

Functional near-infrared spectroscopy for human brain age group classification by machine learning

Martti Ilvesmäki¹, Hany Ferdinando¹, Kai Noponen², Tapio Seppänen², Vesa Korhonen¹, Vesa Kiviniemi¹, and Teemu Myllylä¹

¹Oulun yliopisto

²Oulun Yliopisto Konenaon ja signaalianalyysin tutkimuskeskus

November 15, 2023

Abstract

Aging brain undergoes multiple structural and functional changes. These may contribute to an increased risk of neurodegenerative disease (NDD) and other age-related diseases, highlighting the importance of assessing deviations from healthy brain aging trajectory. In this human brain study, 50 healthy adults were investigated by functional near-infrared spectroscopy (fNIRS). A resting state single channel multiwavelength fNIRS was measured from the forehead in a supine position. The subjects were divided into four age groups. A machine learning approach was utilized for age group classification by using support vector machine and random forest learners with nested cross-validation. The results suggest brain aging effects being more distinct in the oldest age group and a difference in the brain aging for the subjects of the in-between groups. Our study shows high potential for the use of fNIRS in the analysis of brain aging.

RESEARCH ARTICLE

Functional near-infrared spectroscopy for human brain age group classification by machine learning

Martti Ilvesmäki^{1*} | Hany Ferdinando¹ | Kai Noponen² | Tapio Seppänen² | Vesa Korhonen^{1,3} | Vesa Kiviniemi^{1,3} | Teemu Myllylä^{1,4}

1 Research Unit of Health Sciences and Technology, University of Oulu, Oulu, Finland

2 Center for Machine Vision and Signal Analysis Research Unit, University of Oulu, Oulu, Finland

3 Department of Diagnostic Radiology, Oulu University Hospital, Finland

4 Optoelectronics and Measurement Techniques Research Unit, University of Oulu, Oulu, Finland

* Correspondence

Martti Ilvesmäki, University of Oulu, Pentti Kaiteran katu 1, 90570 Oulu, Finland

Martti.ilvesmaeki@oulu.fi Aging brain undergoes multiple structural and functional changes. These may contribute to an increased risk of neurodegenerative disease (NDD) and other age-related diseases, highlighting the importance of assessing deviations from healthy brain aging trajectory. In this human brain study, 50 healthy adults were investigated by functional near-infrared spectroscopy (fNIRS). A resting state single channel multiwavelength fNIRS was measured from the forehead in a supine position. The subjects were divided into four age groups. A machine learning approach was utilized for age group classification by using support vector machine and random forest learners with nested cross-validation. The results suggest brain

aging effects being more distinct in the oldest age group and a difference in the brain aging for the subjects of the in-between groups. Our study shows high potential for the use of fNIRS in the analysis of brain aging.

KEYWORDS

fNIRS, brain aging, aging, machine learning

ABBREVIATIONS AD, Alzheimer’s disease; bacc, balanced accuracy; CSF, cerebrospinal fluid; CV, cross-validation; DPF, differential path length factor; dPhaseTE, differential phase transfer entropy; EEG, electroencephalography; ESSC, envelope-signal spectra correlation; fMRI, functional magnetic resonance imaging; fAPF, fractional amplitude of physiological fluctuations; fNIRS, functional near-infrared spectroscopy; FC, functional connectivity; H₂O, water, HbO, oxygenated-hemoglobin; HbR, deoxygenated-hemoglobin; HbT, total-hemoglobin; LED, light-emitting diodes; LOOCV, leave-one-out-cross-validation; MBBL, modified beer-lambert law; MCCV, Monte Carlo cross-validation; MREG, magnetic resonance encephalography; MRI, magnetic resonance imaging; MRMR, minimum redundancy maximum relevance; MCI, mild cognitive impairment ; NDD, neurodegenerative disease; NIR, near-infrared; PFC, prefrontal cortex; PSD, power spectral density; SampEn, sample entropy; SD, source-detector; SVM, support vector machine; TBI: traumatic brain injury

1 | INTRODUCTION

The life expectancy of the global population is steadily continuing to increase [1]. As aging possess an increased risk of developing neurodegenerative diseases (NDD), such as Alzheimer’s disease (AD), the better understanding of healthy brain aging process is of importance. The assessment of the brain aging trajectory is of interest to detect of deviations from the healthy trajectory, and to monitor the efficacy of the treatments and interventions. The increased understanding of the brain aging process can lead to development of improved interventions, with an aim to delay age-related disease onset and to reduce their severity in the later years of life.

Brain aging has been widely studied, and multiple interconnected structural and functional changes have been reported. The changes can be observed in different scales of interest, such as micro or macro scale. Some notable changes are DNA damage, which is considered as one of the key hallmarks of aging, cerebrovascular changes observed as decreased vessel size, reduced number of capillaries, small infarctions and microbleeds, causing overall decreased cerebral perfusion, changes in functional connectivity (FC), and regional brain atrophy, manifested in overall decrease of brain volume and weight. [2] In recent years, neuroimaging methods such as magnetic resonance imaging (MRI) and electroencephalography (EEG) have been utilized to measure the healthy brain aging trajectory. The methods commonly aim to quantify the biological age of the brain, which is assumed to be congruent with the chronological age of the healthy individuals, while being increased in the presence of aging related disease and decreased in successful aging [3]

Functional near-infrared spectroscopy (fNIRS) presents an intriguing method for the assessment of aging related-changes in the brain due to its good temporal resolution, affordability, and portable light-weight equipment. This enables development of brain health monitoring applications which can be used in a natural environment. The method is based on applying light source-detector (SD) pairs with minimum distance of 2.5 cm to the scalp [4], utilizing wavelengths in the optical window of approximately 650 nm – 1000 nm, where the near-infrared (NIR) light is able to propagate through the superficial biological layers of the skin, skull and cerebrospinal fluid (CSF), reaching the cortical brain layer [5]. Finally, the back scattered light is recorded by the detector pair. By applying modified beer-lambert law (MBBL), the method can be used to quantify chromophore relative concentrations [6]. The method requires selection of at least two wavelengths from the different sides of the isobestic point of the chromophores of interest. Commonly the concentration changes of oxygenated-hemoglobin (HbO) and deoxygenated-hemoglobin (HbR) are monitored, as the cortical changes are related to the brain metabolism caused by neural activity [5]. However, the changes in water (H₂O) can be detected in NIR region as well by selecting wavelength accordingly. The measurement of neurohydrodynamics have been demonstrated recently in the assessment of glymphatic system by fNIRS [7].

fNIRS has been used in multiple aging studies showing evidence for the use of measurement method in the brain aging assessment. In FC studies fNIRS has been used successfully to confirm the finding of increased brain region co-operation, assumed to be due to brain’s adaptation to the structural changes. [8] Furthermore, Nguyen et al. found evidence of age-related decreased FC detected during verbal fluency task, although evidence of age-related changes during oddball and resting state was not found. [9]

