "Deep Residual Learning for Image Recognition," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27-30 June 2016 (IEEE). doi:10.1109/CVPR.2016.90

He, P., Zhang, H., Li, Y., Ren, M., Xiang, J., Zhang, Z., et al. (2020). 1α,25-Dihydroxyvitamin D3-Loaded Hierarchical Titanium Scaffold Enhanced Early Osseointegration. *Mater. Sci. Eng. C* 109, 110551. doi:10.1016/j.msec.2019.110551

Ho, M.-L., Tai, I.-C., Wang, Y.-H., Chen, C.-H., Chuang, S.-C., and Chang, J.-K. (2015). Simvastatin Enhances Rho/Actin/Cell Rigidity Pathway Contributing to Mesenchymal Stem Cells'osteogenic Differentiation. *Int. J. Nanomedicine* 10, 5881–5894. doi:10.2147/IJN.S84273

Khayal, L. A., Grünhagen, J., Provazník, I., Mundlos, S., Kornak, U., Robinson, P. N., et al. (2018). Transcriptional Profiling of Murine Osteoblast Differentiation Based on RNA-Seq Expression Analyses. *Bone* 113, 29–40. doi:10.1016/j.bone.2018.04.006

Kusumoto, D., Seki, T., Sawada, H., Kunitomi, A., Katsuki, T., Kimura, M., et al. (2021). Anti-Senescent Drug Screening by Deep Learning-Based Morphology Senescence Scoring. *Nat. Commun.* 12, 257. doi:10.1038/s41467-020-20213-0

Lamprecht, M. R., Sabatini, D. M., and Carpenter, A. E. (2007). CellProfiler: Free, Versatile Software for Automated Biological Image Analysis. *Biotechniques* 42, 71–75. doi:10.2144/000112257

Li, L., Yang, S., Xu, L., Li, Y., Fu, Y., Zhang, H., et al. (2019). Nanotopography on Titanium Promotes Osteogenesis via Autophagy-Mediated Signaling between YAP and β-Catenin. *Acta Biomater.* 96, 674–685. doi:10.1016/j.actbio.2019.07.007

Li, S., Li, Y., Yao, J., Chen, B., Song, J., Xue, Q., et al. (2021). Label-Free Classification of Dead and Live Colonic Adenocarcinoma Cells Based on 2D

Light Scattering and Deep Learning Analysis. *Cytometry* 99, 1134–1142. doi:10.1002/cyto.a.24475

Lobo, J. L., Laña, I., Del Ser, J., Bilbao, M. N., and Kasabov, N. (2018). Evolving Spiking Neural Networks for Online Learning over Drifting Data Streams. *Neural Networks* 108, 1–19. doi:10.1016/j.neunet.2018.07.014

Lou, Y.-R., Toh, T. C., Tee, Y. H., and Yu, H. (2017). 25-Hydroxyvitamin D3 Induces Osteogenic Differentiation of Human Mesenchymal Stem Cells. *Sci. Rep.* 7, 42816. doi:10.1038/srep42816

Marklein, R. A., Lo Surdo, J. L., Bellayr, I. H., Godil, S. A., Puri, R. K., and Bauer, S. R. (2016). High Content Imaging of Early Morphological Signatures Predicts Long Term Mineralization Capacity of Human Mesenchymal Stem Cells upon Osteogenic Induction. *Stem Cells* 34, 935–947. doi:10.1002/stem.2322

Mauney, J. R., Jaquiéry, C., Volloch, V., Heberer, M., Martin, I., and Kaplan, D. L. (2005). *In Vitro* and *In Vivo* Evaluation of Differentially Demineralized Cancellous Bone Scaffolds Combined with Human Bone Marrow Stromal Cells for Tissue Engineering. *Biomaterials* 26, 3173–3185. doi:10.1016/j.biomaterials.2004.08.020

McBeath, R., Pirone, D. M., Nelson, C. M., Bhadriraju, K., and Chen, C. S. (2004). Cell Shape, Cytoskeletal Tension, and RhoA Regulate Stem Cell Lineage Commitment. *Dev. Cel* 6, 483–495. doi:10.1016/s1534-5807(04)00075-9

Moen, E., Bannon, D., Kudo, T., Graf, W., Covert, M., and Van Valen, D. (2019). Deep Learning for Cellular Image Analysis. *Nat. Methods* 16, 1233–1246. doi:10.1038/s41592-019-0403-1

Ng, H., Ong, S., Foong, K., Goh, P.-S., and Nowinski, W. (2006). "Medical Image Segmentation Using K-Means Clustering and Improved Watershed Algorithm," in IEEE Southwest Symposium on Image Analysis and Interpretation, Denver, CO, USA, 26-28 March 2006 (IEEE). doi:10.1109/SSIAI.2006.1633722

Oei, R. W., Hou, G., Liu, F., Zhong, J., Zhang, J., An, Z., et al. (2019). Convolutional Neural Network for Cell Classification Using Microscope Images of Intracellular Actin Networks. *PLoS One* 14, e0213626. doi:10.1371/journal.pone.0213626

Prasad, A., and Alizadeh, E. (2019). Cell Form and Function: Interpreting and Controlling the Shape of Adherent Cells. *Trends Biotechnology* 37, 347–357. doi:10.1016/j.tibtech.2018.09.007

Qi, C., Deng, Y., Xu, L., Yang, C., Zhu, Y., Wang, G., et al. (2020). A Sericin/Graphene Oxide Composite Scaffold as a Biomimetic Extracellular Matrix for Structural and Functional Repair of Calvarial Bone. *Theranostics* 10, 741–756. doi:10.7150/thno.39502

Redmon, J., Divvala, S., Girshick, R., and Farhadi, A. (2016). "You Only Look once: Unified, Real-Time Object Detection," in Paper Presented at the Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27-30 June 2016 (IEEE), 779–788. doi:10.1109/cvpr.2016.91

Remuzzi, A., Bonandrini, B., Tironi, M., Longaretti, L., Figliuzzi, M., Conti, S., et al. (2020). Effect of the 3D Artificial Nichoid on the Morphology and Mechanobiological Response of Mesenchymal Stem Cells Cultured *In Vitro*. *Cells* 9, 1873. doi:10.3390/cells9081873

Ren, M., Li, Y., Zhang, H., Li, L., He, P., Ji, P., et al. (2021). An Oligopeptide/Aptamer-Conjugated Dendrimer-Based Nanocarrier for Dual-Targeting Delivery to Bone. *J. Mater. Chem. B* 9, 2831–2844. doi:10.1039/d0tb02926b

Ronneberger, O., Fischer, P., and Brox, T. (2015). "U-Net: Convolutional Networks for Biomedical Image Segmentation," in Paper Presented at the International Conference on Medical Image Computing and Computer-Assisted Intervention, 18 November 2015 (Munich, Germany: Springer). doi:10.1007/978-3-319-24574-4_28

Rubin, M., Stein, O., Turko, N. A., Nygate, Y., Roitshtain, D., Karako, L., et al. (2019). TOP-GAN: Stain-Free Cancer Cell Classification Using Deep Learning with a Small Training Set. *Med. image Anal.* 57, 176–185. doi:10.1016/j.media.2019.06.014

Salazar, V. S., Gamer, L. W., and Rosen, V. (2016). BMP Signalling in Skeletal Development, Disease and Repair. *Nat. Rev. Endocrinol.* 12, 203–221. doi:10.1038/nrendo.2016.12

Simonyan, K., and Zisserman, K. (2014). Very Deep Convolutional Networks for Large-Scale Image Recognition. arXiv.20141409.1556.

Simonyan, K., Vedaldi, A., and Zisserman, A. (2014). Deep inside Convolutional Networks: Visualising Image Classification Models and Saliency Maps. arXiv preprint arXiv.1312.6034.

Sirinukunwattana, K., Domingo, E., Richman, S. D., Redmond, K. L., Blake, A., Verrill, C., et al. (2021). Image-Based Consensus Molecular Subtype (imCMS) Classification of Colorectal Cancer Using Deep Learning. *Gut* 70, 544–554. doi:10.1136/gutjnl-2019-319866

Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., and Wojna, Z. (2016). "Rethinking the Inception Architecture for Computer Vision," in Paper Presented at the Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, June 27–June 30, 2016, 2818–2826. doi:10.1109/cvpr.2016.308

Thomas, C. H., Collier, J. H., Sfeir, C. S., and Healy, K. E. (2002). Engineering Gene Expression and Protein Synthesis by Modulation of Nuclear Shape. *Proc. Natl. Acad. Sci.* 99, 1972–1977. doi:10.1073/pnas.032668799

Treiser, M. D., Yang, E. H., Gordonov, S., Cohen, D. M., Androulakis, I. P., Kohn, J., et al. (2010). Cytoskeleton-Based Forecasting of Stem Cell Lineage Fates. *Proc. Natl. Acad. Sci.* 107, 610–615. doi:10.1073/pnas.0909597107

Van der Maaten, L., and Hinton, G. (2008). Visualizing Data Using T-SNE. *Machine Learn.* 9, 2579–2605.

Waisman, A., La Greca, A., Möbbs, A. M., Scarafía, M. A., Santín Velazque, N. L., Neiman, G., et al. (2019). Deep Learning Neural Networks Highly Predict Very Early Onset of Pluripotent Stem Cell Differentiation. *Stem Cel. Rep.* 12, 845–859. doi:10.1016/j.stemcr.2019.02.004

Werner, M., Blanquer, S. B. G., Haimi, S. P., Korus, G., Dunlop, J. W. C., Duda, G. N., et al. (2017). Surface Curvature Differentially Regulates Stem Cell Migration and Differentiation via Altered Attachment Morphology and Nuclear Deformation. *Adv. Sci.* 4, 1600347. doi:10.1002/advs.201600347

Wu, P.-H., Gilkes, D. M., Phillip, J. M., Narkar, A., Cheng, T. W.-T., Marchand, J., et al. (2020). Single-Cell Morphology Encodes Metastatic Potential. *Sci. Adv.* 6, eaaw6938. doi:10.1126/sciadv.aaw6938

Yang, F., Yang, L., Li, Y., Yan, G., Feng, C., Liu, T., et al. (2017). Melatonin Protects Bone Marrow Mesenchymal Stem Cells against Iron Overload-Induced Aberrant Differentiation and Senescence. *J. Pineal Res.* 63, e12422. doi:10.1111/jpi.12422

Yang, X., Wang, Y., Byrne, R., Schneider, G., and Yang, S. (2019). Concepts of Artificial Intelligence for Computer-Assisted Drug Discovery. *Chem. Rev.* 119, 10520–10594. doi:10.1021/acs.chemrev.8b00728

Yelin, I., Snitser, O., Novich, G., Katz, R., Tal, O., Parizade, M., et al. (2019). Personal Clinical History Predicts Antibiotic Resistance of Urinary Tract Infections. *Nat. Med.* 25, 1143–1152. doi:10.1038/s41591-019-0503-6

Zhang, H., Yang, S., Masako, N., Lee, D. J., Cooper, L. F., and Ko, C.-C. (2015). Proliferation of Preosteoblasts on TiO2 Nanotubes Is FAK/RhoA Related. *RSC Adv.* 5, 38117–38124. doi:10.1039/c4ra16803h

Zhang, H., Cooper, L. F., Zhang, X., Zhang, Y., Deng, F., Song, J., et al. (2016). Titanium Nanotubes Induce Osteogenic Differentiation through the FAK/RhoA/YAP Cascade. *RSC Adv.* 6, 44062–44069. doi:10.1039/c6ra04002k

Zhu, Y., Huang, R., Wu, Z., Song, S., Cheng, L., and Zhu, R. (2021). Deep Learning-Based Predictive Identification of Neural Stem Cell Differentiation. *Nat. Commun.* 12, 2614. doi:10.1038/s41467-021-22758-0

Conflict of Interest: The 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.

frontiers in Neuroscience

ORIGINAL RESEARCH
published: 11 February 2022
doi: 10.3389/fnins.2022.782367

Implementation of a Convolutional Neural Network for Eye Blink Artifacts Removal From the Electroencephalography Signal

Marcin Jurczak, Marcin Kołodziej and Andrzej Majkowski*

Institute of Theory of Electrical Engineering, Measurement and Information Systems, Warsaw University of Technology, Warsaw, Poland

Electroencephalography (EEG) signals are disrupted by technical and physiological artifacts. One of the most common artifacts is the natural activity that results from the movement of the eyes and the blinking of the subject. Eye blink artifacts (EB) spread across the entire head surface and make EEG signal analysis difficult. Methods for the elimination of electrooculography (EOG) artifacts, such as independent component analysis (ICA) and regression, are known. The aim of this article was to implement the convolutional neural network (CNN) to eliminate eye blink artifacts. To train the CNN, a method for augmenting EEG signals was proposed. The results obtained from the CNN were compared with the results of the ICA and regression methods for the generated and real EEG signals. The results obtained indicate a much better performance of the CNN in the task of removing eye-blink artifacts, in particular for the electrodes located in the central part of the head.

Keywords: artifacts, electroencephalography, electrooculography, convolutional neural network, independent component analysis

OPEN ACCESS

Edited by:
Mariusz Pelc, University of Greenwich, United Kingdom

Reviewed by:
Aleksandra Dagmara Kawala-Sterniuk, Opole University of Technology, Poland
Izabela Rejer, West Pomeranian University of Technology, Poland
Rahib Abiyev, Near East University, Cyprus

****Correspondence:***
Marcin Kołodziej
marcin.kolodziej@pw.edu.pl

Specialty section:
This article was submitted to Neural Technology, a section of the journal Frontiers in Neuroscience

Received: *24 September 2021*
Accepted: *10 January 2022*
Published: *11 February 2022*

Citation:
Jurczak M, Kołodziej M and Majkowski A (2022) Implementation of a Convolutional Neural Network for Eye Blink Artifacts Removal From the Electroencephalography Signal. Front. Neurosci. 16:782367. doi: 10.3389/fnins.2022.782367

INTRODUCTION

Motivation

Electroencephalography (EEG) is a method of examining brain activity commonly used in medical diagnostics (Levin et al., 2018; Browarska et al., 2021). Unfortunately, in some cases direct analysis of the EEG signal is very difficult or even impossible due to the presence of artifacts (Kilicarslan and Contreras-Vidal, 2017; Kawala-Sterniuk et al., 2020; Zhang C. et al., 2020). There are many types of physiological artifacts, for example those caused by muscle clenching, jaw, tongue movements, or eye movements. One of the strongest artifacts that interfere with the analysis of EEG signals are electrooculography (EOG) artifacts. EOG artifacts are generally high-amplitude patterns in the brain signal caused by blinking of the eyes or low-frequency patterns caused by movements (such as rolling) of the eyes (Anderer et al., 1999). EOG activity has a wide frequency range, being maximal at frequencies below 4 Hz, and is most prominent over the anterior head regions (McFarland et al., 1997). The subject of the article concerns the elimination of EOG artifacts created during blinking (Pham et al., 2017).

Generally, the concept of EOG artifacts is broader and covers both the activity of eye movement and blinking. For the purposes of this article, the authors equate the concept of EOG with eye

blinks (EB). To eliminate them, the authors proposed a deep neural network-based method and compared its operation with the popular methods of artifact elimination – ICA and regression. During the research, the focus was on the analysis of real EEG signals recorded with the use of a professional biomedical signal amplifier. Twenty people participated in the experiment, and each session lasted about 60 min. The authors also used computer-generated signals to train and test the neural network. For this purpose, an algorithm was created to generate EEG/EOG signals.

