Sex classification from resting-state functional brain networks

Lisa Wiersch

HHU Düsseldorf

Institute of Neuroscience and Medicine

Brain and Behaviour (INM-7)

Master Thesis

First supervisor: PD Dr. Susanne Weis

Second supervisor: Prof. Dr. Christian Bellebaum

Author Note

Lisa Wiersch

Lisa.Wiersch@uni-duesseldorf.de

student number: 2324137

Abstract

Sex differences in the brain have received a wide interest in neuroscientific research. Instead of task-based group comparisons, the present study employed a machine-learning (ML)-approach to examine whether the resting-state functional connectivity of 12 meta-analytically defined networks carries enough information to accurately predict the sex of a person. It was hypothesized that especially emotion-related networks should classify well. Sex classification analyses were conducted in the datasets of the healthy brain network (HBN, n = 218) the Rockland Sample of the enhanced Nathan Kline Institute (eNKI, n = 574), the Human Connectome Project (HCP, n = 734) and the 1000BRAINS-dataset (n = 995). The ML-algorithms LASSO, LSVM, Ridge and RVM were used for this classification approach. The results showed that the eNKI- and HCP-datasets as well as the algorithms LASSO and Ridge received on average higher classification accuracies than the other datasets and algorithms. The networks of autobiographical and semantic memory reached the highest accuracies of all networks. Taken together, the results did not support the initial hypothesis. Instead, the results generally displayed a strong dependency on the datasets and ML-algorithms.

Keywords: sex differences, brain networks, resting-state fMRI, functional connectivity, classification, machine-learning

Sex classification from resting-state functional brain networks

Sex differences in the brain have received a wide interest in psychology and neurosciences and many studies already supported the existence of structural and functional differences between men and women (Zaidi, 2010). There was evidence for sex differences in total brain size (Nopoulos, P., Flaum, M., O'Leary, D., & Andreasen, N. C., 2000; Ruigrok et al., 2014), but also in several cortical and subcortical regions (Ruigrok et al., 2014) and also in the ratio of gray and white matter (Sacher et al. 2013; Filkowski et al. 2017; Zaidi 2010). Moreover, sex differences were also evident in lateralization: Women showed more bilateral functioning, whereas men showed stronger lateralization in the domain of language (Baxter et al., 2003) and face processing (Proverbio, Brignone, Matarazzo, Del Zotto, & Zani, 2006). This phenomenon went along with greater interhemispheric connectivity in women and greater intrahemispheric connectivity in men in structural (Ingalhalikar et al., 2014) as well as in functional connectivity (Satterthwaite et al., 2015).

Sex differences were also found in different cognitive domains such as language (Baxter et al., 2003; Burman, Bitan, & Booth, 2008; Weiss, E. M., Kemmler, G., Deisenhammer, E. A., Fleischhacker, W. W., & Delazer, M., 2003) or visual-spatial attention (León, Tascón, & Cimadevilla, 2016; Vaquero, E., Cardoso, M. J., Vazque, M., & Gomez, C. M., 2004), which could be partially attributed to sex-specific functional connectivity in the brain (Kansaku, K., & Kitazawa, S., 2001). Another cognitive domain in which sex differences were found is the processing of emotions: Several studies demonstrated evidence for emotion-related sex differences; for example in form of the recruitment of key regions associated with the processing of emotional materials (Hofer et al., 2006; Stevens & Hamann, 2012), but also in behavior (Whittle, Yücel, Yap, & Allen, 2011).

However, for some other cognitive domains the literature provided not an entirely clear picture regarding sex differences: Especially for working memory, defined as the ability to maintain and manipulate information for a limited period of time (Rottschy et al., 2012; Speck, O., Ernst, T., Braun, J., Koch, C., Miller, E., & Chang, L., 2000), the literature provided partly contradicting results regarding sex differences: Whereas Lynn and Irwing (2002) found behavioral sex differences, other studies could not support this assumption (Solianik, Brazaitis, & Skurvydas, 2016; Teleb, A. A., & Al Awamleh, A. A., 2012). In turn, other studies found sex differences in the functional organization and activation for working memory (Hill, Laird, & Robinson, 2014; Speck, O., Ernst, T., Braun, J., Koch, C., Miller, E., & Chang, L., 2000).

For the domain of attention, the literature is also not entirely clear: For vigilant or sustained attention, which is defined as focussing one´s attention to certain stimuli or simple tasks for a certain period of time (Langner & Eickhoff, 2013; Prinzel & Freeman, 1997), some studies could not found any sex differences (Levy & Hobbes, 1979), whereas other studies stated that sex differences in vigilance performance are evident, but task-specific (Dittmar, M. L., Warm, J. S., Dember, W. N., & Ricks, D. F., 1993; Lin, C. C., Hsiao, C. K., & Chen, W. J., 1999; Prinzel & Freeman, 1997).

Furthermore, also for high-level control processes, the studies did not form an uniform picture, as for example for cognitive action control: Cognitive action control describes an attentional system which mediates goal-oriented behavior by maintaining goal-relevant behavior, which is essential in everyday life (Cieslik, Mueller, Eickhoff, Langner, & Eickhoff, 2015; Mansouri, Fehring, Gaillard, Jaberzadeh, & Parkington, 2016). It could be found that males and females differ in their neural recruitment for cognitive control, but a main effect of sex could not be proven (Liu, Hairston, Schrier, & Fan, 2011; Mansouri et al., 2016).

Altogether, it could be stated that for several cognitive domains, the literature either provided evidence for or against sex differences. These inconsistencies underline the need for further research in the field of sex differences in cognitive domains, which should be also extended to sex differences in functional connectivity (Sacher, Neumann, Okon-Singer, Gotowiec, & Villringer, 2013).

One possible explanation for these inconsistencies could be the phenomenon of the publication bias. This phenomenon means that many studies on a certain topic have been conducted, but were not published because of insignificant findings. Another possible explanation for these ambiguous findings in some cognitive domains could be that these studies were often task-based group comparisons between men and women, often conducted with a small number of participants. To avoid this problem, we chose a different approach to investigate sex differences: The present study employed a machine-learning (ML)-approach that learns sex-specific characteristics in order to predict the sex of a person on the basis of these characteristics.

With this approach we wanted to meet the demand of further research regarding sex differences in cognitive domains by investigating the underlying functional connectivity of different cognitive domains. For this purpose we examined the resting-state (RS) functional connectivity in meta-analytically defined networks. These nodes of the networks were defined by which brain regions showed common activation for a certain cognitive domain in different task-based studies. The RS functional connectivity refers to the temporal correlation of spontaneous fluctuations in the blood oxygen level-dependent (BOLD) signal in fMRI while the subjects rest in the scanner and should not think of anything in particular (Greicius, Supekar, Menon, & Dougherty, 2009). To cover a broad set of cognitive domains, we analyzed networks based on some cognitive domains mentioned before and also some more, in particular the networks which were already examined together in a study by Pläschke et al.

(2017): vigilant attention (VigAtt), theory-of-mind cognition (ToM), reward-related decision making (Rew), motor execution (Motor), mirror neuron system (MNS), cognitive emotion regulation (ER), empathic processing (Empathy), emotional scene and face processing (emoSF), cognitive action control (CogAC), as well as the autobiographical memory (AM), semantic memory (SM) and working memory (WM), further information are provided in table 1.

Taken together, we wanted to investigate the functional connectivity in RS for all the 12 networks; with the ML-approach we examined whether the functional connectivity carried enough sex-specific information to accurately predict the sex of a participant. In doing so, we could finally point out the cognitive domains, whose underlying functional connectivity carried enough information to distinguish well between men and women.

Overall, some of the 12 networks are directly or indirectly related to emotions, such as the network of cognitive emotion regulation. This network is based on studies conducted to cognitive reappraisal, which means an emotion regulation strategy defined as thinking differently about a stimulus than before in order to change its affective impact (Buhle et al., 2014; McRae, Ochsner, Mauss, Gabrieli, & Gross, 2008). The literature provided some contradicting results proposing men to be more successful in this task because they show either higher activity in cognitive areas dealing with emotion regulation strategies (Filkowski, Olsen, Duda, Wanger, & Sabatinelli, 2017), or men showed lower activation in these regions which was also interpreted as more efficient (McRae et al., 2008). This contrast was discussed in a review by Whittle et al. (2011) who proposed a theory of different strategies to succeed in this task: Females were proposed to recruit neural regions associated with emotional processing to transform negative affect into positive affect (McRae et al., 2008) and men recruited more regions related to cognitive processing to regulate emotions. Altogether, the results of the studies did not agree as to whether men or women are more efficient in regulating their emotions, but there was an agreement in sex-specific strategies to deal with this problem.

Sex differences in the recruitment of neural regions were also reported for the emotional processing of faces and scenes: In addition to the fact that sex-specific neuronal responses to emotional stimuli could be already pointed out (Hofer et al., 2006), there were also sex-related differences in face processing: There was female superiority in face processing reported by Batty and Taylor (2006) as well as by Sommer, Hildebrandt, Kunina-Habenicht, Schacht, and Wilhelm (2013). But also in the neural correlates the literature portrayed a uniform picture of a right-hemispheric lateralization for face processing (Kanwisher, McDermott, & Chun, 1997; McCarthy, Puce, Gore, & Allison, 1997; Sergent, J., Ohta, S., & MACDONALD, B., 1992) which was assumed to be more present in men because women showed a more bilateral functioning (Bourne, 2005; Bourne & Maxwell, 2010; Godard & Fiori, 2010; Proverbio et al., 2006).

A cognitive ability that is more concerned with the understanding of emotions is called theory of mind. It describes the understanding of other people’s feelings, thoughts and mental states and that these can differ from one´s own (Premack, D., & Woodruff, G., 1978). This knowledge enables people to explain and predict the behavior of other people, which is an important component in social interactions (Gallagher & Frith, 2003). Multiple studies were conducted in this field and the literature provided an uniform picture of sex differences in form of a female advantage in behavioral measures (Bosacki, 2000; Calero, Salles, Semelman, & Sigman, 2013; Charman, Ruffman, & Clements, 2002; Walker, 2005), but most of the studies were conducted with children and not with adults.

Table 1

Respective publication of the meta-analysis for each network

network	abbreviation	publication
autobiographical memory	AM	Spreng, R. N., Mar, R. A., & Kim, A. S. (2009)
cognitive action control	CogAC	Cieslik et al. (2015)
emotional scene and face processing	emoSF	Sabatinelli et al. (2011)
empathic processing	Empathy	Bzdok et al. (2012)
cognitive emotion regulation	ER	Buhle et al. (2014)
mirror neuron system	MNS	Caspers, Zilles, Laird, and Eickhoff (2010)
motor execution	Motor	Witt, Laird, and Meyerand (2008)
reward-related decision making	Rew	Liu et al. (2011)
semantic memory	SM	Binder, Desai, Graves, and Conant (2009)
theory-of-mind cognition	ToM	Bzdok et al. (2012)
vigilant attention	VigAtt	Langner and Eickhoff (2013)
working memory	WM	Rottschy et al. (2012)

Closely related to theory of mind is the mirror neuron system because it serves action understanding by simultaneous activation for the execution or imitation of actions and action observation (Cheng et al., 2009; Cheng, Y., Decety, J., Lin, C. P., Hsieh, J. C., Hung, D., & Tzeng, O. J, 2007; Rizzolatti & Craighero, 2004). Also for this network the literature showed an uniform picture with women having larger volumes in the brain areas related to the mirror neuron system, higher spinal excitability or stronger activation in task-based studies (Cheng et al., 2009; Cheng, Y. W., Tzeng, O. J., Decety, J., Imada, T., & Hsieh, J. C, 2006; Cheng, Y., Decety, J., Lin, C. P., Hsieh, J. C., Hung, D., & Tzeng, O. J, 2007; Cheng, Y., Lee, P. L., Yang, C. Y., Lin, C. P., Hung, D., & Decety, J., 2008).

The theory of mind and mirror neuron system are in turn essential components for empathic processing, which describes the ability to share the feelings of others and is therefore very important for social interactions (Bernhardt & Singer, 2012; Schulte-Rüther, Markowitsch, Fink, & Piefke, 2007). Behavioral sex differences were found in form of women reporting higher empathy than men (Pohl, Bender, & Lachmann, 2005), but it was also stated that these behavioral differences are not reflected in the neural data (Michalska, Kinzler, & Decety, 2013). Nevertheless, other studies provided evidence for sex-specific recruitment of areas related to empathy: Apart from a common network for empathy shared by both sexes, men were found to recruit more areas related to cognitive evaluation and mentalizing, whereas women recruited more emotion-related areas (Derntl et al., 2010). Schulte-Rüther, Markowitsch, Shah, Fink, and Piefke (2008) supported this assumption and stated that women use areas related to the human mirror neuron system to access the feelings of others. So, sex differences in empathy could be related to sex-related differences in the mirror neuron system (Cheng, Y., Lee, P. L., Yang, C. Y., Lin, C. P., Hung, D., & Decety, J., 2008).