Arterial stiffness has been associated with aging and is one of the leading risk factors for hypertension [10]. As younger subject’s arteries are more compliant, the mechanical stress to the brain caused by pulsatility is reduced in comparison to older adults with stiffer arteries, causing damage to the brain’s microvasculature. The arterial stiffness can be detected from the changed pulse shape form, and the distinguishing of younger and older adults has been shown to be successful by fNIRS by using pulse shape parameters. [11] Additionally, fNIRS has been used with MRI to show association between aging related cortical thinning and regional pulsatility. [12]

fNIRS has been used widely to study aging related changes in prefrontal cortex (PFC) activation, which is involved in the executive function. The results have shown age related differences in the PFC activation [13–16]. Furthermore, fNIRS has been used to characterize differences between younger and older adults, and mild cognitive-impairment (MCI) and AD groups [17,18], and to show age-related differences in within-session trainability of hemodynamic response. [19]

The results of the previous findings suggest potential for the use of fNIRS in the study of aging. To investigate this further, the use of single channel resting state fNIRS to detect aging caused differences in healthy adults is studied by utilizing machine learning (ML) methods in an age group classification task. In addition to traditional analysis of HbO, HbR and total hemoglobin (HbT) changes, the effects of relative H₂O concentration changes and raw fNIRS signals are used. Multiple features based on the latest fNIRS, EEG and functional magnetic resonance imaging (fMRI) studies are applied. The results presented are based on the research conducted for the master’s thesis of Martti Ilvesmäki [20].

2 | EXPERIMENTAL

2.1 | fNIRS measurements and data acquisition

The data collection followed the guidelines established by the Declaration of Helsinki and the study was approved by the regional Ethical Committee of Northern Ostrobothnia Hospital District in Oulu University Hospital. All participants of the study signed informed consent letters before the measurements. The data was collected with hospital multi-modal MRI compatible frequency coded fNIRS utilizing high power light-emitting diodes (LED) coupled to optical fibres, with modifiable wavelength selection [21,22].

Total of 56 healthy controls participated in the study. The participants restrained from the use of alcohol 12 hours prior the measurements. With the exception of one individual, the participants were non-smoking. Data from single optode with 3 cm source detector (SD) separation distance, placed on the left side of the forehead corresponding to the Fp1 of the 10-20 system of the international federation for EEG electrode placement, was used with sampling frequency of 800 Hz. The subjects were in a supine position in resting state during the measurement, which lasted for approximately 5.5 – 10 minutes in total. After visual inspection, the data of six subjects was discarded, as the data of four subjects had gross movement artefacts and for two subjects the length of the measurement was insufficient. The final dataset consisted of 50 subjects with mean age of 42.3 ± 15.2 and age range of 23 - 67 (22F, 28M). The signal lengths were unified to the length 5.5 minutes by selecting signal of corresponding length from the start of the measurement. It is noteworthy, that the chronological and biological age of the brain may differ as pathologies, such as schizophrenia, traumatic brain injury (TBI) and AD, have been shown to affect the aging trajectory of the brain [23–25]. Thus, although the subjects in the study were considered as healthy controls and thus the chronological age is assumed to be similar to the subject’s brain age, the true brain age is not known. Therefore, a potential undiagnosed pathology could affect the results of the study.

The used wavelengths of 690 nm, 830 nm and 980 nm were selected due to the absorption being dominated

by HbR, HbO and H₂O respectively in the corresponding wavelength. Additionally, wavelength of 810 nm was used, which is close to the isobestic point of the hemoglobin.

2.2 | Signal processing and feature extraction

All the signal processing, machine learning and data-analysis methods were implemented by using MATLAB version R2022b. Figure 1 illustrates the signal processing methods used, from signal acquisition until feature extraction.

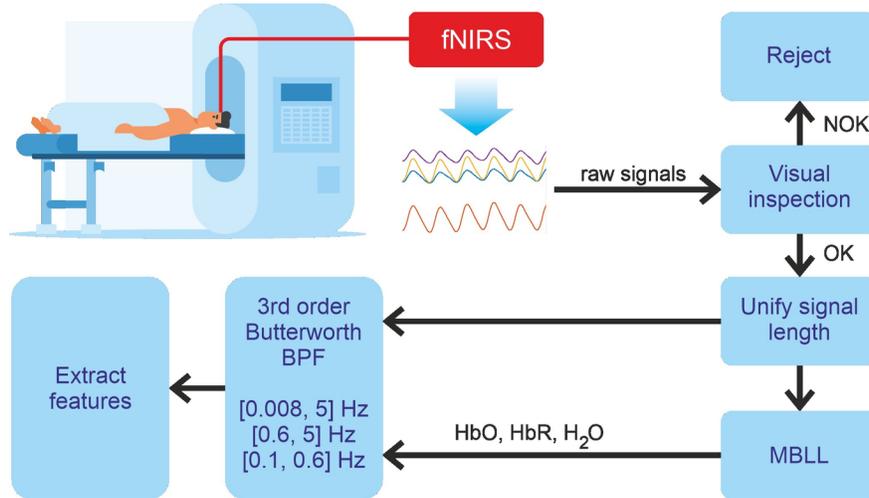


FIGURE 1 Feature extraction pipeline. The four-wavelength fNIRS data is collected from the subjects and several pre-processing methods are applied before extracting the features.

Total of 209 features were extracted from fNIRS and fNIRS derived chromophore relative concentration signals [20]. The differential path length factor (DPF) used in MBLL was 5.97 and the extinction coefficients were based on the work of Cope [26]. The summary of the features is shown in Table 1. Some of the features have been described in the latest fNIRS, EEG and fMRI NDD studies while some are common features, such as signal standard deviation and gender of the subject. The frequency bandwidths of respiratory (0.1 - 0.6 Hz), cardiac (0.6 - 5 Hz) and full (0.008 - 5 Hz) were utilized when extracting fNIRS features. In addition to noise attenuation, the bandwidth specific features are assumed to contain information related to different physiological events and could contain more predictive power in the brain aging assessment.

Spectral entropy, fractional amplitude of physiological fluctuations (fAPF) and relative power derived features have been used in fNIRS and EEG studies related to aging and AD research [18,27–29]. Furthermore, pulse shape derived from fNIRS has been used to distinguish AD subjects from healthy controls [11]. Thus, the features were selected as they are assumed to have potential predictive power in the age group classification task. In addition, the features of differential phase transfer entropy (dPhaseTE) [30,31], sample entropy (sampEn) [32,33], and correlation between HbT and H₂O [34] were utilized. Unique feature of envelope-signal spectra correlation (ESSC) was also used. The feature is calculated by computing the correlation between the power spectral density (PSD) of the signal and its envelope. The feature describes the relationship between the spectral power of the original signal and its low frequency component.

