

## Article

# Resting-State EEG Signal for Major Depressive Disorder Detection: A Systematic Validation on a Large and Diverse Dataset

Chien-Te Wu <sup>1,†</sup>, Hao-Chuan Huang <sup>2,†</sup>, Shiuan Huang <sup>2,†</sup>, I-Ming Chen <sup>3,4</sup>, Shih-Cheng Liao <sup>3</sup>, Chih-Ken Chen <sup>5,6</sup>, Chemin Lin <sup>5,6</sup>, Shwu-Hua Lee <sup>6,7</sup>, Mu-Hong Chen <sup>8,9</sup>, Chia-Fen Tsai <sup>8,9</sup>, Chang-Hsin Weng <sup>2</sup>, Li-Wei Ko <sup>10</sup>, Tzyy-Ping Jung <sup>11,\*</sup> and Yi-Hung Liu <sup>12,\*</sup>

- <sup>1</sup> International Research Center for Neurointelligence (WPI-IRCN), The University of Tokyo Institutes for Advanced Study (UTIAS), The University of Tokyo, Tokyo 113-0033, Japan; wu.chiente@mail.u-tokyo.ac.jp
  - <sup>2</sup> Hipposcreen Neurotech Corp. (HNC), Taipei 114, Taiwan; Alex\_Huang@hipposcreen-nc.com (H.-C.H.); alan\_huang@hipposcreen-nc.com (S.H.); Daniel\_Weng@hipposcreen-nc.com (C.-H.W.)
  - <sup>3</sup> Division of Psychosomatic Medicine, Department of Psychiatry, National Taiwan University Hospital, Taipei 100229, Taiwan; iming@ntuh.gov.tw (I.-M.C.); scliao@ntu.edu.tw (S.-C.L.)
  - <sup>4</sup> Institute of Health Policy and Management, National Taiwan University, Taipei 10617, Taiwan
  - <sup>5</sup> Department of Psychiatry & Community Medicine Research Center, Chang Gung Memorial Hospital, Keelung 204, Taiwan; kenchen@cgmh.org.tw (C.-K.C.); 8902008@cgmh.org.tw (C.L.)
  - <sup>6</sup> College of Medicine, Chang Gung University, Taoyuan 33302, Taiwan; shlee@cgmh.org.tw
  - <sup>7</sup> Department of Psychiatry, Chang Gung Memorial Hospital, Taoyuan 33305, Taiwan
  - <sup>8</sup> Department of Psychiatry, Taipei Veterans General Hospital, Taipei 11217, Taiwan; mhchen12@vghtpe.gov.tw (M.-H.C.); cftsai@vghtpe.gov.tw (C.-F.T.)
  - <sup>9</sup> Faculty of Medicine, National Yang Ming Chiao Tung University, Taipei 11217, Taiwan
  - <sup>10</sup> Department of Bio Science & Tech., National Yang Ming Chiao Tung University, Hsinchu 30010, Taiwan; lwko@nctu.edu.tw
  - <sup>11</sup> Institute for Neural Computation, University of California, San Diego, CA 92093, USA
  - <sup>12</sup> Department of Mechanical Engineering, National Taiwan University of Science and Technology, Taipei 106335, Taiwan
- \* Correspondence: tpjung@ucsd.edu (T.-P.J.); lyh@mail.ntust.edu.tw (Y.-H.L.)
- † These authors contributed equally to this paper.

**Citation:** Wu, C.-T.; Huang, H.-C.; Huang, S.; Chen, I.-M.; Liao, S.-C.; Chen, C.-K.; Lin, C.; Lee, S.-H.; Chen, M.-H.; Tsai, C.-F.; et al. Resting-State EEG Signal for Major Depressive Disorder Detection: A Systematic Validation on a Large and Diverse Dataset. *Biosensors* **2021**, *11*, 499. <https://doi.org/10.3390/bios11120499>

Received: 15 October 2021

Accepted: 4 December 2021

Published: 6 December 2021

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).

**Abstract:** Major depressive disorder (MDD) is a global healthcare issue and one of the leading causes of disability. Machine learning combined with non-invasive electroencephalography (EEG) has recently been shown to have the potential to diagnose MDD. However, most of these studies analyzed small samples of participants recruited from a single source, raising serious concerns about the generalizability of these results in clinical practice. Thus, it has become critical to re-evaluate the efficacy of various common EEG features for MDD detection across large and diverse datasets. To address this issue, we collected resting-state EEG data from 400 participants across four medical centers and tested classification performance of four common EEG features: band power (BP), coherence, Higuchi's fractal dimension, and Katz's fractal dimension. Then, a sequential backward selection (SBS) method was used to determine the optimal subset. To overcome the large data variability due to an increased data size and multi-site EEG recordings, we introduced the conformal kernel (CK) transformation to further improve the MDD as compared with the healthy control (HC) classification performance of support vector machine (SVM). The results show that (1) coherence features account for 98% of the optimal feature subset; (2) the CK-SVM outperforms other classifiers such as K-nearest neighbors (K-NN), linear discriminant analysis (LDA), and SVM; (3) the combination of the optimal feature subset and CK-SVM achieves a high five-fold cross-validation accuracy of 91.07% on the training set (140 MDD and 140 HC) and 84.16% on the independent test set (60 MDD and 60 HC). The current results suggest that the coherence-based connectivity is a more reliable feature for achieving high and generalizable MDD detection performance in real-life clinical practice.

**Keywords:** electroencephalographic (EEG) signal; healthcare; major depressive disorder (MDD); machine learning; support vector machine; feature selection

## 1. Introduction

Major depressive disorder (MDD) is a prevalent mood disorder characterized by persistent sadness, psychomotor retardation, and loss of interest in daily activities [1]. MDD is the leading cause of disability [2], with significant implications for the social functioning and work productivity of individuals [3,4]. Over 700,000 people kill themselves every year as a result of MDD, a disease that affects 320 million people worldwide [5]. In particular, the prevalence of MDD in the United States has increased substantially from 7% prior to the COVID-19 pandemic to 27% during the first year of the pandemic (April 2020–March 2021) [6]. Although treatments for MDD are promising, accurate diagnosis may be difficult due to its heterogeneous etiologies and various psychopathological manifestations [7]. Currently, the diagnosis of MDD is still based on clinical interviews and self-reports. There is an urgent need to develop complementary diagnostic tools that differentiate MDD from healthy persons based on neurophysiological changes.

### 1.1. Related Works

Many studies have tried to apply machine learning (ML) methods to classify resting-state (e.g., [8–23]) versus task-related (e.g., [21,24,25]) electroencephalography (EEG) signals between MDD and healthy control (HC) groups. A few studies have also proposed multi-model systems that fuse EEG signals and other physiological data (e.g., [26]).

Various features of EEG signals have been investigated. For example, in a study by Bachmann et al. [9], a classification accuracy of 92% on single-electrode EEGs of 26 participants was reported based on the use of a combination of spectral features and complexity features such as Higuchi's fractal dimension (HFD) and detrended fluctuation analysis (DFA). Hosseinifard et al. [23] reported that EEG alpha and beta power features achieved 70% and 73% accuracies, respectively, and that the use of complexity features could achieve greater than 80% accuracy. A study reported 99.6% accuracy [18] by using a combination of statistical, spectral, function connectivity, and complexity features on a public dataset containing 19-channel EEGs from 64 participants.

The reported MDD-HC classification accuracy based on EEG and ML varies greatly across studies (~60–99%). Such variations result from the differences in terms of chosen algorithms (features, classifiers, etc.), and also the differences in terms of apparatus setting, EEG collection procedure, and participant demographics. Even though some of previous studies have reported high classification accuracy, the true generalization performance of previous methods is still unknown, because the EEG datasets used in most of the previous studies were quite small (ours included [20,21]).

### 1.2. Problem Description

The sample sizes in these ML-based studies [9–24] were all less than 100 (24–90), except for two studies by Cai et al. [12,15] in which the subject numbers were 267 and 213, respectively (however, they only measured a very low-density EEGs from three electrodes at FP1, FP2, and Fz). Previous studies, due to limited data size, have only performed cross validation procedures on limited EEG data to estimate the generalization performance of their proposed algorithms/models, without using an independent test set of sufficiently large size. As a result of this limitation, the cross validation in previous studies often suffered from overfitting due to information and data leakage (more details are explained in Section 2). Furthermore, MDD is a highly heterogeneous disorder [27]. Due to the heterogeneity, it is difficult for a dataset of limited size to faithfully represent the real-life varia-

tions in EEG signals. All these factors including the use of small EEG dataset, the overfitting due to the data leakage in the cross validation, and the lack of an independent EEG test set, could lead to an overly optimistic or even biased estimate of performance. Therefore, to develop an EEG-based MDD-HC classification model with high generalizability, increasing the size and diversity of EEG datasets by recruiting a large enough group of participants is critical, as suggested in a recent review by Čukić et al. [8].

Indeed, developing a classification model that is resistant to large data variability due to highly diversified data sources is extremely challenging but critical. Aside from individual differences, data diversity can also result from differences in study designs, equipment and laboratory settings (EEG amplifier, room condition, etc.), EEG acquisition settings (reference electrode, sampling rate, etc.), and data collection procedures (preparation, instruction to participants, etc.) [28]. Any inconsistency in any factor across data collection sites will eventually result in higher inter-site data variability. Although it is possible to control for these factors across recording sites by applying rigorously standardized procedures (e.g., using the same equipment, experimental setting, and sequence of data collection), such control is not possible for a few other factors. For example, it is extremely difficult to maintain consistency in the recording environment (e.g., nature of lighting, room temperature, and ambient noise) consistent across hospitals. Another difficult factor to control is the technician's skill, which is also a significant factor in signal variability. Assume one technician is very skilled in operating an EEG system (including applying electric gel to the electrodes) at one location, while another is qualified but not so skilled at another location. The time allotted for EEG cap preparation may differ between the two sites, and as a result, participants may potentially experience fatigue as a result of a longer and more uncomfortable preparation [28]. Even if the EEG recording is in a resting state, this will have an effect on the participants' mental states. These inevitable factors of EEG variability have been considered to be a challenge to real-world brain-computer interface (BCI) [29,30] and EEG-based applications (e.g., seizure detection [31]). In this view, most of the previous EEG studies [9–25] have lacked a validation on datasets with high diversity (or large EEG variability) because their data were all collected from a single site.

### *1.3. Proposed Work*

In summary, previous studies on EEG-based MDD detection using the ML approach suffer from the common limitation of small data size and diversity, which makes the effectiveness of the previously used EEG features or classification models in MDD detection still inconclusive (e.g., the frontal alpha asymmetry (FAA) [32]). To address this limitation, the present study launched a multi-site EEG data collection project. The large-scale EEG dataset used in this study contains the high-density (26 electrodes) resting-state EEG signals recorded from 400 age- and gender-matched participants (200 MDD and 200 HC from four different medical centers using the same recording protocols such as equipment, settings, and data collection procedures). As a result, the collected dataset has a relatively large data size and high diversity as compared with prior studies which, to the best of our knowledge, did not meet the two following criteria: (1) more than 300 participants and (2) more than two EEG recording locations.

A dataset with a large size and high diversity may result in a relatively large variation in the EEG data distribution. This could further make the MDD-HC classification problem even more challenging. To address this issue, selecting a set of discriminative EEG features and designing a robust ML classifier with improved generalization ability are crucial to detecting MDD with high sensitivity. Because this is the first study to use ML on a large EEG dataset, systematically re-examining the classification performance of commonly used EEG features is critical for proposing an effective and generalizable solution. To this end, in this study, we test the classification performance by using band powers (BP), coherence, and two types of fractal dimensions (FDs), i.e., Higuchi's FD (HFD) and Katz' FD (KFD). We investigate the performance of all possible BP and FD features at all electrodes

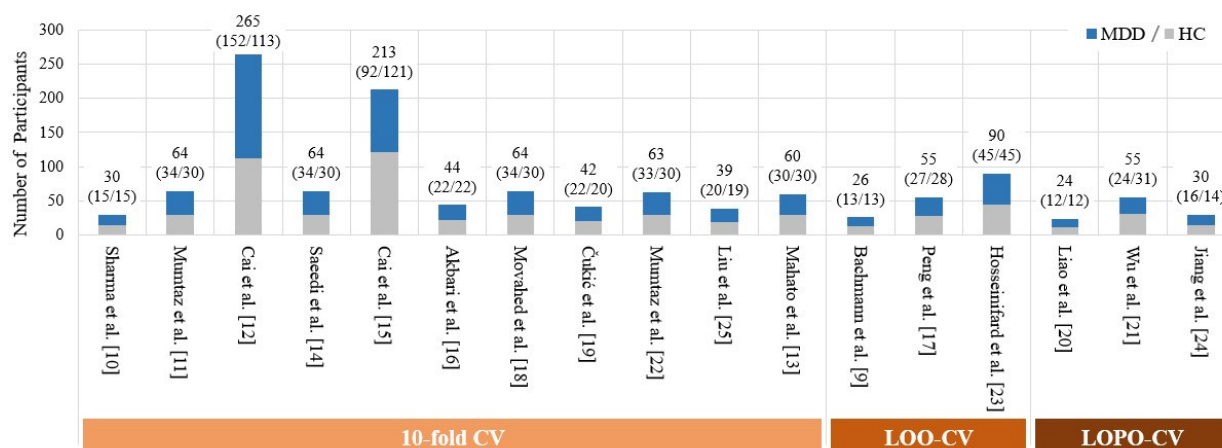
and all frequency bands (we consider five different bands for BPs) for classifying MDD and HC groups. We also examine coherence features from all possible (a) regional inter-hemispheric, (b) cross-regional inter-hemispheric, and (c) intra-hemispheric electrode pairs at various frequency bands. This thorough analysis results in a total of 1859 features. Then, we used a wrapper-based feature selection approach, i.e., the sequential backward selection (SBS) algorithm [33], to find the optimal feature subset among all of the 1859 feature candidates for different classifiers, through which we can examine what combination of the commonly used EEG features would be the most sensitive to the detection of MDD.

