



# Identifying neuroimaging biomarkers of major depressive disorder from cortical hemodynamic responses using machine learning approaches

Zhifei Li,<sup>a,b</sup> Roger S. McIntyre,<sup>c,d</sup> Syeda F. Husain,<sup>e</sup> Roger Ho,<sup>a,f,\*</sup> Bach X. Tran,<sup>g,h</sup> Hien Thu Nguyen,<sup>ij</sup> Shuenn-Chiang Soo,<sup>k</sup> Cyrus S. Ho,<sup>f,k</sup> and Nanguang Chen<sup>a,b,l</sup>

<sup>a</sup>Institute of Health Innovation and Technology (iHealthtech), National University of Singapore, Singapore

<sup>b</sup>Department of Biomedical Engineering, Faculty of Engineering, National University of Singapore, Singapore

<sup>c</sup>Mood Disorders Psychopharmacology Unit, University Health Network, University of Toronto, Toronto, ON, Canada

<sup>d</sup>Canadian Rapid Treatment Center of Excellence, Mississauga, ON, Canada

<sup>e</sup>Department of Paediatrics, Yong Loo Lin School of Medicine, NUS, Singapore

<sup>f</sup>Department of Psychological Medicine, Yong Loo Lin School of Medicine, National University of Singapore, Singapore

<sup>g</sup>Institute for Preventive Medicine and Public Health, Hanoi Medical University, Viet Nam

<sup>h</sup>Bloomberg School of Public Health, Johns Hopkins University, USA

<sup>i</sup>Institute for Global Health Innovations, Duy Tan University, Viet Nam

<sup>j</sup>Faculty of Medicine, Duy Tan University, Da Nang, Viet Nam

<sup>k</sup>Department of Psychological Medicine, National University Hospital, Singapore

<sup>l</sup>NUS (Suzhou) Research Institute, Suzhou, China

## Summary

**Background** Early diagnosis of major depressive disorder (MDD) could enable timely interventions and effective management which subsequently improve clinical outcomes. However, quantitative and objective assessment tools for the suspected cases who present with depressive symptoms have not been fully established.

**Methods** Based on a large-scale dataset ( $n = 363$  subjects) collected with functional near-infrared spectroscopy (fNIRS) measurements during the verbal fluency task (VFT), this study proposed a data representation method for extracting spatiotemporal characteristics of NIRS signals, which emerged as candidate predictors in a two-phase machine learning framework to detect distinctive biomarkers for MDD. Supervised classifiers (e.g., support vector machine (SVM), k-nearest neighbors (KNN)) cooperated with cross-validation were implemented to evaluate the predictive capability of selected features in a training set. Another test set that was not involved in developing the algorithms enabled the independent assessment of the model's generalization.

**Findings** For the classification with the optimal fusion features, the SVM classifier achieved the highest accuracy of  $75.6\% \pm 4.7\%$  in the nested cross-validation, and the correct prediction rate of  $78.0\%$  with a sensitivity of  $75.0\%$  and a specificity of  $81.4\%$  in the test set. Moreover, the multiway ANOVA test on clinical and demographic factors confirmed that twenty out of 39 optimal features were significantly correlated with the MDD-distinctive consequence.

**Interpretation** The abnormal prefrontal activity of MDD may be quantified as diminished relative intensity and inappropriate activation timing of hemodynamic response, resulting in an objectively measurable biomarker for assessing cognitive deficits and screening MDD at the early stage.

**Funding** This study was funded by NUS iHealthtech Other Operating Expenses (R-722-000-004-731).

**Copyright** © 2022 The Author(s). Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>)

**Keywords:** Functional near-infrared spectroscopy; Depressive disorder; Feature selection; Supervised learning; Biomarkers discovery; Depression

\*Corresponding author at: Department of Psychological Medicine, Yong Loo Lin School of Medicine L9 NUHS Tower Block, 1E Kent Ridge Road, National University of Singapore.

E-mail address: [pcmrhcm@nus.edu.sg](mailto:pcmrhcm@nus.edu.sg) (R. Ho).

## Introduction

Major depressive disorder (MDD) is a chronic mood disorder and negatively affects workplace performance, physical health, and quality of life.<sup>1</sup> Approximately 3.8%

eBioMedicine 2022;79:  
104027

Published online xxx

[https://doi.org/10.1016/j.](https://doi.org/10.1016/j.ebiom.2022.104027)

[ebiom.2022.104027](https://doi.org/10.1016/j.ebiom.2022.104027)

### Research in context

#### *Evidence before the study*

Major depression is a chronic illness of considerable morbidity, with high rates of relapse and recurrence. As a non-invasive, continuous and economical technique for measurement of brain hemodynamic activity, functional near infrared spectroscopy (fNIRS) is being increasingly applied for investigation of the association between the social cognition of people with depression and their prefrontal cortex activation. A search in PubMed with the terms, ((MDD) OR (depression) OR (depressive) OR (depressed)) AND ((fNIRS) OR (NIRS) OR (functional near infrared spectroscopy)) AND ((machine learning) OR (classification) OR (pattern recognition) OR (discriminant analysis)) AND ((biomarker) OR (marker)), was conducted to investigate the previous studies using machine learning for recognizing depression-distinctive pattern in fNIRS data. Among these reports on prior to April 9th, 2022, four relevant studies were identified, but the sample sizes are relatively small to provide a definitive conclusion. Few evidence to support a quantitative neuroimaging biomarker and pathophysiology elucidation for depression, particularly involving the influence from confounding variables.

#### *Added value of the study*

Based on the topographic fNIRS technology that allows sensitive and real-time detection of cerebral hemoglobin changes, this study was performed on a time-series signals dataset with the standard measurement procedure regarding probe setting and task paradigm. A two-phase machine learning framework was developed to identify the depression-distinctive biomarkers, which presented distinguishing features of hemodynamic intensity and activation timing in depressed patients. A large-scale dataset collected with the same experimental paradigm not only enables the development of advanced machine learning approaches to analyze spatiotemporal characteristics from fNIRS signals at single-channel level, but also makes a more comprehensive validation on the model performance due to the independent test set and sample diversity.

#### *Implications of all the available evidence*

The high accuracy discrimination results demonstrated the multi-dimensional profile of cognitive deficits captured in the fNIRS patterns could objectively confirm and distinguish the underlying neuropathological symptoms of depression. Furthermore, two validation approaches were applied to explore the prediction performances and hemodynamic responses affected by various demographic factors. This analysis could draw attention to the heterogeneity observed amongst depressed cases and contribute to developing confounder-free biomarkers for aiding differential diagnosis.

of the world's general population was estimated to suffer from depression, which was ranked by the World Health Organization (WHO) as the largest contributor to global disability.<sup>2,3</sup> Individuals with MDD are at increased risk for developing several medical disorders, e.g., hypertension, dementia, and cerebrovascular disease.<sup>4</sup> Moreover, severe cases of depression carry a high risk of suicide if it is not caught early.