Another domain indirectly associated with emotions is autobiographical memory, which is defined as the episodic memory for personal events from one´s own life (Spreng, R. N., Mar, R. A., & Kim, A. S., 2009; Young, Bellgowan, Bodurka, & Drevets, 2013). Piefke and Fink (2005) reported sex-specific differential neural activation while recollecting autobiographical memories supporting the hypothesis of distinct cognitive memory styles. Differences in the neural recruitment for autobiographical memories were also found by

Young et al. (2013). Some other studies reported behavioral differences in form of women reporting autobiographical memories in a more detailed, social- and emotion-related way (Buckner & Fivush, 1998; Grysman & Hudson, 2013) or women reported longer, richer and evaluative autobiographical narratives (Schulkind, Schoppel, & Scheiderer, 2012).

Another domain of memory, which also describes knowledge that is acquired through one's own experience, is semantic memory. In contrast to autobiographical memory, this domain does not describe personal events, but rather knowledge about people, objects, relations, objects and culture acquired through own experience (Binder et al., 2009). The regions related to semantic memory are found to deal with highly processed multimodal and supramodal input, so they serve high-level integrative processes (Binder et al., 2009). A few studies investigated sex differences in semantic memory with tests about general knowledge in different fields which resulted in male or female superiority in different sub-categories (Lynn & Irwing, 2002; Tran, Hofer, & Voracek, 2014). But the results could not provide an entirely clear picture to sex differences in semantic memory.

Just like for semantic memory, the literature to sex differences in reward-related decision making was relatively sparse: Sex differences were found in the recruitment of different neural regions to form the reward-related neural function, but this was partially only evident under certain conditions like stressful conditions and not in general (Dreher et al., 2007; Lighthall et al., 2012).

Much more information were available for the motor domain regarding sex differences: Several studies reported that men performed better than women in several motor tasks (Davies & Rose, 2000; Dorfberger, Adi-Japha, & Karni, 2009; Jiménez-Jiménez et al., 2011) and Junaid and Fellowes (2006) proposed that males and females perform differently well in different sub-categories of motor tasks. These behavioral differences could also be related to a neural basis: Amunts, Jäncke, Mohlberg, Steinmetz, and Zilles (2000) provided evidence for sex-dependent differences in the organization of the motor cortex. However, also all these results together did not result in a uniform picture regarding sex differences in the domain of motor execution.

In the present study we wanted to find out whether these sex differences from task-based studies could be connected to the underlying functional connectivity in the respective neural networks. It could be assumed that in the cognitive domain in which many sex differences could be already found, there might be also sex differences in the underlying functional connectivity within the respective neuronal network. The ML-algorithm should then be able to learn well from this set of sex differences in a network and make an accurate prediction of the sex of a person. This ML-approach for sex classification is relatively new and currently very little literature employing similar approaches exists. One of the few studies that used a ML-approach for sex classification was the study by Zhang, Dougherty, Baum, White, and Michael (2018): They employed predefined functional templates of different network and found that the features that contributed the most to the prediction were in the default-mode network, fronto-parietal and sensorimotor network. So, in their study sex could be reliably predicted using rfMRI data with an accuracy up to 87%.

Because of the fact, that this approach was not conducted very often, it is not clear which ML-algorithm is the best to use for the question of classification. There are many ML-algorithms which differ by different features. For example, there are some approaches that classify by a linear separation of classes, while others take a non-linear approach (Müller, Anderson, & Birch, 2003). Again other approaches create a sparse model, which means that in these models only a few variables are considered that are not set to zero (James, G., Witten, D., Hastie, T., & Tibshirani, R, 2013; Tipping, 2001). On the other hand, other approaches create models that consider all possible variables and are therefore non-sparse. In statistics, it is well known that there is not one single algorithm that is the best for a specific question. This phenomenon is known as ‘no free lunch theorem’ (James, G., Witten, D., Hastie, T., & Tibshirani, R, 2013; Wolpert, D. H., & Macready, W. G., 1997). Thus, an algorithm may be suitable for a certain question and dataset, while for a different kind of question and data another algorithm works better. For this reason, we applied in this study four different ML-algorithms exploratively to four big different datasets to see if the classification results for the networks are independent of the dataset and the ML-algorithm.

According to the literature regarding the cognitive domains and the assumption that task-based sex differences could manifest themselves also in the underlying functional connectivity, we stated the following hypotheses for the 12 networks that are related to the cognitive domains: For the networks of theory of mind, mirror neuron system and empathy, we hypothesized that the ML-algorithms should classify well between men and women because of the big amount of studies that already provided evidence sex differences in these domains. We also expected a good classification accuracy for the network of emotional scene and face processing which could result from the reported sex differences in the lateralization of face processing. Same is true for the networks of cognitive emotion regulation and autobiographical memory. Here we expected that the sex-specific processing styles should also lead in the RS functional connectivity to a good distinction between the sexes. Due to the large amount of sex differences that could be found in the literature in motor tasks, we also hypothesized a high classification accuracy for the network of motor execution. For the cognitive domains of working memory, vigilant attention, semantic memory, reward-related decision making and cognitive action control the literature either provided not a clear picture regarding sex differences or was relatively sparse. For this reason, we did not directly expect a high classification accuracy in these respective neural networks.

To sum it up, we expected higher classification accuracies especially in the networks that are related to emotions, because the processing of emotions is known to be different between men and women (Sacher et al., 2013). Nevertheless it must also kept in mind that this ML-approach is different to most of the studies reported before performing group comparisons, wherefore it could also lead to different results than the studies before.

Methods

Datasets

The four datasets, which were analyzed for this investigation, are presented in the following:

Healthy Brain Network (HBN). The HBN has been initiated by the Child Mind Institute. It is an open resource for transdiagnostic research and contains information about different types of phenotypes as well as, among others, RS fMRI (Alexander et al., 2017). The authors reported that the participants for HBN were measured in a Siemens 3 Tesla (T) Tim Trio MRI scanner with a CMRR simultaneous multi-slice echo planar imaging sequence in the Brain Imaging Center of the Rutgers University with the following parameters: 60 slices, resolution of 2.4 x 2.4 x 2.4 mm^3, Field of View (FoV) = 64 x 2.4 mm^2, Repetition Time (TR) = 800 ms, Echo time (TE) = 30.00 ms and flip angle = 31°. The subjects of the HBN-dataset included in this study were a total of 218, of whom 109 are men. The age of these subjects ranged from 5.04 to 19.50 ($M = 10.52$, $SD = 3.46$).

Enhanced Nathan Kline Institute-Rockland Sample (eNKI). The eNKI is a deeply phenotyped large-scale dataset containing subjects with ages from childhood to late adulthood (Nooner et al., 2012). For this study we used the RS fMRI data of the eNKI acquired with the Siemens Magnetom Trio Tim sygno which had the following parameters:

38 slices, resolution = 3.0 x 3.0 x 3.0 mm^3, FoV = 256 x 200 mm^2, Multi-slice mode = interleaved; TR = 2500 ms, TE = 30 ms, flip angle = 80°. The present study contains 574 subjects (212 males) of the eNKI, aging from 6.00 to 85.00 (*M* = 40.23, *SD* = 21.70).

Human Connectome Project (HCP). The dataset of the HCP contains a large cohort of healthy adults and is freely available to use in the neuroimaging community (Smith et al., 2013; van Essen et al., 2012). The HCP contains RS fMRI data, acquired with a Siemens Skyra 3 T MRI scanner with multiband accelerated echo-planar imaging. Further parameters were: 72 slices, resolution = 2 x 2 x 2 mm^3 and FoV = 208 x 180 mm^2, TR = 720ms , TE = 33.1 ms and flip angle = 52° (Glasser et al., 2013; Smith et al., 2013). We included 734 subjects (366 males) of the HCP-dataset in this study, with their age ranges from 22.00 to 37.00 (*M* = 28.55, *SD* = 3.46).

1000BRAINS. The 1000BRAINS study by Caspers et al. (2014) investigated the variability in the human brain during aging in a population-based German cohort. These subjects were measured in a Siemens Tim-TRIO 3 T MR scanner using gradient-echo echoplanar imaging (EPI). The corresponding parameters were the following: 36 slices, resolution = 3.1 x 3.1 x 3.1 mm^3, FoV = 200 x 200 mm^2, TR = 2.2s and TE = 30 ms, flip angle = 90°. We included 995 subjects (547 males) of the 1000BRAINS-dataset in this study, whose age ranges from 18.50 to 88.20 (*M* = 63.44, *SD* = 12.41).

A re-analysis of the anonymized data of these samples was approved by the Ethics Commission of the Medical Faculty of the Heinrich-Heine-University Düsseldorf (study numbers: 2018-317-RetroDEuA and 5193).

Preprocessing

HBN & eNKI. First of all, the data of the two datasets were realigned and unwarped by using the Phase map created with the SPM12 Fieldmap Toolbox v2.1 (https://www.fil.ion.ucl.ac.uk/spm/toolbox/fieldmap/). The step of estimation contained the following parameters: quality of 0.95, separation of 3, register to the mean and an interpolation of the 7th Degree of B-Spline. Unwarping reslice was also performed with the interpolation with the 7th Degree of B-Spline. In the following step of intersubject registration (normalization) the mean EPI was co-registered to the gray matter probability map in the Old Segmentation toolbox in SPM (SPM12 v6685, Wellcome Centre for Human Neuroimaging, 2018) by using the normalized mutual information with the EPI time-series kept aligned. The mean EPI was normalized to the template space in SPM MNI 152 (Holmes, C. J., Hoge, R., Collins, L., Woods, R., Toga, A. W., & Evans, A. C., 1998) using the classic Unified Segmentation approach in the Old Segment function in SPM. In order to mitigate overfitting (Overfitting: The model has adapted to too many parameters and noise of the training data, some of which are not relevant for the general prediction of unknown data (James, G., Witten, D., Hastie, T., & Tibshirani, R, 2013)) the warp frequency cutoff was set to 45 to limit the discrete cosine transform (DCT) bases and the sampling distance was set to 2. The resulting deformation field was then applied to the EPI time-series and the mean EPI. In the last step of spatial smoothing, the preprocessed data of each dataset passed through a 5 mm smoothing with a fixed kernel in the MNI152 space.

HCP. As described in the paper of Smith et al. (2013), the data of the HCP are available in raw and preprocessed form. For this study, the preprocessed data were chosen which have the following corrections: In the spatial preprocessing, the data were corrected for spatial distortions, head motions, B_0 distortion and the data is registered to the T1-weighted structural image. Linking the previous corrections with the nonlinear warp field, results in a single warp per timepoint. By applying this single warp to the original time series provides single resampling in the 2 mm MNI space. Furthermore, global intensity normalization was applied and voxels, which don´t belong to the brain, are masked out. Noisy voxels were excluded and the data were regularized with 2 mm Full width at half maximum (FWHM) surface smoothing (Smith et al., 2013, further details in Glasser et al., 2013). Further steps of the temporal preprocessing were for one thing a weak highpass temporal filtering (> 2000s FWHM) and for another thing an Independent Component Analysis (ICA) by using the Multivariate Exploratory Linear Optimized Decomposition into Independent Components method (MELODIC) provided by the FSL toolbox (https://fsl.fmrib.ox.ac.uk/fsl/fslwiki/MELODIC). "MELODIC estimates how many components the given quality and quantity of data will support being separated from each other" (Smith et al., 2013, p. 157). These components were classified by FMRIB´s ICA-based X-noisifier (FIX, Salimi-Khorshidi et al., 2014) and artificial components as well as motion-related timecourses are removed from the data.

1000BRAINS. The RS fMRI data of the 1000BRAINS-dataset were cleaned from noise also by the MELODIC method provided by the FSL toolbox. These denoised data were further pre-processed with SPM12 using Matlab R2014a (Mathworks, Natick, MA). The first four EPI volumes were discarded for each participant. The remaining EPI images were then corrected for head movement by affine registration in a two-step procedure: First of all, the images were aligned to the first image. In the second step, the images were aligned to the mean of all volumes. Then, the resulting mean EPI image was spatially normalized to the MNI152 template using the "unified segmentation" approach to take into account inter-individual differences in the brain morphology (Ashburner & Friston, 2005).

Analysis

Networks. In this study we analyzed the RS functional connectivity in the 12 following task-based networks: autobiographical memory, cognitive action control, emotional scene and face processing, empathic processing, cognitive emotion regulation, mirror neuron system, motor execution, reward-related decision making, semantic memory, theory-of-mind cognition, vigilant attention and working memory. The respective networks were defined in individual meta-analyses and examined together in a study by (Pläschke et al., 2017). For each network, brain regions and their corresponding coordinates were defined in individual meta-analyses, as provided in table 1. In this study, the activation in RS in the whole brain was measured. It was extracted for the previously defined regions of each network and summarized as a time course. These time courses of the activations were correlated with each other resulting in a matrix of correlation values for all of the defined regions of a functional network, what is called the connectome. This procedure was repeated over all of the participants, so that we got a connectome for each participant as input for the further ML-analyses.

In the ML-part of the analysis, the data was split up in 90% as training data and 10% to test the predictions at unknown data. So, a ML-classifier was trained to learn the relationship between the connectivity patterns of the received correlation-matrices of 90% of the participants and their respective sex. For the remaining 10%, the classifier should predict the sex of a person according their connectivity pattern.

