NeuroImage: Clinical 27 (2020) 102304

Contents lists available at ScienceDirect

NeuroImage: Clinical

journal homepage: www.elsevier.com/locate/ynicl

Identifying psychosis spectrum youth using support vector machines and cerebral blood perfusion as measured by arterial spin labeled fMRI

Dawson J. Overton[a,b], Nikhil Bhagwat[a,b,e,f,g], Joseph D. Viviano[a,b], Grace R. Jacobs[a,b,c], Aristotle N. Voineskos[a,b,d,*]

[a] Kimel Family Translational Imaging Genetics Research Laboratory, The Centre for Addiction and Mental Health, University of Toronto, Toronto, Canada
[b] Campbell Family Mental Health Research Institute, The Centre for Addiction and Mental Health, University of Toronto, Toronto, Canada
[c] Institute of Medical Science, University of Toronto, Toronto, Canada
[d] Department of Psychiatry, University of Toronto, Toronto, Canada
[e] Cerebral Imaging Centre, Douglas Mental Health University Institute, Montreal, Quebec, Canada
[f] Department of Biological and Biomedical Engineering, McGill University, Montreal, Quebec, Canada
[g] Institute of Biomaterials and Biomedical Engineering, University of Toronto, Toronto, Ontario, Canada

A B S T R A C T

Altered cerebral blood flow (CBF), as measured by arterial spin labelling (ASL), has been observed in several psychiatric conditions, but is a generally underutilized MRI technique, especially in the study of psychosis spectrum (PS) symptoms. We aimed to determine group differences in ASL resting state functional connectivity (rsFC) between PS and non-PS youth, and the reliability of a support vector machine (SVM) classifier trained on ASL rsFC features to differentiate PS and non-PS youth, especially compared to blood oxygen level dependent (BOLD) fMRI. 1146 youth aged 8–22 with ASL and BOLD data from the Philadelphia Neurodevelopmental Cohort were analyzed. Widespread ASL hyperconnectivity was found in the left cuneus, precuneus, and dorsolateral prefrontal cortex, and hypoconnectivity was found in the left cingulate cortex and orbitofrontal area (multiple linear regression, FDR corrected). An SVM trained on ASL and BOLD features outperformed either modality alone (AUC_{BOTH} = 0.72 versus AUC_{ASL} = 0.68 and AUC_{BOLD} = 0.67). Classification performance and precision improved when the non-PS group had no psychiatric comorbidities. The relative success of the classifier suggests ASL rsFC changes exist in PS individuals that differ from BOLD rsFC changes, and extends previous findings of CBF dysregulation in PS.

1. Introduction

Effective interventions are available for early psychosis (Birchwood et al., 1998). However, efforts at primary or secondary prevention of a first episode of psychosis have proved challenging (Fusar-Poli, 2013), including in the “clinical high-risk” (CHR) population, which even if successful would apply to a small minority of incident first episode cases. Therefore, the need to identify individuals before symptoms and more severe functioning impairments fully manifest remains a high priority. The use of machine learning classification coupled with advanced brain imaging techniques offers promise to detect subtle, high-dimensional differences in functional brain activity for PS youth, and may provide insight into underlying neurobiology and mechanisms of psychosis risk. Further, assessing performance of classification models in PS youth provides a more realistic appraisal of the potential for future generation of identification tools early in impairment, and well before schizophrenia, first episode psychosis, or even the clinical high-risk state.

Some success in applying methods to identify CHR individuals has been achieved (Cannon et al., 2016; Koutsouleris et al., 2012). Work by Koutsouleris and colleagues has shown that support vector machines (SVMs) applied to structural MRI data can achieve high accuracy in differentiating between healthy controls (HCs) and individuals who exhibit both early and late at-risk mental states (ARMS) for psychosis (Koutsouleris et al., 2012, 2009). However, the majority of work using machine learning in the early stages of psychosis is based on structural neuroimaging (Satterthwaite et al., 2016): there is little functional neuroimaging, and to our knowledge no use of arterial spin labeling (ASL) to classify those who may be suffering from dimensional psychosis symptoms (a group that has significant overlap with but is distinct from the CHR group), despite evidence that it may be a useful biomarker for regional resting state brain function (Detre et al., 2009).

The Philadelphia Neurodevelopmental Cohort (PNC) is a community-based sample of children and youth (of which ~20% meet criteria for having PS symptoms) with multimodal neuroimaging data (n = 1445) including both resting state blood-oxygen-level dependent

* Corresponding author at: Kimel Family Translational Imaging Genetics Research Laboratory, The Centre for Addiction and Mental Health, 250 College St., Toronto, Ontario M5T 1R8, Canada.
E-mail address: Aristotle.Voineskos@camh.ca (A.N. Voineskos).

https://doi.org/10.1016/j.nicl.2020.102304
Received 21 February 2020; Received in revised form 15 May 2020; Accepted 1 June 2020
Available online 04 June 2020

(BOLD) fMRI as well as ASL data (Calkins et al., 2014). Such a large sample provides an opportunity to study cerebral blood flow differences in PS youth and test prediction accuracy with a greater level of confidence. Given the understudied nature of ASL fMRI and the fact that it possesses signal characteristics which may lend themselves to improved detection of network differences in PS youth even beyond what BOLD can detect (Wang et al., 2003), new knowledge may be generated by investigating classification performance using this modality.

