Metabolic connectivity-based single subject classification by multi-regional linear approximation in the rat ☆

Maximilian Grosch [a,b,*], Leonie Beyer [b], Magdalena Lindner [a,b], Lena Kaiser [b], Seyed-Ahmad Ahmadi [a], Anna Stockbauer [b], Peter Bartenstein [b,d], Marianne Dieterich [a,c,d], Matthias Brendel [b,d], Andreas Zwergal [a,c,1], Sibylle Ziegler [b,1]

[a] German Center for Vertigo and Balance Disorders, DSGZ, University Hospital, Ludwig-Maximilians-University Munich, Marchioninistrasse 15, D-81377 Munich, Germany
[b] Department of Nuclear Medicine, University Hospital, LMU Munich, Munich Germany
[c] Department of Neurology, University Hospital, LMU Munich, Munich, Germany
[d] Munich Cluster of Systems Neurology, SyNergy, Munich, Germany

ARTICLE INFO

ABSTRACT

Metabolic connectivity patterns on the basis of [^{18}F]-FDG positron emission tomography (PET) are used to depict complex cerebral network alterations in different neurological disorders and therefore may have the potential to support diagnostic decisions. In this study, we established a novel statistical classification method taking advantage of differential time-dependent states of whole-brain metabolic connectivity following unilateral labyrinthectomy (UL) in the rat and explored its classification accuracy.

The dataset consisted of repeated [^{18}F]-FDG PET measurements at baseline and 1, 3, 7, and 15 days (= maximum of 5 classes) after UL with 17 rats per measurement day. Classification in different stages after UL was performed by determining connectivity patterns for the different classes by Pearson's correlation between uptake values in atlas-based segmented brain regions. Connections were fitted with a linear function, with which different thresholds on the correlation coefficient (r = [0.5, 0.85]) were investigated. Rats were classified by determining the congruence of their PET uptake pattern with the fitted connectivity patterns in the classes.

Overall, the classification accuracy with this method was 84.3% for 3 classes, 75.0% for 4 classes, and 54.1% for 5 classes and outperformed random classification as well as machine learning classification on the same dataset. The optimal classification thresholds of the correlation coefficient and distance-to-fit were found to be |r| > 0.65 and d = 4 when using Siegel's slope estimator for fitting.

This connectivity-based classification method can compete with machine learning classification and may have methodological advantages when applied to support PET-based diagnostic decisions in neurological network disorders (such as neurodegenerative syndromes).

1. Introduction

Machine-supported diagnostic evaluation procedures are receiving increasing attention in all fields of medicine and have already been used in various imaging techniques (Cai et al., 2020). Especially in three-dimensional imaging, computer-based diagnostic approaches are capable of analyzing whole volume datasets and therefore promise to enhance the limited visual evaluation on two-dimensional slices used in clinical practice, especially in neurological use cases (Klöppel et al., 2012). Another interesting application may be the quantification of visually inaccessible interregional brain connectivity network alterations in different neurological diseases. Recent studies suggested that neuro-

logical diseases, such as dementia or neurodegenerative movement disorders have their origin and associated functional impairment not in distinct brain regions but in a network of connected brain regions created by complex interactions (Huber et al., 2020; Jalilianhasanpour et al., 2019). Dynamic whole-brain connectivity changes also appear during adaptive brain plasticity following peripheral and central nervous system lesions (Grosch et al., 2020). Most studies investigating network connectivity used functional magnetic resonance imaging (fMRI) (Galvin et al., 2011; Kenny et al., 2012; Peraza et al., 2016) and positron emission tomography (PET) (Grosch et al., 2020; Sanabria-Diaz et al., 2013) for data acquisition. Whereas fMRI connectivity patterns can in principle be determined on a single subject level by measuring time series, metabolic connectivity information based on PET tracer measurements is often derived in a population-based approach, as already shown for different pathologies such as Alzheimer's disease (Morbelli et al., 2013), Parkinson's disease (Ge et al., 2020), or Lewy body dementia (Huber et al., 2020). Population-based metabolic connectivity patterns can be determined by Pearson's correlation or sparse inverse covariance estimation (SICE) (Grosch et al., 2020; Huang et al., 2010). A recent study suggested Kullback-Leibler divergence similarity estimation (KLSE) as a method to derive single subject connectivity (Wang et al., 2020). Generally, metabolic connectivity patterns describe large-scale cerebral networks and therefore, they could be used for image-based classification of pathological or functional states. SICE and KLSE were investigated with respect to their applicability to classification tasks (Huang et al., 2010; Wang et al., 2020). However, to our knowledge, no metabolic connectivity classification relying on Pearson's correlation has been described. On the other hand, different approaches implementing machine learning classification based on image-derived features were reported (Liu et al., 2015; Liu et al., 2018). In this study, we derived the cerebral connectome from [^{18}F]-FDG PET by Pearson's correlation and evaluated a novel statistical classification method, which could be capable of classifying single subjects by their regional cerebral glucose metabolism (rCGM). For benchmarking, we compared its performance to state-of-the-art machine learning classifiers. Therefore, we used an established rat model of unilateral labyrinthectomy, which has been shown to induce a process of central vestibular compensation characterized by different stages of cerebral metabolic connectivity over time to define distinct classes (Grosch et al., 2020). Single subjects were classified by testing the congruence of their connectivity patterns to those of the different classes. It was hypothesized that the incorporation of connective information into classification could yield higher accuracies than classifying on univariate PET uptake values.

2. METHODS

2.1. Experimental setup

The dataset used in this study was taken from two previous studies (Lindner et al., 2017; Lindner et al., 2019) investigating the time course of vestibular compensation after unilateral labyrinthectomy (UL) in the rat. Data included five PET measurements, at baseline and at days 1, 3, 7, and 15 after surgery. These time points were defined as five classes. The experimental procedures, protocols, and environment were identical and only animals without specific drug treatment were used. Measurements with incorrect tracer application, misplacement of the rat, or image artifacts were discarded, resulting in 17 successful imaged animals per time point. Male Sprague-Dawley rats (mean 400 +- 20 g, age 3 months at time of UL, Charles River Ltd, UK), were housed two animals per cage and granted free access to food and water. The housing room was temperature- and humidity-controlled with a 12 h light/dark cycle. Female rats exhibit cycle-dependent changes in cerebral activity and were therefore not investigated in this study (Barth et al., 2015; Strauss, 2020).

All animal experiments were approved by the government of Upper Bavaria and performed in accordance with the guidelines for the use of living animals in scientific studies and the German Law for the Protection of Animals.