State of the Art

Many methods are used to remove artifacts from the EEG signal (Mumtaz et al., 2021). The simplest of them just reject those fragments of EEG signals with artifacts. Unfortunately, this approach results in the loss of all information from the rejected signal fragments (Hasasneh et al., 2018; Khatwani et al., 2018; Nejedly et al., 2019; Tosun and Kasım, 2020; Iaquinta et al., 2021; Placidi et al., 2021). In addition, we must have a very good artifact detection algorithm that will allow us to identify them. Artifacts can also be selected by an expert by visual inspection. This approach is not always possible and usually applies to off-line analyzes. Artifact removal approaches may require so-called reference channels (Mumtaz et al., 2021). The regression method requires such a reference channel, that is, the one based on which artifacts of the remaining channels are removed (Mannan et al., 2018). Usually, one of the channels from the "frontal" position or the EOG signal is chosen as the reference channel. Then, with the use of signals from the reference channel, the regression method eliminates artifacts from successive electrodes (propagated from the reference electrode to the others). This means that the artifacts are not removed from the reference electrode (it only serves to eliminate artifacts from other electrodes). When artifacts are removed using a reference electrode (Mannan et al., 2018), it is assumed that neural activity (EEG) and electro-oculographic signals (EOG) are not correlated. In turn, the independent component analysis method (ICA) (Jiang et al., 2019) does not require a reference channel. The ICA method allows for the determination of the signal components (statistically independent), which enables the rejection of artifacts and disturbances. This method allows the removal of artifacts from all electrodes. In the ICA method, rejected components are often selected on the basis of their visualization. It requires expert knowledge (Mannan et al., 2018) and signal recording with the use of multiple channels. However, there are methods that allow for automatic selection of rejected components (Li et al., 2017). Hybrid methods are also used to remove artifacts (Li et al., 2017; Mumtaz et al., 2021). Their idea is to use more than one algorithm to remove artifacts. An example is the use of the combination of wavelet transform and blind signal separation (BSS) (Rakibul Mowla et al., 2015). By means of BBS, signals are decomposed into components, and then the components are subjected to the wavelet transform. The next step is to remove components that contain artifacts based on thresholding and then reconstruct the signal. Other examples of hybrid methods are the combination of adaptive filtering and BSS and the combination of BSS and supporting vector machine (SVM).

Deep learning methods are becoming more and more popular every year. An example of this method may be the convolutional neural network (CNN), which has a very wide application in many different fields of science (Arora et al., 2020). An example may be the field related to computer vision and image recognition (Chen et al., 2019; Lou and Shi, 2020). CNN has also found application in neuroinformatics to recognize emotions (Zhang Y. et al., 2020) and detect mild depression (Li et al., 2020) using encephalography. Another application is the detection of myocardial infarction based on the ECG signal (Natesan et al., 2020). On the basis of existing applications, it is assumed that convolutional networks can also work well in tasks related to cleaning biomedical signals from artifacts. Moreover, CNN offers very wide possibilities to select structures and hyperparameters (Arora et al., 2020).

In work (Garg et al., 2017) a 10-layer convolutional neural network (CNN) is presented, which directly labels eye-blink artifacts. Thirty subjects were tested. The classification accuracy achieved was 99.67%, the sensitivity was 97.62%, the specificity was 99.77%, and the ROC AUC was 98.69%. The authors also showed that the learned spatial features correspond to those that human experts typically use, which corroborated the validity of the model. In work (Placidi et al., 2021) independent component analysis (ICA) is used to split the signal into independent components (ICs) whose re-projections on 2D scalp topographies (images), also called topoplots, allow to separate artifacts and useful brain signals (UBS). In the article, a completely automatic and effective framework for EEG artifact recognition by IC topoplots is presented, based on 2D convolutional neural networks (CNNs), capable of dividing topoplots into four classes: three types of artifacts and UBS. Experiments carried out on public EEG datasets showed an overall accuracy of more than 98%. In Iaquinta et al. (2021) a reliable and user-independent algorithm is presented to detect and remove eye blink in EEG signals using CNN. For training and validation, three sets of public EEG data were used. All three sets contain samples obtained while the recruited subjects performed assigned tasks that included blinking voluntarily at specific moments, watching a video, and reading an article. The model used in this study was able to have an embracing understanding of all the features that distinguish a trivial EEG signal from a signal contaminated with eye blink artifacts. In Sun et al. (2020) a one-dimensional residual convolutional neural network (1D-ResCNN) model for raw waveform-based EEG denoising is proposed. An end-to-end (i.e., waveform in and waveform out) manner was used to map a noisy EEG signal to a clean EEG signal. The proposed model was evaluated on the EEG signal from the CHB-MIT Scalp EEG Database, and the added noise signals were obtained from the database. The proposed model was compared with independent component analysis (ICA), fast independent composite analysis (FICA), recursive least squares (RLS) filter, wavelet transform (WT), and deep neural network (DNN) models. Experimental results show that the proposed model can produce cleaner waveforms and achieve a significant improvement in SNR and *RMSE*. Meanwhile, the proposed model can also preserve the nonlinear characteristics of the EEG signals. In Yang et al. (2018) the use of the deep learning network (DLN) to remove ocular

artifacts (OA) in EEG signals was investigated. The proposed method consists of an offline stage and an online stage. In the offline stage, training samples without OAs were intercepted and used to train a DLN to reconstruct the EEG signals. In the online stage, trained DLN was used as a filter to automatically remove OAs from contaminated EEG signals. The advantages of the proposed method are the non-use of additional EOG reference signals, the possibility of analyzing any number of EEG channels, time savings, and strong generalizability. The proposed method was compared with the classic independent component analysis (ICA), kurtosis-ICA (K-ICA), second-order blind identification (SOBI), and a shallow network method. Experimental results show that the proposed method performs better even for very noisy EEG.

A large number of teaching examples are needed to train the CNN. Unfortunately, the number of recorded EEG signal examples is often too small. Therefore, there is a need to use a technique called augmentation to increase the number of training examples. Various methods of augmentation of EEG signals are presented in Lashgari et al. (2020). In Lashgari et al. (2020) the authors indicate that the most popular methods of augmentation are those based on noise addition, GAN networks, sliding window, sampling, Fourier transform, recombination of segmentation. Wang et al. (2018) added Gaussian white noise to training data (in the time domain) to obtain new samples for an emotion-recognition task. Differential entropy (DE) features were used to train classifiers. For EEG signals, the DE features are equivalent to the logarithm of the energy spectrum in the delta (1–3 Hz), theta (4–7 Hz), alpha (8–13 Hz), beta (14–30 Hz), and gamma (31–50 Hz) frequency bands. The authors opted for Gaussian noise due to concerns that adding some local noise, such as Poisson or salt-and-pepper, may change the intrinsic features of EEG signals. In Wang et al. (2018) two basic data augmentation approaches used in image processing were implemented: geometric transformation and noise addition. In Luo et al. (2020) methods based on two deep generative models, variational autoencoder (VAE) and generative adversarial network (GAN), and two data augmentation strategies were proposed. To evaluate the effectiveness of these methods, a systematic experimental study was carried out on two public EEG datasets for emotion recognition, namely SEED and DEAP. First, realistic-like EEG training data in two forms were generated: power spectral density and differential entropy. Then, the original training data sets were augmented with a different number of realistic-like generated EEG data. Finally, support vector machines and deep neural networks with shortcut layers were trained to build affective models using the original and augmented training datasets. In Bao et al. (2021) a data augmentation model named VAE-D2GAN was proposed for EEG-based emotion recognition using a generative adversarial network. EEG features representing different emotions were extracted as topological maps of differential entropy (DE) in five classical frequency bands. The proposed model was designed to learn the distributions of these features for real EEG signals and generate artificial samples for training. The variational autoencoder (VAE) architecture can learn the spatial distribution of the actual data through a latent vector and is introduced into the dual discriminator GAN to improve the diversity of the generated artificial samples.

Aim of the Paper

We propose a method based on the convolutional neural network (CNN) that allows the removal of eye blink artifacts from the EEG signal. The results obtained with the use of CNN were compared with the most popular methods of artifact removal – ICA and regression. For the implementation of the CNN-based method, it was necessary to achieve the intermediate goal, which was the implementation of the EEG signal and EOG artifact generators. The use of only a real EEG signal does not give the possibility of direct evaluation of the obtained results because we do not have a reference (it is not known what the real EEG signal is). Generated signals also enable better training of the neural network.

Signal fragments from 2 channels are fed to the CNN input. The first channel contains the eye blink signal and the second channel contains the EEG signal from which we want to remove the eye blink artifacts. The idea is presented in **Figure 1**. In this case, at the CNN input, fragments of the signal from the Fp1 electrode (eye blink artifacts) and the signal from which we want to eliminate blinks (the C3 electrode) are fed. CNN eliminates the eye blink signal (C3 – CNN). Then the input signals are shifted. This operation can be performed for each electrode.

The article is organized as follows. In the section "Materials," two types of EEG/EOG signals used during the experiments were presented: real and generated signals containing eye blink artifacts. Details on generating artificial EEG signals with eye blink artifacts are also provided. The section "Methods" describes the structure of the CNN proposed to remove artifacts and details of training the network. Furthermore, two commonly used methods for removing eye blink artifacts are presented, i.e., independent component analysis and regression. The section "Results and Discussion" presents the results of the comparison of ICA, REG, and CNN methods for removing eye blink artifacts. The advantages and disadvantages of using CNN for this task are discussed.

MATERIALS

To develop and evaluate eye blink artifact removal algorithms, we decided to use two datasets. The first set contains the real signals recorded for the N-back experiment. The N-back task is a standard method used to examine memory and attention (Salimi et al., 2020). This data set has a long duration and contains registrations from multiple users. Thanks to our algorithm, it was possible to generate a second set of artificial EEG/EOG signals. This data set was of particular importance for CNN training and testing.

Real Electroencephalography Signals

Real EEG signals were recorded during an EEG test conducted with 20 people during an N-back task. EEG signals were recorded for the purposes of previous research related to the detection of user fatigue (Kołodziej et al., 2020). However, its use for research on methods to remove EB artifacts was not accidental.

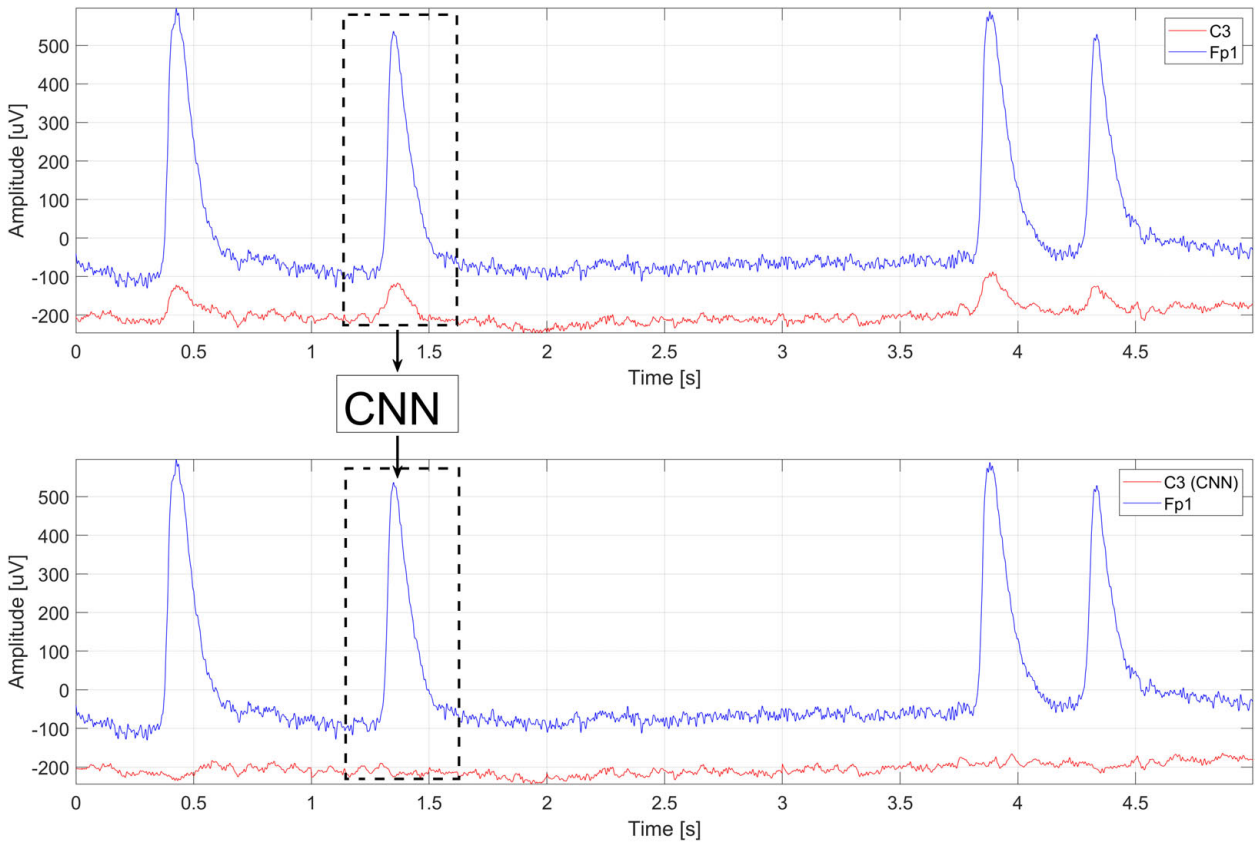

FIGURE 1 | The idea of using the CNN to eliminate artifacts.

EEG signals were recorded for a relatively long time. There are numerous eye blink (EB) artifacts in the EEG signal. Participants (women and men) were 19–25 years old. They were informed about the overall purpose and organization of the experiment. The whole experiment lasted about 80 min and the experiment was always carried out at 10:00 am in a single session. Participants were recruited through an advertisement on the Internet and social media. During recruitment, they were asked to complete a survey via the Internet to collect basic information about them, such as age, sex, education, and presence/absence of neurological and psychiatric diseases. We only invited those who met the basic requirements (including, but not limited to, written permission to participate in the experiment and confirmation of no medical burden).

The letters were presented to the participants on a computer screen (one at a time). The task was to indicate whether the letter presented currently is the same as $N = 2$ letters back. Each participant completed the N-back task for 60 min. To register the EEG signal, we used a professional biomedical signal amplifier g.USBamp and an EEG cap equipped with 16 electrodes. The distribution of electrodes and their names are presented in **Figure 2**. The sampling frequency was 512 Hz. Electrodes were arranged according to the international 10–20 system: Oz, O2, O1, Pz, P4, P3, C4, C3, Cz, F8, F7, Fz, F4, F3, Fp1, and F9. No preprocessing methods were used.

A fragment of the EEG signal from one of the subjects is shown in **Figure 3**. Eye blink artifacts are very clearly visible, located around 10 and 12.5 s. The highest amplitudes of artifacts were recorded on the Fp1 and P3 electrodes. The propagation of artifacts to other electrodes is also visible.

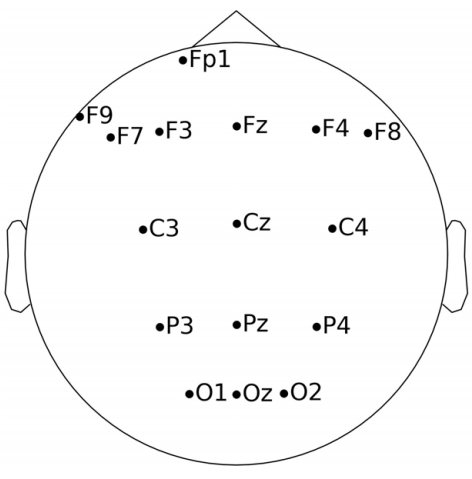

FIGURE 2 | Electrodes and their location on the head.