Additionally, developmental sex differences related to brain structure (Gennatas et al., 2017) and connectivity (Gur and Gur, 2016) have been observed in typically developing youth, including changes in cerebral perfusion as measured by ASL (Kaczkurkin et al., 2018; Satterthwaite et al., 2014). A variety of sex differences are present in schizophrenia (Mendrek and Mancini-Marïe, 2016) that extend to earlier stages of psychosis risk (Barajas et al., 2015), including sex and age dependent differences in brain structure and resting-state fMRI connectivity that have been found in PS youth (Jacobs et al., 2019). Thus, analyses examining sex differences in ASL may provide insight into the neurobiological basis of PS symptoms in youth.

The first aim of this study is to describe differences in connectivity between psychosis spectrum (PS) and non-PS groups using ASL resting state functional connectivity (rsFC), and determine the ability of a support vector machine classifier trained on ASL rsFC data to differentiate between these groups. Performance will also be compared with a classifier trained on BOLD rsFC data from the same subjects, as well as one trained on a concatenation of BOLD and ASL rsFC. We hypothesize that classification of PS youth using ASL will be as successful as that using BOLD, but may reveal that some individuals are more accurately classified using one technique or the other. Given that developmental sex differences have been observed in ASL regionally, the second aim of this study is to investigate sex-based differences in ASL rsFC, and compare performance of models trained on exclusively male or female ASL rsFC data. Here, we hypothesize that the ASL classifier may perform differently for boys and girls (when differentiating between PS vs. non-PS).

2. Materials and methods

2.1. Participant recruitment and characterization

As described in detail elsewhere e.g., (Calkins et al., 2015), the Philadelphia Neurodevelopmental Cohort (PNC) is a large (n = 9428), cross-sectional population sample of youth (aged 8–22) that includes a wide range of neuroimaging, genetic, neurocognitive, and demographic data. Recruitment procedures for this catchment area cohort have been previously described (Calkins et al., 2014). All participants provided written informed consent. For youth under 18, both written informed consent and parental/legal guardian permission were acquired. The University of Pennsylvania and the Children's Hospital of Philadelphia Institutional Review Boards approved all procedures. 1445 of these subjects were stratified for age and sex and randomly selected for neuroimaging. Of these subjects, a subset was excluded on the basis of missing or poor quality fMRI data, as well as missing demographic or instrument data. 243 subjects had missing or unusable BOLD fMRI data, and a further 56 subjects had missing or unusable ASL fMRI data (some excluded subjects were also missing demographic data, but all subjects with usable BOLD and ASL data had demographic data). This resulted in n = 1146 with usable ASL and BOLD fMRI data, in addition to all necessary demographic data (see Table 1).

Participants were defined as being part of the PS group if they scored sufficiently high on the PRIME Screen-Revised for sub-psychosis symptoms (Miller et al., 2003); a modified version of the Kiddie Schedule for Affective Disorders and Schizophrenia (K-SADS) (Ambrosini, 2000) developed for collection of the PNC data (Calkins et al., 2015), or the Scale of Prodromal Symptoms (SOPS). If a subject met one or more of the following sets of criteria, they were considered part of the PS group:

PRIME Screen-Revised	- Overall z-score > 2 in age group - > = 1 rating of 6 (strongly agree) - > = 3 ratings of 5 (somewhat agree)
K-SADS Psychosis Screen	- > 1 day, > 5 distress rating
SOPS	- Overall z-score > 2 in age group

2.2. MRI acquisition

All scans were performed using a 3 T Siemens TIM Trio whole-body scanner (VB17 software revision) located at the Hospital of the University of Pennsylvania. Gradient performance was 45 mT/m, with a maximum slew rate of 200 T/m/s. A quadrature body coil was used for signal transmission and a 32-channel head coil for signal reception (Satterthwaite et al., 2014).

2.3. BOLD resting state scan

Resting state BOLD acquisitions were obtained using a single-shot, interleaved, multislice, gradient-echo, echo planar imaging (GE-EPI) sequence. A voxel resolution of $3 \times 3 \times 3$ mm with 124 slices was used, with TR = 3000 ms, TE = 32 ms, flip angle = 90°, bandwidth/pixel = 2056 Hz.

During the BOLD resting state scans, a fixation cross was displayed and subjects were instructed to stay awake, keep their eyes open, remain still and fixate on the cross. Resting state scan duration was 6 min 18 s (Satterthwaite et al., 2014).

2.4. ASL perfusion scan

ASL acquisitions were collected using a pseudo-continuous arterial spin labeling (pCASL) sequence (Wu et al., 2007) with a single-shot spin-echo EPI readout (label duration = 1500 ms, post label delay = 1200 ms, labeling plane = 90 mm inferior to center slice, 40 label and 40 control slices). Slices were acquired in ascending, non-interleaved order (Satterthwaite et al., 2014), with TR = 4000 ms, TE = 15 ms, flip angle = 90°/180°, bandwidth/pixel = 2604 Hz.

2.5. Image preprocessing - ASL and BOLD fMRI

The ASL data as received from the Philadelphia group was already motion corrected. Beyond this, data was preprocessed in a way similar to previous ASL connectivity studies (Loggia et al., 2013), using FSL (FMRIB's Software Library) (Smith et al., 2004) and AFNI (Analysis of Functional Neuroimages) (Cox, 1996).

