Diagnostic Machine Learning Approaches on Clinical Populations and Clinical Conditions using Functional Near Infrared Spectroscopy : A Review

Aykut Eken[1], Farhad Nassehi[1], Osman Eroğul[1]

1. Biomedical Engineering Department, TOBB University of Economics and Technology, Ankara, Turkey

Corresponding Author

Aykut Eken

TOBB University of Economics and Technology

Faculty of Engineering

Biomedical Engineering Department

Söğütözü, Söğütözü St. No:43

06510

Çankaya / Ankara

Tel: +90 312 292 40 00 / 4268

Fax : +90 312 287 19 46

e-mail: aykuteken@etu.edu.tr

Abstract

Objective : There is a growing interest in functional near-infrared spectroscopy (fNIRS) among researchers due to its remarkable advantages such as ease of use, being inexpensive and less tolerance to motion artifacts compared to other neuroimaging modalities. Also, its interaction with machine learning (ML) approaches was inevitable like other neuroimaging modalities for different purposes such as diagnostic classification of diseases or prediction of disease severity due to the lack of robust and objective biomarkers. A review of literature is carried out to understand the evolution of biomarker research on clinical populations and clinical states by combining fNIRS and ML.

Approach : In this review, article search was carried out in accordance with the Preferred Reporting Items for Systematic reviews and Meta-Analyses (PRISMA) standard and 62 studies were evaluated and we provided a general overview by using fNIRS, particularly in clinical populations and some clinically relevant conditions on healthy populations. Also, potential biomarkers that were found in these studies and some popular ML algorithms that had been used for the prediction or classification of fNIRS data were discussed.

Main Results : There is an increasing trend to perform ML applications on fNIRS data on biomarker research related to different clinical fields. Among these studies, few were able to have a notable number of participants for classification and clinical state prediction. Oxy-hemoglobin was used more than deoxy-hemoglobin in ML-based studies as a potential feature source.

Significance : Using ML on fNIRS data might be a promising approach to revealing specific biomarkers for either diagnostic classification of diseases or prediction of clinical conditions.

Keywords: fNIRS, Machine Learning, Psychiatry, Neurological, Biomarkers

1. Introduction

Functional neuroimaging approaches are used to understand functional responses either against a stimulus or during resting-condition. However, in recent years, their usage as a diagnostic tool is widely discussed (Henderson et al., 2020). In this perspective, functional neuroimaging approaches such as Functional Magnetic Resonance Imaging (fMRI), Electroencephalography (EEG), Magnetoencephalography (MEG), Positron Emission Tomography (PET), Functional Near Infrared Spectroscopy (fNIRS) provide functional brain data that can be utilized to discriminate diseases that have common symptoms and diseases from healthy individuals (Gao et al., 2018).

There is a strong interest in recent years in machine learning applications in neuroimaging. Previous reviews that cover a combination of ML techniques for the prediction of several diseases by using EEG (see review (Craik et al., 2019)), resting-state fMRI (see review (Khosla et al., 2019)) and PET (see review (Duffy et al., 2019)) showed that neuroimaging techniques might have a future on individual diagnosis with ML. Machine learning applications in neuroimaging allow researchers to investigate clinical populations or clinical conditions from different perspectives. Because, compared to conventional statistical approaches such as t-test, ANOVA, Kruskal-Wallis or Friedman test. Machine learning provides us individual or trial-level answers rather than average sense. This is quite remarkable in medicine. Because, (i) Many diseases/disorders/syndromes have common symptoms that make them complicated to distinguish each other (ii) While diagnosing them, self-reporting of patients which is the conventional approach and also gold-standard for diagnosis of many disorders, might provide unreliable results due to having the potential to be easily manipulated. Therefore, there is a great necessity to reveal robust and objective biomarkers that provides individual accurate diagnosis (iii) In general, vast majority of behavioral and neuroimaging studies that focus on differences between patients and healthy individuals show these differences in average sense. However, these differences might not be valid for some individual cases due to huge variability across participants. At this point, the combination of neuroimaging approaches and ML techniques plays an important role to provide us some answers related to individual diagnoses rather than populations.

Among these functional neuroimaging techniques, fNIRS is relatively new and promising approach due to its advantages (see reviews (Baskak, 2018; Ehlis et al., 2014; Gao et al., 2018; Irani et al., 2007)) and it has almost a contemporary history with artificial intelligence applications in medicine. However, due to lack of data and computational cost, ML usage in fNIRS studies was limited until recent years. After overcoming these limitations, ML usage has increased greatly through the last decade among fNIRS researchers. Compared to other neuroimaging modalities such as fMRI and PET, it is less expensive, portable, easy to apply and has more tolerance to motion artifacts. When compared to EEG, it has higher spatial resolution that allows the researchers to focus on a specific region of interest (ROI). In addition to these advantages, it also provides information about concentration changes of oxy-hemoglobin (ΔHbO), deoxy-hemoglobin (ΔHb) and total-hemoglobin (ΔHbT= ΔHbO + ΔHb) by using at least two different wavelengths. These advantages feature fNIRS as a potential alternative tool for the diagnosis of psychiatric diseases. It has widely been preferred by researchers and clinicians from many different fields such as infant development, cognition, anesthesia, motor control and psychiatric disorders (see review (Boas et al., 2014)).

Systematically, an fNIRS-based ML system consists of several components as it is shown in Figure 1. A specific task or a resting-state procedure is conducted for data acquisition via a multi or single-channel fNIRS system. After data acquisition, a pre-processing step is carried out. In pre-processing step, several types of artifacts such as physiological noise (heartbeat, respiration, mayer waves (Fekete et al., 2011a)), motion artifacts and very low-frequency noise (<0.1 Hz) need to be filtered out. For this purpose, band-pass filtering, signal detrending and motion artifact algorithms (Brigadoi et al., 2014) are used. Having carefully filtered the data, feature extraction is carried out. Feature extraction step directly affects the performance of classifiers. Due to this reason, *a priori* knowledge in either temporal or spatial behavior of hemodynamic response might be essential. Depending on the type of data (resting-state or task), extracted feature types might be different. Feature selection should also be carried out if the number of features is high. This may lead to a dimensionality problem which may cause an overfitting or underfitting problem. In this step, there are several algorithms that might be used such as Principal Component Analysis (PCA), Least Absolute Shrinkage and Selection Operator (LASSO), t-test and Recursive Feature Elimination (RFE). Cross-validation types (Hold-Out, Leave-one-out (LOOCV) and K-fold) are generally selected depending on the amount of data. In some studies, hyperparameter optimization techniques such as grid-search, random-search or Bayesian are used to improve the performance of classifiers or predictors. For classification or prediction, methods such as Support Vector Machine (SVM), K-nearest neighborhood (KNN), linear discriminant analysis (LDA), Gaussian process classifier (GPC), Random Forest (RaF), Linear regression (LR) and Convolutional Neural Network (CNN) as a deep learning model are used.

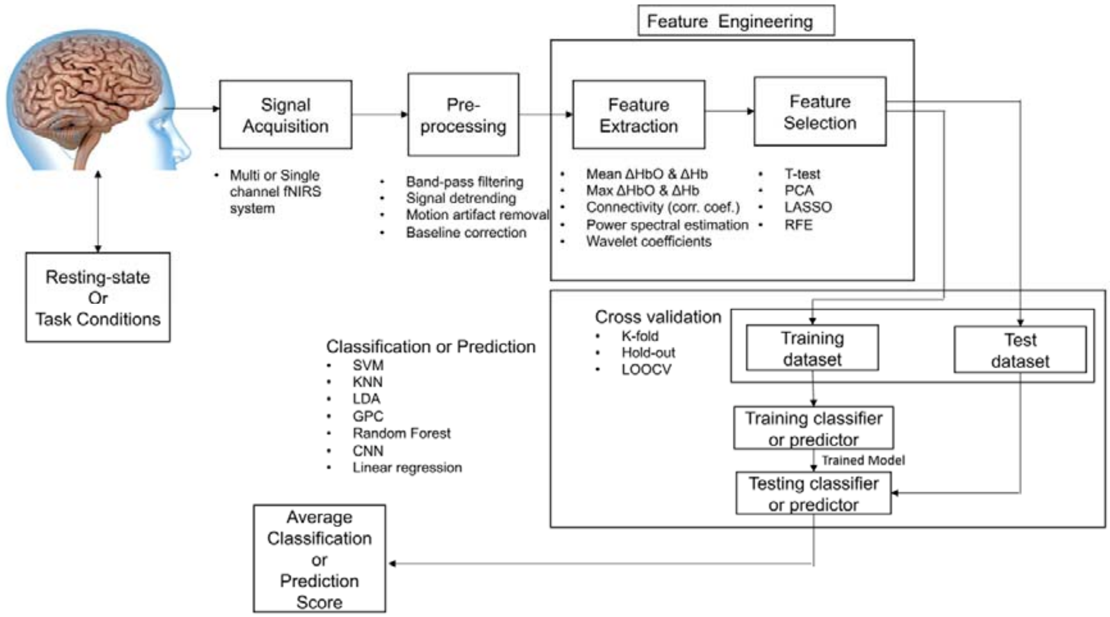

Feature Engineering
Signal Acquisition
Pre-processing
Feature Extraction
Feature Selection
Multi or Single channel fNIRS system
Band-pass filtering
Signal detrending
Motion artifact removal
Baseline correction
Mean ΔHbO & ΔHb
Max ΔHbO & ΔHb
Connectivity (corr. coef.)
Power spectral estimation
Wavelet coefficients
T-test
PCA
LASSO
RFE
Resting-state Or Task Conditions
Cross validation
K-fold
Hold-out
LOOCV
Training dataset
Test dataset
Training classifier or predictor
Trained Model
Testing classifier or predictor
Classification or Prediction
SVM
KNN
LDA
GPC
Random Forest
CNN
Linear regression
Average Classification or Prediction Score

Figure 1. A framework for Classification or Prediction of Clinical Populations or Clinical Conditions using fNIRS

Our primary objective to review fNIRS-based ML studies is to emphasize the potential of fNIRS for developing new clinical biomarkers. We also provide insights into the potential problems for diagnostic usage of fNIRS and suggest questions for further studies. This review includes a general overview of these applications on clinical populations, applications for clinical conditions such as pain and anesthesia monitoring, also some popular methods being used in these studies and discussion about the future of fNIRS in biomarker research. To our best knowledge, this is the first review that covers machine learning studies focusing on biomarker research using fNIRS. There is a recent review focusing on deep learning (Eastmond et al., 2022), however, as we stated above we also discussed the regions and potential biomarkers used as a feature that provides high diagnostic accuracy.

2. Methodology

The present study was performed according to the "Preferred Reporting Items for Systematic reviews and Meta-Analyses" (PRISMA) statement (Page et al., 2021), shown as a schema in Figure 2. The search procedure was initiated by using Web of Science and PubMed databases. We used the keywords ("Functional Near Infrared Spectroscopy" OR "Near Infrared Spectroscopy" OR "Diffuse Optical Imaging") AND ("Machine Learning" OR "Prediction" OR "Classification") that describe in Table 1 in detail. Original research papers published from starting 2010 until end of September 2022 were included. A total of 1350 (Pubmed: 727, Web of Science:679) search results that were published in Science Citation Indexing and Science Citation Indexing-Expanded, were reached. After removing the duplicate results, 1350 articles were left. Articles Conference proceedings and reviews excluding, 1314 articles were left. We also excluded the non-clinical studies and brain-computer interface (BCI) studies and studies closely related to BCI such as motor and mental arithmetic tasks since it has been extensively reviewed by Naseer and Hong (Naseer & Hong, 2015).

----- Add Figure 2 Here----

----- Add Table 1 Here----

We scanned and reported 62 articles that were suitable for our context. All included studies are summarized in Table 2. Extracted data types from publications were first author and year of the publication, populations, objective of the study, experiment type (task/resting), used fNIRS system, region of interest with 10-20 position if available, sample size, used features to train and test the model, used machine learning algorithm, cross-validation technique, hyperparameter optimization type, obtained the highest accuracy, other classification scores and comments related to the study. Studies were grouped according to the focused clinical population or clinical condition. Some studies include two or more clinical populations. For these studies, we included them in the related section.

----- Add Table 2 Here----

3. Clinical Populations

In this review, we considered studies from many different clinical populations such as Attention Deficit and Hyperactivity Disorder (ADHD), Alzheimer's Disease (AD), Autism Spectrum Disorder (ASD), Amyotrophic Lateral Sclerosis (ALS), Depressive Disorder (MDD), Epilepsy, Fibromyalgia (FM), Mild Cognitive Impairment (MCI), Parkinson's Disease (PD), Schizophrenia (SCZ), Somatic Symptom Disorder (SSD), Stuttering and Traumatic Brain Injury (TBI). Among those clinical populations, the vast majority of studies are related to ADHD and SCZ populations. While explaining the studies, we focused on five main points related to the studies. These points are;

- Experimental design: We focused on the experimental paradigm due to inducing the specific response that leads to discriminate the clinical population or condition.
- The number of participants: It is one of the most critical factors of the reliability of the study.
- Extracted features on the fNIRS signal: Depending on the performance of the classifier, extracted features might be considered as a potential biomarker related to the clinical population or clinical condition.
- Focused Region/s: Related region/s that features were extracted from, to involve in the model.
- ML algorithm: Preferred ML algorithm to develop the diagnostic or predictive model.
- Model Performance: Result depends on all the factors above.

3.1. Attention Deficit and Hyperactivity Disorder (ADHD)

3.1.1. ADHD vs Healthy Controls (HC)

ADHD is a highly prevalent disorder particularly among children populations. It is mainly represented by lack of attention, excessive activity and impulsivity and its prevalence were reported as 3.6 % according to the DSM-V (American Psychiatric Association, 2013). ADHD patients are generally treated by medication (Chan et al., 2016; Geffen & Forster, 2018), psychotherapy (Gentile & Atiq, 2006) and neurofeedback systems (Razoki, 2018). In general, all these methods are used as complementary to each other. However, the underlying reason is still unclear. Previous, fNIRS studies proposed that Right Pre Frontal Cortex (R PFC) can be an objective biomarker in ADHD (Jourdan Moser et al., 2009; Monden et al., 2012). Considering this information, input data for classifiers were mainly extracted from the frontal region (Crippa et al., 2017; Gu et al., 2018; Ishii-Takahashi et al., 2015; Sutoko et al., 2019; Yasumura et al., 2017). Studies have generally low sample sizes (min-max: 44-50) except for Yasumura and colleagues (Yasumura et al., 2017). This study is a multi-center study performed to validate the reliability of a classifier. It includes the highest number of subjects (Training data; ADHD: 108, HC: 108. Validation data; ADHD: 62, HC: 37) among all ADHD classification studies using fNIRS. fNIRS data that was acquired from PFC via a reverse Stroop task from different centers were used as input data with behavioral and physiological features. 86.25 % accuracy was found by using Radial Basis Function (RBF)-SVM and reverse stoop task-induced PFC activation was suggested as a critical biomarker for ADHD diagnosis.

Among other ADHD studies, Ishii-Takahashi and her colleagues first tried to use the fNIRS to predict the efficiency of methylphenidate (MPH) on ADHD-diagnosed children by measuring inferior frontal cortex (IFC) during a stop signal task (SST) and they could be able to classify the groups that were previously separated according to the Clinical Global Impressions-

Severity (CGI-S) score (CGI-S >4 & CGI-S <3) after 4-8 weeks of MPH administration with 81% accuracy by using LDA (Ishii-Takahashi et al., 2015).

Crippa and his colleagues used both ΔHb and ΔHbO data together for ADHD classification (Crippa et al., 2017). In this study, a visuospatial N-back task was performed and multi-modal measurements were obtained including clinical measures, neurophysiological data and fNIRS signals (ΔHb and ΔHbO) from frontotemporal areas. The surprising result was that ΔHb-based features were significantly effective in classifying ADHD participants. By using SVM, classification accuracy of only ΔHb and only ΔHbO-based features was 78 % and 57 % respectively. When features of both responses were fused classification accuracy was reported as 72 %. This finding is quite interesting since ΔHbO was generally preferred as a dependent variable for fNIRS analysis due to its high signal-to-noise ratio (SNR) relative to ΔHb (Homae et al., 2010; Montero-Hernandez et al., 2018; Niu et al., 2011; Zhang et al., 2010).

Different strategies for feature extraction such as MVPA were also studied in ADHD research using fNIRS (Gu et al., 2018). 40 participants (20 ADHD, 20 HC) were included. Using N-back task-induced mean ΔHbO values extracted from bilateral dorsolateral prefrontal cortex (DLPFC), inferior middle frontal cortex (IMFC), right posterior prefrontal cortex (PPFC) and right temporal cortex (TC) were given to Multi-Voxel Pattern Analysis (MVPA) to reveal the hemodynamic patterns as input for SVM classifier. Maximum accuracy was found as 86 %. Except for frontal regions that we addressed before, ΔHbO activity in TC might be used as an efficient biomarker for ADHD classification.

On the other hand, the first fNIRS/EEG multimodal study to classify ADHD patients was performed by Güven and her colleagues (Güven et al., 2020). Model development was performed on 44 participants (23 ADHD, 21 HC) and the integral value of pre-frontal ΔHbO induced by the oddball paradigm was used. Maximum accuracy was found as 77.27 via naïve Bayes classifier. The best performance was obtained by combining ΔHbO and EEG/ERP features which yield 93.18 % accuracy.

3.1.2. ADHD vs ASD

ADHD vs ASD classification is also a challenging problem since both have higher prevalence compared to other neurodevelopmental disorders (Hansen et al., 2018) and it was suggested that both have the same origin and share common symptoms which make the discrimination harder (Kern et al., 2015). Two fNIRS studies focused on ADHD / ASD classification. One of those studies focused on hemodynamic biomarkers in the occipital region induced by a face-familiarity task, however, their sample size is relatively quite small (N=17, ADHD=9, ASD=8) compared to other ADHD classification studies (Ichikawa et al., 2014). They found 84 % accuracy by using SVM. The other study focused on MPH medication and hypothesized that hemodynamic response after MPH medication can be a discriminative biomarker between ADHD and ASD (Sutoko et al., 2019). 32 subjects were recruited (ADHD=21, ASD=11) and the Go/No-Go task-induced ΔHbO and ΔHb activation of the right middle frontal gyrus (MFG), right angular gyrus (AG) and right precentral gyrus (PreCG) was used as input for classifiers. After pooling results of six different classifiers (Simple, AND, OR, LDA, quadratic discriminant analysis, SVM), 82 % accuracy with 93 % sensitivity and 86 % specificity was found and the most optimum feature was reported as increased MPH–evoked ΔHbO response in ADHD group and decreased activation in the right hemisphere in ASD group.

3.2. Alzheimer’s Disease (AD)

AD is one of the subpopulations of dementia and the vast majority of the dementia population (>70 %) are of this population (Alzheimer’s Association, 2019). It has a prevalence among populations over 60-year-old aged between 4-6.4 % (Mayeux & Stern, 2012). Cognitive impairment and depression are the most common symptoms (Bature et al., 2017). Several functional neuroimaging studies tried to focus on diagnostic biomarkers of AD (see reviews (Ruan et al., 2016; Varghese et al., 2013)). Particularly, several fNIRS studies were published to understand the underlying mechanism of AD (Arai et al., 2006; Araki et al., 2014; Herrmann et al., 2008; Hock et al., 1997; R. Li et al., 2018). However, few ML applications were performed by using fNIRS on AD populations. Among all AD classification studies, Ho and colleagues’ study has the highest number of participants and they proposed a deep learning framework for sub-population classification of AD (T. K. K. Ho et al., 2022). 140 subjects including 53 HC, 28 asymptomatic AD, 50 prodromal AD and 9 AD dementia attended an fNIRS session focusing on prefrontal cortex activation during Oddball, 1-back and VFT. The highest accuracy that was reached in this study was 90% ± 1.2% when the CNN-LSTM classifier was used and raw data was selected as inputs of architecture.

On the other hand, two fNIRS/EEG study focusing on AD was reported in 2021. One includes four clinical groups with 29 participants in total (MCI=6, mild AD=8, moderate/severe AD=8, HC=7) (Cicalese et al., 2020). Random digit encoding-retrieval task-induced mean ΔHbO and ΔHb were used for classification using linear discriminant analysis (LDA). While fNIRS-based features and EEG-based features were separately giving 58.62 % and 65.52 % accuracy respectively, using fNIRS and EEG-based features together resulted as 79.31 % accuracy. Cortical regions including the right PFC and left parietal area was the most discriminative hybrid features that were found in this study. A similar multimodal EEG-fNIRS approach to classify Alzheimer's disease was introduced by Chiarelli and colleagues (Chiarelli et al., 2021). 35 participants (17 AD, 18 HC) were attended to a resting state experiment and average concentrations of ΔHbO and ΔHb were used as features. A multivariate regression analysis was used to classify AD and control subjects by using EEG, fNIRS and Neurovascular uncoupling metrics and 90.5% accuracy was reported.