The rats received a left-sided chemical UL by injection of bupivacaine and p-arsanilic acid inducing an acute vestibular syndrome as described earlier (Anniko and Wersall, 1977; Beck et al., 2014; Magnusson et al., 2002; Vignaux et al., 2012; Zwergal et al., 2017). Animals were anesthetized with 1.5% isoflurane in O$_2$ delivered up to 1.2 l/min O$_2$ via a mask. For surgical analgesia, 1.5 mg/kg meloxicam was injected s.c. before and 3 days after surgery. An additional 5 ml saline was injected s.c. as a bolus. After local anesthesia with 1% bupivacaine hydrochloride, a left paramedian incision was made to expose the lamboidal ridge and the external ear canal. The external ear canal was opened just anterior to the exit point of the facial nerve. With a 26-gauge needle, the tympanic membrane was perforated caudally to the hammer shaft, and about 0.150 ml of a 20% bupivacaine solution was instilled into the tympanic cavity. For about 2 min the bupivacaine solution was aspirated and instilled slowly again multiple times. The same procedure was followed to instill 0.150 ml of a 10% solution of p-arsanilic acid, which irreversibly desensitized the primary sensory cells of the inner ear (Vignaux et al., 2012). After the last thorough aspiration, the wound was closed by skin suture, and for preventive antibiosis, 2 mg/kg marbofloxacin was injected s.c. for 3 days.

2.2. PET imaging

Sequential [^{18}F]-FDG PET images of the rats were acquired at baseline and 1, 3, 7, and 15 days post UL. Anesthesia was induced with isoflurane (as described above) and a cannula was placed in a lateral tail vein for [^{18}F]-FDG bolus injection (in 0.5 ml saline). Then, the animals were awakened and were allowed to move freely for an uptake period of 30 min until anesthesia was induced again with isoflurane (1.8%) for the PET scan. Two animals per scan were positioned in the Siemens Inveon PET scanner (Siemens Healthineers, Erlangen, Germany) and were kept warm with a heating pad. In order to avoid head movement, the head position was fixed using a custom-made head-holder. Emission data were recorded for 30 min followed by a 7-min transmission scan using a rotating [^{57}Co] point source. Upon recovery from anesthesia, the rats were returned to their home cages (Beck et al., 2014; Lindner et al., 2019).

2.3. Image processing

Emission recordings were reconstructed using an Ordered Subsets Expectation Maximization (OSEM-3D) algorithm with decay correction, scatter correction, attenuation correction, dead time correction, and sensitivity normalization (Siemens Healthineers, Erlangen, Germany) yielding a 128 × 128 × 159 image matrix (Fig. 1a). For attenuation correction, the corresponding transmission measurement at the end of the emission scan was used. The voxel dimensions of the reconstructed images were 0.78 × 0.78 × 0.80 mm^3 (Zwergal et al., 2016). Radioactivity distribution in the reconstructed images was used as a surrogate of the regional cerebral glucose metabolism (rCGM). Cropping and a template-based rigid coregistration into Px Rat (W. Schiffer) atlas space was performed (Schiffer et al., 2006) using PMOD medical image analysis software (PMOD Technologies LLC, RRID: SCR_016547, v4.004) (Fig. 1b). An isotropic Gaussian filter with 0.8 mm FWHM and whole-brain normalization was applied in order to achieve comparability between all images. Images were segmented into 57 brain regions using W. Schiffer's rat brain atlas (Fig. 1c). The brain blood flow region of this atlas depicts a brain vessel and was therefore not included for further processing. Mean normalized activity values were extracted for each brain region using PyRadiomics (Van Griethuysen et al., 2017) Python package (Fig. 1d) and the dataset tested for normal distribution with Jarque-Bera test to proof applicability of Pearson's correlation.
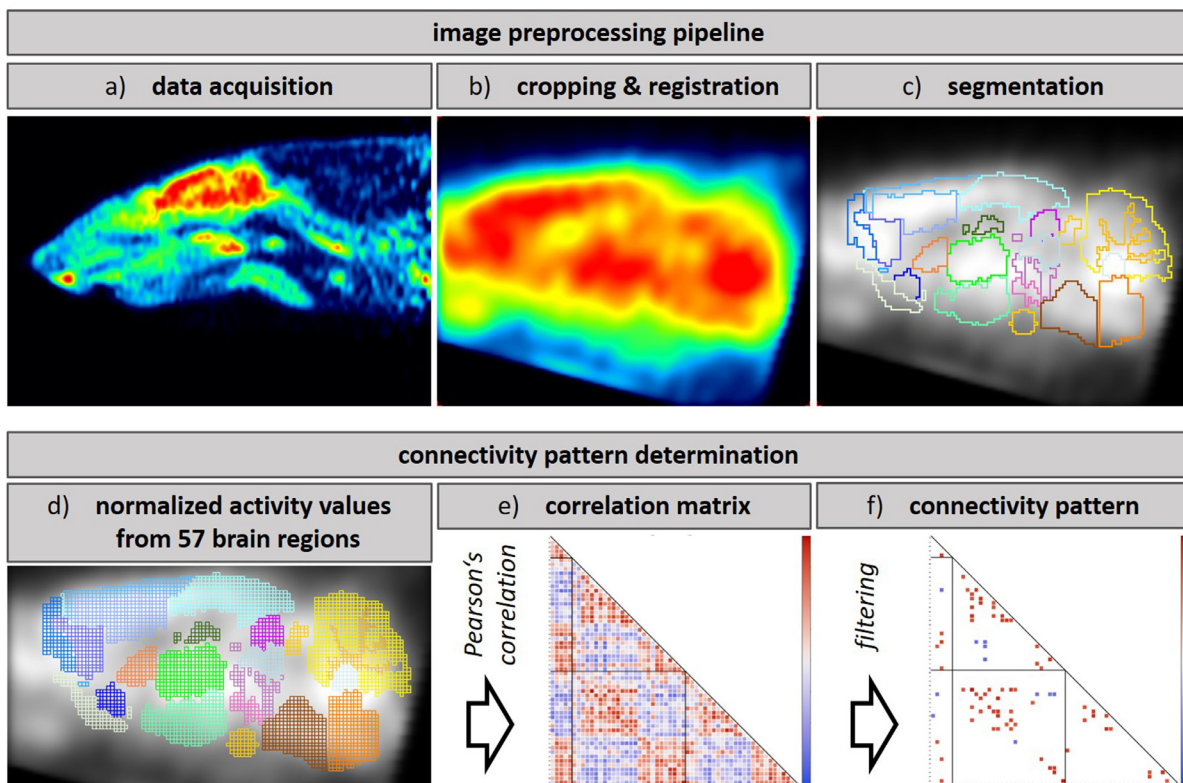
Fig. 1. Data processing. a-c) Images were acquired, reconstructed, cropped, registered, filtered, normalized, and segmented. d-f) Normalized mean activity values for 57 brain regions were extracted, a correlation matrix calculated using Pearson's correlation, and the results filtered by r- and p-thresholds or sparsity to determine the class-specific connectome (adapted from Grosch et al. 2020).

Table 1

Parameter sets for thresholds of correlation coefficient r, connection density s (sparsity), linear fitting algorithms, and thresholds on the single subject distance d from fitted functions for classification.