Accuracy diagnosis and early intervention in the initial stages of MDD is essential for prompt treatment to reduce severe morbidity and mortality. However, the diagnosis of MDD mainly relies on clinical interviews and subjective evaluation of depressive symptoms, which are not applicable for large scale population assessment or continuous monitoring of the disease progression due to medical burden and shortage of mental health professionals.<sup>5</sup> There seems to be an increasing view from the field of neuroscience that the pathophysiology of psychiatric disorders may be reflected by abnormal activities in the cerebral cortex.<sup>6,7</sup> To measure the physiological markers in people suffering from MDD, functional near-infrared spectroscopy (fNIRS) offers the most suitable modality by detecting functional changes in cerebral activity.<sup>8</sup> The topographic fNIRS is able to continuously measure brain hemodynamic parameters in the form of oxygen-, deoxygen- and total hemoglobin concentration changes (i.e.,  $\Delta\text{HbO}$ ,  $\Delta\text{HbR}$  &  $\Delta\text{HbT}$ ), which are considered as sensitive indicators for the investigation of psychiatric disorders.<sup>9</sup>

Previous research found that fNIRS showed abnormal functions of the prefrontal cortex (PFC) in people with MDD during the verbal fluency test (VFT).<sup>10</sup> The decreased brain activity correlated with symptom severity is statistically recognized as low integral values of fNIRS waveform patterns in the frontal and temporal regions for MDD.<sup>11,12</sup> Discriminant analysis with statistical approaches was able to reveal significant discrepancies in the probability distributions of signal patterns between people with MDD (MDDs) and healthy controls (HCs), however, it remains insufficient to use fNIRS to establish the diagnosis of MDD.<sup>13</sup> In recent years, machine learning (ML) has been explored from classification to treatment outcomes prediction for psychiatric disorders due to its ability in automatically learning from empirical data to recognize complex patterns.<sup>14,15</sup> Despite early concept studies did not include external validation due to the lack of an independent dataset,<sup>16,17</sup> the preliminary results still indicated that machine learning is a promising analysis tool for discovering biomarkers and elucidating pathophysiology of MDD. In our previous work,<sup>18,19</sup> the frontotemporal hemodynamic signals collected from 210 subjects were investigated based on the group-level and channel-cluster-level statistics test, and results revealed the significant difference in the intensity of prefrontal

activity during VFT test between MDD and Healthy Controls (HCs). With 164 new subjects extending to the sample set, this study aims to use ML models to implement effective feature selection and to explore MDD-distinctive patterns from a large-scale dataset, the impact of confounding factors would be evaluated on the model robust for application of NIRS measure in establishing the diagnosis.

## Methods

### Participants and ethics

In this study, people with stable newly diagnosed or existing major depressive disorder from the outpatient psychiatric clinic at the National University Hospital, Singapore were recruited. Age and gender-matched healthy controls were also recruited from the community. People are regarded as suitable to participate in the experiment if they fulfill the following criteria: (i) age between 21 years and 65, (ii) right-handed, (iii) English-speaking and (iv) participants who have the capacity to consent to the research. The exclusion criteria are: (i) mental retardation, (ii) comorbid other psychiatric disorders such as schizophrenia, anxiety disorders, personality disorders, substance abuse, (iii) history of seizure, stroke or head injury, (iv) currently psychotic or suicidal, (v) visual, auditory and speech impairment and (vi) unstable medical illnesses.

By examination of valid fNIRS channels, we selected an effective sample set consisting of 177 people with MDD [male/female: 80/97; mean age: 39.0; standard deviation (SD): 13.9]; and 186 healthy subjects (male/female: 75/111; mean age: 36.4; SD: 14.2) from the community. Psychiatric diagnosis was made by a psychiatrist with professional qualifications, based on Structured Clinical Interview for DSM Disorders (SCID) for both patients and controls.<sup>20</sup> The cutoff value for scores was determined by: Five (or more) of the symptoms including depressed mood, loss of interest, poor sleep, weight loss, fatigue, psychomotor retardation, poor concentration, suicidal thought, social impairment) have been present during the same 2-week period and represent a change from previous functioning; at least one of the symptoms is either (1) depressed mood or (2) loss of interest or pleasure. In addition, the evaluation for severity of depressive symptoms and psychosocial functioning was based on the day of participation using the 17-item Hamilton rating scale for depression (HAM-D) and global assessment of functioning (GAF).<sup>21,22</sup> All participants gave written informed consent and were recruited from May 2017 to June 2020. The study was in accordance with the Declaration of Helsinki, and the ethical principles in the Belmont Report. It was approved by the Domain Specific Review Board of the National Healthcare Group, Singapore (protocol number 2017/00509).