This procedure is repeated for 10 times, whereby the sample is always divided up differently into training and test data, which is called 10-fold cross validation (Hawkins, Basak, & Mills, 2003). For each new division, stratified sampling was applied, which means that the samples have the same sex ratio as it is in the whole dataset (Neyman, 1934). This process was in turn repeated 100 times for each individual analysis. The mean value of the 100 iterations was then calculated for the results. This value describes the mean classification accuracy for each analysis, which is the accuracy of the prediction, the classifier reached for the unknown 10% of the data. This information was used to determine how much sex-specific information to differentiate between the sexes was available in each of the 12 functional networks. In order to take into account imbalanced sex ratios, as it was the case in the eNKI, HCP and 1000BRAINS-datasets, the comparison between the predicted and actual sex was calculated separately for men and women and these values were than averaged, so that the overall chance level for each data set was 50%.

In order to ensure that the results were not only based on a specific ML-method, but were independent of it, we applied four different ML-algorithms in this study, which are presented below.

ML-algorithms

Least Absolute Shrinkage and Selection Operator (LASSO). This supervised ML-method offers an automatic feature selection (EliteDataScience.com, 2016-2018) by selecting a subset of variables. LASSO specifies a limit for the sum of the absolute variables, whereby some variables shrink to zero. This identifies the variables, that are the closest to the target variable and the model can be interpreted better (James, G., Witten, D., Hastie, T., & Tibshirani, R, 2013). In the case of this study, it means that the variables, which were most likely to identify sex-specific activity, were used to build a model. As a result, a simple model is obtained, which doesn´t necessarily contain all of the coefficients which leads to an increase of prediction accuracy (Tibshirani, 1996).

Linear Support Vector Machine (LSVM). SVM is a supervised classification technique with a two-class classifier and creates a model based on the input of training data sets with these two classes. This method tries to separate the different classes by the maximum possible wide gap, which is called hyperplane (Rafi & Shahid Shaikh, 2013). LSVM means that a linear hyperplane can be used to separate the dataset into two classes. In the case of this study it means that male datasets were on the one side and female datasets on the other side of the hyperplane. For the unseen data, LSVM tries to predict to which one of the two classes (male or female), so to which side of the hyperplane a person´s dataset can be assigned to according to the previously learned connectomes with the respective sex of a person (Bradley & Mangasarian, 2000).

Ridge Regression (Ridge). Ridge is a method to stabilize regression estimates in case of extreme collinearity (Frank & Friedman, 1993). It provides a solution by a penalized least squares criterion, meaning that feature shrinkage is offered by minimizing the sum of squared residuals, which again leads to smaller coefficients, but they are not forced to zero (Frank & Friedman, 1993; James, G., Witten, D., Hastie, T., & Tibshirani, R, 2013). This ML-method reduces the dimensionality of the factor space to select the subset of predictors which fit the most (Hoerl & Kennard, 1970).

Relevance Vector Machine (RVM). RVM is a statistical learning technique, which is based on Bayesian estimation for learning sparse regression models and classifiers (Saarela M., Elomaa T., Ruohonen K., 2010; Wei, L., Yang, Y., Nishikawa, R. M., & Jiang, Y., 2005). In contrast to SVM, RVM uses probabilistic measure to define the hyperplane. This ML-algorithm yields a solution function based only on a small number of training examples, which are called relevance vectors (Rafi & Shahid Shaikh, 2013).

Results

Datasets

HBN. Considering all the four different ML-algorithms for the HBN-dataset, the classification accuracies ranged between 42.51% and 56.16% (all raw values for each dataset and ML-algorithm are provided in the appendix in table 3-6). Averaging about the algorithms and all the networks, the HBN-dataset reached a mean classification accuracy of 49.58% (*SD* = 3.35%). As can be also seen in figure 1, especially the ML-algorithms LASSO and Ridge mostly showed higher or lower accuracy values for the same networks. The highest mean classification accuracy of all ML-algorithms was achieved by the LSVM-algorithm for this dataset.

eNKI. The highest classification accuracy that could be reached in the eNKI-dataset regarding all ML-algorithms was 71.69% and the lowest 49.98%. The mean classification accuracy of this dataset averaged over the four ML- algorithms was about 62.21% (*SD* = 5.29%). Figure 2 demonstrates some quite similar peaks and low points in the classification accuracies for all the four ML-algorithms for the 12 networks, but Ridge achieved the highest mean classification accuracy of all 4 algorithms.

HCP. In the HCP-dataset, the classification accuracies ranged between 49.73% and 66.03% with regard to all the four ML-algorithms. The mean classification accuracy in this dataset was about 56.42% (*SD* = 4.74%) when averaging about all the 12 networks and ML-algorithms. As can be seen in figure 3, the ML-algorithms LASSO, LSVM and Ridge produced for all networks classification accuracies that were higher than chance level and also showed the same pattern of higher or lower classification accuracies for the same networks, whereas RVM produced for all networks accuracies that ranged close around the 50%-chance level. So, when averaging the four algorithms together, all networks reached a mean classification accuracy that is higher than chance level. The ML-algorithm LSVM reached also in this dataset the highest mean classification accuracy of all algorithms.

1000BRAINS. The classification accuracies in the 1000BRAINS-dataset ranged between 49.73% and 62.03% with regard to all four ML-algorithms. The mean classification accuracy for this dataset was about 53.43% (*SD* = 4.24%). Regarding the four ML-algorithms, LSVM and RVM produced classification accuracies that ranged in most cases close around 50%, whereas Ridge and LASSO showed mostly similar accuracy values for the 12 networks which is also displayed in figure 4. Out of the four ML-algorithms, Ridge produced on average the highest classification accuracy in this dataset.

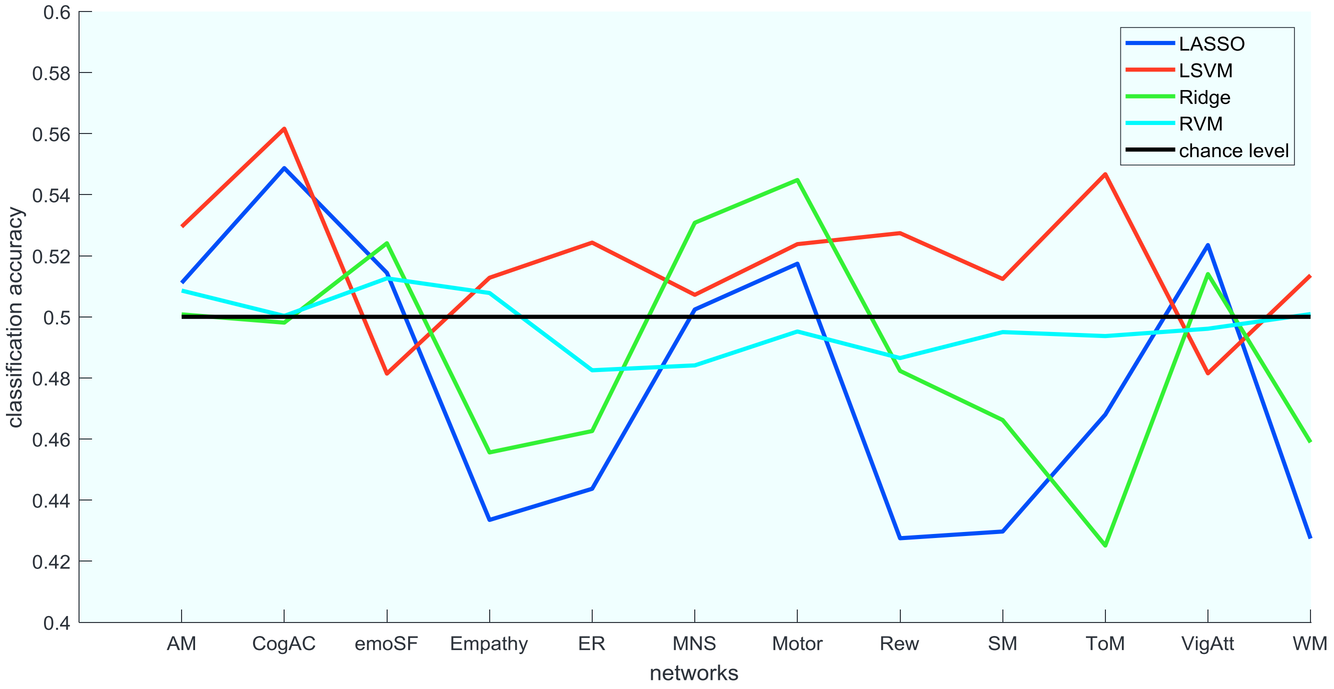

Figure 1. Classification accuracies for the 12 networks in the HBN-dataset for the four different ML-algorithms. All plots are created with Matlab 2017a (Mathworks, Natick, MA).

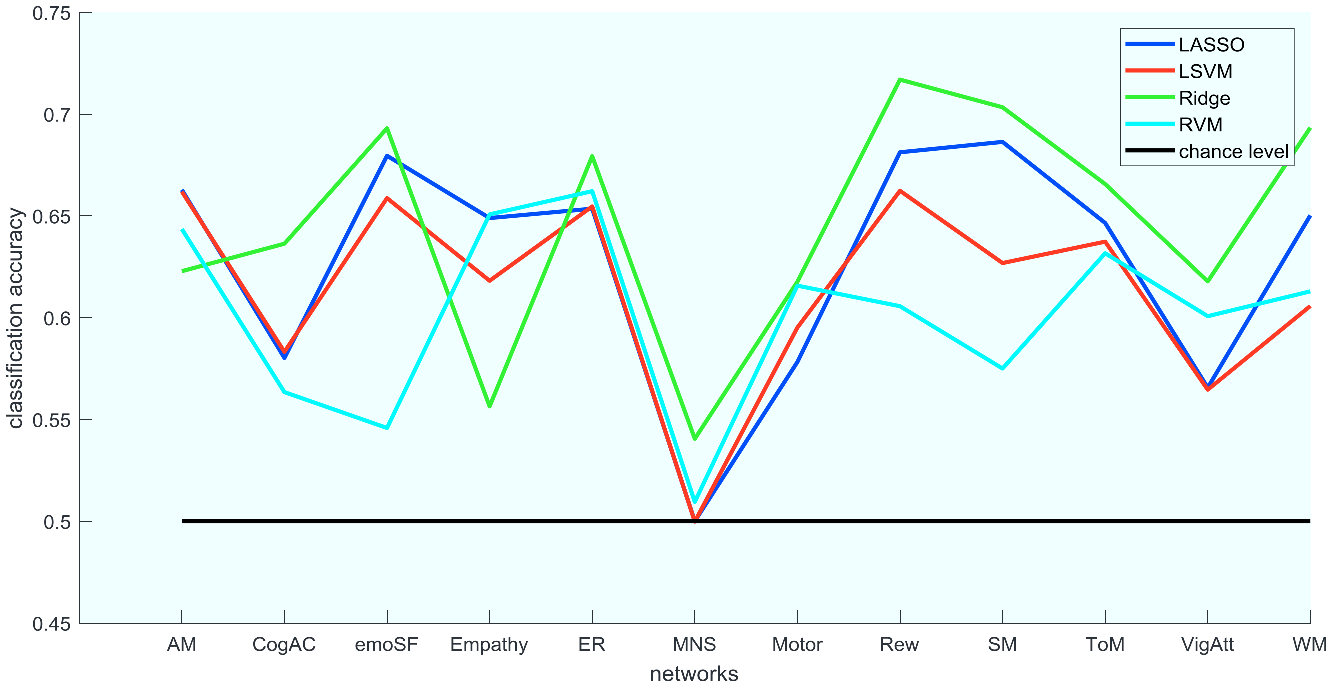

Figure 2. Displayed are the classification accuracies for the 12 networks in the eNKI-dataset for all of the four ML-algorithms.

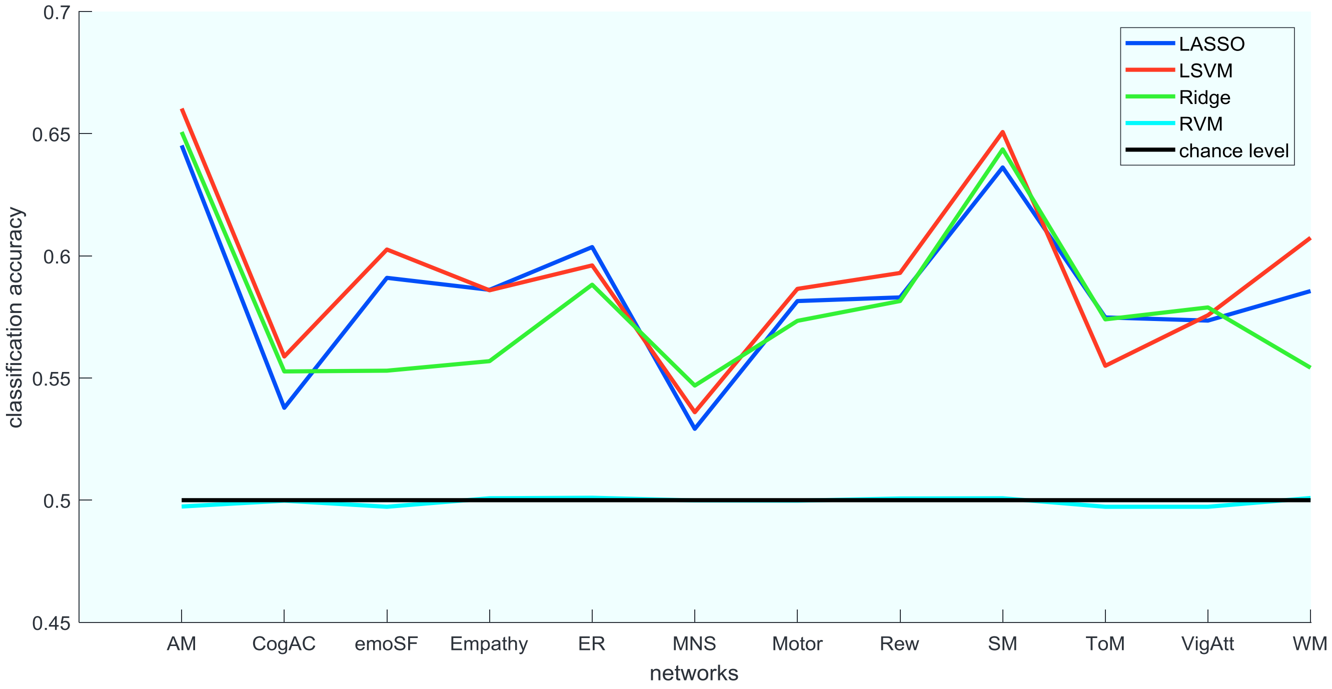

Figure 3. Classification accuracies for the 12 networks in the HCP-dataset for the four different ML-algorithms.