ASL data was skull-stripped using the FSL brain extraction tool (BET). Adjacent tag and control images were subtracted using a custom script to obtain a perfusion signal. FSL FLIRT was used to register the ASL data to a high resolution anatomical volume (T1), and then transform the data to MNI152 standard space. Data was also high-pass filtered at 0.008 Hz using a 4th order Butterworth filter and spatially smoothed at FWHM = 5 mm using AFNI, consistent with preprocessing in other studies (Loggia et al., 2013).

BOLD data was preprocessed using standard methods (Satterthwaite et al., 2015). The first 4 volumes were removed to allow for signal stabilization, and the remaining volumes were skull-stripped using BET, registered to a T1 volume, and transformed to MNI152 standard space using FSL FLIRT. Data was band-pass filtered to allow frequencies between 0.01 and 0.08 Hz using a 4th order Butterworth filter. Time points (TRs) with a framewise displacement of > 0.25 mm/TR were replaced with a linear interpolation of the two closest surviving TRs.

In all cases, visual quality control (QC) was performed. In addition, image quality was assessed according to temporal signal-to-noise ratio (tSNR), subject motion, global signal spike rate, and global signal drift (Satterthwaite et al., 2014). High-motion subjects were identified on the basis of number of TRs retained in the BOLD scans after TR

Table 1
Summary statistics of psychosis spectrum and non-psychosis spectrum youth with usable ASL and BOLD imaging data.

Demographic/Clinical Characteristics	nonPS	PS	p value
n (%)	882 (76.96)	264 (23.04)	
Age (years, mean ± SD)	14.71 (3.46)	14.92 (2.88)	0.363
Sex (% female)	54.4	51.9	
Ethnicity (% European American)	51.3	33.1	
WRAT-4 Standard Reading Score (mean (SD))	53.23 (9.30)	50.69 (8.91)	< 0.001
PRIME Screen Revised total score (mean (SD))	3.85 (6.00)	23.33 (13.99)	< 0.001
SOPS total score (mean (SD))	1.44 (1.82)	5.04 (4.50)	< 0.001
Children's Global Assessment Scale, current (mean (SD))	82.58 (9.82)	72.51 (12.82)	< 0.001
Maternal education (years, mean (SD))	14.46 (2.50)	13.72 (2.24)	< 0.001
Paternal education (years, mean (SD))	14.16 (2.74)	13.38 (2.53)	< 0.001

scrubbing. Out of 120 TRs, if fewer than 100 TRs were retained, the subject was considered to be "high-motion" and was excluded (although this TR scrubbing step was performed only on BOLD scans, subjects that did not pass the 100 TR threshold were also excluded from ASL analyses).

2.6. Connectivity matrix construction

All connectivity analyses were performed using Python and associated packages available in Anaconda (https://continuum.io/). A whole-brain, graph-theory-based parcellation (Shen et al., 2013) was used to define volumetric regions of interest (ROIs) for connectivity analyses. This atlas was chosen due to its high degree of parcellation reproducibility at the group level, and its superior subunit spatial homogeneity compared to some competing approaches (e.g., Craddock et al., 2012). The atlas was downsampled to match the resolution of the functional images ($91 \times 109 \times 91$). A symmetric correlation matrix between ROIs for each modality and subject was calculated using the mean BOLD and ASL signal in each ROI across the timeseries.

After registration of the functional data to MNI152 space, some ROIs had a mean signal of exactly 0 for certain subjects (54 out of 268), and it was found that the masking procedure excluded these ROIs. To ensure that each subject had the same number of connections for comparison, these ROIs were excluded from all subjects for subsequent analyses (leaving 214; see Supplementary Figure S2 for visualization of excluded ROIs).

2.7. Feature selection

Given the large number of features (i.e., connection strengths between ROIs) in both connectivity matrices (22,791), and the fact that many connections are not functionally meaningful, it was important to reduce the number of features to reduce noise and prevent model overfitting. To remove redundant features, pairwise correlations were calculated across all subjects within each fold for all features, and connections that had a correlation coefficient of greater than 0.85 with at least one other connection were dropped (the number of features dropped this way ranged from 2122 to 2865 for each fold). Other groups have investigated feature selection in fMRI data (Zeng et al., 2012; Pereira et al., 2009), and have found that an ANOVA F-value ranking of features based on differences between classes can work well for BOLD fMRI classification. Correspondingly, features were ranked according to their F-value on the training data within each of 10 cross-validation folds after the removal of correlated features, and the top 50% of features were retained in each case (resulting in approximately 10,000 features in each fold; range: 9963 to 10,334).

2.8. ASL rsFC group comparison

Subjects were divided by sex and diagnosis (PS and non-PS). A multiple linear regression with PS and non-PS groups as a dummy variable for each unique correlation (connection) between ROIs (22,791 connections) was performed, separately for each sex and for the entire sample. FDR multiple comparison correction was performed, and the mean values of connections with a p-value < 0.05 after correction were calculated, resulting in a single ROI-specific connection value for significant connections. As well, the analysis was repeated with age as a covariate in the regression. These ROIs were mapped to anatomical areas, and the connectivity differences between PS and non-PS groups were visualized on a connectivity plot. For comparison, the same procedures were also performed for BOLD rsFC.