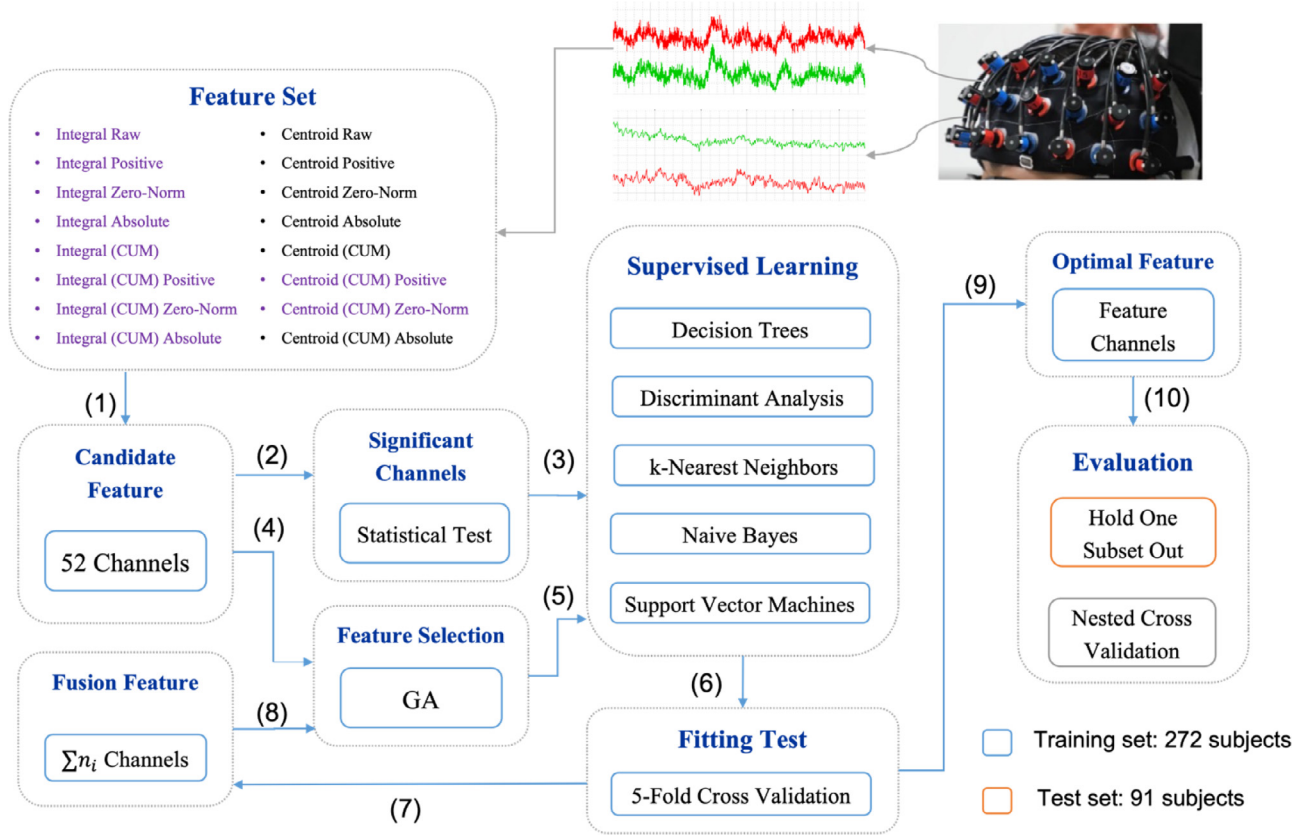
### Data preparation

Hemoglobin changes in the frontal-temporal regions were quantified by the NIRS signals recorded during the verbal fluency task (VFT), which is considered a neuropsychological test to elicit functional abnormalities relevant to major psychiatric disorders.<sup>23</sup> The NIRS apparatus and measurement procedure were fully described in our earlier research.<sup>18,19</sup> According to the modified Beer-Lambert law,<sup>24</sup> the concentration changes in oxygen-hemoglobin ( $\Delta\text{HbO}$ ) and deoxyhemoglobin ( $\Delta\text{HbR}$ ) could be derived from NIRS signals measured during the task period. Further analyses focused on the spatiotemporal characteristics of  $\Delta\text{HbO}$ , which were considered to more directly reflect task-related cortical activation than other signals, as evidenced by the strong correlation to the blood oxygenation level-dependent signal measured by fMRI and by the findings of animal studies.<sup>25,26</sup> The time course of  $\Delta\text{HbO}$  was normalized by linear fitting between 10 s baseline at the end of the pre-task period, and 5 s in the post-task period. A moving average with a window width of 50 sampling points (5s) was applied to remove high frequency noise from  $\Delta\text{HbO}$  signals. Noise channels with saturation or low intensity were filtered out, and the subject data with at least 30 available channels were included for further analysis.<sup>27</sup> Pre-processing of NIRS signals was conducted in Brainstorm and Nirstorm plugin.<sup>28,29</sup>

### Machine learning framework

In order to develop a quantitative tool for individual diagnostic in practical utility, we performed extensive research efforts on specialized machine learning (ML) approaches to identify neuroimaging biomarkers of MDD. Figure 1 showed an outline of the proposed ML framework, which involved a sequence process of the fNIRS feature extraction, selection, classification, and validation. In the feature extraction, the time-series  $\Delta\text{HbO}$  and  $C_{\text{HbO}}$  across the 60s task period of word production during the VFT was chosen at each channel and each subject. Subsequently, a total of 16 variables were generated, e.g., integral raw, centroid (CUM) positive, etc., to represent the spatiotemporal characteristics of the hemodynamic response (see *Supplementary: Candidate Feature Extraction*).

By adopting the statistical test or genetic algorithm (GA)<sup>30</sup> as the feature selection method, our ML framework identifies the most informative decision variables from the extracted features while reducing the space of possible solutions. Then we established five supervised classification algorithms (i.e., k-nearest neighbors (KNN), support vector machine (SVM), discriminant analysis (DA), decision tree (TREE), and Naïve Bayes (NB)) for modeling the correlation between the selected features and the corresponding diagnostic outcomes. The explanation of these models and the specific



**Figure 1.** The process of ML framework, including feature extraction, statistics-based or GA-based feature selection, supervised learning models, and validation. Statistical criteria-based feature ranking steps: (1) → (2) → (3) → (6) → (9) → (10). GA-based feature searching steps: (1) → (4) → (5) → (6) → (7) → (8) → (5) → (6) → (9) → (10).

parameters assigned to each classifier could be found in *Supplementary: Feature Selection and Tables 1, 2*. In the process of searching for the optimal features, the loss function of each model was estimated by five-fold cross-validation to effectively avoid over-fitting. Moreover, an independent test and the nested cross-validation (Nested CV) were implemented to validate the generalizability of the classifiers as well as the potential biomarkers.

All fNIRS data, including 363 subjects, were randomly shuffled; then 80% of them, a total of 272 subjects (129 MDDs and 143 HCs) were treated as the training set, and the remaining 20%, 91 subjects (48 MDDs and 43 HCs) as the test set. As shown in [Figure 1](#), in order to objectively evaluate the classification performance, the analysis methods were only performed in the training set (blue box) while the final selected features were validated in the test set (orange box). As for the five-fold Nested CV (grey box) for calculating the mean and standard deviations of prediction metrics, the training (141 MDDs and 149 HCs) and test (36 MDDs and 37 HCs) data at comparable sample distribution were randomly selected for each split to validate the replicability of results across different sub-datasets.

### Evaluate the effect of demographic heterogeneity on classification models

The effect of demographic factors on cortical activation patterns of MDD was studied in terms of the predictive robustness of classifier and statistics test. Firstly, MDDs and HCs were divided respectively into different groups by a specific factor, as shown in *Supplementary Table 5*. We sought to confirm the evaluation reliability which targeted at the groups including more than 10 samples, thus the evaluation metric applied in potential confounders was set as: (1) Factors could be involved in the analysis of accuracy, including gender, age, years of education, smoking, alcohol misuse, and past medical history; (2) Only sensitivity was assessed for exclusive factors in people with MDD, including duration of MDD, antidepressant use, and family psychiatric history.