Parameter	Options
r-thresholds	|0.5|, |0.55|, |0.6|, |0.65|, |0.7|, |0.75|, |0.8|, |0.85|
s-thresholds	0.01, 0.02, 0.03, 0.04, 0.05, 0.06, 0.07, 0.08, 0.09, 0.1, 0.15, 0.2, 0.25
linear fit	least-square / Sen's slope / Siegel's slope
d-thresholds	0.5, 1.0, 1.5, 2.0, 2.5, 3.0, 3.5, 4.0, 4.5, 5.0,5.5, 6.0, 6.5, 7.0, 7.5, 8.0, 8.5, 9.0, 9.5, 10.0,10.5, 11.0, 11.5, 12.0, 12.5, 13.0, 13.5, 14.0, 14.5, 15.0

2.4. Connectivity pattern-based classification

Metabolic brain connectivity patterns can be derived with various methods (Yakushev et al., 2017). In this work, they were determined in a population-based approach by pairwise correlation of normalized mean uptake values in 57 segmented brain regions with Pearson's correlation as described earlier (Grosch et al., 2020), resulting in $57 \times 56 / 2 = 1596$ individual correlation pairs stored in a correlation matrix (Fig. 1e). The correlation matrix was filtered by two different approaches: by applying a threshold to the correlation coefficient r and second by applying a density s (sparsity) threshold (Tsai, 2018). Subsequently, the entries left in the correlation matrix were filtered according to their statistical significance level p. For this purpose, a false discovery rate (FDR) correction under dependency was applied to correct the initial α level of 0.05 for multiple comparisons (Benjamini and Yekutieli, 2001). All correlations meeting the requirements were considered as connections (Fig. 1f) in the classification procedure described below.

2.4.1. r- and s-thresholds

The classification performance was evaluated using different threshold values on the correlation coefficient r and the connection density s as indicated in Table 1. For the r-threshold, the same tests were performed without taking the absolute values in the r-thresholds, thus just allowing positive correlations to be recognized as connections. On the other hand, the sparsity threshold s permits a percentage of the total number of correlations to be seen as connected, whereby the correlations with the highest absolute value of the correlation coefficient were used preferentially. As sparsity generally does not set requirements on the correlation coefficient, an r-threshold of r > |0.5| was applied additionally to ensure reasonability of the following linear fitting.

2.4.2. Linear fitting

Pearson's correlation coefficient r is a measure of the linear relationship between the uptake values from two different brain regions (Benesty et al., 2009). All connected pairs of brain regions were fitted with a linear function as they meet the applied thresholds on the correlation coefficient (Fig. 2a). Fitting was performed with ordinary least-square approximation (LS) and two algorithms more robust to outliers in the datasets, Sen's slope estimator (Sen) (Sen, 1968) and Siegel's slope estimator (Siegel) (Siegel, 1982). The results were tested for independent and identical distributed and normal distributed residual errors with the Durbin-Watson test and the Jarque-Bera test, respectively. The classification performance was evaluated for all three algorithms by determining the accuracy and the area under the receiver operating characteristic curve (ROC-AUC). The latter was determined as one-versus-rest macro average to enable multi class evaluation.
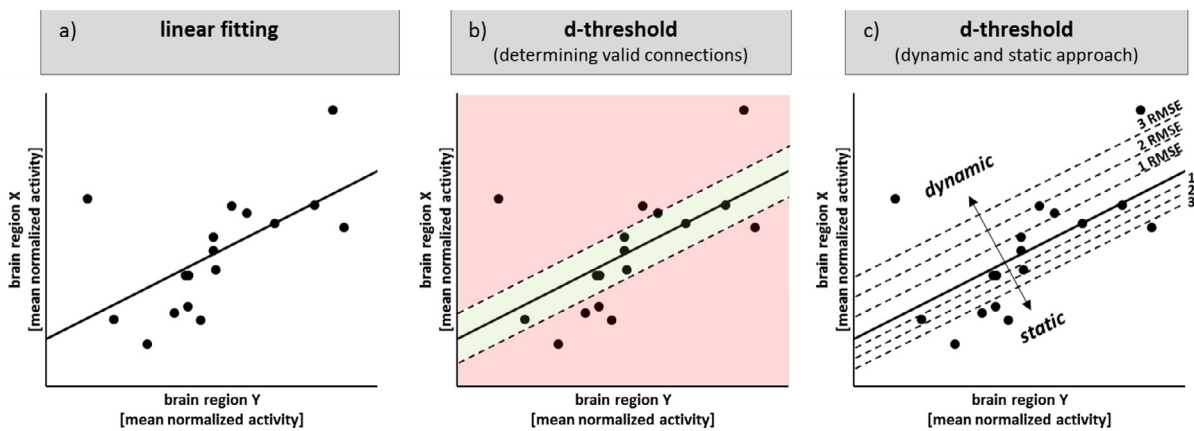
Fig. 2. Examining connectivity on single subject level by evaluating the distance from a linear fit in connections determined in a population based approach. a) Fitting a linear function to the mean normalized activity values from two correlating brain regions. b) Check validity of a connection by applying a distance d-threshold. c) Threshold could be static or dynamic, whereby the dynamic threshold depends on the root-mean-square error (RMSE) of the underlying distribution.

2.4.3. Classification procedure

Pattern-based classification was performed by comparing the uptake values in segmented brain regions of a single subject to the connectivity patterns of all available classes. Therefore, different sets of linear functions were fitted according to the connectivity patterns in the classes as described above. Subsequently, a new subject was represented by points in the coordinate systems of the fits and its Euclidean distances d from the linear functions were determined. In order to extract the matching connections on a single-subject level, these distances were filtered by a threshold (Fig. 2b). The subject was classified by evaluating the percentage of valid connections in each class, where the class with the highest percentage of matching connections was assigned to the subject.

2.4.4. d-thresholds

The classification performance was investigated by two approaches to the distance-to-fit threshold d, first by using static thresholds and second by using dynamic thresholds, depending on the root-mean-square error (RMSE) of the fits. The RMSE is an error measure describing the precision of a fitted function to the underlying data, and thus is unique for each fit. The static d-thresholds are indicated in table 1 and dynamic d-thresholds were achieved by multiplying the static d-thresholds with the individual RMSE of each fit (Fig. 2c).

2.4.5. Classification performance

Each measurement day represents a class with 17 subjects per class. We investigated the algorithm's capability to classify into three classes (baseline, days 3 and 15), into four classes (baseline, days 1, 3, and 15), and into five classes (baseline, days 1, 3, 7, and 15). Classification performance was measured in a leave-one-out manner by fitting the linear functions in each class on 16 subjects and using the omitted subjects (one from each class) to test the classification results. The fitting and testing were repeated 17 times and the overall classification accuracy was reported. The theoretical accuracies when classifying by chance were 33%, 25%, and 20% for the 3-class, 4-class, and 5-class classification, respectively.

2.5. Machine learning classification

Another approach to image classification is provided by machine learning methods. In this study we used a support vector machine (SVM) and a random forest classifier (RFC) to benchmark our connectivity-based classification results, which were set up with SciKit-learn (Pedregosa et al., 2011) in Python. SVM and RFC are well-established supervised learning classifiers and popular in the field of

brain disorders (Du et al., 2018; Sarica et al., 2017). Supervised learning algorithms take labelled samples as an input for their learning routine and later aim to find the correct label on previously unseen samples (Pisner and Schnyer, 2020). In this study, the subjects are the samples and represented by the mean normalized activity values in 57 different brain regions, the so-called features. The labels are the measurement days corresponding to the images. Linear, polynomial, radial-basis-function (RBF) and sigmoidal kernels were investigated for the SVM. RFC hyperparameters ranged from 50 to 500 trees and maximal depth from 10 to unlimited.