2.9. Support vector machine training and validation

Classifier training was performed using scikit-learn (version 0.18.1) (Pedregosa et al., 2011), which includes a Python implementation of Libsvm (Chang and Lin, 2011). For both BOLD and ASL data, a linear-kernel SVM classifier was trained using the selected features. L2 regularization that occurs during SVM training makes SVMs relatively less sensitive to overfitting (which would otherwise be a danger due to a small number of samples with a much larger number of highly variable features) (Scholkopf and Smola, 2001). The PNC is a very large dataset in the context of neuroimaging, but relatively small compared to a typical machine learning classification problem, and thus SVMs were considered to be a good model choice to ameliorate the effects of a lack of samples. Grid search hyperparameter selection was performed for the value of C, where $C \in \{10^{-6}, 10^{-5}, \ldots, 10^{5}, 10^{6}\}$. A class imbalance existed between PS and non-PS groups (with PS making up 23.04% of the entire sample with roughly equal males and females), and thus classes were weighted inversely proportional to frequency during training, which resulted in increased sensitivity of the classifier at the expense of specificity. 10-fold stratified cross validation was performed, and area under the Receiver Operating Characteristic curve (AUC) across all folds was calculated for each model, as well as a precision-recall curve and a confusion matrix.

Three models were trained: one with ASL rsFC data, one with BOLD rsFC data, and a third using a concatenation of the ASL and BOLD rsFC data (using the same set of ROIs for each). All models were trained in 15.42 s on a computer with a Intel Xeon X5687 3.60 GHz 16-core processor, 24 GB of RAM, and running Ubuntu 14.04. A method described by Hanley and McNeil was used to detect differences in performance between classifiers based on AUC (Hanley and McNeil, 1982), which corrects the calculation of standard error of the difference between AUC values for paired data in a way analogous to a paired *t*-test. A permutation importance method described by Breiman and implemented by Raschka (Raschka, 2018) was used to determine the most significant features for classification (Breiman, 2001).

We also investigated if performance could be improved in ASL rsFC classification by considering only a subset of the non-PS youth with no other neurological or mental health concerns (see exclusion criteria below). This resulted in a TD group of 409 youth, and a model was trained using the same procedure as previous analyses. Finally,

classifiers were trained separately on males and females, and performance was compared between sexes.

2.10. TD group exclusion criteria

Youth with any of the following were removed from the non-PS group to yield the TD group: evidence of epilepsy, lead poisoning, meningitis, multiple sclerosis, serious head injury, skull fracture, or autism spectrum disorder (ASD). Youth were also removed if they met criteria for bipolar disorder, severe phobia, major depressive disorder, post-traumatic stress disorder, obsessive compulsive disorder, generalized anxiety disorder, oppositional-defiant disorder, or attention deficit hyperactivity disorder.

3. Results

3.1. ASL functional connectivity differences were found between PS and non-PS youth, and were more pronounced in females

When accounting for age through the use of a multiple linear regression model for each FC, ASL hyperconnectivity and hypoconnectivity were found in many regions, including hyperconnectivity in the cuneus, precuneus, and dorsolateral prefrontal cortex, and hypoconnectivity in the left cingulate cortex and orbitofrontal area (see Supplementary Figure S1; PS > non-PS; $p < 0.05$, FDR corrected). The ASL rsFC group differences were more pronounced in females, with a larger number of significant FCs and additional hyperconnectivity in parietal regions ($p < 0.05$, FDR corrected). Using the same procedure, group differences were also found in BOLD rsFC, including hyperconnectivity in the fusiform gyrus, inferior temporal gyrus, and precuneus, and hypoconnectivity in the cuneus and anterior prefrontal cortex, though fewer significant FCs were found compared to ASL rsFC. As with ASL rsFC, a greater number of significant FCs were found when the regression was performed only on female subjects. The FCs with the strongest group differences for both modalities are shown in Supplementary Figure S1.

3.2. A multivariate predictive model trained on ASL rsFC performed comparatively to BOLD rsFC, and a concatenation of ASL and BOLD outperformed either rsFC model alone

Given differences in ASL signal characteristics compared to BOLD, and a lack of studies exploring the predictive value of ASL rsFC in PS, it was not previously known how effective a modality it would be to differentiate between PS and non-PS groups. Classifiers were trained on ASL and BOLD rsFC separately, and a concatenation of both. The performance of the models after 10-fold cross-validation is summarized below (Fig. 1).

An additional classifier trained on ASL rsFC with randomly permuted labels did not perform better than chance, strongly suggesting that the performance seen in the classifiers with the correct class labels was indeed due to group differences in the ASL rsFC and BOLD rsFC.

Performance of the BOLD rsFC classifier was similar to ASL rsFC. The Hanley and McNeil method was used to compare the performance of both models, and the ASL classifier was not found to be superior to a significant degree ($p = 0.7972$). The concatenated model, however, resulted in significantly improved performance than classifiers trained on either modality alone ($p = 0.0443$).

3.3. Sex differences exist in ASL classification performance, but are at least partially explained by participant motion

The 1146 youth were split by sex (617 females, 529 males), and a separate SVM classifier was trained on only male and only female ASL rsFC, BOLD rsFC, and ASL + BOLD rsFC data. The performance of these models is summarized below (Fig. 3):

Performance for the female subjects was superior to the male subjects for all three models (ASL: $p = 0.0380$, BOLD: $p = 0.0482$, ASL + BOLD: $p = 0.02229$).

Since it is known that head motion can impact several metrics related to functional connectivity (Van Dijk et al., 2012; Satterthwaite et al., 2012); we investigated whether these sex differences could potentially be explained by head motion. We used mean framewise displacement (FD) to quantify head motion, under the assumption that high motion individuals during the BOLD scan would also be high motion during the ASL scan.