TABLE 1 List of all features extracted from the fNIRS and derived relative chromophore concentration signals used in the study.

Feature name	Number of features per subject	Frequency bands	fNIRS signals
HbT vs. H ₂ O correlation	3	Full, cardiac, resp*	Raw, conc**
dPhaseTE	4 x 3 x 2***	Full, cardiac, resp	Conc
ESSC	8 x 3	Full, cardiac, resp	Raw, conc
fAPF	8 x 2	Cardiac, resp	Raw, conc
Gender	1	-	-
Pulse Shape of 830nm (mean, min, max, std, iqr)	5	Cardiac	830 nm
Relative power (min, mean, 95th quantile, std)	8 x 2 x 4	Cardiac, resp	Raw, conc
Sample entropy	8 x 3	Full, cardiac, resp	Raw, conc
Spectral entropy	8 x 3	Full, cardiac, resp	Raw, conc
Standard deviation	8 x 3	Full, cardiac, resp	Raw, conc

* Respiratory bandwidth, ** Relative concentration of the chromophores, *** dPTE features computed for two directions, e.g., HbO - Water and Water - HbO.

2.3 | Machine learning methods

Random forest ensemble and support vector machines (SVM) with linear, polynomial and gaussian kernels were selected as learning algorithms. The feature selection for SVM learners was implemented by using minimum redundancy maximum relevance (MRMR) algorithm by MATLAB’s `fscmr` function. The hyperparameter optimization was conducted by using Bayesian optimization with MATLAB’s function `bayesopt` with 60 iterations.

Due to relatively small sample size used in the study, each test utilized nested cross-validation (CV) to provide robust and unbiased performance estimate [35]. The nested CV protocol is illustrated in Figure 2. The outer loop consisted of stratified Monte Carlo cross-validation (MCCV) or K-fold CV depending on the test, while the inner loop used leave-one-out cross-validation (LOOCV). The inner loop contained hyperparameter tuning for each model, including hyperparameters for the feature selection, which in the case of MRMR was number of selected features based on MRMR feature ranking. For all SVM learners box constraint was tuned with default search space, while for polynomial SVM the polynomial order was tuned using search space of [2,3], and for gaussian SVM the kernel scale was tuned using default search space. For random forest, the tuned hyperparameters were minimum leaf size and number of predictors to sample with default search space, and number of trees with [5, 1000] search space. The overall CV performance was recorded by using balanced accuracy (bacc). When analysing the results, the bacc is evaluated either as poor (bacc [?] 60 %), moderate (< 60 % < bacc [?] 70 %), good (70 % < bacc [?] 80 %) or excellent (bacc > 80 %).

2.3.1 | Classification to age groups with pre-defined age groups

The first classification test was conducted by dividing the data into four pre-defined age groups. The age thresholds for each group were selected by aiming for approximately equal sized groups, while still containing wide age distribution. The age group division is illustrated in Figure 3. Stratified MCCV with 20-80 % test-train-split and 1000 iterations was used for the outer CV.

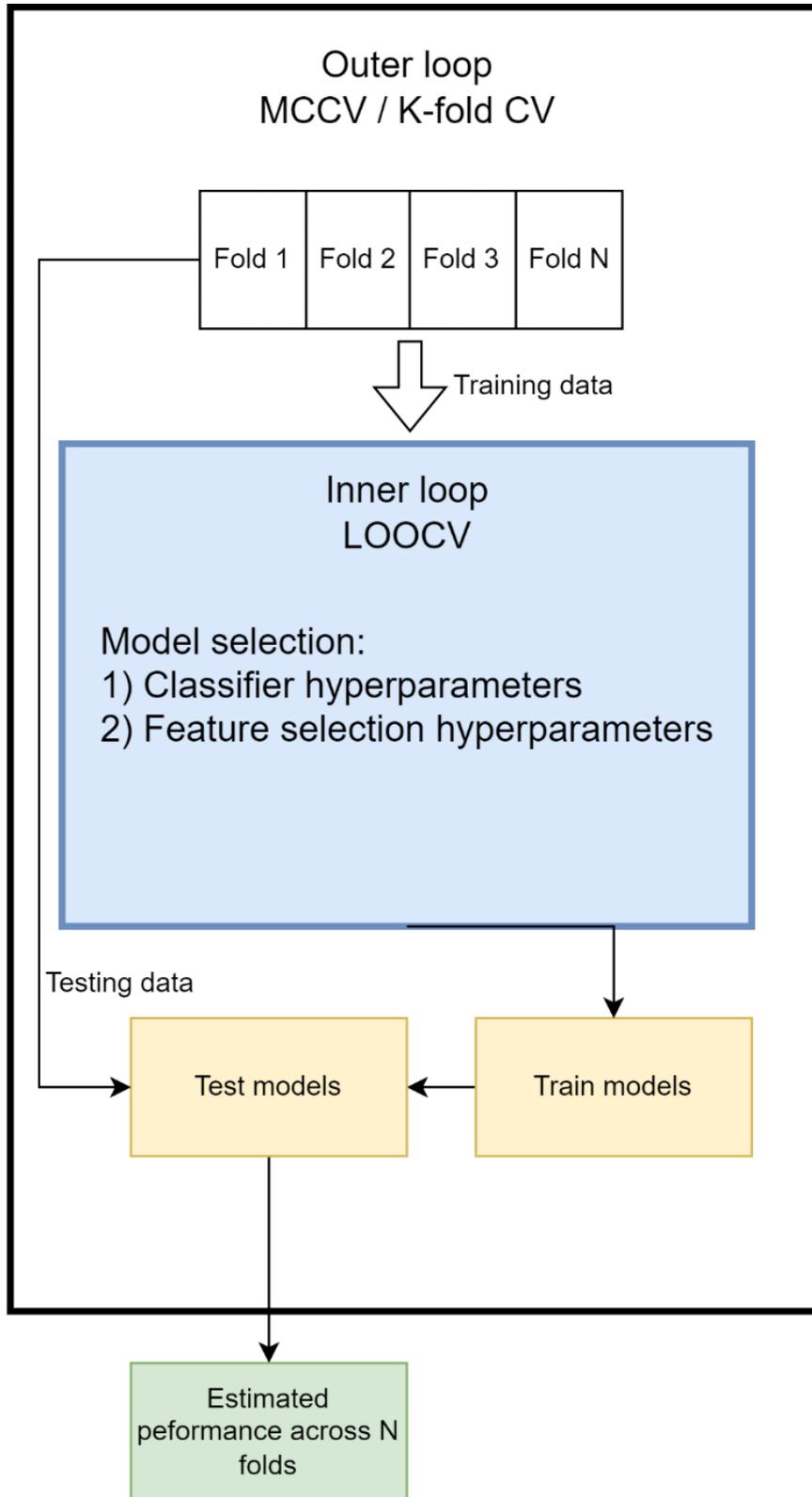


FIGURE 2 Nested cross-validation protocol used in the study. The inner loop (light blue) is used for the model selection, while in the outer loop selected models are trained based on the tuned hyperparameters. The procedure is repeated N times based on the outer loop CV method, resulting in the output (green) of estimated performance. For the outer loop, MCCV was used with pre-defined age groups classification and K-fold CV in the age threshold tests. The inner loop model selection included feature selection hyperparameters selection for the MRMR algorithm used by SVM learners.