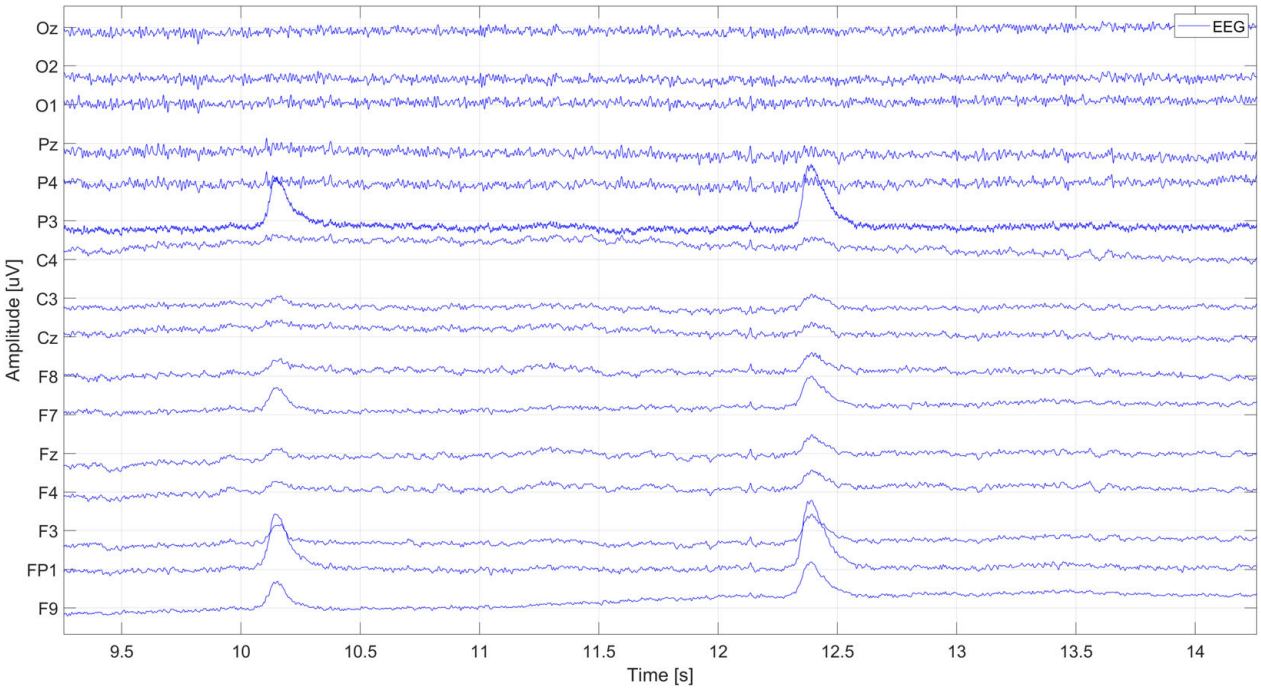

FIGURE 3 | Recorded EEG signal fragment.

The 1-s window presenting the signal fragment from **Figure 3** is shown in **Figure 4**. It is a zoom-in on the eye blink artifact occurrence.

Generated Electroencephalography/ Electrooculography Signals

The real EEG signal contains various types of artifacts that appear when the test is performed. However, we do not have a reference signal from which to conclude what the EEG signal should look like after the artifacts have been removed. To enable such an evaluation, we have developed software that allows the generation of artificial EEG signals without artifacts and the addition of EOG artifacts to them. Due to this, it is possible to compare the performance of each of the analyzed methods (we have an EEG signal contaminated with artifacts and a clean EEG signal that should be obtained after cleaning). Statistical parameters of the generated signals were determined on the basis of observations of real signals recorded during the tests. These are the standard deviation (5–15 μV), the peak-to-peak value (45–100 μV), the interval between the appearance of eye blinks (0.5–4 s), the amplitude of eye blinks (0–650 μV), and the length of the signal in seconds.

The first step is to generate an EEG signal and then add artifacts with the appropriate electrode-dependent gain to it. The EEG signal can be generated in many ways. One of them is to get pink noise with given parameters. Pink noise, also known as 1/f noise, is a random stochastic process or a signal whose mean power spectral density is inversely proportional to frequency (Isar and Gajitzki, 2016). According to Voytek et al. (2015), many natural phenomena, including electroencephalography, can be described by 1/f noise. We decided to use a different method of EEG signal generation (Sakai et al., 2017). This method generates an EEG signal based on random modifications of the spectrum of a real signal. As the EEG reference signal, the signal fragment from an electrode subjected to minimal EOG interference (e.g., the Cz electrode) is selected. Thus, the generated signal corresponds best to the undisturbed EEG signal. The signal generation process begins with the calculation of the spectrum of a given fragment of the real EEG signal using the fast Fourier transform (FFT). Then, random coefficients are generated and a modified spectrum is created by adding/subtracting the random values to the FFT coefficients of the real EEG signal. The spectrum-modifying coefficients are in the range of ± 2 μV. The last step is to apply the inverse Fourier transform (IFFT), which enables us to obtain a time-domain signal similar to the real EEG signal. The generated EEG signal has a spectrum similar to pink noise.

The generated EEG signal is in the form of a 1-s window that can be combined into a signal with a predetermined number of seconds. The generation algorithm ensures that the amplitude differences at the border of the joined windows are not too large. The incoming signal can differ up to 7 μV from the last sample of the signal already created – this value was determined based on the observation of real signals. The generated signal (on different electrodes) based on the real EEG signal from the Cz electrode is shown in **Figure 5**. The spectra of the individual signals are also shown there.

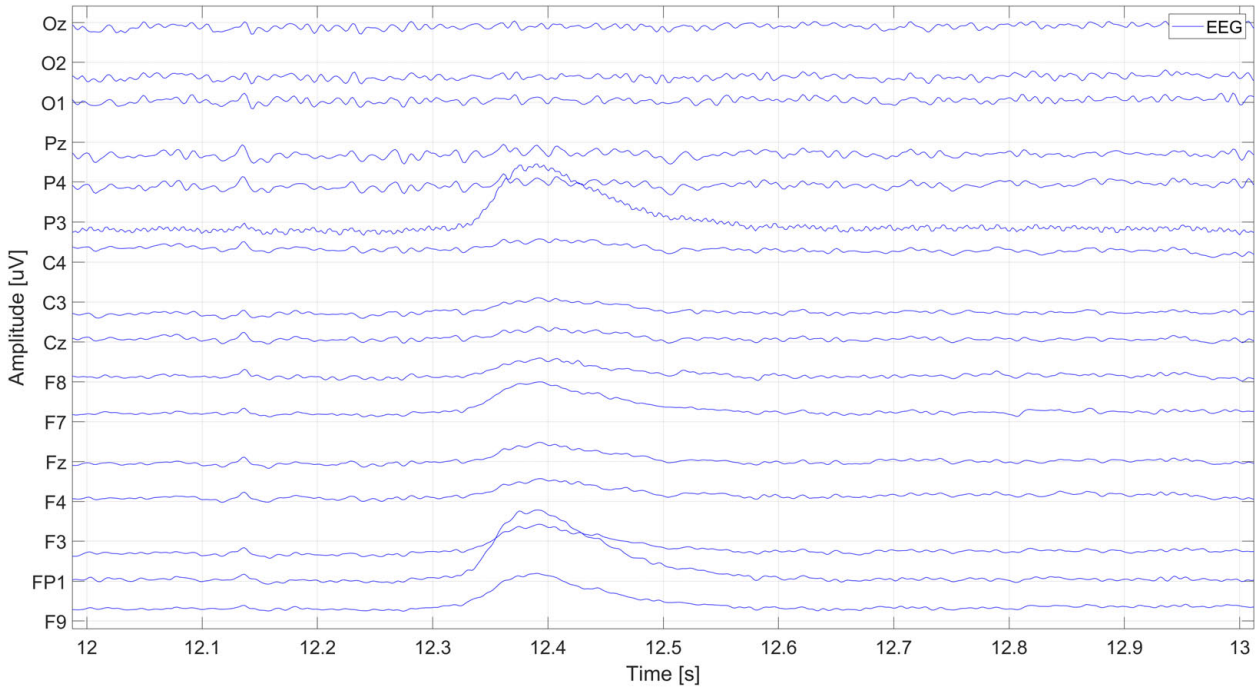

FIGURE 4 | The EEG signal fragment containing the occurrence of an eye blink artifact.

The next step is to generate eye blink artifacts, according to the parameters determined on the basis of observation of real signals. The propagation of artifacts on the EEG signal on individual electrodes is very important here. For this purpose, the range of coefficients responsible for artifact propagation was established for each of the electrodes, depending on their position. Artifacts were added to the pure signals after appropriate amplification or attenuation, depending on the coefficient specified for the given electrode. The eye blink artifact resembles the shape of a Gaussian window and this shape was used to generate the artifacts (Alquran et al., 2019). Eye blink artifacts were generated and added to the clean signal (with an appropriate time interval). **Figure 6** presents 5 s of pure EEG signal and EEG signal with eye blink artifacts propagated on individual electrodes.

Zoom in on the EEG signal (**Figure 6**) containing the occurrence of the EOG artifact and its propagation to the remaining electrodes is shown in **Figure 7**.

The generated signals allow the check and comparison of individual methods because we have a reference in the form of a pure EEG signal. In the case of cleaning the real EEG signal from artifacts, it is not possible to compare the waveforms with the reference signal (pure EEG) because it is not known. Artificially generated signals were also used to train CNN.

METHODS

In our research, we compared the use of the CNN method to remove EB artifacts with the two best-known methods: regression (REG) and independent component analysis (ICA). Each method works differently. The ICA method tries to find the most independent components. Based on expert knowledge or quantitative measures, we are able to identify ICA components responsible for artifacts and remove them. The regression method removes artifacts from individual channels. For this, it is required to indicate the signal in which the artifacts are found. We assumed that this is the signal from the Fp1 electrode.

Independent Component Analysis

The ICA method (Cheng et al., 2019; Jiang et al., 2019) allows the removal of artifacts from the EEG signal without the need for a reference channel. The ICA method works by decomposing the recorded signal into independent components. In principle, the components will include those responsible for the sources of artifacts. Such artifact-containing components are rejected automatically or by an expert, and the signal is then reconstructed by mixing the remaining components. As a result, we get signals without artifacts. The problem can be described by the equation (Jiang et al., 2019):

$$\boldsymbol{X} = \boldsymbol{W} * \boldsymbol{S}$$

We assume that $\boldsymbol{X}$ is the matrix of signals recorded by the measuring electrodes, $\boldsymbol{W}$ is the mixing matrix, and $\boldsymbol{S}$ is the matrix of source signals. After transforming the equation, we get the unknown source signals:

$$\boldsymbol{S} = \boldsymbol{W}^{-1} * \boldsymbol{X}$$

The assumption of the ICA method (Jiang et al., 2019) is the statistical independence of the source signals. The aim of ICA is to find such a mixing matrix $\boldsymbol{W}$ that allows one to obtain the

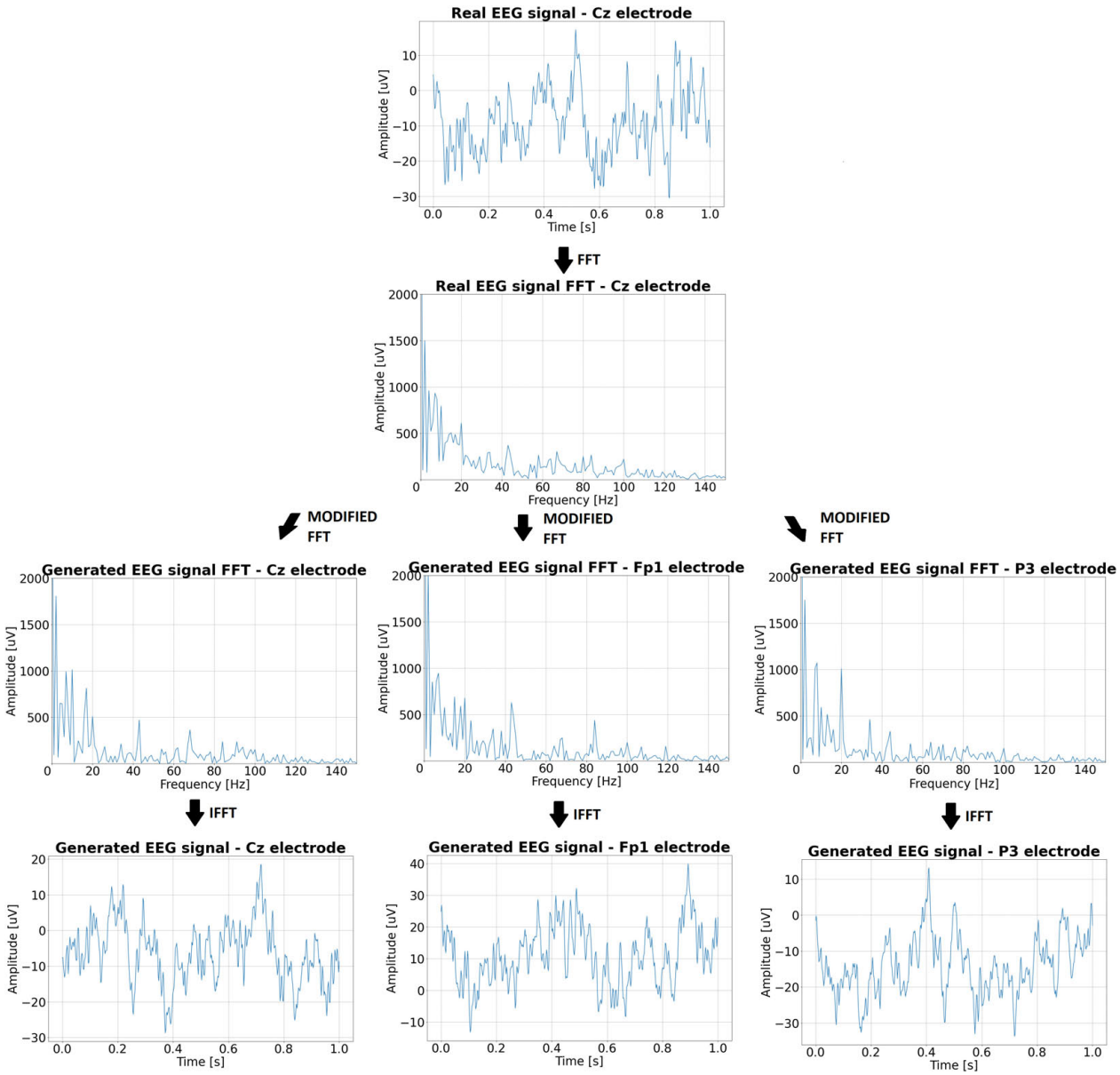

FIGURE 5 | The real EEG signal from the Cz electrode, its spectrum, and the artificial signals generated on its basis along with the spectra.

most independent result signals. If the components responsible for artifacts (eye blink or other) are found, it is enough to reset the appropriate weights of the $\boldsymbol{W}$ matrix and then mix the other components. The matrix W_m is the modified mixing matrix.

$$\boldsymbol{X} = \boldsymbol{W_m} * \boldsymbol{S}$$

Before making the transformation, the low frequencies were filtered out (high-pass filtering, cutoff frequency 1 Hz, Butterworth filter order 6). We decided to choose 15 ICA components. Two components 0 and 1 (associated with artifacts) were removed and then the signal without these components was reconstructed. Examples of the components calculated for the real signal are shown in **Figure 8**. Each component was visually assessed and the eye blink components were selected.