Another study that tried to classify AD, MCI and HC subjects was conducted by Kim and colleagues (E. Kim et al., 2021). In this study 60 participants(18 AD, 11 MCI and 31 HC) attended fNIRS recording sessions during two different working memory tasks including Delayed Match-To-Sample (DMTS) task and Digit Span Test (DST). Functional connectivity maps from PFC were calculated using ΔHbO. FC values were used as input of artificial neural network (ANN) classifier to classify disease state and found the highest accuracy was 93.7%.

3.3. Autism Spectrum Disorder (ASD)

According to the DSM-5, ASD is a neurodevelopmental disorder characterized by troubles with communication and social interaction with people, limited interests, persistent behaviors and difficulties that affect social and work life (American Psychiatric Association, 2013). It has a prevalence of 2.6 % and the ratio between males and females in the general population are 2.6:1 (Y. S. Kim et al., 2011). Autism Diagnostic Observation Schedule (ADOS) is admitted as the gold standard approach for ASD diagnosis (Lord et al., 2000). Previous studies on determining ASD biomarkers focused on biochemical and clinical signatures (Bridgemohan et al., 2019), structural MRI (Pagnozzi et al., 2018) resting-state fMRI (Chen et al., 2019) and genetics (Goldani et al., 2014).

All reported ASD classification studies were done by using a similar dataset. In this dataset, 47 children (Typical developing (TD)=22, ASD = 25) were recruited and an 8 min of resting-state measurement from bilateral temporal regions was performed. In the first study, researchers utilized resting state functional connectivity and power spectra (RSFC) as measures for both ΔHbO and ΔHb (J. Li et al., 2016). 81.6 % sensitivity and 94.6 % specificity were found by using a linear SVM classifier. There are also two critical findings in this study. First, weaker RSFC and second higher ΔHbO and ΔHb power spectra values in the ASD group. Both findings were suggested as discriminative biomarkers.

A re-analyzed version of this data was reported in another study (Cheng et al., 2019). In addition to the features used in the previous study, a specific frequency of interest for both ΔHbO (0.02 Hz) and ΔHb (0.0267 & 0.0333 Hz) in TC was also added as a feature and used as an input for an SVM classifier. With this new feature set, 92.7 % accuracy was found. The major difference between the two groups was reported as in the frequency band of 0.02-0.03 Hz. Also, a recent study on resting state-based classification of ASD individuals used sample entropy as a feature (Xu, Hua, et al., 2020). Using k-means classification, 97.6 % accuracy was found. After performing machine learning studies, two deep learning studies on similar data were recently reported (Xu et al., 2019; Xu, Liu, et al., 2020). In the first study (Xu et al., 2019), a convolutional neural network (CNN) with a gate-recurrent unit (GRU) was trained and tested via hold-out cross-validation and 92.2 % accuracy with 85 % sensitivity and 99.4 % specificity was found. In the other study (Xu, Liu, et al., 2020), CNN and long-short term memory (LSTM) were trained and tested via hold-out cross-validation and 95.7 % accuracy was reported. Compared to conventional machine learning approaches, deep learning approaches (CNN + LSTM) outperformed previously reported machine learning scores. The last study using the same dataset was reported in 2021 (Xu et al., 2021). From raw data, they found 90.6 % sensitivity and 97.5 % specificity by using LSTM with an attention mechanism. These studies also show the efficiency of deep learning approaches in classifying fNIRS signals.

Dahan and colleagues conducted a study to classify Autism Spectrum Quotient (AQ) patients according to their severity (Dahan et al., 2020). In this purpose, 26 ASD patients were attended to the synchronization task. Portable 23-channel fNIRS device (Brite 23 Artinis Medical Systems) was used to acquire fNIRS signals. Ɛ-complexity coefficients of signals were extracted and used as inputs of classification algorithms. SVM and Random Forest algorithms were used as classification algorithms. As validation techniques 5-fold cross-validation and Leave one out cross validation were used for SVM and only k-fold cross-validation technique for RF. The highest accuracy that was reached in this study was reported as 96.3% when RF was used as a classifier.

3.4. Amyotrophic Lateral Sclerosis (ALS)

Amyotrophic lateral sclerosis (ALS) is a neurodegenerative disease that is characterized by the loss of upper and lower motor neurons of corticospinal tract, brain stem and spinal anterior horns (Proudfoot et al., 2018). In addition to applicability of fNIRS to decode signals from ALS patients (Borgheai et al., 2020), Deligani and colleagues proposed a classification framework by using multimodal EEG-fNIRS imaging technique to classify ALS patients (Deligani et al., 2021). For this purpose, multimodal data acquisition was done during a visuo-mental task from 18 participants (9 ALS, 9 HC). Peak value and AUC of ΔHbO were used as features for SVM classifier. 60.19% and 62.64% accuracies were found by using only fNIRS-based features. These accuracies increased to 87.32% and 87.51% for multimodal approach.

3.5. Major Depressive Disorder (MDD)

MDD is a highly prevalent (%7 in the U.S.) psychiatric disorder (American Psychiatric Association, 2013) which is characterized by abnormal affect and mood change with several physiological disturbances in sleep, appetite, and psychomotor agitation (Belmaker & Agam, 2008). It is generally diagnosed by psychiatric assessment which is accepted as the "gold standard" method, however, the reliability of this approach and self-reporting is quite controversial (Davison et al., 2009). Due to this reason, strong biomarkers are needed to classify this process (Kennis et al., 2020). For MDD classification, several studies have been reported and in the first study, a physical rehabilitation task was applied to 31 participants (14 HC and 17 MDD) (Zhu et al., 2020). Ten features were extracted from ΔHbO of DLPFC and VLPFC and five of those features (ΔHbO variance from left DLPFC, mean ΔHbO from left VLPFC, FWHM of ΔHbO from medial PFC, mean ΔHbO from right VLPFC and Kurtosis of ΔHbO from right DLPFC) gave the highest accuracy for both XG Boost (92.6 %) and Random Forest (91.13 %) classifiers.

Another study for MDD classification was conducted by Chao and colleagues (Chao et al., 2021). 32 participants (16 MDD and 16 HC) attended to emotional sound test. Statistical-based features such as mean, standard deviation, AUC and slope were calculated from ΔHbO signals. Besides these features, 4 vector-based features were calculated using ΔHb and ΔHbO orthogonal plane vectors. These are Cerebral Blood Volume (ΔCBV), Cerebral Oxygen Change (ΔCOE), angle K which represents the ratio of ΔCOE to ΔCBV, degree of oxygen exchange and the magnitude of L that represents change in 4 hemoglobin indices ΔHb, ΔHbO, ΔCOE and ΔCBV. Four simple neural networks architecture such as multilayer neural network (MNN), feedforward neural network (FNN), cascade forward neural network (CFNN) and recurrent neural network (RNN) were performed to classify MDD and HCs groups. Highest accuracies achieved using only statistical-based features were reported as 89.74% for fear emotion when researchers used RNN and 99.86% was reported for only vector-based features and used CFNN classifier for fear emotion. Also, this study claimed that AUC and angle K of fNIRS signals recorded from the prefrontal cortex (PFC) are specific neurological biomarkers for detecting MDD.

In the study performed by Wang and colleagues (Wang et al., 2021), 96 subjects (79 MDD and 17 HC subjects) attended 3 tasks including the before-task silent session (30 s), the on-task session (60 s) and after task silent session (60 s). Channel selection step was done according to SNR and t-test results. In addition to raw data and correlation maps of channels, 67 features from three indicators of ΔHbO, ΔHb and ΔHbT were calculated and used as input for deep recognition framework. RestNet18, AlexNet and machine learning algorithms like gradient boosting decision tree and SVM were used as classifiers. In this study highest accuracy of 90% was achieved when correlation maps were used as input and AlexNet was selected as classifier.

Two studies used verbal fluency task (VFT) in MDD classification studies. VFT is a popular task in MDD research to reveal potential differences between MDD and HC groups (Henry & Crawford, 2005). First of these studies was performed by Li and colleagues (Z. Li et al., 2022). Also, among all these reported studies for MDD classification, the highest number of participants were reported in this study. 363 participants (177 MDD and 186 HC) were applied to verbal fluency task (VFT). Extracted features were based on integral and centroid values of ΔHbO obtained from the pre-frontal region. By using features, SVM classifier achieved the highest accuracy of 75.6% ± 4.7.

Another study that used VFT was done by Ho and colleagues (C. S. Ho et al., 2022). In addition to fNIRS data, they used clinical and demographic information of 133 subjects (65 MDD and 68 HC) that attended to verbal fluency task. 14 time domain features FC Pearson's correlation coefficients that were calculated from fronto-temporo-parietal based ΔHbO and ΔHb. SVM was used as classifier and the highest accuracy was reached with an 87.98 ± 8.84 rate when clinical, demographic and features given as inputs to SVM algorithm.

3.6. Epilepsy

Epilepsy is one of the most common neurological diseases that more than 70 million people worldwide are suffering (Thijs et al., 2019). It was characterized as one or more seizures with a relatively high recurrence risk (Fisher et al., 2014; St Louis & Cascino, 2016). EEG is the most popular and widely known technique for the diagnosis of epilepsy and MRI is also critical to understand the etiology and localization of seizures (St Louis & Cascino, 2016). fNIRS has recently emerged as an alternative tool for both these techniques for epileptic seizure prediction (Binder & Haut, 2013).

Three studies focus on the prediction of epileptic seizures using fNIRS. The first study used fNIRS signals acquired with EEG data with a deep learning approach to predict epileptic seizures, however by only using fNIRS data (Rosas-Romero et al., 2019). Detection of ictal, pre-ictal and inter-ictal seizures was performed by using a convolutional neural network (CNN) for five epilepsy patients due to showing epileptic seizures during the measurement. Input data was three-dimensional tensors that were created by using ΔHbO and ΔHb via more than 100 fNIRS channels that cover the fronto-temporo-parietal region. For all patients, classification accuracy varies between 96.9-100 %. Both fNIRS and EEG were also used by using deep learning (long-short term memory -LSTM) to predict the epileptic seizures (Sirpal et al., 2019). In this study, 40 patients with 89 epileptic seizures were tried to detect by using three-dimensional tensors created using ΔHbO and ΔHb. Accuracies were found as 97.6 % and 97.0 % by using only EEG and only fNIRS, respectively. Using both modalities slightly increased the accuracy to 98.3 %.

Unlike the previous studies, the most recent study on epileptic seizure detection was performed by using conventional machine learning algorithms such as Multi-layer Perceptron (MLP) and SVM (Guevara et al., 2020). To classify only the pre-ictal state which corresponds to the stage before the onset of the seizure, similar EEG, ΔHbO and ΔHb time series in a previous study (Rosas-Romero et al., 2019) were used to train these algorithms with seven different combinations (EEG, ΔHbO, ΔHb, EEG + ΔHbO, EEG + ΔHb, ΔHbO + ΔHb, EEG +ΔHbO + ΔHb). Among those feature combinations, EEG-based features resulted an accuracy between 58.3-82.75 % by using MLP and between 90-97.1 % by using SVM. However, using ΔHbO and ΔHb-based features gave accuracy between 98.05 -100 % and 90.16 -100 % respectively by using MLP and SVM. Other feature combinations reached 100 % accuracy in both classifiers. However, the most critical finding was that fNIRS-based classification performance outperformed the classification using EEG data which is generally considered as the “gold standard” for epileptic seizures (Shellhaas, 2015).

3.7. Impulsive

Despite not being a specific disorder or disease, impulsivity is a critical trait that is used in the diagnosis of psychiatric disorders (Moeller et al., 2001). Therefore, its existence might provide support to clinicians in the diagnostic decision. The gold standard of impulsivity diagnosis is behavioral questionnaires and clinical interviews that are highly subjective. To overcome this problem, Erdogan and colleagues proposed a computer-based decision support algorithm (Erdogan et al., 2021)using fNIRS signals that were recorded from 71 participants (38 impulsive adolescents and 33 HC) during a Stroop task. Connectivity-based features were extracted from fNIRS signals and fused with behavioral features. SVM and ANN were used as classifiers to classify impulsive adolescents from HC subjects in this study. Accuracies of SVM and ANN for behavioral, clinical test and fNIRS features were both above 90%, but SVM performed better accuracy (92.2%).

3.8. Fibromyalgia (FM)

FM is a widely known disease that can be identified by chronic and widespread pain, tenderness and several cognitive dysfunctions with has 2-8 % prevalence in the population (Clauw, 2014). Several systemic conditions have similar symptoms to FM (Hochberg et al., 2003). Patients with FM can usually have irritable bowel syndrome, functional gastrointestinal disorders, chronic fatigue, somatoform disorders and other regional pain diseases (Clauw, 2014). There are several classification studies for individual diagnosis of Fibromyalgia using structural MRI (Robinson et al., 2015), resting state fMRI (Sundermann et al., 2014) and pain task-induced fMRI (Lopez-Sola et al., 2017). To our best knowledge, there is only one fNIRS study for the classification of FM disease (Gokcay et al., 2019). In this study, 36 participants (19 FM and 17 HC) were recruited and three different experimental paradigms for both hands were performed (i) finger tapping, (ii) median nerve stimulation task using transcutaneous electrical nerve stimulation (TENS) and (iii) painful stimulation with TENS. Extracted features were functional connectivity values using correlation coefficients and mean ΔHbO from the six regions including postcentral gyrus (Post CG), middle frontal gyrus (MFG), precentral gyrus (Pre CG), superior parietal gyrus (SPG), supramarginal gyrus (SMG) and angular gyrus (AG). After selecting the statistically significant features using a t-test, several different classifiers SVM, K-nearest neighborhood (KNN), and

Linear Discriminant Analysis (LDA) with different parameters were trained and tested. A decision-based-fusion model using the results of different classifiers was also proposed in this study. After fusing the decision, 100 % accuracy was found. It was also suggested that such a decision-level framework by using fNIRS data can also be used for diagnostic classification purposes.

3.9. Mild Cognitive Impairment (MCI)

MCI is a syndrome that is defined as cognitive decline that is associated to aging (Lopez, 2013). It is a phase before AD that the treatment might reduce its further effects (Yeung et al., 2016). Due to this, its early identification and predicting conversion to AD is quite critical. Several fMRI (Hojjati et al., 2017, 2018; A. Khazaee et al., 2016; A Khazaee et al., 2017) and PET (Cabral et al., 2015) studies are mainly focused on these two critical points. Until now, only two studies were reported for MCI classification using fNIRS. The first study on MCI classification using fNIRS and machine learning was performed by Yang and colleagues (Yang et al., 2019). 24 participants (15 MCI: 9 HC) were recruited for this study and three task (N-back, Verbal Fluency and Stroop) were applied to these participants. Extracted features were mean, slope, peak, skewness and kurtosis of ΔHbO & ΔHb with t-map and correlation maps of all channels from left, middle and right PFC. T-map and correlation maps were used to train and test a CNN and other features were used to train an LDA. CNN resulted as a 90 % accuracy using t-maps from the N-back task and LDA resulted as 76.67% using N-back and Stroop-task.

The other study that used a deep learning approach to diagnose MCI was conducted by (D. Yang & Hong, 2020). In this study, 15 MCI patients and 9 HC subjects attended resting state sessions with 9 different durations (30s, 60s, 90s, 120s, 180s, 210s and 240s). Mean, standard deviation and variance of both ΔHb and ΔHbO were calculated besides functional connectivity matrices for all time intervals. Mean, standard deviation and variance of both ΔHb and ΔHbO were calculated besides functional connectivity coefficients for all time intervals. The highest accuracy that was reached in this study was 97.01% when the classification-based transfer learning method using the VGG19 pre-trained CNN model with 30s time duration.

3.10. Parkinson's Disease (PD)

PD is a neurological disorder that is characterized by a noticeable tremor in hands and non-motor symptoms (Kalia & Lang, 2015). It has a prevalence of 1% among the population above 60 years (Tysnes & Storstein, 2017). Its diagnosis was performed by considering a comprehensive history and physical examination (DeMaagd & Philip, 2015). Neuroimaging studies using MRI have revealed some biomarkers such as basal ganglia connectivity and substantia nigra morphology (see review(He et al., 2018)). fNIRS has also recently gained importance in PD research (see review (Stuart et al., 2018)). In addition to this, the first PD classification study using fNIRS and EEG was recently published (Abtahi et al., 2020). 18 participants (PD:9, HC:9) were recruited and fNIRS measurements were conducted during 4 different tasks that were bilaterally applied (Right and left-hand finger tapping, right and left-hand flipping, right and left arm movement, right and left foot stomping). Mean block average ΔHbO and mean power on EEG theta, alpha, beta frequency bands and activity-based sensor data were used as features. SVM was preferred as a classification algorithm with Linear, Polynomial and RBF kernels. By using only fNIRS based, EEG based, fNIRS + EEG based, fNIRS + EEG + Sensor based features, 81.23 %, 92.79 %, 92.27 %, 93.40 % accuracy were found respectively. Compared to other scenarios, classification using fNIRS-based features resulted as lower accuracy. Delay in response was suggested as a potential reason for this classification.

3.11. Schizophrenia (SCZ)

SCZ is a severe mental illness that is characterized by symptoms such as hallucinations and delusions with severe cognitive problems such as working memory impairment (McCutcheon et al., 2019). It has a prevalence of %1-2 (Saha et al., 2005). The vast majority of previous biomarker research on SCZ mainly focused on neuroimaging, genetic and protein biomarkers (Rodrigues-Amorim et al., 2017). Among neuroimaging studies, EEG (Johannesen et al., 2016; Sabeti et al., 2009; Shim et al., 2016) and structural and functional MRI-based biomarkers have promising results (de Filippis et al., 2019).

SCZ is the most studied population with ADHD using fNIRS (Koike et al., 2013). In addition to conventional experimental studies since 1994 (Okada et al., 1994), eleven machine learning studies have been performed by utilizing fNIRS since 2010. The vast majority of those studies focused on the prefrontal cortex (PFC) based on differences between two populations. The first one was performed by recruiting 120 participants (SCZ =60, HC =60) and four different experimental paradigms were carried out; VFT, Tower of Hanoi, Sternberg and Stroop task (Azechi et al., 2010). Among these 120 participants, 60 of them (30 HC, 30 SCZ) were used for training the LDA classifier and the remaining participants (30 HC, 30 SCZ) were used for testing. Mean ΔHbO from the frontal region and task performance data were used as features. Classification results by using only fNIRS measurements showed a 78.3 % accuracy for the training group. For the testing group, 65 % accuracy was observed. By including task performance, 88.3 % accuracy was observed for the training group and 75 % accuracy was observed for the testing group. This study showed that combining fNIRS and task measures increases the diagnostic success rate of SCZ.

Another study, that focused on frontal region-based biomarkers was performed by utilizing a probabilistic method for classification (Hahn et al., 2013). Compared to other studies, prior information of a class (such as the prevalence of disease) was taken into account in this study because PPV and NPV of disease were stated as more important measures for diagnosis which are sensitive to prior class probability. In this study, the N-back task was applied to 80 participants (SCZ =40, HC=40). Whole ΔHbO response from the frontotemporal region was used as input for Gaussian Process Classifier (GPC). As a result, 76% accuracy with was found.

Chuang et al. also focused on both PFC-based biomarkers in SCZ, however, tried to classify them using a k-means classification (Chuang et al., 2014). In this study, 99 participants (SCZ =53, HC=46) were recruited and mean value of ΔHbO response was extracted from 52 channels located in the bilateral PFC and temporal lobe by performing a verbal fluency task. The highest accuracy was found as 71. 72 % with 77 % sensitivity and 65 % specificity by using 6 channels located on left IFG (5 of them) and right IFG (one of them).

In addition to the efficiency of PFC biomarkers, performance comparison of different classifiers is also another open question. Li et al. performed a verbal fluency task by focusing on the frontotemporal region (Z. Li et al., 2015). In this study, 240 participants (SCZ=120, HC=120) were recruited and four different classifiers (LDA, SVM, KNN, GPC) were trained using the mean value of ΔHbO. The highest accuracy was found by using Radial Basis Function (RBF) SVM (83.37 %).

PFC oriented specific channel selection approach was also used to classify the SCZ (Einalou et al., 2016). In this study, 27 participants (SCZ:16, HC :11) was recruited and participants were requested to complete stroop task. By using wavelet transform, 0.003-0.11 Hz frequencies were found critical for classification and genetic algorithm was used to select channels in PFC. Using SVM, they found 83.59 % accuracy.

fNIRS-based functional connectivity was also used as a biomarker in SCZ discrimination (Song et al., 2017). In this study, a one-back working memory task was applied to 76 participants (SCZ =42, HC=34) and activity from the frontotemporal region was recorded. After creating connectivity matrices for ΔHbO, ΔHb and ΔHbT, eigenvectors extracted from the degree of node, clustering coefficient, local efficiency and global efficiency of three concentration changes were extracted as features and given as input to RBF – SVM classifier. Higher accuracies were reported by using ΔHbO and ΔHb (85.5 %) compared to ΔHbT (80.3 %). Sensitivity scores were similar for all three concentration changes (92.8 %). However, the specificity score of ΔHbT (64.7 %) was found lower than ΔHbO and ΔHb (76.5 %).