Classifier design is also critical for the MDD detection performance. Previous EEG-based MDD detection studies have used a variety of classifiers, including the basic K-nearest neighbors (K-NN), linear discriminant analysis (LDA), logistic regression (LR), and the more advanced support vector machine (SVM). These classifiers have been extensively tested, and overall, the SVM consistently outperforms others [17,21–23]. However, when dealing with a large EEG dataset with a high degree signal variation, implying that the pattern distribution between the two classes of MDD and HC is likely to be highly non-separable, SVM's performance might be comprised even if the optimal feature subset is used. To overcome this limitation, we use the conformal kernel (CK) technique [34] to improve the classification performance of the SVM. The success of CK-SVM has been shown in various EEG-based applications (e.g., [21,35]) that used relatively small EEG datasets. Because the CK transformation increases the spatial resolution around the SVM's separation hyperplane in a kernel-induced feature space, we expect the CK-SVM can outperform the SVM on the larger-size and higher-diversity dataset used in this study.

Finally, we tested the proposed methods on an independent test set comprised of EEG data from 120 participants. We further systematically compare the performance of MDD detection using several popular but inconclusive EEG markers, with the goal of providing a thorough evaluation and reference of optimal EEG features for developing a reliable and effective EEG-based MDD-HC classification.

## 2. Cross Validation Used in Previous Works: More Detailed Review and Analysis

Previous studies [9–25] had a very limited number of participants (see Figure 1 for detail). Therefore, it is difficult to partition the available EEG dataset of small sample size into a training dataset used for training classifiers and an independent test dataset of sufficient size. Due to this limitation, previous studies have only estimated the efficacy of their method in MDD detection by performing a k-fold CV on the entire EEG dataset, where  $k = 10$  [10–16,18,19,22,25] or  $k = \text{number of EEG samples}$  [9,17,23] (i.e., leave-one-out (LOO)).



**Figure 1.** Comparison of the number of subjects in previous EEG- and ML-based studies.

Directly performing a 10-fold CV on the group of participants (i.e., partitioning the entire group into 10 folds, with no EEGs of the same participants appearing in different folds at the same time) is nearly impossible due to participant size limitations. As a result, the k-fold CV is often performed on the EEG samples rather than participants. As an example, Movahed et al. [18] used a public EEG dataset (34 MDD, 30 HC) [36] and, for each participant, two EEG episodes of 5 min (eyes open and eyes closed) were collected. Their analysis procedure consisted of three steps: (1) data augmentation, in which the two signals of each participant were segmented into 10 EEG samples with 1 min lengths, resulting in a total of 6400 EEG samples at most; (2) data cleaning, in which noisy samples are removed; and (3) evaluation: by performing 10-fold CV on the remaining 510 samples (249 MDD and 261 HC EEG samples). They did not specify how many samples were saved from each participant after the second step. Nevertheless, it is reasonable to conclude that different samples of the same participant(s) were likely included in both training and test sets at the same time in each iteration of the 10-fold CV. Personal information leakage, which can be considered to be a type of data leakage known in the ML community [37], often results in the overfitting problem. As a result, such intra-participant EEG models often lead to higher classification performance, as emphasized in a recent review by Roy et al. [38].

To evaluate the participant-independent classification performance, a few studies used the leave-one-participant-out CV (LOPO-CV) strategy [20,21,24]. In each iteration of the LOPO-CV procedure, the EEG data of one participant were used for testing, while the data of remaining participants were used as the training set. Although the LOPO-CV avoids participant information leakage, another type of data leakage could occur. Take our previous studies [20,21] as examples, where the test data were actually used in the grid search-based feature selection and hyperparameter optimization. As a result, the test data in each iteration were not truly independent, but rather validation data. All these problems point to the same issue, i.e., a lack of an independent test set that meets the following two criteria: (1) containing a set of EEGs from a sufficiently large number of participants and (2) all these participants have never participated in the model training and validation.

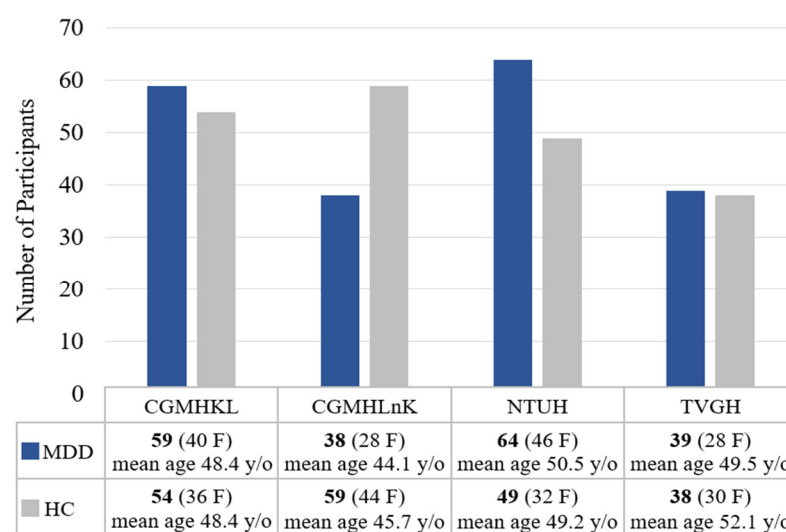
### 3. Materials and Methods

#### 3.1. Participants

A total of 200 individuals with MDD (142 females, 52.85 years old on average and 58 males, 54.90 years old on average) and 200 age- and gender-matched healthy controls (142 females, mean age of 49.87 years and 58 males, 54.59 years old on average) were included in the study. Individuals with MDD were recruited from four tertiary medical centers: National Taiwan University Hospital (NTUH), Taipei Veterans General Hospital (TVGH), Chang-Gung Memorial Hospital KeeLung (CGMHKL), and Chang-Gung Memorial Hospital LinKou (CGMHLnK). Healthy controls were recruited from local communities surrounding each medical center. The current study included 64 MDD and 49 HC from NTUH, 39 MDD and 38 HC from TVGH, 59 MDD and 54 HC from CGMHKL, and 38 MDD and 59 HC from CGMHLnK. Figure 2 summarizes the gender distribution and mean age of participants from each medical center. All research protocols were reviewed and approved by the corresponding Research Ethic Committee from each medical center (IRB no: 201908038RSC for NTUH, 2019-09-004C for TVGH, and 201901342A3 for CGMH). All participants were informed of the study procedure and signed the consent forms before evaluations of medical history, family history, and neuropsychological tests.

The inclusion criteria for individuals with MDD were: (1) Age of 20 years or older, (2) with a diagnosis of major depressive disorder by two board-certified psychiatrists using the DSM-5 [39], (3) without any other psychiatry diagnoses based on evaluation of the Mini International Neuropsychiatric Interview (MINI) [40], and (4) the ability to comprehend the contents of consent forms. The inclusion criteria for the HC group were: (1) Age

of 20 years or older, (2) no history of mental disorder as confirmed by a board-certified psychiatrist, and (3) capable of understanding the contents of consent forms. The general exclusion criteria for the current study were: (1) Confirmed pregnancy; (2) inability to understand the contents of consent form; (3) with a diagnosis of other types of mental disorder, including schizophrenia, psychotic disorder, bipolar disorder, obsessive and compulsive disorder, autism spectrum disorder, neurodevelopmental disorder, substance use disorder, neurocognitive disorder; (4) with a diagnosis of neurological disorder, including stroke, seizure, neurocognitive diseases, brain tumor; (5) a history of medication with anticonvulsant (e.g., Clonazepam) within 7 days before experiment; (6) severe vision or hearing impairment that leads to inability to perform cognitive tests; (7) a history of sustained loss of consciousness due to traumatic brain injury; (8) a history of other diseases that produce metabolic encephalopathy.



**Figure 2.** Gender distribution and mean age of participants from each medical center.

Each participant was interviewed by a psychiatrist for personal medical history, family medical history, and diagnosis. In addition, all participants completed the following questionnaires: Patient Health Questionnaire (PHQ-9) [41], Becks Depression Inventory–II (BDI-II) [42], and Hospital Anxiety and Depression Scale (HADS) [43]. Table 1 summarizes the demographic data and questionnaire scores for all participants. As can be observed, females were the gender majority on both MDD and HC groups. In many epidemiological studies, women were twice as likely to have depression as men [44]. The female predominance in our sample was mainly a result of gender difference in the epidemiological nature of MDD. Furthermore, since the gender distribution in the MDD and HC groups were similar, gender effect on electrophysiological biomarkers would not affect our findings.

**Table 1.** Demographic data and results of questionnaires.

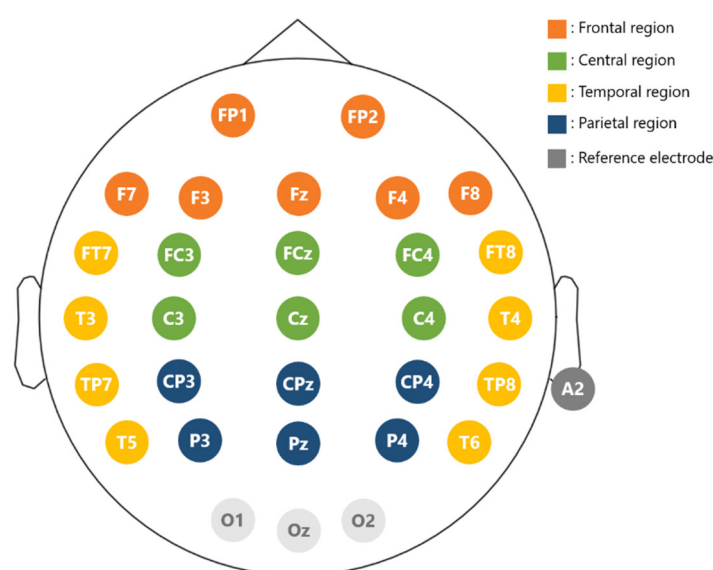
Variable	MDD, <i>n</i> = 200	HC, <i>n</i> = 200	<i>p</i> -Values	Effect Size
Gender	142 F, 58 M	142 F, 58 M	1.000	0
Age	53.44 (±16.44)	51.24 (±17.85)	0.1969	0.1293
BDI-II	25.79 (±14.30)	4.18 (±6.74)	$5.60 \times 10^{-59}$	1.9282
PHQ-9	12.64 (±7.30)	2.09 (±3.73)	$4.51 \times 10^{-54}$	1.8150
HADS-A	10.46 (±4.91)	3.53 (±3.37)	$1.18 \times 10^{-46}$	1.6425
HADS-D	10.29 (±5.19)	2.50 (±2.85)	$7.91 \times 10^{-56}$	1.8556

Note: BDI-II, Becks Depression Inventory–II; PHQ-9, Patient Health Questionnaire; HADS, Hospital Anxiety and Depression Scale.



### 3.2. Apparatus

Experimental stimuli were presented on a 24" flat screen monitor. Participants were sitting with their eyes 70 cm away from the monitor. The EEG data were recorded with a 32-channel Quick-cap (NeuroScan Inc., Charlotte, NC, USA) attached to four HNC amplifiers (HippoScreen Neurotech Corp.) with gain 50, A/D resolution 24 bit, and sampling rate 500 Hz. Electrodes were made of Al/AgCl material, and the montage was based on the extended international 10–20 system (Figure 3). Electrode gel (Quick-Gel by Compumedics Inc., Charlotte, NC, USA) was applied to the electrodes to keep the impedance below 10 K Ohm. Specifically for each electrode, we gently rubbed the surface of the scalp region while injecting gels so as to reduce the impedance level to be below the designated level of 10 k Ohm. EEG signals were on-line referenced to the A2 electrode (right mastoid).



**Figure 3.** Electrode layout. Electrode positions follow the extended 10–20 international system. Reference is at A2. The entire scalp region is divided into four regions for analysis: frontal (brown), central (green), parietal (blue), and temporal (yellow).

### 3.3. Data Collection and Preprocessing

We recorded 90 s of eyes-open resting-state EEG signals from all participants. A timer was used to signal the start of the resting-state recording. Following the presentation of a fixation cross, participants were required to maintain a gentle fixation on the fixation cross for 90 s with their eyes open. During preprocessing, signals were re-referenced to vertex Cz and digitally filtered with a 2–50 Hz bandpass filter. Signals from O1, Oz, and O2 were excluded from analyses due to the constant signal instability caused by the cap fitness with Asian population (the back of heads of Asians are usually flatter and these electrodes were more often detached during recording). Then, an independent component analysis (ICA) conducted based on the Picard algorithm [45] was used to remove artifacts caused by eye movements and eye blinks [46], which was implemented with MNE-Python. Then, the 90 s resting-state EEG data from the remaining 26 electrodes were divided into 15 non-overlapped 6 s EEG epochs.

### 3.4. Feature Extraction

#### 3.4.1. Band Power (BP)

The power spectrum density (PSD) of each of the 15 6 s EEG epochs from the same electrode was computed using a Fast Fourier Transform (FFT), and the powers of delta (2–4 Hz), theta (4–8 Hz), alpha (8–13 Hz), beta (13–30 Hz), and gamma (30–45 Hz) frequency bands were calculated separately. Then, the 15 band powers (BPs) of the same

electrode's 15 EEG epochs were averaged for each band. Given the very large values of the calculated BP values, we took the logarithm to the averaged BP value (a log BP). We obtained a total of 130 log BP features from each participant's EEG because there were five frequency bands and 26 electrodes. The BPs in the following refer to the log BPs. It should be noted that the delta band used in this study is narrower than the commonly used 0.5–4 Hz. As a result, the delta band in this study refers to the fast delta [47].