Estimation of the accuracy or sensitivity of a classifier would be repeated ten times using leave-one-out cross-validation and reporting the mean and standard deviation results across different grouping samples. The KNN algorithm can highly be overfitting to the training set and affect the evaluation conclusion, thus SVM and DA classifiers with preferable performances were chosen to assess the factors. Furthermore, to investigate the association between the characteristics of hemodynamic response and clinical/demographic factors, the multiway analysis of variance (ANOVA) was applied for quantifying the significance level of identified biomarkers due to the influencing factors.

### Role of funding source

No funding source or sponsor has a role in study design; data collection, analyses or interpretation; preparation, review or approval of the report. The corresponding author had full access to all of the data and the final responsibility to submit for publication.

## Results

### Classification results of two-phase feature selection model

In the two-phase feature selection, the genetic algorithm (GA) approach was applied to find the optimal feature channels from the training dataset and adopted a specific classification model to estimate the fitness of selected features. After 300 iterations, the searching algorithm can be substantially converged at an informative and reduced feature set. Two examples of convergence with SVM and DA are respectively shown in *Supplementary Figure 3*, where both the loss of five-fold cross-validation and the prediction error rate gradually decreased to stable minimums. The metric of test accuracy was not involved in the GA searching process so that the effectiveness of method could be fairly validated on the final features. Furthermore, the feature dimension was reduced from 520 to 39 while employing SVM as classifier and up to 60 with KNN, which also demonstrated better dimension reduction ability and efficiency in MDD-related biomarkers discovery.

The performances of GA feature selection method using different classifiers for the identification of MDD cases were demonstrated in [Table 1](#). Both the five-fold cross-validation and independent test accuracy revealed that the model SVM cooperated with 39 features pattern could differentiate people with MDD from controls with the best performance. For the classification of training set, the accuracy of correctly classified cases and non-cases by SVM was 82.4% with a sensitivity of 85.3% (true positive = 110 of the 129 MDDs) and a specificity of 79.7% (true negative = 114 of the 143 HCs). As for the test set, the accuracy of correctly classified was 78.0% with a sensitivity of 75.0% (true positive = 36 of the 48 MDDs) and a specificity of 81.4% (true negative = 35 of the 43 HCs). In the outer loop of nested CV, the averaged accuracy with standard deviation of  $0.756 \pm 0.047$  was obtained over five splits of dataset.

The 39 feature channels selected by the GA and SVM classifier could be recognized as neurophysiological biomarkers that accurately identify the hemodynamic pattern of people with MDD. The group-level comparisons of discriminative features between the MDDs and HCs were plotted on a common scale in [Figure 2](#), wherein the fusion features belong to nine feature variables. [Figure 2](#) (b) and (c) illustrate the  $\Delta\text{HbO}$  and  $C_{\text{HbO}}$  mapping of group-level statistics on averaging of HCs and MDDs. Six out of the seven integral variants of time-

Classifier	Feature Number		KNN 60	SVM 39	DA 32	TREE 8	NB 21
GA Conv.	Train Set (3/4)	5-Fold CV	81.6%	76.9%	75.4%	67.3%	75.0%
		Accuracy	100.0%	82.4%	82.0%	82.7%	79.4%
		Sensitivity	100.0%	85.3%	85.3%	75.2%	86.1%
		Specificity	100.0%	79.7%	79.0%	89.5%	73.4%
	Test Set (1/4)	Accuracy	78.0%	78.0%	75.8%	65.9%	72.5%
		Sensitivity	79.2%	75.0%	75.0%	60.4%	68.8%
		Specificity	76.7%	81.4%	76.7%	72.1%	76.7%
Nested CV	Train Set (4/5)	5-Fold CV	0.77 ± 0.02	0.72 ± 0.02	0.72 ± 0.02	0.67 ± 0.01	0.71 ± 0.02
		Accuracy	1.0 ± 0.0	0.79 ± 0.01	0.78 ± 0.01	0.80 ± 0.00	0.78 ± 0.01
		Sensitivity	1.0 ± 0.0	0.80 ± 0.01	0.80 ± 0.01	0.71 ± 0.03	0.84 ± 0.01
		Specificity	1.0 ± 0.0	0.79 ± 0.02	0.76 ± 0.01	0.89 ± 0.03	0.73 ± 0.02
	Test Set (1/5)	Accuracy	0.72 ± 0.04	0.76 ± 0.05	0.73 ± 0.05	0.67 ± 0.06	0.72 ± 0.04
		Sensitivity	0.74 ± 0.03	0.77 ± 0.06	0.76 ± 0.08	0.59 ± 0.12	0.78 ± 0.07
		Specificity	0.69 ± 0.06	0.74 ± 0.06	0.71 ± 0.04	0.75 ± 0.06	0.65 ± 0.09

Table 1: Classification results with fusion features and different classifiers.

series data measured from MDDs were significantly lower than HCs during the VFT. Figure 2 (a) shows the C<sub>HbO</sub> mapping of group-level statistics on averaging of two groups, two of centroid variants, i.e., ‘Centroid (CUM) Positive’ and ‘Centroid (CUM) Zero-Norm’, for MDDs were observed slightly smaller mean value (earlier activation timing) and larger standard deviation (scattered in wider range) than HCs. The statistical correlations on the selected variables can be found in Supplementary Table 4.

The locations of 39 feature channels were superimposed on a cerebral cortex atlas using the probabilistic registrations for NIRS channels (NFRI functions toolbox)<sup>31</sup> and shown in Figure 3. According to the spatial probability of probes and Brodmann’s map,<sup>32</sup> it was observed that these channels were mainly distributed on the left anterior-dorsolateral prefrontal cortex as well as part of the inferior frontal gyrus. The color gradient of channels indicated the number of features selected for the SVM model, e.g., CH-1 in red represented three feature variants (‘Integral Raw’, ‘Integral (CUM)’ and ‘Centroid (CUM) Positive’) at this channel were included to improve outcomes of individual classification.