Both classifiers were implemented in a stratified 5-fold nested cross-validation with univariate select k-best feature selection to reduce the 57 features to 10. Hyperparameter tuning was performed with randomized search in 100 iterations and scoring was based on accuracy or ROC-AUC. The reported accuracy was calculated as the mean accuracy from the five outer folds and the ROC-AUC as one-versus-rest macro average over the five classes accordingly.

2.6. Data availability

Data and code reported in this article will be shared with any appropriately qualified investigator upon request.

3. Results

The classification performance was evaluated with different approaches to reduce the number of connections used for classification, namely different thresholds on the correlation coefficient r and the numeric sparsity s. The classification accuracies and ROC-AUCs were determined for different values of each threshold individually. Furthermore, the impact of the distance-to-fit threshold d was investigated with static and dynamic (RMSE-dependent) values. Here, two different connectivity filter setups were used: First, applying an r-threshold of r > 0.65 (r-filtered) and second applying r- and s-thresholds of r > 0.5 and s = 0.05 (s-filtered). All tests were performed with three different linear fitting algorithms (Least square fitting / Sen's slope estimator / Siegel's slope estimator) and with varying complexity with regard to the number of classes (3, 4, and 5 classes).

For 3-class classification, the classification accuracy with different thresholds of the correlation coefficient r reached 74.5% (LS), 70.6% (Sen), and 74.5% (Siegel) and deteriorated for lower and higher thresholds (Fig. 3a). Rejecting negative correlations resulted in maximum classification accuracies of 72.5% (LS), 68.6% (Sen), and 76.5% (Siegel) (Fig. 3b). Reducing the number of connections via sparsity thresholds resulted in classification accuracies of up to 72.5% (LS), 70.6% (Sen), and 76.5% (Siegel) (Fig. 3c).
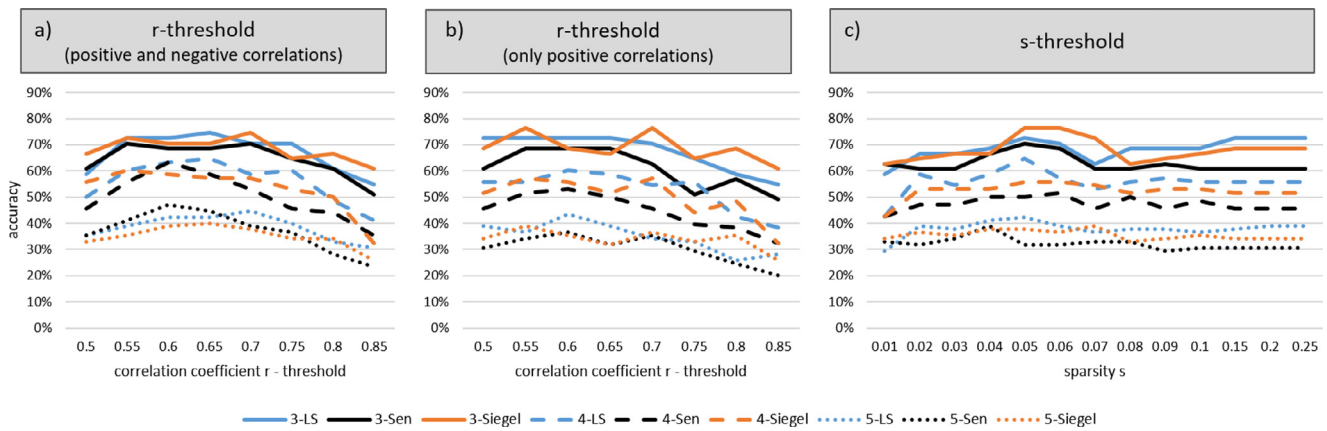
Fig. 3. Classification accuracy in dependence on the correlation filter thresholds for three, four, and five classes and with different fitting algorithms. a) Absolute value of the correlation coefficient r with static d = 2. b) Signum of the correlation coefficient r with static d = 2. c) Percentage sparsity of the total number of correlation pairs with r > 0.5 and static d = 2. Three-class, four-class, and five-class classification are indicated as solid line, dashed line, and dotted line, respectively. Least-square fitting (LS), Sen's slope estimator (Sen), and Siegel's slope estimator (Siegel) are colored in blue, black, and orange, respectively.
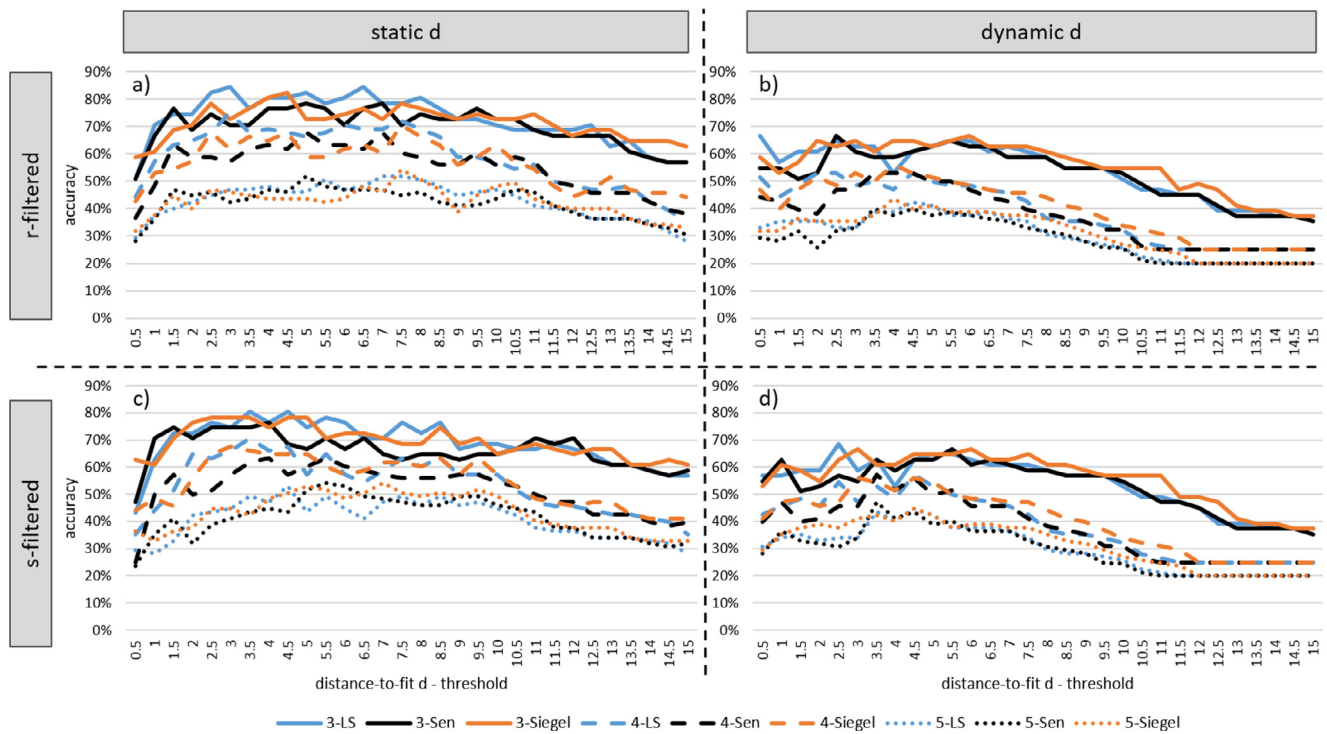
Fig. 4. Classification accuracy in dependence on the distance-to-fit threshold d for three, four, and five classes and with different fitting algorithms. a-b) static and dynamic d with r-filter of r > 0.65, c-d) static and dynamic d with s-filter of r > 0.5 and s = 0.05. Three-class, four-class, and five-class classification are indicated as solid line, dashed line, and dotted line, respectively. Least-square fitting (LS), Sen's slope estimator (Sen), and Siegel's slope estimator (Siegel) are colored in blue, black, and orange, respectively.