### 3.4.2. Coherence

EEG coherence was also calculated by the FFT method. For each pair of electrodes and for each band  $f$ , the squared band coherence value between two EEG epochs  $i$  and  $j$  was calculated spectrally as:

$$COH_{ij}^2(f) = \frac{|G_{ij}(f)|^2}{G_{ii}(f)G_{jj}(f)} \quad (1)$$

where  $COH_{ij}(f)$  is the coherence on band  $f$ ,  $G_{ij}(f)$  is the averaged cross-power density, and  $G_{ii}(f)$  and  $G_{jj}(f)$  are respective averaged auto-power spectral densities across the frequency band  $f$ . The coherence value lies in the interval  $[0, 1]$ . A coherence value that is close to 1 corresponds to a strong linear association between the EEG signals. Then, we computed the averaged coherence values across 15 epochs (referred to as coherence hereafter). We extracted a total of 1625 coherence features from each participant's EEG ( $26 \times 25/2$  electrode pairs  $\times$  5 bands).

### 3.4.3. Higuchi's Fractal Dimension (HFD)

Fractal dimensions (FD) have been used to measure complexity of a time series. The FD of a time series can be estimated by time domain or phase space domain approaches [48]. The former estimates the FD directly in the waveform of a signal, such as the Higuchi and Katz's methods used in the current study. The latter estimates the FD of an attractor in the state-space domain. Compared to the phase space domain approach, the time-domain approach has the advantage of fast computation.

In this study, each EEG epoch is a 6-second signal containing 3000 sample points (sampling rate = 500 Hz). For such a short-time segment, HFD has been shown to be a suitable FD estimator [49,50]. We first reconstruct  $k$  new signals from a time series of  $N_s$  samples ( $X = [X(1), X(2), \dots, X(N_s)]$ ) as:

$$X_m^k = \left[ X(m), X(m+k), X(m+2k), \dots, X\left(m + \text{int}\left(\frac{N_s - m}{k}\right)k\right) \right], m = 1, 2, \dots, k \quad (2)$$

where  $m$  is the initial time point,  $k$  is the time interval, and  $\text{int}()$  denotes integer function. Here, we reset the definition of the symbol  $k$  from the number of folds in a CV procedure to the number of reconstructed signals, in order to be consistent with the symbol used in previous HFD studies. For a starting time point  $m$ , the length  $L_m(k)$  of the corresponding curve (i.e., reconstructed signal)  $y_m^r$  is given by:

$$L_m(k) = \frac{1}{k} \left[ \frac{N_s - 1}{\text{int}\left(\frac{N_s - m}{k}\right)k} \left( \sum_{i=1}^{\text{int}\left(\frac{N_s - m}{k}\right)} |X(m + ik) - X(m + (i-1)k)| \right) \right] \quad (3)$$

which represents the normalized sum of absolute values of differences between two consecutive points separated by a distance of  $k$ . For each time interval  $k$ , the average curve length is then calculated by:

$$L(k) = \frac{1}{k} \sum_{m=1}^k L_m(k) \quad (4)$$

The HFD value of the time series is the slope of  $\log(L(k))$  against  $\log(1/k)$  within the range of  $k = 1, 2, \dots, k_{\max}$ , which can be calculated using a least-squares linear fitting.



The HFD value lies within the range between 1 and 2. The HFD value is equal to 1 for a deterministic signal, and a signal with a higher HFD value corresponds to a higher irregularity. The HFD calculation involves one free parameter, the maximum value of  $k$ ,  $k_{\max}$ . It has been suggested in a motor imagery study [51] that the setting of  $k_{\max} = 100$  is appropriate for the detection of change in brain dynamics. Here, we tried three different settings ( $k_{\max} = 50, 100$ , and  $150$ ) to seek for a better solution of the HFD for MDD-HC classification. In the present study, for each electrode, we calculated the HFD value of each EEG epoch, and then computed the averaged HFD values across 15 epochs. Finally, a total of 78 HFD features were extracted from each participant ( $26 \text{ electrodes} \times 3 \text{ parameter values}$ ).

#### 3.4.4. Katz's Fractal Dimension (KFD)

Katz's method [52] estimates the FD of a time series directly from its waveform by

$$\frac{\log\left(\frac{L_w}{A}\right)}{\log\left(\frac{D_w}{L_w}\right) + \log\left(\frac{L_w}{A}\right)}, \quad (5)$$

where  $L_w$  and  $D_w$  are the total length and the planar extent of the waveform, respectively, and  $A$  is the average of the distances between two successive sample points of the time series. The value of KFD also lies within the range between 1 and 2. The complexity of a time series increases with the KFD value. In this study, the KFDs of the 15 EEG epochs from the same electrode were calculated, and then averaged. Therefore, we obtained a total of 26 KFD features from each participant.

In total, we extracted a total of 1859 features ( $130 \text{ BPs} + 1625 \text{ coherences} + 78 \text{ HFDs} + 26 \text{ KFDs}$ ) from each participant's 90 s long, 26-channel EEG signals. In addition, since we used a feature averaging strategy across epochs in the feature extraction process, there is only one value for each feature eventually. As a result, for example, there is only one 26-dimensional feature vector (also known as a data or a data point in the following) is extracted from each participant if all of the alpha BP features from the 26 electrodes are used for classification.

### 3.5. Classification

Four different classifiers were compared: K-NN, LDA, SVM, and CK-SVM. Here, we briefly introduce the LDA and SVM, and then show how the conformal kernel transformation improves the conventional SVM's performance, i.e., CK-SVM.

#### 3.5.1. K-NN

K-NN is an instance-based learning method. A test data point is compared to K-nearest data points in the training set. Then, the class label of the test data point is assigned according to majority voting among the classes of the K-nearest training data points. The Euclidean metric is used to calculate the distance between data points in this study.

#### 3.5.2. LDA

Suppose that  $n$  features are used for classification. The LDA classifier finds a hyper-plane  $w_{lda}^t x + b_{lda} = 0$  to separate the two classes of MDD and HC in the original space of patterns  $R^n$  [21], where:

$$w_{lda} = (m_{MDD} - m_{HC})^t \Sigma^{-1} \quad (6)$$

is the weight vector.

$$b_{lda} = \frac{1}{2} (m_{MDD} - m_{HC})^t \Sigma^{-1} (m_{MDD} + m_{HC}) - \ln \left( \frac{C_{MDD} p_{HC}}{C_{HC} p_{MDD}} \right) \quad (7)$$

is the bias of the separating hyperplane;  $m_{MDD}$  and  $m_{HC}$  are the mean vectors of the training data of MDD and HC classes, respectively;  $\Sigma$  is the covariance matrix of the set containing both classes' training data;  $C_{MDD}$  and  $C_{HC}$  are penalties assigned to the MDD and HC classes, respectively; and  $p_{HC}$  and  $p_{MDD}$  are the a priori probabilities of the HC and the MDD classes, respectively. An unseen data  $x \in R^n$  is classified as MDD if  $D_{LDA}(x) = w_{lda}^t x + b_{lda} > 0$ ; otherwise, it is classified as HC.

### 3.5.3. SVM

An SVM maps the training data  $x_i \in R^n$ ,  $i = 1, \dots, M$ , into a higher dimensional feature space  $F$  via a nonlinear mapping  $\varphi: R^n \rightarrow F$ , and then finds an optimal separating hyperplane  $H = w^t \varphi(x) + b = 0$  that maximizes the margin of separation and minimizes the training error. The decision function of SVM is given by:

$$D_{SVM}(x) = \sum_{x_i \in SV} \alpha_i y_i K(x_i, x) + b \quad (8)$$

where  $0 \leq \alpha_i \leq C$  are Lagrange multipliers;  $y_i \in [-1, +1]$  are class labels of the training data;  $C$  is the penalty weight that needs to be determined experimentally; SV stands for the support vectors, i.e., the training data whose  $\alpha_i$ 's satisfying the condition  $0 < \alpha_i \leq C$ ; the bias  $b$  is computed using the Kuhn–Tucker conditions of SVM; and  $K(\cdot)$  is the kernel function that computes the inner product of two mapped data points. Here, we adopted the Gaussian kernel defined by  $K(x_i, x_j) = \exp(-\gamma \|x_i - x_j\|^2)$ , where  $\gamma$  is the kernel parameter. The class label of an unseen data  $x$  is predicted as MDD if  $D_{SVM}(x) > 0$ ; otherwise  $x$  is classified as the HC group.

### 3.5.4. CK-SVM

A Gaussian kernel-induced feature space  $F$  is of infinite dimension, and the Gaussian kernel embeds the original space  $R^n$  as a Riemannian manifold distributed on the unit sphere centered at the origin of the space  $F$  (please refer to the author's study [53] for a detailed explanation). Thus, the relationship between a small shift of the Riemannian distance  $ds$  and a small displacement in the original space  $dx$  can be described by the Riemannian metric  $g_{ij}(x)$  as follows:

$$ds^2 = \sum_{i,j} g_{ij}(x) dx_i dx_j = \sum_{i,j} \left. \frac{\partial K(x, x')}{\partial x_i \partial x'_j} \right|_{x'=x} dx_i dx_j \quad (9)$$

where  $x_i$  and  $x_j$  denote the  $i$ th and  $j$ th elements of the vector  $x$ , respectively. This shows that a local volume's size in  $R^n$  can be enlarged or shrink in  $F$  by conformally transforming the mapping as  $\tilde{\varphi}(x) = c(x)\varphi(x)$  and choosing a proper conformal function as [34]:

$$c(x) = \sum_{x_i \in T} \exp \left( \frac{-\|\varphi(x) - \varphi(x_i)\|^2}{\frac{1}{n_e} \sum_j \|\varphi(x_j) - \varphi(x_i)\|^2} \right) \quad (10)$$

where the term  $\sum_{j \neq i} \|\varphi(x_j) - \varphi(x_i)\|^2 / n_e$  is the average squared distance from the mapped data point  $\varphi(x_i)$  to its  $n_e$  nearest neighbors  $\varphi(x_i)$ . Notice that the expression of Equation (10) can be further simplified by kernel tricks, and  $K(x_i, x_i) = 1$ ,  $\forall i$  in the present study, because of the use of Gaussian kernel. The conformal function decreases exponentially with the distance from  $\varphi(x)$  to the image of the set  $T$ . By this setting, the spatial resolution is enlarged around the images of  $x_i$  and contracted elsewhere in the feature space of  $F$ . The inner product of two conformally transformed data points  $\tilde{\varphi}(x)$  and  $\tilde{\varphi}(x')$  further leads to the form of a conformal transform of a kernel:

$$\tilde{K}(x, x') = c(x)c(x')K(x, x') \quad (11)$$

where  $\tilde{K}$  is the so-called conformal kernel (CK).

The design of the function  $c(x)$  is crucial to the performance of the CK method. In the original work by Wu and Amari [34], they proposed that the set  $T$  should include the

support vectors, based on the assumption that most support vectors within the margin of separation. By doing so, increasing the spatial resolution around the mapped support vectors is equivalent to enlarging the spatial resolution of the margin of separation in the feature space  $F$ . This leads to a better class separability, further resulting in increased generalization. However, in a highly non-separable case, many support vectors would fall outside the margin rather than within it. To remedy this problem, Wu et al. [21] proposed that the set  $T$  should only include the support vectors within the margin (SVWM):

$$\text{SVWM} = \{x_i | -1 \leq w^t \varphi(x_i) + b \leq 1\} \quad (12)$$

and that the number of nearest neighbors in the set  $T$  can be set as  $n_e = 3$ .

In the current study, we use this strategy to increase the spatial resolution around the images of the SVWM to further increase the spatial resolution in the vicinity of the SVM's separating hyperplane in the feature space. However, we set  $n_e = 5$ , for the reason that it is expected that the SVWM would be much more in the current study, because of the use of a much larger and more diverse EEG dataset. In summary, the CK-SVM has two training steps,

**Step 1** Train the SVM with a training set and a chosen kernel  $K(\cdot, \cdot)$ , and determine the optimal hyperparameter.

**Step 2** Use the conformal kernel  $\tilde{K}(\cdot, \cdot)$  calculated by Equation (11) to retrain the SVM with the same training set and the same optimal hyperparameter of the SVM.

It should be noted that using the same training set in Step 2 means that the features are also the same as those used in the training set used for the SVM in Step 1. Finally, the retrained SVM is called CK-SVM.

### 3.6. Hyperparameter Optimization Procedure

For the K-NN classifier, we set  $K = 3$ . In LDA, we assigned the same penalties to both MDD and HC classes by setting  $C_{MDD}/C_{HC} = 1$ . The SVM involves two free parameters, i.e., the penalty weight  $C$  and the kernel parameter  $\gamma$ . The two parameters were optimized using a 5-fold CV procedure combined with a grid search. Initially, we searched for the suboptimal value of  $\gamma$  within a wide range with a rough grid, and then searched the optimal solution with smaller parameter grids in the set  $\{(C, \gamma) | C = 1000, 10000; \gamma = 0.0001:0.0001:0.01\}$  containing a total of 200 grids. Given a training set  $S$ , the entire hyperparameter optimization consists of two steps: (1) For every grid, the 5-fold CV is performed once on  $S$  and (2) the optimal grid is the one giving the highest 5-fold CV classification accuracy.

### 3.7. Determine the Optimal Feature Subset Using Sequential Backward Selection

The total number of features in the current study is quite large. The best combination of features for classification must be determined. In this case, wrapper methods are preferable to filter methods (e.g., ranking each feature by Fisher's criterion [54]) for feature selection. The wrapper approach is classifier specific, which determines the optimal feature combination directly by using the generalization performance (e.g., cross-validation accuracy) as the evaluation criterion. As a result, the wrapper approach often outperforms the filter approach in terms of feature selection [55]. Therefore, we adopted a popular wrapper method, i.e., the sequential backward selection (SBS) [33], to select the optimal feature set from the 1859 candidates.