**Prediction quality impacted by demographic factors**

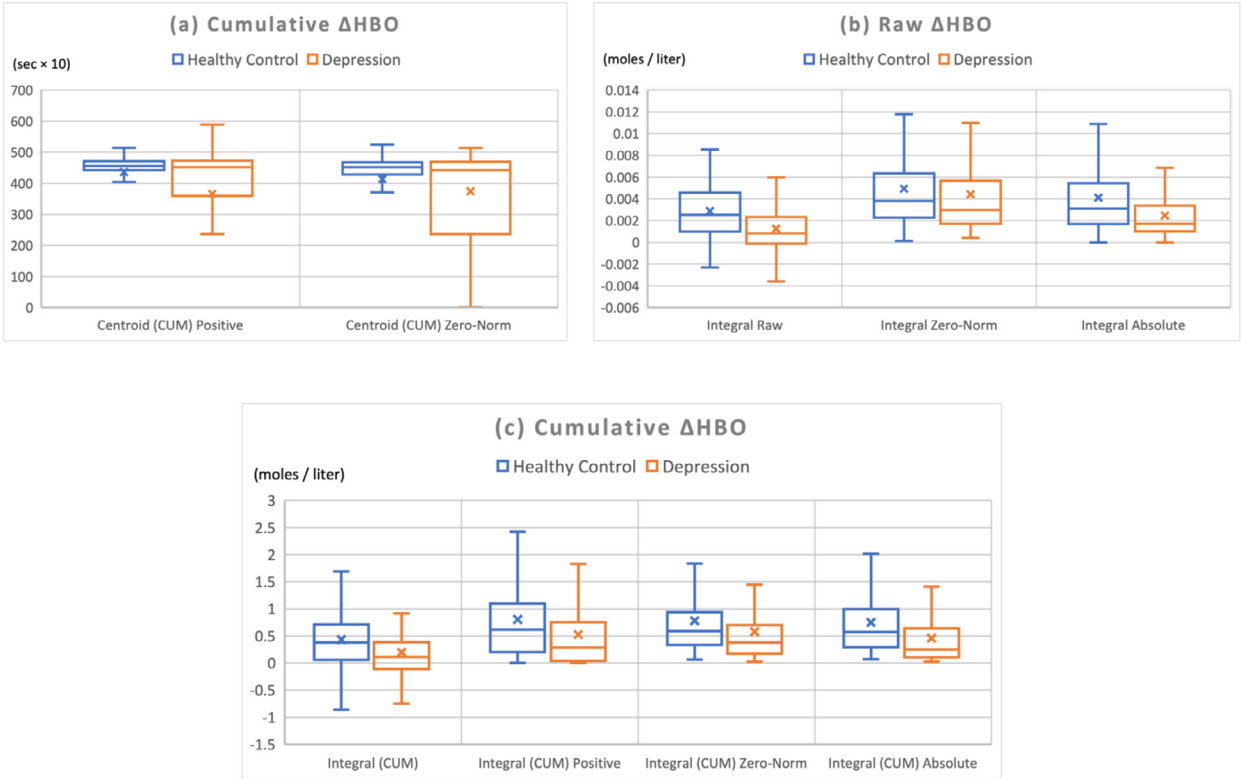
To assess the prediction robustness of classifier, we compared the accuracy or sensitivity of SVM and DA models across various demographic factors. Figure 4 (a–c) showed the accuracy comparisons in gender, past medical history, and years of education respectively. There was no obvious or consistent discrepancy of prediction accuracy for both classifiers, which implied these factors did not account for the classification rate.

As for the factor of age shown in Figure 4 (d), it is noticed that SVM presented higher accuracy than DA

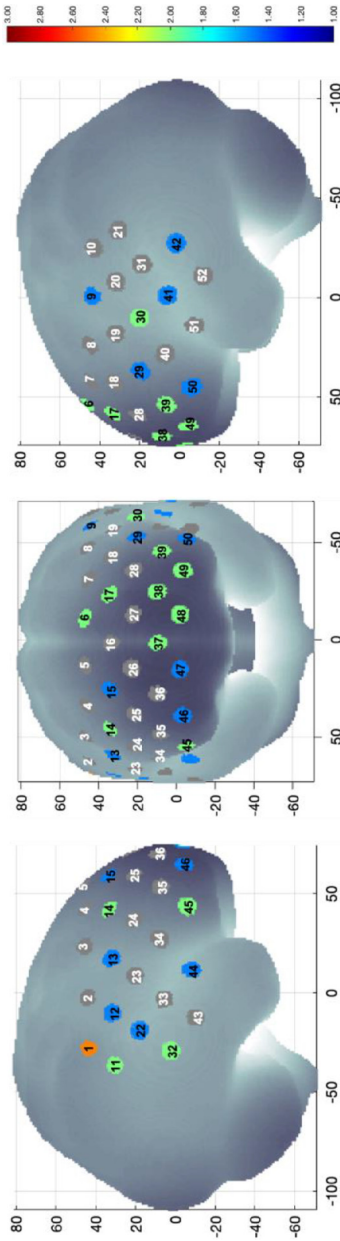
among subjects under 50 whereas DA was more effective to classify the older group with age ≥ 50. In Figure 4 (e), (f), the accuracies in chronic smokers were higher than that in non-smokers for both classifiers, while ~10% higher accuracy was achieved in chronic drinkers than non-drinkers for SVM. These performance differences were mainly due to the increasing recognition rate of people with MDD who smoke or alcohol misuse.

The influences of family history of MDD and antidepressant use on classification sensitivity were examined respectively and shown in Figure 4 (g), (h). The resulting sensitivity in patients with family history of MDD was ~7.5% higher than those without family history for DA classifier, while the sensitivity classified for patients in antidepressant treatment was ~8.9% higher than for unmedicated patients using SVM. To perform analysis on sensitivity differences among durations of MDD, we divided the people with MDD into five groups by the years of illness, each group with comparable sample size. As illustrated in Figure 4 (i), the sensitivity trend for SVM follows a decreasing pattern, which is not applicable to DA classifier. The averaged sensitivity of 80.0% in the MDDs with a duration of < 12 months indicates that DA is more promising for early detection of MDD.

In addition, the ANOVA test results on statistical correlation between the identified biomarkers and clinical/demographic factors were summarized in Supplementary Figure 4. Among the optimal 39 features applied in SVM classifier, twenty of the features were significantly correlated with MDD symptoms, which contributed to the main characteristics of hemodynamic response for differing MDDs from HCs. Less than five features were confirmed significant differences due to the effects of other factors, e.g., gender, age, medical history, etc. The statistics result from DA classifier also demonstrated



**Figure 2.** Box-plot comparisons of fusion features selected by GA and SVM between HC and MDD groups.



**Figure 3.** A total of 39 feature channels were determined to be applied for SVM classifier after GA optimization. Channels that did not include a feature variant were represented in gray. The color gradient indicates the number of feature variants in a specific channel contribute to the best classification model.

that the determined features could effectively represent the MDD-distinctive pattern in cortical activity. All the involved features and the respective statistical significance (*p*-value) were listed in *Supplementary Tables 7, 8*.