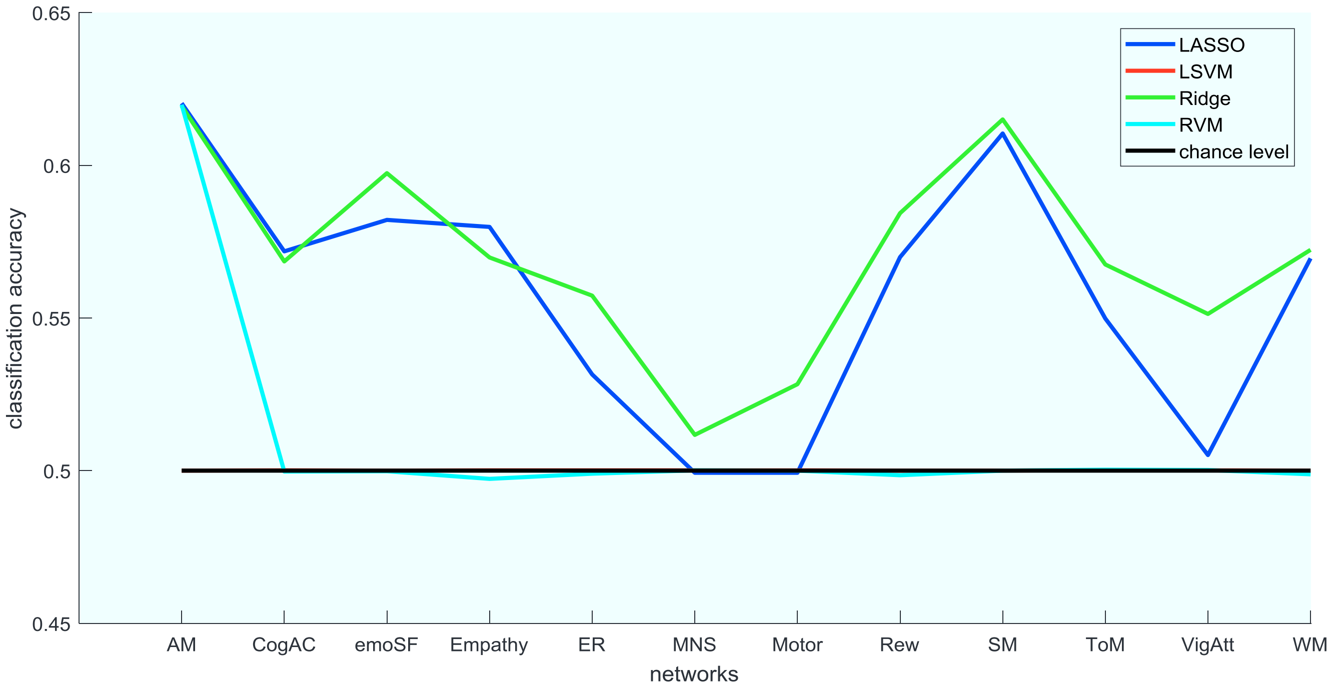

Figure 4. Displayed are the classification accuracies for the 12 networks in the 1000BRAINS-dataset for the four different ML-algorithms.

ML-algorithms

LASSO. The classification accuracies with regard to all four datasets ranged between 42.74% and 68.63% and the mean classification accuracy for LASSO was about 56.24% (*SD* = 6.95%) when averaging about all networks and datasets. LASSO produced on average the highest classification accuracy and the highest variance for the eNKI-dataset with 62.77% (*SD* = 5.56%), slightly lower accuracies for the HCP- (*M* = 58.56%, *SD* = 3.19%) and the 1000BRAINS-dataset (*M* = 55.74%, *SD* = 3.94%) and the lowest accuracy was reached on average for the HBN-dataset (*M* = 47.89%, *SD* = 4.31%). Averaging about the four datasets, the network of autobiographical memory achieved the highest classification accuracy of all networks with 60.99% (*SD* = 5.90%) with this algorithms.

LSVM. Regarding all the four datasets, the highest classification accuracy produced by LSVM was about 66.23% and the minimum was about 48.14%. When averaging about all networks and datasets, LSVM produced a mean classification accuracy of 55.62% (*SD* = 5.71%). This ML-method produced the highest classification accuracy and variance for the eNKI-dataset (*M* = 61.40%, *SD* = 4.64%) and also a relatively high accuracy for the HCP-dataset (*M* = 59.23%, *SD* = 3.47%). The values of classification accuracy and variance were lower for the HBN- (*M* = 51.85%, *SD* = 2.21%) and the 1000BRAINS-dataset (*M* = 50.01%, *SD* = 0.01%). Also for this ML-algorithm, autobiographical memory (*M* = 58.79%, *SD* = 7.39%) achieved the highest classification accuracy of all networks when averaging about the four datasets.

Ridge. The highest classification produced by the ML-method Ridge was about 71.69% when considering all four datasets, whereas the minimum was 42.51%. When averaging about the four datasets, the mean classification accuracy produced by Ridge was about 57.09% (*SD* = 6.81%). Also in this case, the highest classification accuracy and variance were in the eNKI-dataset (*M* = 64.52%, *SD* = 5.45%). For the HCP- (*M* = 57.95%, *SD* = 3.29%) and the 1000BRAINS-dataset (*M* = 57.02%, *SD* = 3.04%), Ridge produced nearly similar values for classification accuracy and variance. For the HBN-dataset, Ridge produced on average the lowest classification accuracy (*M* = 48.86%, *SD* = 3.45%). The network that reached on average the highest classification accuracy in this ML-method was semantic memory with 60.70% (*SD* = 8.73%).

RVM. This ML-method produced classification accuracies in a range from 48.25% to 66.21% with regard to all the datasets. When averaging about the datasets and the networks, RVM produced a mean classification accuracy of 52.68% (*SD* = 5.15%). RVM produced the highest classification accuracy for the eNKI-dataset (*M* = 60.14%, *SD* = 4.36%), as well as the other algorithms did. However, the classification accuracies for the HBN- (*M* = 49.69%,

SD = 0.92%), HCP- (*M* = 49.94%, *SD* = 0.15%) and 1000BRAINS-dataset (*M* = 50.95%, *SD* = 3.32%) ranged close around 50%. With regard to the networks, the highest classification accuracy that was produced on average for the autobiographical memory with 56.73% (*SD* = 6.50%).

Networks

Under consideration of the four datasets and the four ML- algorithms, autobiographical and semantic memory were the networks which had the highest classification accuracies of all networks, further information are displayed in table 2.

Table 2

Classification accuracies of the 12 networks

	M	SD
networks:		
AM	59.09%	6.61%
CogAC	54.76%	3.83%
emoSF	56.45%	6.58%
Empathy	54.75%	6.21%
ER	55.87%	7.43%
MNS	51.23%	1.77%
Motor	54.73%	4.32%
Rew	56.25%	7.72%
SM	57.20%	8.23%
ToM	55.21%	6.69%
VigAtt	54.04%	4.15%
WM	55.32%	6.96%

Note. Displayed are the mean classification accuracies for each of the 12 networks averaged over the four datasets and the four ML- algorithms.

Discussion

The purpose of the present study was to investigate whether the RS functional connectivity in certain brain networks that are known for sex-specific processing strategies differs also according to sex. With an ML-approach we examined whether the RS functional connectivity within the brain networks carried enough information to accurately predict the sex of a person. For this purpose, the RS connectivity patterns in 12 different meta-analytically defined networks in four datasets were extracted and analyzed with four different ML-algorithms. Many studies already provided evidence for sex-specific processing strategies in domains dealing with emotions (Sacher et al., 2013). Therefore, higher classification accuracies were expected especially in emotion-related networks.

Higher classification accuracies and the highest variance were attained in the eNKI-dataset, the HCP-dataset showed on average the second-highest accuracies. The classification accuracies were on average lower for the 1000BRAINS-dataset and the HBN-dataset provided the lowest mean accuracies of all datasets. With regard to the four different ML-algorithms, LASSO and Ridge provided on average the highest classification accuracies and a higher variance. LSVM achieved similar high classification accuracies on average, whereas the accuracies for RVM were lower in contrast to the other three algorithms. With regard to the 12 networks, two of them could be highlighted in particular: autobiographical and semantic memory showed the highest classification accuracies on average within the four datasets and the four ML-algorithms. Altogether, the hypothesis of higher classification accuracies in emotion-related networks cannot be supported by the present results. Instead, the results were strongly dependent on the datasets and algorithms. Different possibilities, how these results could have been achieved, are discussed in the following:

Datasets. First, concerning the differing results in each of the datasets, an effect of age might be involved here. Although the analysis was controlled for the variable age, this effect might be still so strong that it influenced the results. In the eNKI- and HCP-dataset, the accuracies were higher than in the other datasets and these two datasets had in common, that their mean age was middle-aged in contrast to the others. Therefore, it might be hypothesized that the sex-specific information the basis of which the ML-algorithms create their model is most readily available in middle-aged participants. For the 1000BRAINS-dataset, the mean classification accuracy was slightly lower than for the other two mentioned before. The reason may be that this dataset contains less middle-aged and more elderly people because it was a study to investigate the variability of the human brain during aging (Caspers et al., 2014). According to the fact, that the time of adolescence and also aging is associated with considerable changes in the brain (Blakemore & Choudhury, 2006; Tomasi & Volkow, 2012), it can be assumed that the functional connectivity patterns, that differentiate well between men and women, are developed to a full extend only in middle-aged persons. Therefore, it seems reasonable that the classification accuracies in the 1000BRAINS-dataset are not as high as in the eNKI- and HCP-dataset. However, in the HBN-dataset, there were on the one hand only relatively young participants, who were all under 20 years old. Since the lowest mean classification accuracies were achieved in this dataset, it would agree with the assumption that the sex-specific information are not yet available at this age to the extent that it is possible to differentiate sufficiently between men and women. On the other hand, the HBN-dataset was the one with the lowest number of participants of all four datasets. So, another possible explanation for the lowest classification accuracies in this dataset might be that the ML-algorithms had not enough information to learn the relationship between sex-specific features and the respective sex of a person, because of the lower number of subjects.

Consequently, the algorithms are not able to differentiate so well between men and women which have then led to the low accuracies in the HBN-dataset.

ML-algorithms. Second, the four different ML-algorithms did not all show a uniform picture regarding the classification accuracies for the 12 networks. In all four datasets, LASSO and Ridge showed a similar pattern for the 12 networks in form of similarly high or low classification accuracies for the same networks. Both algorithms have in common that they shrink regression coefficients, so that they identify a subset of variables that fit the most to the target variable, which means that both algorithms build sparse models (James, G., Witten, D., Hastie, T., & Tibshirani, R, 2013). Since these two algorithms achieved on average the highest classification accuracies of all four ML-algorithms, it seem that the sparse models are quite suitable for this type of classification task. In contrast to Ridge and LASSO, the algorithms LSVM and RVM had in common that they separate the two classes of sex by the widest possible gap, the hyperplane (Rafi & Shahid Shaikh, 2013). However, it is not possible to state that these approaches are generally worse than the algorithms that built sparse models, because LSVM achieved on average the highest classification accuracies in two of the four datasets. Altogether, it can be said that the sparse algorithms seem quite suitable for the classification task, but they were not the best algorithms in all cases of the datasets. This empathizes the ‘no free lunch theorem’, so also these results display that not a single algorithms works best for all possible datasets as already stated by James, G., Witten, D., Hastie, T., & Tibshirani, R (2013).