SBS is a heuristic search method. Let  $N$  represent the number of feature candidates. Initially, a 5-fold CV is performed on the dataset with all the  $N$  features. Following that, each feature is removed, one at a time. The 5-fold CV is applied to all  $N-1$  features subsets. Then, the worst feature is discarded. Here, the worst feature means that without it, the subset with the remaining  $N-1$  features gives the highest cross-validation accuracy among all the  $N-1$  subsets. In the next round, each of the  $N-1$  features is deleted one at a time, and the worst feature is discarded, using the same logic. A subset with  $N-2$  features is formed.

This procedure is repeated until none of the  $N$  features is left. Finally, the optimal feature subset is the one that produces the highest 5-fold CV accuracy. It should be noted that if the chosen classifier is SVM, the SBS-based feature selection must be performed together with the classifier's hyperparameter optimization procedure.

#### 4. Results

To provide a reliable result with high generalizability, the collected dataset (Table 1) was divided into a training set and a test set. The training set included the EEG data of 280 participants (140 MDD and 140 HC), while the test set included the EEG data of the remaining 120 participants (60 MDD and 60 HC). The training set was used for feature selection as well as classifier training and validation. The test set was used to see how well the trained and optimized methods generalized to another independent group of 120 participants. Tables 2 and 3 summarize the demographic data of the training and the test sets, respectively.

**Table 2.** Demographic data and scores of questionnaires for participants in the training set.

Variable	MDD ( $n = 140$ )	HC ( $n = 140$ )	$p$ -Values	Effect Size
Gender	100 F, 40 M	97 F, 43 M	0.7935	0.0474
Age	53.06 ( $\pm 16.31$ )	50.83 ( $\pm 17.64$ )	0.2734	0.1312
BDI-II	25.82 ( $\pm 14.23$ )	3.69 ( $\pm 5.98$ )	$1.43 \times 10^{-44}$	2.0201
PHQ-9	12.85 ( $\pm 7.31$ )	2.00 ( $\pm 3.52$ )	$1.77 \times 10^{-40}$	1.8855
HADS-A	10.52 ( $\pm 4.97$ )	3.19 ( $\pm 3.05$ )	$4.85 \times 10^{-37}$	1.7720
HADS-D	10.29 ( $\pm 5.24$ )	2.32 ( $\pm 2.71$ )	$4.63 \times 10^{-41}$	1.9046

**Table 3.** Demographic data and scores of questionnaires for participants in the test set.

Variable	MDD ( $n = 60$ )	HC ( $n = 60$ )	$p$ -Values	Effect Size
Gender	42 F, 18 M	45 F, 15 M	0.6826	0.1091
Age	54.32 ( $\pm 16.70$ )	52.18 ( $\pm 17.40$ )	0.4983	0.1240
BDI-II	25.70 ( $\pm 14.46$ )	5.30 ( $\pm 8.41$ )	$4.22 \times 10^{-16}$	1.7237
PHQ-9	12.13 ( $\pm 7.25$ )	2.30 ( $\pm 4.18$ )	$4.05 \times 10^{-15}$	1.6475
HADS-A	10.32 ( $\pm 4.76$ )	4.32 ( $\pm 3.91$ )	$1.42 \times 10^{-11}$	1.3662
HADS-D	10.27 ( $\pm 5.08$ )	2.90 ( $\pm 3.12$ )	$3.16 \times 10^{-16}$	1.7334

##### 4.1. Electrode- and Region-Specific Comparison of 5-Fold CV Classification Accuracy Between Features Using LDA Classifier

Here, we compare the 5-fold CV-based MDD-HC classification accuracy (referred to simply as accuracy in this subsection) between commonly used features in terms of different scalp regions (region level) and different electrodes (electrode level). Prior to using a nonlinear classifier such as SVM, first, we use an LDA classifier to examine whether the features in different scalp regions can be well separated by the linear decision boundary learned from LDA to examine the linear class separability of the features in different electrodes and scalp regions.

Here, we use the standard accuracy as the performance metric:

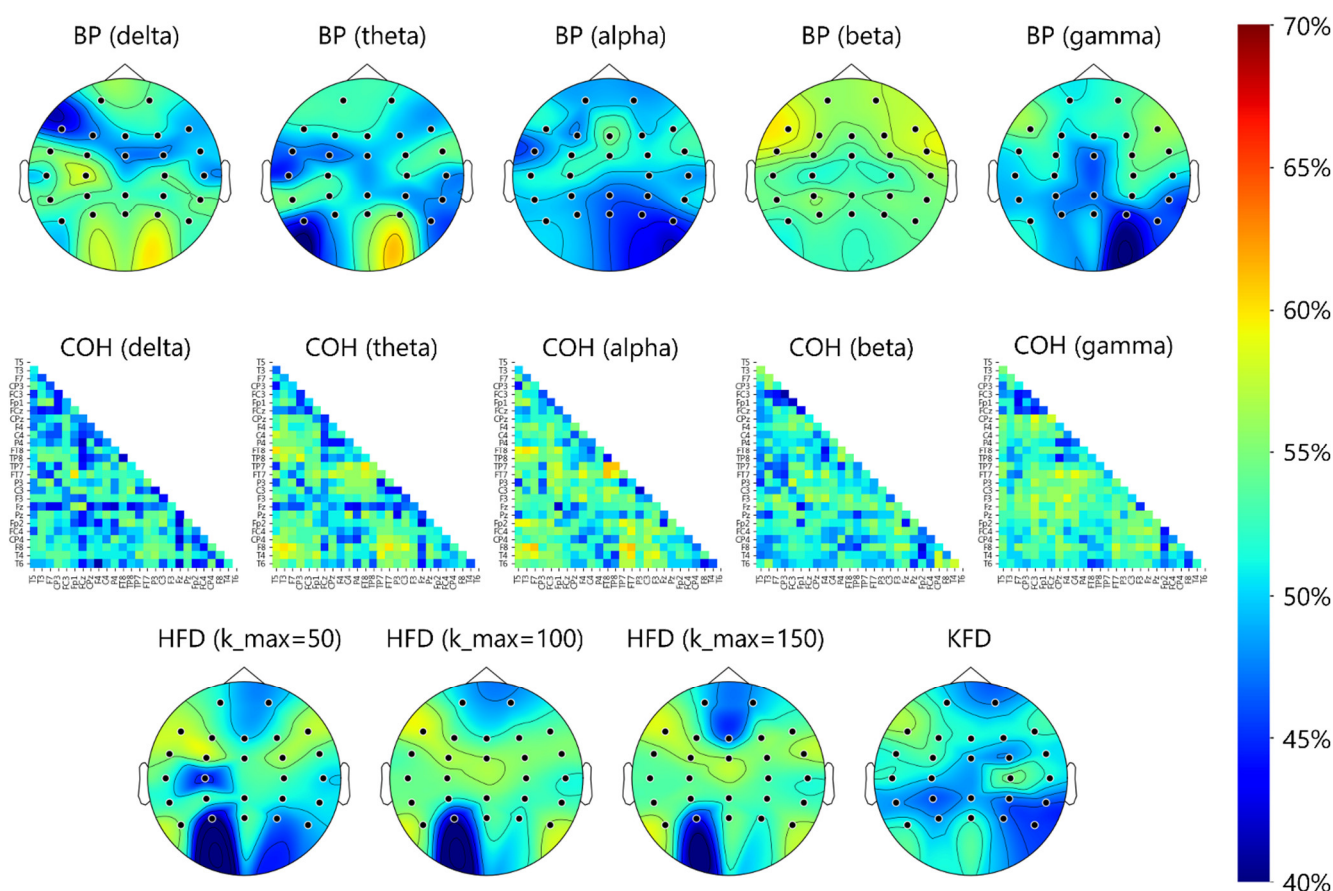
$$\text{Accuracy} = (TP + TN) / (TP + FN + TN + FP)$$

where  $TP$ ,  $TN$ ,  $FP$ , and  $FN$  represent true positive, true negative, false positive, and false negative, respectively. Because both the training and test sets have equal numbers of MDD and HC data, the classification accuracy would not result in biased classifier performance. Moreover, because the dataset is balanced, the accuracy identical to the average of sensitivity (true positive rate) and specificity (true negative rate), which are two measures in clinical evaluation of medical devices.

Figure 4 shows the topoplots that depict the electrode-specific results. Overall, for all bands, the classification accuracy based on each electrode's BP feature was below 60%. The HFD and KFD features yielded the same results. The best accuracy for a single electrode was 59.29% (beta BP at F7). For coherence, a few electrode pairs provided accuracies slightly higher than 60%. The best accuracy of 61.71% was obtained from the alpha-band coherence feature of the F8-FT7 pair, which is only slightly higher than the chance level of 50%. Table 4 shows the scalp-region-specific results. Take the beta BP of the frontal region as an example, the seven beta-band BP features from seven frontal electrodes (Figure 3) were all fed into the LDA classifier without feature selection and the accuracy was 56.07%. In the same frontal region, the accuracy increased slightly (61.07%) when the best two beta-band BP features (FP1 and FP2) determined by the SFS algorithm. The results under the "ALL" category were based on features extracted from the 26 electrodes or  $26 \times 25/2$  electrode pairs in the frontal, central, parietal, and temporal regions.

**Table 4.** Comparison of 5-fold cross validation accuracies among different features, frequency bands, and scalp regions using LDA (in %). In each grid of this table, the numbers inside and outside the brackets denote the accuracies obtained from the features with and without the SBS-based feature selection.

EEG Features		Frontal	Central	Temporal	Parietal	ALL
BP	$\delta$	49.53 $\pm$ 6.80 (55.82 $\pm$ 3.19)	56.35 $\pm$ 4.24 (60.77 $\pm$ 4.71)	53.86 $\pm$ 3.50 (61.19 $\pm$ 5.49)	51.80 $\pm$ 3.81 (58.09 $\pm$ 5.30)	48.80 $\pm$ 6.76 (63.07 $\pm$ 3.11)
		49.63 $\pm$ 11.93 (58.91 $\pm$ 4.74)	51.24 $\pm$ 6.88 (55.26 $\pm$ 2.08)	52.62 $\pm$ 7.25 (57.25 $\pm$ 5.37)	45.83 $\pm$ 4.57 (55.44 $\pm$ 4.97)	52.58 $\pm$ 8.61 (59.82 $\pm$ 9.68)
	$\alpha$	47.06 $\pm$ 7.21 (56.28 $\pm$ 3.85)	51.67 $\pm$ 7.20 (54.01 $\pm$ 1.82)	44.49 $\pm$ 7.36 (49.16 $\pm$ 2.90)	46.16 $\pm$ 6.84 (54.93 $\pm$ 3.19)	42.09 $\pm$ 9.95 (57.65 $\pm$ 4.32)
		52.53 $\pm$ 5.36 (58.61 $\pm$ 5.71)	55.00 $\pm$ 7.35 (55.42 $\pm$ 3.64)	53.87 $\pm$ 6.51 (56.29 $\pm$ 3.11)	50.76 $\pm$ 7.64 (53.96 $\pm$ 2.62)	51.74 $\pm$ 9.00 (64.21 $\pm$ 4.43)
	$\gamma$	48.28 $\pm$ 7.21 (54.63 $\pm$ 4.43)	55.51 $\pm$ 2.19 (56.88 $\pm$ 4.13)	46.62 $\pm$ 8.46 (51.61 $\pm$ 3.64)	43.94 $\pm$ 6.00 (53.13 $\pm$ 5.46)	44.15 $\pm$ 11.66 (60.29 $\pm$ 3.81)
COH	$\delta$	53.44 $\pm$ 8.04 (61.94 $\pm$ 2.53)	54.38 $\pm$ 5.86 (59.88 $\pm$ 9.68)	46.70 $\pm$ 5.95 (59.08 $\pm$ 4.32)	52.85 $\pm$ 3.23 (62.83 $\pm$ 3.57)	52.49 $\pm$ 4.85 (81.27 $\pm$ 3.64)
		56.30 $\pm$ 7.58 (64.24 $\pm$ 6.83)	58.64 $\pm$ 7.65 (61.49 $\pm$ 5.30)	51.86 $\pm$ 5.64 (61.53 $\pm$ 5.60)	52.47 $\pm$ 10.46 (63.58 $\pm$ 4.43)	49.16 $\pm$ 9.78 (75.36 $\pm$ 3.46)
	$\alpha$	52.67 $\pm$ 8.48 (66.11 $\pm$ 6.43)	56.08 $\pm$ 4.45 (62.44 $\pm$ 3.11)	56.36 $\pm$ 8.28 (65.28 $\pm$ 2.62)	52.34 $\pm$ 13.05 (63.25 $\pm$ 6.64)	51.32 $\pm$ 6.17 (79.06 $\pm$ 6.25)
		55.43 $\pm$ 6.91 (66.28 $\pm$ 4.60)	58.68 $\pm$ 2.92 (62.97 $\pm$ 5.30)	55.78 $\pm$ 7.18 (64.49 $\pm$ 1.96)	53.15 $\pm$ 6.28 (64.17 $\pm$ 6.53)	53.37 $\pm$ 7.41 (85.65 $\pm$ 10.68)
	$\gamma$	50.05 $\pm$ 6.86 (62.17 $\pm$ 6.74)	58.60 $\pm$ 8.13 (63.82 $\pm$ 4.29)	46.13 $\pm$ 5.57 (63.26 $\pm$ 8.07)	55.66 $\pm$ 5.33 (59.42 $\pm$ 4.43)	49.70 $\pm$ 5.41 (78.11 $\pm$ 4.71)
HFD	$k_{\max} = 50$	51.28 $\pm$ 6.06 (59.01 $\pm$ 8.10)	57.38 $\pm$ 3.30 (59.47 $\pm$ 4.07)	52.23 $\pm$ 7.21 (60.09 $\pm$ 5.82)	48.52 $\pm$ 6.25 (55.50 $\pm$ 4.97)	50.25 $\pm$ 6.29 (62.08 $\pm$ 5.91)
		52.93 $\pm$ 6.93 (59.46 $\pm$ 6.02)	59.42 $\pm$ 3.90 (60.10 $\pm$ 2.42)	55.31 $\pm$ 5.98 (60.17 $\pm$ 4.92)	51.47 $\pm$ 3.54 (57.35 $\pm$ 4.92)	54.33 $\pm$ 4.38 (63.81 $\pm$ 6.39)
	$k_{\max} = 100$	52.93 $\pm$ 6.93 (59.24 $\pm$ 4.60)	58.93 $\pm$ 3.86 (60.66 $\pm$ 3.75)	54.14 $\pm$ 8.45 (60.21 $\pm$ 3.64)	50.56 $\pm$ 5.60 (58.04 $\pm$ 5.71)	54.04 $\pm$ 4.73 (64.96 $\pm$ 6.04)
KFD		44.60 $\pm$ 8.37 (58.36 $\pm$ 4.60)	41.05 $\pm$ 10.18 (63.01 $\pm$ 2.90)	35.65 $\pm$ 9.99 (58.23 $\pm$ 5.58)	50.75 $\pm$ 10.15 (54.41 $\pm$ 6.87)	46.38 $\pm$ 8.46 (61.89 $\pm$ 4.87)