### Discussion

We proposed data transformation and feature selection methods to discriminately analyze the pattern of time-series NRIS signals. For the classification on the optimal features by five supervised models, the linear SVM model achieved the highest nested cross-validation accuracy of  $75.6\% \pm 4.7\%$  and the prediction rate of  $78.0\%$  (*Table 1*), which were superior to that using rank-based significant features and NB classifier ( $66.9\% \pm 5.3\%$  and  $76.9\%$ , *Supplementary Table 6*). However, the number of samples used for training and testing strongly affects the outcome of a classifier. Most previous studies usually produced results either on a small number of samples or have not yet been validated in external samples due to the lack of independent datasets.<sup>15</sup> In this study, the results were achieved on 363 subjects with comparable case-control samples, and the outcomes of classifiers were comprehensively verified by nested cross-validation and independent test set, which presented an outstanding generalization capability of this proposed framework amongst the classification studies conducted in a large-scale dataset, e.g., a correct prediction rate of  $74.6\%$  for recognizing people with MDD in fNIRS data<sup>33</sup> and accuracies  $60.8\sim 61.7\%$  at independent subgroup with high depression severity in resting-state fMRI data.<sup>34</sup>

The determined 39 features by GA and SVM were verified to contain 19 features without the presence of significant intra-class correlation, which revealed that the present feature selection approach was capable of identifying the features that were weak biomarkers by standalone but possessed a strong joint power of classifying people with MDD. Taken together, in comparison with other imaging modalities, e.g., positron emission tomography (PET) and functional magnetic resonance imaging (fMRI), fNIRS possesses advantages of non-ionizing radiation, long time monitoring, real-time measurement, easy operation, and lower cost.<sup>10</sup> The proposed ML approach further enables this optical technique to offer patient-oriented monitoring benefits with accurate, objective and individual assessment of MDD.

Cortical oxygen-hemoglobin changes measured by fNIRS offer a direct and sensitive indicator of cerebral neurophysiological function. In this study, as quantified with diminished relative intensity and inappropriate activation timing in the specific prefrontal regions of people with MDD, the abnormal hemodynamic response provides further evidence for the biological basis to diagnose MDD. More specifically, consistent with the results from prior research,<sup>11,12</sup> the determined MDD-related biomarkers in our findings revealed that



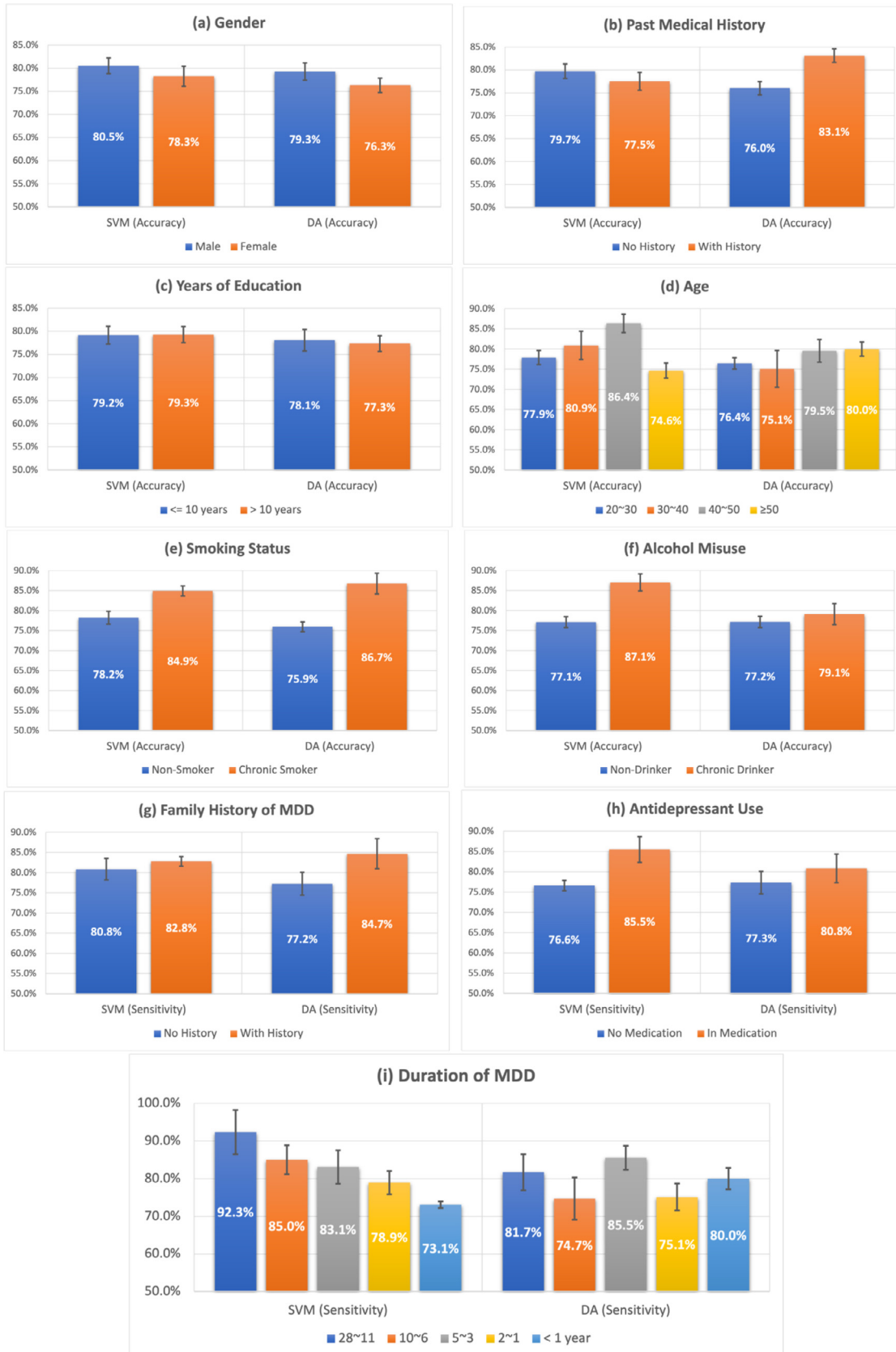


Figure 4. Effects of demographic factors on classification performances of SVM and DA classifiers.

the defined integral variants representing the intensity of cortical activity were significantly lower in MDDs than in HCs. In addition, to settle several disputes about the limited replication in the anomaly activation timing of MDD,<sup>18,35</sup> two centroid variants indicating the efficiency of hemodynamic response were first proposed and observed to be significantly dispersed among MDDs. As another independent indicator for evaluating the cognitive behaviour from the VFT outcome, the number of words generated by people with MDD ( $16.2 \pm 6.5$ ) was fewer compared to controls ( $19.8 \pm 5.7$ ) with statistical difference ( $p$ -value  $< 0.001$ , 95% confidence interval by two-tailed  $t$ -test). In brief, our analyses provide converging evidence to support that those who suffered from MDD would show objectively measurable cognitive deficits, i.e., trouble concentrating and difficulty generating words, when performing the VFT task.