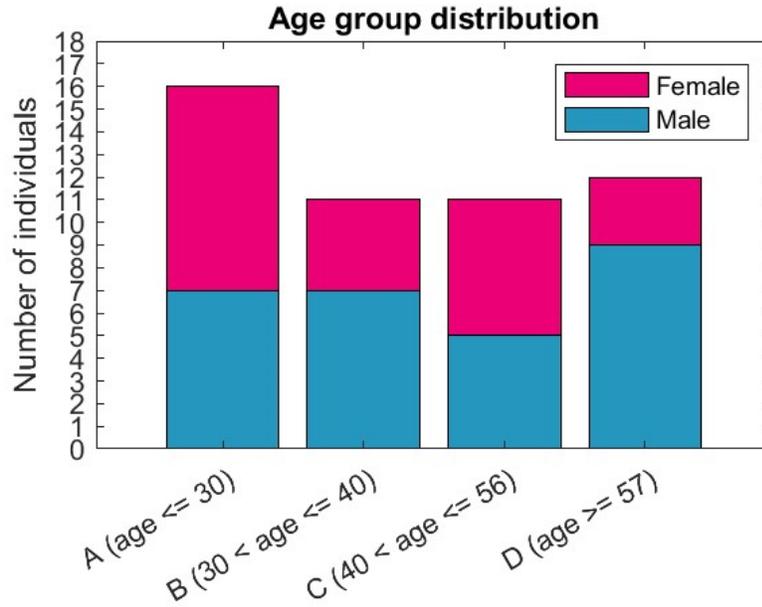


FIGURE 3 Age group division. The bars represent number of subjects in each age group, while the colours illustrate the within-class gender distribution.

2.3.2 | Classification to age groups with variable age group thresholds

The effect of varying the age group thresholds was further analysed. Age grid of 26 to 63 was used to evaluate different thresholds. Instead of using all the ages between minimum and maximum in the age grid, only the ages appearing in the data were considered as part of the grid. As a result, the best corresponding thresholds were found for each classifier based on the achieved balanced accuracy. For each threshold, stratified K-fold CV was used with K=5. Figure 4 illustrates the age threshold finding process for each test.

In the age threshold test 1, all the data was used and split into two age groups of young and old adults using the aforementioned age grid. For the test 2, two age thresholds were used and the in-between data was excluded from the ML pipeline, while the beyond threshold subjects were sorted in to young and old adult groups. Test 3 used similar method as test 2, but the in-between data was used as a third rejection class. The goal of the tests 2 and 3 was to identify how the exclusion or inclusion of the in-between groups affects the classifiers performance.

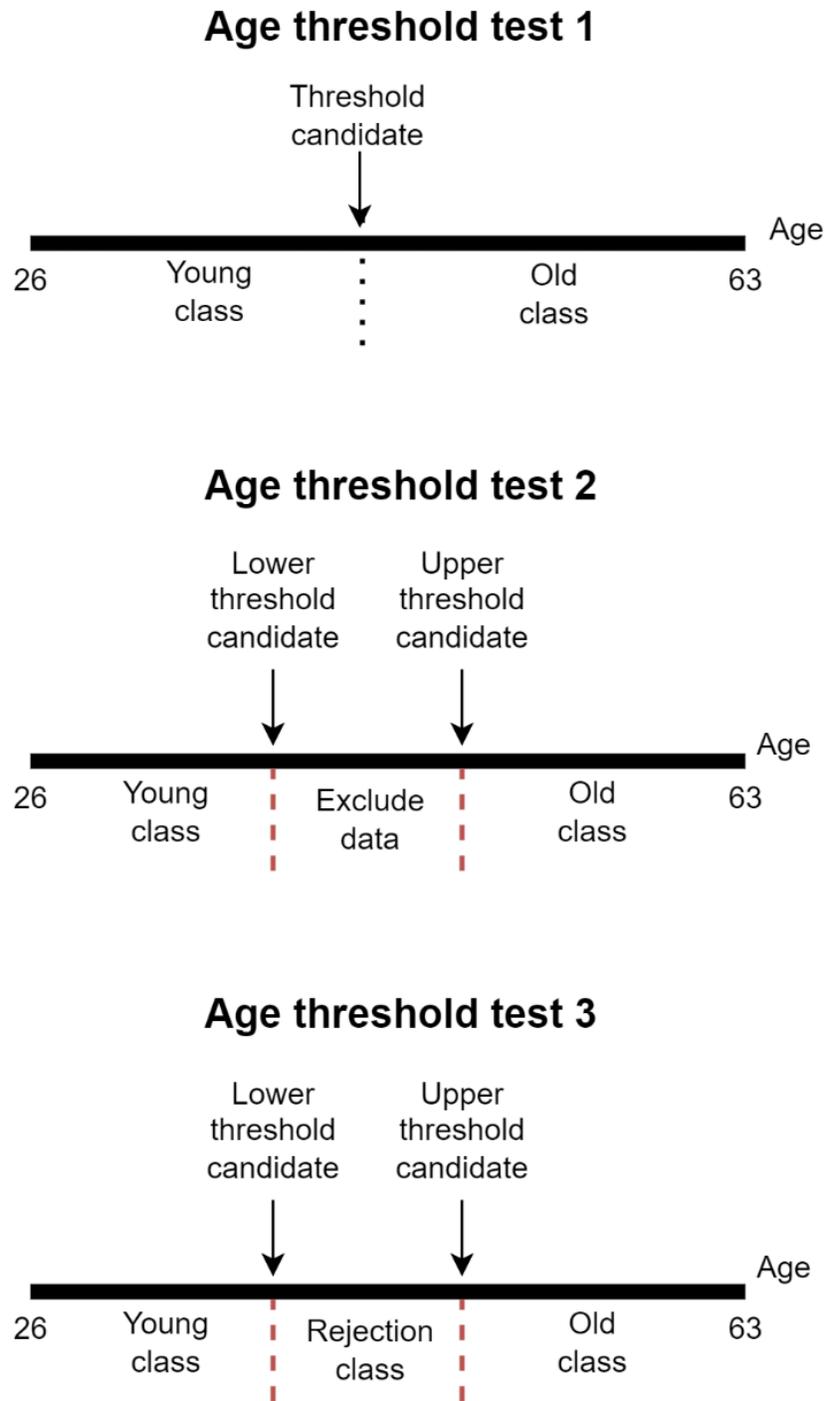


FIGURE 4 Illustration of the age group threshold search protocol for different age threshold tests. The test 1 was used to find a single threshold for dividing data into two classes, while the tests 2 and 3 utilized two thresholds for either excluding the data or creating a third rejection class.