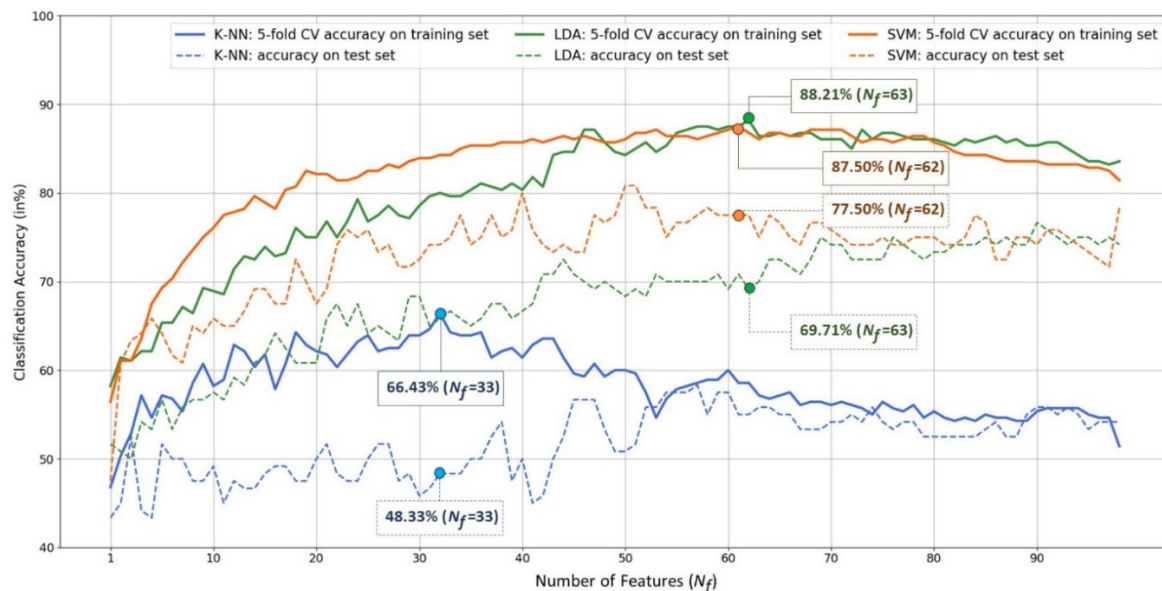
**Figure 4.** Topoplots of 5-fold CV classification accuracies in BP, coherence (COH), HFD, and KFD features. The accuracy distribution shown around the occipital area should be ignored, because the EEGs from O1, O2, and Oz were not analyzed in the present study.

Table 4 shows that in the condition of “without feature selection”, all the combinations of features and regions do not show any discriminative ability (<60%). For most of the combinations, the accuracy improves after feature selection. However, the improvement is limited (slightly greater than 60%) for all intra-regional and the ALL-regional features, with the exception of the ALL-scalp coherence. The results show that after feature selection, the accuracy of using the coherence features was greater than 80% for all bands except for the beta coherence features. On the one hand, a set of optimal theta-band coherence features extracted from the entire scalp’s electrode pairs, in particular, achieved the highest accuracy of 83.21%. BP and FD (including KFD and HFD) features, on the other hand, did not perform well even when using their corresponding optimal feature subsets for each scalp region.

#### 4.2. Feature Fusion and Feature Selection Results

Next, we ran the SBS feature selection on all 1859 feature candidates composed of all BP, coherence, HFD, and KFD features, to examine if the feature fusion and selection strategy could achieve high 5-fold CV accuracy on the training set and satisfactory accuracy on the test set. This procedure was performed separately for the three classifiers, i.e., K-NN, LDA, and SVM, because this wrapper approach of SBS is classifier specific. Figure 5 shows the results. For each classifier, a 5-fold CV classification accuracy was obtained for a given number of features. As a result, the range of the number of features,  $N_f$ , was supposed to be between 1 and 1859. However, it would be unnecessary to show the change in the accuracy with respect to all possible numbers of features from 1859 to 1, because,

for the three classifiers, their corresponding numbers of best features were all less than 70. As a result, in Figure 5, we only show the accuracy curves within the range of  $N_f \in [1, 99]$ , so that the subtle variation of accuracy within this range can be seen more clearly. Note that the number of features  $N_f$  is in fact identical to the dimension  $n$  of an input data described in Section 3.5.2 (LDA) and Section 3.5.3 (SVM).



**Figure 5.** Plots of classification accuracy (vertical) versus the number of features determined by the SBS-based feature selection procedure (horizontal). The solid curves represent the MDD-HC classification accuracy obtained by performing the 5-fold cross validation on the training set using different classifiers. The dotted curves represent the classification accuracy on the test set.

There are four observations shown in Figure 5. (1) In terms of 5-fold CV classification accuracy performed on the training set, the best accuracy of the K-NN, LDA, and SVM classifiers are 66.43%, 88.21%, and 87.50%, respectively. Furthermore, the best numbers of features for the three classifiers are 33, 63, and 62, respectively. LDA achieved the highest CV classification accuracy. (2) However, when applying the trained models (the classifiers trained with the corresponding optimal feature subsets) to the test set of 120 independent participants (each misclassified participant results in a  $0.83\% = 100/120$  accuracy drop), on the one hand, the LDA classifier's performance dropped significantly from 88.21% to 69.71%, indicating poor generalization. SVM, on the other hand, outperformed the others and can maintain an MDD-HC classification accuracy close to 80% (77.5%) on the test set. K-NN performed the worst of the three classifiers, with a testing accuracy of 48.33%. (3) The feature subset achieving the highest 5-fold SV classification accuracy does not necessarily guarantee the best performance when confronted with unseen data. For example, as  $N_f = 51$ , although SVM's 5-fold CV training accuracy of 86.07% was not the best, it achieved the highest classification accuracy of 80.83% on the unseen test set (i.e., 97 out of the 120 participants in the test set were correctly classified). (4) According to the 5-fold CV accuracy on the training set, the SVM's optimal feature subset determined by the SBS search consists of 62 features, which are all coherence features except one beta-band BP feature of the Pz electrode. Figure 6 depicts the 61 optimal coherence features in various bands, suggesting that the functional connectivity in all bands is critical in classifying MDD and HC groups when the SVM is used as a classifier. Furthermore, coherence features make up the majorities of the 33 and 63 optimal features for K-NN and LDA classifiers (see Figures A1, A2, and A3 for a detailed feature performance ranking of the optimal features for the three classifiers in Appendix A). Table 5 shows the numbers and the percentages of BP, COH, HFD, and KFD in the optimal feature subset for each classifier. It





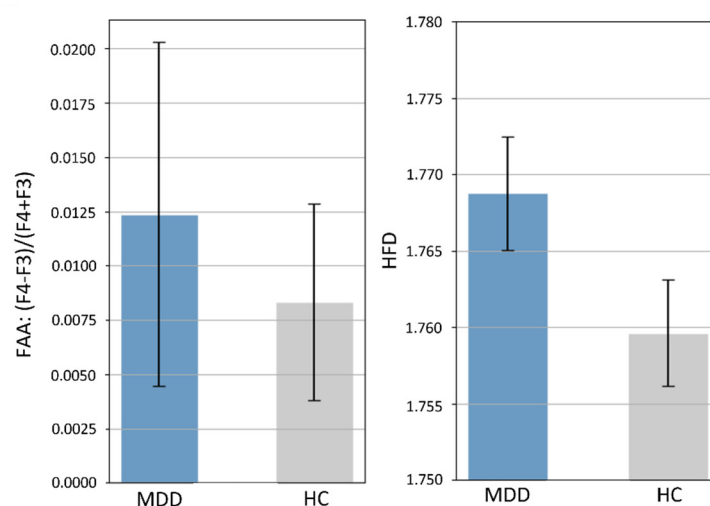
outperformed conventional SVM in the MDD-HC classification. For example, in the case of 5-fold CV performed on the training set, the CK-SVM outperformed the SVM in both conditions of  $N_f = 51$  and  $N_f = 62$  by 5% (91.07–86.07%) and 1.78% (89.28–87.50%), respectively.

**Table 6.** Comparison of training and testing classification accuracy among different classifiers (in %).

Classifier (number of Features)	Training Set		Test Set	
	5-Fold CV Accuracy	(Accuracy	Sensitivity	Specificity)
K-NN ( $N_f = 33$ )	$66.43 \pm 7.79$	48.33	48.33	48.33
LDA ( $N_f = 63$ )	$88.21 \pm 5.60$	69.17	75.00	63.33
SVM ( $N_f = 51$ )	$86.07 \pm 4.71$	80.83	86.67	75.00
SVM ( $N_f = 62$ )	$87.50 \pm 4.92$	77.50	85.00	70.00
CK-SVM ( $N_f = 51$ )	$91.07 \pm 3.43$	84.16	88.33	80.00
CK-SVM ( $N_f = 62$ )	$89.28 \pm 3.29$	80.83	88.33	73.33

#### 4.4. Comparison with Previous Literature: Statistical Analysis on Frontal Alpha Asymmetry (FAA) and HFD Complexity

Finally, we compared our results to previous research using two commonly used features: the frontal alpha asymmetry (FAA) and HFD complexity. Specifically, we used a Student's t-test to statistically test for group differences in FAA and HFD complexity of EEG from all electrodes (mean of 26 electrodes' HFD) on the 400 participants. Figure 7 shows that the average FAA in the MDD group ( $0.0124 \pm 0.1123$ ) was higher than that in the HC group ( $0.0083 \pm 0.064$ ), but the difference was not statistically significant ( $p = 0.66$ ). There was also no significant difference in HFD-based EEG complexity between groups (MDD,  $1.7688 \pm 0.0522$  and HC:  $1.7596$ ,  $p = 0.07$ ).



**Figure 7.** Means and 95% error bars (standard error of the mean) for MDD and HC participants. Left: frontal alpha asymmetry (FAA) calculated by the ratio  $(F4 - F3)/(F4 + F3)$  of alpha power. Right: EEG complexity represented by the HFD with  $k_{max} = 50$ . There were no significant differences in FAA ( $p = 0.66$ ) and HFD-based complexity ( $p = 0.07$ ) between the MDD and HC groups.

## 5. Discussion

In this section, we compare our results with previous studies in terms of the four EEG markers: FAA, HFD-based complexity, BP, and coherence features. It should be noted that a direct comparison between previous findings and ours could be a bit unfair, because there are many differences in terms of EEG protocols and, particularly, the nature of EEG

datasets. Nevertheless, such comparison can still provide important insight into some worthwhile discussion issues.

**FAA.** Several studies have suggested that the FAA of resting-state EEG could be a neurosignature for the diagnosis of MDD. For example, Debener et al., 2000 (15 MDD, 22 HC) [56] and Roh et al., 2020 (67 MDD, 60 HC) [57] reported a significant difference in FAA between the MDD and HC groups. Our result is inconsistent with theirs, but is consistent with the findings of Knott et al., 2001 (69 MDD, 23 HC) [32], Vinne et al., 2017 (1008 MDD, 336 HC) [58], and Smith et al., 2018 (143 MDD, 163 HC) [59]. However, some variables may influence the FAA, such as the severity of depression. In the study by Roh et al. [57], 23 of the 67 patients with MDD had suicidal ideation, which has been shown to be a clinically important moderator of FAA in patients with MDD.

**Complexity of EEG Activity.** There is substantial evidence showing that increased complexity of resting-state EEG activity can be measured in MDD (e.g., [23,60,61], see [62] for a review on the EEG complexity in MDD detection). In addition, the EEG HFD complexity has also been shown to be capable of producing a high MDD-HC classification [61]. Although our results also show increased HFD values of individuals with MDD as compared with HC (no statistical significance), the HFD feature did not provide satisfactory classification performance (maximum 62.86% based on LDA, see Table 4) in the current study. This low and inconsistent (as compared with previous works) accuracy level could be due to a number of factors. One possible factor is the condition of data collection; our EEG data were collected during the eye-open resting-state, whereas previous studies [23,61] collected EEGs during the eye-closed resting-state. For example, using HFD as features, a study by Hosseinifard et al. [23] showed a LOO-CV accuracy of 73.3% with the LDA classifier, whereas a study by Čukić et al. [61] showed an average accuracy of 89.90% across different classifiers (multilayer perception, logistic regression, SVM, decision tree, random forest, and Naïve Bayes). Other factors such as differences in the EEG amplifiers (dB, gain, and input noise specifications) and the signal preprocessing methods (band-pass filtering, artifact removal, signal epoch length, etc.) may also affect the signal-to-noise ratio of EEG signals. As a result, due to these possible factors and the fact that HFD is sensitive to noise [48], the effectiveness of HFD in detecting MDD may differ. More importantly, the current study collected EEG data from four different medical centers, whereas previous studies only collected EEG data from a single location. The high diversity of the current EEG dataset may result in highly variable noise levels.