There is another study that distinguishes SCZ from other psychiatric disorders using fNIRS and ML (Koike et al., 2017). In this study, four different populations including 143 participants (ultra-high risk psychosis (UHRP) =47, First episode psychosis (FEP)=30, Chronic Schizophrenia (SCZ)= 34, healthy controls (HC)=33) were recruited and after a 12 month of follow-up, another measurement was performed on 34 UHRP, 21 FEP and 33 healthy participants. Two different markers were used; integral and centroid values. Integral value is the sum of ΔHbO signal change during task period and centroid value is the time value of hemodynamic response that corresponds to half of the integral value after the stimulus onset. By using discriminant analysis,

A recent study on SCZ classification by using fNIRS was performed by utilizing wavelet-based decomposition (WBD) (Dadgostar et al., 2018). 27 participants (HC=11, SCZ =16) were recruited and fNIRS measurement was performed during a color-word matching Stroop task. ΔHbO wavelet-based energy values for 0-0.108 Hz were extracted using WBD for 16 channels and channel selection was performed by using a genetic algorithm and this input was given an RBF-SVM classifier. By using all channels, 74.31 % accuracy was reported, however, 87.31 % accuracy was reported by using only 6 channels.

In addition to Song et al. (Song et al., 2017) , another functional connectivity-based SCZ classification study was performed by using fNIRS (Ji et al., 2020). 200 SCZ patients and 100 HC were recruited for this study and a Chinese verbal fluency task was applied during the measurement. Seed-based functional connectivity was used as an input for the RBF-SVM classifier. Accuracy, sensitivity and specificity results were found as 89.67 %, 93.00 % and 86.00 % respectively.

Chou and colleagues proposed an algorithm to classify and detect first episode Schizophrenia patients (Chou et al., 2021). In this study, 67 participants (33 first episode SCZ and 34 HC) attended to VFT session during the fNIRS recording. Integral and centroid values of oxyhemoglobin changes were computed from fNIRS signals. SVM and DNN were used as classifiers. DNN reached better accuracy than SVM, with 79.9% while SVM accuracy was 68.8%.

In another study, fNIRS measurements were performed on 200 participants (100 SCZ and 100 HC) during a VFT task (J. Yang et al., 2020). Data was collected from bilateral prefrontal and temporal regions. FC matrices were calculated as features. LDA, GPC, KNN and SVM classifiers were used in this study. The highest accuracy that was reached in this study was 84.67%. This accuracy was reached when FC matrices from three channels on the medial prefrontal and left ventrolateral prefrontal cortices were used as inputs of the classifier.

SCZ and BP groups were also comparatively classified with HC by Eken and colleagues (Eken et al., 2022). In this study, 83 participants (23 SCZ, 30 BP and 30 HC) attended to fNIRS recording session during reading the mind in the eyes (RMET) task. ΔHbO time series that used to calculate dynamic functional connectivity maps using sliding window correlation method from bilateral frontopolar area, bilateral Broca's area, bilateral dorsolateral prefrontal cortex, bilateral inferior frontal gyrus, bilateral premotor cortex bilateral middle temporal gyrus and left superior temporal gyrus regions. As a machine learning strategy, one vs one technique was used to classify groups from each other. SVM, DA and KNN classifiers were used to classify BP & SCZ, BP & HC and SCZ &HC groups. In BP & HC and SCZ & BP classification problems highest average accuracy reached in SVM algorithm with 0.825 ± 0.051 and 0.755 ± 0.066 ratios. In SCZ & HC classification problems highest average accuracy was reached in DA with 0.790 ± 0.064. This study also is one of the rare studies that try to investigate comparative biomarkers between different psychiatric disorders.

Erodoğan and colleagues were conducted a study that try to classify 4 classes including HC, SCZ, OCD and migraine patients (Erdogan et al., 2021). In this purpose 13 HC subjects and 67 patients were attended to stroop task session when their prefrontal hemodynamic response was recorded by ARGES-CEREBRO fNIRS device. Cognitive and functional connectivity features were used as inputs of classification algorithms.10 run 10-fold CV was used to split data. NB, SVM and LDA algorithms were to classify groups. SVM method shown the highest performance with accuracy 85.1 ± 1.77%, sensitivity 84 ± 1.7%, specificity 95 ± 0.5% rates.

3.12. Somatic Symptom Disorder (SSD)

Somatic symptom disorder (SSD) or somatization disorder is a collection of medically unexplained symptoms that causes distress and discomfort. According to the DSM-5, it includes a group of disorders such as illness anxiety disorder, conversion disorder, factitious disorder and it has a prevalence of 5-7 % in the population (American Psychiatric Association, 2013). Previous studies proposed several neuroimaging biomarkers to identify SSD such as the right insula and left superior occipital gyrus (Pan et al., 2019) and MPFC – Anterior Cingulate Cortex (ACC) connection (Ou et al., 2018). fNIRS has previously been used in SSD research (Ren et al., 2017) and the first ML study using fNIRS was performed by Eken et al. (Eken et al., 2019). In this study, 40 participants (HC=21, SSD = 19) were recruited and fNIRS measurement was performed during two different painful stimulation task (Individual pain threshold level painful stimulation and constant level painful stimulation) with a brush stimulation as a control condition that was presented in both tasks. Dynamic functional connectivity using sliding window correlation (Sakoglu et al., 2010) was used to reveal the task-related connectivity changes. LDA and SVM classification was performed and 82 % accuracy with 81 % sensitivity and 85 % specificity were found using the data obtained by applying pain in levels of individual pain threshold and using connections right superior temporal – left angular gyri, right middle frontal – left supramarginal gyri and right middle temporal – left middle frontal gyri as input to an SVM classifier. Compared to only an fMRI study that utilizes ML (Ou et al., 2018), similar accuracy results were obtained.

3.13. Stuttering

Stuttering, which is mentioned as “Childhood-Onset Fluency Disorder” in DSM-5 (American Psychiatric Association, 2013) is a highly prevalent (5-8% among children (Månsson, 2000)) speech disorder that causes disturbances in fluently speaking (Perez & Stoeckle, 2016) and can be observed in every age. Its pathophysiology is still unclear. There are several functional neuroimaging studies that aimed to understand this disorder (see review (Etchell et al., 2018)) and several fNIRS studies (Jackson et al., 2019; Tellis et al., 2015; Walsh et al., 2017). Previous fMRI- ML studies using intrinsic brain connectivity (Qiao et al., 2017) and speech task-based hemodynamic response (J. Jiang et al., 2012) showed notable accuracy values for classifying patients suffering from the stuttering disorder.

To our best knowledge, there is only one stuttering study that combines fNIRS and ML (Hosseini et al., 2018). In this study, 32 children (stuttering:16, HC: 16) with an additional test group that includes 16 children who recovered from stuttering were recruited. fNIRS data were collected during a speech production task for three groups. Statistical (mean, variance, skewness, kurtosis) and morphological features (curve length, number of peaks, zero-crossings) with Normalized Area Under Signal (NAUS), Hjorth Mobility, Hjorth Complexity, Autocorrelation, Bicorrelation were used as input to SVM, LDA, KNN, decision tree and ensemble classifiers. The highest accuracy was found using SVM as 87.5 %.

3.14. Traumatic Brain Injury (TBI)

TBI is the condition of brain damage that may cause death or disability (Ghajar, 2000). It generally occurs in car accidents, sports injuries or assaults. It can be sub-grouped as mild, moderate and severe. There is only published research that focused on the classification of TBI using fNIRS (Karamzadeh et al., 2016). In this study, 61 participants (TBI =30, HC =31) were recruited and fNIRS measurement was performed during an event-related complexity task (Krueger et al., 2009). After collecting data, statistical features of averaged ΔHbO activity such as mean, variance, kurtosis, area under curve (AUC), Full-Width Half Maximum (FWHM) with Discrete Fourier Transform coefficients were extracted. Among SVM, LDA and decision tree classifiers, the highest accuracy was found as 84 % by using AUC, DFT coefficients and FWHM of ΔHbO activity.

3.15. Tinnitus

Tinnitus is relevance symptom that basically defined as the hearing of sounds in one or both ears when there is no evident in environment (Han et al., 2009). According to our knowledge first study that was tried to classify tinnitus and HC subjects was conducted by Shoustarian and colleagues (Shoushtarian et al., 2020). 25 tinnitus patients and 21 HC subjects were attended to visual and auditory task sessions besides resting state tasks. Subjects NIRS signals was acquired by multi-channel continuous wave NIRScout device. Connectivity measures and evoked response amplitudes derived from both HbO and HbR were used as inputs of classification algorithms. 4 Different classification algorithms including Naïve Bayes, KNN, Rule introduction and ANN were used as classifier and 10-fold cross validation technique was used to split data. In classification case of patients and HC, the highest accuracy 78.3% (sensitivity 72.33% and specificity 64.25%) was reached when auditory response features were used in Naïve Bayes method. In predicting severity of tinnitus case, the highest accuracy 87.32% (sensitivity 51.23% and specificity 95.12%) was reached when functional connectivity features were used in ANN method.

4. Clinical Conditions

In this review, we also included studies that consist of clinically relevant conditions to understand the efficiency of fNIRS while separating from each other. Under this title, we provided an overview of studies that aims to distinguish different clinical conditions (e.g. stress vs non-stress, high-pain vs low-pain) and non-clinical populations.

4.1. Anesthesia Monitoring

During surgical operations, anesthesia monitoring is a vital topic to detect sudden changes. Two studies were reported to classify the anesthesia phases. In the first study, fNIRS data collected under different anesthesia conditions from 19 patients under surgery was used for analysis (Hernandez-Meza et al., 2017). In this study, anesthesia conditions were identified as maintenance (MC) and emergence (EC). 248 (MC=229, EC=29) maintenance and emergence conditions were extracted and classified using RBF-SVM. Used features were local mean of ΔHbO, standard deviation of ΔHbO, local minimum of ΔHb and ΔHbO and range of ΔHb and ΔHbO. 94.8 % accuracy with 94.8 % sensitivity, 94.7 % specificity, 99.5 % PPV and 60 % NPV were found. The second study also used a similar paradigm with similar features except for different number of MC and EC conditions (MC =19, EC=19) from 19 patients under surgery (Hernandez-Meza et al., 2018). Using similar features like the previous study, LDA and RBF-SVM classifiers were trained. RBF-SVM gave the highest classification score as 94.7 accuracy. Both studies showed that automatic anesthesia phase detection was carried out by using fNIRS-based features with high classification scores.

4.2. Pain

Pain perception is a complex function that several brain regions are involved and its mechanism includes affective, sensory and cognitive processing networks in the brain. According to the International Association for the Study of Pain (IASP), pain is defined as “an unpleasant sensory and emotional experience associated with actual or potential tissue damage or described in terms of such damage” (Treede, 2018). Pain perception in the brain has been investigated for several years via several neuroimaging modalities such as fMRI (Apkarian et al., 2005) and fNIRS (L. Becerra et al., 2009; L. Becerra et al., 2008; Franceschini et al., 2003; Koch et al., 2010). There are several regions related to pain perception and processing such that the primary somatosensory cortex (SI), secondary somatosensory cortex, ACC and insula (Bornhovd et al., 2002; Buchel et al., 2002; Bushnell et al., 1999; Coghill et al., 1999; Derbyshire et al., 1997; Moulton et al., 2005; Poineau et al., 2012; Porro et al., 2003; Ringler et al., 2003). Also, PFC is found to be closely related to pain processing (L. Becerra et al., 2008; L. R. Becerra et al., 1999; Derbyshire et al., 1997). Functional neuroimaging of pain perception was generally conducted in terms of painful and non-painful stimuli comparison (Apkarian et al., 1999; L. R. Becerra et al., 1999; K. Li et al., 2002; Lui et al., 2008). Pain perception mechanisms were also analyzed using fNIRS, but due to the physical limitations, only SI and PFC were considered as a region of interest (L. Becerra et al., 2008). fNIRS studies that compare painful and non-painful stimuli show that there are significant bilateral SI activation and contralateral activation was found greater than ipsilateral one (L. Becerra et al., 2009; L. Becerra et al., 2008; Franceschini et al., 2003). A recent study focusing on this comparison of 11 healthy participants showed that responses for both stimuli were easily distinguished (Yucel et al., 2015).

For the classification of pain using fNIRS, the first study was published by Pourshoghi et al. (Pourshoghi et al., 2016). In this study, the main objective was to classify painful and non-painful heat stimulation conditions as low-pain and high-pain. To achieve this, a cold pressor test that includes four different temperatures (0-5-10-15 C°) was applied to 19 healthy control participants and acquired hemodynamic responses from the prefrontal region were grouped as high-pain and low-pain according to the self-reporting pain scores. Finally, 61 trials were classified and Functional Data Analysis (fDA) (Ramsay et al., 2009) framework was used to extract the features. 94 % accuracy was found by using SVM.

Two more studies focused on classifying four different painful stimuli (low-hot, low-cold, high-hot and high-cold) (Fernandez Rojas et al., 2017, 2019). In both studies, fNIRS signals collected during a heat-based painful stimulation paradigm from the SI region were used. In the study (Fernandez Rojas et al., 2017), wavelet coefficients were used as an input to both KNN and SVM and in the other study (Fernandez Rojas et al., 2019) peak time, mean value and Fourier coefficients were used in addition to wavelet coefficients as an input to LDA, SVM (linear, RBF, polynomial kernels) and KNN. In the first study, the highest accuracy was found as 92.08% with KNN and 91.25 % using SVM. However, in the second study a slightly lower accuracy 89.44 % was found by using RBF – SVM. These high accuracies revealed that fNIRS might provide discriminative biomarkers for identifying painful conditions.

4.3. Stress and Anxiety Prediction

First study focused on anxiety level prediction using fNIRS was published by Sato et al. (Sato et al., 2013). In this study, 19 participants aged between 19-24 were recruited and fNIRS signals collected from PFC during a 3-min resting state were used as input to a Bayesian approach to predict the State-Trait Anxiety Inventory (STAI) score. Mean values, all combinations of covariance and correlation coefficient of ΔHbO and ΔHb of R and L PFC were used. Average prediction error was found lower than 5 that STAI scores vary between 20-80. A similar study was also performed on older (61-79 age) population including 17 participants

(Y. Fukuda et al., 2014). Using the same paradigm and feature sets, average prediction error of STAI values was found 5.27.

In addition to anxiety level prediction, stress-based studies have recently become popular. Four studies were reported related to the classification of stress conditions. First study was published by Al-Shargie et al. (Al-Shargie et al., 2017b). In this study, 20 healthy participants were recruited and both fNIRS and EEG measurements were performed during a Montreal Imaging Stress Task that includes a control condition (mental arithmetic task without any time restriction) and a stress condition (a mental arithmetic task with time restriction). Mean ΔHbO and mean power of alpha band (8-12 Hz) from DLPFC, ventrolateral PFC (VLPFC) and Frontopolar area were used as input. Using RBF-SVM, two classifiers (one for fNIRS features and one for EEG features) were separately run and resulting decisions were fused. Only using fNIRS-based features resulted as 87 % accuracy. However, fusing both classifiers resulted as 96.48 % accuracy. Also, a similar study, carried out on 25 participants was performed by utilizing the same paradigm and ROIs (Al-Shargie et al., 2017a). In this study, mean ΔHbO and wavelet coefficients of EEG alpha band (8-12 Hz) were used together as potential stress markers. The highest classification accuracy was found as 97.7 % in channels located in VLPFC by fusing the EEG and fNIRS features and VLPFC was suggested as a potential ROI in stress detection.

Another study focusing on acute stress with mental workload was published by Parent and colleagues (Parent et al., 2019). In this study, a working memory task (N-back) and an auditory stress task were conducted on 18 healthy participants. During the task, simultaneous fNIRS on the frontal region and electrocardiography (ECG) measurement for assessing heart rate variability (HRV) were performed. For classification, mean and slope of ΔHbO & ΔHb and HRV features were extracted from fNIRS and ECG respectively. However, only the classification of stress conditions could slightly exceed the chance rate (50%). Using fNIRS-based features, HRV-based features and both, stress conditions were classified with 63 %, 53 % and 62 % accuracy respectively.

Also, a recent study using both hemodynamic response and heart rate-based features for stress detection was reported by Hakimi and his colleagues (Hakimi et al., 2020). Using Montreal Imaging Stress Task (MIST), 20 right-handed participants were measured using fNIRS. From the fNIRS signal, both heart rate and hemodynamic response-based features were extracted and given as input to CNN, dense neural network, SVM and RF. The highest accuracy was found by using heart rate-based features as 98.69 % using CNN and in general, classification results using heart rate-based features outperformed the classification results using fNIRS-based features.

4.4. Surgical Skill Assessment

Surgical training has vital importance for patient safety and to understand the training level, objective measures are strongly needed (Dawe et al., 2014). To overcome this problem, a recent study on an objective assessment of surgical training using fNIRS and ML was published (Nemani et al., 2019). The main objective was to automatically classify virtually (VST) or physically trained (PST) medical students from untrained medical students (UMS). In this study, 18 medical students were recruited and grouped into three. 5 of them were never trained, 7 of them were trained by Fundamentals of Laparoscopic Surgery Trainer (FLS) which is a physical simulator and 6 of them were trained by virtual basic laparoscopic skills trainer (VBLaST) which is the replication of FLS pattern cutting task on a computer. Both trained groups attended training for 12 consecutive days. After this period, fNIRS measurement was conducted on all groups during a surgical transfer task. Mean ΔHbO values were extracted from bilateral PFC, Supplementary Motor Area and Primary Motor Cortex (MI) and used as features with performance metrics of FLS and VBLaST. Using LDA classifier, the classification of PST vs UMS resulted in 97.8 % sensitivity and 97.3 % specificity. For VST vs UMS classification, 91.1 % sensitivity and 90.9 specificity % were found. Surgical skill assessment is a new field to show the efficiency of fNIRS.

Another study related to surgical skill assessment was published by (Keles et al., 2021).33 participants (11 surgeons, 5 surgery residents and 17 medical students) were applied to two FLS tasks (Peg transfer and Threading) and these participants were sub-grouped according to their NASA-TLX score (High vs Low) which is a subjective multi-dimensional scale that provides mental workload level score during the task and skill levels (Student vs Attending surgeon). Windowed standard deviation in each channel was extracted as features and given to linear SVM as input. For both classification tasks, ~90% accuracy was found and fNIRS showed great efficiency in determining surgical levels and surgical complexity.

5. Machine Learning in fNIRS

Machine learning usage by utilizing fNIRS data allows the researchers to focus on multivariate patterns that are hard to detect by using conventional statistical approaches. In this section, we will only mention the supervised learning approaches and the two most popular algorithms, SVM and LDA.

5.1. Supervised Learning

Supervised learning is a type of machine learning approach that provides information about the type of input and output. For instance, considering fNIRS, while training a classifier, input data occurs with different features such as mean ΔHbO, variance and its output is also labeled as a patient or healthy control (also can be patient from group 1 or patient from group 2). If this labeling was done according to the distinct classes, the procedure that is conducted is called “Classification”. Or if the labels were numerical values such as a disease severity score, this procedure is called “Prediction”. In supervised learning, the main objective is to create a model that maps input and output responses by using as much data as possible.