Linear Regression

The regression method, according to Urigüen and Zapirain (2015), was very often used to remove EOG artifacts in the 1990s due to its simplicity and low computational requirements. Regression is still a popular and commonly used method of artifact removal (Ranjan et al., 2021). The method requires a reference channel, which was chosen by us as Fp1 (the electrode closest to the eye). The regression method assumes (Urigüen and Zapirain, 2015) that each EEG measurement channel is the sum of a certain clean source signal and a reference signal (containing

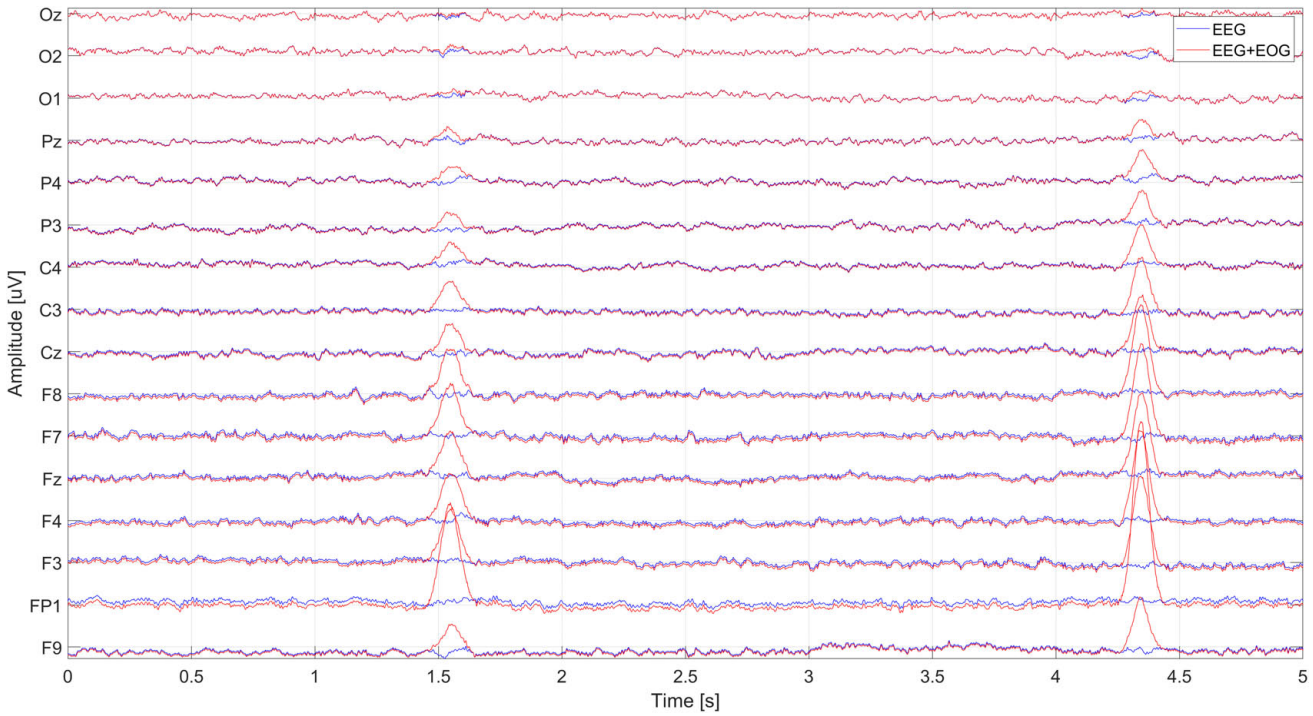

FIGURE 6 | Artificially generated EEG and EEG + EOG signals.

artifacts). The aim of regression is to estimate the optimal value of the propagation coefficient for each of the electrodes (except the reference one), allowing proper removal of artifacts. Removal of artifacts is the subtraction of a certain amount of the EOG reference signal (from the Fp1 electrode) from the contaminated EEG. As a result, we obtained a cleaned signal. In general, the regression equation (Jiang et al., 2019) can be written as:

$$EEG_{clear} = EEG_{noised} - B * EOG_{ref}$$

EEG_{clear} is the artifact-cleaned EEG signal, EEG_{noised} is the signal before artifact removal, B is the propagation factor, and EEG_{ref} is the EOG reference signal. With multiple regression (Urigüen and Zapirain, 2015), the signals measured on individual electrodes are influenced by more than one reference signal, for example, horizontal, vertical, or radial ocular artifacts.

The linear regression method used by us takes the data from the Fp1 electrode (EOG_{ref}) and subtracts from each sample the mean of the signal recorded on that channel (EOG_{ref_maen}). Assuming that we have N samples, we can represent this as an equation:

$$\forall n \in \{1, \ldots, N\} \; EOG_{ref}(n) = EOG_{ref}(n) - EOG_{ref\_maen}$$

The signal is then multiplied by its transposition to compute the cov factor.

$$cov = EOG_{ref} * {EOG_{ref}}^{T}$$

In the next step, for each of the channels (separately, in order to reduce memory use), data is collected, and the average is calculated, which is subtracted from the entire signal for a given channel, similarly to the reference channel. To remove artifacts, from a channel (other than Fp1) containing N samples, the recorded EEG_{noised} signal with an average equal to EOG_{noised_maen} is transformed as follows:

$$\forall n \in \{1, \ldots, N\} \; EOG_{noised}(n) = EOG_{noised}(n) - EOG_{noised\_maen}$$

Then the coefficient B is calculated as a solution to the linear equation:

$$cov * B = EOG_{ref} * {EOG_{noised}}^{T}$$

After transforming the equation we get:

$$B = cov^{-1} * EOG_{ref} * {EOG_{noised}}^{T}$$

In the next step, the reference signal multiplied by factor B is subtracted from the signal from the analyzed channel (electrode). In this way, we obtain a signal cleaned of artifacts for a given electrode (EEG_{clear}).

$$EEG_{clear} = EOG_{noised} - B * EOG_{ref}$$

The algorithm works in this way until all channels are cleared (apart from the reference channel, which in our case is Fp1).

Convolutional Neural Network

Convolutional neural networks (CNN) (Arora et al., 2020) are most often used in problems related to computer vision. These networks can be used not only for classification but also for regression problems. A characteristic feature of CNN compared to the traditional neural network is the fact that during its

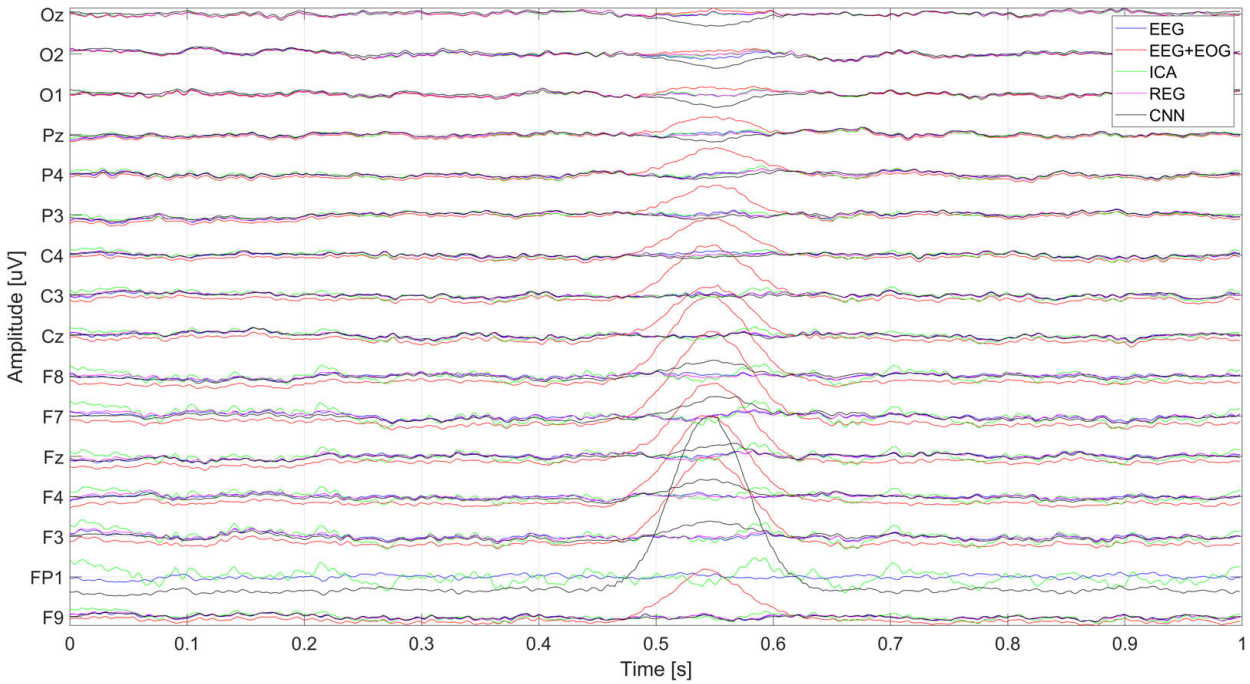

FIGURE 7 | Close up on the eye blink artifact.

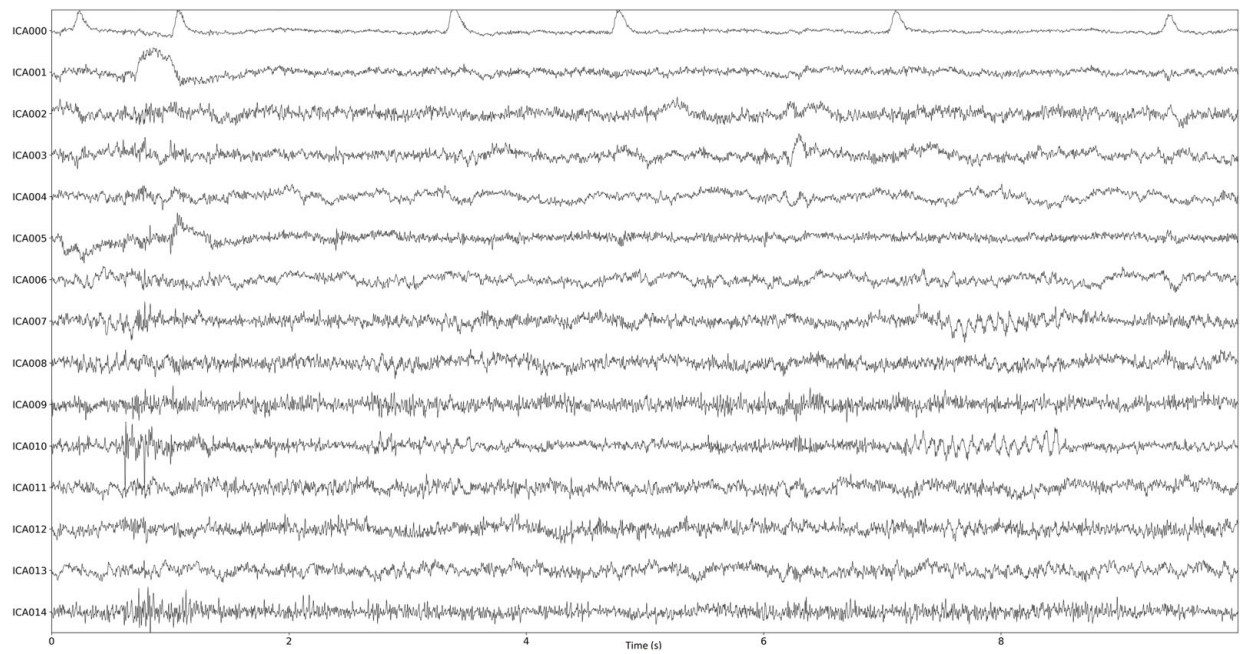

FIGURE 8 | ICA components for the real signal sample.

operation it focuses on the extraction of features (Arora et al., 2020). Each CNN consists of four basic layers – the convolution layer (filters with given shapes that allow for the extraction of features), pooling layer (they are used to reduce the size of analyzed data, we distinguish several types of pooling, for example, MaxPooling or Average Pooling), fully connected layer and loss function (responsible for calculating errors between the current and the desired network output). There are many CNN structures, they can vary in the number of layers, shape and size of filters, activation functions, and other parameters. Examples

of very popular networks are AlexNet, GoogLeNet, and VGGNet (Arora et al., 2020; Mutasa et al., 2021).

The operation of CNN is broken down into several stages (Mutasa et al., 2021). First, filters allow the designation of a feature map. This is done by the convolution layer (Vinayakumar et al., 2017). It is a key component of CNN. The process is repeated several times to filter the feature maps obtained with the use of subsequent convolutional kernels. Characteristic parameters of the convolution layer are the number and size of filters in individual layers, the step by which the window corresponding to the filter is moved (Murata et al., 2018). The pooling layer is usually placed between two convolutional layers (Zhao and Wang, 2019). The layer performs the pooling operation on feature maps, i.e., the reduction of data size while maintaining the most important features. For this purpose, the data is divided into cells of equal size and a certain value is kept for each cell (maximal – Max Pooling, average – Average Pooling). The pooling layer has two main hyperparameters. These are the size of the cell into which we divide the data and the step by which individual cells will be separated from each other. The ReLU correction layer allows you to convert all negative values to zero. It comes as an activation function (Tan and Pan, 2019). The fully connected layer is often the last layer of CNN. A feature vector is fed as an input, which is transformed into a new vector using a linear combination and an activation function. The network output is compared with the training data set and the resulting loss, depending on its degree, causes the network weights to be updated using gradient and backpropagation. During the training of the neural network, this process is repeated many times to improve the quality of the model.

Selecting the correct CNN structure requires a lot of research. We focus on ensuring a compromise between operating time and the effectiveness of cleaning the signal from artifacts. Due to the one-dimensional input data (signal to be cleaned and reference signal from Fp1), one-dimensional filters were used. Two convolutional layers were created. In the first one, the number of filters was 20, and the kernel size was assumed to be 40. The filter shift step was set to 2. The second convolution layer contained 10 filters and the kernel size was set to 20. The shift step was 1. In both convolutional layers, the activation function was the ReLU. Next, a densely connected layer was added, which also defined the size of the output data (1-s window, 512 samples). The ADAM optimizer was used in the training process. **Table 1** shows the structure of the convolutional network. The network training set contained 70,000 1-s EEG/EOG signal windows, which were broken down into training data (80%) and validation data (20%).

Training the CNN required the determination of the number of epochs and examples that were fed to the input during subsequent iterations (batch size). The selected batch size was 128. This allowed for the use of less memory. Furthermore, more frequent updates of the network weights were performed, which accelerated training. The number of epochs used in CNN training was 10. We considered adding batch normalization layers, but it did not improve the performance of the network. Therefore, we decided to omit them. The network structure generated with the use of the *tensorflow* packet is shown in **Figure 9**.

The ADAM optimization algorithm was used in the learning process. The parameters selected during the training are summarized in **Table 2**. The chosen loss function was the mean square error.

With the use of *Learning rate* we can determine how much weights will be modified in subsequent training iterations (Yoo et al., 2019). Beta_1 and Beta_2 are hyperparameters used for first- and second-order moment estimation, respectively. Thanks to them, it is possible to correct the moments by removing the bias (Şen and Özkurt, 2020). The epsilon parameter is responsible

TABLE 1 | The structure of the proposed CNN.

Layer	Parameters
Input Layer	Input shape (512,2)
Convolutional_1D_1	20 filters 40 × 1 convolutions with stride 2 and padding same
Relu_1	ReLU
Convolutional_1D_2	10 filters 20 × 1 convolutions with stride 1 and padding same
Relu_2	ReLU
Flatten_3	Flatten layer – flattens the input to fully connected layer
Dense Output Layer	Output layer with desired clear signal shape (512)

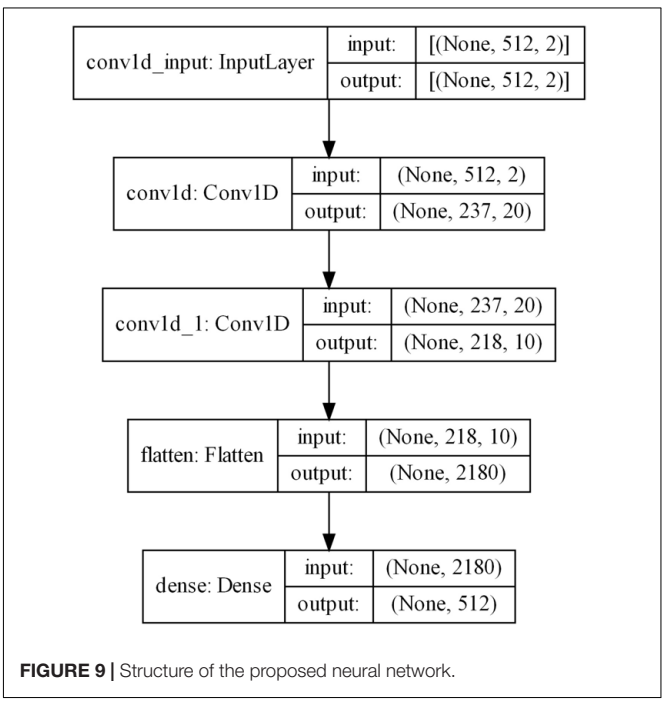

FIGURE 9 | Structure of the proposed neural network.