**BP.** A few previous studies have suggested that alpha BP could be an MDD indicator. Our previous work [20] showed a nearly 70% (68.98%) classification accuracy using all-electrode alpha BP features and a K-NN classifier. However, in the current study, the all-electrode alpha BP (without feature selection) could only achieve an accuracy of 47.50% (Table 4). Another study applied an LDA-based classifier to the alpha BP features at C3, P3, O1, O2, and F7 [23] and reported a classification accuracy of 73%. In our study, the accuracy of the alpha BP remained below 60% (58.21%) even after feature selection. A discrepancy in alpha BP's ability to classify appears to exist between studies. The empirical results of this study suggested that including the BP feature in the EEG protocols alone may not be an accurate and reliable way to detect MDD.

Intra-participant variability in EEG amplitudes could contribute to the inconsistent results in the classification performance of BP. Often, every EEG study would keep the impedance of every electrode below a desired level before recordings. However, in practice, maintaining impedance stability throughout the recording period is quite challenging. The slight loss of an electrode could cause the impedance to change, resulting in changes in EEG amplitudes. Note that in some studies, noisy EEG segments were often removed from analyses (e.g., [18]). In the current study, however, we did not adopt this data cleaning strategy because, for the current study, an important objective was to verify the effectiveness of a feature in MDD detection, and also the robustness of a feature that could overcome the technical instability that may occur in real-world applications. Moreover, maintaining a stable impedance/signal quality across multiple recording sites is even

more challenging [28]. As a result, both intra-subject and inter-site EEG variabilities could have contributed to the poor performance of the BP feature.

In the experiments, we did not perform any EEG normalization because, as previously noted, the aim was to test the performance of the commonly used EEG features in previous MDD studies and, for example, the band powers in those studies are absolute powers without any normalization. However, it is worthwhile to investigate whether performing an EEG normalization can affect the BP performance. To this end, we normalize the power of each channel by subtracting the average power of the 26 channels in each frequency band, which can be viewed as a BP-based common average referencing (CAR). Table 7 summarizes the findings. As compared with the BP performance listed in Table 4 (the best performance is 64.64% without normalization), the normalized BP achieved comparable performance (the best is 63.57%). This comparison shows that there is almost no difference after using the CAR normalization method. One possible reason is as follows.

**Table 7.** Five-fold cross validation accuracies among different frequency bands and scalp regions using a normalized BP and a LDA classifier (in %). The numbers inside and outside the brackets denote the accuracies obtained from the features with and without the SBS-based feature selection.

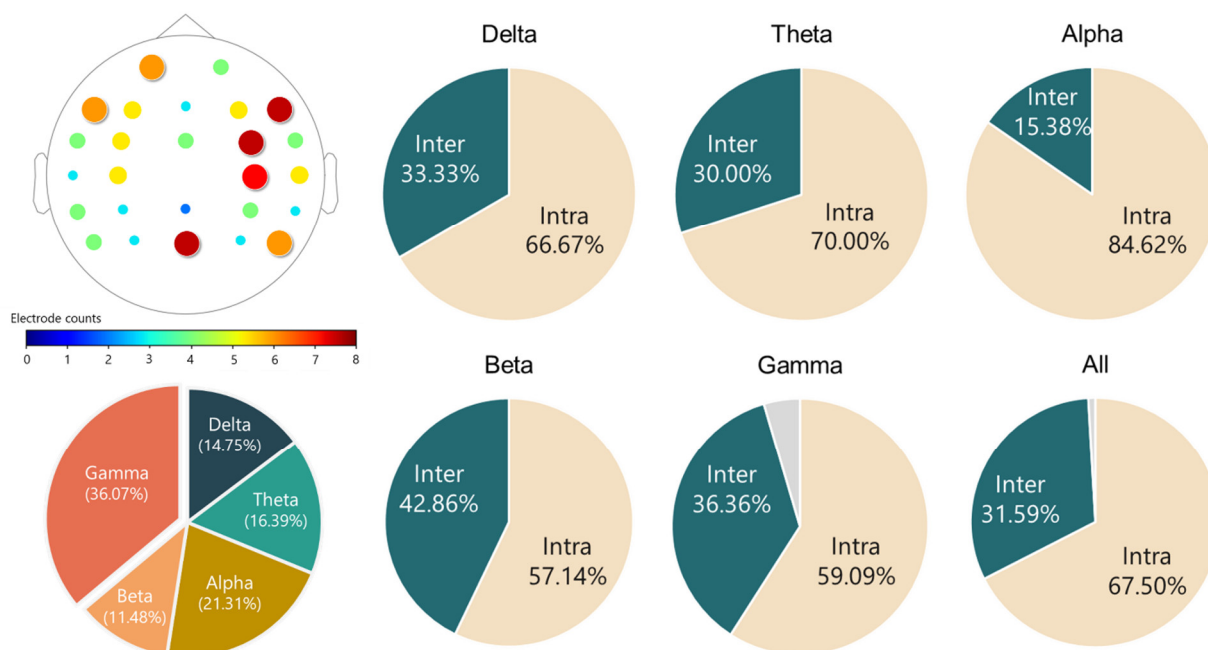
EEG Features		Frontal	Central	Temporal	Parietal	ALL
Normalized BP	$\delta$	52.14 (56.43)	57.14 (58.57)	53.35 (58.21)	42.86 (51.43)	50.00 (63.57)
	$\theta$	53.93 (58.21)	52.86 (54.64)	57.14 (58.21)	43.93 (54.64)	54.29 (60.36)
	$\alpha$	51.43 (58.57)	51.43 (52.86)	46.43 (53.93)	46.43 (53.93)	49.29 (57.50)
	$\beta$	54.29 (56.07)	56.43 (57.86)	56.07 (57.86)	51.43 (55.00)	52.50 (62.86)
	$\gamma$	54.29 (58.21)	59.29 (59.29)	50.00 (56.43)	46.79 (55.00)	48.21 (63.21)

The topographical BP distribution of EEG signals are mainly determined by the relative differences in BP between channels. Although the BP-based CAR normalization may reduce the variability caused by possible baseline BP differences between subjects, it does not affect the EEG BP distribution over the scalp for each individual because the average power is subtracted equally from all channels. This was nicely depicted as the effect of the rising or receding water levels of a lake in a mountainous area, which changed the location of the zero water level mark, but not the landscape [63]. For example, in the situation where one electrode's impedance changes during the EEG recording process, the BP will still vary even if such normalization is performed afterwards. Unfortunately, this happens in real-world applications, as mentioned above.

*Coherence.* The coherence (more precisely, the spectral coherence) has long been proposed as a fundamental mechanism for communication within brains [64] and is not affected by the amplitude changes of EEG oscillations [65], and therefore this property plays a key role in improving classification in the dataset with large intra-participant and inter-site EEG variability [66]. This could explain why 61 of the 62 optimal features selected from all 1859 feature candidates are all coherence features in the current study.

Previous literature has shown that MDD affects the functional connectivity of the brain in the resting state, but these findings were still inconclusive. For example, both functional MRI [67] and EEG [68] studies showed that MDD participants had increased connectivity in different brain regions, whereas some EEG studies [69,70] reported that a decreased functional connectivity was found in patients with MDD. This study combined all these coherence features as a feature vector for the MDD-HC classification, and therefore we did not further test whether each coherence had a significant difference between the two groups (MDD versus HC). Recently, Peng et al. [17] classified MDD and HC using the EEG connectivity features from four different bands (delta, theta, alpha, and beta). They concluded that the connectivity features of the four frequency bands are all crucial to the high-performance classification. Their results also suggested that depression affects the connectivity across the whole brain, as nearly all electrode pairs were involved in the connectivity-based classification. Our findings are in line with theirs, as evidenced by the

61 optimal coherence features summarized in Figure 8. There are three main findings. First, all 26 electrodes were used in the 61 optimal features. Second, all of the frequency bands contributed to the optimal coherence feature set. Third, the intra-hemispheric coherence features make up the majority of the optimal coherence feature set.



**Figure 8.** Characteristics of the 61 optimal coherence features selected by the SBS and SVM classifier. Upper-left: The occurring times of each electrode in the 61 features. Bottom-left: Proportion of each band in the 61 features. Remaining: Proportions of the intra-hemispheric (Intra) and inter-hemispheric (Inter) coherence features in each band. The gray part represents the regional connectivity of the electrode pair FCz and Pz.

**Feature Selection.** In the current study, we applied the wrapper approach, i.e., the SBS, to identify the optimal feature subset among the 1859 feature candidates. However, there is a limitation of the SBS algorithm. It is unable to re-evaluate the utility of the features that have been discarded. In other words, the optimal feature subset selected by the SBS algorithm may not really be the optimal. Other greedy and randomized feature selection methods, such as the genetic algorithm (GA), can be used to overcome this limitation. It should also be noted that the optimal feature subset determined by SBS may differ from those identified by other methods like GA.

As previously noted, the aim of the current study is to identify the best features among the 1859 feature candidates. The obtained 62 optimal features are explainable, which have already been able to provide a useful reference for researchers looking to investigate the relationship between them and the pathological mechanism of the brain in the future. If a dimension reduction method is further applied to the 62 features to obtain a lower-dimensional feature representation, not only can the data be visualized, but also the classification accuracy can be improved. A variety of advanced methods, including principal component analysis (PCA) and its nonlinear version, i.e., the kernel PCA, manifold learning approaches: the locally linear embedding (LLE), *t*-distributed stochastic neighbor embedding (*t*-SNE), the uniform manifold approximation and projection (UMAP), and neural network approaches such as autoencoders and self-organizing map (SOM). Although such research is beyond the scope of this study, it is definitely a worthwhile topic for future research.

## 6. Conclusions

The main goal of the current study was to provide a systematic validation of the efficacy of common and standard EEG features used in previous ML-based MDD-HC classification studies on a relatively large and diverse dataset. To this end, we thoroughly and systematically analyzed these features at the electrode, region, and optimal feature subset levels. Our results revealed that the coherence feature is the most reliable, effective and generalizable solution for the EEG-based MDD detection in real-world applications, where there might be high data variability caused by uncontrollable or unavoidable conditions. We also showed that the CK-SVM, a variant of SVM, was able to improve the performance generalization of the challenging MDD-HC classification, achieving a high “individual-level” classification accuracy of 84% on a large and independent dataset. According to the promising results obtained on a large dataset across multiple EEG recording sites, we have demonstrated the effectiveness and applicability of the EEG- and ML-based computer-aided diagnosis (CADx) for MDD.

Several critical issues remain to be solved. One issue is the establishment of an age-optimized MDD-HC classification model. Since MDD and mild cognitive impairment (MCI) or early dementia share some behavior signs in their early stage (i.e., cognitive decline), diagnosis of geriatric MDD has been a challenging task in clinical practice. Therefore, to develop an age-optimized CADx that is sensitive to MDD symptoms and disregards potential signal contamination from aging-related changes will be the next challenge. In summary, our results provide a reliable and valuable reference for future research to develop a more accurate and efficient CADx system.

**Author Contributions:** Conceptualization, C.-T.W. and Y.-H.L.; study design, C.-T.W., Y.-H.L., S.-C.L., C.-K.C., S.-H.L., and C.-F.T.; data collection, H.-C.H., S.H., I.-M.C., S.-C.L., C.-K.C., C.L., S.-H.L., M.-H.C., C.-F.T., and C.-H.W.; methodology, Y.-H.L., C.-T.W., H.-C.H., S.H., and C.-H.W.; formal analysis, H.-C.H., S.H., Y.-H.L., C.-T.W., and C.-H.W.; data discussion, C.-T.W., Y.-H.L., H.-C.H., S.H., I.-M.C., S.-C.L., C.-K.C., C.L., S.-H.L., M.-H.C., C.-F.T., C.-H.W., L.-W.K., and T.-P.J.; writing—original draft preparation, Y.-H.L., C.-T.W., H.-C.H., and S.H.; writing—review and editing, C.-T.W., Y.-H.L., and T.-P.J.; supervision, Y.-H.L. and T.-P.J.; project administration, Y.-H.L. and C.-H.W. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by the HippoScreen Neurotech Corp. (Taipei, Taiwan).

**Institutional Review Board Statement:** The study was conducted according to the guidelines of the Declaration of Helsinki, and approved by the Ethics Committee of NTUH (IRB no 201908038RSC, 2019/11/13), of TVGH (2019-09-004C, 2019/10/21), and of CGMH (201901342A3, 2019/10/3).

**Informed Consent Statement:** Informed consent was obtained from all subjects involved in the study.

**Acknowledgments:** Chih-Ken Chen and Chemin Lin were partly supported in this work by Chang Gung Memorial Hospital Research Projects (CRRPG2K0012 and CRRPG2K0022).