Networks. Third, the results according to the networks were different than hypothesized: There were no generally higher classification accuracies in the emotion-related networks and the motor-network. Based on the results, it can be concluded that the sex-specific processing strategies, which were reported for the emotion-related domains, do not manifest themselves in general in the underlying RS functional connectivity or at least it could not be demonstrated in this study. Instead, the results displayed the on average highest classification accuracies for the networks of autobiographical and semantic memory. For the autobiographical memory, some studies already found sex differences e.g. in neural recruitment (Piefke & Fink, 2005; Young et al., 2013). Therefore it was already hypothesized that the classification accuracies in this network should display a higher classification accuracy. However, it was expected that this would go along with higher classification accuracies also in other emotion-related networks. Instead, also the semantic memory displayed higher classification accuracies which was not hypothesized in the introduction. So, it seems that there is another component instead of the emotional component, which shows particularly sex-specific patterns that make it possible to differentiate well between males and females. Both cognitive domains, the autobiographical and semantic memory have in common that they belong to the declarative memory system (Renoult, Davidson, Palombo, Moscovitch, & Levine, 2012). Moreover, both memory system are acquired through personal experience as mentioned in the introduction: Whereas the cognitive domain of autobiographical memory is defined as the episodic memory for personal events in one´s own life (Spreng, R. N., Mar, R. A., & Kim, A. S., 2009; Young et al., 2013), semantic memory describes the knowledge about people, objects, culture etc., which was acquired through own experience (Binder et al., 2009). Since this component is present in both networks, it can be assumed that there are sex differences in the way that knowledge about these situations is stored in long-term memory. As already stated by Piefke and Fink (2005), sex-specific cognitive styles might exist regarding encoding, rehearsing and thinking about personal experiences, although the authors only referred to emotionally laden personal experiences. This assumption would also agree with the results of the study by Maitland, S. B., Herlitz, A., Nyberg, L., Bäckman, L., & Nilsson, L. G. (2004), who could provide evidence for sex-dependent differences in the performance for episodic and the semantic memory as well as for their subfactors e.g. recognition and recall. Taken together, the results provided evidence for sex-specific processing strategies in cognitive domains that belong to the declarative memory. Since the analysis of their underlying RS functional connectivity patterns resulted in the highest classification accuracies of all the networks that were investigated.

Considering the accuracy values of the results in general, it also needs to be mentioned that the classification accuracies did not reach the high level of classification accuracies that was reached in the study by Zhang et al. (2018) who also used an ML-approach for sex classification. In contrast to the present study, Zhang et al. (2018) examined the functional connectivity of the whole brain to classify sex and to identify the features of functional connectivity that are most predictive for sex. For this purpose, they investigated the RS fMRI data of 820 subjects of the HCP-dataset with a partial least squares regression modeling approach. They achieved classification accuracies of 80% and higher for sex classification based on RS-fMRI functional connectivity data. The present study, in turn, investigated the RS functional connectivity in predefined networks to predict sex, not the whole brain connectivity. Another difference to the study by Zhang et al. (2018) is that the present study employed different ML-algorithms and apart from the HCP-datasets also further datasets were used. Whereas the results of Zhang et al. (2018) could clearly point out the existence of sex differences in brain connectivity, the results of the present study could not show this in such a clear way because the classification accuracies ranged in most cases between 50% and 60%.

One possible explanation that the classification accuracies in the present were not as high as in the study by Zhang et al. (2018) might be that the functional connectivity within predefined networks was examined instead of the whole brain connectivity. The functional connectivity of the whole brain contains more information than the connectivity within a certain network. So, the amount of information to classify sex was limited in the present study. It might be that the information to predict the sex of a person are very complex and dispensed in the brain. For this reason, the networks may partially not contain sufficient information for sex classification, as these are more likely to be captured by the whole brain connectivity. This fact might have led to the overall relatively low classification accuracies. Still, the results provided indications which networks contain enough sex-specific information to classify sex with a better performance than chance level as pointed out before.

Another possible reason for these results might be that the 12 networks are based on task-related studies, whereas the present study examined the underlying RS functional connectivity of several networks. On the one hand, it was already mentioned in the introduction that the ML-approach is different to the task-based studies on which the hypotheses were based. Therefore, it is quite possible that the classification approach leads to different results than were expected according to these task-based group comparisons reported in the introduction. On the other hand, it can be assumed that RS functional connectivity shows different patterns of connectivity at different neural areas than the activation that is shown in relation to a specific task. Consequently, sex-specific differences in RS functional connectivity that can lead to a good distinction between men and women may not necessarily show up in the defined neural networks. The same networks as were investigated in the present study were already examined by Pläschke et al. (2017) for the phenotypes of certain diseases and age. This study also achieved higher classification for these phenotypes than the present study for the phenotype of sex. As a consequence, it can be hypothesized that the functional connectivity in these certain networks is more likely to reflect phenotypes like age and disease. The functional connectivity that might be predictive for sex in turn might be too complex to be covered by a single network. This in turn, may have led to the present classification accuracies that ranged mostly between 50% and 60% in the present study.

Another possible explanation for the fact that the distinction between men and women was not as good to reach high classification accuracies might be a greater variance within one sex than between the sexes in several cognitive domains. If at least one of the sexes shows a greater variability in the RS-connectivity patterns of the networks it is hard to impossible to distinguish between males and females. If this was the case in the present study it might be a good explanation for why the classification accuracies were partly just above chance level and as a consequence the distinction between men and women did not work very well. This explanation would be in line with a hypothesis that was already stated in 1894 by Ellis (quoted from Feingold, 1992): Males were intellectually more variable than females and Ellis also concluded that this greater male variability was innate. This assumption has been examined several times and Feingold (1992) could prove this in his study only partially: Males showed in his study a greater variability in quantitative and spatial abilities but not in verbal abilities. However, in a cross-cultural quantitative review by Feingold (1994) it could be pointed out that the findings of Feingold (1992) were mostly dependent of the culture because the variance ratios were not consistent across different countries. The finding of greater male variability in mathematical and spatial abilities is well-established only for the US population, but this is represented in three (HBN-, eNKI- and HCP-dataset) of the four datasets that were used in this study.

However, the studies mentioned before tested this hypothesis only for the three cognitive abilities that are related to language, mathematics and spatial abilities. In consideration of the results of the present study, it can also be assumed that different variability between the sexes also manifests itself in other cognitive domains, such as domains related to emotions or attention. If this greater variability in the cognitive domains also applies to the underlying RS-connectivity patterns of a cognitive domain, it can be another explanation for the results of the present study: If either the males or the females showed a greater variability than the other sex, then the distinction between men and women do not work so well. As a consequence the classification accuracies are not so high, as it was the case in the present study.

It might also be important to note that this study only examined the relationship between the RS-connectivity patterns of a person in specific networks and their respective sex and not their gender identity. The terms ‘sex’ and ‘gender’ are often used synonymously, but there are decisive differences: The ‘sex’ of a person describes the biological status of a person according to their anatomical characteristics as either ‘male’ or ‘female’, but the ‘gender’ of a person is related to socially constructed roles and cultural representations (Diamond, 2002; Newman, 2002). ‘Gender roles’ in turn refer to socially characteristics and expectations what is associated with being male or female which means a certain behavior pattern according to a social script (Diamond, 2002; Newman, 2002). The gender roles are of course influenced by culture. In most cases, male roles are associated with strength or the financial provision of the family, while female roles are more concerned with taking care of the family. The terms of sex and gender identity do not necessarily have to agree with each other, which becomes apparent in phenomena such as intersexuality or transsexuality (Diamond, 2002). These terms describe on the one hand the possibility of ambiguous biological structures of the sex of a person. On the other hand it also describes people that do not identify themselves with their biological sex and the respective gender roles and rather feel like the opposite kind of sex.

According to Newman (2002), the ‘brain sex’ of a person is the foundation of gender in social and cultural forms. In consideration of this statement, it can be supposed that the brain reflects more the ‘gender identity’ than the biologically based ‘sex’ of a person. Otherwise, it would not be possible that phenomena as transsexuality are possible to occur. For this reason, it can speculated that the RS connectivity patterns of a network show more specialized patterns according to the gender identity of a person than for their biological sex. Therefore, it should be considered to take the gender identity as a further variable of interest into account, in addition to the biological sex. This has been already done for example in a study by Bourne and Maxwell (2010): Besides the biological sex also the psychological gender identity was examined and the authors could outline an interaction between these two variables. Taking both variables into account might also help for a better understanding of gender variation according to cultural variations and in the broader community. This can also help to improve social tolerance (Newman, 2002).

Another approach by Joel et al. (2015) might yield a different perspective on the present results but also corresponds to the explanations discussed before: In their approach of the human brain mosaic, Joel et al. (2015) proposed that the brain is not dimorphic in the meaning of not having two distinct categories of ‘males’ and ‘females’. This approach does not deny the existence of sex differences on the group level; but in their investigations the authors could outline a great overlap between the distributions of men and women for gray and white matter as well as for connectivity measures (Joel & Fausto-Sterling, 2016). Furthermore, internal consistency in form of a brain having only ‘male’ or only ‘female’ features was lower than variability. These findings fit to studies that proposed that women as well as men may have ‘typical male’ and ‘typical female’ characteristics. Consequently, Joel et al. (2015) suggested a shift from the dimorphic view of the brain to a perspective that sees each human brain as a unique mosaic of features. Some of these features are more common in females in contrast to males, some features are more common in males and some features as pieces of the whole mosaic are common in males as well as in females. So the authors see the brain more multi-morphic than dimorphic. The authors also stated that the brain is highly variable in form of an heterogeneous mosaic of ‘male’ and ‘female’ characteristics that is always changing (Joel, 2011). These changes might be also modulated by social influences and certain gender roles. This multi-morphic approach fits well to the explanatory approach discussed before, namely that the complexity of the phenotype sex is more likely to be represented in the whole connectivity of the brain. According to the authors, this ever-changing variability is generated by a complex interaction of genetic and environmental factors as well as hormonal influences. Again, there can be drawn a connection to a previously discussed approach; namely that the variability within one sex may be higher than between the sexes, which in turn may have led to the lower accuracies. As reported by Joel et al. (2015), the sex of a person would be enough to predict whether a person have more 'typical females' or more 'typical males', it would not be possible to predict the specific combination this person have in the 'typical male' and 'typical female' mosaics. In another article, Joel and Fausto-Sterling (2016) also stated that the sex differences which might be found in brain structure are not necessarily transferable to functional differences, so that brain function would not be necessarily characterized by sexual dimorphism.

On the one hand, this statement of Joel and Fausto-Sterling (2016) contrasts with our original assumption mentioned in the introduction that the functional connectivity in a network carries enough sex-specific information to accurately predict the sex of a person in a classification approach. However, since the results could not portray an overall accurate prediction of sex, this would show support for the statement of Joel and Fausto-Sterling (2016) that the sex differences in the cognitive domains of the networks investigated in this study are not transferable to brain function.

On the other hand, the results showed that the four datasets and ML-algorithms produced differently high accuracies and for example the eNKI-dataset and the Ridge-algorithm can produce accuracies that are definitely higher than chance level. For the one thing, this indicates that the classification accuracies are highly dependent on the datasets and the ML-algorithms. Then again, it can also be speculated that in some datasets is a higher variability in the human brain mosaics than a clear distinction between persons carrying more ‘typical male’ or ‘typical female’ features. This higher variability would result in the ML-algorithms not being able to differentiate so well between men and women, which would explain the partially low classification accuracies near chance level in the HBN-dataset for example. According to Joel (2011), the variability is created by several factors like genetics or hormones. Since all subjects in the HBN-dataset were relatively young, meaning under 20 years, it can be assumed that in this dataset the hormones during puberty might produce a great variability in their functional brain patterns, because it is already known that there are considerable changes in brain function during adolescence (Sturman & Moghaddam, 2011). This would make it more difficult to distinguish between the sexes, which is reflected again in the low classification accuracies of the HBN-dataset. So, this perspective would explain why this dataset produced the lowest accuracies of all datasets.

This approach can also be used to explain the classification accuracies in the 1000BRAINS-dataset, where the ML-algorithms LSVM and RVM produced accuracies near chance level: The subjects in this dataset were on average relatively old, especially in comparison to the remaining three datasets. According to Geerligs, Renken, Saliasi, Maurits, and Lorist (2014) there are changes within and between functional networks in the brain with increasing age and also structural changes. The age range Geerligs et al. (2014) investigated in their study is nearly similar to the age range of the 1000BRAINS-dataset. Therefore, it can be assumed that these age-related changes also lead to a larger variability in sex-specific features, which would make it again more difficult for the ML-algorithms to distinguish between the sexes. Consequently, the classification accuracies were partly not higher than chance level. Since the eNKI- and HCP-datasets could produce on average higher classification accuracies than the HBN- and 1000BRAINS-dataset, it can be assumed that in these datasets that contain mostly middle-aged persons, there are fewer factors like hormones and environmental influences that can produce more variability in the sex-specific features in the brain. As a consequence, the distinction between men and women could have been achieved on a higher level, which would lead to the higher classification accuracies in the eNKI- and HCP-dataset in contrast to the HBN- and 1000BRAINS-dataset.

This approach can also explain why the ML-algorithms LASSO and Ridge produced on average higher classification accuracies than LSVM and RVM: As mentioned before, LASSO and Ridge seem to be better suited for this classification approach with their sparse models, because they do not focus on the entire variability. In contrast, LSVM and RVM try to separate the two sexes with the hyperplane. As suggested by Joel et al. (2015), there is rather a large overlap of the distributions of men and women in several features and less the two distinct dimorphic classes ‘male’ and ‘female’. So, the working basis for the algorithms LSVM and RVM would be missing or is only slightly available according to the approach of the human brain mosaic. For this reason, it can explain how the on average lower classification accuracies of these algorithms might have come about.