TABLE 2 | ADAM optimization function parameters.

Parameter	Value
Learning rate	0.001
Beta_1	0.9
Beta_2	0.999
Epsilon	1e-07

for preventing a possible division by zero when updating the weights. Therefore, very low epsilon values should be chosen in such a way as not to affect the result and, at the same time, to ensure no division by zero.

RESULTS AND DISCUSSION

To evaluate the effectiveness of the proposed CNN method for removing eye blink artifacts, comparisons were made with the ICA and regression methods. To be able to compare the methods, a set of EEG signals and a set of signals containing eye blink artifacts were generated. The pure EEG signal served as the reference signal. Then, ten 1-s windows containing EOG artifacts were generated. For each window, statistical coefficients (C_{kk}, C_{Fp1} *MAPE*, *RMSE,* and *Skewness)* were calculated, allowing a comparison of the effectiveness of artifact removal.

The C_{kk} is the correlation between the cleaned signal (with the use of one of the methods – CNN, ICA, and regression) and the original signal on the electrode k. The measure used is the Pearson correlation. The higher the absolute value of the C_{kk}, the better, because the signal after cleaning is closer to the real signal. C_{Fp1} is the correlation between the samples of the signal from the Fp1 (reference) electrode and the samples of the cleaned signal for a specific electrode. In general, it is better to keep the C_{Fp1} value as low as possible. *MAPE* determines the mean percentage error between the reference signal (EEG) and the one cleared by an algorithm. It is calculated as the arithmetic mean of the sum of the absolute values of the differences between the samples from the real signal and the cleaned signal, related to the real signal.

$$MAPE\left(y,\hat{y}\right) = \frac{1}{n_{samples}} \sum_{i=0}^{n_{samples}-1} \frac{|y_i - \widehat{y_i}|}{|y_i|}$$

The number of inputs is denoted by $n_{samples}$, y_i is the value for the i-th sample, and $\widehat{y_i}$ is the model predicted value for the i-th sample. *RMSE* is the root mean square error. It is calculated as the root of the arithmetic mean of the sum of the squares of the differences between the samples of the raw signal (EEG) and the signal cleaned by the given method.

$$MSE\left(y,\hat{y}\right) = \frac{1}{n_{samples}} \sum_{i=0}^{n_{samples}-1} \left(y_i - \widehat{y_i}\right)^2$$

$$RMSE\left(y,\hat{y}\right) = \sqrt{MSE\left(y,\hat{y}\right)}$$

MAPE and *MSE* errors should be kept as low as possible. The *Skewness* describes the skewness calculated for the cleaned signal using a given method. For normally distributed data (perfectly symmetric distribution), the skewness should be zero. If skewness is greater than zero, the largest number of data is on the left side of the curve representing the probability distribution. A skewness that is different from zero may indicate an existing eye blink artifact (Xiang et al., 2020).

We calculated C_{kk}, C_{Fp1}, *MAPE*, *RMSE,* and *Skewness* for all electrodes and all subjects. Detailed results are presented in **Supplementary Appendix 1**. The calculated values of the coefficients are plotted on the head surface (**Figures 10**–**14**). Such a representation allows for easier comparison of the results.

Figure 10 shows the C_{kk} value plotted on the head surface. A great similarity can be observed in terms of the distribution and values for the CNN and ICA methods. On the other hand, higher values of the C_{kk} coefficient occur for the REG method.

Figure 11 shows the values of C_{Fp1} plotted on the head surface. The lowest values of the coefficients are observed for the REG method. This is due to the principle of the REG method, that is, minimizing the correlation between individual electrodes and the Fp1 electrode (associated with eye blink artifacts). We can observe an increase in the C_{Fp1} value for the electrodes at the front of the head for the CNN method. On the other hand, lower values of C_{Fp1} can be observed for electrodes placed in the back of the head. For the ICA method, the distribution of the C_{Fp1} coefficient is more homogeneous. In this case we did not observe negative values of the C_{Fp1} coefficient.

Figure 12 shows the *Skewness* plotted on the surface of the head. The smallest disproportions of the coefficient values (close to zero) are observed for the ICA and REG methods. However, we can observe significant disproportions for the CNN method. For the CNN method, we can observe positive values of the skewness coefficient for the electrodes at the front of the head, while negative values for the electrodes at the back of the head, in particular for the electrodes O1, O2, and Oz.

Figures 13, **14** show the *RMSE* and *MAPE* errors. We can observe an increase in the values of the errors for the electrodes in the front of the head for the ICA and REG methods. Much lower values of the *RMSE* and *MAPE* errors can be observed for the CNN method, especially in the front part of the head. We observe lower values of the *MAPE* error for the entire head area for the CNN method compared to the ICA and REG methods. In our opinion, the *MAPE*/*RMSE* measure best describes the effectiveness of artifact removal as it relates to the reference signal.

To discuss in more detail the values obtained for C_{kk}, C_{Fp1}, *MAPE*, *RMSE* and *Skewness*, four electrodes were selected, located in the central, parietal, frontal, and occipital parts of the head: Cz, P3, F3, and Oz. The C_{kk}, C_{Fp1}, *MAPE*, *RMSE,* and *Skewness* values for electrodes Cz, P3, F3, and Oz are presented in **Tables 3**–**6**.

Table 3 presents the coefficients related to cleaning the signal from the Cz electrode. This electrode is located in the center of the head. In this case, very good results achieved by the CNN method can be observed. The correlation C_{kk} (0.93) is very high. The errors *MAPE* (0.805) and *RMSE* (2.935) have low values. The C_{Fp1} coefficient (−0.027) is low, which confirms the correct elimination of eye blink artifacts.

Table 4 shows the coefficients related to cleaning the signal from the P3 electrode. The electrode is located on the left side of the central part of the head. You can also notice very good removal of artifacts using the CNN method. The cleaned and real EEG signals are strongly, positively correlated – the C_{kk} coefficient is 0.869. The *MAPE* (1.219) and *RMSE* (4.381) errors for the CNN method are the lowest among the methods analyzed. The *Skewness* coefficient (−0.018) is also the smallest – it proves

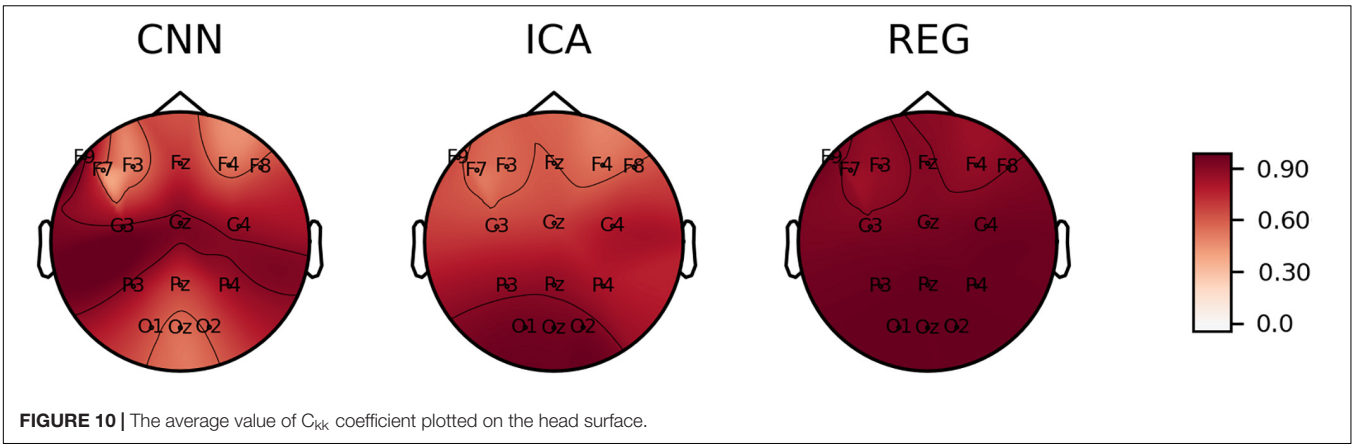

FIGURE 10 | The average value of C_{kk} coefficient plotted on the head surface.

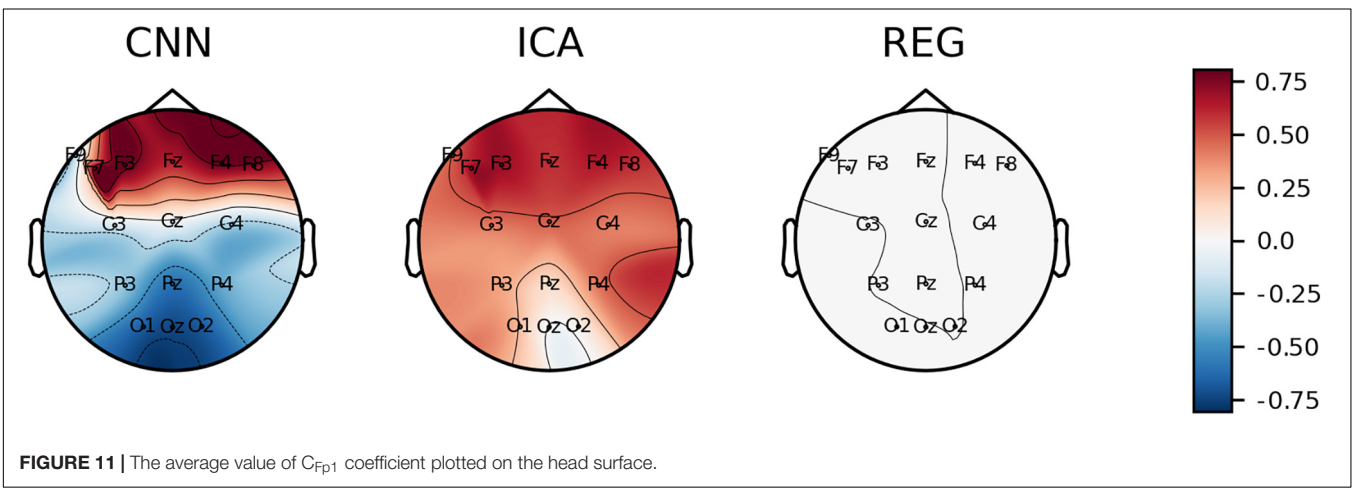

FIGURE 11 | The average value of C_{Fp1} coefficient plotted on the head surface.

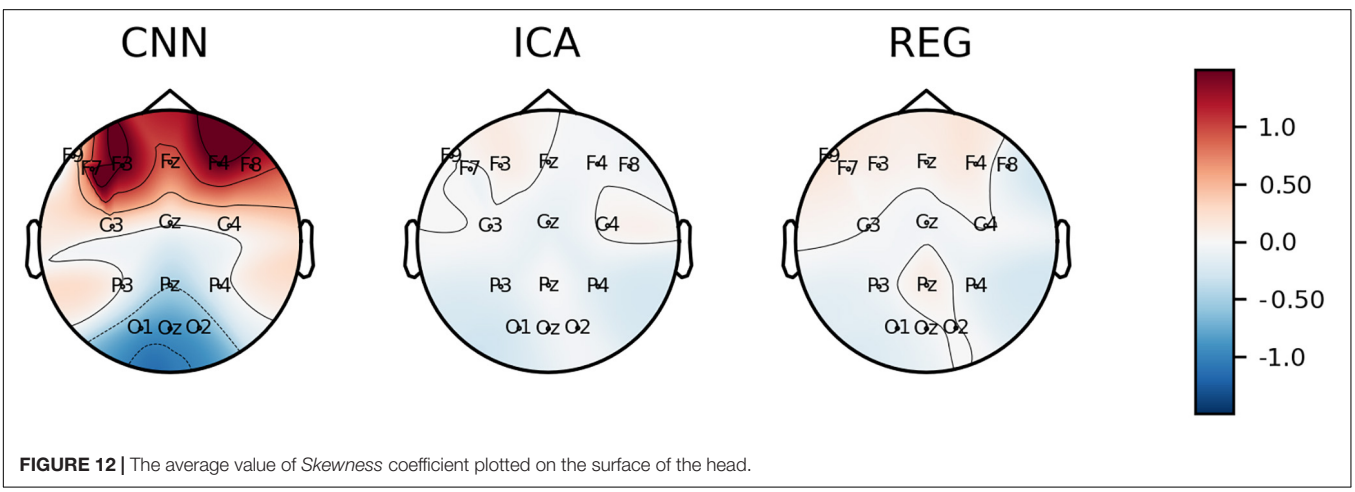

FIGURE 12 | The average value of *Skewness* coefficient plotted on the surface of the head.

that the distribution is even. The correlation with the Fp1 electrode is negative and reaches values close to the ICA method (C_{Fp1} equal to −0.321). In this case, the CNN method turned out to be comparable (and even better in terms of errors) to the regression method. Additionally, the ICA method introduced changes to the signal skewness, which is not desirable for proper signal cleaning.

Table 5 presents the coefficients related to the cleaning of the signals from the F3 electrode. This electrode is located in the left front of the head. In this case, the CNN cleaning results are comparable to those of ICA. The CNN method achieved significantly smaller *MAPE* (2.712) and *RMSE* (11.975) errors compared to the other methods. However, the obtained values of C_{kk} (0.508) and a relatively high positive correlation with

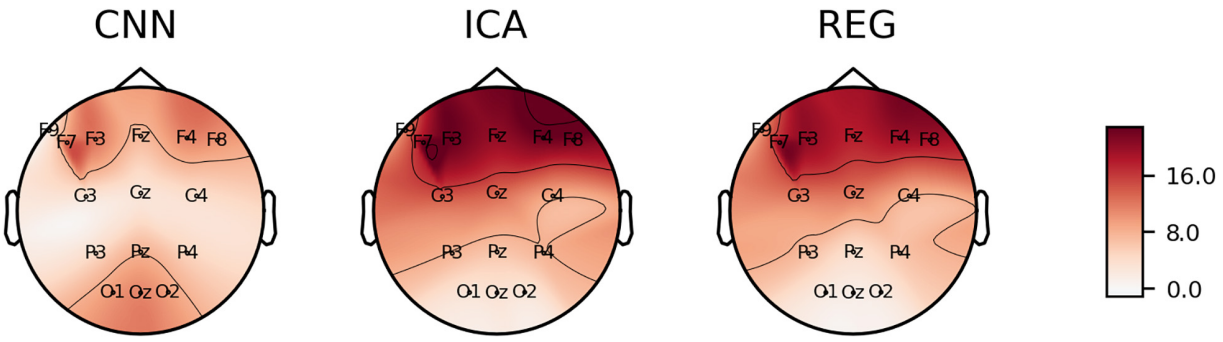

FIGURE 13 | The average value of *RMSE* coefficient plotted on the surface of the head.