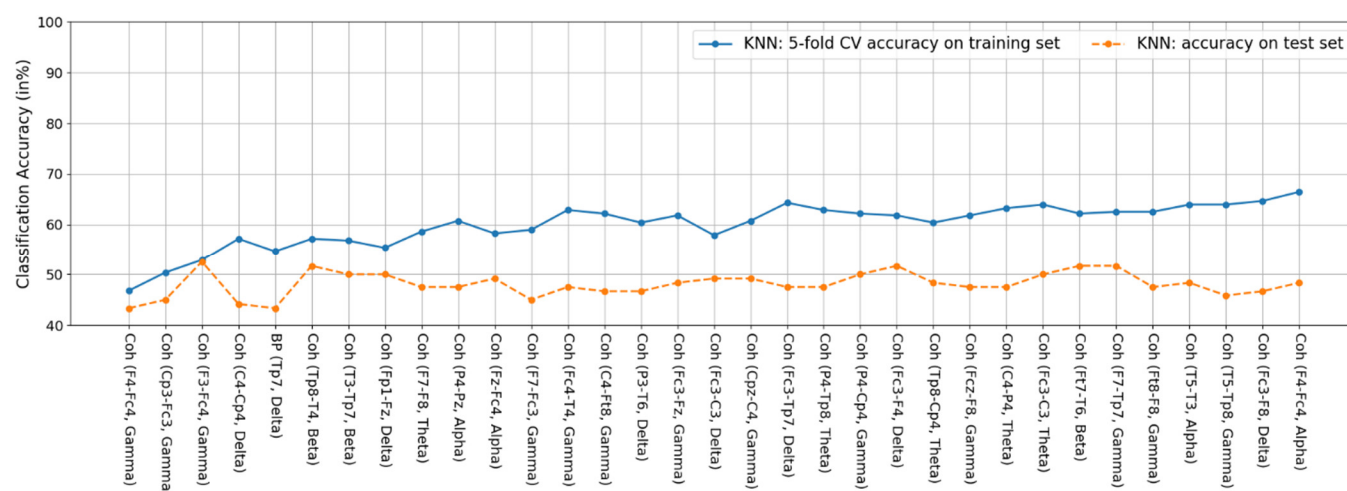
**Conflicts of Interest:** The authors declare no conflict of interest.

## Appendix A

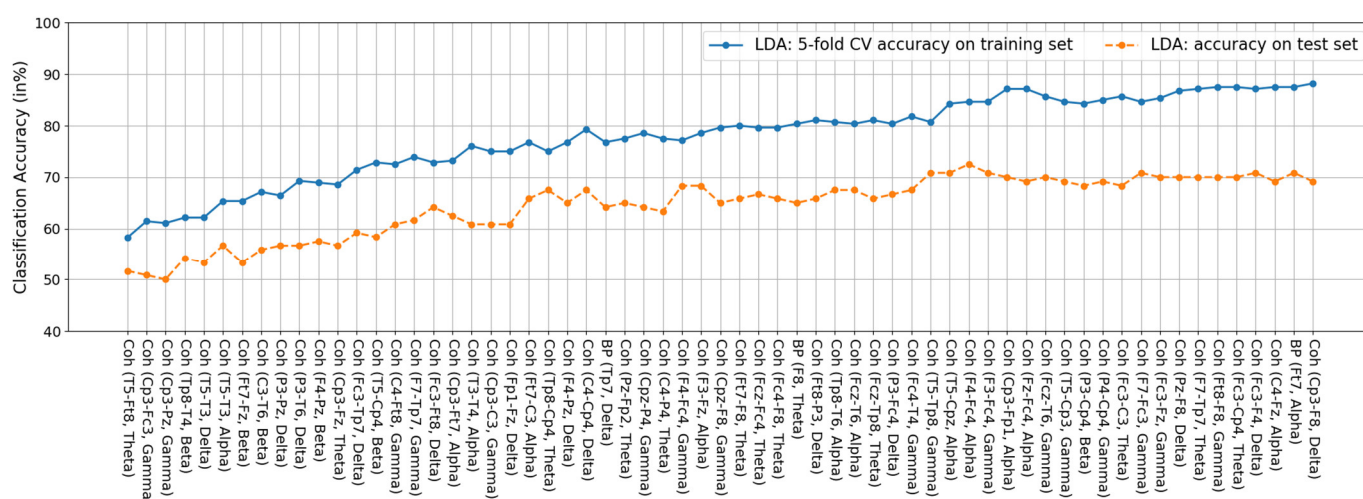
Here, we provide the feature performance ranking results of the features in the optimal feature subsets for the K-NN (33 optimal features), LDA (63 optimal features), and SVM (62 optimal features) classifiers in Figures A1–A3, respectively, where the accuracies are the same as those shown in Figure 5 (Section 4.2). For example, as the total 33 optimal features are used, the 5-fold CV accuracy (66.43) for the K-NN classifier is exactly the same as the one shown in Figure 5. It should also be noted that the feature performance ranking does not refer to the individual feature ranking, because the SBS algorithm is a wrapper feature selection approach, which does not rank each feature but evaluates the classification performance of a group of features. For example, in the case of the K-NN classifier, the accuracy associated with the second feature Coh (CP3-FC3, gamma) is the accuracy



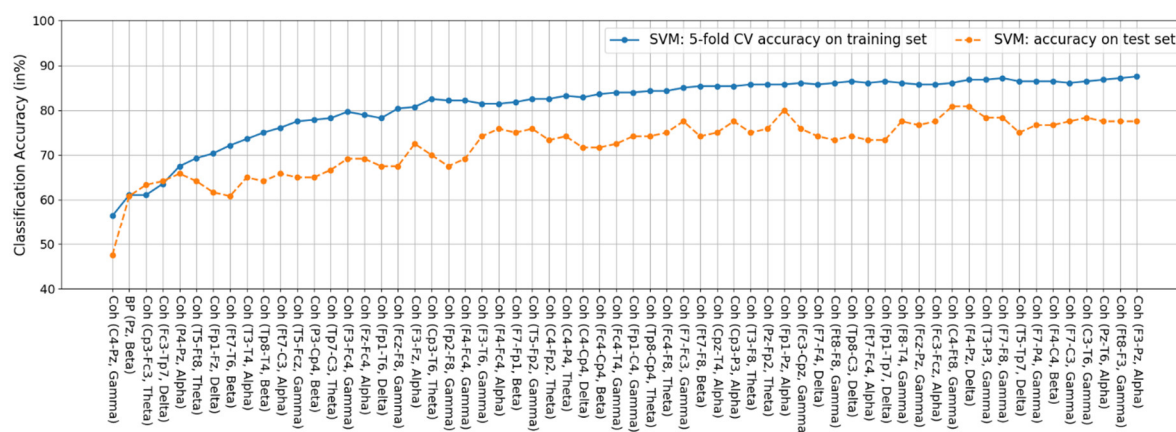
obtained by using the top two features together, i.e., Coh (F4-FC4, gamma) and Coh (CP3-FC3, gamma).



**Figure A1.** Features of the K-NN classifier. Among the 33 features, 32 are coherence features and the remaining one feature is the delta BP feature at TP7.



**Figure A2.** Features of the LDA classifier. Among the 63 features, 60 are coherence features and the remaining three are all BP features.



**FigureA3.** Features of the SVM classifier. Among the 62 features, 61 are coherence features and the remaining one is the beta BP feature at Pz.



## References

- Kupfer, D.J.; Frank, E.; Phillips, M.L. Major depressive disorder: New clinical, neurobiological, and treatment perspectives. *Lancet* **2012**, *379*, 1045–1055. [https://doi.org/10.1016/s0140-6736\(11\)60602-8](https://doi.org/10.1016/s0140-6736(11)60602-8).
- Friedrich, M.J. Depression Is the Leading Cause of Disability around the World. *JAMA* **2017**, *317*, 1517. <https://doi.org/10.1001/jama.2017.3826>.
- Greenberg, P.E.; Fournier, A.-A.; Sisitsky, T.; Simes, M.; Berman, R.; Koenigsberg, S.H.; Kessler, R.C. The Economic Burden of Adults with Major Depressive Disorder in the United States (2010 and 2018). *Pharmacoeconomics* **2021**, *39*, 653–665. <https://doi.org/10.1007/s40273-021-01019-4>.
- Vos, T.; Abajobir, A.A.; Abate, K.H.; Abbafati, C.; Abbas, K.M.; Abd-Allah, F.; Abdulkader, R.S.; Abdulle, A.M.; Abebo, T.A.; Abera, S.F.; et al. Global, regional, and national incidence, prevalence, and years lived with disability for 354 diseases and injuries for 195 countries and territories, 1990–2017: A systematic analysis for the Global Burden of Disease Study. *Lancet* **2017**, *392*, 1789–1858.
- World Health Organization (WHO). Depression and Other Common Mental Disorders: Global Health Estimates. 2017. Available online: <https://repository.gheli.harvard.edu/repository/11487/> (accessed on 5 December 2021).
- Proudman, D.; Greenberg, P.; Nellesen, D. The Growing Burden of Major Depressive Disorders (MDD): Implications for Researchers and Policy Makers. *Pharmacoeconomics* **2021**, *39*, 619–625.
- Goldberg, D. The heterogeneity of “major depression”. *World Psychiatry* **2011**, *10*, 226–228.
- Čukić, M.; López, V.; Pavón, J. Classification of Depression through Resting-State Electroencephalogram as a Novel Practice in Psychiatry: Review. *J. Med. Internet Res.* **2020**, *22*, e19548. <https://doi.org/10.2196/19548>.
- Bachmann, M.; Päske, L.; Kalev, K.; Aarma, K.; Lehtmets, A.; Ööpik, P.; Lass, J.; Hinrikus, H. Methods for classifying depression in single channel EEG using linear and nonlinear signal analysis. *Comput. Methods Programs Biomed.* **2018**, *155*, 11–17.
- Sharma, M.; Achuth, P.; Deb, D.; Puthankattil, S.D.; Acharya, U.R. An automated diagnosis of depression using three-channel bandwidth-duration localized wavelet filter bank with EEG signals. *Cogn. Syst. Res.* **2018**, *52*, 508–520. <https://doi.org/10.1016/j.cogsys.2018.07.010>.
- Mumtaz, W.; Ali, S.S.A.; Yasin, M.A.M.; Malik, A.S. A machine learning framework involving EEG-based functional connectivity to diagnose major depressive disorder (MDD). *Med. Biol. Eng. Comput.* **2018**, *56*, 233–246.
- Cai, H.; Chen, Y.; Han, J.; Zhang, X.; Hu, B. Study on Feature Selection Methods for Depression Detection Using Three-Electrode EEG Data. *Interdiscip. Sci. Comput. Life Sci.* **2018**, *10*, 558–565. <https://doi.org/10.1007/s12539-018-0292-5>.
- Mahato, S.; Paul, S. Classification of Depression Patients and Normal Subjects Based on Electroencephalogram (EEG) Signal Using Alpha Power and Theta Asymmetry. *J. Med. Syst.* **2020**, *44*, 28. <https://doi.org/10.1007/s10916-019-1486-z>.
- Saeedi, M.; Saeedi, A.; Maghsoudi, A. Major depressive disorder assessment via enhanced k-nearest neighbor method and EEG signals. *Phys. Eng. Sci. Med.* **2020**, *43*, 1007–1018. <https://doi.org/10.1007/s13246-020-00897-w>.
- Cai, H.; Han, J.; Chen, Y.; Sha, X.; Wang, Z.; Hu, B.; Yang, J.; Feng, L.; Ding, Z.; Chen, Y.; et al. A Pervasive Approach to EEG-Based Depression Detection. *Complexity* **2018**, *2018*, 5238028. <https://doi.org/10.1155/2018/5238028>.
- Akbari, H.; Sadiq, M.T.; Rehman, A.U.; Ghazvini, M.; Naqvi, R.A.; Payan, M.; Bagheri, H.; Bagheri, H. Depression recognition based on the reconstruction of phase space of EEG signals and geometrical features. *Appl. Acoust.* **2021**, *179*, 108078. <https://doi.org/10.1016/j.apacoust.2021.108078>.
- Peng, H.; Xia, C.; Wang, Z.; Zhu, J.; Zhang, X.; Sun, S.; Li, X. Multivariate pattern analysis of EEG-based functional connectivity: A study on the identification of depression. *IEEE Access* **2019**, *7*, 92630–92641.
- Movahed, R.A.; Jahromi, G.P.; Shahyad, S.; Meftahi, G.H. A major depressive disorder classification framework based on EEG signals using statistical, spectral, wavelet, functional connectivity, and nonlinear analysis. *J. Neurosci. Methods* **2021**, *358*, 109209.
- Čukić, M.; Stokić, M.; Simić, S.; Pokrajac, D. The successful discrimination of depression from EEG could be attributed to proper feature extraction and not to a particular classification method. *Cogn. Neurodyn.* **2020**, *14*, 443–455. <https://doi.org/10.1007/s11571-020-09581-x>.
- Liao, S.C.; Wu, C.T.; Huang, H.C.; Cheng, W.T.; Liu, Y.H. Major Depression Detection from EEG Signals Using Kernel Eigen-Filter-Bank Common Spatial Patterns. *Sensors* **2017**, *17*, 1385.
- Wu, C.-T.; Dillon, D.G.; Hsu, H.-C.; Huang, S.; Barrick, E.; Liu, Y.-H. Depression Detection Using Relative EEG Power Induced by Emotionally Positive Images and a Conformal Kernel Support Vector Machine. *Appl. Sci.* **2018**, *8*, 1244. <https://doi.org/10.3390/app8081244>.
- Mumtaz, W.; Xia, L.; Ali, S.S.; Mohd, Y.; Mohd, A.; Hussain, M.; Malik, A. Electroencephalogram (EEG)-based computer-aided technique to diagnose major depressive disorder (MDD). *Biomed. Signal Process. Control* **2017**, *31*, 108–115.
- Hosseinifard, B.; Moradi, M.H.; Rostami, R. Classifying depression patients and normal subjects using machine learning techniques and nonlinear features from EEG signal. *Comput. Methods Programs Biomed.* **2013**, *109*, 339–345. <https://doi.org/10.1016/j.cmpb.2012.10.008>.
- Jiang, C.; Li, Y.; Tang, Y.; Guan, C. Enhancing EEG-Based Classification of Depression Patients Using Spatial Information. *IEEE Trans. Neural Syst. Rehabil. Eng.* **2021**, *29*, 566–575. <https://doi.org/10.1109/tnsre.2021.3059429>.
- Liu, W.; Zhang, C.; Wang, X.; Xu, J.; Chang, Y.; Ristaniemi, T.; Cong, F. Functional connectivity of major depression disorder using ongoing EEG during music perception. *Clin. Neurophysiol.* **2020**, *131*, 2413–2422. <https://doi.org/10.1016/j.clinph.2020.06.031>.