The presented biomarkers of MDD were verified to be robust across demographic variables through two validations. The prediction performances evaluated by considering the influence of independent factors indicated that both SVM and DA classifiers were able to achieve the averaged accuracy  $> 74.6\%$  or sensitivity  $> 73.1\%$  at any specific factor-of-interest groups. In addition, the multiway ANOVA test on clinical and demographic factors confirmed the identified biomarkers were strongly associated with the MDD-distinctive consequence. Therefore, the results further evidenced that these promising models and biomarkers may be used to elucidate the underlying neural-activity heterogeneity for depression.

The comparisons of classification performance in factor-of-interest allowed for analyzing the potential implication of demographic heterogeneity on the hemodynamic pattern presented in fNIRS signals, which also offered improvement strategies to establish a more precise model. It was observed that people with MDD who were chronic smokers, misused alcohol, having long duration of illness, or psychiatric family history were more sensitive to recognition. This finding implies MDD interacted with these factors probably aggravate cognitive deficits of brain function and exhibit significant biomarkers. As mental health professionals in clinical settings require differentiating MDDs from HCs as accurately as possible. Further development on integrating multiple classifiers and other biological parameters to form a complementary structure would improve prediction quality and enhance anti-confounding effects. For example, the performance difference related to age groups suggests that we could choose SVM to recognize MDD in the young group while DA is preferable for diagnosing the elderly group. Moreover, previous study identified inflammatory biotypes derived from peripheral cytokine measurements as the potential biological predictors for personalizing depression treatments,<sup>36</sup> so data fusion between neuroimaging and inflammatory biomarkers provides comprehensive indicators to evaluate the progress of depression.

It should be noted though, that this study has several limitations. Firstly, the diagnosis data was obtained by a single psychiatrist, it would be desirable that the diagnostic process is performed by plural clinicians and cross-site to check the classification robustness across different clinical settings. Secondly, in the case of the opportunistic screening for depression, the suspected cases ask for medical attention can range in seriousness from mild, temporary episodes of sadness to severe, persistent depression. However, the studied sample set involved only two groups of subjects, i.e., MDDs and HCs, and recognizing different subtypes of depression would be beyond the ability of this algorithm. Therefore, increasing the diversity of samples and developing sensitive biomarkers aimed at severity assessment for depression are required before applying the fNIRS-assisted diagnosis in practical settings. Thirdly, the fNIRS measures and machine learning techniques may have potential applications in other important but more challenging diagnostic problems. For instance, MDD commonly co-occurs with borderline personality disorder (BPD). Since people with BPD often present depressive symptoms, it can be difficult to distinguish between MDD and BPD. Future research on appropriate cognitive tasks and analysis models is required for pattern mining from the brain activity of people with BPD and co-occurring depression. Lastly, the continuous-wave fNIRS technology is not yet useful for depth-sensitive, tomographic, and concentration measurements. Another promising NIRS modality, time-domain diffuse optical spectroscopy has recently been introduced in tracking the microvascular blood flow of multi-layered brain model.<sup>37,38</sup> With the application of this cost-effective system, researchers may obtain more cortical hemodynamic information for evaluating the pathological changes, locations, and causes of depression.

In conclusion, early and accurate identification of MDD contributes to a better treatment outcome, exploration of cost-effective biomarkers as part of the diagnostic criteria can be helpful in the fast case-finding as well as monitoring the course of the disorder. This study developed a machine learning framework to detect the distinctive pattern of MDD from a large-scale, case-control NIRS dataset and established a biomarkers-based objective analysis tool for the assessment of suspected cases. The statistical and discriminant analysis results further demonstrated the possibility of neuroimaging technique in clarifying the pathophysiological features of MDD and future data-driven diagnostics.

#### Contributors

Z.L., S.F.H. conducted the fNIRS scans, verified and analyzed the data, developed machine learning framework, and drafted the manuscript. R.H., S.S. and C.S.H. recruited participants. N.C., T.H.N., and B.T. provided

advice for fNIRS signal, machine learning and statistical analysis. R.S.M edited and provided a critical review of data analysis and manuscript. Z. L. and R.C.H. designed and supervised the study. All authors contributed to the preparation of this manuscript and approved the final article.

### Code and data availability

The raw data that support the findings of this study are available on request from the corresponding author. The raw data are not publicly available due to privacy or ethical restrictions. The MATLAB code in this study are available at <https://www.mathworks.com/matlabcentral/fileexchange/99804-machine-learning-framework-for-identification-of-depression>.

### Declaration of interests

All authors report no financial interests or potential conflicts of interest.

### Acknowledgments

This study is supported by NUS iHeattech Other Operating Expenses (R-722-000-004-731) and NUS Department of Psychological Medicine Other Operating Expenses (R-177-000-003-001), the Ministry of Education under HICOE scheme to CISIR, UTP, the Vingroup Innovation Foundation in project code VINIF.2019.DA14. We also gratefully acknowledge the FUJIFILM Healthcare Asia Pacific (RCA-2017-0763) for providing the functional near infrared spectroscopy machine, and the National University of Singapore Engineering in Medicine pitch-for-fund for providing the workstation to run machine learning algorithms.

### Supplementary materials

Supplementary material associated with this article can be found in the online version at doi:[10.1016/j.ebiom.2022.104027](https://doi.org/10.1016/j.ebiom.2022.104027).