Nevertheless, the results provided indications for sex-specific processing strategies in the RS functional connectivity patterns of the networks of autobiographical and semantic memory. Because these networks achieved the overall highest classification accuracies of all investigated networks. Also this part of the results can be linked to the approach of the human brain mosaic by Joel et al. (2015): It might be assumed that these two networks represent these pieces of the mosaic that show more sex-specific features than the other networks. Consequently, in these networks the two sexes could be easier to distinct which then leads to the higher classification accuracies that are displayed in the results.

Joel and Fausto-Sterling (2016) also posed the question whether sex should still be taken up as a variable of interest in future studies if the brain cannot be seen in the dimorphic classes and more as a specific composition of ‘typical male’ and ‘typical female’ features according to the approach of the human brain mosaic. First, the authors said by themselves that the variable sex might correlate with other variables and therefore it should still be taken up to clarify variance. Second, the results of the present showed that the distinction and classification between men and women worked differently well: We saw that the choice of a certain dataset or ML-algorithm might influence the classification accuracies. Additionally, the results also showed that in certain memory networks men and women may be better classified than for example in the hypothesized emotion-related networks. For this reason, it can be suggested that sex should still be used in future studies as a variable of interest. This might become also relevant to improve the understanding how sex may influence the brain also in clinical aspects, since sex differences may also play a role in the occurrence of certain diseases as stated for example by Bijl, Graaf, Ravelli, Smit, and Vollebergh (2002). As suggested before, it should generally be considered to take sex and also the gender identity as variables of interest into account in future studies to improve the understanding of the influence of these variables.

To sum it up, the results of the present study did not support the hypotheses of higher classification accuracies in especially emotion-related networks. Instead, the classification accuracies were generally not so high as in another study that also employed a ML-approach for sex classification. A possible explanation might be that this study investigated whole-brain connectivity whereas the present study examined the connectivity within specific networks. Another possible reason can be that the RS-functional connectivity patterns which were analyzed in this study are not so easily to compare to the task-based group comparisons on which the hypotheses were built on. Another explanation might be that this ML-classification approach was another approach than the group comparisons that were conducted in the past. Therefore it is not inconceivable that different results were received with a different approach. Moreover, a higher variability within than between the sexes could have led to a worse distinction and consequently to lower classification accuracies. This explanation goes hand in hand with the idea of the human brain mosaic that suggests a greater variability in form of a greater overlap between 'male' and 'female' features than a clear distinction between the two classes. In line with this approach, we also considered the idea, that gender identity might be more represented in the brain than the biological sex of a person. It is possible to receive other results when considering the gender identity as variable of interest which can be a suggestion for future studies. Still, the results provided indications for sex-specific processing strategies in the networks of autobiographical and semantic memory. This suggests that their common component of personal experience is particularly processed in sex-specific ways that leads to a good distinction between men and women.

Overall, it can be generally outlined that although the present study examined in all cases RS-functional connectivity from fMRI-studies, the results showed a strong dependency of the respective datasets and ML-algorithms. This dependency should be taken into account for future studies, because the present study showed how different the results may be according to possible age or variability effects.

References

Alexander, L. M., Escalera, J., Ai, L., Andreotti, C., Febre, K., Mangone, A., . . . Milham, M. P. (2017). *An open resource for transdiagnostic research in pediatric mental health and learning disorders*. https://doi.org/10.1101/149369

Amunts, K., Jäncke, L., Mohlberg, H., Steinmetz, H., & Zilles, K. (2000). Interhemispheric asymmetry of the human motor cortex related to handedness and gender. *Neuropsychologia*, *38*(3), 304–312. https://doi.org/10.1016/S0028-3932(99)00075-5

Ashburner, J., & Friston, K. J. (2005). Unified segmentation. *NeuroImage*, *26*(3), 839–851. https://doi.org/10.1016/j.neuroimage.2005.02.018

Batty, M., & Taylor, M. J. (2006). The development of emotional face processing during childhood. *Developmental Science*, *9*(2), 207–220. https://doi.org/10.1111/j.1467-7687.2006.00480.x

Baxter, L. C., Saykin, A. J., Flashman, L. A., Johnson, S. C., Guerin, S. J., Babcock, D. R., & Wishart, H. A. (2003). Sex differences in semantic language processing: A functional MRI study. *Brain and Language*, *84*(2), 264–272. https://doi.org/10.1016/S0093-934X(02)00549-7

Bernhardt, B. C., & Singer, T. (2012). The neural basis of empathy. *Annual Review of Neuroscience*, *35*, 1–23. https://doi.org/10.1146/annurev-neuro-062111-150536

Bijl, R. V., Graaf, R. de, Ravelli, A., Smit, F., & Vollebergh, W. A. M. (2002). Gender and age-specific first incidence of DSM-III-R psychiatric disorders in the general population. Results from the Netherlands Mental Health Survey and Incidence Study (NEMESIS). *Social Psychiatry and Psychiatric Epidemiology*, *37*(8), 372–379. https://doi.org/10.1007/s00127-002-0566-3

Binder, J. R., Desai, R. H., Graves, W. W., & Conant, L. L. (2009). Where is the semantic system? A critical review and meta-analysis of 120 functional neuroimaging studies. *Cerebral Cortex (New York, N.Y. : 1991)*, *19*(12), 2767–2796. https://doi.org/10.1093/cercor/bhp055

Blakemore, S.-J., & Choudhury, S. (2006). Development of the adolescent brain: Implications for executive function and social cognition. *Journal of Child Psychology and Psychiatry, and Allied Disciplines*, *47*(3-4), 296–312. https://doi.org/10.1111/j.1469-7610.2006.01611.x

Bosacki, S. L. (2000). Theory of mind and self-concept in preadolescents: Links with gender and language. *Journal of Educational Psychology*, *92*(4), 709–717. https://doi.org/10.1037//0022-0663.92.4.709

Bourne, V. J. (2005). Lateralised processing of positive facial emotion: Sex differences in strength of hemispheric dominance. *Neuropsychologia*, *43*(6), 953–956. https://doi.org/10.1016/j.neuropsychologia.2004.08.007

Bourne, V. J., & Maxwell, A. M. (2010). Examining the sex difference in lateralisation for processing facial emotion: Does biological sex or psychological gender identity matter? *Neuropsychologia*, *48*(5), 1289–1294. https://doi.org/10.1016/j.neuropsychologia.2009.12.032

Bradley, P. S., & Mangasarian, O. L. (2000). Massive data discrimination via linear support vector machines. *Optimization Methods and Software*, *13*(1), 1–10. https://doi.org/10.1080/10556780008805771

Buckner, J. P., & Fivush, R. (1998). Gender and self in children's autobiographical narratives. *Applied Cognitive Psychology: The Official Journal of the Society for Applied Research in*

Memory and Cognition, *12*(4), 407–429. https://doi.org/10.1002/(SICI)1099-0720(199808)12:4<407::AID-ACP575>3.0.CO;2-7

Buhle, J. T., Silvers, J. A., Wager, T. D., Lopez, R., Onyemekwu, C., Kober, H., . . . Ochsner, K. N. (2014). Cognitive reappraisal of emotion: A meta-analysis of human neuroimaging studies. *Cerebral Cortex*, *24*(11), 2981–2990. https://doi.org/10.1093/cercor/bht154

Burman, D. D., Bitan, T., & Booth, J. R. (2008). Sex differences in neural processing of language among children. *Neuropsychologia*, *46*(5), 1349–1362. https://doi.org/10.1016/j.neuropsychologia.2007.12.021

Bzdok, D., Schilbach, L., Vogeley, K., Schneider, K., Laird, A. R., Langner, R., & Eickhoff, S. B. (2012). Parsing the neural correlates of moral cognition: Ale meta-analysis on morality, theory of mind, and empathy. *Brain Structure & Function*, *217*(4), 783–796. https://doi.org/10.1007/s00429-012-0380-y

Calero, C. I., Salles, A., Semelman, M., & Sigman, M. (2013). Age and gender dependent development of Theory of Mind in 6- to 8-years old children. *Frontiers in Human Neuroscience*, *7*, 281. https://doi.org/10.3389/fnhum.2013.00281

Caspers, S., Moebus, S., Lux, S., Pundt, N., Schütz, H., Mühleisen, T. W., . . . Amunts, K. (2014). Studying variability in human brain aging in a population-based German cohort-rationale and design of 1000BRAINS. *Frontiers in Aging Neuroscience*, *6*, 149. https://doi.org/10.3389/fnagi.2014.00149

Caspers, S., Zilles, K., Laird, A. R., & Eickhoff, S. B. (2010). Ale meta-analysis of action observation and imitation in the human brain. *NeuroImage*, *50*(3), 1148–1167. https://doi.org/10.1016/j.neuroimage.2009.12.112

Charman, T., Ruffman, T., & Clements, W. (2002). Is there a Gender Difference in False Belief Development? *Social Development*, *11*(1), 1–10. https://doi.org/10.1111/1467-9507.00183

Cheng, Y., Chou, K.-H., Decety, J., Chen, I.-Y., Hung, D., Tzeng, O. J.-L., & Lin, C.-P. (2009). Sex differences in the neuroanatomy of human mirror-neuron system: A voxel-based morphometric investigation. *Neuroscience*, *158*(2), 713–720. https://doi.org/10.1016/j.neuroscience.2008.10.026

Cheng, Y. W., Tzeng, O. J., Decety, J., Imada, T., & Hsieh, J. C. (2006). Gender differences in the human mirror system: a magnetoencephalography study. *Neuroreport*, *17*(11), 1115–1119.

Cheng, Y., Decety, J., Lin, C. P., Hsieh, J. C., Hung, D., & Tzeng, O. J. (2007). Sex differences in spinal excitability during observation of bipedal locomotion. *Neuroreport*, *18*(9), 887–890.

Cheng, Y., Lee, P. L., Yang, C. Y., Lin, C. P., Hung, D., & Decety, J. (2008). Gender differences in the mu rhythm of the human mirror-neuron system. *PloS One*, *3*(5), e2113. https://doi.org/10.1371/journal.pone.0002113

Cieslik, E. C., Mueller, V. I., Eickhoff, C. R., Langner, R., & Eickhoff, S. B. (2015). Three key regions for supervisory attentional control: Evidence from neuroimaging meta-analyses. *Neuroscience and Biobehavioral Reviews*, *48*, 22–34. https://doi.org/10.1016/j.neubiorev.2014.11.003

Davies, P. L., & Rose, J. D. (2000). Motor skills of typically developing adolescents: awkwardness or improvement? *Physical & Occupational Therapy In Pediatrics*, *20*(1), 19–42. https://doi.org/10.1080/J006v20n01_03

Derntl, B., Finkelmeyer, A., Eickhoff, S., Kellermann, T., Falkenberg, D. I., Schneider, F., & Habel, U. (2010). Multidimensional assessment of empathic abilities: Neural correlates and gender differences. *Psychoneuroendocrinology*, *35*(1), 67–82. https://doi.org/10.1016/j.psyneuen.2009.10.006

Diamond, M. (2002). Sex and Gender are Different: Sexual Identity and Gender Identity are Different. *Clinical child psychology and psychiatry*, *7*(3), 320–334.

Dittmar, M. L., Warm, J. S., Dember, W. N., & Ricks, D. F. (1993). Sex differences in vigilance performance and perceived workload. *The Journal of General Psychology*, 309–322.

Dorfberger, S., Adi-Japha, E., & Karni, A. (2009). Sex differences in motor performance and motor learning in children and adolescents: An increasing male advantage in motor learning and consolidation phase gains. *Behavioural Brain Research*, *198*(1), 165–171. https://doi.org/10.1016/j.bbr.2008.10.033

Dreher, J.-C., Schmidt, P. J., Kohn, P., Furman, D., Rubinow, D., & Berman, K. F. (2007). Menstrual cycle phase modulates reward-related neural function in women. *Proceedings of the National Academy of Sciences of the United States of America*, *104*(7), 2465–2470. https://doi.org/10.1073/pnas.0605569104

EliteDataScience.com. (2016-2018). Data Science Primer: Chapter 5: Algorithm Selection in Machine Learning - Data Science Primer. Retrieved from https://elitedatascience.com/algorithm-selection

Ellis, H. (1916). *Man and woman: A study of human secondary sexual characters*. Walter Scott Publishing Company.

Feingold, A. (1992). Sex Differences in Variability in Intellectual Abilities: A New Look at an Old Controversy. *Review of Educational Research*, *62*(1), 61–84.

Feingold, A. (1994). Gender differences in variability in intellectual abilities: A cross-cultural perspective. *Sex roles*, *30*(1-2), 81–92.

Filkowski, M. M., Olsen, R. M., Duda, B., Wanger, T. J., & Sabatinelli, D. (2017). Sex differences in emotional perception: Meta analysis of divergent activation. *NeuroImage*, *147*, 925–933. https://doi.org/10.1016/j.neuroimage.2016.12.016

Frank, I. E., & Friedman, J. H. (1993). A Statistical View of Some Chemometrics Regression Tools. *Technometrics*, *35*(2), 109–135.