3 | RESULTS AND DISCUSSION

3.1 | Classification with pre-defined age groups

Table 2 illustrates the results of the classification to pre-defined age groups. The classification between all four groups shows moderate results with all classifiers reaching bacc between 66 – 67 %. Interestingly, the highest performance is reached in the classification task of groups A vs. D, where the age difference between the groups is the largest. The best performing model confusion matrices for the previously mentioned tasks are shown in the Figure 6. Furthermore, the performance in binary classification task of younger groups A, B, C vs. the oldest group D can be seen to decline linearly as the age gap between groups decreases as shown in the Figure 5. The result indicates that distinguishing the younger groups from the oldest group is better when the age gap between the groups is higher. On the other hand, the binary classification results between the groups A, B and C show poor performance, suggesting similarity between the groups. The majority of B and C classifications is done to the group A, which is likely affected by the class imbalance, group A being the largest. With more balanced class sizes, more equal proportion of miss-classified samples between groups A, B and C would be expected in the case of suspected homogenous samples. Overall, the results suggest no great difference between the younger groups A, B and C in this study setup, while the age-related differences become more distinct with the oldest age group.

TABLE 2 . The classifiers performance for the age group classification test with pre-defined age groups. The results include balanced accuracy and 95 % confidence intervals inside the brackets. Polynomial SVM in A vs. D age group classification task achieved the highest performance (highlighted in bold).

Task	Linear SVM	Polynomial SVM	Gaussian SVM	Random Forest
A vs. B vs. C vs. D	67.05 % (66.67 – 67.43)	66.58 % (66.21 – 66.94)	66.38 % (66.05 – 66.70)	66.09 % (65.75 – 66.43)
A vs. D	77.23 % (76.02 – 78.45)	78.98 % (77.80 – 80.15)	73.68 % (72.44 – 74.91)	73.46 % (72.33 – 74.58)
B vs. D	64.23 % (62.87 – 65.58)	63.15 % (61.78 – 64.52)	62.63 % (61.25 – 64.00)	69.33 % (67.97 – 70.68)
C vs. D	54.30 % (53.01 – 55.59)	52.35 % (51.15 – 53.55)	54.68 % (53.45 – 55.90)	57.90 % (56.59 – 59.20)
B vs. C	40.08 % (38.90 – 41.25)	41.50 % (40.29 – 42.71)	41.68 % (40.41 – 42.94)	44.50 % (43.11 – 45.89)
A vs. C	45.76 % (44.93 – 46.58)	53.49 % (52.23 – 54.75)	50.04 % (49.23 – 50.86)	46.03 % (45.15 – 46.92)
A vs. B	46.24 % (45.41 – 47.07)	45.86 % (44.87 – 46.85)	47.79 % (47.21 – 48.37)	46.25 % (45.23 – 47.27)

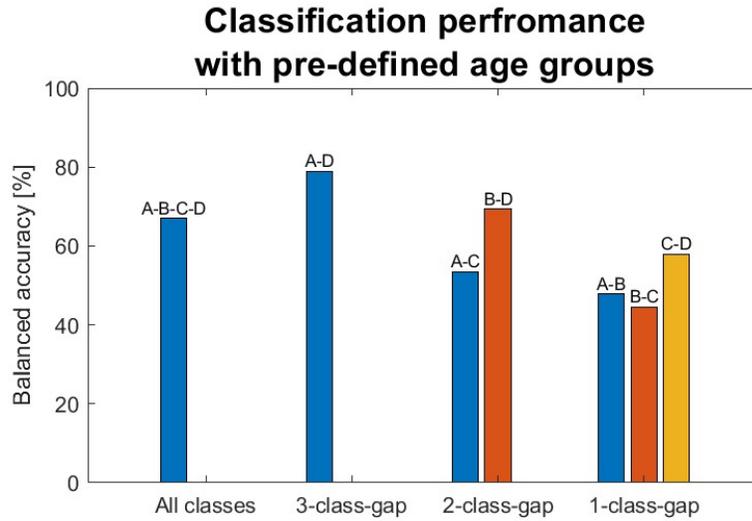


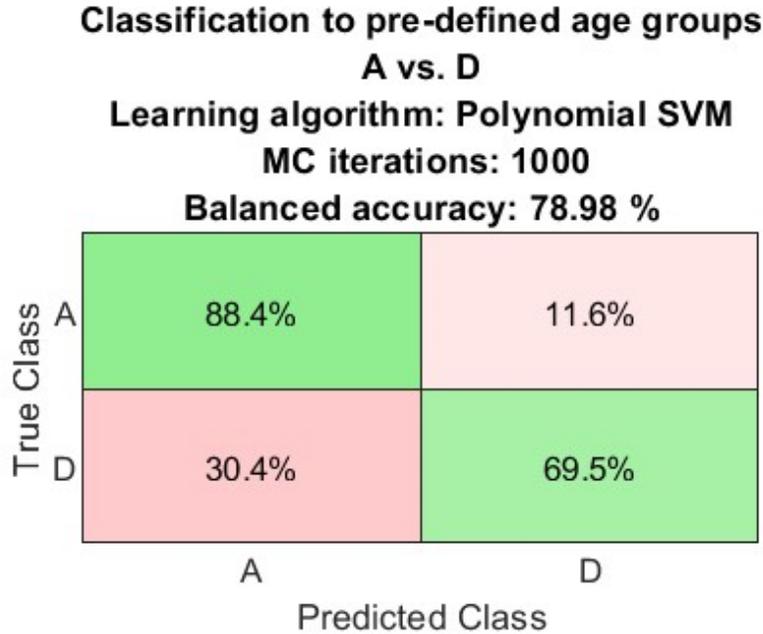
FIGURE 5 The best performing models for the classification to pre-defined age groups A, B, C and D. The chart illustrates the performance with low and high age gap between the groups.

Classification to pre-defined age groups A vs. B vs. C vs. D Learning algorithm: Linear SVM MC iterations: 1000 Balanced accuracy: 67.05 %

True Class	A	B	C	D
A	59.5%	17.7%	13.4%	9.4%
B	61.3%	8.8%	17.3%	12.6%
C	50.5%	18.7%	10.5%	20.3%
D	33.8%	9.3%	10.7%	46.2%

Predicted Class

(A)



(B)

FIGURE 6 The best performing model normalized confusion matrices for the classification to pre-defined age groups for the tasks A vs. B vs. C vs. D (chart A) and A vs. D (chart B).