Moreover, the classification accuracy with respect to the distance-to-fit d reached 84.3% (LS), 78.4% (Sen), and 82.4% (Siegel) when using static thresholds and r-filtering (Fig. 4a). With dynamic d thresholds and r-filtering, accuracies of 66.7% (LS), 66.7% (Sen), and 66.7% (Siegel) were achieved (Fig. 4b). Using s-filtered connectivity patterns resulted in classification accuracies of up to 80.4% (LS), 76.5% (Sen), and 78.4% (Siegel) for static d thresholds (Fig. 4c) and 68.6% (LS), 66.7% (Sen), and 66.7% (Siegel) for dynamic d thresholds (Fig. 4d). For r-filtering and static d-thresholds, the ROC-AUC values showed a course related to those of the corresponding accuracies and reached 0.883 (LS), 0.871 (Sen), and 0.889 (Siegel) (Fig. 5a). The ROC curves for the best performing parameters (|r| > 0.65 and static d = 4) were plotted in Fig. 5b. As depicted in Figs. 3, 4, and 5, classification with 4 and 5 classes mirrored

the results of 3-class classification at a lower level (see appendix A for details).

As a benchmark, the classification accuracy and the ROC-AUC of the SVM and the RFC were evaluated on the identical input features. The results are shown in Table 2.

4. Discussion

In this study, we investigated a novel, [18F]-FDG PET-based classification approach relying on class differences in metabolic brain connectivity patterns. Classes were defined as compensation stages after unilateral labyrinthectomy in the rat. The main findings were: 1) Single-subject classification based on metabolic connectivity patterns deter-
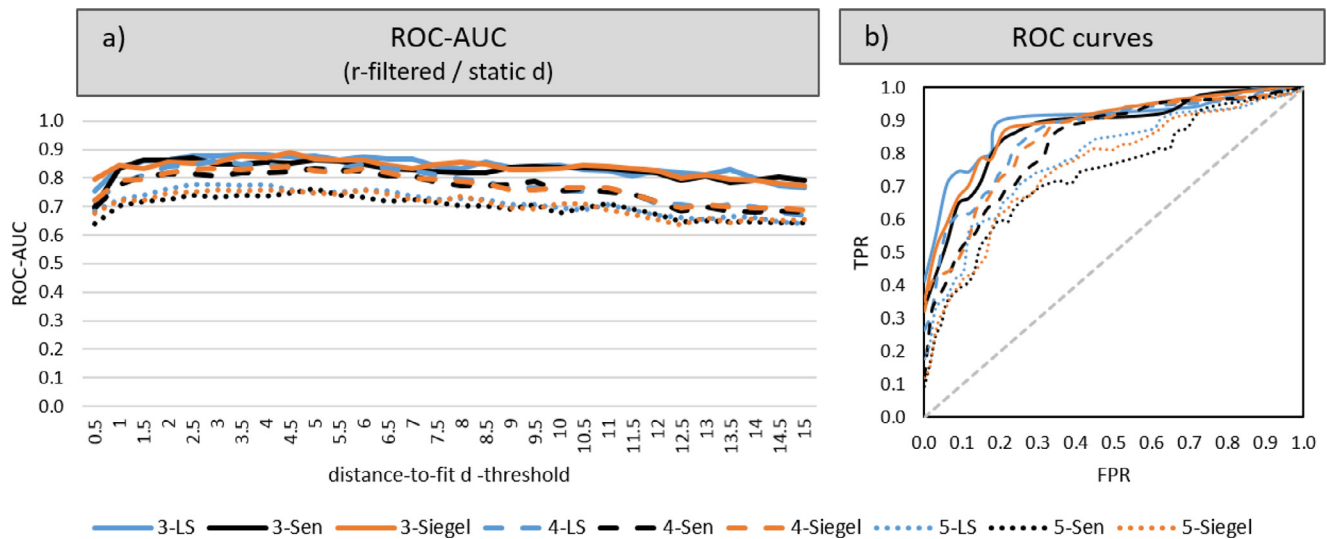
Fig. 5. Receiver operating characteristic (ROC). a) Area under the ROC curve (ROC-AUC) in dependence on the static distance-to-fit threshold d with |r| > 0.65. b) ROC curves for the best performing parameters |r| > 0.65 and static d = 4 (FPR = false positive rate, TPR = true positive rate). Three-class, four-class, and five-class classification are indicated as solid line, dashed line, and dotted line, respectively. Least-square fitting (LS), Sen's slope estimator (Sen), and Siegel's slope estimator (Siegel) are colored in blue, black, and orange, respectively.

Table 2
Classification performance measured as accuracy and area under the receiver operating characteristic curve (ROC-AUC) with different kernels of a support vector machine (SVM) and a random forest classifier (RFC).

Classifier	3 classes		4 classes		5 classes	
	Accuracy	ROC-AUC	Accuracy	ROC-AUC	Accuracy	ROC-AUC
SVM (linear)	76.5%	0.915	65.9%	0.839	49.4%	0.786
SVM (polynomial)	70.9%	0.875	66.2%	0.851	50.6%	0.778
SVM (RBF)	74.5%	0.935	69.0%	0.861	51.8%	0.795
SVM (sigmoidal)	74.5%	0.916	65.9%	0.847	54.1%	0.773
RFC	74.4%	0.891	64.8%	0.847	50.6%	0.792

mined with Pearson's correlation is feasible. 2) Taking negative correlation coefficients into account yields higher classification accuracies. 3) Robust linear fitting with Siegel's slope estimator results in overall higher classification accuracies than using least-square approximation or robust Sen's slope estimator. 4) Static d-thresholds reach higher classification accuracies than dynamic d-thresholds. 5) Metabolic brain connectivity-based classification can outperform state-of-the art machine learning methods.