A significant difference of mean FD was found between sexes ($\overline{x}_{Males} = 0.1693110$, $\overline{x}_{Females} = 0.1369225$, $t = 3.4687$, $df = 560.5$, $p = 0.0005632$). Using a logistic regression model predicting correct classification by the ASL model versus sex, the effect of sex was found to be significant ($p = 0.0328$). When mean FD and a sex/mean FD interaction term were included in the model, the effect of sex was no longer significant ($p = 0.8606$), suggesting the observed sex differences could at least partially be attributed to head motion. It is important to note, however, that no significant difference in mean FD between PS and non-PS women was found ($\overline{x}_{PS} = 0.1429584$, $\overline{x}_{non\text{-}PS} = 0.1353748$, $p = 0.5768$).

No significant correlation was found between mean FD and total PRIME score ($r^2 = 0.02628924$, $p = 0.4812$), SOPS score ($r^2 = 0.01917064$, $p = 0.6076$), or GOASSESS score ($r^2 = 0.01214009$, $p = 0.7567$).

3.4. Classification performance is superior when comparing PS with TD youth

In previous analyses, all subjects that did not meet criteria for the PS label were considered part of the non-PS group, even if they had other symptoms. Since these subjects could make the classification task more difficult, we investigated if performance could be improved by comparing only the most “typically developing” (TD) subjects in the non-PS group (Fig. 4).

The performance of the classifier trained on ASL rsFC, BOLD rsFC, and ASL + BOLD rsFC data to differentiate between PS and TD individuals performed better than differentiating between PS and non-PS individuals in all three cases; though was only significant for the concatenated ASL and BOLD model ($p = 0.0164$).

4. Discussion

In this study, we found that an SVM classifier trained on ASL connectivity data could differentiate between PS and non-PS youth with reasonable accuracy (mean AUC = 0.68), and to a level not statistically different than BOLD connectivity (mean AUC = 0.67). When ASL and BOLD rsFC were concatenated prior to feature selection, the classifier performed better than either modality alone (mean AUC = 0.72). To our knowledge, this study is the first to investigate the discriminative power of ASL connectivity in youth with early psychosis symptoms, who are largely unmedicated and absent many confounds of clinical samples. The combined ASL and BOLD classifier’s performance improved further when PS youth were compared to typically developing youth (mean AUC = 0.72, with a much more balanced precision-recall curve). An AUC of 0.72 with a balanced precision-recall curve comparing youth with subthreshold symptoms to healthy youth is relatively excellent, and may in part be due to our use of network connectivity of fMRI and ASL, rather than region of interest estimates.

Our results provide some evidence that ASL and BOLD could be complementary modalities to further improve prediction. Our model trained on a concatenation of ASL and BOLD rsFC performed significantly better than BOLD rsFC after feature selection, suggesting some mutually exclusive information relevant to PS is being captured by both modalities. Most machine learning classification work in groups at-risk for psychosis has been conducted using largely structural and