3.2 | Classification with variable age group thresholds

Table 3 shows the results of the age threshold tests for each classifier. The performance in the age threshold test 1, classification of the subjects to young and old groups using different age thresholds, achieved moderate to good performance (bacc: 63.44 – 71.05 %). Three of the classifiers found similar best age threshold (39 – 41), while the overall best performing model of linear SVM performed best with an age threshold of 56. However, it is noteworthy that there is a clear uneven distribution of subjects when using the suggested age threshold of 56 (N_young: 36, N_old: 14). The class imbalance is similar to one observed in classification to pre-defined age groups but pronounced more strongly. On the other hand, it can be observed that the second-best performing classifier of Gaussian SVM reached the highest performance with an age threshold of 42, where the achieved performance can be considered moderate with more balanced classes (N_young: 27, N_old: 23). The found best age threshold of 56 matches the threshold used for group D when using pre-defined age groups. It is visible from Figure 6 A that in the most cases subjects B and C are classified to group A. The result indicates that with the current dataset the differences in the study population seem to be most pronounced with the threshold of 56. Interestingly, the threshold was in agreement with the class distribution in Figure 3A.

In the age threshold test 2, where two age thresholds were selected, and the in-between subjects were excluded from the ML loop, the results show good to excellent performance with all the classifiers (bacc: 79.09 – 81.50 %). The lower threshold varies between 27 to 37, while the upper threshold was between 57 to 62. The performance of the age threshold test 3 with rejection class can be considered good with all the classifiers (micro averaged bacc: 72 – 74.68 %). Polynomial and Gaussian SVM, and random forest classifiers performed best with the same age thresholds (44 – 62), while linear SVM used different range for the rejection class (27 – 57). As visible in Figure 7 confusion matrix of the best performing model polynomial SVM, most of the predictions were done in favour of the biggest group of the young subjects, as noticed as well in the previous tests. The closer examination of class sizes showed that the biggest class had approximately three times more

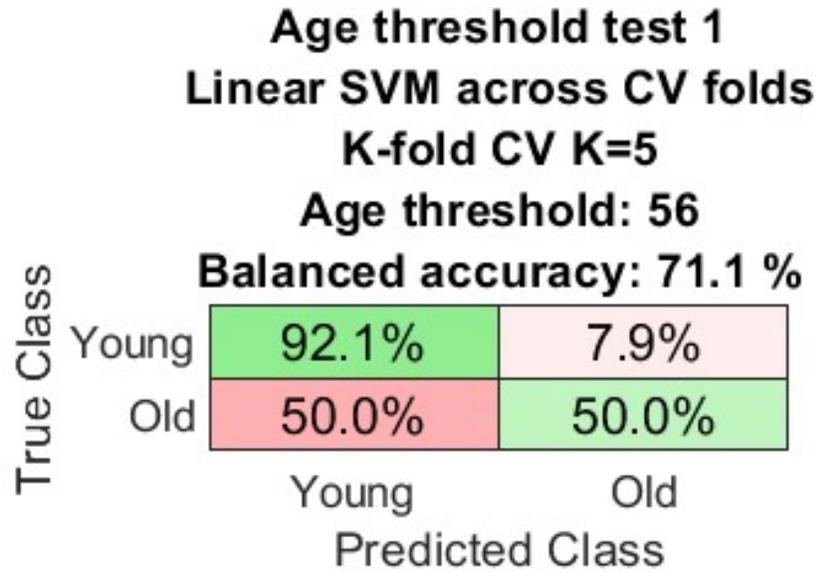
subjects in comparison to the other two groups (N_young: 30, N_rejection: 9 , N_old: 11). The uneven class distribution can result in case where when the classifier is in doubt it will favour the majority class. The more balanced case was found with linear SVM model with similar performance as seen in Figure 7 (D), where the age thresholds for rejection class of 27 – 57 yielded more balanced class distribution (N_young: 14, N_old: 12, N_rejection: 24), although still favouring one class over the others in the predictions.

The results of the age threshold test 2 and 3 present two interesting observations. Firstly, the classification performance is higher when the subjects in-between the young and old age group are excluded from the ML loop. Secondly, the upper age threshold is fairly constant (57, 62), while the lower age threshold has more variation (27, 33, 37, 44). The variation in the age threshold between the classifiers is expected, as the different classifiers perform differently in the tasks. The between tests variation with younger threshold with same classifiers could be explained by the effect caused by homogeneity in the data of the younger subjects, which is further supported by the results of the classification task using pre-defined age groups. These findings suggest a difference in the rate of aging between the subjects, and that the age-related differences are more distinctly pronounced as the aging process has progressed further.

TABLE 3 The performance of the classifiers in the age threshold test. The results consist of balanced accuracy of K-fold (K=5) cross-validation with the selected age thresholds. The results of the best performing models are bolded.

Test	Linear SVM	Polynomial SVM	Gaussian SVM	Random Forest
Test 1	71.05 % , th*: 56	66.51 %, th: 41	67.37 %, th: 42	63.44 %, th: 39
Test 2	81.37 %, ths: 37 - 57	81.50 % , ths: 33 - 62	79.09 %, ths: 33 - 62	80.13 %, ths: 27 - 62
Test 3	72.00 %, ths: 27 - 57	74.68 % , ths: 44 - 62	73.33 %, ths: 44 - 62	72.00 %, ths. 44 - 62

*age threshold



(A)

Age threshold test 2
Polynomial SVM across CV folds
K-fold CV, K=5
Ages excluded: 33 - 62
Balanced accuracy: 81.5 %

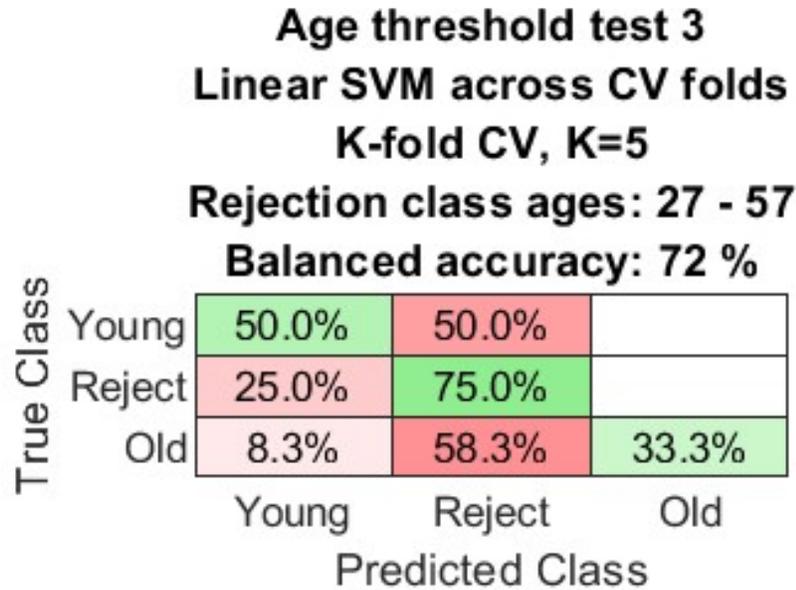
True Class	Young	93.8%	6.2%
	Old	30.8%	69.2%
		Young	Old
		Predicted Class	