Gallagher, H. L., & Frith, C. D. (2003). Functional imaging of ‘theory of mind’. *Trends in Cognitive Sciences*, *7*(2), 77–83. https://doi.org/10.1016/S1364-6613(02)00025-6

Geerligs, L., Renken, R. J., Saliasi, E., Maurits, N. M., & Lorist, M. M. (2014). A Brain-Wide Study of Age-Related Changes in Functional Connectivity. *Cerebral Cortex*, *25*(7), 1987–1999. https://doi.org/10.1093/cercor/bhu012

Glasser, M. F., Sotiropoulos, S. N., Wilson, J. A., Coalson, T. S., Fischl, B., Andersson, J. L., . . . Jenkinson, M. (2013). The minimal preprocessing pipelines for the Human Connectome Project. *NeuroImage*, *80*, 105–124. https://doi.org/10.1016/j.neuroimage.2013.04.127

Godard, O., & Fiori, N. (2010). Sex differences in face processing: Are women less lateralized and faster than men? *Brain and Cognition*, *73*(3), 167–175. https://doi.org/10.1016/j.bandc.2010.04.008

Greicius, M. D., Supekar, K., Menon, V., & Dougherty, R. F. (2009). Resting-state functional connectivity reflects structural connectivity in the default mode network. *Cerebral Cortex*, *19*(1), 72–78. https://doi.org/10.1093/cercor/bhn059

Grysman, A., & Hudson, J. A. (2013). Gender differences in autobiographical memory: Developmental and methodological considerations. *Developmental Review*, *33*(3), 239–272. https://doi.org/10.1016/j.dr.2013.07.004

Hawkins, D. M., Basak, S. C., & Mills, D. (2003). Assessing model fit by cross-validation. *Journal of Chemical Information and Computer Sciences*, *43*(2), 579–586. https://doi.org/10.1021/ci025626i

Hill, A. C., Laird, A. R., & Robinson, J. L. (2014). Gender differences in working memory networks: A BrainMap meta-analysis. *Biological Psychology*, *102*, 18–29. https://doi.org/10.1016/j.biopsycho.2014.06.008

Hoerl, A. E., & Kennard, R. W. (1970). Ridge Regression: Biased Estimation for Nonorthogonal Problems. *Technometrics*, *12*(1), 55–67. https://doi.org/10.1080/00401706.1970.10488634

Hofer, A., Siedentopf, C. M., Ischebeck, A., Rettenbacher, M. A., Verius, M., Felber, S., & Fleischhacker, W. W. (2006). Gender differences in regional cerebral activity during the perception of emotion: A functional MRI study. *NeuroImage*, *32*(2), 854–862. https://doi.org/10.1016/j.neuroimage.2006.03.053

Holmes, C. J., Hoge, R., Collins, L., Woods, R., Toga, A. W., & Evans, A. C. (1998). Enhancement of MR images using registration for signal averaging. *Journal of computer assisted tomography*, *22*(2), 324–333.

Ingalhalikar, M., Smith, A., Parker, D., Satterthwaite, T. D., Elliott, M. A., Ruparel, K., . . . Verma, R. (2014). Sex differences in the structural connectome of the human brain. *Proceedings of the National Academy of Sciences*, *111*(2), 823–828. https://doi.org/10.1073/pnas.1316909110

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). *An introduction to statistical learning*. New York: Springer.

Jiménez-Jiménez, F. J., Calleja, M., Alonso-Navarro, H., Rubio, L., Navacerrada, F., Pilo-de-la-Fuente, B., . . . Agúndez, J. A. G. (2011). Influence of age and gender in motor performance in healthy subjects. *Journal of the Neurological Sciences*, *302*(1-2), 72–80. https://doi.org/10.1016/j.jns.2010.11.021

Joel, D. (2011). Male or Female? Brains are Intersex. *Frontiers in Integrative Neuroscience*, *5*, 57. https://doi.org/10.3389/fnint.2011.00057

Joel, D., Berman, Z., Tavor, I., Wexler, N., Gaber, O., Stein, Y., . . . Assaf, Y. (2015). Sex beyond the genitalia: The human brain mosaic. *Proceedings of the National Academy of Sciences*, *112*(50), 15468–15473. https://doi.org/10.1073/pnas.1509654112

Joel, D., & Fausto-Sterling, A. (2016). Beyond sex differences: New approaches for thinking about variation in brain structure and function. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *371*(1688), 20150451. https://doi.org/10.1098/rstb.2015.0451

Junaid, K. A., & Fellowes, S. (2006). Gender Differences in the Attainment of Motor Skills on the Movement Assessment Battery for Children. *Physical & Occupational Therapy In Pediatrics*, *26*(1-2), 5–11. https://doi.org/10.1080/J006v26n01_02

Kansaku, K., & Kitazawa, S. (2001). Imaging studies on sex differences in the lateralization of language. *Neuroscience research*, *41*(4), 333–337.

Kanwisher, N., McDermott, J., & Chun, M. M. (1997). The Fusiform Face Area: A Module in Human Extrastriate Cortex Specialized for Face Perception. *Journal of neuroscience*, *17*(11), 4302–4311. https://doi.org/10.1523/JNEUROSCI.17-11-04302.1997

Langner, R., & Eickhoff, S. B. (2013). Sustaining attention to simple tasks: A meta-analytic review of the neural mechanisms of vigilant attention. *Psychological Bulletin*, *139*(4), 870–900. https://doi.org/10.1037/a0030694

León, I., Tascón, L., & Cimadevilla, J. M. (2016). Age and gender-related differences in a spatial memory task in humans. *Behavioural Brain Research*, *306*, 8–12. https://doi.org/10.1016/j.bbr.2016.03.008

Levy, F., & Hobbes, G. (1979). The influences of social class and sex on sustained attention (vigilance) and motor inhibition in children. *The Australian and New Zealand Journal of Psychiatry*, *13*(3), 231–234. https://doi.org/10.3109/00048677909159141

Lighthall, N. R., Sakaki, M., Vasunilashorn, S., Nga, L., Somayajula, S., Chen, E. Y., . . . Mather, M. (2012). Gender differences in reward-related decision processing under stress. *Social Cognitive and Affective Neuroscience*, *7*(4), 476–484. https://doi.org/10.1093/scan/nsr026

Lin, C. C., Hsiao, C. K., & Chen, W. J. (1999). Development of sustained attention assessed using the continuous performance test among children 6–15 years of age. *Journal of abnormal child psychology*, *27*(5), 403–412.

Liu, X., Hairston, J., Schrier, M., & Fan, J. (2011). Common and distinct networks underlying reward valence and processing stages: A meta-analysis of functional neuroimaging studies. *Neuroscience and Biobehavioral Reviews*, *35*(5), 1219–1236. https://doi.org/10.1016/j.neubiorev.2010.12.012

Lynn, R., & Irwing, P. (2002). Sex differences in general knowledge, semantic memory and reasoning ability. *British Journal of Psychology*, *93*(4), 545–556. https://doi.org/10.1348/000712602761381394

Maitland, S. B., Herlitz, A., Nyberg, L., Bäckman, L., & Nilsson, L. G. (2004). Selective sex differences in declarative memory. *Memory & cognition*, *32*(7), 1160–1169.

Mansouri, F. A., Fehring, D. J., Gaillard, A., Jaberzadeh, S., & Parkington, H. (2016). Sex dependency of inhibitory control functions. *Biology of Sex Differences*, *7*(1), 11. https://doi.org/10.1186/s13293-016-0065-y

McCarthy, G., Puce, A., Gore, J. C., & Allison, T. (1997). Face-specific processing in the human fusiform gyrus. *Journal of Cognitive Neuroscience*, *9*(5), 605–610. https://doi.org/10.1162/jocn.1997.9.5.605

McRae, K., Ochsner, K. N., Mauss, I. B., Gabrieli, J. J. D., & Gross, J. J. (2008). Gender Differences in Emotion Regulation: An fMRI Study of Cognitive Reappraisal. *Group Processes & Intergroup Relations : GPIR*, *11*(2), 143–162. https://doi.org/10.1177/1368430207088035

Michalska, K. J., Kinzler, K. D., & Decety, J. (2013). Age-related sex differences in explicit measures of empathy do not predict brain responses across childhood and adolescence. *Developmental Cognitive Neuroscience*, *3*, 22–32. https://doi.org/10.1016/j.dcn.2012.08.001

Müller, K.-R., Anderson, C. W., & Birch, G. E. (2003). Linear and nonlinear methods for brain-computer interfaces. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, *11*(2), 165–169. https://doi.org/10.1109/TNSRE.2003.814484

Newman, L. K. (2002). Sex, Gender and Culture: Issues in the Definition, Assessment and Treatment of Gender Identity Disorder. *Clinical child psychology and psychiatry*, *7*(3), 352–359.

Neyman, J. (1934). On the two different aspects of the representative method: the method of stratified sampling and the method of purposive selection. *Journal of the Royal Statistical Society*, *97*(4), 558–625.

Nooner, K. B., Colcombe, S. J., Tobe, R. H., Mennes, M., Benedict, M. M., Moreno, A. L., . . . Milham, M. P. (2012). The NKI-Rockland Sample: A Model for Accelerating the Pace of Discovery Science in Psychiatry. *Frontiers in Neuroscience*, *6*, 152. https://doi.org/10.3389/fnins.2012.00152

Nopoulos, P., Flaum, M., O'Leary, D., & Andreasen, N. C. (2000). Sexual dimorphism in the human brain: evaluation of tissue volume, tissue composition and surface anatomy using magnetic resonance imaging. *Psychiatry Research: Neuroimaging*, *98*(1), 1–13.

Piefke, M., & Fink, G. R. (2005). Recollections of one's own past: The effects of aging and gender on the neural mechanisms of episodic autobiographical memory. *Anatomy and Embryology*, *210*(5-6), 497–512. https://doi.org/10.1007/s00429-005-0038-0

Pläschke, R. N., Cieslik, E. C., Müller, V. I., Hoffstaedter, F., Plachti, A., Varikuti, D. P., . . . Eickhoff, S. B. (2017). On the integrity of functional brain networks in schizophrenia, Parkinson's disease, and advanced age: Evidence from connectivity-based single-subject classification. *Human Brain Mapping*, *38*(12), 5845–5858. https://doi.org/10.1002/hbm.23763

Pohl, R. F., Bender, M., & Lachmann, G. (2005). Autobiographical memory and social skills of men and women. *Applied Cognitive Psychology: The Official Journal of the Society for Applied Research in Memory and Cognition*, *19*(6), 745–759. https://doi.org/10.1002/acp.1104

Premack, D., & Woodruff, G. (1978). Does the chimpanzee have a theory of mind? *Behavioral and brain sciences*, *1*(4), 515–526.

Prinzel, L. J., & Freeman, F. G. (1997). Task-specific sex differences in vigilance performance: Subjective workload and boredom. *Perceptual and Motor Skills*, *85*(3 Pt 2), 1195–1202. https://doi.org/10.2466/pms.1997.85.3f.1195

Proverbio, A. M., Brignone, V., Matarazzo, S., Del Zotto, M., & Zani, A. (2006). Gender differences in hemispheric asymmetry for face processing. *BMC Neuroscience*, *7*(1), 44. https://doi.org/10.1186/1471-2202-7-44

Rafi, M., & Shahid Shaikh, M. (2013). A comparison of SVM and RVM for Document Classification.

Renoult, L., Davidson, P. S. R., Palombo, D. J., Moscovitch, M., & Levine, B. (2012). Personal semantics: At the crossroads of semantic and episodic memory. *Trends in Cognitive Sciences*, *16*(11), 550–558. https://doi.org/10.1016/j.tics.2012.09.003

Rizzolatti, G., & Craighero, L. (2004). The mirror-neuron system. *Annual Review of Neuroscience*, *27*, 169–192. https://doi.org/10.1146/annurev.neuro.27.070203.144230

Rottschy, C., Langner, R., Dogan, I., Reetz, K., Laird, A. R., Schulz, J. B., . . . Eickhoff, S. B. (2012). Modelling neural correlates of working memory: A coordinate-based meta-analysis. *NeuroImage*, *60*(1), 830–846. https://doi.org/10.1016/j.neuroimage.2011.11.050

Ruigrok, A. N. V., Salimi-Khorshidi, G., Lai, M.-C., Baron-Cohen, S., Lombardo, M. V., Tait, R. J., & Suckling, J. (2014). A meta-analysis of sex differences in human brain structure. *Neuroscience and Biobehavioral Reviews*, *39*, 34–50. https://doi.org/10.1016/j.neubiorev.2013.12.004

Saarela M., Elomaa T., Ruohonen K. (2010). *An Analysis of Relevance Vector Machine Regression: In Advances in Machine Learning I*. (pp. 227-246). Springer, Berlin, Heidelberg. Retrieved from https://link.springer.com/chapter/10.1007/978-3-642-05177-7_11

Sabatinelli, D., Fortune, E. E., Li, Q., Siddiqui, A., Krafft, C., Oliver, W. T., . . . Jeffries, J. (2011). Emotional perception: Meta-analyses of face and natural scene processing. *NeuroImage*, *54*(3), 2524–2533. https://doi.org/10.1016/j.neuroimage.2010.10.011

Sacher, J., Neumann, J., Okon-Singer, H., Gotowiec, S., & Villringer, A. (2013). Sexual dimorphism in the human brain: Evidence from neuroimaging. *Magnetic Resonance Imaging*, *31*(3), 366–375. https://doi.org/10.1016/j.mri.2012.06.007

Salimi-Khorshidi, G., Douaud, G., Beckmann, C. F., Glasser, M. F., Griffanti, L., & Smith, S. M. (2014). Automatic denoising of functional MRI data: Combining independent component analysis and hierarchical fusion of classifiers. *NeuroImage*, *90*, 449–468. https://doi.org/10.1016/j.neuroimage.2013.11.046

Satterthwaite, T. D., Wolf, D. H., Roalf, D. R., Ruparel, K., Erus, G., Vandekar, S., . . . Gur, R. C. (2015). Linked Sex Differences in Cognition and Functional Connectivity in Youth. *Cerebral Cortex*, *25*(9), 2383–2394. https://doi.org/10.1093/cercor/bhu036

Schulkind, M., Schoppel, K., & Scheiderer, E. (2012). Gender differences in autobiographical narratives: He shoots and scores; she evaluates and interprets. *Memory & Cognition*, *40*(6), 958–965. https://doi.org/10.3758/s13421-012-0197-1

Schulte-Rüther, M., Markowitsch, H. J., Fink, G. R., & Piefke, M. (2007). Mirror neuron and theory of mind mechanisms involved in face-to-face interactions: A functional magnetic resonance imaging approach to empathy. *Journal of Cognitive Neuroscience*, *19*(8), 1354–1372. https://doi.org/10.1162/jocn.2007.19.8.1354

Schulte-Rüther, M., Markowitsch, H. J., Shah, N. J., Fink, G. R., & Piefke, M. (2008). Gender differences in brain networks supporting empathy. *NeuroImage*, *42*(1), 393–403. https://doi.org/10.1016/j.neuroimage.2008.04.180

Sergent, J., Ohta, S., & MACDONALD, B. (1992). Functional neuroanatomy of face and object processing: a positron emission tomography study. *Brain*, *115*(1), 15–36.