5.2. Algorithms

Many different machine learning algorithms were used in fNIRS studies. These algorithms were reported in Figure 4. Support Vector Machine (SVM) and Linear Discriminant Analysis (LDA) are the most preferred ones among all. SVM is a robust supervised learning algorithm that aims to discriminate the input data by creating a hyperplane with maximum margin according to their previously defined labels as we addressed in the section 5.1 (Vapnik, 1995). It was the most preferred classifier among fNIRS studies (Abtahi et al., 2020; Al-Shargie et al., 2017b; Cheng et al., 2019; Crippa et al., 2017; Dadgostar et al., 2018; Einalou et al., 2016; Eken et al., 2019; Fernandez Rojas et al., 2017, 2019; Gokcay et al., 2019; Gu et al., 2018; Guevara et al., 2020; Hernandez-Meza et al., 2018; Hernandez-Meza et al., 2017; Hosseini et al., 2018; Ichikawa et al., 2014; Karamzadeh et al., 2016; J. Li et al., 2016; Z. Li et al., 2015; Naseer & Hong, 2015; Pourshoghi et al., 2016; Song et al., 2017; Sutoko et al., 2019; Xu et al., 2019; Yasumura et al., 2017). The data points that are used to define this margin are called Support Vectors. This hyperplane can either be a linear kernel or a non-linear kernel (such as polynomial or gaussian (also known as radial basis function)). For an input $x \in R^{p\ x\ d}$, for a single observation x_i and its associated class y_i is represented as;

$$x_i(x_i \in R) \rightarrow y_i(y_i \in \{-1, +1\}) \qquad (1)$$

For this association, -1 and +1 represents binary classes. An ideal decision hyperplane can be written as by using a linear kernel.

$$w^T x + b = 0 \qquad (2)$$

In this equation, w is the orthogonal weight vector to the hyperplane, x is input and b is represented as bias. For the best separable case of both classes, optimal margin should be $\frac{2}{\|\boldsymbol{w}\|}$. By using quadratic programming, this margin is maximized. w and b are used for this purpose. So, the objective function becomes in Eq. 3, subject to $y_i(w^T x + b) \geq 1 - \varphi_i$ and $\varphi_i \geq 0$.

$$\min_{w,b,\varphi_i} \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{n} \varphi_i \qquad (3)$$

In Eq. 3, w is the weight matrix, C is the regularization parameter. C is used to avoid overfitting. If C value become larger, SVM classifier assigns less support vectors, however, this causes a long training duration. φ_i represents the slack variables that SVM uses them to penalize the data points that if ith sample exceeds the margins. If $\varphi_i = 0$, it means ith sample did not exceed the margin. Otherwise, it becomes $\varphi_i \geq 0$. Algorithm uses the Lagrangian multipliers to find the optimum values. If $a_1, a_2 \ldots \ldots, a_n$ are the Lagrangian coefficients, Eq. 4 should be minimized according to the these coefficients subject to Karush-Kuhn-Tucker rule for separable case $\sum \alpha_i y_i = 0\,, \alpha_i \geq 0$, and non-separable case $\sum \alpha_i y_i = 0\,, 0 \leq \alpha_i \leq C$ for linear SVM.

$$\min_{\alpha_1,\ldots\ldots\alpha_n} \sum_i \alpha_i - \frac{1}{2}\sum_i\sum_j \alpha_i\alpha_j y_i y_j x_j{}' x_k \qquad (4)$$

The score function can be estimated according the Eq. (4).

$$\hat{y} = \sum_{i=1}^{n} \hat{a}_i\, y_i x' x_k + \hat{b} \qquad (5)$$

In Eq. 5, $\hat{b}$ is bias estimate, $\hat{a}_i$ is the ith estimate of $\hat{a}$ vector. For non-linear SVM, instead of the variables $x_j{}'x_k$, a non-linear function K is involved into equation and the SVM function is minimized subject to Karush- Kuhn-Tucker rule for non-separable case $\sum \alpha_i y_i = 0\,, 0 \leq \alpha_i \leq C$ as shown in Eq. 6.

$$\min_{\alpha_1,\ldots\ldots\alpha_n} \sum_i \alpha_i - \frac{1}{2}\sum_i \sum_j \alpha_i \alpha_j y_i y_j \, K(x_j, x_k) \quad (6)$$

Score function is shown in Eq. 7.

$$\hat{y} = \sum_{i=1}^{n} \hat{a}_i \, y_i K(x_j, x_k) + \hat{b} \quad (7)$$

For polynomial kernels, d as degree of polynomial and c as a constant, $K(x_j, x_k) = (c + x_j{}'x_k)^d$ and for radial basis kernels as σ^2 is the variance of kernel, RBF kernel is represented as $K(x_j, x_k) = e^{(\frac{-\|x_j - x_k\|^2}{2\sigma^2})}$.

Another popular algorithm is LDA. LDA is also a quite popular supervised learning algorithm and it is also widely used for fNIRS based classification studies (Azechi et al., 2010; Cicalese et al., 2020; Eken et al., 2019; Fernandez Rojas et al., 2019; Gokcay et al., 2019; Hernandez-Meza et al., 2018; Hosseini et al., 2018; Ishii-Takahashi et al., 2015; Karamzadeh et al., 2016; Z. Li et al., 2015; Nemani et al., 2019; Sutoko et al., 2019; Xu et al., 2019; Yang et al., 2019). Simplicity and low computation cost are the greatest advantages of LDA. Unlike SVM, the hyperplane is estimated by projecting the data via maximizing the distance between classes. LDA can be used for both dimensionality reduction and classification. In LDA, if we assume that data in every class is normally distributed, as Σ is the common covariance matrix and k is the class, our multivariate normal density equation for class k becomes;

$$p(x|k) = \frac{1}{(2\pi)^{d/2}|\Sigma|^{1/2}} e^{-\frac{1}{2}((x-\mu_k)^T \Sigma^{-1}(x-\mu_k))}$$

our linear discriminant function is;

$$D_k(x) = x^T \Sigma^{-1} \mu_k - \frac{1}{2}\, \mu_k^T \Sigma^{-1} \mu_k + log p_k$$

In this equation μ_k is the mean value of every class and p_k is the prior probability of every class. According to this equation, linear discriminant classifier function $C_k(x)$ is;

$$C_k(x) = \arg\max_k D_k(x)$$

5.3. Feature Engineering

Feature engineering is a general term that represents the process to create highly predictive features from raw data using domain knowledge. It plays an important role while identifying a biomarker. In this review, we will focus on two important parts of feature engineering; Feature extraction and feature selection or reduction. In conventional machine learning approaches except for deep learning methods, feature extraction is the initial step that domain knowledge is strongly involved. A priori information related to data is strongly essential. Therefore, relevant features that can increase the prediction or classification accuracy are extracted. After completing the Feature Extraction step, performing feature selection or feature reduction is dependent on dataset dimensions. Although there isn't any rule of thumb about determining an ideal dimension of a feature vector -except for the number of sample sizes >> the number of dimensions-, it is remarkable to use feature selection methods to avoid overfitting due to high dimensionality in neuroimaging datasets. There are several feature selection (or reduction) methods such as Principal Component Analysis (PCA), t-test, t-distributed stochastic neighbor embedding (t-SNE), Least Absolute Shrinkage and Selection Operator (LASSO) (Tibshirani, 1996), Recursive Feature Elimination (RFE), ReliefF (Kenji & Larry, 1992).

For fNIRS studies, feature types can be grouped under two different categories; Features extracted from task-based hemodynamic response and functional connectivity-based features. After excluding deep learning-based studies that generally do not require any feature engineering steps, the vast majority of those studies used mean ΔHbO as a feature from different cortical regions depending on the research question. As we mentioned above, ΔHbO is the main dependent variable for fNIRS analysis due to its high SNR compared to ΔHb (Homae et al., 2010; Montero-Hernandez et al., 2018; Niu et al., 2011; Zhang et al., 2010). It is also preferred in BCI studies (Naseer & Hong, 2015). However, some surprising results can be encountered such as finding higher accuracy by using ΔHb than using ΔHbO (Crippa et al., 2017; Xu et al., 2019). This is a controversial issue. Although there are some exceptional cases (Strangman et al., 2002), common agreement is that decrease in ΔHb is highly correlated with blood-oxygenation-level-dependent (BOLD) signal (Mehnert et al., 2013; Steinbrink et al., 2006).

Connectivity-based features have also emerged as another alternative input for ML algorithms. Due to its nature, resting-state-based classification studies using fNIRS utilize these features (Cheng et al., 2019; J. Li et al., 2016; Xu et al., 2019; Xu, Liu, et al., 2020). In addition to this, some task-based studies also use connectivity-based features (Eken et al., 2019; Gokcay et al., 2019; Song et al., 2017; Yang et al., 2019).

5.4. Optimizing Hyperparameters

To improve the performance of classifiers, optimizing hyperparameters using different approaches is an option. Among 62 fNIRS studies, only 17 of them utilized hyperparameter optimization. Vast majority of these studies used Grid-search parameter optimization (Fernandez Rojas et al., 2019; Z. Li et al., 2015; Yang et al., 2019; Yasumura et al., 2017) and Bayesian optimization (Eken et al., 2019; Hosseini et al., 2018; Parent et al., 2019). The grid-search algorithm creates all combinations of parameters and trains the classifier by using these parameters. After training all, it gives the optimum parameter set that provides the lowest validation error. Grid- search is computationally expensive both for time and space.

Also, as the number of parameters increases, computational complexity becomes high. On the other hand, Bayesian optimization is a sequential iterative optimization process that aims to find the global optimum set of parameters using minimum iterations. Compared to grid search, it uses less training time but, considers fewer options. For deep learning studies, Adam (adaptive moment estimation) optimizer is the most popular method for parameter optimization and is generally preferred in several fNIRS-based deep learning studies (Xu, Liu, et al., 2020; Yang et al., 2019).

5.5. Cross-Validation Techniques

Cross-validation (CV) is the resampling approach to generalize the ML results. Three main CV techniques are used in ML studies. Leave-one-out cross-validation (LOOCV), Hold-Out CV and K-fold CV. In LOOCV, only a single observation from data is used for the test and the rest is used for training. This operation was done for every observation. Therefore, you have n test scores and then the average score is estimated. It provides less bias since all data is used for testing. However, for the same reason, variation is high in scores. Also, for larger samples (e.g. > 100-1000) computational cost is high. For hold-out CV, data is separated as training and test set. Percentages vary around for training 60-90 % and test 10-40 %. Training and testing are done only once. This is ideal for a large dataset that requires more computational power and time. However, results are highly biased due to less generalization because training and testing samples might not represent the whole data. Among analyzed studies, few of them used hold-out validation

K-fold CV is the most popular CV method. In this method, observations are divided into K number of training and test folds that both training and test folds were stratified. For every fold, a classifier is trained by using training fold and tested by using test fold. This is done by K times. After having a classification score from every classifier, all these scores were averaged. It is ideal for moderate-sized (e.g.$(N \approx 50 - 100)$) datasets. However, for larger datasets, it causes computational complexity.

For some cases, nested CV is also used (Crippa et al., 2017; Eken et al., 2019). Nested CV consists of two nested loops. The outer loop is always for generalization of ML models and the inner loop is either for hyperparameter optimization or rarely feature selection (Parvandeh et al., 2020). It is used for having an unbiased estimate of classification scores. To optimize classification results with unbiased results, nested CV is a highly reliable approach.

6. Biomarker Research on fNIRS

Biomarker identification for a disease has always been a challenging topic for researchers and clinicians. According to the Biomarkers Definitions Working Group, biomarker is defined as “a characteristic that is objectively measured and evaluated as an indicator of normal biological processes, or pharmacologic responses to an intervention” (Biomarkers Definitions Working, 2001). Weickert et al addressed three features of a biomarker; 1) Diagnostic (ability to classify the disease) 2) Prognostic (ability to predict the progress of the disease), 3) Theranostic (ability to predict a treatment pathway) (Weickert et al., 2013). Biomarkers should also have some additional features such as interpretability, deployability and generalizability (Woo & Wager, 2015). As Huss stated, depending on several criteria, biomarkers can be classified as either imaging biomarkers such as MRI, PET, Computed Tomography, EEG or molecular biomarkers such as genetic variations, mutations, proteins and lipids (Huss, 2015). Interpretable biomarkers in neuroimaging should be meaningful and associated to other parameters such as genetic or clinical. For this review, few studies include correlation or regression analysis between selected features for ML and clinical or genetic parameters

For ML studies, the vast majority of the studies reported performance results by utilizing ΔHbO. However, notable number studies also considers about ΔHb as a critical feature source (Cheng et al., 2019; Chiarelli et al., 2021; Cicalese et al., 2020; Crippa et al., 2017; Y. Fukuda et al., 2014; Guevara et al., 2020; Hernandez-Meza et al., 2018; Hernandez-Meza et al., 2017; Koike et al., 2017; J. Li et al., 2016; Parent et al., 2019; Rosas-Romero et al., 2019; Sato et al., 2013; Sirpal et al., 2019; Song et al., 2017; Sutoko et al., 2019; Xu et al., 2019; Xu, Hua, et al., 2020; Xu, Liu, et al., 2020; Xu et al., 2021; Yang et al., 2019; D. Yang et al., 2020). In some cases, depending on the measure, ΔHb might provide better classification accuracies compared to ΔHbO (Eken, 2021).

6.1. Current Problems in ML Applications On fNIRS

As we stated in the Introduction section, ML applications have a growing popularity in healthcare and particularly in diagnostic imaging (F. Jiang et al., 2017). Among these studies that we reviewed, we noticed that fNIRS provided us promising results about the usage of ML for diagnosis. In fact, in 2014, the Ministry of Health, Labour and Welfare in Japan approved fNIRS for health insurance as a supportive tool in laboratories for differential diagnosis of the depressive state of BP and SCZ from that of major depression (M. Fukuda, 2015). This approval was received after a multi-center study was carried out on patients with depressive state (Takizawa et al., 2014). However, this decision was strictly criticized afterward and it was stated that fNIRS is still immature to provide such a diagnosis due to less scientific evidence (Kato et al., 2017). At this point, lack of data diminishes the reliability of these studies. After having enormous amount of high-quality data with accurate and precise spatial information, it will be possible to develop more accurate ML models for diagnostic purposes. A very common problem in low sample size and high dimension datasets is; they tend to cause overfitting or underfitting. Low sample size in neuroimaging studies led to several problems in replicability (Turner et al., 2018) and cause high variance (Mumford, 2012). Low sample size with circular analysis cause higher classification accuracies which is possibly a misleading signature for diseases such as ADHD (see review (Pulini et al., 2019)). Also, cross-validation will cause a large error bias when the sample size is low (Varoquaux, 2018). Previous studies reported that low sample size-based classification studies reach higher accuracy when higher sample sizes lead lower accuracies (Schnack & Kahn, 2016).

Related to the sample size problem, lack of databases for fNIRS prevents to create big sample cohorts. Compared to fNIRS, there are several fMRI and MRI databases such as Alzheimer's Disease Neuroimaging Initiative (ADNI) (Jack et al., 2008), openfMRI (Poldrack et al., 2013; Poldrack & Gorgolewski, 2017). That allows data sharing among research groups which is critical for reaching big data cohorts. However, it is necessary to standardize some critical procedures such as anatomical positioning, and data acquisition parameters on common templates such as MNI (Tsuzuki et al., 2007) to achieve this. At this point, either utilizing MRI data of subjects or using 3D digitizers can be considered valid options to perform an accurate channel localization (Tsuzuki & Dan, 2014). Also, to assess regional biomarkers for every individual, cortical ROIs should be precisely defined and corresponding coordinates of this ROI should be reported. Some toolboxes provide anatomical information of channels by using MRI or 3D optode coordinate data such as AtlasViewer (Aasted et al., 2015), NIRS-SPM (Ye et al., 2009), NAP(Fekete et al., 2011a, 2011b) and fOLD (Zimeo Morais et al., 2018). This also will gain insight into further studies particularly comparing the results. Like databases, more multi-center studies should also be performed to generalize the performance of ML for diagnostic purposes. Until now, only two multi-center studies were reported for ADHD (Yasumura et al., 2017) and depression (Takizawa et al., 2014).

To reach more data, datasets with a standard near-infrared data format (.snirf) that includes spatial information are necessary. Therefore, not only the studies related to specific diseases or disorders but also meta-analyses related to these disorders might be published. Like fMRI, resting-state is a critical measurement technique in fNIRS (Niu & He, 2014). Also, several fNIRS-ML studies utilize resting-state measurements (Cheng et al., 2019; J. Li et al., 2016; Xu et al., 2019; Xu, Hua, et al., 2020; Xu, Liu, et al., 2020).

Another critical point is the pre-processing of fNIRS signals. A recent study that compares different pre-processing approaches revealed that ignoring the removal of task-evoked physiological noise led to different statistical results (Pfeifer et al., 2017). Also, a recent review showed that there is high variability among pre-processing methods carried out in fNIRS studies (Pinti et al., 2018). Standard pre-processing pipelines that cover all potential noise sources should be developed.

6.2. What should the further steps be to improve the biomarker research using fNIRS and ML?

Biomarkers can be classified as imaging and molecular biomarkers. The vast majority of fNIRS research mainly focuses on trying to associate the brain and behavior (or clinical measures). However, reliability of behavioral measures such as psychiatric paper-based test scores, and experimental responses is quite controversial. While defining a biomarker, one of the most critical criteria is to be subject-independent. A strong association between an objective feature with another one can be a valid indicator about the reliability of this biomarker. For instance, genetic features to associate with hemodynamic response can be used in addition to clinical variables (Ohi et al., 2011; Ohi et al., 2009; Reif et al., 2011; Takizawa, Hashimoto, et al., 2009; Takizawa, Tochigi, et al., 2009).

Multimodal approaches such as using EEG can also increase the classification accuracy like in BCI (see review (Naseer & Hong, 2015)). On the other hand, Diffuse Correlation Spectroscopy (DCS) which is the close cousin of fNIRS and allows us to measure cerebral blood flow (CBF) has recently been active in functional studies (see reviews (Durduran et al., 2010; Durduran & Yodh, 2014)) and recent studies reported that it has a three times higher brain to scalp sensitivity compared to fNIRS (Selb et al., 2014). Also, it was reported that an approximate 2 % change in Hb/HbO corresponds to 40 % change in blood flow (Durduran et al., 2004). Simultaneous measurement of CBF with Hb and HbO allows us to measure the cerebral metabolic rate of oxygen ($CMRO_2$). Although there are several fNIRS-based ML studies focused on clinical purposes, only two DCS-based ML studies are available and these studies both focused on predicting intracranial pressure using CBF data on non-human primates (Ruesch et al., 2020) and humans (Fischer et al., 2020). Not only neuroimaging-based measures but also other physiological measures such as ECG might also help to increase classification scores (Parent et al., 2019). As functional studies using DCS are gaining popularity and showing promising results for clinical approaches, ML applications will also be started to talk about DCS or its combination with fNIRS.

7. Conclusion

In this paper, we have reviewed the fNIRS-based ML studies that focus on biomarker research. Studies related to more than 15 different clinical populations and conditions were reviewed. It is widely known that fNIRS has several challenges such as data standardization, lack of data, and preprocessing problems. However, despite these pitfalls, there is a growing interest to understand the potential biomarkers to be used as discriminative parameters for different populations or conditions via fNIRS by utilizing ML approaches.

Acknowledgement

We would like to thank to Prof. Dr. Turgut Durduran from the Institute of Photonic Sciences (ICFO, Barcelona, Spain) for his valuable and constructive suggestions during the planning and development of this review.

Conflict of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

Data Availability Statement

No new data were created or analyzed in this study

Funding

There is no funding received related to this study.

8. References

Aasted, C. M., Yucel, M. A., Cooper, R. J., Dubb, J., Tsuzuki, D., Becerra, L., . . . Boas, D. A. (2015). Anatomical guidance for functional near-infrared spectroscopy: AtlasViewer tutorial. *Neurophotonics, 2*(2), 020801. doi:10.1117/1.NPh.2.2.020801

Abtahi, M., Borgheai, S. B., Jafari, R., Constant, N., Diouf, R., Shahriari, Y., & Mankodiya, K. (2020). Merging fNIRS-EEG Brain Monitoring and Body Motion Capture to Distinguish Parkinson's Disease. *IEEE Trans Neural Syst Rehabil Eng*. doi:10.1109/TNSRE.2020.2987888

Al-Shargie, F., Tang, T. B., & Kiguchi, M. (2017a). Assessment of mental stress effects on prefrontal cortical activities using canonical correlation analysis: an fNIRS-EEG study. *Biomed Opt Express, 8*(5), 2583-2598. doi:10.1364/BOE.8.002583

Al-Shargie, F., Tang, T. B., & Kiguchi, M. (2017b). Stress Assessment Based on Decision Fusion of EEG and fNIRS Signals. *IEEE Access, 5*, 19889-19896. doi:10.1109/access.2017.2754325

Alzheimer's Association. (2019). 2019 Alzheimer's disease facts and figures. *Alzheimer's & Dementia, 15*(3), 321-387. doi:10.1016/j.jalz.2019.01.010

American Psychiatric Association. (2013). *Diagnostic and Statistical Manual of Mental Disorders* (5th ed.). Washington, DC.

Apkarian, A. V., Bushnell, M. C., Treede, R. D., & Zubieta, J. K. (2005). Human brain mechanisms of pain perception and regulation in health and disease. *Eur J Pain, 9*(4), 463-484. doi:10.1016/j.ejpain.2004.11.001

Apkarian, A. V., Darbar, A., Krauss, B. R., Gelnar, P. A., & Szeverenyi, N. M. (1999). Differentiating cortical areas related to pain perception from stimulus identification: temporal analysis of fMRI activity. *J Neurophysiol, 81*(6), 2956-2963.

Arai, H., Takano, M., Miyakawa, K., Ota, T., Takahashi, T., Asaka, H., & Kawaguchi, T. (2006). A quantitative near-infrared spectroscopy study: a decrease in cerebral hemoglobin oxygenation in Alzheimer's disease and mild cognitive impairment. *Brain Cogn, 61*(2), 189-194. doi:10.1016/j.bandc.2005.12.012

Araki, T., Wake, R., Miyaoka, T., Kawakami, K., Nagahama, M., Furuya, M., . . . Horiguchi, J. (2014). The effects of combine treatment of memantine and donepezil on Alzheimer's disease patients and its relationship with cerebral blood flow in the prefrontal area. *Int J Geriatr Psychiatry, 29*(9), 881-889. doi:10.1002/gps.4074

Azechi, M., Iwase, M., Ikezawa, K., Takahashi, H., Canuet, L., Kurimoto, R., . . . Takeda, M. (2010). Discriminant analysis in schizophrenia and healthy subjects using prefrontal activation during frontal lobe tasks: a near-infrared spectroscopy. *Schizophr Res, 117*(1), 52-60. doi:10.1016/j.schres.2009.10.003

Baskak, B. (2018). The Place of Functional Near Infrared Spectroscopy in Psychiatry. *Noro Psikiyatr Ars, 55*(2), 103-104. doi:10.29399/npa.23249

Bature, F., Guinn, B. A., Pang, D., & Pappas, Y. (2017). Signs and symptoms preceding the diagnosis of Alzheimer's disease: a systematic scoping review of literature from 1937 to 2016. *BMJ Open, 7*(8), e015746. doi:10.1136/bmjopen-2016-015746

Becerra, L., Harris, W., Grant, M., George, E., Boas, D., & Borsook, D. (2009). Diffuse optical tomography activation in the somatosensory cortex: specific activation by painful vs. non-painful thermal stimuli. *PLoS One, 4*(11), e8016. doi:10.1371/journal.pone.0008016

Becerra, L., Harris, W., Joseph, D., Huppert, T., Boas, D. A., & Borsook, D. (2008). Diffuse optical tomography of pain and tactile stimulation: activation in cortical sensory and emotional systems. *Neuroimage, 41*(2), 252-259. doi:10.1016/j.neuroimage.2008.01.047

Becerra, L. R., Breiter, H. C., Stojanovic, M., Fishman, S., Edwards, A., Comite, A. R., . . . Borsook, D. (1999). Human brain activation under controlled thermal stimulation and habituation to noxious heat: an fMRI study. *Magn Reson Med, 41*(5), 1044-1057.