(B)

Age threshold test 3
Polynomial SVM across CV folds
K-fold CV, K=5
Rejection class ages: 44 - 62
Balanced accuracy: 74.67 %

True Class	Young	93.3%	3.3%	3.3%
	Reject	88.9%	11.1%	
	Old	81.8%		18.2%
		Young	Reject	Old
		Predicted Class		

(C)



(D)

FIGURE 7 Normalized confusion matrices for the best performing models in the age threshold tests 1 (A), 2 (B) and 3 (C). Due to notable class imbalance with the age threshold 3 with the best performing model, the confusion matrix for the best linear SVM model shown additionally (D).

4 | CONCLUSION

The two key findings of the study are 1) fNIRS has potential to classify young and old adults to corresponding age groups with good performance, and 2) the inclusion and exclusion of the subjects in the young, in-between and old age groups affects the classification performance significantly, suggesting different aging rate within these three groups, i.e., young, in-between, and old groups. As the wavelengths used in our fNIRS device are sensitive to hemoglobin and water dynamics, the noticed changes are potentially related to the cerebrovascular and neurohydrodynamic events. However, it is not clear whether water or hemoglobin change caused effects are more significant. Furthermore, the fNIRS was measured from Fp1, and thus the detected changes are only related to the frontal lobe activity and to the hemodynamic and neurohydrodynamic differences in the brain. The effect of superficial layers in fNIRS study of brain aging would be of interest to study more in detail in the future, by utilizing short SD pair in addition to the long SD pair.

Several factors accelerating and decelerating brain aging process have been identified in the literature [2]. The existence of these factors could explain the observed changes in classification performance, when including the subjects in the in-between age groups. Although the study population was considered healthy, the brain aging related accelerative and decelerative factors information, such as alcohol consumption habits and amount of physical activity per week, were not considered in this study. The analysis of these factors would be intriguing, as they may delay or expedite the brain aging process. Furthermore, inclusion of individuals of age over 70 would be of interest as it has been suggested that brain aging accelerates significantly afterwards [36].

Another intriguing prospect is to use fNIRS in the aging study related to sleep. Recently it was found that the brain's glymphatic system has been found to be more active during the sleep [37] and the aging has been suggested to affect its function negatively [38]. As the system consists of CSF waste clearance, the use of fNIRS to measure the brain pulsations, which are one of the drivers of the glymphatic system [39], and CSF dynamics with water sensitive fNIRS setup could provide potential for gaining further insight into

the function of the system in humans, and the effects caused by aging. Furthermore, the EEG of sleep has been successfully utilized in the brain age prediction task. [28] As EEG measured neural activity and fNIRS measured hemodynamic response are connected by neurovascular coupling effect, it could be of interest to study if age-related differences become more pronounced during sleep.

The key limitation of our study is the used relatively small sample size, as in general the use of large datasets increases the statistical power in pattern recognition, [35] and therefore the used datasets in the brain age studies are usually of large size [40]. It would be of great interest in the future to conduct fNIRS brain aging studies with large, well age distributed, sample population to confirm our findings. Additionally, use of multi-channel fNIRS could give boarded picture on the global effects of the aging on brain. Although the used multi-modal measurement setup included different brain measurement modalities, such as magnetic encephalography (MREG) and EEG [21], in this study we focused on the analysis of fNIRS and brain aging specifically due to its novelty in the context. Thus, another area for future exploration in the fNIRS brain aging studies is to analyse combination of brain measurement modalities data, such as fNIRS combined with EEG or MREG. The analysis of different modalities combination would enhance the information content and provide a comparison of the methods performance in the brain aging context. However, although suffering from the lack of data, our study protocol utilized the best practices for ML research with limited data by utilizing nested cross-validation framework, with feature and model selection conducted within the inner loop, for estimating unbiased performance of the models [35]. Additionally, we utilized multiple fNIRS signal features, introduced previously in the study of brain by different measurement modalities. The findings of our study demonstrate the high potential for the use of fNIRS in the study of brain aging.

ACKNOWLEDGMENTS

The authors would like to acknowledge CSC – IT Center for Science, Finland, for providing the computational resources. The study was financed by the Academy of Finland Profi6 funding, 6G-Enabling Sustainable Society (6GESS) programme. Academy of Finland (Terva grant 335723), BF (11146/31/2022), JPND/AoF (357805).

FINANCIAL DISCLOSURE

The authors have no financial conflicts of interest in connection with this article.

CONFLICT OF INTEREST

The authors have no conflicts of interest in connection with this article.

DATA AVAILABILITY STATEMENT

Research data are not shared due to the hospital policy.

REFERENCES

- [1] Kontis, V., Bennett, J.E., Mathers, C.D., Li, G., Foreman, K., Ezzati, M., *The Lancet* , 2017, **389** , 1323.
- [2] Turrini, S., Wong, B., Eldaief, M., Press, D.Z., Sinclair, D.A., Koch, G., Avenanti, A., Santarnecchi, E., *Ageing Research Reviews* , 2023, **88** , 101939.
- [3] Higgins-Chen, A.T., Thrush, K.L., Levine, M.E., *Seminars in Cell and Developmental Biology* , 2021, **116** , 180.
- [4] Korhonen, V.O., Myllylä, T.S., Kirillin, M.Y., Popov, A.P., Bykov, A.V., Gorshkov, A.V., Sergeeva, E.A., Kinnunen, M., Kiviniemi, V., *IEEE Journal of Selected Topics in Quantum Electronics* , 2014,**20** , 289.
- [5] Ferrari, M., Quaresima, V., *NeuroImage* , 2012,**63** , 921.
- [6] Boas, D.A., Gaudette, T., Strangman, G., Cheng, X., Marota, J.J.A., Mandeville, J.B., *NeuroImage* , 2001, **13** , 76.