Smith, S. M., Beckmann, C. F., Andersson, J., Auerbach, E. J., Bijsterbosch, J., Douaud, G., . . . Glasser, M. F. (2013). Resting-state fMRI in the Human Connectome Project. *NeuroImage*, *80*, 144–168. https://doi.org/10.1016/j.neuroimage.2013.05.039

Solianik, R., Brazaitis, M., & Skurvydas, A. (2016). Sex-related differences in attention and memory. *Medicina*, *52*(6), 372–377. https://doi.org/10.1016/j.medici.2016.11.007

Sommer, W., Hildebrandt, A., Kunina-Habenicht, O., Schacht, A., & Wilhelm, O. (2013). Sex differences in face cognition. *Acta Psychologica*, *142*(1), 62–73. https://doi.org/10.1016/j.actpsy.2012.11.001

Speck, O., Ernst, T., Braun, J., Koch, C., Miller, E., & Chang, L. (2000). Gender differences in the functional organization of the brain for working memory. *Neuroreport*, *11*(11), 2581–2585.

Spreng, R. N., Mar, R. A., & Kim, A. S. (2009). The common neural basis of autobiographical memory, prospection, navigation, theory of mind, and the default mode: a quantitative meta-analysis. *Journal of cognitive neuroscience*, *21*(3), 489–510.

Stevens, J. S., & Hamann, S. (2012). Sex differences in brain activation to emotional stimuli: A meta-analysis of neuroimaging studies. *Neuropsychologia*, *50*(7), 1578–1593. https://doi.org/10.1016/j.neuropsychologia.2012.03.011

Sturman, D. A., & Moghaddam, B. (2011). The neurobiology of adolescence: Changes in brain architecture, functional dynamics, and behavioral tendencies. *Neuroscience and Biobehavioral Reviews*, *35*(8), 1704–1712. https://doi.org/10.1016/j.neubiorev.2011.04.003

Teleb, A. A., & Al Awamleh, A. A. (2012). Gender differences in cognitive abilites. *Current Research in Psychology*, *3*(1), 33–39. https://doi.org/10.3844/crpsp.2012.33.39

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

Tipping, M. E. (2001). Sparse Bayesian learning and the relevance vector machine. *Journal of machine learning research*, *1*(Jun), 211–244.

Tomasi, D., & Volkow, N. D. (2012). Aging and functional brain networks. *Molecular Psychiatry*, *17*(5), 549-58. https://doi.org/10.1038/mp.2011.81

Tran, U. S., Hofer, A. A., & Voracek, M. (2014). Sex differences in general knowledge: Meta-analysis and new data on the contribution of school-related moderators among high-school students. *PloS One*, *9*(10), e110391. https://doi.org/10.1371/journal.pone.0110391

Van Essen, D. C., Ugurbil, K., Auerbach, E., Barch, D., Behrens, T. E. J., Bucholz, R., . . . Yacoub, E. (2012). The Human Connectome Project: A data acquisition perspective. *NeuroImage*, *62*(4), 2222–2231. https://doi.org/10.1016/j.neuroimage.2012.02.018

Vaquero, E., Cardoso, M. J., Vazque, M., & Gomez, C. M. (2004). Gender differences in event-related potentials during visual-spatial attention. *International Journal of Neuroscience*, *114*(4), 541–557. https://doi.org/10.1080/00207450490422056

Walker, S. (2005). Gender differences in the relationship between young children's peer-related social competence and individual differences in theory of mind. *The Journal of Genetic Psychology*, *166*(3), 297–312. https://doi.org/10.3200/GNTP.166.3.297-312

Wei, L., Yang, Y., Nishikawa, R. M., & Jiang, Y. (2005). A study on several Machine-learning methods for classification of Malignant and benign clustered microcalcifications. *IEEE TRANSACTIONS ON MEDICAL IMAGING*, *24*(3), 371–380. Retrieved from https://ieeexplore.ieee.org/abstract/document/1397824

Weiss, E. M., Kemmler, G., Deisenhammer, E. A., Fleischhacker, W. W., & Delazer, M. (2003). Sex differences in cognitive functions. *Personality and Individual Differences*, *35*(4), 863–875. https://doi.org/10.1016/S0264-3707(03)00061-9

Wellcome Centre for Human Neuroimaging. (2018). SPM - Statistical Parametric Mapping. Retrieved from https://www.fil.ion.ucl.ac.uk/spm/

Whittle, S., Yücel, M., Yap, M. B. H., & Allen, N. B. (2011). Sex differences in the neural correlates of emotion: Evidence from neuroimaging. *Biological Psychology*, *87*(3), 319–333. https://doi.org/10.1016/j.biopsycho.2011.05.003

Witt, S. T., Laird, A. R., & Meyerand, M. E. (2008). Functional neuroimaging correlates of finger-tapping task variations: An ALE meta-analysis. *NeuroImage*, *42*(1), 343–356. https://doi.org/10.1016/j.neuroimage.2008.04.025

Wolpert, D. H., & Macready, W. G. (1997). No free lunch theorems for optimization. *IEEE transactions on evolutionary computation*, *1*(1), 67–82.

Young, K. D., Bellgowan, P. S. F., Bodurka, J., & Drevets, W. C. (2013). Functional neuroimaging of sex differences in autobiographical memory recall. *Human Brain Mapping*, *34*(12), 3320–3332. https://doi.org/10.1002/hbm.22144

Zaidi, Z. F. (2010). Gender Differences in Human Brain: A Review. *The Open Anatomy Journal*, *2*(1), 37–55. https://doi.org/10.2174/1877609401002010037

Zhang, C., Dougherty, C. C., Baum, S. A., White, T., & Michael, A. M. (2018). Functional connectivity predicts gender: Evidence for gender differences in resting brain connectivity. *Human Brain Mapping*, *39*(4), 1765–1776. https://doi.org/10.1002/hbm.23950

Appendix

Table 3

Classification accuracies of the HBN-dataset

	machine learning method			
	LASSO	LSVM	Ridge	RVM
networks:				
AM	51.11%	52.95%	50.08%	50.86%
CogAC	54.87%	56.16%	49.81%	50.03%
emoSF	51.44%	48.14%	52.41%	51.26%
Empathy	43.35%	51.28%	45.56%	50.78%
ER	44.37%	52.43%	46.26%	48.25%
MNS	50.24%	50.72%	53.08%	48.41%
Motor	51.74%	52.38%	54.48%	49.52%
Rew	42.75%	52.74%	48.23%	48.65%
SM	42.97%	51.24%	46.62%	49.50%
ToM	46.80%	54.67%	42.51%	49.37%
VigAtt	52.35%	48.15%	51.40%	49.61%
WM	42.74%	51.36%	45.89%	50.09%

Note. Displayed are the individual result values of the analyses for each network in the HBN-dataset with regard to the four different ML-algorithms.

Table 4

Classification accuracies of the eNKI-dataset

	machine learning method			
	LASSO	LSVM	Ridge	RVM
networks:				
AM	66.28%	66.19%	62.28%	64.35%
CogAC	58.02%	58.32%	63.63%	56.34%
emoSF	67.95%	65.87%	69.30%	54.58%
Empathy	64.89%	61.81%	55.64%	65.07%
ER	65.35%	65.46%	67.93%	66.21%
MNS	50.00%	49.98%	54.05%	50.95%
Motor	57.82%	59.51%	61.76%	61.57%
Rew	68.12%	66.23%	71.69%	60.56%
SM	68.63%	62.68%	70.33%	57.50%
ToM	64.65%	63.73%	66.56%	63.16%
VigAtt	56.56%	56.46%	61.78%	60.07%
WM	65.02%	60.57%	69.33%	61.29%

Note. Displayed are the individual result values of the analyses for each network in the eNKI-dataset with regard to the four different ML-algorithms.

Table 5

Classification accuracies of the HCP-dataset

	machine-learning method			
	LASSO	LSVM	Ridge	RVM
networks:				
AM	64.52%	66.03%	65.07%	49.74%
CogAC	53.78%	55.88%	55.27%	49.98%
emoSF	59.10%	60.26%	55.30%	49.73%
Empathy	58.61%	58.59%	55.69%	50.08%
ER	60.36%	59.61%	58.82%	50.10%
MNS	52.92%	53.60%	54.69%	49.99%
Motor	58.15%	58.65%	57.34%	49.98%
Rew	58.30%	59.30%	58.15%	50.07%
SM	63.62%	65.07%	64.36%	50.08%
ToM	57.48%	55.50%	57.40%	49.73%
VigAtt	57.35%	57.57%	57.89%	49.73%
WM	58.56%	60.74%	55.42%	50.09%

Note. Displayed are the individual result values of the analyses for each network in the HCP-dataset with regard to the four different ML-algorithms.

Table 6

Classification accuracies of the 1000BRAINS-dataset

	machine-learning method			
	LASSO	LSVM	Ridge	RVM
networks:				
AM	62.03%	50.00%	61.94%	61.98%
CogAC	57.18%	50.01%	56.85%	49.97%
emoSF	58.21%	50.00%	59.74%	49.98%
Empathy	57.98%	50.01%	56.98%	49.73%
ER	53.15%	50.01%	55.73%	49.90%
MNS	49.93%	50.01%	51.17%	50.01%
Motor	49.93%	50.01%	52.83%	50.00%
Rew	56.99%	50.00%	58.43%	49.85%
SM	61.04%	50.00%	61.50%	50.00%
ToM	54.98%	50.02%	56.75%	50.03%
VigAtt	50.51%	50.01%	55.13%	50.02%
WM	56.95%	50.00%	57.23%	49.88%

Note. Displayed are the individual result values of the analyses for each network in the 1000BRAINS-dataset with regard to the four ML-algorithms.

HEINRICH HEINE
UNIVERSITÄT DÜSSELDORF

Erklärung

Hiermit versichere ich an Eides Statt, dass ich die Arbeit mit dem Titel

Sex classification from resting-state functional brain networks

selbstständig und ohne unzulässige fremde Hilfe verfasst habe. Ich versichere insbesondere:

(1) Ich habe keine anderen als die angegebenen Quellen und Hilfsmittel benutzt.

(2) Alle wörtlich oder dem Sinn nach aus anderen Texten entnommenen Stellen habe ich als solche kenntlich gemacht; dies gilt für gedruckte Texte ebenso wie für elektronische Ressourcen (z.B. im Internet verfügbare Dokumente).

(3) Die Arbeit habe ich in der vorliegenden oder einer modifizierten Form noch nicht als Abschlussarbeit oder in einer Lehrveranstaltung als Studienleistung vorgelegt – sei es in einem Studiengang an der Heinrich-Heine-Universität oder an einer anderen Universität.

(4) Die Arbeit enthält keine Teile (Textpassagen, Grafiken usw.) aus Arbeiten anderer Studierender.

(5) Ich weiß, dass eine Täuschung im Sinne von §18 der Prüfungsordnungen für den Bachelor- bzw. Master-of-Science-Studiengang Psychologie vorliegt, wenn mein tatsächliches Handeln nicht meiner Versicherung an Eides Statt entspricht; diese Täuschung stellt eine Ordnungswidrigkeit dar und kann nach §63 des Hochschulgesetzes NRW mit einer Geldbuße von bis zu 50.000 € geahndet werden.

Datum: 02.07.2019

Name: Lisa Wiersch

Unterschrift: L. Wiersch