4.1. Optimization of classification parameters

Various connectivity-based classification approaches have been reported in the field of fMRI (Anderson et al., 2011; Du et al., 2018; Nielsen et al., 2013). fMRI provides multiple images from the same patient by measuring time series, which can be used to determine connectivity patterns on a single-subject level by Pearson's correlation (Liu et al., 2015). In molecular imaging, especially in PET, classification approaches relying on sparse inverse covariance estimation (SICE) or Kullback-Leibler divergence similarity estimation (KLSE) have been investigated on Alzheimer's disease datasets (Huang et al., 2010; Wang et al., 2020). Another common method to analyze brain connectivity patterns is to apply a population-based approach with static PET images and Pearson's correlation (Grosch et al., 2020; Sanabria-Diaz et al., 2013). In comparison to the usually reported population-based approaches to metabolic brain connectivity, KLSE provides the opportunity to determine connectivity patterns on a single-subject level without the assumption that brain regions have linearly correlating uptake patterns, which on the other hand is the basis for the proposed

classification method. As Pearson's correlation coefficient is a measure of the linear correlation of two variables, fitting a linear function to the normalized mean uptake values of pairs of brain regions is reasonable as long as sufficiently strong requirements on the correlation coefficient and the corresponding statistical significance are met (Benesty et al., 2009). In this study, the application of different r thresholds resulted in an approximately concave-shaped curve of classification accuracies, which declined for higher r-thresholds as the number of connections dropped and for lower thresholds as the linearity of the used connections decreased (Fig. 3a, 3b). This effect was visible for 3-class, 4-class, and 5-class classifications, which indicates robustness of the method for varying r-thresholds while statistical significance of the correlation is given. Furthermore, taking negative correlations into account reduced noise in the classification accuracies as the number of connections used for classification has an important impact on the classifier performance. Nevertheless, connections without sufficient statistical significance were filtered by their p-value after FDR correction for multiple comparisons to exclude random effects. We performed further investigations on distance-to-fit thresholds d with |r| > 0.65. On the other hand, sparsity filtering (Fig. 3c) yielded constant accuracies for higher values as the number of connections is not determined by the sparsity anymore, but by the additionally applied filters on the correlation coefficient and the statistical significance. The decrease of accuracies for stricter s-thresholds is again related to the lower number of connections included for classification, indicating that a reasonable number of connections needs to be contained in the connectivity pattern in order to reach higher classification performance independent of the filtering method. In the end, the choice of the thresholds needs to be a trade-

off between sufficiently high requirements on the correlation coefficient and the number of connections passing the filter. These settings depend on the dataset (number of samples, number of brain regions) and have to be tuned for every application individually.

The accuracy in relation to the distance-to-fit threshold d revealed a roughly concave course over the observed parameter range in all investigated conditions (Fig. 4), whereby the peak positions were different, indicating the necessity of tuning this threshold to an optimum for every application and every dataset. As seen in Fig. 4a and 4c, the accuracies dropped for very low d-thresholds rapidly because hardly any connections were determined as valid on the single-subject level. For higher d-thresholds the accuracies decreased steadily because randomness of valid connection increased. The r-filtered connectivity patterns resulted in an overall higher maximum performance for all fitting algorithms than using sparsity filtering. The overall accuracy using dynamic d-thresholds (Fig. 4b, d) was lower than that using static values; nevertheless dynamic thresholding outperformed classification by chance significantly (33% in 3-classes, 25% in 4-classes, and 20% in 5-classes). As uptake values were normalized to the whole brain prior to classification, a static threshold is reasonable and avoids potential uncertainties in the RMSE-based threshold induced by outliers. When comparing the three linear fitting algorithms, Siegel's slope estimator and least-square fitting on average performed better than Sen's slope estimator. As robustness to outliers might be less important in highly controlled preclinical experimental setups, we assume that robust fitting performs better than least-square fitting in clinical datasets.

We state that in this dataset r-filtering with $|r| > 0.65$ in combination with a static d-threshold of $d = 4$ and Siegel's slope estimator for linear fitting yield the most robust and reliable classification results with this novel technique. These values ensure adequate linearity in the underlying dataset and sufficient statistical significance to exclude randomness.

4.2. Benchmarking classification performance

Machine learning was successfully applied to and evaluated on various PET datasets for disease classification (Garraux et al., 2013; Liu et al., 2018; Vandenberghe et al., 2013). In order to benchmark our novel method, we set up SVM and RFC classifiers with 5-fold nested cross validation, as described above, and evaluated the classification performance on the same set of features as used for connectivity analysis. Machine learning classification relies on individual features, the normalized uptake in individual brain regions, whereas our method instead takes the metabolic connectome into account. The connectivity-based classification achieved performances slightly above machine learning classifiers (e.g. accuracy 80.4% versus 76.5% for 3 classes). An increase in the number of classes is equivalent to rising demands on the algorithm and therefore classification into more classes yielded lower classification accuracies. Nevertheless, both methods performed well compared to classification by chance. This indicates that the calculated connectivity patterns were well suited as the basis for classification and comparably simple linear fitting of a population could outperform non-linear machine learning methods especially in small datasets where machine learning methods tend to underfit. The main advantage of our method is the incorporation of PET-based metabolic connectivity information into classification tasks. Nevertheless, the identification of proper cohorts with low class-label noise in the fitting procedure is essential and limits the classification performance in both machine learning and connectivity classification. Furthermore, our method determines the connectome based on linear relationships between the brain regions, whereas machine learning classifiers could learn non-linear relationships as well. This limits the performance of our method as possible non-linear connections could be missed. The non-linearity of the connectome needs to be investigated in future studies to possibly use other curve fitting models for classification. One advantage of machine learning classification is that the dataset does not need to fulfil the assumptions for linear fitting. Our dataset fulfils the assumptions, but that is not necessarily the

case for other datasets. The small sample size of our UL dataset, which is in line with the governmental animal protection regulations, also limits the possibility to evaluate the algorithms on test datasets. That is why we implemented cross-validation procedures.

4.3. Clinical relevance of connectivity-based classification in nuclear medicine

The proposed algorithm aims to distinguish classes by analyzing metabolic brain connectivity patterns. Classes could, for example, represent different diseases related to brain network dysfunction (such as distinct neurodegenerative syndromes), but also different cerebral network states during disease progression or recovery in defined disease cohorts. In this study, we investigated the classification performance on a longitudinally recorded [18F]-FDG PET dataset in a group of rats after UL as a well-established paradigm. Acute unilateral vestibular damage results in a rapid process of adaptive brain plasticity, called vestibular compensation, which is reflected in dynamic changes of connectivity patterns in parallel with behavioral recovery (Grosch et al., 2020; Zwergal et al., 2016). The varying connectivity patterns at five time points before and after UL enabled us to consider each day of PET imaging as a different class. The highly controlled experimental environment and procedures led to zero class-label noise; therefore, the dataset is perfectly suited to investigating classification techniques. A potential drawback of preclinical imaging are the small dimensions of the experimental subjects (e.g. rats), and the limited spatial resolution of the small-animal PET scanner (Visser et al., 2009). Even though we tried to minimize error sources (e.g. positioning errors in the scanner were countered with a custom-made head-holder), classification performance could be improved in human imaging by enhancing connectivity calculations through better image acquisition conditions. Furthermore, clinical PET scanning does not require patient anesthesia, which could influence the imaged brain metabolism in our rat model. We tried to minimize this potential influence by waking up the animals for the 30 min tracer uptake period. Another potential limitation in the translation of the proposed technique to clinical datasets could be the influence of sex on the classification results, as this was not investigated in this study due to the lack of female animals in the experimental setup.