26. Zhu, J.; Wang, Z.; Gong, T.; Zeng, S.; Li, X.; Hu, B.; Li, J.; Sun, S.; Zhang, L. An Improved Classification Model for Depression Detection Using EEG and Eye Tracking Data. *IEEE Trans. NanoBiosci.* **2020**, *19*, 527–537. <https://doi.org/10.1109/tnb.2020.2990690>.
27. Fried, E. Moving forward: How depression heterogeneity hinders progress in treatment and research. *Expert Rev. Neurother.* **2017**, *17*, 423–425. <https://doi.org/10.1080/14737175.2017.1307737>.
28. Farzan, F.; Atluri, S.; Frehlich, M.; Dhami, P.; Kleffner, K.; Price, R.; Lam, R.W.; Frey, B.N.; Milev, R.; Ravindran, A.; et al. Standardization of electroencephalography for multi-site, multi-platform and multi-investigator studies: Insights from the canadian biomarker integration network in depression. *Sci. Rep.* **2017**, *7*, 7473.
29. Zao, J.K.; Gan, T.-T.; You, C.-K.; Chung, C.-E.; Wang, Y.-T.; Rodríguez Méndez, S.J.; Mullen, T.; Yu, C.; Kothe, C.; Hsiao, C.-T.; et al. Pervasive brain monitoring and data sharing based on multi-tier distributed computing and linked data technology. *Front. Hum. Neurosci.* **2014**, *8*, 370. <https://doi.org/10.3389/fnhum.2014.00370>.
30. Yeh, S.C.; Hou, C.L.; Peng, W.H.; Wei, Z.Z.; Huang, S.; Kung, E.Y.C.; Lin, L.; Liu, Y.H. A multiplayer online car racing virtual-reality game based on internet of brains. *J. Syst. Archit.* **2018**, *89*, 30–40.
31. Yang, Y.; Truong, N.D.; Eshraghian, J.K.; Maher, C.; Nikpour, A.; Kavehei, O. A multimodal AI system for out-of-distribution generalization of seizure detection. *bioRxiv* **2021**, doi: <https://doi.org/10.1101/2021.07.02.450974>
32. Knott, V.; Mahoney, C.; Kennedy, S.; Evans, K. EEG power, frequency, asymmetry and coherence in male depression. *Psychiatry Res. Neuroimaging* **2001**, *106*, 123–140. [https://doi.org/10.1016/S0925-4927\(00\)00080-9](https://doi.org/10.1016/S0925-4927(00)00080-9).
33. Pudil, P.; Novovičová, J.; Kittler, J. Floating search methods in feature selection. *Pattern Recognit. Lett.* **1994**, *15*, 1119–1125. [https://doi.org/10.1016/0167-8655\(94\)90127-9](https://doi.org/10.1016/0167-8655(94)90127-9).
34. Wu, S.; Amari, S.-I. Conformal Transformation of Kernel Functions: A Data-Dependent Way to Improve Support Vector Machine Classifiers. *Neural Process. Lett.* **2002**, *15*, 59–67. <https://doi.org/10.1023/a:1013848912046>.
35. Liu, Y.-H.; Wu, C.-T.; Cheng, W.-T.; Hsiao, Y.-T.; Chen, P.-M.; Teng, J.-T. Emotion Recognition from Single-Trial EEG Based on Kernel Fisher's Emotion Pattern and Imbalanced Quasiconformal Kernel Support Vector Machine. *Sensors* **2014**, *14*, 13361–13388. <https://doi.org/10.3390/s140813361>.
36. Mumtaz, W.; Xia, L.; Yasin, M.A.M.; Ali, S.S.A.; Malik, A.S. A wavelet-based technique to predict treatment outcome for Major Depressive Disorder. *PLoS ONE* **2017**, *12*, e0171409. <https://doi.org/10.1371/journal.pone.0171409>.
37. Brownlee, J. Data Preparation for Machine Learning: Data Cleaning, Feature Selection, and Data Transforms in Python; Machine Learning Mastery: Melbourne, Victoria, Australia, 2020.
38. Roy, Y.; Banville, H.; Albuquerque, I.; Gramfort, A.; Falk, T.H.; Faubert, J. Deep learning-based electroencephalography analysis: A systematic review. *J. Neural Eng.* **2019**, *16*, 051001. <https://doi.org/10.1088/1741-2552/ab260c>.
39. American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders*, 5th ed.; DSM-5; American Psychiatric Publishing: Washington, DC, USA, 2013.
40. Lecrubier, Y.; Sheehan, D.V.; Weiller, E.; Amorim, P.; Bonora, I.; Sheehan, K.H.; Janavs, J.; Dunbar, G. The Mini International Neuropsychiatric Interview (MINI). A short diagnostic structured interview: Reliability and validity according to the CIDI. *Eur. Psychiatry* **1997**, *12*, 224–231. [https://doi.org/10.1016/S0924-9338\(97\)83296-8](https://doi.org/10.1016/S0924-9338(97)83296-8).
41. Kroenke, K.; Spitzer, R.L.; Williams, J.B. The PHQ-9: Validity of a brief depression severity measure. *J. Gen. Intern Med.* **2001**, *16*, 606–613.
42. Beck, A.T.; Steer, R.A.; Brown, G.K. *Manual for the Beck Depression Inventory-II*; Psychological Corporation: San Antonio, TX, USA, 1996.
43. Zigmond, A.S.; Snaith, R.P. The hospital anxiety and depression scale. *Acta Psychiatr. Scand.* **1983**, *67*, 361–370.
44. Belmaker, R.H.; Agam, G. Major Depressive Disorder. *N. Engl. J. Med.* **2008**, *358*, 55–68. <https://doi.org/10.1056/nejmra073096>.
45. Ablin, P.; Cardoso, J.; Gramfort, A. Faster ICA under orthogonal constraint. Faster ICA under orthogonal constraint. In Proceedings of the International Conference on Acoustics, Speech & Signal Processing, Calgary, AB, Canada, 15–20 April 2018.
46. Stapleton, P.; Dispenza, J.; McGill, S.; Sabot, D.; Peach, M.; Raynor, D. Large effects of brief meditation intervention on EEG spectra in meditation novices. *IBRO Rep.* **2020**, *9*, 290–301. <https://doi.org/10.1016/j.ibror.2020.10.006>.
47. Benoit, O.; Daurat, A.; Prado, J. Slow (0.7–2 Hz) and fast (2–4 Hz) delta components are differently correlated to theta, alpha and beta frequency bands during NREM sleep. *Clin. Neurophysiol.* **2000**, *111*, 2103–2106.
48. Esteller, R.; Vachtsevanos, G.; Echauz, J.; Litt, B. A comparison of waveform fractal dimension algorithms. *IEEE Trans. Circuits Syst. I Regul. Pap.* **2001**, *48*, 177–183. <https://doi.org/10.1109/81.904882>.
49. Higuchi, T. Approach to an irregular time series on the basis of the fractal theory. *Phys. D Nonlinear Phenom.* **1988**, *31*, 277–283. [https://doi.org/10.1016/0167-2789\(88\)90081-4](https://doi.org/10.1016/0167-2789(88)90081-4).
50. Phothisonothai, M.; Nakagawa, M. Fractal-Based EEG Data Analysis of Body Parts Movement Imagery Tasks. *J. Physiol. Sci.* **2007**, *57*, 217–226. <https://doi.org/10.2170/physiolsci.rp006307>.
51. Liu, Y.-H.; Huang, S.; Huang, Y.-D. Motor Imagery EEG Classification for Patients with Amyotrophic Lateral Sclerosis Using Fractal Dimension and Fisher's Criterion-Based Channel Selection. *Sensors* **2017**, *17*, 1557. <https://doi.org/10.3390/s17071557>.
52. Katz, M.J. Fractals and the analysis of waveforms. *Comput. Biol. Med.* **1988**, *18*, 145–156. [https://doi.org/10.1016/0010-4825\(88\)90041-8](https://doi.org/10.1016/0010-4825(88)90041-8).
53. Liu, Y.H.; Liu, Y.C.; Chen, Y.J. Fast support vector data descriptions for novelty detection. *IEEE Trans. Neural Netw.* **2010**, *21*, 1296–1313.

54. Fang, L.; Zhao, H.; Wang, P.; Yu, M.; Yan, J.; Cheng, W.; Chen, P. Feature selection method based on mutual information and class separability for dimension reduction in multidimensional time series for clinical data. *Biomed. Signal Process. Control* **2015**, *21*, 82–89. <https://doi.org/10.1016/j.bspc.2015.05.011>.
55. Jović, A.; Brkić, K.; Bogunović, N. A review of feature selection methods with applications. In Proceedings of the 2015 38th International Conference on Information and Communication Technology, Electronics and Microelectronics, Opatija, Croatia, 25–29 May 2015.
56. Debener, S.; Beauducel, A.; Nessler, D.; Brocke, B.; Heilemann, H.; Kayser, J. Is resting anterior EEG alpha asymmetry a trait marker for depression? Findings for healthy adults and clinically depressed patients. *Neuropsychobiology* **2000**, *41*, 31–37.
57. Roh, S.-C.; Kim, J.S.; Kim, S.; Kim, Y.; Lee, S.-H. Frontal Alpha Asymmetry Moderated by Suicidal Ideation in Patients with Major Depressive Disorder: A Comparison with Healthy Individuals. *Clin. Psychopharmacol. Neurosci.* **2020**, *18*, 58–66. <https://doi.org/10.9758/cpn.2020.18.1.58>.
58. van der Vinne, N.; Vollebregt, M.; van Putten, M.; Arns, M. Frontal alpha asymmetry as a diagnostic marker in depression: Fact or fiction? A meta-analysis. *NeuroImage Clin.* **2017**, *16*, 79–87.
59. Smith, E.E.; Cavanagh, J.F.; Allen, J.J.B. Intracranial source activity (eLORETA) related to scalp-level asymmetry scores and depression status. *Psychophysiology* **2018**, *55*, e13019. <https://doi.org/10.1111/psyp.13019>.
60. Bachmann, M.; Lass, J.; Suhhova, A.; Hinrikus, H. Spectral asymmetry and Higuchi's fractal dimension measures of depression electroencephalogram. *Comput. Math. Methods Med.* **2013**, *2013*, 251638.
61. Čukić, M.; Pokrajac, D.; Stokić, M.; Simić, S.; Radivojević, V.; Ljubisavljević, M. EEG machine learning with Higuchi fractal dimension and sample entropy as features for successful detection of depression. *arXiv* **2018**, arXiv:1803.05985.
62. Čukić, M.; Stokić, M.; Radenković, S.; Ljubisavljević, M.; Simić, S.; Savić, D. Nonlinear analysis of EEG complexity in episode and remission phase of recurrent depression. *Int. J. Methods Psychiatr. Res.* **2019**, *29*, e1816. <https://doi.org/10.1002/mpr.1816>.
63. Chella, F.; Pizzella, V.; Zappasodi, F.; Marzetti, L. Impact of the reference choice on scalp EEG connectivity estimation. *J. Neural Eng.* **2016**, *13*, 036016. <https://doi.org/10.1088/1741-2560/13/3/036016>.
64. Fries, P. A mechanism for cognitive dynamics: Neuronal communication through neuronal coherence. *Trends Cogn. Sci.* **2005**, *9*, 474–480. <https://doi.org/10.1016/j.tics.2005.08.011>.
65. Nunez, P.L. *Electric Fields of the Brain: The Neurophysics of EEG*; Oxford University Press: New York, USA, 2006.
66. La Rocca, D.; Campisi, P.; Vegso, B.; Cserti, P.; Kozmann, G.; Babiloni, F.; Fallani, F.D.V. Human Brain Distinctiveness Based on EEG Spectral Coherence Connectivity. *IEEE Trans. Biomed. Eng.* **2014**, *61*, 2406–2412. <https://doi.org/10.1109/tbme.2014.2317881>.
67. Sheline, Y.I.; Price, J.L.; Yan, Z.; Mintun, M.A. Resting-state functional MRI in depression unmasks increased connectivity between networks via the dorsal nexus. *Proc. Natl. Acad. Sci. USA* **2010**, *107*, 11020–11025. <https://doi.org/10.1073/pnas.1000446107>.
68. Tas, C.; Cebi, M.; Tan, O.; Hızlı-Sayar, G.; Tarhan, N.; Brown, E. EEG power, cordance and coherence differences between unipolar and bipolar depression. *J. Affect. Disord.* **2015**, *172*, 184–190. <https://doi.org/10.1016/j.jad.2014.10.001>.
69. Connolly, C.; Wu, J.; Ho, T.C.; Hoeft, F.; Wolkowitz, O.; Eisendrath, S.; Frank, G.; Hendren, R.; Max, J.E.; Paulus, M.; et al. Resting-State Functional Connectivity of Subgenual Anterior Cingulate Cortex in Depressed Adolescents. *Biol. Psychiatry* **2013**, *74*, 898–907. <https://doi.org/10.1016/j.biopsych.2013.05.036>.
70. Alalade, E.; Denny, K.; Potter, G.; Steffens, D.; Wang, L. Altered cerebellar-cerebral functional connectivity in geriatric depression. *PLoS ONE* **2011**, *6*, e20035.