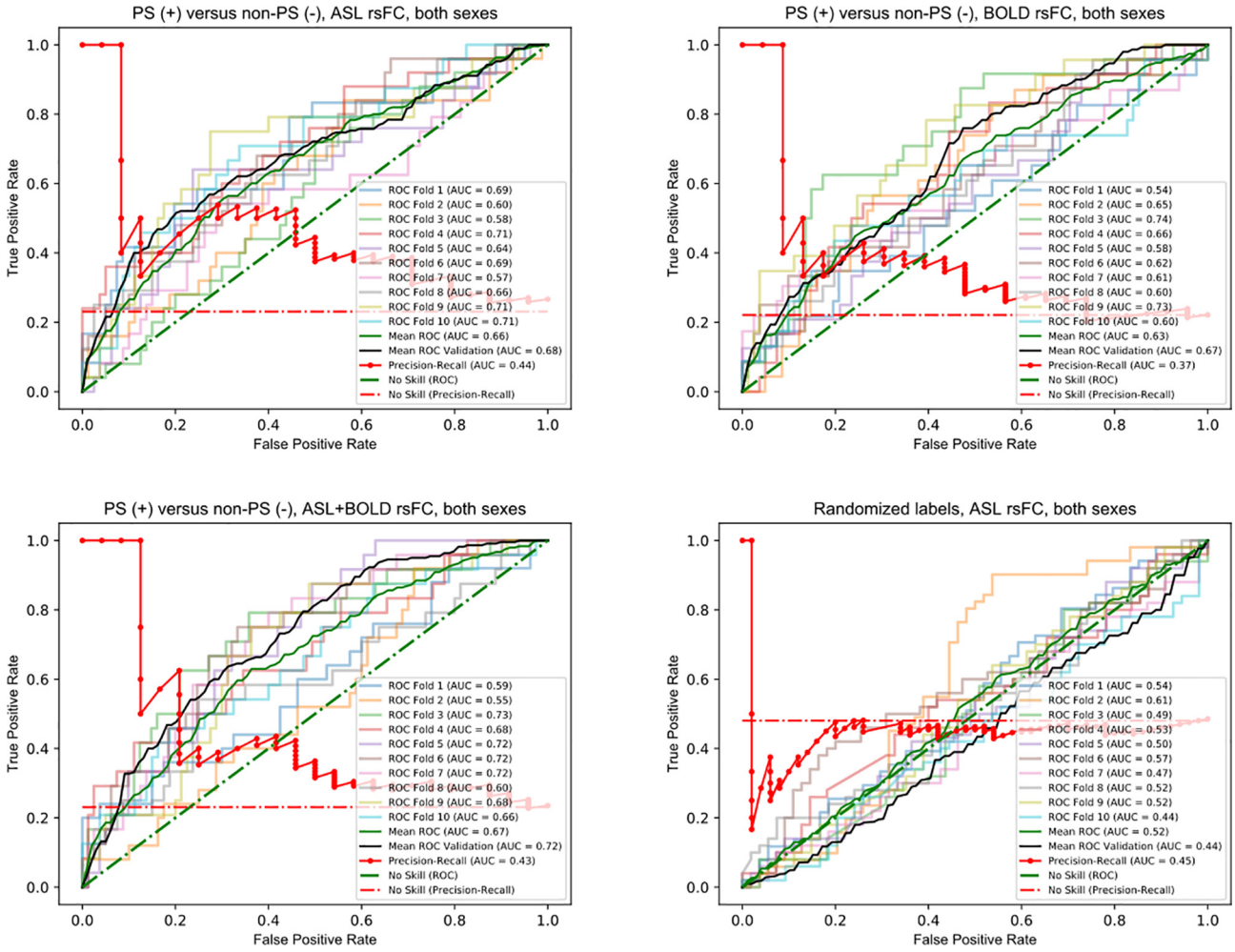

Fig. 1. ROC and precision-recall curves for 10-fold CV, linear kernel SVM classification using ASL or BOLD rsFC as the feature vector. In each subplot, the average ROC curve across all folds against the fold's "test" data is shown in green. Additionally, the average ROC curve for performance against a completely isolated validation set is shown in black. The precision-recall curve is shown is red (and must be compared to the No Skill curve shown in dotted red). **Top left**: ASL rsFC feature vector (mean validation AUC = 0.68). **Top right**: BOLD rsFC feature vector (mean validation AUC = 0.67). **Bottom left**: Concatenated ASL and BOLD rsFC feature vector (mean validation AUC = 0.72). **Bottom right**: Randomized class labels (mean validation AUC = 0.44). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

neuropsychological data, with comparatively little using fMRI, and essentially none using ASL despite its ease of use and brief acquisition time. Calhoun's group found that some network properties (connectivity strength and clustering coefficients) of BOLD rsFC differed between a group of 24 healthy controls and a group of 24 schizophrenia patients, suggesting classification could be possible, and others have confirmed this (Yu et al., 2011; Shen et al., 2010). Though the modalities have similarities, ASL fMRI has different signal characteristics compared to BOLD fMRI (Wang et al., 2003). It has been shown to have a more stable noise profile across the frequency spectrum (whereas BOLD is skewed toward low frequency noise), and lower inter-subject variability for task and resting state activation (see Supplementary Discussion). Further, the present study's large size and opportunity to compare PS youth to a population sample provides a higher confidence that the results are generalizable.

In our results, the most discriminative connections across folds in the ASL and BOLD models appear to involve different brain regions (see Fig. 2): the important ASL connections are within the dorsolateral prefrontal cortex (dlPFC) and several areas within visual and premotor cortex, while the important BOLD connections are found in superior temporal gyrus (Wernicke's area), several areas within the frontal and parietal lobe, and the posterior and anterior cingulate. Some of these areas are consistent with previous network dysfunction work in schizophrenia: altered connectivity in the cingulate is consistently found, as is altered connectivity in parietal regions (Bluhm et al., 2007; Zhou et al., 2008; Garrity et al., 2007; Lynall et al., 2010). At a network level, the important BOLD features seem to represent altered fronto-parietal and default mode network connectivity, while the important ASL features seem to represent mainly altered salience and visual attention network connectivity. Interestingly, the important connections are slightly different in the combined model and draw from both modalities.

There are several reasons why connections may have different feature importances depending on modality. Firstly, the aforementioned ASL signal differences may be more sensitive to different network dysfunction than BOLD; altered CBF functional connectivity has recently been found in regions which include the salience and visual attention networks (Oliveira et al., 2018), and this could be characteristic of CBF in schizophrenia (see Supplementary Discussion). Secondly, although collected at the same site, differences in the data collection protocol between modalities could have resulted in the emphasis or de-emphasis of certain connections (notably, 124 BOLD volumes were collected for each subject, but only 40 spin-labeled and 40 control ASL volumes were collected).

Beyond differences in signal characteristics, ASL and BOLD capture fundamentally different phenomena, and the distinction between these