On the other hand, automated data analysis to support clinical diagnosis in human datasets may be challenging because of lower labelling accuracy and potential multiple neurological pathologies present in one patient. For a future clinical application of connectivity-based classification approaches in [18F]-FDG PET neuroimaging, it is advantageous that the practically relevant disorders such as different dementia or hypokinetic syndromes, are increasingly considered as complex network disorders. For neurodegenerative disorders, such as Alzheimer's disease (AD), Lewy body dementia (LBD), or Parkinson's disease (PD), a contribution of network changes to functional decline (e.g. in cognitive domains or motor domains) was convincingly shown (Ge et al., 2020; Huber et al., 2020; Morbelli et al., 2013). Furthermore, in clinical practice [18F]-FDG PET still is the cornerstone of diagnostic imaging assessment in dementia (e.g., AD versus non-AD syndromes, or PD versus atypical parkinsonian syndromes). For those network disorders, our classification method based on metabolic connectivity incorporates knowledge about interactions within the brain, which accesses more information than solely uptake values in brain regions or single voxels. It thereby represents a whole brain 'meta'-view of [18F]-FDG PET data, which surpasses visual or semi-automated (e.g. z-score) analysis. In consequence, as the investigated classes exhibit strong variations in the connectome due to the underlying disease pattern, classification relying on metabolic connectivity could potentially enhance clinical diagnosis in neurological network diseases, whereby this method extends the field of metabolic connectome-based disease classification. Other methods (SICE and KLSE) were investigated on AD datasets, including only brain regions especially affected by the disease. There may be other neurological disorders in which the a priori reduction of brain regions based on pathophysiological studies

is more challenging. In this instance, a data driven detection of brain connections in the whole brain such as in our study may have advantages. On the other hand, SICE and KLSE but not the proposed method were validated on imbalanced datasets, which regularly appear in clinical settings (Huang et al., 2010; Wang et al., 2020). Transferring our preclinical results to human datasets of patients suffering from neurological diseases could potentially add another powerful imaging biomarker of metabolic changes to the diagnostic evaluation in a wide range of entities.

5. Conclusion

This novel Pearson's correlation-based method enables classification of single subjects by metabolic connectivity analysis with a diagnostic accuracy that outperforms machine learning classifiers in a preclinical dataset. Classes were derived from functional network changes due to adaptive brain plasticity following peripheral vestibular damage in the rat. Further evaluation in human datasets of propagated neuropsychiatric network disorders is needed to potentially transfer this technique to clinical diagnostics in nuclear medicine applications.

Acknowledgements

CRediT Author Statement

Maximilian Grosch: Conceptualization, Methodology, Software, Validation, Formal analysis, Data Curation, Writing – Original Draft, Visualization. Leonie Beyer: Conceptualization, Methodology, Formal analysis, Writing – Original Draft. Magdalena Lindner: Investigation, Project administration. Lena Kaiser: Software, Writing – Review & Editing. Seyed-Ahmad Ahmadi: Methodology, Writing – Review & Editing. Anna Stockbauer: Writing – Review & Editing. Peter Bartenstein: Resources, Writing – Review & Editing. Marianne Dieterich: Resources, Writing – Review & Editing. Matthias Brendel: Conceptualization, Methodology, Formal analysis, Writing – Original Draft. Andreas Zwergal: Resources, Writing – Review & Editing, Supervision, Project administration, Funding acquisition. Sibylle Ziegler: Resources, Formal analysis, Writing – Original Draft, Supervision

Data and code availability statement

Data and code reported in this article will be shared with any appropriately qualified investigator on request.

APPENDIX A

A.1. 4-class classification

The classification accuracy with different thresholds of the correlation coefficient r reached 64.7% (LS), 63.2% (Sen), and 60.3% (Siegel) and deteriorated for lower and higher thresholds (Fig. 3a). Rejecting negative correlations resulted in maximum classification accuracies of 60.3% (LS), 52.9% (Sen), and 57.4% (Siegel) (Fig. 3b). Reducing the number of connections via sparsity thresholds resulted in classification accuracies of up to 64.7% (LS), 51.5% (Sen), and 55.9% (Siegel) (Fig. 3c).

Moreover, the classification accuracy with respect to the distance-to-fit d reached 75.0% (LS), 67.6% (Sen), and 70.6% (Siegel) when using static thresholds and r-filtering (Fig. 4a). With dynamic d thresholds and r-filtering, accuracies of 52.9% (LS), 52.9% (Sen), and 55.9% (Siegel) were achieved (Fig. 4b). Using s-filtered connectivity patterns resulted

in classification accuracies of up to 70.6% (LS), 63.2% (Sen), and 67.6% (Siegel) for static d thresholds (Fig. 4c) and 57.4% (LS), 57.4% (Sen), and 55.9% (Siegel) for dynamic d thresholds (Fig. 4d). For r-filtering and static d-thresholds, the ROC-AUC values showed a course related to those of the corresponding accuracies and reached 0.863 (LS), 0.835 (Sen), and 0.842 (Siegel) (Fig. 5a).

A.2. 5-class classification

For 3-class classification, the classification accuracy with different thresholds of the correlation coefficient r reached 44.7% (LS), 47.1% (Sen), and 40.0% (Siegel) and deteriorated for lower and higher thresholds (Fig. 3a). Rejecting negative correlations resulted in maximum classification accuracies of 43.5% (LS), 36.5% (Sen), and 38.8% (Siegel) (Fig. 3b). Reducing the number of connections via sparsity thresholds resulted in classification accuracies of up to 42.4% (LS), 38.8% (Sen), and 38.8% (Siegel) (Fig. 3c).

Moreover, the classification accuracy with respect to the distance-to-fit d reached 51.8% (LS), 51.8% (Sen), and 54.1% (Siegel) when using static thresholds and r-filtering (Fig. 4a). With dynamic d thresholds and r-filtering, accuracies of 42.4% (LS), 40.0% (Sen), and 43.5% (Siegel) were achieved (Fig. 4b). Using s-filtered connectivity patterns resulted in classification accuracies of up to 52.9% (LS), 54.1% (Sen), and 54.1% (Siegel) for static d thresholds (Fig. 4c) and 44.7% (LS), 47.1% (Sen), and 44.7% (Siegel) for dynamic d thresholds (Fig. 4d). For r-filtering and static d-thresholds, the ROC-AUC values showed a course related to those of the corresponding accuracies and reached 0.777 (LS), 0.762 (Sen), and 0.761 (Siegel) (Fig. 5a).

References

Anderson, J.S., et al., 2011. Functional connectivity magnetic resonance imaging classification of autism. Brain 134, 3742–3754. doi:10.1093/brain/awr263.

Anniko, M., Wersall, J., 1977. Experimentally (atoxyl) induced ampullar degeneration and damage to the maculae utriculi. Acta Otolaryngol. 83, 429–440. doi:10.3109/00016487709128868.

Barth, C., et al., 2015. Sex hormones affect neurotransmitters and shape the adult female brain during hormonal transition periods. Front. Neurosci. 9, 37.

Beck, R., et al., 2014. The mixed blessing of treating symptoms in acute vestibular failure–evidence from a 4-aminopyridine experiment. Exp. Neurol. 261, 638–645. doi:10.1016/j.expneurol.2014.08.013.

Benesty, J., et al., 2009. Pearson Correlation Coefficient. Noise Reduction in Speech Processing. Springer, pp. 1–4.