Belmaker, R. H., & Agam, G. (2008). Major depressive disorder. *N Engl J Med, 358*(1), 55-68. doi:10.1056/NEJMra073096

Binder, D. K., & Haut, S. R. (2013). Toward new paradigms of seizure detection. *Epilepsy Behav, 26*(3), 247-252. doi:10.1016/j.yebeh.2012.10.027

Biomarkers Definitions Working, G. (2001). Biomarkers and surrogate endpoints: preferred definitions and conceptual framework. *Clin Pharmacol Ther, 69*(3), 89-95. doi:10.1067/mcp.2001.113989

Boas, D. A., Elwell, C. E., Ferrari, M., & Taga, G. (2014). Twenty years of functional near-infrared spectroscopy: introduction for the special issue. *Neuroimage, 85, Part 1*, 1-5. doi:http://dx.doi.org/10.1016/j.neuroimage.2013.11.033

Borgheai, S. B., McLinden, J., Zisk, A. H., Hosni, S. I., Deligani, R. J., Abtahi, M., . . . Shahriari, Y. (2020). Enhancing Communication for People in Late-Stage ALS Using an fNIRS-Based BCI System. *IEEE Trans Neural Syst Rehabil Eng, 28*(5), 1198-1207. doi:10.1109/TNSRE.2020.2980772

Bornhovd, K., Quante, M., Glauche, V., Bromm, B., Weiller, C., & Buchel, C. (2002). Painful stimuli evoke different stimulus-response functions in the amygdala, prefrontal, insula and somatosensory cortex: a single-trial fMRI study. *Brain, 125*, 1326-1336.

Bridgemohan, C., Cochran, D. M., Howe, Y. J., Pawlowski, K., Zimmerman, A. W., Anderson, G. M., . . . Neumeyer, A. M. (2019). Investigating Potential Biomarkers in Autism Spectrum Disorder. *Front Integr Neurosci, 13*, 31. doi:10.3389/fnint.2019.00031

Brigadoi, S., Ceccherini, L., Cutini, S., Scarpa, F., Scatturin, P., Selb, J., . . . Cooper, R. J. (2014). Motion artifacts in functional near-infrared spectroscopy: a comparison of motion correction techniques applied to real cognitive data. *Neuroimage, 85 Pt 1*, 181-191. doi:10.1016/j.neuroimage.2013.04.082

Buchel, C., Bornhovd, K., Quante, M., Glauche, V., Bromm, B., & Weiller, C. (2002). Dissociable neural responses related to pain intensity, stimulus intensity, and stimulus awareness within the anterior cingulate cortex: A parametric single-trial laser functional magnetic resonance imaging study. *Journal of Neuroscience, 22*(3), 970-976.

Bushnell, M. C., Duncan, G. H., Hofbauer, R. K., Ha, B., Chen, J. I., & Carrier, B. (1999). Pain perception: is there a role for primary somatosensory cortex? *Proc Natl Acad Sci U S A, 96*(14), 7705-7709.

Cabral, C., Morgado, P. M., Campos Costa, D., Silveira, M., & Alzheimers Disease Neuroimaging, I. (2015). Predicting conversion from MCI to AD with FDG-PET brain images at different prodromal stages. *Comput Biol Med, 58*, 101-109. doi:10.1016/j.compbiomed.2015.01.003

Chan, E., Fogler, J. M., & Hammerness, P. G. (2016). Treatment of Attention-Deficit/Hyperactivity Disorder in Adolescents: A Systematic Review. *JAMA, 315*(18), 1997-2008. doi:10.1001/jama.2016.5453

Chao, J., Zheng, S., Wu, H., Wang, D., Zhang, X., Peng, H., & Hu, B. (2021). fNIRS Evidence for Distinguishing Patients with Major Depression and Healthy Controls. *IEEE Trans Neural Syst Rehabil Eng, PP*. doi:10.1109/TNSRE.2021.3115266

Chen, C. M., Yang, P., Wu, M. T., Chuang, T. C., & Huang, T. Y. (2019). Deriving and validating biomarkers associated with autism spectrum disorders from a large-scale resting-state database. *Sci Rep, 9*(1), 9043. doi:10.1038/s41598-019-45465-9

Cheng, H., Yu, J., Xu, L., & Li, J. (2019). Power spectrum of spontaneous cerebral homodynamic oscillation shows a distinct pattern in autism spectrum disorder. *Biomed Opt Express, 10*(3), 1383-1392. doi:10.1364/BOE.10.001383

Chiarelli, A. M., Perpetuini, D., Croce, P., Filippini, C., Cardone, D., Rotunno, L., . . . Merla, A. (2021). Evidence of Neurovascular Un-Coupling in Mild Alzheimer's Disease through Multimodal EEG-fNIRS and Multivariate Analysis of Resting-State Data. *Biomedicines, 9*(4). doi:10.3390/biomedicines9040337

Chou, P. H., Yao, Y. H., Zheng, R. X., Liou, Y. L., Liu, T. T., Lane, H. Y., . . . Wang, S. C. (2021). Deep Neural Network to Differentiate Brain Activity Between Patients With First-Episode Schizophrenia and Healthy Individuals: A Multi-Channel Near Infrared Spectroscopy Study. *Front Psychiatry, 12*, 655292. doi:10.3389/fpsyt.2021.655292

Chuang, C. C., Nakagome, K., Pu, S., Lan, T. H., Lee, C. Y., & Sun, C. W. (2014). Discriminant analysis of functional optical topography for schizophrenia diagnosis. *J Biomed Opt, 19*(1), 011006. doi:10.1117/1.JBO.19.1.011006

Cicalese, P. A., Li, R., Ahmadi, M. B., Wang, C., Francis, J. T., Selvaraj, S., . . . Zhang, Y. (2020). An EEG-fNIRS hybridization technique in the four-class classification of alzheimer's disease. *J Neurosci Methods, 336*, 108618. doi:10.1016/j.jneumeth.2020.108618

Clauw, D. J. (2014). Fibromyalgia: a clinical review. *JAMA, 311*(15), 1547-1555. doi:10.1001/jama.2014.3266

Coghill, R. C., Sang, C. N., Maisog, J. M., & Iadarola, M. J. (1999). Pain intensity processing within the human brain: a bilateral, distributed mechanism. *J Neurophysiol, 82*(4), 1934-1943.

Craik, A., He, Y., & Contreras-Vidal, J. L. (2019). Deep learning for electroencephalogram (EEG) classification tasks: a review. *J Neural Eng, 16*(3), 031001. doi:10.1088/1741-2552/ab0ab5

Crippa, A., Salvatore, C., Molteni, E., Mauri, M., Salandi, A., Trabattoni, S., . . . Castiglioni, I. (2017). The Utility of a Computerized Algorithm Based on a Multi-Domain Profile of Measures for the Diagnosis of Attention Deficit/Hyperactivity Disorder. *Front Psychiatry, 8*, 189. doi:10.3389/fpsyt.2017.00189

Dadgostar, M., Setarehdan, S. K., Shahzadi, S., & Akin, A. (2018). CLASSIFICATION OF SCHIZOPHRENIA USING SVM VIA fNIRS. *Biomedical Engineering: Applications, Basis and Communications, 30*(02), 1850008. doi:10.4015/S1016237218500084

Dahan, A., Dubnov, Y. A., Popkov, A. Y., Gutman, I., & Probolovski, H. G. (2020). Brief Report: Classification of Autistic Traits According to Brain Activity Recoded by fNIRS Using epsilon-Complexity Coefficients. *J Autism Dev Disord*. doi:10.1007/s10803-020-04793-w

Davison, T. E., McCabe, M. P., & Mellor, D. (2009). An examination of the "gold standard" diagnosis of major depression in aged-care settings. *Am J Geriatr Psychiatry, 17*(5), 359-367. doi:10.1097/JGP.0b013e318190b901

Dawe, S. R., Windsor, J. A., Broeders, J. A., Cregan, P. C., Hewett, P. J., & Maddern, G. J. (2014). A systematic review of surgical skills transfer after simulation-based training: laparoscopic cholecystectomy and endoscopy. *Ann Surg, 259*(2), 236-248. doi:10.1097/SLA.0000000000000245

de Filippis, R., Carbone, E. A., Gaetano, R., Bruni, A., Pugliese, V., Segura-Garcia, C., & De Fazio, P. (2019). Machine learning techniques in a structural and functional MRI diagnostic approach in schizophrenia: a systematic review. *Neuropsychiatr Dis Treat, 15*, 1605-1627. doi:10.2147/NDT.S202418

Deligani, R. J., Borgheai, S. B., McLinden, J., & Shahriari, Y. (2021). Multimodal fusion of EEG-fNIRS: a mutual information-based hybrid classification framework. *Biomed Opt Express, 12*(3), 1635-1650. doi:10.1364/BOE.413666

DeMaagd, G., & Philip, A. (2015). Parkinson's Disease and Its Management: Part 1: Disease Entity, Risk Factors, Pathophysiology, Clinical Presentation, and Diagnosis. *P T, 40*(8), 504-532.

Derbyshire, S. W., Jones, A. K., Gyulai, F., Clark, S., Townsend, D., & Firestone, L. L. (1997). Pain processing during three levels of noxious stimulation produces differential patterns of central activity. *Pain, 73*(3), 431-445.

Duffy, I. R., Boyle, A. J., & Vasdev, N. (2019). Improving PET Imaging Acquisition and Analysis With Machine Learning: A Narrative Review With Focus on Alzheimer's Disease and Oncology. *Mol Imaging, 18*, 1536012119869070. doi:10.1177/1536012119869070

Durduran, T., Choe, R., Baker, W. B., & Yodh, A. G. (2010). Diffuse Optics for Tissue Monitoring and Tomography. *Rep Prog Phys, 73*(7). doi:10.1088/0034-4885/73/7/076701

Durduran, T., & Yodh, A. G. (2014). Diffuse correlation spectroscopy for non-invasive, micro-vascular cerebral blood flow measurement. *Neuroimage, 85 Pt 1*, 51-63. doi:10.1016/j.neuroimage.2013.06.017

Durduran, T., Yu, G., Burnett, M. G., Detre, J. A., Greenberg, J. H., Wang, J., . . . Yodh, A. G. (2004). Diffuse optical measurement of blood flow, blood oxygenation, and metabolism in a human brain during sensorimotor cortex activation. *Opt Lett, 29*(15), 1766-1768.

Eastmond, C., Subedi, A., De, S., & Intes, X. (2022). Deep learning in fNIRS: a review. *Neurophotonics, 9*(4), 041411. doi:10.1117/1.NPh.9.4.041411

Ehlis, A. C., Schneider, S., Dresler, T., & Fallgatter, A. J. (2014). Application of functional near-infrared spectroscopy in psychiatry. *Neuroimage, 85 Pt 1*, 478-488. doi:10.1016/j.neuroimage.2013.03.067

Einalou, Z., Maghooli, K., Setarehdan, S. K., & Akin, A. (2016). Effective channels in classification and functional connectivity pattern of prefrontal cortex by functional near infrared spectroscopy signals. *Optik, 127*(6), 3271-3275. doi:https://doi.org/10.1016/j.ijleo.2015.12.090

Eken, A. (2021). Assessment of flourishing levels of individuals by using resting-state fNIRS with different functional connectivity measures. *Biomedical Signal Processing and Control, 68*, 102645. doi:https://doi.org/10.1016/j.bspc.2021.102645

Eken, A., Akaslan, D. S., Baskak, B., & Munir, K. (2022). Diagnostic classification of schizophrenia and bipolar disorder by using dynamic functional connectivity: An fNIRS study. *J Neurosci Methods, 376*, 109596. doi:10.1016/j.jneumeth.2022.109596

Eken, A., Colak, B., Bal, N. B., Kusman, A., Kizilpinar, S. C., Akaslan, D. S., & Baskak, B. (2019). Hyperparameter-tuned prediction of somatic symptom disorder using functional near-infrared spectroscopy-based dynamic functional connectivity. *J Neural Eng, 17*(1), 016012. doi:10.1088/1741-2552/ab50b2

Erdogan, S. B., Yukselen, G., Yegul, M. M., Usanmaz, R., Kiran, E., Derman, O., & Akin, A. (2021). Identification of impulsive adolescents with a functional near infrared spectroscopy (fNIRS) based decision support system. *J Neural Eng, 18*(5). doi:10.1088/1741-2552/ac23bb

Etchell, A. C., Civier, O., Ballard, K. J., & Sowman, P. F. (2018). A systematic literature review of neuroimaging research on developmental stuttering between 1995 and 2016. *J Fluency Disord, 55*, 6-45. doi:10.1016/j.jfludis.2017.03.007

Fekete, T., Rubin, D., Carlson, J. M., & Mujica-Parodi, L. R. (2011a). The NIRS Analysis Package: noise reduction and statistical inference. *PLoS One, 6*(9), e24322. doi:10.1371/journal.pone.0024322

Fekete, T., Rubin, D., Carlson, J. M., & Mujica-Parodi, L. R. (2011b). A stand-alone method for anatomical localization of NIRS measurements. *Neuroimage, 56*(4), 2080-2088. doi:10.1016/j.neuroimage.2011.03.068

Fernandez Rojas, R., Huang, X., & Ou, K. L. (2017). Toward a functional near-infrared spectroscopy-based monitoring of pain assessment for nonverbal patients. *J Biomed Opt, 22*(10), 1-12. doi:10.1117/1.JBO.22.10.106013

Fernandez Rojas, R., Huang, X., & Ou, K. L. (2019). A Machine Learning Approach for the Identification of a Biomarker of Human Pain using fNIRS. *Sci Rep, 9*(1), 5645. doi:10.1038/s41598-019-42098-w

Fischer, J. B., Ghouse, A., Tagliabue, S., Maruccia, F., Rey-Perez, A., Baguena, M., . . . Durduran, T. (2020). Non-invasive estimation of intracranial pressure by diffuse optics - a proof-of-concept study. *J Neurotrauma*. doi:10.1089/neu.2019.6965

Fisher, R. S., Acevedo, C., Arzimanoglou, A., Bogacz, A., Cross, J. H., Elger, C. E., . . . Wiebe, S. (2014). ILAE official report: a practical clinical definition of epilepsy. *Epilepsia, 55*(4), 475-482. doi:10.1111/epi.12550

Franceschini, M. A., Fantini, S., Thompson, J. H., Culver, J. P., & Boas, D. A. (2003). Hemodynamic evoked response of the sensorimotor cortex measured noninvasively with near-infrared optical imaging. *Psychophysiology, 40*(4), 548-560.

Fukuda, M. (2015). [Optical Topography as an Auxiliary Laboratory Test for Differential Diagnosis of Depressive State: Clinical Application of Near-infrared Spectroscopy (NIRS) as the First Trial for Approved Laboratory Tests in Psychiatry]. *Seishin Shinkeigaku Zasshi, 117*(2), 79-93.

Fukuda, Y., Ishikawa, W., Kanayama, R., Matsumoto, T., Takemura, N., & Sakatani, K. (2014). Bayesian prediction of anxiety level in aged people at rest using 2-channel NIRS data from prefrontal cortex. *Adv Exp Med Biol, 812*, 303-308. doi:10.1007/978-1-4939-0620-8_40

Gao, S., Calhoun, V. D., & Sui, J. (2018). Machine learning in major depression: From classification to treatment outcome prediction. *CNS Neurosci Ther, 24*(11), 1037-1052. doi:10.1111/cns.13048

Geffen, J., & Forster, K. (2018). Treatment of adult ADHD: a clinical perspective. *Ther Adv Psychopharmacol, 8*(1), 25-32. doi:10.1177/2045125317734977

Gentile, J. P., & Atiq, R. (2006). Psychotherapy for the patient with adult ADHD. *Psychiatry (Edgmont), 3*(8), 31-35.

Ghajar, J. (2000). Traumatic brain injury. *Lancet, 356*(9233), 923-929. doi:10.1016/S0140-6736(00)02689-1

Gokcay, D., Eken, A., & Baltaci, S. (2019). Binary Classification Using Neural and Clinical Features: An Application in Fibromyalgia With Likelihood-Based Decision Level Fusion. *IEEE J Biomed Health Inform, 23*(4), 1490-1498. doi:10.1109/JBHI.2018.2844300

Goldani, A. A., Downs, S. R., Widjaja, F., Lawton, B., & Hendren, R. L. (2014). Biomarkers in autism. *Front Psychiatry, 5*, 100. doi:10.3389/fpsyt.2014.00100

Gu, Y., Miao, S., Han, J., Liang, Z., Ouyang, G., Yang, J., & Li, X. (2018). Identifying ADHD children using hemodynamic responses during a working memory task measured by functional near-infrared spectroscopy. *J Neural Eng, 15*(3), 035005. doi:10.1088/1741-2552/aa9ee9

Guevara, E., Flores-Castro, J. A., Peng, K., Nguyen, D. K., Lesage, F., Pouliot, P., & Rosas-Romero, R. (2020). Prediction of epileptic seizures using fNIRS and machine learning. *Journal of Intelligent & Fuzzy Systems, 38*(2), 2055-2068. doi:10.3233/jifs-190738

Güven, A., Altınkaynak, M., Dolu, N., İzzetoğlu, M., Pektaş, F., Özmen, S., . . . Batbat, T. (2020). Combining functional near-infrared spectroscopy and EEG measurements for the diagnosis of attention-deficit hyperactivity disorder. *Neural Computing and Applications, 32*(12), 8367-8380. doi:10.1007/s00521-019-04294-7

Hahn, T., Marquand, A. F., Plichta, M. M., Ehlis, A. C., Schecklmann, M. W., Dresler, T., . . . Fallgatter, A. J. (2013). A novel approach to probabilistic biomarker-based classification using functional near-infrared spectroscopy. *Hum Brain Mapp, 34*(5), 1102-1114. doi:10.1002/hbm.21497

Hakimi, N., Jodeiri, A., Mirbagheri, M., & Setarehdan, S. K. (2020). Proposing a convolutional neural network for stress assessment by means of derived heart rate from functional near infrared spectroscopy. *Comput Biol Med, 121*, 103810. doi:https://doi.org/10.1016/j.compbiomed.2020.103810

Han, B. I., Lee, H. W., Kim, T. Y., Lim, J. S., & Shin, K. S. (2009). Tinnitus: characteristics, causes, mechanisms, and treatments. *J Clin Neurol, 5*(1), 11-19. doi:10.3988/jcn.2009.5.1.11

Hansen, B. H., Oerbeck, B., Skirbekk, B., Petrovski, B. E., & Kristensen, H. (2018). Neurodevelopmental disorders: prevalence and comorbidity in children referred to mental health services. *Nord J Psychiatry, 72*(4), 285-291. doi:10.1080/08039488.2018.1444087

He, R., Yan, X., Guo, J., Xu, Q., Tang, B., & Sun, Q. (2018). Recent Advances in Biomarkers for Parkinson's Disease. *Front Aging Neurosci, 10*, 305. doi:10.3389/fnagi.2018.00305

Henderson, T. A., van Lierop, M. J., McLean, M., Uszler, J. M., Thornton, J. F., Siow, Y. H., . . . Cohen, P. (2020). Functional Neuroimaging in Psychiatry-Aiding in Diagnosis and Guiding

Treatment. What the American Psychiatric Association Does Not Know. *Front Psychiatry, 11*, 276. doi:10.3389/fpsyt.2020.00276

Henry, J., & Crawford, J. R. (2005). A meta-analytic review of verbal fluency deficits in depression. *J Clin Exp Neuropsychol, 27*(1), 78-101. doi:10.1080/138033990513654

Hernandez-Meza, G., Izzetoglu, M., Osbakken, M., Green, M., Abubakar, H., & Izzetoglu, K. (2018). Investigation of optical neuro-monitoring technique for detection of maintenance and emergence states during general anesthesia. *J Clin Monit Comput, 32*(1), 147-163. doi:10.1007/s10877-017-9998-x

Hernandez-Meza, G., Izzetoglu, M., Sacan, A., Green, M., & Izzetoglu, K. (2017). Investigation of data-driven optical neuromonitoring approach during general anesthesia with sevoflurane. *Neurophotonics, 4*(4), 041408. doi:10.1117/1.NPh.4.4.041408

Herrmann, M. J., Langer, J. B., Jacob, C., Ehlis, A. C., & Fallgatter, A. J. (2008). Reduced prefrontal oxygenation in Alzheimer disease during verbal fluency tasks. *Am J Geriatr Psychiatry, 16*(2), 125-135. doi:10.1097/JGP.0b013e3180cc1fbc

Ho, C. S., Chan, Y. L., Tan, T. W., Tay, G. W., & Tang, T. B. (2022). Improving the diagnostic accuracy for major depressive disorder using machine learning algorithms integrating clinical and near-infrared spectroscopy data. *J Psychiatr Res, 147*, 194-202. doi:10.1016/j.jpsychires.2022.01.026

Ho, T. K. K., Kim, M., Jeon, Y., Kim, B. C., Kim, J. G., Lee, K. H., . . . Gwak, J. (2022). Deep Learning-Based Multilevel Classification of Alzheimer's Disease Using Non-invasive Functional Near-Infrared Spectroscopy. *Front Aging Neurosci, 14*, 810125. doi:10.3389/fnagi.2022.810125

Hochberg, M. C., Silman, A. J., Smolen, J. S., Weinblatt, M. E., & Weisman, M. H. (2003). *Rheumatology* (3 ed.). New York: Mosby.