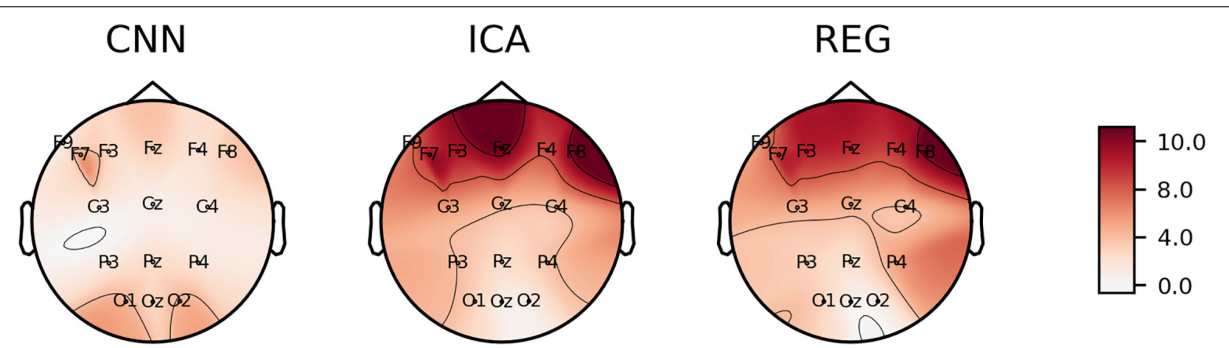

FIGURE 14 | The mean value of *MAPE* coefficient plotted on the head surface.

the F_{p1} reference electrode (C_{Fp1} equal to 0.790) indicate partial removal of artifacts. Furthermore, the *Skewness* index for the CNN method is high (1.499), which may indicate the existence of artifacts in the signal despite attempts to clean it.

Table 6 presents the average values of the coefficients for the Oz electrode. This electrode is located on the back of the subject's head. In this case, the advantage of the ICA and regression methods over CNN can be observed. The C_{kk} coefficient that describes the correlation between the cleaned and the original signal is much lower for CNN (0.556) than for the other methods (0.944 for ICA and 0.980 for regression). This means that there is a discrepancy between the cleaned signal and the original one. The other coefficients, *MAPE* equal to 2.810 and *RMSE* 10.978, are also high for the CNN method. The C_{Fp1} coefficient indicates a high content of artifacts in the cleaned signal – the correlation of the cleaned signals on individual electrodes with the F_{p1} reference electrode is high (−0.725). The *Skewness* for the CNN method (−0.875) also indicates a higher occurrence of artifacts than for the ICA (−0.073) and regression methods (−0.036).

TABLE 3 | The averaged values of C_{kk}, C_{Fp1}, *MAPE*, *RMSE,* and *Skewness* for the Cz electrode.

Method	C_{kk}	C_{Fp1}	*MAPE*	*RMSE*	*Skewness*
CNN	0.930	−0.027	0.805	2.935	0.037
ICA	0.692	0.481	4.485	13.140	−0.113
REG	0.934	<0.001	4.795	12.145	−0.051

TABLE 4 | The averaged values of C_{kk}, C_{Fp1}, *MAPE*, *RMSE,* and *Skewness* for the P3 electrode.

Method	C_{kk}	C_{Fp1}	*MAPE*	*RMSE*	*Skewness*
CNN	0.869	−0.321	1.219	4.381	−0.018
ICA	0.872	0.313	3.952	7.763	−0.157
REG	0.974	<0.001	3.010	6.832	−0.079

TABLE 5 | The averaged values of C_{kk}, C_{Fp1}, *MAPE*, *RMSE,* and *Skewness* for the F3 electrode.

Method	C_{kk}	C_{Fp1}	*MAPE*	*RMSE*	*Skewness*
CNN	0.508	0.790	2.712	11.975	1.499
ICA	0.567	0.685	10.650	22.356	0.092
REG	0.872	<0.001	10.115	19.954	0.084

TABLE 6 | The averaged values of *Ckk*, C_{Fp1}, *MAPE*, *RMSE*, and *Skewness* for the Oz electrode.

Method	C_{kk}	C_{Fp1}	*MAPE*	*RMSE*	*Skewness*
CNN	0.556	−0.725	2.810	10.978	−0.875
ICA	0.944	−0.006	1.062	3.402	−0.073
REG	0.980	<0.001	0.645	1.719	−0.036

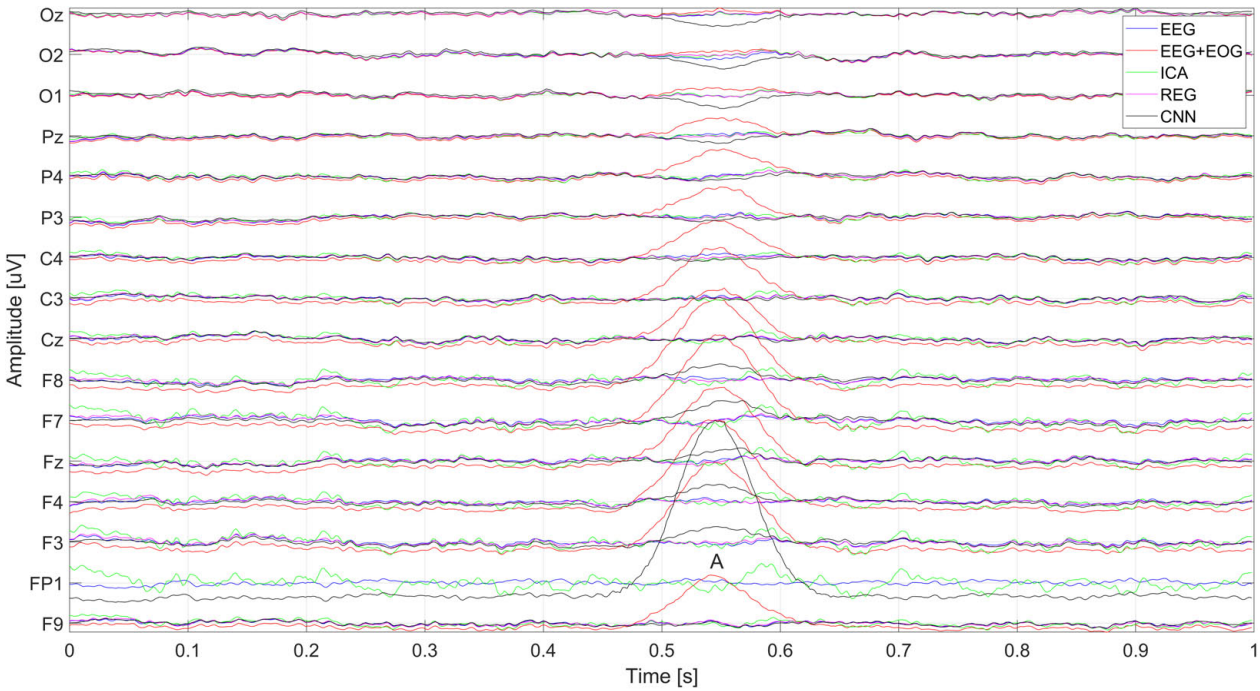

FIGURE 15 | Artificially generated 1-s window of pure EEG, EEG + EOG, and cleaned signal using ICA, regression, CNN methods. The letter A represents the moment the blink artifact occurred.

Analyzing **Tables 3–6**, it can be seen that for the CNN method, the C_{kk} coefficient is high for the Cz (0.93) and P3 (0.869) electrodes, which means that the signals are cleaned properly. A high correlation indicates a strong similarity between the cleaned signal and the original. The ICA and CNN methods are distinguished by relatively low *MAPE* and *RMSE* errors (for 3 out of 4 electrodes CNN achieves much lower errors). The cleaned signals are strongly correlated with the original, and the C_{kk} coefficients are high (for the Cz electrode −0.93, for the P3 electrode −0.869). Furthermore, cleaned signals are poorly correlated with the Fp1 reference electrode.

Figure 15 shows the artificially generated 1-s window (512 samples) of EEG, EEG + EOG, and cleaned signals using each of the tested methods – CNN, ICA, and regression. In the case of generated signals, the CNN changes the polarization of the Oz, O2, and O1 electrodes at the place of the artifacts (marked A in **Figure 15**) – this is not the desired phenomenon. **Figure 5** shows the differences in the cleaned signals obtained with the use of the tested methods. You can observe discrepancies in the cleaned signal using the ICA method in relation to the others, for example, electrodes F7, F8, and Fz.

Figure 16 shows the spectrum of the signal from the Cz electrode (presented in **Figure 15**). EOG signals (Banerjee et al., 2013) are in the range of 0.1–20 Hz. In **Figure 16**, it can be seen that all the methods eliminated low-frequency amplitudes. The CNN-based method performed very well. The spectrum of the cleaned signal is closest to the original one. It should be noted that the ICA method introduced a significant distortion of the spectrum for 5–10 Hz.

Figure 17 shows a close-up of the signal from the Cz electrode. There is a noticeable difference in the operation of ICA and other methods visible in the times A, B, and C marked in **Figure 17**. Changes in the signals for A and C are caused by the presence of a constant component – in many cases of EEG signal analysis, it does not matter. It can also be seen that the highest coverage of signals with the original EEG (correct cleaning) is in the case of the regression method and CNN. In the part marked B, we can observe a significant modification of the signal using the ICA method.

Figure 18 shows the cleaning effect on the real EEG signal (user S03). It can be seen that artifacts from the real EEG signal are correctly removed. The ICA method, as described above, also cleans the signal on the reference electrode. It can be seen that the removal of artifacts from the F_{p1} reference electrode with ICA is much worse than the cleaning of signals on other electrodes.

Figure 19 shows the spectrum of the signal from the Cz electrode (presented in **Figure 18**). There is a visible decrease in the amplitudes of successive bands of the spectrum in the low-frequency range, which indicates the correct operation of the methods used to eliminate eye blink artifacts. It can be seen that all the methods allow us to obtain a similar spectrum.

Figure 20 shows a close-up of the signals recorded on electrode Cz (presented in **Figure 18**). The figure shows two eye blink artifacts labeled A and B. The first was correctly removed with each of the analyzed methods. In the case of B, it can be seen that the artifact removal using the ICA method was not complete. Much better results were obtained using CNN and the regression method.

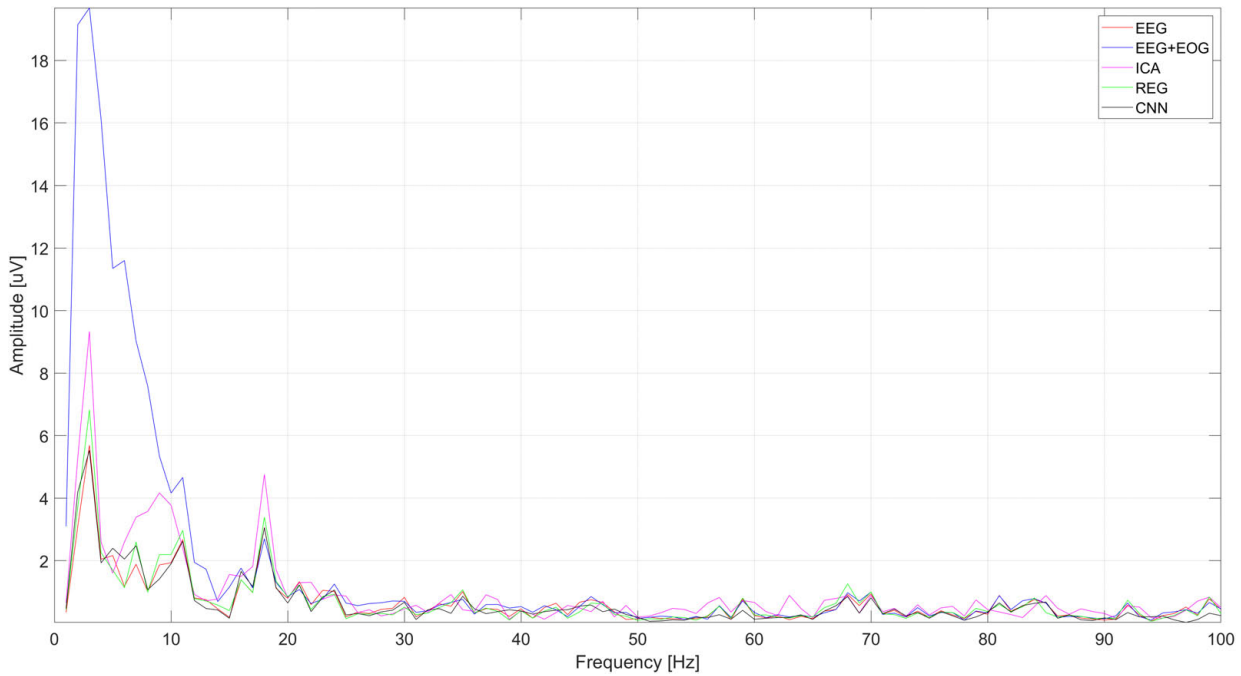

FIGURE 16 | The spectrum of the signal from the Cz electrode.

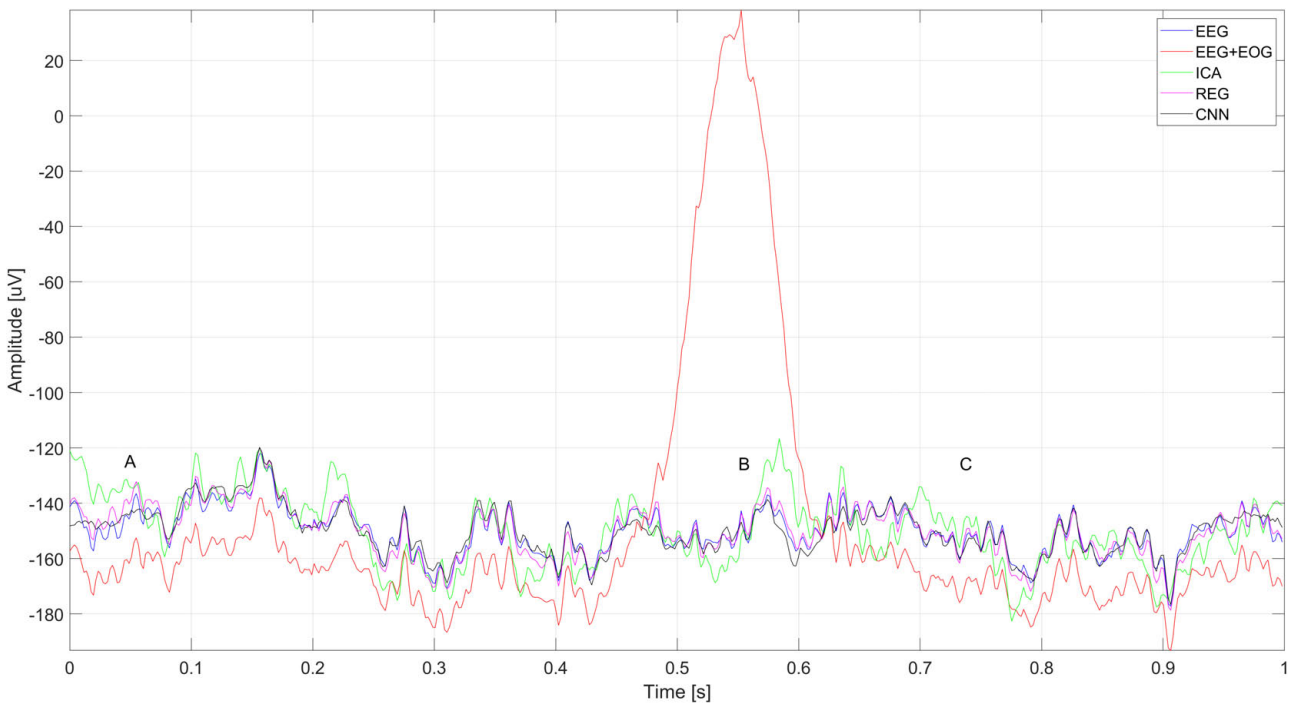

FIGURE 17 | A fragment of the simulated EEG signal for the Cz electrode. The letters A–C represent selected moments: A and C the EEG signal fragment without artifacts and B with an eye blink artifact.