Benjamini, Y., Yekutieli, D., 2001. The control of the false discovery rate in multiple testing under dependency. Ann. Stat. 1165–1188.

Cai, L., et al., 2020. A review of the application of deep learning in medical image classification and segmentation. Ann. Transl. Med. 8.

Du, Y., et al., 2018. Classification and prediction of brain disorders using functional connectivity: promising but challenging. Front. Neurosci. 12, 525. doi:10.3389/fnins.2018.00525.

Galvin, J.E., et al., 2011. Resting bold fMRI differentiates dementia with Lewy bodies vs Alzheimer disease. Neurology 76, 1797–1803. doi:10.1212/WNL.0b013e31821ccc83.

Garraux, G., et al., 2013. Multiclass classification of FDG PET scans for the distinction between Parkinson's disease and atypical parkinsonian syndromes. NeuroImage: Clinical 2, 883–893.

Ge, J., et al., 2020. Metabolic network as an objective biomarker in monitoring deep brain stimulation for Parkinson's disease: a longitudinal study. EJNMMI Res 10, 131. doi:10.1186/s13550-020-00722-1.

Grosch, M., et al., 2020. Dynamic whole-brain metabolic connectivity during vestibular compensation in the rat. Neuroimage, 117588 doi:10.1016/j.neuroimage.2020.117588.

Huang, S., et al., 2010. Learning brain connectivity of Alzheimer's disease by sparse inverse covariance estimation. Neuroimage 50, 935–949. doi:10.1016/j.neuroimage.2009.12.120.

Huber, M., et al., 2020. Metabolic correlates of dopaminergic loss in dementia with lewy bodies. Mov. Disord. 35, 595–605. doi:10.1002/mds.27945.

Jalilianhasanpour, R., et al., 2019. Functional Connectivity in Neurodegenerative Disorders: Alzheimer's Disease and Frontotemporal Dementia. Top Magn Reson Imaging 28, 317–324. doi:10.1097/RMR.0000000000000223.

Kenny, E.R., et al., 2012. Functional connectivity in cortical regions in dementia with Lewy bodies and Alzheimer's disease. Brain 135, 569–581. doi:10.1093/brain/awr327.

Klöppel, S., et al., 2012. Diagnostic neuroimaging across diseases. Neuroimage 61, 457–463.

Lindner, M., et al., 2017. Betahistine improves vestibular compensation after unilateral labyrinthectomy: a [18F] FDG-μPET study in the rat. EANM Conference 2017.

Lindner, M., et al., 2019. Ginkgo biloba Extract EGb 761 Improves Vestibular Compensation and Modulates Cerebral Vestibular Networks in the Rat. Front Neurol 10, 147. doi:10.3389/fneur.2019.00147.

Liu, F., et al., 2015. Multivariate classification of social anxiety disorder using whole brain functional connectivity. Brain Struct. Funct. 220, 101–115. doi:10.1007/s00429-013-0641-4.

Liu, M., et al., 2018. Classification of Alzheimer's disease by combination of convolutional and recurrent neural networks using FDG-PET images. Front. Neuroinform. 12, 35.

Magnusson, A.K., et al., 2002. Early compensation of vestibulo-oculomotor symptoms after unilateral vestibular loss in rats is related to GABA(B) receptor function. Neuroscience 111, 625–634. doi:10.1016/s0306-4522(01)00618-2.

Morbelli, S., et al., 2013. Metabolic networks underlying cognitive reserve in prodromal Alzheimer disease: a European Alzheimer disease consortium project. J. Nucl. Med. 54, 894–902. doi:10.2967/jnumed.112.113928.

Nielsen, J.A., et al., 2013. Multisite functional connectivity MRI classification of autism: ABIDE results. Front. Hum. Neurosci. 7, 599. doi:10.3389/fnhum.2013.00599.

Pedregosa, F., et al., 2011. Scikit-learn: Machine learning in Python. J. Mach. Learn. Res. 12, 2825–2830.

Peraza, L.R., et al., 2016. Regional functional synchronizations in dementia with Lewy bodies and Alzheimer's disease. Int. Psychogeriatr. 28, 1143–1151. doi:10.1017/S1041610216000429.

Pisner, D.A., Schnyer, D.M., 2020. Support Vector Machine. Machine Learning. Elsevier, pp. 101–121.

Sanabria-Diaz, G., et al., 2013. Glucose metabolism during resting state reveals abnormal brain networks organization in the Alzheimer's disease and mild cognitive impairment. PLoS One 8, e68860. doi:10.1371/journal.pone.0068860.

Sarica, A., et al., 2017. Random forest algorithm for the classification of neuroimaging data in Alzheimer's disease: a systematic review. Front. Aging Neurosci. 9, 329. doi:10.3389/fnagi.2017.00329.

Schiffer, W.K., et al., 2006. Serial microPET measures of the metabolic reaction to a microdialysis probe implant. J. Neurosci. Methods 155, 272–284. doi:10.1016/j.jneumeth.2006.01.027.

Sen, P.K., 1968. Estimates of the regression coefficient based on Kendall's tau. J. Am. Stat. Assoc. 63, 1379–1389.

Siegel, A.F., 1982. Robust regression using repeated medians. Biometrika 69, 242–244.

Strauss, F., 2020. Der weibliche Sexualzyklus. Walter de Gruyter GmbH & Co KG.

Tsai, S.Y., 2018. Reproducibility of structural brain connectivity and network metrics using probabilistic diffusion tractography. Sci. Rep. 8, 11562. doi:10.1038/s41598-018-29943-0.

Van Griethuysen, J.J., et al., 2017. Computational radiomics system to decode the radiographic phenotype. Cancer Res. 77, e104–e107.

Vandenberghe, R., et al., 2013. Binary classification of 18F-flutemetamol PET using machine learning: comparison with visual reads and structural MRI. Neuroimage 64, 517–525.

Vignaux, G., et al., 2012. Evaluation of the chemical model of vestibular lesions induced by arsanilate in rats. Toxicol. Appl. Pharmacol. 258, 61–71. doi:10.1016/j.taap.2011.10.008.

Visser, E.P., et al., 2009. Spatial resolution and sensitivity of the Inveon small-animal PET scanner. J. Nucl. Med. 50, 139–147. doi:10.2967/jnumed.108.055152.

Wang, M., et al., 2020. Individual brain metabolic connectome indicator based on Kullback-Leibler Divergence Similarity Estimation predicts progression from mild cognitive impairment to Alzheimer's dementia. Eur. J. Nuclear Med. Mol. Imaging 47, 2753–2764.

Yakushev, I., et al., 2017. Metabolic connectivity: methods and applications. Curr. Opin. Neurol. 30, 677–685. doi:10.1097/WCO.0000000000000494.

Zwergal, A., et al., 2017. In vivo imaging of glial activation after unilateral labyrinthectomy in the rat: a [(18)F]GE180-PET Study. Front. Neurol. 8, 665. doi:10.3389/fneur.2017.00665.

Zwergal, A., et al., 2016. Sequential [(18)F]FDG microPET whole-brain imaging of central vestibular compensation: a model of deafferentation-induced brain plasticity. Brain Struct. Funct. 221, 159–170. doi:10.1007/s00429-014-0899-1.