### References

- Choo CC, Chew PKH, Ho CS, Ho RC. Quality of life in patients with a major mental disorder in Singapore. *Front Psychiatry*. 2018;9:727.
- Institute of Health Metrics and Evaluation. Global Health Data Exchange (GHDx). Available from: <http://ghdx.healthdata.org/gbd-results-tool?params=gbd-api-2019-permalink/d780dffbe8a381b25e1416884959e88b>. Accessed 9 April 2022.
- World Health Organization (WHO). *Depression and other Common Mental Disorders: Global Health Estimates*. Geneva: World Health Organization; 2017. Available from: <http://apps.who.int/iris/bitstream/handle/10665/254610/WHOMSD?sequence=1>. Accessed 9 April 2022.
- Rubio-Guerra AF, Rodriguez-Lopez L, Vargas-Ayala G, et al. Depression increases the risk for uncontrolled hypertension. *Exp Clin Cardiol*. 2013;18(1):10.
- Bilello JA. Seeking an objective diagnosis of depression. *Biomark Med*. 2016;10(8):861–875.
- Cui Y, Yang Y, Ni Z, et al. Astroglial Kir4.1 in the lateral habenula drives neuronal bursts in depression. *Nature*. 2018;554(7692):323–327.
- Kumar V, Shivakumar V, Chhabra H, et al. Functional near infrared spectroscopy (fNIRS) in schizophrenia: a review. *Asian J Psychiatry*. 2017;27:18–31.
- Rupawala M, Dehghani H, Lucas SJE, et al. Shining a light on awareness: a review of functional near-infrared spectroscopy for prolonged disorders of consciousness. *Front Neurol*. 2018;350.
- Herold F, Wiegel P, Scholkmann F, et al. Applications of functional near-infrared spectroscopy (fNIRS) neuroimaging in exercise–cognition science: a systematic, methodology-focused review. *J Clin Med*. 2018;7(12):466.
- Lai CYY, Ho CSH, Lim CR, et al. Functional near-infrared spectroscopy in psychiatry. *BJPsych Adv*. 2017;23(5):324–330.
- Fukuda M, Mikuni M. Clinical application of near-infrared spectroscopy (NIRS) in psychiatry: the advanced medical technology for differential diagnosis of depressive state. *Seishin Shinkeigaku Zasshi Psychiatr Neurol Jpn*. 2012;114(7):801–806.
- Zhang H, Dong W, Dang W, et al. Near-infrared spectroscopy for examination of prefrontal activation during cognitive tasks in patients with major depressive disorder: a meta-analysis of observational studies. *Psychiatry Clin Neurosci*. 2015;69(1):22–33.
- Tak S, Ye JC. Statistical analysis of fNIRS data: a comprehensive review. *Neuroimage*. 2014;85:72–91.
- Chekroud AM, Bondar J, Delgadillo J, et al. The promise of machine learning in predicting treatment outcomes in psychiatry. *World Psychiatry*. 2021;20(2):154–170.
- Grzenda A, Kraguljac NV, McDonald WM, et al. Evaluating the machine learning literature: a primer and user's guide for psychiatrists. *Am J Psychiatry*. 2021;178(8):715–729.
- Koutsouleris N, Dwyer DB, Degenhardt F, et al. Multimodal machine learning workflows for prediction of psychosis in patients with clinical high-risk syndromes and recent-onset depression. *JAMA Psychiatry*. 2021;78(2):195–209.
- Zhu Y, Jayagopal JK, Mehta RK, et al. Classifying major depressive disorder using fNIRS during motor rehabilitation. *IEEE Trans Neural Syst Rehabil Eng*. 2020;28(4):961–969.
- Husain SF, Yu R, Tang TB, et al. Validating a functional near-infrared spectroscopy diagnostic paradigm for Major Depressive Disorder. *Sci Rep*. 2020;10(1):1–9.
- Husain SF, Tang TB, Yu R, et al. Cortical haemodynamic response measured by functional near infrared spectroscopy during a verbal fluency task in patients with major depression and borderline personality disorder. *EBioMedicine*. 2020;51:102586.
- First M., Williams J., Karg R., Spitzer R. Structured clinical interview for DSM-5 disorders. Clinical Trials Vers (SCID-5-CT) 2015.
- Hamilton M. A rating scale for depression. *J Neurol Neurosurg Psychiatry*. 1960;23(1):56.
- Edition F. Diagnostic and statistical manual of mental disorders. *Am Psychiatr Assoc*. 2013;21:591–643.
- Zanelli J, Reichenberg A, Morgan K, et al. Specific and generalized neuropsychological deficits: a comparison of patients with various first-episode psychosis presentations. *Am J Psychiatry*. 2010;167(1):78–85.
- Kocsis L, Herman P, Eke A. The modified Beer-Lambert law revisited. *Phys Med Biol*. 2006;51(5):N91–N98.
- Strangman G, Culver JP, Thompson JH, et al. A quantitative comparison of simultaneous BOLD fMRI and NIRS recordings during functional brain activation. *Neuroimage*. 2002;17(2):719–731.
- Hoshi Y, Kobayashi N, Tamura M. Interpretation of near-infrared spectroscopy signals: a study with a newly developed perfused rat brain model. *J Appl Physiol*. 2001;90(5):1657–1662.
- Herold F, Wiegel P, Scholkmann F, et al. Functional near-infrared spectroscopy in movement science: a systematic review on cortical activity in postural and walking tasks. *Neurophotonics*. 2017;4(4):041403.
- Tadel F, Baillet S, Mosher JC, Pantazis D, Leahy RM. “Brainstorm: A user-friendly application for MEG/EEG analysis”. *Comput Intell Neurosci*. 2011;879716. <https://doi.org/10.1155/2011/879716>. 13 pages. <https://www.hindawi.com/journals/cin/2011/879716/>.
- Nirstorm - Brainstorm plugin for fNIRS data analysis. Available online: <https://github.com/Nirstorm/nirstorm>. Accessed 9 April 2022.
- Oluleye B, Leisa A. A genetic algorithm-based feature selection. *Int J Electron Commun Comput Eng*. 2014;5(4):889–905. <https://ro.ecu.edu.au/cgi/viewcontent.cgi?article=1654&context=ecuworks-post2013>.

- 31 Singh AK, Okamoto M, Dan H, et al. Spatial registration of multi-channel multi-subject fNIRS data to MNI space without MRI. *Neuroimage*. 2005;27:842–851.
- 32 Chowdhury A, Liu C, Yu R. The neural correlates of reaching focal points. *Neuropsychologia*. 2020;140:107397.
- 33 Takizawa R, Fukuda M, Kawasaki S, et al. Neuroimaging-aided differential diagnosis of the depressive state. *Neuroimage*. 2014;85(1):498–507.
- 34 Gao S, Calhoun VD, Sui J. Machine learning in major depression: From classification to treatment outcome prediction. *CNS Neurosci Ther*. 2018;24(11):1037–1052.
- 35 Wei YY, Chen Q, Curtin A, et al. Functional near-infrared spectroscopy (fNIRS) as a tool to assist the diagnosis of major psychiatric disorders in a Chinese population. *Eur Arch Psychiatry Clin Neurosci*. 2021;271(4):745–757.
- 36 Lee Y, Mansur RB, Brietzke E, et al. Peripheral inflammatory biomarkers define biotypes of bipolar depression. *Mol Psychiatry*. 2021;26(7):3395–3406.
- 37 Hasnain A, Mehta K, Zhou X, et al. Laplace-domain diffuse optical measurement. *Sci Rep*. 2018;8(1):1–8.
- 38 Mehta KB, Hasnain A, Zhou X, et al. Spread spectrum time-resolved diffuse optical measurement system for enhanced sensitivity in detecting human brain activity. *J Biomed Opt*. 2017;22:(4) 045005.