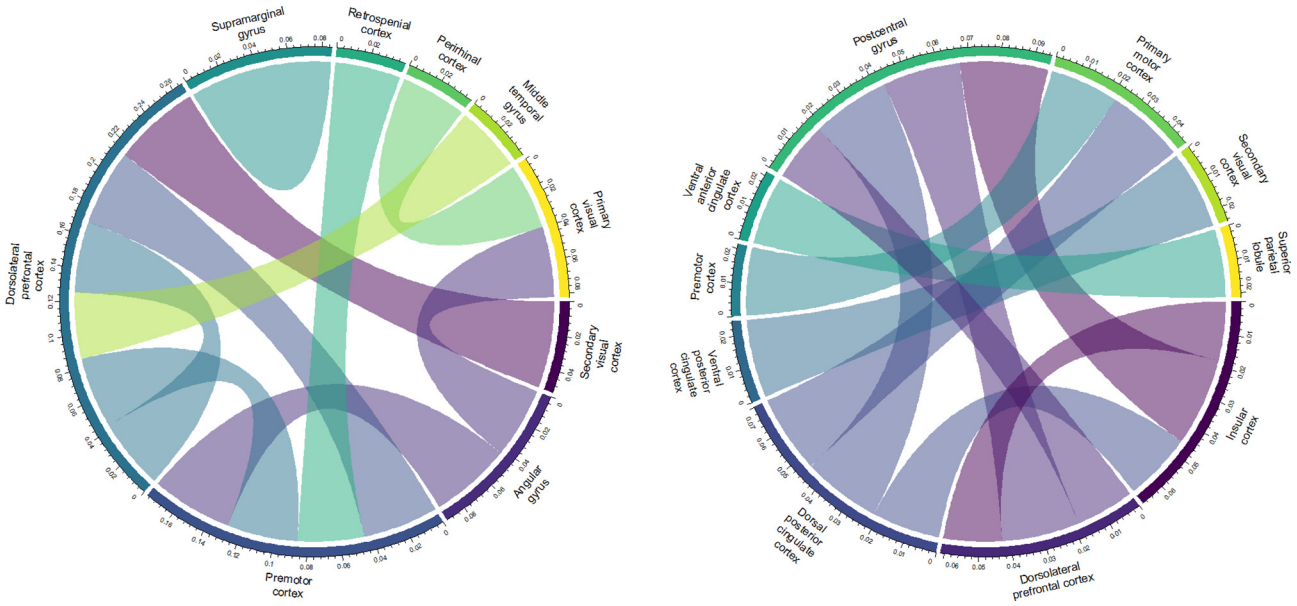

Fig. 2. Most important functional connections for rsFC classification, as determined by permutation importance. ROIs have been mapped to anatomical areas. **Left**: ASL rsFC permutation importance for PS vs non-PS classification. **Right**: BOLD rsFC permutation importance for PS vs non-PS classification. Values are of relative importance only.

phenomena is especially relevant in the developmental period. Indeed, this is a large motivation for developing simultaneous ASL and BOLD MRI acquisition sequences for developmental populations, and authors have speculated that given the rapid rate of development both neurovascularly and neurocognitively in these populations, having access to both at a single timepoint could be very useful for isolating true differences in neurocognition (Schmithorst et al., 2014). Theoretically, the combination of both modalities allows one to determine the extent to which an increased BOLD signal is due to increased CBF versus decreased neuronal oxygen demand at that location. Other authors have done this explicitly, by estimating oxygen metabolic rate using the BOLD signal, ASL signal, blood volume, and Fick's law (see e.g., Davis et al. where this was done in humans; Davis et al., 1998). Though we have chosen not to make any a priori assumptions regarding the importance of each modality type and have included all connections in our model, it is possible that we are also implicitly taking advantage of the BOLD signal while taking CBF into account as a confound, through our feature selection process which includes the dropping of correlated features.

We also examined model classification performance in males and females separately, given that sex-based developmental differences in CBF have previously been observed (Kaczkurkin et al., 2018; Satterthwaite et al., 2014; Taki et al., 2011). We found that females were classified more accurately in all three models. A follow-up analysis, however, revealed that the female group had significantly lower head motion (as measured by mean FD), and that the statistical significance of sex in a logistic regression model of correct classification disappeared when mean FD was taken into account. Thus, the observed differences originally thought to be due to sex are likely due to decreased motion artifacts in females; however, it cannot be ruled out that performance differences would still be observed even if the male group had minimal average head motion. Satterthwaite and colleagues have shown that CBF has a marked age-sex interaction in the inferior parietal lobe, dorsolateral prefrontal cortex (dlPFC), insula, and other areas, with developmental trajectories between males and females diverging after puberty. Since psychosis spectrum disorders typically have a post-puberty onset and a marked sex difference (with prevalence in males being higher than females) (Aleman et al., 2003), one might expect classification performance to differ between sexes especially post-puberty, but we were not able to confirm this due to the group difference in head motion. Even so, our results show that especially in low motion individuals, regardless of sex, ASL has reasonable predictive accuracy for differentiating between PS and non-PS groups.

Some limitations of our results should be kept in mind. Firstly, the data we used from the PNC was cross-sectional and represents only a single "snapshot" of brain function, and thus we must be cautious with extending our results to the prediction of future states in the same individual. Perhaps more importantly, the definition of the psychosis spectrum group was based on scores from the PRIME, KSADS and SOPS and is not a formal diagnosis (as explained, this may have actually hampered classifier performance). If more information about the severity of symptoms existed for each subject (especially data that was not self-reported), it is likely that a cleaner distinction could be made between PS and non-PS groups. Additionally, as mentioned, a group difference in head motion between sexes impaired our ability to investigate functional connectivity differences between sexes in more detail. Strengths of the PNC dataset include the fact that the sample is largely unmedicated and substance-naive, and the age range (8–22) represents a demographic that has not been studied extensively in the transition to psychosis literature (while also being an especially important demographic to investigate in order to understand risk of transitioning to psychosis in individuals that do not pass through the "clinical high-risk" state).

It should be emphasized that all subjects in the PNC have not been formally diagnosed with a psychosis spectrum disorder or as CHR, and thus our task is different (and fundamentally more difficult) than if we had access to ASL and BOLD fMRI data from individuals with more severe psychosis symptoms and a diagnosis. By repeating our methods on such a population, it is likely that differentiability would increase. Along a similar line of reasoning, we saw that performance classifying PS versus TD youth was slightly superior compared to PS versus non-PS youth, even with less data. If a greater amount of TD youth data could be obtained, it is likely this improvement would become statistically significant.