The above discussion shows that the application of the CNN method gives very good results in the removal of eye blink artifacts, in particular for the electrodes placed in the central part of the head. Therefore, the application of the proposed method may be useful as a pre-processing in the analysis of the P300 potential or other event-related potential (ERP) occurring in the

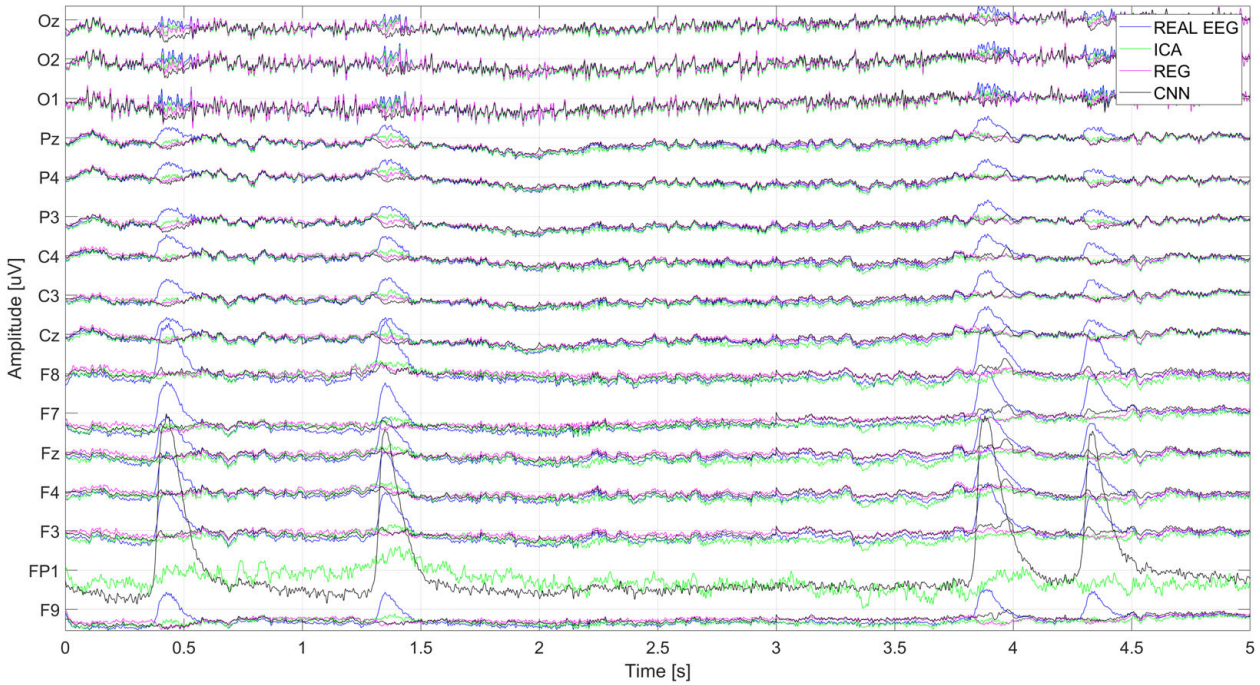

FIGURE 18 | The real EEG signal fragment recorded during the test and cleaned by ICA, regression, CNN method.

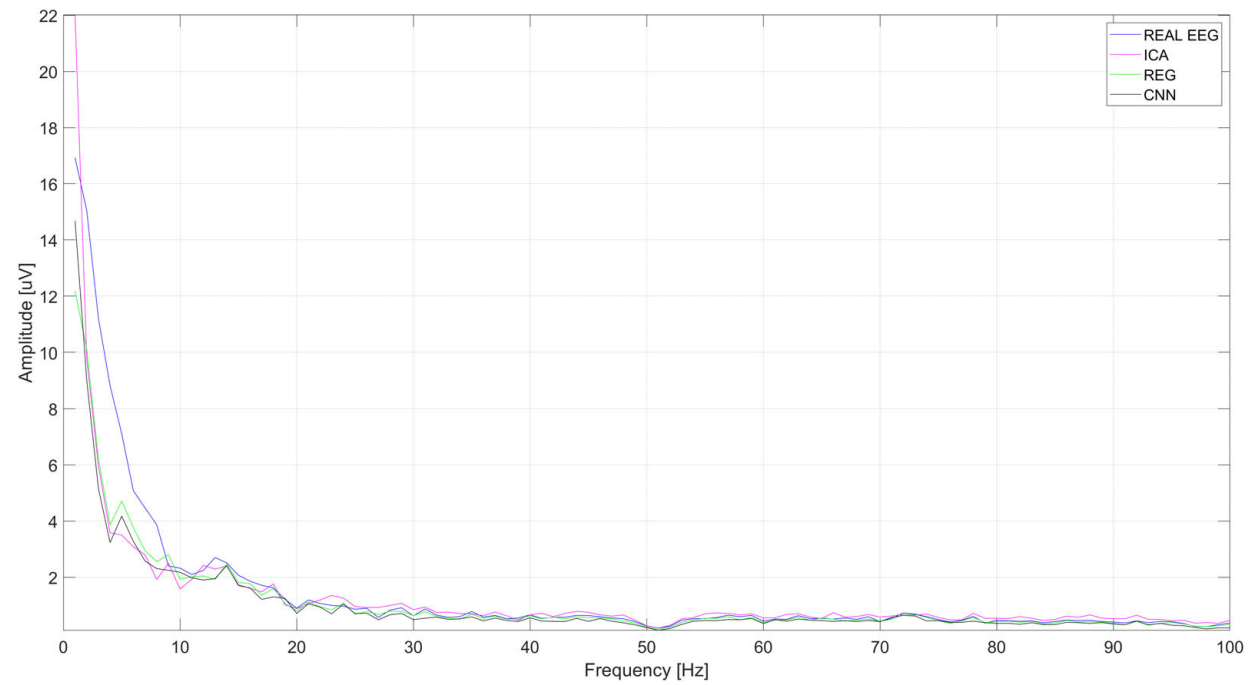

FIGURE 19 | An example of the spectrum of the signal from the Cz electrode.

central part of the head. To verify the usefulness of the method to eliminate eye blink artifacts, we cleared the EEG signals from the Cz electrode for the signals registered during the experiments with the N-back task.

Table 7 presents the signal statistics – parameters describing the real signal and signals after artifact removal for the Cz electrode. Average values for standard deviation and peak-to-peak values are shown for all 20 users. The results obtained

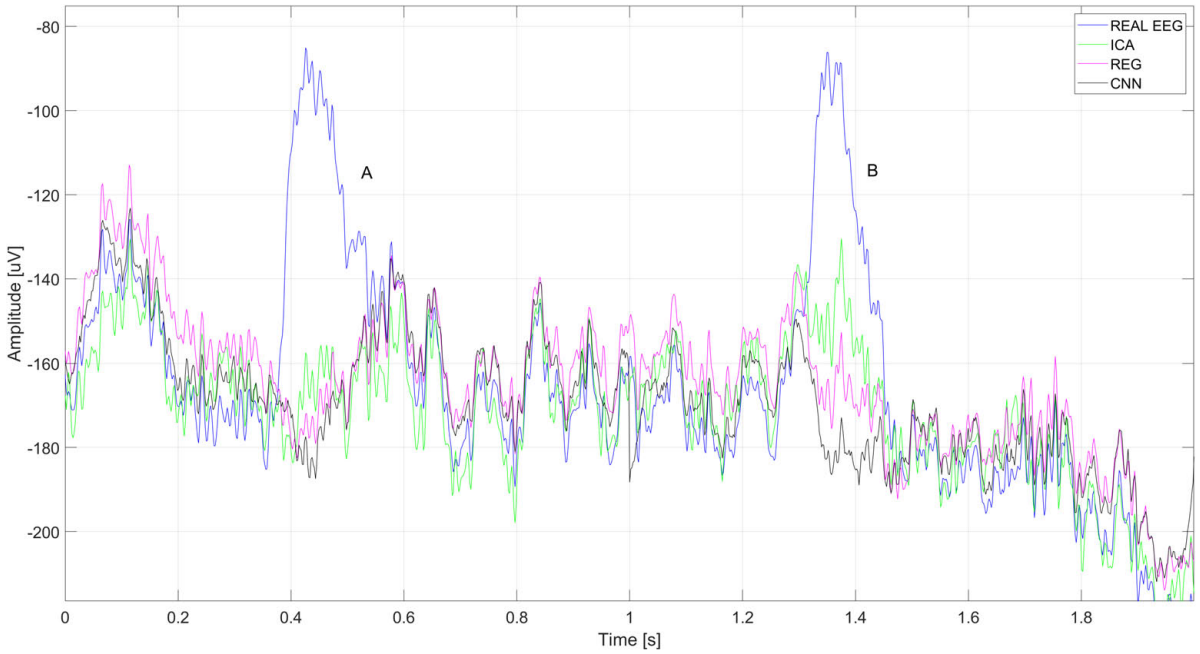

FIGURE 20 | A fragment of the EEG signal recorded for the Cz electrode. The letters A and B represent the times when the blink artifact occurred.

indicate that the artifacts are correctly removed. The peak-to-peak value for the tested methods is lower than that for the raw signal containing the artifacts. The peak-to-peak value of the signal before cleaning is 141.76 μV. After cleaning, it decreases for each method (CNN – 88.8 μV, ICA – 93.73 μV, regression – 101 μV). These indicate a good performance of the CNN and ICA methods and slightly worse for the regression method. In addition, a reduction in the standard deviation can be seen for each method. This indicates a reduction in the scattering of samples in the cleaned signal compared to the raw signal. The decrease is most noticeable for the CNN method (from 25.89 μV to 15.197 μV) and regression (from 25.89 μV to 16.58 μV).

Eye blink artifacts produce much larger amplitudes than potentials of interest in the EEG signal. This is especially true for ERP. During the N-back task, the users watched the computer monitor. Stimuli that are presented for a long time can cause discomfort in the examined person and force the eyes to blink. This is a natural activity. It happens that such blinks provoked by the presented stimuli can be easily mistaken for the desired potentials. Such an example can be observed in the case of recorded signals. For user S03, about 0.4 s after the stimulus presented, blinking of the eyes occurs very frequently and regularly. It is observable on the FP1 electrode but also on Cz, where we would expect, e.g., the P300 potential. **Figure 21** shows an example of averaged ERP after the N-back stimulus. ERP without filtration (real) is shown in blue, orange – after removing artifacts using CNN, green – after removing artifacts using the ICA method, and red – after removing artifacts using regression. Even averaging, which is standard in this type of analysis, does not eliminate the problem of repetitive artifacts. This may result in incorrect interpretation of potentials.

Next, we check the operating times of the CNN, ICA, and regression algorithms implemented. We used the real signal (S03_EEG), fragments of various lengths were selected – 10 s, 60 s, 10 min, 30 min, and 50 min. The operation of the methods

TABLE 7 | Statistical parameters describing 3 s of the real and cleaned signals (using CNN, ICA, and regression methods) for the Cz electrode.

Method	Standard deviation	Peak-peak
Real EEG	25.89	141.76
CNN	15.197	88.80
ICA	18.22	93.73
REG	16.58	101.00

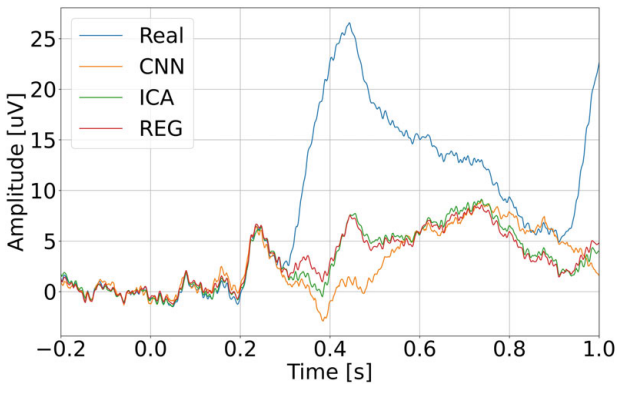

FIGURE 21 | Averaged ERP potential for the user S03 – electrode Cz.

TABLE 8 | Real signal cleaning times using CNN, ICA, and regression.

Duration	CNN (s)	ICA (s)	REG (s)
10 s	5.29	0.437	0.043
60 s	32.414	0.662	0.0799
10 min	335.13	5.33	0.489
30 min	984.10	18.681	1.373
50 min	1578.91	19.690	2.176

was tested using a computer equipped with an Intel Core i7-9750H 2.60 GHz processor, 32 GB RAM, and a GeForce GTX 1660 Ti graphics card with 6 GB GDDR6 memory. **Table 8** shows the operating times of each method needed to clean EEG signals of various lengths.

According to the data in **Table 8**, it can be seen that the CNN method is the slowest method. It takes about 26 min to clean 1 h of an EEG signal with 16 channels. The fastest method is regression – for a signal lasting 50 min, the cleaning lasted 2.176 s. The time differences are due to the computational complexity of the individual methods. Despite long training and long operating time, the CNN method gave very good results in cleaning the signals from the electrodes located in the center and slightly on the back of the head. The *RMSE* and *MAPE* errors for these electrodes are much lower than those obtained when using other methods. In the case of analysis of real signals, the CNN method does not introduce distortion into the cleaned signal, which shows its advantage over the ICA method.

It should be noted that in experiments we used a signal database recorded previously with a fixed sampling frequency (fs = 512 Hz). The trained CNN for set conditions cannot be used for differently recorded EEG signals. Changing the sampling rate or changing the amplifier has to be associated with retraining the CNN. However, the results presented show that it is a promising method of artifact removal. In future experiments, the authors plan to record EEG signals, EOG signals, and muscle activity. The network could then be trained not only to remove EB-type artifacts, but also artifacts related to the movement of the eye, facial muscles, and neck. Future research should also include optimization of the number of samples of the EEG signal fed the CNN. Currently, the number of samples is 512. This number of samples is somewhat of a compromise between the signal time, which may contain a blink pattern, and the number of samples at the input of the network. Too many samples make it difficult to train the network, but too few samples could not take into account the shape of the eye blink.

CONCLUSION

Experiments have shown that the use of CNN method gives better results in the task of removing eye blink artifacts than regression methods or independent component analysis. The mean value of the *MAPE* error for the CNN method was 4.69, for the ICA method it was 7.84, and for the REG method it was 7.76. The CNN method better removes eye blink artifacts, especially in the central and parietal parts of the head. An example can be the electrode Cz. In that case, for the CNN method, errors such as *MAPE* (0.805) and *RMSE* (2.935) are much lower than for ICA (*MAPE* = 4.485 *RMSE* = 13.140) and regression (*MAPE* = 4.795 *RMSE* = 12.145). Furthermore, visual inspection showed that the ICA method introduces distortion in the shape of the EEG signal. No such changes were observed for the regression method and CNN. On the other hand, better artifact removal results were obtained for ICA and regression methods when it comes to electrodes placed in the occipital area of the head (O1, O2, and Oz). In this case, the use of the CNN method is questionable. It should be noted that the CNN method is much better suited for offline removal of artifacts than online removal. This is because we need to have a set of signals that are needed to train the network. In addition, we need to train the network. The time required on the CNN method to work on short EEG signals is acceptable (a few minutes). For EEG signals that last several hours, the analysis may be too time-consuming. Further research should also consider other CNN neural network structures and training the network using more examples and types of artifacts.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/**Supplementary Material**, further inquiries can be directed to the corresponding author/s.