Hock, C., Villringer, K., Muller-Spahn, F., Wenzel, R., Heekeren, H., Schuh-Hofer, S., . . . Villringer, A. (1997). Decrease in parietal cerebral hemoglobin oxygenation during performance of a verbal fluency task in patients with Alzheimer's disease monitored by means of near-infrared spectroscopy (NIRS)--correlation with simultaneous rCBF-PET measurements. *Brain Res, 755*(2), 293-303. doi:10.1016/s0006-8993(97)00122-4

Hojjati, S. H., Ebrahimzadeh, A., Khazaee, A., Babajani-Feremi, A., & Alzheimer's Disease Neuroimaging, I. (2017). Predicting conversion from MCI to AD using resting-state fMRI, graph theoretical approach and SVM. *J Neurosci Methods, 282*, 69-80. doi:10.1016/j.jneumeth.2017.03.006

Hojjati, S. H., Ebrahimzadeh, A., Khazaee, A., Babajani-Feremi, A., & Alzheimer's Disease Neuroimaging, I. (2018). Predicting conversion from MCI to AD by integrating rs-fMRI and structural MRI. *Comput Biol Med, 102*, 30-39. doi:10.1016/j.compbiomed.2018.09.004

Homae, F., Watanabe, H., Otobe, T., Nakano, T., Go, T., Konishi, Y., & Taga, G. (2010). Development of global cortical networks in early infancy. *J Neurosci, 30*(14), 4877-4882. doi:10.1523/JNEUROSCI.5618-09.2010

Hosseini, R., Walsh, B., Tian, F., & Wang, S. (2018). An fNIRS-Based Feature Learning and Classification Framework to Distinguish Hemodynamic Patterns in Children Who Stutter. *IEEE Trans Neural Syst Rehabil Eng, 26*(6), 1254-1263. doi:10.1109/TNSRE.2018.2829083

Huss, R. (2015). Chapter 19 - Biomarkers. In A. Atala & J. G. Allickson (Eds.), *Translational Regenerative Medicine* (pp. 235-241). Boston: Academic Press.

Ichikawa, H., Kitazono, J., Nagata, K., Manda, A., Shimamura, K., Sakuta, R., . . . Kakigi, R. (2014). Novel method to classify hemodynamic response obtained using multi-channel fNIRS measurements into two groups: exploring the combinations of channels. *Frontiers in Human Neuroscience, 8*, 480-480. doi:10.3389/fnhum.2014.00480

Irani, F., Platek, S. M., Bunce, S., Ruocco, A. C., & Chute, D. (2007). Functional near infrared spectroscopy (fNIRS): an emerging neuroimaging technology with important applications for the study of brain disorders. *Clin Neuropsychol, 21*(1), 9-37. doi:10.1080/13854040600910018

Ishii-Takahashi, A., Takizawa, R., Nishimura, Y., Kawakubo, Y., Hamada, K., Okuhata, S., . . . Kano, Y. (2015). Neuroimaging-Aided Prediction of the Effect of Methylphenidate in Children with Attention-Deficit Hyperactivity Disorder: A Randomized Controlled Trial. *Neuropsychopharmacology, 40*(12), 2676-2685. doi:10.1038/npp.2015.128

Jack, C. R., Jr., Bernstein, M. A., Fox, N. C., Thompson, P., Alexander, G., Harvey, D., . . . Weiner, M. W. (2008). The Alzheimer's Disease Neuroimaging Initiative (ADNI): MRI methods. *J Magn Reson Imaging, 27*(4), 685-691. doi:10.1002/jmri.21049

Jackson, E. S., Wijeakumar, S., Beal, D. S., Brown, B., Zebrowski, P., & Spencer, J. P. (2019). A fNIRS Investigation of Speech Planning and Execution in Adults Who Stutter. *Neuroscience, 406*, 73-85. doi:10.1016/j.neuroscience.2019.02.032

Ji, X., Quan, W., Yang, L., Chen, J., Wang, J., & Wu, T. (2020). Classification of Schizophrenia by Seed-based Functional Connectivity using Prefronto-Temporal Functional Near Infrared Spectroscopy. *J Neurosci Methods*, 108874. doi:10.1016/j.jneumeth.2020.108874

Jiang, F., Jiang, Y., Zhi, H., Dong, Y., Li, H., Ma, S., . . . Wang, Y. (2017). Artificial intelligence in healthcare: past, present and future. *Stroke Vasc Neurol, 2*(4), 230-243. doi:10.1136/svn-2017-000101

Jiang, J., Lu, C., Peng, D., Zhu, C., & Howell, P. (2012). Classification of types of stuttering symptoms based on brain activity. *PLoS One, 7*(6), e39747. doi:10.1371/journal.pone.0039747

Johannesen, J. K., Bi, J., Jiang, R., Kenney, J. G., & Chen, C. A. (2016). Machine learning identification of EEG features predicting working memory performance in schizophrenia and healthy adults. *Neuropsychiatr Electrophysiol, 2*, 3. doi:10.1186/s40810-016-0017-0

Jourdan Moser, S., Cutini, S., Weber, P., & Schroeter, M. L. (2009). Right prefrontal brain activation due to Stroop interference is altered in attention-deficit hyperactivity disorder - A functional near-infrared spectroscopy study. *Psychiatry Res, 173*(3), 190-195. doi:10.1016/j.pscychresns.2008.10.003

Kalia, L. V., & Lang, A. E. (2015). Parkinson's disease. *Lancet, 386*(9996), 896-912. doi:10.1016/S0140-6736(14)61393-3

Karamzadeh, N., Amyot, F., Kenney, K., Anderson, A., Chowdhry, F., Dashtestani, H., . . . Gandjbakhche, A. H. (2016). A machine learning approach to identify functional biomarkers in human prefrontal cortex for individuals with traumatic brain injury using functional near-infrared spectroscopy. *Brain Behav, 6*(11), e00541. doi:10.1002/brb3.541

Kato, T., Sakai, N., Watanabe, Y., & Nomura, S. (2017). Possibility of over-diagnosis of bipolar disorder due to near-infrared spectroscopy. *Psychiatry Clin Neurosci, 71*(12), 843. doi:10.1111/pcn.12561

Keles, H. O., Cengiz, C., Demiral, I., Ozmen, M. M., & Omurtag, A. (2021). High density optical neuroimaging predicts surgeons's subjective experience and skill levels. *PLoS One, 16*(2), e0247117. doi:10.1371/journal.pone.0247117

Kenji, K., & Larry, A. R. (1992). The feature selection problem: traditional methods and a new algorithm %@ 0262510634 *Proceedings of the tenth national conference on Artificial intelligence* (pp. 129–134). San Jose, California: AAAI Press.

Kennis, M., Gerritsen, L., van Dalen, M., Williams, A., Cuijpers, P., & Bockting, C. (2020). Prospective biomarkers of major depressive disorder: a systematic review and meta-analysis. *Mol Psychiatry, 25*(2), 321-338. doi:10.1038/s41380-019-0585-z

Kern, J. K., Geier, D. A., Sykes, L. K., Geier, M. R., & Deth, R. C. (2015). Are ASD and ADHD a Continuum? A Comparison of Pathophysiological Similarities Between the Disorders. *J Atten Disord, 19*(9), 805-827. doi:10.1177/1087054712459886

Khazaee, A., Ebrahimzadeh, A., & Babajani-Feremi, A. (2016). Application of advanced machine learning methods on resting-state fMRI network for identification of mild cognitive impairment and Alzheimer's disease. *Brain Imaging Behav, 10*(3), 799-817. doi:10.1007/s11682-015-9448-7

Khazaee, A., Ebrahimzadeh, A., & Babajani-Feremi, A. (2017). Classification of patients with MCI and AD from healthy controls using directed graph measures of resting-state fMRI. *Behav Brain Res, 322*, 339-350. doi:https://doi.org/10.1016/j.bbr.2016.06.043

Khosla, M., Jamison, K., Ngo, G. H., Kuceyeski, A., & Sabuncu, M. R. (2019). Machine learning in resting-state fMRI analysis. *Magn Reson Imaging, 64*, 101-121. doi:10.1016/j.mri.2019.05.031

Kim, E., Yu, J. W., Kim, B., Lim, S. H., Lee, S. H., Kim, K., . . . Choi, J. W. (2021). Refined prefrontal working memory network as a neuromarker for Alzheimer's disease. *Biomed Opt Express, 12*(11), 7199-7222. doi:10.1364/BOE.438926

Kim, Y. S., Leventhal, B. L., Koh, Y. J., Fombonne, E., Laska, E., Lim, E. C., . . . Grinker, R. R. (2011). Prevalence of autism spectrum disorders in a total population sample. *Am J Psychiatry, 168*(9), 904-912. doi:10.1176/appi.ajp.2011.10101532

Koch, S. P., Habermehl, C., Mehnert, J., Schmitz, C. H., Holtze, S., Villringer, A., . . . Obrig, H. (2010). High-resolution optical functional mapping of the human somatosensory cortex. *Front Neuroenergetics, 2*, 12. doi:10.3389/fnene.2010.00012

Koike, S., Nishimura, Y., Takizawa, R., Yahata, N., & Kasai, K. (2013). Near-infrared spectroscopy in schizophrenia: a possible biomarker for predicting clinical outcome and treatment response. *Front Psychiatry, 4*, 145. doi:10.3389/fpsyt.2013.00145

Koike, S., Satomura, Y., Kawasaki, S., Nishimura, Y., Kinoshita, A., Sakurada, H., . . . Kasai, K. (2017). Application of functional near infrared spectroscopy as supplementary examination for diagnosis of clinical stages of psychosis spectrum. *Psychiatry Clin Neurosci, 71*(12), 794-806. doi:10.1111/pcn.12551

Krueger, F., Spampinato, M. V., Barbey, A. K., Huey, E. D., Morland, T., & Grafman, J. (2009). The frontopolar cortex mediates event knowledge complexity: a parametric functional MRI study. *Neuroreport, 20*(12), 1093-1097. doi:10.1097/WNR.0b013e32832e7ea5

Li, J., Qiu, L., Xu, L., Pedapati, E. V., Erickson, C. A., & Sunar, U. (2016). Characterization of autism spectrum disorder with spontaneous hemodynamic activity. *Biomed Opt Express, 7*(10), 3871-3881. doi:10.1364/BOE.7.003871

Li, K., Wang, L., Cheng, J., Zhang, L., Duan, H., Lu, Y., . . . Chen, J. (2002). [Screening and cloning gene of hepatocyte protein interacting with hepatitis C virus core protein]. *Zhonghua Shi Yan He Lin Chuang Bing Du Xue Za Zhi, 16*(4), 351-353.

Li, R., Rui, G., Chen, W., Li, S., Schulz, P. E., & Zhang, Y. (2018). Early Detection of Alzheimer's Disease Using Non-invasive Near-Infrared Spectroscopy. *Front Aging Neurosci, 10*, 366. doi:10.3389/fnagi.2018.00366

Li, Z., McIntyre, R. S., Husain, S. F., Ho, R., Tran, B. X., Nguyen, H. T., . . . Chen, N. (2022). Identifying neuroimaging biomarkers of major depressive disorder from cortical hemodynamic responses using machine learning approaches. *EBioMedicine, 79*, 104027. doi:10.1016/j.ebiom.2022.104027

Li, Z., Wang, Y., Quan, W., Wu, T., & Lv, B. (2015). Evaluation of different classification methods for the diagnosis of schizophrenia based on functional near-infrared spectroscopy. *J Neurosci Methods, 241*, 101-110. doi:10.1016/j.jneumeth.2014.12.020

Lopez, O. L. (2013). Mild cognitive impairment. *Continuum (Minneap Minn), 19*(2 Dementia), 411-424. doi:10.1212/01.CON.0000429175.29601.97

Lopez-Sola, M., Woo, C. W., Pujol, J., Deus, J., Harrison, B. J., Monfort, J., & Wager, T. D. (2017). Towards a neurophysiological signature for fibromyalgia. *Pain, 158*(1), 34-47. doi:10.1097/j.pain.0000000000000707

Lord, C., Risi, S., Lambrecht, L., Cook, E. H., Jr., Leventhal, B. L., DiLavore, P. C., . . . Rutter, M. (2000). The autism diagnostic observation schedule-generic: a standard measure of social and communication deficits associated with the spectrum of autism. *J Autism Dev Disord, 30*(3), 205-223.

Lui, F., Duzzi, D., Corradini, M., Serafini, M., Baraldi, P., & Porro, C. A. (2008). Touch or pain? Spatio-temporal patterns of cortical fMRI activity following brief mechanical stimuli. *Pain, 138*(2), 362-374. doi:10.1016/j.pain.2008.01.010

Månsson, H. (2000). Childhood stuttering: Incidence and development. *Journal of Fluency Disorders, 25*(1), 47-57. doi:https://doi.org/10.1016/S0094-730X(99)00023-6

Mayeux, R., & Stern, Y. (2012). Epidemiology of Alzheimer disease. *Cold Spring Harb Perspect Med, 2*(8). doi:10.1101/cshperspect.a006239

McCutcheon, R. A., Reis Marques, T., & Howes, O. D. (2019). Schizophrenia-An Overview. *JAMA Psychiatry*, 1-10. doi:10.1001/jamapsychiatry.2019.3360

Mehnert, J., Akhrif, A., Telkemeyer, S., Rossi, S., Schmitz, C. H., Steinbrink, J., . . . Neufang, S. (2013). Developmental changes in brain activation and functional connectivity during response inhibition in the early childhood brain. *Brain Dev, 35*(10), 894-904. doi:10.1016/j.braindev.2012.11.006

Moeller, F. G., Barratt, E. S., Dougherty, D. M., Schmitz, J. M., & Swann, A. C. (2001). Psychiatric aspects of impulsivity. *Am J Psychiatry, 158*(11), 1783-1793. doi:10.1176/appi.ajp.158.11.1783

Monden, Y., Dan, H., Nagashima, M., Dan, I., Tsuzuki, D., Kyutoku, Y., . . . Momoi, M. Y. (2012). Right prefrontal activation as a neuro-functional biomarker for monitoring acute effects of methylphenidate in ADHD children: An fNIRS study. *Neuroimage Clin, 1*(1), 131-140. doi:10.1016/j.nicl.2012.10.001

Montero-Hernandez, S., Orihuela-Espina, F., Sucar, E. L., Pinti, P., Hamilton, A., Burgess, P., & Tachtsidis, I. (2018). Estimating Functional Connectivity Symmetry between Oxy- and Deoxy-Haemoglobin: Implications for fNIRS Connectivity Analysis. *Algorithms, 11*(5). doi:10.3390/a11050070

Moulton, E. A., Keaser, M. L., Gullapalli, R. P., & Greenspan, J. D. (2005). Regional intensive and temporal patterns of functional MRI activation distinguishing noxious and innocuous contact heat. *J Neurophysiol, 93*(4), 2183-2193. doi:10.1152/jn.01025.2004

Mumford, J. A. (2012). A power calculation guide for fMRI studies. *Soc Cogn Affect Neurosci, 7*(6), 738-742. doi:10.1093/scan/nss059

Naseer, N., & Hong, K. S. (2015). fNIRS-based brain-computer interfaces: a review. *Front Hum Neurosci, 9*, 3. doi:10.3389/fnhum.2015.00003

Nemani, A., Kruger, U., Cooper, C. A., Schwaitzberg, S. D., Intes, X., & De, S. (2019). Objective assessment of surgical skill transfer using non-invasive brain imaging. *Surg Endosc, 33*(8), 2485-2494. doi:10.1007/s00464-018-6535-z

Niu, H., & He, Y. (2014). Resting-state functional brain connectivity: lessons from functional near-infrared spectroscopy. *Neuroscientist, 20*(2), 173-188. doi:10.1177/1073858413502707

Niu, H., Khadka, S., Tian, F., Lin, Z. J., Lu, C., Zhu, C., & Liu, H. (2011). Resting-state functional connectivity assessed with two diffuse optical tomographic systems. *J Biomed Opt, 16*(4), 046006. doi:10.1117/1.3561687

Ohi, K., Hashimoto, R., Yasuda, Y., Fukumoto, M., Yamamori, H., Umeda-Yano, S., . . . Takeda, M. (2011). The SIGMAR1 gene is associated with a risk of schizophrenia and activation of the prefrontal cortex. *Prog Neuropsychopharmacol Biol Psychiatry, 35*(5), 1309-1315. doi:10.1016/j.pnpbp.2011.04.008

Ohi, K., Hashimoto, R., Yasuda, Y., Kiribayashi, M., Iike, N., Yoshida, T., . . . Takeda, M. (2009). TATA Box-Binding Protein gene is associated with risk for schizophrenia, age at onset and prefrontal function. *Genes, Brain and Behavior, 8*(4), 473-480. doi:10.1111/j.1601-183X.2009.00497.x

Okada, F., Tokumitsu, Y., Hoshi, Y., & Tamura, M. (1994). Impaired interhemispheric integration in brain oxygenation and hemodynamics in schizophrenia. *Eur Arch Psychiatry Clin Neurosci, 244*(1), 17-25. doi:10.1007/bf02279807

Ou, Y., Liu, F., Chen, J., Pan, P., Wu, R., Su, Q., . . . Guo, W. (2018). Increased coherence-based regional homogeneity in resting-state patients with first-episode, drug-naive somatization disorder. *J Affect Disord, 235*, 150-154. doi:10.1016/j.jad.2018.04.036

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., . . . Moher, D. (2021). Updating guidance for reporting systematic reviews: development of the PRISMA 2020 statement. *J Clin Epidemiol, 134*, 103-112. doi:10.1016/j.jclinepi.2021.02.003

Pagnozzi, A. M., Conti, E., Calderoni, S., Fripp, J., & Rose, S. E. (2018). A systematic review of structural MRI biomarkers in autism spectrum disorder: A machine learning perspective. *Int J Dev Neurosci, 71*, 68-82. doi:10.1016/j.ijdevneu.2018.08.010

Pan, P., Ou, Y., Su, Q., Liu, F., Chen, J., Zhao, J., & Guo, W. (2019). Voxel-based global-brain functional connectivity alterations in first-episode drug-naive patients with somatization disorder. *J Affect Disord, 254*, 82-89. doi:10.1016/j.jad.2019.04.099

Parent, M., Peysakhovich, V., Mandrick, K., Tremblay, S., & Causse, M. (2019). The diagnosticity of psychophysiological signatures: Can we disentangle mental workload from acute stress with ECG and fNIRS? *Int J Psychophysiol, 146*, 139-147. doi:10.1016/j.ijpsycho.2019.09.005

Parvandeh, S., Yeh, H. W., Paulus, M. P., & McKinney, B. A. (2020). Consensus Features Nested Cross-Validation. *Bioinformatics*. doi:10.1093/bioinformatics/btaa046

Perez, H. R., & Stoeckle, J. H. (2016). Stuttering: Clinical and research update. *Can Fam Physician, 62*(6), 479-484.