- [7] Myllylä, T., Harju, M., Korhonen, V., Bykov, A., Kiviniemi, V., Meglinski, I., *Journal of Biophotonics* , 2018, **11** , e201700123.
- [8] Liang, J., Huang, J., Luo, Z., Wu, Y., Zheng, L., Tang, Z., Li, W., Ou, H., *Frontiers in Human Neuroscience* , 2023, **17** .
- [9] Nguyen, T., Kim, M., Gwak, J., Lee, J.J., Choi, K.Y., Lee, K.H., Kim, J.G., *Journal of Biophotonics* , 2019, **12** .
- [10] Boutouyrie, P., Chowienczyk, P., Humphrey, J.D., Mitchell, G.F., *Circulation Research* , 2021, **128** , 864.
- [11] Ferdinando, H., Myllylä, T., in *Saratov Fall Meeting 2019: Optical and Nano-Technologies for Biology and Medicine* , SPIE, 2020, **11457** , 22.
- [12] Mohammadi, H., Peng, K., Kassab, A., Nigam, A., Bherer, L., Lesage, F., Joannette, Y., *Neurobiology of Aging* , 2021,**106** , 103.
- [13] Stojan, R., Mack, M., Bock, O., Voelcker-Rehage, C.,*NeuroImage* , 2023, **273** .
- [14] Huang, W., Li, X., Xie, H., Qiao, T., Zheng, Y., Su, L., Tang, Z.-M., Dou, Z., *Frontiers in Aging Neuroscience* , 2022,**14** .
- [15] Nóbrega-Sousa, P., Gobbi, L.T.B., Orcioli-Silva, D., Conceição, N.R.D., Beretta, V.S., Vitório, R., *Neurorehabilitation and Neural Repair* , 2020, **34** , 915.
- [16] Lucas, M., Wagshul, M.E., Izzetoglu, M., Holtzer, R.,*Journals of Gerontology - Series A Biological Sciences and Medical Sciences* , 2019, **74** , 435.
- [17] Zeller, J.B.M., Katzorke, A., Müller, L.D., Breunig, J., Haeussinger, F.B., Deckert, J., Warrings, B., Lauer, M., Polak, T., Herrmann, M.J., *Brain Imaging and Behavior* , 2019, **13** , 283.
- [18] Ferdinando, H., Moradi, S., Korhonen, V., Helakari, H., Kiviniemi, V., Myllylä, T., *The European Physical Journal Special Topics* , 2022, **232** , 655.
- [19] Kober, S.E., Spörk, R., Bauernfeind, G., Wood, G.,*Neurobiology of Aging* , 2019, **81** , 127.
- [20] Iivesmäki, M., Assessment of human brain ageing by functional near-infrared spectroscopy. University of Oulu, 2023.
- [21] Korhonen, V., Hiltunen, T., Myllylä, T., Wang, X., Kantola, J., Nikkinen, J., Zang, Y.-F., LeVan, P., Kiviniemi, V., *Brain Connectivity* , 2014, **4** , 677.
- [22] H.s.s, S., Myllylä, T.S., Kirillin, M.Y., Sergeeva, E.A., Myllylä, R.A., Elseoud, A.A., Nikkinen, J., Tervonen, O., Kiviniemi, V.,*Quantum Electronics* , 2010, **40** , 1067.
- [23] Lieslehto, J., Jääskeläinen, E., Kiviniemi, V., Haapea, M., Jones, P.B., Murray, G.K., Veijola, J., Dannowski, U., Grotegerd, D., Meinert, S., Hahn, T., Ruef, A., Isohanni, M., Falkai, P., Miettunen, J., Dwyer, D.B., Koutsouleris, N., *npj Schizophrenia* , 2021,**7** , 1.
- [24] Cole, J.H., Leech, R., Sharp, D.J., Initiative, for the A.D.N.,*Annals of Neurology* , 2015, **77** , 571.
- [25] Löwe, L.C., Gaser, C., Franke, K., Initiative, for the A.D.N.,*PLOS ONE* , 2016, **11** , e0157514.
- [26] Cope, M., The Application of Near Infrared Spectroscopy to Non Invasive Monitoring of Cerebral Oxygenation in the Newborn Infant. University College London, 1991.
- [27] Helakari, H., Kananen, J., Huotari, N., Raitamaa, L., Tuovinen, T., Borchardt, V., Rasila, A., Raatikainen, V., Starck, T., Hautaniemi, T., Myllylä, T., Tervonen, O., Rytty, S., Keinänen, T., Korhonen, V., Kiviniemi, V., Ansakorpi, H., *NeuroImage: Clinical* , 2019,**22** , 101763.

- [28] Sun, H., Paixao, L., Oliva, J.T., Goparaju, B., Carvalho, D.Z., van Leeuwen, K.G., Akeju, O., Thomas, R.J., Cash, S.S., Bianchi, M.T., Westover, M.B., *Neurobiology of Aging* , 2019, **74** , 112.
- [29] Ferdinando, H., Moradi, S., Korhonen, V., Kiviniemi, V., Myllylä, T., in *Neural Imaging and Sensing 2023* , SPIE, 2023,**12365** , 18.
- [30] Lobier, M., Siebenhühner, F., Palva, S., Palva, J.M.,*NeuroImage* , 2014, **85** , 853.
- [31] Matteo, F., Arjan, H., Phase Transfer Entropy in Matlab. figshare, 2017.
- [32] Richman, J.S., Moorman, J.R., *American Journal of Physiology-Heart and Circulatory Physiology* , 2000, **278** , H2039.
- [33] Flood, M.W., Grimm, B., *PLOS ONE* , 2021, **16** , e0259448.
- [34] Borhardt, V., Korhonen, V., Helakari, H., Nedergaard, M., Myllylä, T., Kiviniemi, V., *The European Physical Journal Plus* , 2021, **136** , 497.
- [35] Vabalas, A., Gowen, E., Poliakoff, E., Casson, A.J., *PLOS ONE* , 2019, **14** , e0224365.
- [36] Scahill, R.I., Frost, C., Jenkins, R., Whitwell, J.L., Rossor, M.N., Fox, N.C., *Archives of Neurology* , 2003, **60** , 989.
- [37] Xie, L., Kang, H., Xu, Q., Chen, M.J., Liao, Y., Thiyagarajan, M., O'Donnell, J., Christensen, D.J., Nicholson, C., Iliff, J.J., Takano, T., Deane, R., Nedergaard, M., *Science* , 2013,**342** , 373.
- [38] Kress, B.T., Iliff, J.J., Xia, M., Wang, M., Wei, H.S., Zeppenfeld, D., Xie, L., Kang, H., Xu, Q., Liew, J.A., Plog, B.A., Ding, F., Deane, R., Nedergaard, M., *Annals of Neurology* , 2014,**76** , 845.
- [39] Bohr, T., Hjorth, P.G., Holst, S.C., Hrabětová, S., Kiviniemi, V., Lilius, T., Lundgaard, I., Mardal, K.-A., Martens, E.A., Mori, Y., Nägerl, U.V., Nicholson, C., Tannenbaum, A., Thomas, J.H., Tithof, J., Benveniste, H., Iliff, J.J., Kelley, D.H., Nedergaard, M.,*iScience* , 2022, **25** .
- [40] Cole, J.H., Franke, K., *Trends in Neurosciences* , 2017,**40** , 681.