ETHICS STATEMENT

Ethical approval was not provided for this study on human participants because the studies on volunteers were non-interventional. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

MJ was responsible for the implementation of deep learning methods (CNN), program generating EEG/EOG signals, and software comparing the operation of the methods, and preparation of the text of the article – introduction and results. MK was responsible for recording the EEG signal during the presentation of the N-back task for 20 people, research concept, developing the CNN method to remove artifacts, and developing a methodology for comparing the operation of the methods. AM was responsible for the preparation of the software for the presentation of the N-back task and EEG signal registration, substantive evaluation of the results, editing the summary of the article, and developing a method for comparing artifacts removal algorithms: CNN, ICA, and regression. All authors contributed to the article and approved the submitted version.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fnins.2022.782367/full#supplementary-material

REFERENCES

Alquran, H., Alqudah, A., Abuqasmieh, I., and Almashaqbeh, S. (2019). "Gaussian model of electrooculogram signals," in *Proceeding of the 2019 IEEE Jordan International Joint Conference on Electrical Engineering and Information Technology (JEEIT)*, doi: 10.1109/JEEIT.2019.8717499

Anderer, P., Roberts, S., Schlögl, A., Gruber, G., Klösch, G., Herrmann, W., et al. (1999). Artifact processing in computerized analysis of sleep EEG – a review. *Neuropsychobiology* 40, 150–157. doi: 10.1159/000026613

Arora, D., Garg, M., and Gupta, M. (2020). "Diving deep in deep convolutional neural network," in *Proceeding of the 2020 2nd International Conference on Advances in Computing, Communication Control and Networking (ICACCCN)*, 749–751. doi: 10.1109/ICACCCN51052.2020.9362907

Banerjee, A., Datta, S., Pal, M., Konar, A., Tibarewala, D. N., and Janarthanan, R. (2013). Classifying electrooculogram to detect directional eye movements. *Proc. Technol.* 10, 67–75. doi: 10.1016/j.protcy.2013.12.338

Bao, G., Yan, B., Tong, L., Shu, J., Wang, L., Yang, K., et al. (2021). Data augmentation for EEG-based emotion recognition using generative adversarial networks. *Front. Comput. Neurosci.* 15:115. doi: 10.3389/fncom.2021.723843

Browarska, N., Zygarlicki, J., Pelc, M., Niemczynowicz, M., Zygarlicka, M., and Kawala-Sterniuk, A. (2021). "Pilot study on using innovative counting peaks method for assessment purposes of the EEG data recorded from a single-channel non-invasive brain-computer interface," in *Proceeding of the 2021 25th International Conference on Methods and Models in Automation and Robotics (MMAR)* (Miêdzyzdroje, Poland: IEEE), 68–72. doi: 10.1109/MMAR49549.2021.9528447

Chen, G., Chen, Y., Yuan, Z., Lu, X., Zhu, X., and Li, W. (2019). "Breast Cancer Image Classification based on CNN and Bit-Plane slicing," in *Proceeding of the 2019 International Conference on Medical Imaging Physics and Engineering (ICMIPE)*, 1–4. doi: 10.1109/ICMIPE47306.2019.9098216

Cheng, J., Li, L., Li, C., Liu, Y., Liu, A., Qian, R., et al. (2019). Remove diverse artifacts simultaneously from a single-channel EEG based on SSA and ICA: a semi-simulated study. *IEEE Access* 7:2019. doi: 10.1109/ACCESS.2019.2915564

Garg, P., Davenport, E., Murugesan, G., Wagner, B., Whitlow, C., Maldjian, J., et al. (2017). "Using convolutional neural networks to automatically detect eye-blink artifacts in magnetoencephalography without resorting to electrooculography," in *Medical Image Computing and Computer Assisted Intervention - MICCAI 2017 Lecture Notes in Computer Science*, eds M. Descoteaux, L. Maier-Hein, A. Franz, P. Jannin, D. L. Collins, and S. Duchesne (Cham: Springer International Publishing), 374–381. doi: 10.1007/978-3-319-66179-7_43

Hasasneh, A., Kampel, N., Sripad, P., Shah, N. J., and Dammers, J. (2018). Deep Learning Approach for Automatic Classification of Ocular and Cardiac Artifacts in MEG Data. *J. Eng.* 2018, 1–10. doi: 10.1155/2018/1350692

Iaquinta, A. F., Silva, A. C., de, S., Júnior, A. F., de Toledo, J. M., and von Atzingen, G. V. (2021). EEG multipurpose eye blink detector using convolutional neural network. *ArXiv* [preprint]. ArXiv210714235.

Isar, D., and Gajitzki, P. (2016). "Pink noise generation using wavelets," in *Proceeding of the 2016 12th IEEE International Symposium on Electronics and Telecommunications (ISETC)*, 261–264. doi: 10.1109/ISETC.2016.7781107

Jiang, X., Bian, G.-B., and Tian, Z. (2019). Removal of artifacts from EEG signals: a review. *Sensors* 19:987. doi: 10.3390/s19050987

Kawala-Sterniuk, A., Podpora, M., Pelc, M., Blaszczyszyn, M., Gorzelanczyk, E. J., Martinek, R., et al. (2020). Comparison of smoothing filters in analysis of EEG data for the medical diagnostics purposes. *Sensors* 20:807. doi: 10.3390/s20030807

Khatwani, M., Hosseini, M., Paneliya, H., Mohsenin, T., Hairston, W. D., and Waytowich, N. (2018). "Energy efficient convolutional neural networks for EEG artifact detection," in *Proceeding of the 2018 IEEE Biomedical Circuits and Systems Conference (BioCAS)*, 1–4. doi: 10.1109/BIOCAS.2018.8584791

Kilicarslan, A., and Contreras-Vidal, J. L. (2017). "Full characterization and removal of motion artifacts from scalp EEG recordings," in *Proceeding of the 2017 International Symposium on Wearable Robotics and Rehabilitation (WeRob)*, 1–1. doi: 10.1109/WEROB.2017.8383881

Kołodziej, M., Tarnowski, P., Sawicki, D. J., Majkowski, A., Rak, R. J., Bala, A., et al. (2020). Fatigue detection caused by office work with the use of EOG signal. *IEEE Sens. J.* 20, 15213–15223. doi: 10.1109/JSEN.2020.3012404

Lashgari, E., Liang, D., and Maoz, U. (2020). Data augmentation for deep-learning-based electroencephalography. *J. Neurosci. Methods* 346:108885. doi: 10.1016/j.jneumeth.2020.108885

Levin, A. R., Méndez Leal, A. S., Gabard-Durnam, L. J., and O'Leary, H. M. (2018). BEAPP: the batch electroencephalography automated processing platform. *Front. Neurosci.* 12:513. doi: 10.3389/fnins.2018.00513

Li, P., Chen, Z., and Hu, Y. (2017). "A method for automatic removal of EOG artifacts from EEG based on ICA-EMD," in *Proceeding of the 2017 Chinese Automation Congress (CAC)*, 1860–1863. doi: 10.1109/CAC.2017.8243071

Li, X., La, R., Wang, Y., Hu, B., and Zhang, X. (2020). A deep learning approach for mild depression recognition based on functional connectivity using electroencephalography. *Front. Neurosci.* 14:192. doi: 10.3389/fnins.2020.00192

Lou, G., and Shi, H. (2020). Face image recognition based on convolutional neural network. *China Commun.* 17, 117–124. doi: 10.23919/JCC.2020.02.010

Luo, Y., Zhu, L.-Z., Wan, Z.-Y., and Lu, B.-L. (2020). Data augmentation for enhancing EEG-based emotion recognition with deep generative models. *ArXiv* [preprint]. ArXiv200605331, doi: 10.1088/1741-2552/abb580

Mannan, M. M. N., Kamran, M. A., and Jeong, M. Y. (2018). Identification and removal of physiological artifacts from electroencephalogram signals: a review. *IEEE Access* 6, 30630–30652. doi: 10.1109/ACCESS.2018.2842082

McFarland, D. J., McCane, L. M., David, S. V., and Wolpaw, J. R. (1997). Spatial filter selection for EEG-based communication. *Electroencephalogr. Clin. Neurophysiol.* 103, 386–394. doi: 10.1016/S0013-4694(97)00022-2

Mumtaz, W., Rasheed, S., and Irfan, A. (2021). Review of challenges associated with the EEG artifact removal methods. *Biomed. Signal Process. Control* 68:102741. doi: 10.1016/j.bspc.2021.102741

Murata, K., Mito, M., Eguchi, D., Mori, Y., and Toyonaga, M. (2018). "A Single Filter CNN Performance for Basic Shape Classification," in *Proceedings of the 2018 9th International Conference on Awareness Science and Technology (iCAST)* (Fukuoka: IEEE), 139–143. doi: 10.1109/ICAwST.2018.8517219

Mutasa, S., Sun, S., and Ha, R. (2021). Understanding artificial intelligence based radiology studies: CNN architecture. *Clin. Imaging* 80, 72–76. doi: 10.1016/j.clinimag.2021.06.033

Natesan, P., Priya, V. V., and Gothai, E. (2020). "Classification of multi-lead ECG signals to predict myocardial infarction using CNN," in *Proceeding of the 2020 Fourth International Conference on Computing Methodologies and Communication (ICCMC)*, 1029–1033. doi: 10.1109/ICCMC48092.2020.ICCMC-000192

Nejedly, P., Cimbalnik, J., Klimes, P., Plesinger, F., Halamek, J., Kremen, V., et al. (2019). Intracerebral EEG artifact identification using convolutional neural networks. *Neuroinformatics* 17, 225–234. doi: 10.1007/s12021-018-9397-6

Pham, V. T., Dang, N. D., Nguyen, D. A., Nguyen, T. A., Chu, D. H., and Tran, D.-T. (2017). "Automatic removal of EOG artifacts using SOBI algorithm combined with intelligent source identification technique," in *Proceeding of the 2017 International Conference on Advanced Technologies for Communications (ATC)*, 260–264. doi: 10.1109/ATC.2017.8167629

Placidi, G., Cinque, L., and Polsinelli, M. (2021). Convolutional neural networks for automatic detection of artifacts from independent components represented in scalp topographies of EEG signals. *Comput. Biol. Med.* 132:104347. doi: 10.1016/j.compbiomed.2021.104347

Rakibul Mowla, M., Ng, S.-C., Zilany, M. S. A., and Paramesran, R. (2015). Artifacts-matched blind source separation and wavelet transform for multichannel EEG denoising. *Biomed. Signal Process. Control* 22, 111–118. doi: 10.1016/j.bspc.2015.06.009

Ranjan, R., Chandra Sahana, B., and Kumar Bhandari, A. (2021). Ocular artifact elimination from electroencephalography signals: a systematic review. *Biocybern. Biomed. Eng.* 41, 960–996. doi: 10.1016/j.bbe.2021.06.007

Sakai, A., Minoda, Y., and Morikawa, K. (2017). "Data augmentation methods for machine-learning-based classification of bio-signals," in *Proceeding of the 2017 10th Biomedical Engineering International Conference (BMEiCON)*, 1–4. doi: 10.1109/BMEiCON.2017.8229109

Salimi, N., Barlow, M., and Lakshika, E. (2020). "Towards potential of N-back task as protocol and EEG net for the EEG-based biometric," in *Proceeding of the 2020 IEEE Symposium Series on Computational Intelligence (SSCI)*, 1718–1724. doi: 10.1109/SSCI47803.2020.9308487

Şen, S. Y., and Özkurt, N. (2020). "Convolutional neural network hyperparameter tuning with adam optimizer for ECG classification," in *Proceeding of the 2020 Innovations in Intelligent Systems and Applications Conference (ASYU)*, 1–6. doi: 10.1109/ASYU50717.2020.9259896

Sun, W., Su, Y., Wu, X., and Wu, X. (2020). A novel end-to-end 1D-ResCNN model to remove artifact from EEG signals. *Neurocomputing* 404, 108–121. doi: 10.1016/j.neucom.2020.04.029

Tan, Z., and Pan, P. (2019). "Network fault prediction based on CNN-LSTM hybrid neural network," in *Proceeding of the 2019 International Conference on Communications, Information System and Computer Engineering (CISCE)* (Haikou: IEEE), 486–490. doi: 10.1109/CISCE.2019.00113

Tosun, M., and Kasım, Ö (2020). *Novel Eye-Blink Artefact Detection Algorithm from Raw EEG Signals Using FCN-Based Semantic Segmentation Method - Tosun - 2020 - IET Signal Processing - Wiley Online Library*. Available online at: https://ietresearch.onlinelibrary.wiley.com/doi/10.1049/iet-spr.2019.0602 (accessed December 21, 2021).

Urigüen, J., and Zapirain, B. (2015). EEG artifact removal – state-of-the-art and guidelines. *J. Neural Eng.* 12:031001. doi: 10.1088/1741-2560/12/3/031001

Vinayakumar, R., Soman, K. P., and Poornachandran, P. (2017). "Applying convolutional neural network for network intrusion detection," in *Proceeding of the 2017 International Conference on Advances in Computing, Communications and Informatics (ICACCI)* (Udupi: IEEE), 1222–1228. doi: 10.1109/ICACCI.2017.8126009

Voytek, B., Kramer, M. A., Case, J., Lepage, K. Q., Tempesta, Z. R., Knight, R. T., et al. (2015). Age-related changes in 1/f neural electrophysiological noise. *J. Neurosci.* 35, 13257–13265. doi: 10.1523/JNEUROSCI.2332-14.2015

Wang, F., Zhong, S., Peng, J., Jiang, J., and Liu, Y. (2018). "Data augmentation for EEG-based emotion recognition with deep convolutional neural networks," in *MultiMedia Modeling Lecture Notes in Computer Science*, eds K. Schoeffmann, T. H. Chalidabhongse, C. W. Ngo, S. Aramvith, N. E. O'Connor, Y.-S. Ho, et al. (Cham: Springer International Publishing), 82–93. doi: 10.1007/978-3-319-73600-6_8

Xiang, J., Maue, E., Fan, Y., Qi, L., Mangano, F. T., Greiner, H., et al. (2020). Kurtosis and skewness of high-frequency brain signals are altered in paediatric epilepsy. *Brain Commun.* 2:fcaa036. doi: 10.1093/braincomms/fcaa036

Yang, B., Duan, K., Fan, C., Hu, C., and Wang, J. (2018). Automatic ocular artifacts removal in EEG using deep learning. *Biomed. Signal Process. Control* 43, 148–158. doi: 10.1016/j.bspc.2018.02.021

Yoo, J.-H., Yoon, H., Kim, H.-G., Yoon, H.-S., and Han, S.-S. (2019). "Optimization of Hyper-parameter for CNN Model using Genetic Algorithm", in *Proceedings of the 2019 1st International Conference on Electrical, Control and Instrumentation Engineering (ICECIE)* (Kuala Lumpur, Malaysia: IEEE), 1–6. doi: 10.1109/ICECIE47765.2019.8974762

Zhang, C., Lian, Y., and Wang, G. (2020). "ARDER: an automatic EEG artifacts detection and removal system," in *Proceeding of the 2020 27th IEEE International Conference on Electronics, Circuits and Systems (ICECS)*, 1–2. doi: 10.1109/ICECS49266.2020.9294865

Zhang, Y., Chen, J., Tan, J. H., Chen, Y., Chen, Y., Li, D., et al. (2020). An investigation of deep learning models for EEG-based emotion recognition. *Front. Neurosci.* 14:1344. doi: 10.3389/fnins.2020.622759

Zhao, J., and Wang, Z. (2019). "Study on the influence of the distribution of convolution kernels," in *Proceeding of the 2019 IEEE 9th International Conference on Electronics Information and Emergency Communication (ICEIEC)* (Beijing: IEEE), 1–4. doi: 10.1109/ICEIEC.2019.8784457

Conflict of Interest: The 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.

Advantages of publishing in Frontiers

OPEN ACCESS
Articles are free to read for greatest visibility and readership

FAST PUBLICATION
Around 90 days from submission to decision

HIGH QUALITY PEER-REVIEW
Rigorous, collaborative, and constructive peer-review

TRANSPARENT PEER-REVIEW
Editors and reviewers acknowledged by name on published articles

Frontiers
Avenue du Tribunal-Fédéral 34
1005 Lausanne | Switzerland

Visit us: www.frontiersin.org
Contact us: frontiersin.org/about/contact

REPRODUCIBILITY OF RESEARCH
Support open data and methods to enhance research reproducibility

DIGITAL PUBLISHING
Articles designed for optimal readership across devices

FOLLOW US
@frontiersin

IMPACT METRICS
Advanced article metrics track visibility across digital media

EXTENSIVE PROMOTION
Marketing and promotion of impactful research

LOOP RESEARCH NETWORK
Our network increases your article's readership