Pfeifer, M. D., Scholkmann, F., & Labruyere, R. (2017). Signal Processing in Functional Near-Infrared Spectroscopy (fNIRS): Methodological Differences Lead to Different Statistical Results. *Front Hum Neurosci, 11*, 641. doi:10.3389/fnhum.2017.00641

Pinti, P., Scholkmann, F., Hamilton, A., Burgess, P., & Tachtsidis, I. (2018). Current Status and Issues Regarding Pre-processing of fNIRS Neuroimaging Data: An Investigation of Diverse Signal Filtering Methods Within a General Linear Model Framework. *Front Hum Neurosci, 12*, 505. doi:10.3389/fnhum.2018.00505

Poineau, F., Johnstone, E. V., Weck, P. F., Forster, P. M., Kim, E., Czerwinski, K. R., & Sattelberger, A. P. (2012). beta-Technetium trichloride: formation, structure, and first-principles calculations. *Inorg Chem, 51*(9), 4915-4917. doi:10.1021/ic300612k

Poldrack, R. A., Barch, D. M., Mitchell, J. P., Wager, T. D., Wagner, A. D., Devlin, J. T., . . . Milham, M. P. (2013). Toward open sharing of task-based fMRI data: the OpenfMRI project. *Front Neuroinform, 7*, 12. doi:10.3389/fninf.2013.00012

Poldrack, R. A., & Gorgolewski, K. J. (2017). OpenfMRI: Open sharing of task fMRI data. *Neuroimage, 144*(Pt B), 259-261. doi:10.1016/j.neuroimage.2015.05.073

Porro, C. A., Cettolo, V., Francescato, M. P., & Baraldi, P. (2003). Functional activity mapping of the mesial hemispheric wall during anticipation of pain. *Neuroimage, 19*(4), 1738-1747.

Pourshoghi, A., Zakeri, I., & Pourrezaei, K. (2016). Application of functional data analysis in classification and clustering of functional near-infrared spectroscopy signal in response to noxious stimuli. *J Biomed Opt, 21*(10), 101411. doi:10.1117/1.JBO.21.10.101411

Proudfoot, M., Bede, P., & Turner, M. R. (2018). Imaging Cerebral Activity in Amyotrophic Lateral Sclerosis. *Front Neurol, 9*, 1148. doi:10.3389/fneur.2018.01148

Pulini, A. A., Kerr, W. T., Loo, S. K., & Lenartowicz, A. (2019). Classification Accuracy of Neuroimaging Biomarkers in Attention-Deficit/Hyperactivity Disorder: Effects of Sample Size and Circular Analysis. *Biol Psychiatry Cogn Neurosci Neuroimaging, 4*(2), 108-120. doi:10.1016/j.bpsc.2018.06.003

Qiao, J., Wang, Z., Zhao, G., Huo, Y., Herder, C. L., Sikora, C. O., & Peterson, B. S. (2017). Functional neural circuits that underlie developmental stuttering. *PLoS One, 12*(7), e0179255. doi:10.1371/journal.pone.0179255

Ramsay, J. O., Giles, H., & Spencer, G. (2009). *Functional Data Analysis with R and MATLAB %@ 0387981845*: Springer Publishing Company, Incorporated.

Razoki, B. (2018). Neurofeedback versus psychostimulants in the treatment of children and adolescents with attention-deficit/hyperactivity disorder: a systematic review. *Neuropsychiatr Dis Treat, 14*, 2905-2913. doi:10.2147/NDT.S178839

Reif, A., Schecklmann, M., Eirich, E., Jacob, C. P., Jarczok, T. A., Kittel-Schneider, S., . . . Ehlis, A. C. (2011). A functional promoter polymorphism of neuronal nitric oxide synthase moderates prefrontal functioning in schizophrenia. *Int J Neuropsychopharmacol, 14*(7), 887-897. doi:10.1017/S1461145710001677

Ren, X., Lu, J., Liu, X., Shen, C., Zhang, X., Ma, X., . . . Liu, P. (2017). Decreased prefrontal brain activation during verbal fluency task in patients with somatoform pain disorder: An exploratory multi-channel near-infrared spectroscopy study. *Prog Neuropsychopharmacol Biol Psychiatry, 78*, 153-160. doi:10.1016/j.pnpbp.2017.05.006

Ringler, R., Greiner, M., Kohlloeffel, L., Handwerker, H. O., & Forster, C. (2003). BOLD effects in different areas of the cerebral cortex during painful mechanical stimulation. *Pain, 105*(3), 445-453.

Robinson, M. E., O'Shea, A. M., Craggs, J. G., Price, D. D., Letzen, J. E., & Staud, R. (2015). Comparison of machine classification algorithms for fibromyalgia: neuroimages versus self-report. *J Pain, 16*(5), 472-477. doi:10.1016/j.jpain.2015.02.002

Rodrigues-Amorim, D., Rivera-Baltanas, T., Lopez, M., Spuch, C., Olivares, J. M., & Agis-Balboa, R. C. (2017). Schizophrenia: A review of potential biomarkers. *J Psychiatr Res, 93*, 37-49. doi:10.1016/j.jpsychires.2017.05.009

Rosas-Romero, R., Guevara, E., Peng, K., Nguyen, D. K., Lesage, F., Pouliot, P., & Lima-Saad, W. E. (2019). Prediction of epileptic seizures with convolutional neural networks and functional near-infrared spectroscopy signals. *Comput Biol Med, 111*, 103355. doi:10.1016/j.compbiomed.2019.103355

Ruan, Q., D'Onofrio, G., Sancarlo, D., Bao, Z., Greco, A., & Yu, Z. (2016). Potential neuroimaging biomarkers of pathologic brain changes in Mild Cognitive Impairment and Alzheimer's disease: a systematic review. *BMC Geriatr, 16*, 104. doi:10.1186/s12877-016-0281-7

Ruesch, A., Yang, J., Schmitt, S., Acharya, D., Smith, M. A., & Kainerstorfer, J. M. (2020). Estimating intracranial pressure using pulsatile cerebral blood flow measured with diffuse correlation spectroscopy. *Biomed Opt Express, 11*(3), 1462-1476. doi:10.1364/BOE.386612

Sabeti, M., Katebi, S., & Boostani, R. (2009). Entropy and complexity measures for EEG signal classification of schizophrenic and control participants. *Artif Intell Med, 47*(3), 263-274. doi:10.1016/j.artmed.2009.03.003

Saha, S., Chant, D., Welham, J., & McGrath, J. (2005). A systematic review of the prevalence of schizophrenia. *PLoS Med, 2*(5), e141. doi:10.1371/journal.pmed.0020141

Sakoglu, U., Pearlson, G. D., Kiehl, K. A., Wang, Y. M., Michael, A. M., & Calhoun, V. D. (2010). A method for evaluating dynamic functional network connectivity and task-modulation: application to schizophrenia. *MAGMA, 23*(5-6), 351-366. doi:10.1007/s10334-010-0197-8

Sato, M., Ishikawa, W., Suzuki, T., Matsumoto, T., Tsujii, T., & Sakatani, K. (2013). Bayesian STAI anxiety index predictions based on prefrontal cortex NIRS data for the resting state. *Adv Exp Med Biol, 765*, 251-256. doi:10.1007/978-1-4614-4989-8_35

Schnack, H. G., & Kahn, R. S. (2016). Detecting Neuroimaging Biomarkers for Psychiatric Disorders: Sample Size Matters. *Front Psychiatry, 7*, 50. doi:10.3389/fpsyt.2016.00050

Selb, J., Boas, D. A., Chan, S. T., Evans, K. C., Buckley, E. M., & Carp, S. A. (2014). Sensitivity of near-infrared spectroscopy and diffuse correlation spectroscopy to brain hemodynamics: simulations and experimental findings during hypercapnia. *Neurophotonics, 1*(1). doi:10.1117/1.NPh.1.1.015005

Shellhaas, R. A. (2015). Continuous long-term electroencephalography: the gold standard for neonatal seizure diagnosis. *Semin Fetal Neonatal Med, 20*(3), 149-153. doi:10.1016/j.siny.2015.01.005

Shim, M., Hwang, H. J., Kim, D. W., Lee, S. H., & Im, C. H. (2016). Machine-learning-based diagnosis of schizophrenia using combined sensor-level and source-level EEG features. *Schizophr Res, 176*(2-3), 314-319. doi:10.1016/j.schres.2016.05.007

Shoushtarian, M., Alizadehsani, R., Khosravi, A., Acevedo, N., McKay, C. M., Nahavandi, S., & Fallon, J. B. (2020). Objective measurement of tinnitus using functional near-infrared spectroscopy and machine learning. *PLoS One, 15*(11), e0241695. doi:10.1371/journal.pone.0241695

Sirpal, P., Kassab, A., Pouliot, P., Nguyen, D. K., & Lesage, F. (2019). fNIRS improves seizure detection in multimodal EEG-fNIRS recordings. *J Biomed Opt, 24*(5), 1-9. doi:10.1117/1.JBO.24.5.051408

Song, H., Chen, L., Gao, R., Bogdan, I. I. M., Yang, J., Wang, S., . . . Yu, X. (2017). Automatic schizophrenic discrimination on fNIRS by using complex brain network analysis and SVM. *BMC Med Inform Decis Mak, 17*(Suppl 3), 166. doi:10.1186/s12911-017-0559-5

St Louis, E. K., & Cascino, G. D. (2016). Diagnosis of Epilepsy and Related Episodic Disorders. *Continuum (Minneap Minn), 22*(1 Epilepsy), 15-37. doi:10.1212/CON.0000000000000284

Steinbrink, J., Villringer, A., Kempf, F., Haux, D., Boden, S., & Obrig, H. (2006). Illuminating the BOLD signal: combined fMRI-fNIRS studies. *Magn Reson Imaging, 24*(4), 495-505. doi:10.1016/j.mri.2005.12.034

Strangman, G., Culver, J. P., Thompson, J. H., & Boas, D. A. (2002). A quantitative comparison of simultaneous BOLD fMRI and NIRS recordings during functional brain activation. *Neuroimage, 17*(2), 719-731.

Stuart, S., Vitorio, R., Morris, R., Martini, D. N., Fino, P. C., & Mancini, M. (2018). Cortical activity during walking and balance tasks in older adults and in people with Parkinson's disease: A structured review. *Maturitas, 113*, 53-72. doi:10.1016/j.maturitas.2018.04.011

Sundermann, B., Burgmer, M., Pogatzki-Zahn, E., Gaubitz, M., Stuber, C., Wessolleck, E., . . . Pfleiderer, B. (2014). Diagnostic classification based on functional connectivity in chronic pain: model optimization in fibromyalgia and rheumatoid arthritis. *Acad Radiol, 21*(3), 369-377. doi:10.1016/j.acra.2013.12.003

Sutoko, S., Monden, Y., Tokuda, T., Ikeda, T., Nagashima, M., Kiguchi, M., . . . Dan, I. (2019). Distinct Methylphenidate-Evoked Response Measured Using Functional Near-Infrared Spectroscopy During Go/No-Go Task as a Supporting Differential Diagnostic Tool Between Attention-Deficit/Hyperactivity Disorder and Autism Spectrum Disorder Comorbid Children. *Front Hum Neurosci, 13*, 7. doi:10.3389/fnhum.2019.00007

Takizawa, R., Fukuda, M., Kawasaki, S., Kasai, K., Mimura, M., Pu, S., . . . Joint Project for Psychiatric Application of Near-Infrared Spectroscopy, G. (2014). Neuroimaging-aided differential diagnosis of the depressive state. *Neuroimage, 85 Pt 1*, 498-507. doi:10.1016/j.neuroimage.2013.05.126

Takizawa, R., Hashimoto, K., Tochigi, M., Kawakubo, Y., Marumo, K., Sasaki, T., . . . Kasai, K. (2009). Association between sigma-1 receptor gene polymorphism and prefrontal hemodynamic response induced by cognitive activation in schizophrenia. *Prog Neuropsychopharmacol Biol Psychiatry, 33*(3), 491-498. doi:10.1016/j.pnpbp.2009.01.014

Takizawa, R., Tochigi, M., Kawakubo, Y., Marumo, K., Sasaki, T., Fukuda, M., & Kasai, K. (2009). Association between catechol-O-methyltrasferase Val108/158Met genotype and prefrontal hemodynamic response in schizophrenia. *PLoS One, 4*(5), e5495. doi:10.1371/journal.pone.0005495

Tellis, G. M., Vitale, C., & Murgallis, T. (2015). Near Infrared Spectroscopy (NIRS): A Pilot Study to Measure Hemoglobin Concentration Changes in the Brains of Persons who Stutter and Typically Fluent Speakers. *Procedia - Social and Behavioral Sciences, 193*, 261-265. doi:https://doi.org/10.1016/j.sbspro.2015.03.269

Thijs, R. D., Surges, R., O'Brien, T. J., & Sander, J. W. (2019). Epilepsy in adults. *Lancet, 393*(10172), 689-701. doi:10.1016/S0140-6736(18)32596-0

Tibshirani, R. (1996). Regression Shrinkage and Selection via the Lasso. *Journal of the Royal Statistical Society. Series B (Methodological), 58*(1), 267-288.

Treede, R. D. (2018). The International Association for the Study of Pain definition of pain: as valid in 2018 as in 1979, but in need of regularly updated footnotes. *Pain Rep, 3*(2), e643. doi:10.1097/PR9.0000000000000643

Tsuzuki, D., & Dan, I. (2014). Spatial registration for functional near-infrared spectroscopy: from channel position on the scalp to cortical location in individual and group analyses. *Neuroimage, 85 Pt 1*, 92-103. doi:10.1016/j.neuroimage.2013.07.025

Tsuzuki, D., Jurcak, V., Singh, A. K., Okamoto, M., Watanabe, E., & Dan, I. (2007). Virtual spatial registration of stand-alone fNIRS data to MNI space. *Neuroimage, 34*(4), 1506-1518. doi:10.1016/j.neuroimage.2006.10.043

Turner, B. O., Paul, E. J., Miller, M. B., & Barbey, A. K. (2018). Small sample sizes reduce the replicability of task-based fMRI studies. *Commun Biol, 1*, 62. doi:10.1038/s42003-018-0073-z

Tysnes, O. B., & Storstein, A. (2017). Epidemiology of Parkinson's disease. *J Neural Transm (Vienna), 124*(8), 901-905. doi:10.1007/s00702-017-1686-y

Vapnik, V. N. (1995). The Nature of Statistical Learning. *Theory*.

Varghese, T., Sheelakumari, R., James, J. S., & Mathuranath, P. (2013). A review of neuroimaging biomarkers of Alzheimer's disease. *Neurol Asia, 18*(3), 239-248.

Varoquaux, G. (2018). Cross-validation failure: Small sample sizes lead to large error bars. *Neuroimage, 180*(Pt A), 68-77. doi:10.1016/j.neuroimage.2017.06.061

Walsh, B., Tian, F., Tourville, J. A., Yucel, M. A., Kuczek, T., & Bostian, A. J. (2017). Hemodynamics of speech production: An fNIRS investigation of children who stutter. *Sci Rep, 7*(1), 4034. doi:10.1038/s41598-017-04357-6

Wang, R., Hao, Y., Yu, Q., Chen, M., Humar, I., & Fortino, G. (2021). Depression Analysis and Recognition Based on Functional Near-Infrared Spectroscopy. *IEEE J Biomed Health Inform, 25*(12), 4289-4299. doi:10.1109/JBHI.2021.3076762

Weickert, C. S., Weickert, T. W., Pillai, A., & Buckley, P. F. (2013). Biomarkers in schizophrenia: a brief conceptual consideration. *Dis Markers, 35*(1), 3-9. doi:10.1155/2013/510402

Woo, C. W., & Wager, T. D. (2015). Neuroimaging-based biomarker discovery and validation. *Pain, 156*(8), 1379-1381. doi:10.1097/j.pain.0000000000000223

Xu, L., Geng, X., He, X., Li, J., & Yu, J. (2019). Prediction in Autism by Deep Learning Short-Time Spontaneous Hemodynamic Fluctuations. *Front Neurosci, 13*, 1120. doi:10.3389/fnins.2019.01120

Xu, L., Hua, Q., Yu, J., & Li, J. (2020). Classification of autism spectrum disorder based on sample entropy of spontaneous functional near infra-red spectroscopy signal. *Clin Neurophysiol, 131*(6), 1365-1374. doi:10.1016/j.clinph.2019.12.400

Xu, L., Liu, Y., Yu, J., Li, X., Yu, X., Cheng, H., & Li, J. (2020). Characterizing autism spectrum disorder by deep learning spontaneous brain activity from functional near-infrared spectroscopy. *J Neurosci Methods, 331*, 108538. doi:10.1016/j.jneumeth.2019.108538

Xu, L., Sun, Z., Xie, J., Yu, J., Li, J., & Wang, J. (2021). Identification of autism spectrum disorder based on short-term spontaneous hemodynamic fluctuations using deep learning in a multi-layer neural network. *Clin Neurophysiol, 132*(2), 457-468. doi:10.1016/j.clinph.2020.11.037

Yang, D., Hong, K. S., Yoo, S. H., & Kim, C. S. (2019). Evaluation of Neural Degeneration Biomarkers in the Prefrontal Cortex for Early Identification of Patients With Mild Cognitive Impairment: An fNIRS Study. *Front Hum Neurosci, 13*, 317. doi:10.3389/fnhum.2019.00317

Yang, D., Huang, R., Yoo, S.-H., Shin, M.-J., Yoon, J. A., Shin, Y.-I., & Hong, K.-S. (2020). Detection of Mild Cognitive Impairment Using Convolutional Neural Network: Temporal-Feature Maps of Functional Near-Infrared Spectroscopy. *Frontiers in Aging Neuroscience, 12*(141). doi:10.3389/fnagi.2020.00141

Yang, J., Ji, X., Quan, W., Liu, Y., Wei, B., & Wu, T. (2020). Classification of Schizophrenia by Functional Connectivity Strength Using Functional Near Infrared Spectroscopy. *Front Neuroinform, 14*, 40. doi:10.3389/fninf.2020.00040

Yasumura, A., Omori, M., Fukuda, A., Takahashi, J., Yasumura, Y., Nakagawa, E., . . . Inagaki, M. (2017). Applied Machine Learning Method to Predict Children With ADHD Using Prefrontal Cortex Activity: A Multicenter Study in Japan. *J Atten Disord*, 1087054717740632. doi:10.1177/1087054717740632

Ye, J. C., Tak, S., Jang, K. E., Jung, J., & Jang, J. (2009). NIRS-SPM: statistical parametric mapping for near-infrared spectroscopy. *Neuroimage, 44*(2), 428-447. doi:10.1016/j.neuroimage.2008.08.036

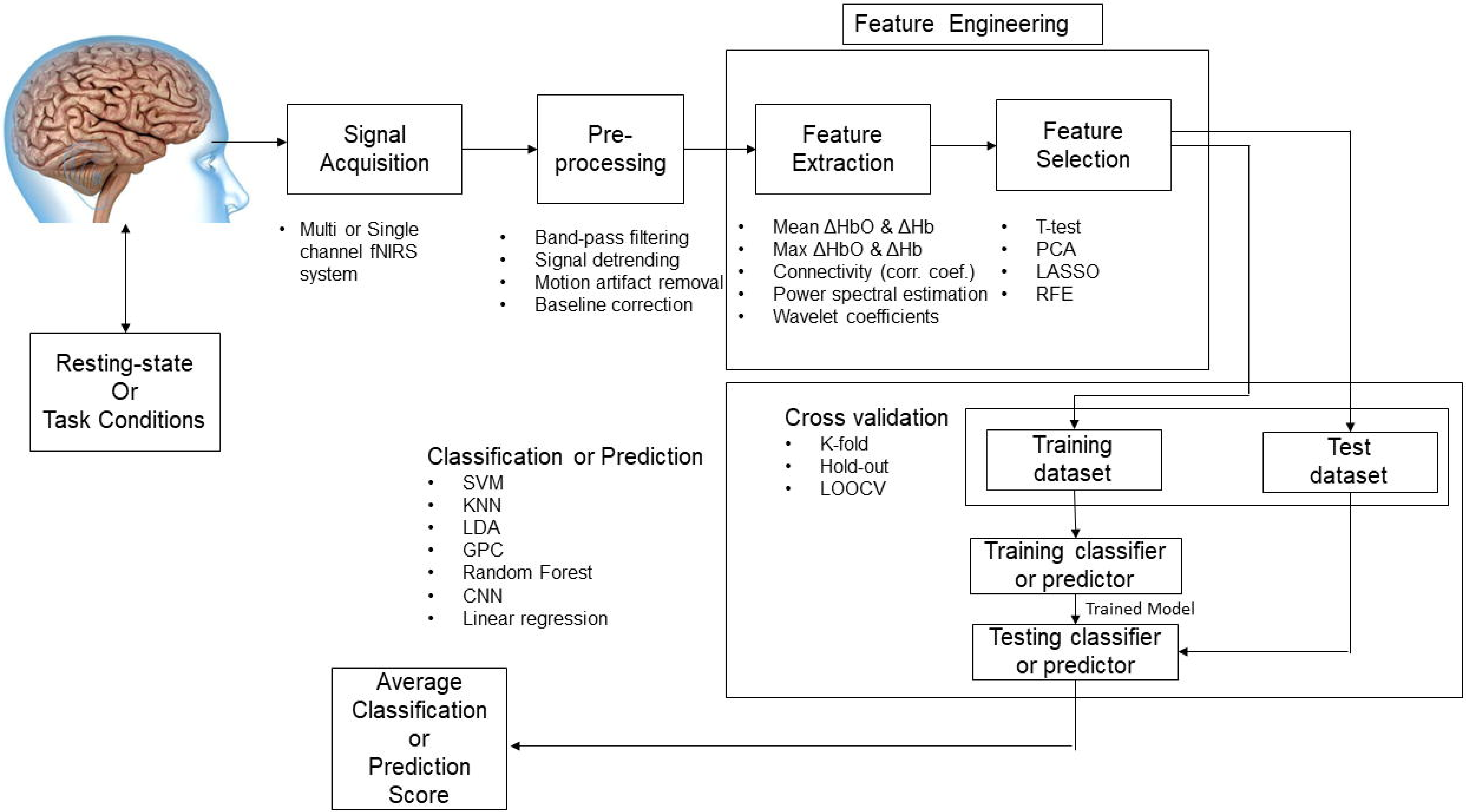
Yeung, M. K., Sze, S. L., Woo, J., Kwok, T., Shum, D. H., Yu, R., & Chan, A. S. (2016). Altered Frontal Lateralization Underlies the Category Fluency Deficits in Older Adults with Mild Cognitive Impairment: A Near-Infrared Spectroscopy Study. *Front Aging Neurosci, 8*, 59. doi:10.3389/fnagi.2016.00059

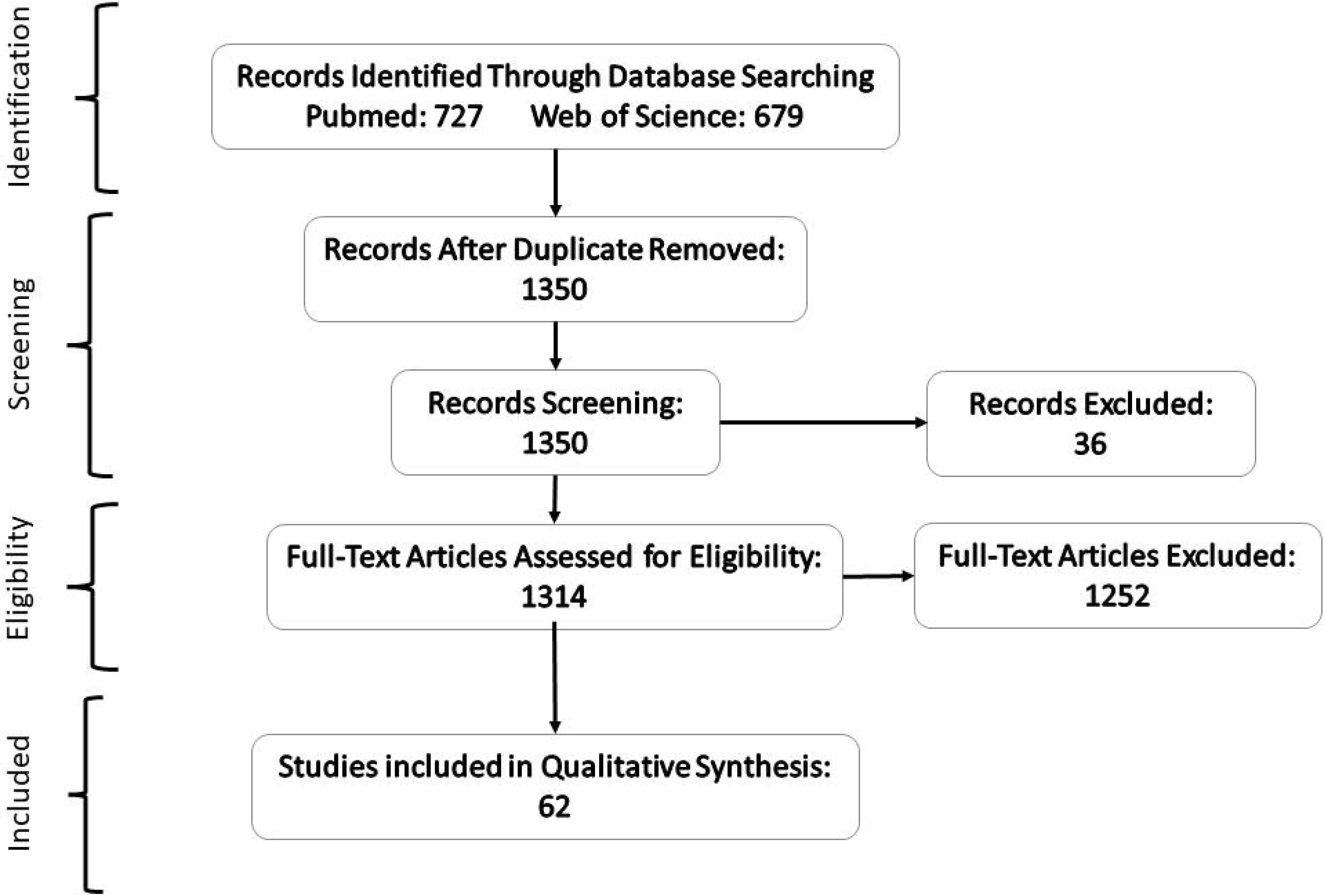
Yucel, M. A., Aasted, C. M., Petkov, M. P., Borsook, D., Boas, D. A., & Becerra, L. (2015). Specificity of hemodynamic brain responses to painful stimuli: a functional near-infrared spectroscopy study. *Sci Rep, 5*, 9469. doi:10.1038/srep09469

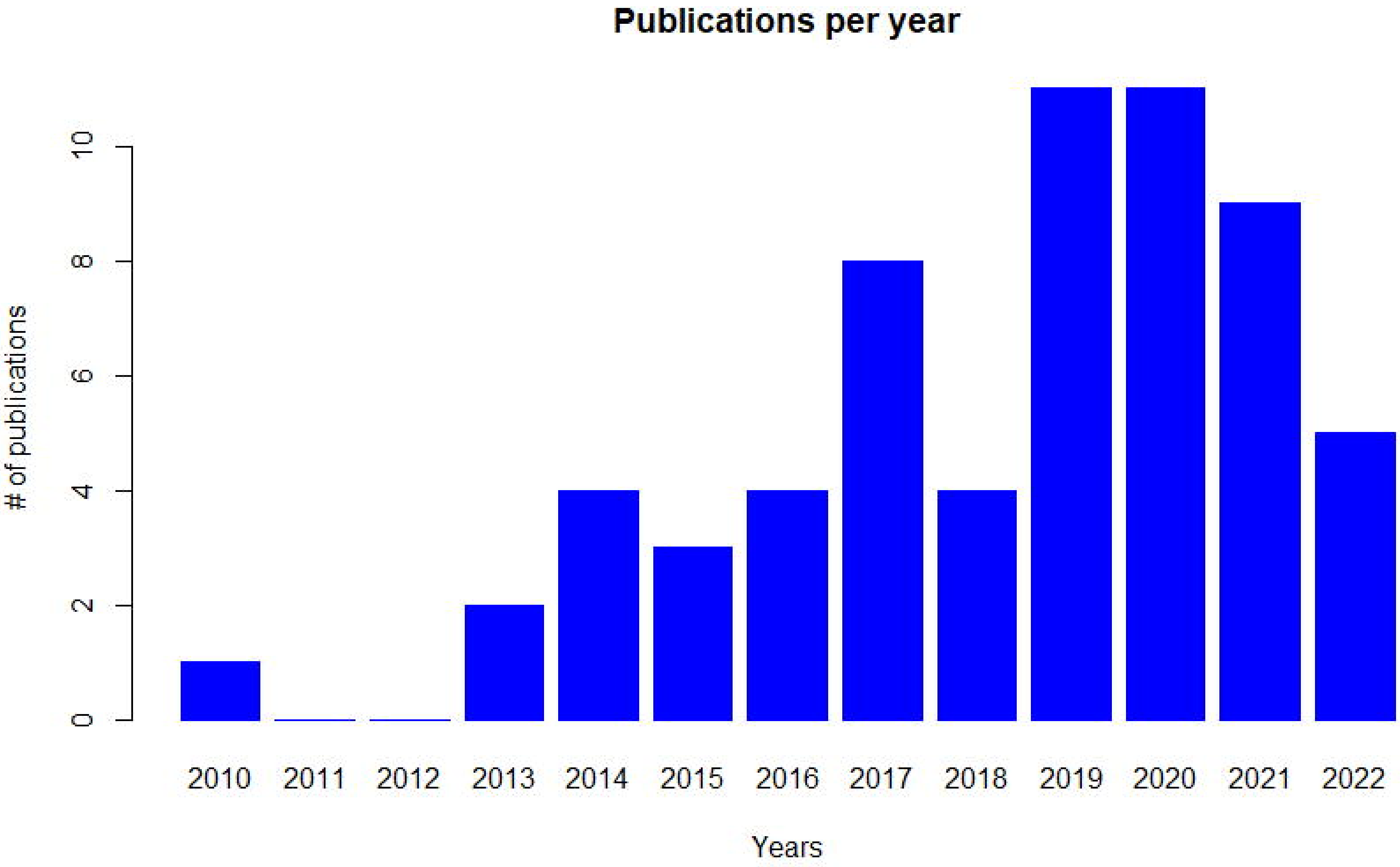
Zhang, Y. J., Lu, C. M., Biswal, B. B., Zang, Y. F., Peng, D. L., & Zhu, C. Z. (2010). Detecting resting-state functional connectivity in the language system using functional near-infrared spectroscopy. *J Biomed Opt, 15*(4), 047003. doi:10.1117/1.3462973

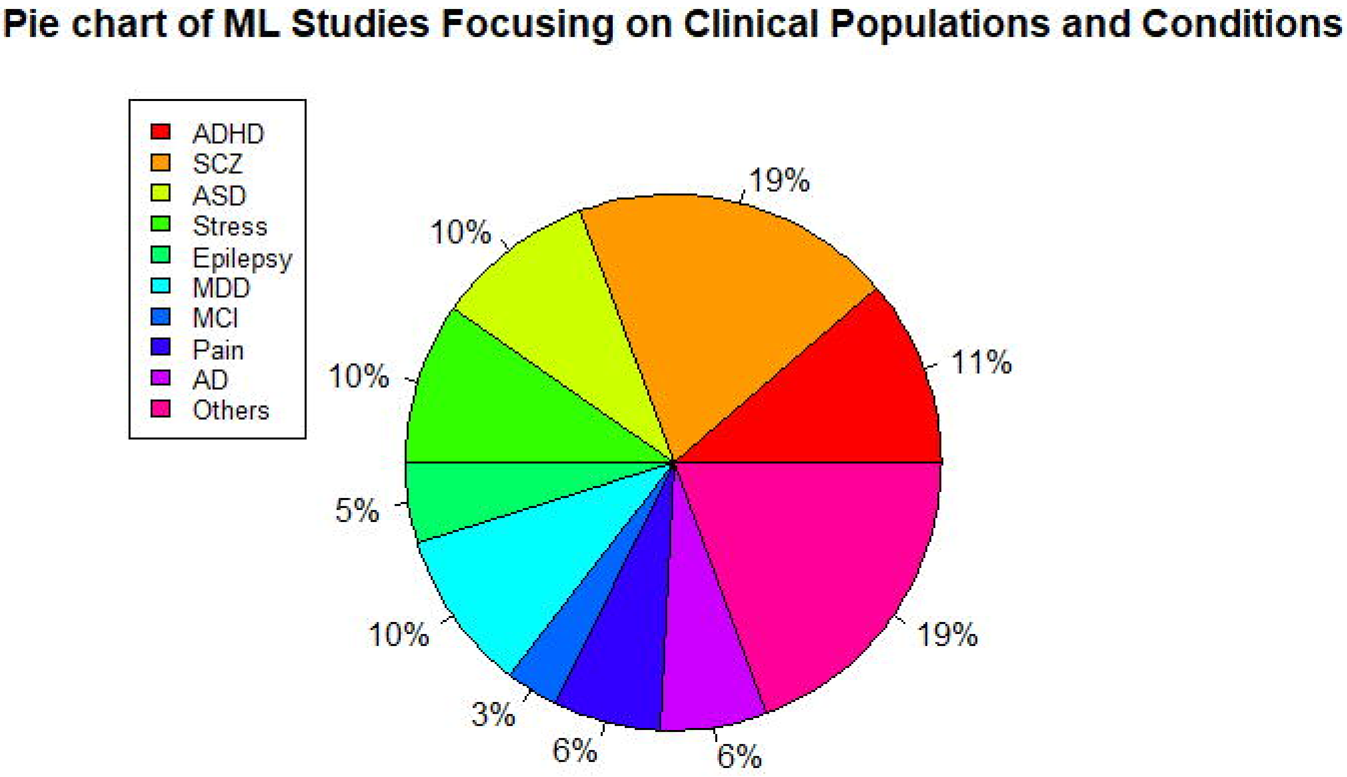
Zhu, Y., Jayagopal, J. K., Mehta, R. K., Erraguntla, M., Nuamah, J., McDonald, A. D., . . . Chang, S. (2020). Classifying Major Depressive Disorder using fNIRS during Motor Rehabilitation. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 1-1. doi:10.1109/TNSRE.2020.2972270

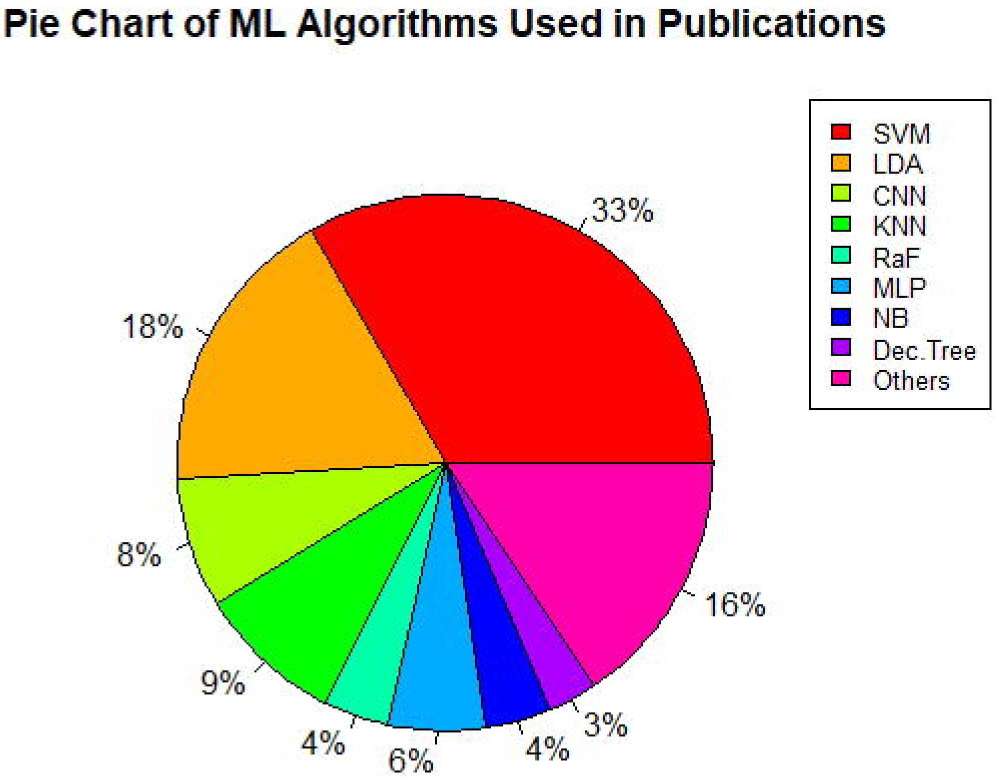
Zimeo Morais, G. A., Balardin, J. B., & Sato, J. R. (2018). fNIRS Optodes' Location Decider (fOLD): a toolbox for probe arrangement guided by brain regions-of-interest. *Sci Rep, 8*(1), 3341. doi:10.1038/s41598-018-21716-z

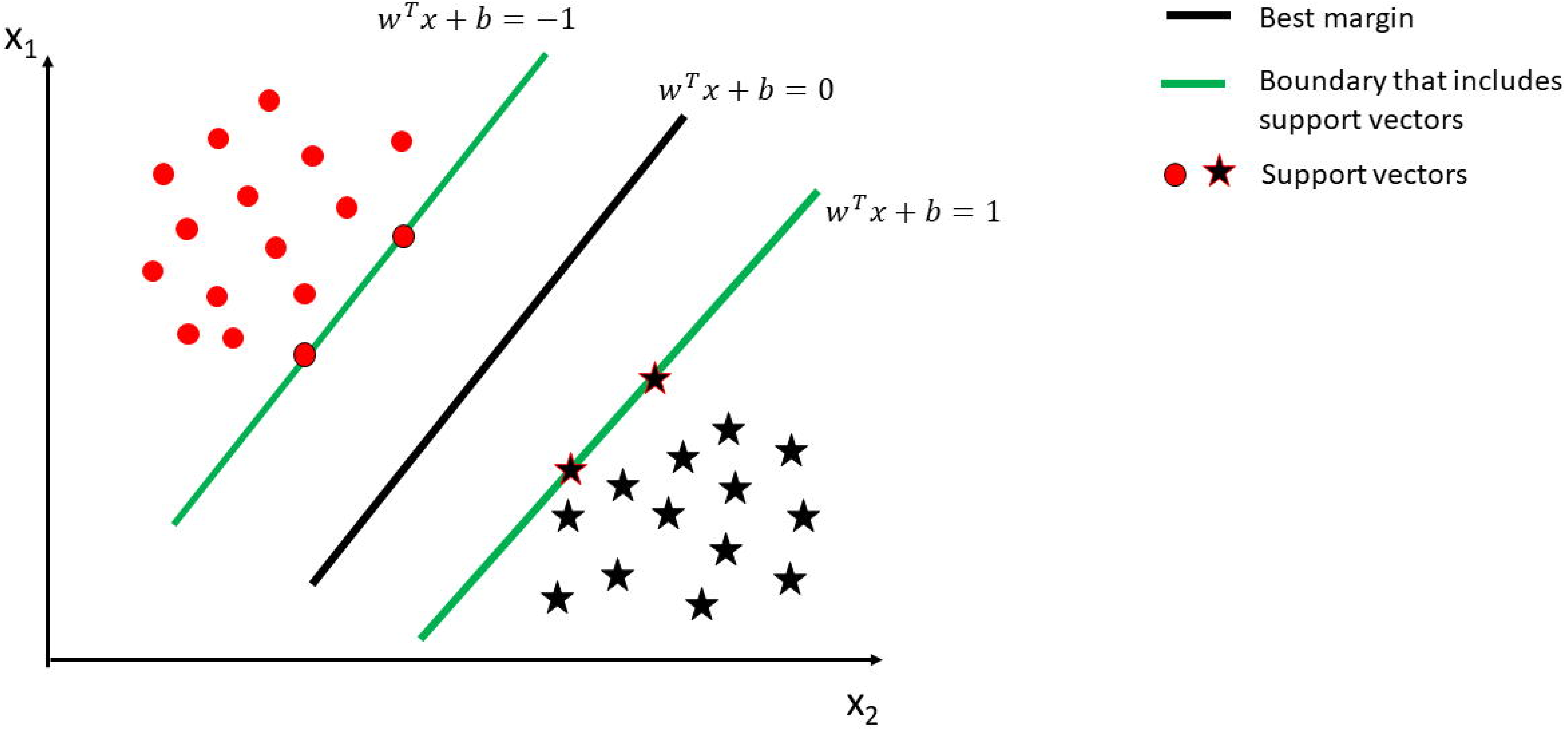

Feature Engineering
Signal Acquisition
Pre-processing
Feature Extraction
Feature Selection
Multi or Single channel fNIRS system
Band-pass filtering
Signal detrending
Motion artifact removal
Baseline correction
Mean ΔHbO & ΔHb
Max ΔHbO & ΔHb
Connectivity (corr. coef.)
Power spectral estimation
Wavelet coefficients
T-test
PCA
LASSO
RFE
Resting-state Or Task Conditions
Cross validation
K-fold
Hold-out
LOOCV
Training dataset
Test dataset
Classification or Prediction
SVM
KNN
LDA
GPC
Random Forest
CNN
Linear regression
Training classifier or predictor
Trained Model
Testing classifier or predictor
Average Classification or Prediction Score


Identification
Screening
Eligibility
Included
Records Identified Through Database Searching
Pubmed: 727 Web of Science: 679
Records After Duplicate Removed:
1350
Records Screening:
1350
Records Excluded:
36
Full-Text Articles Assessed for Eligibility:
1314
Full-Text Articles Excluded:
1252
Studies included in Qualitative Synthesis:
62


Publications per year
# of publications
0
2
4
6
8
10
2010
2011
2012
2013
2014
2015
2016
2017
2018
2019
2020
2021
2022
Years


Pie chart of ML Studies Focusing on Clinical Populations and Conditions
ADHD
SCZ
ASD
Stress
Epilepsy
MDD
MCI
Pain
AD
Others
19%
10%
10%
11%
5%
10%
19%
3%
6%
6%


Pie Chart of ML Algorithms Used in Publications
33%
18%
8%
9%
4%
6%
4%
3%
16%
SVM
LDA
CNN
KNN
RaF
MLP
NB
Dec.Tree
Others


$w^T x + b = -1$
$w^T x + b = 0$
$w^T x + b = 1$
x_1
x_2
Best margin
Boundary that includes support vectors
Support vectors