Our results can be extended in several ways. Our classification analysis between PS and non-PS groups could be extended to a regression analysis, predicting not only class membership but symptom severity or Global Functioning score. This would be useful for identifying the highest risk individuals within the PS group and connectivity features differentiating them, and is a more representative metric of risk

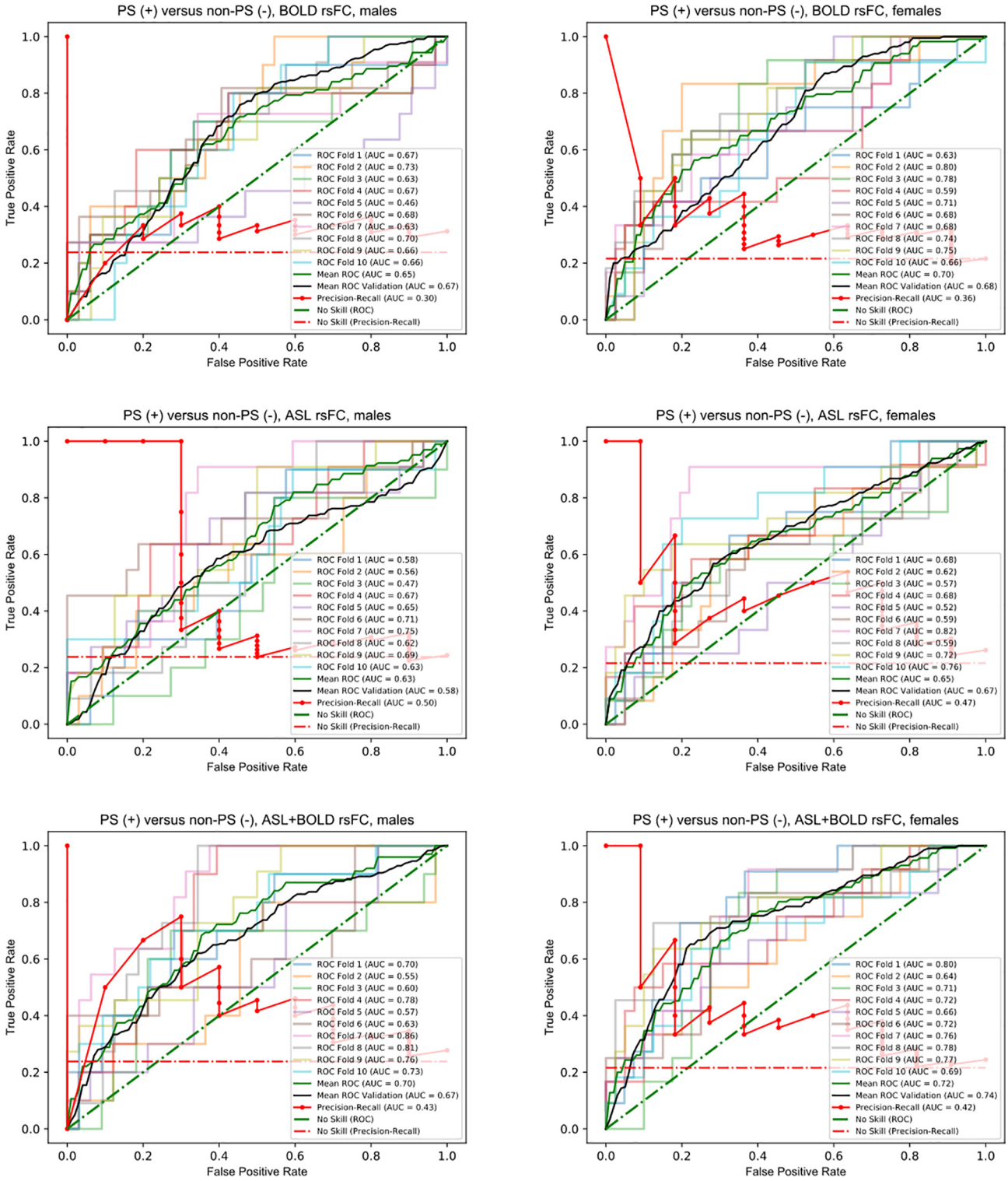

Fig. 3. ROC and precision-recall curves for 10-fold CV, linear kernel SVM classification using ASL or BOLD rsFC as the feature vector, and separated by sex. In each subplot, the average ROC curve across all folds against the fold's "test" data is shown in green. Additionally, the average ROC curve for performance against a completely isolated validation set is shown in black. The precision-recall curve is shown is red (and must be compared to the No Skill curve shown in dotted red). **Left column; male youth**. **Top:** BOLD rsFC feature vector (mean validation AUC = 0.67). **Middle**: ASL rsFC feature vector (mean validation AUC = 0.58). **Bottom**: Concatenated ASL and BOLD rsFC feature vector (mean validation AUC = 0.67). **Right column; female youth. Top:** BOLD rsFC feature vector (mean validation AUC = 0.68). **Middle**: ASL rsFC feature vector (mean validation AUC = 0.67). **Bottom**: Concatenated ASL and BOLD rsFC feature vector (mean validation AUC = 0.74). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

for individuals close to the decision boundary. Additionally, it is possible both classification and regression performance could be improved with different model types and feature selection techniques, especially involving deep neural networks (DNNs), which have been used in classification of schizophrenia from BOLD fMRI (Kim et al., 2016). Though fMRI data is not currently collected for this kind of risk

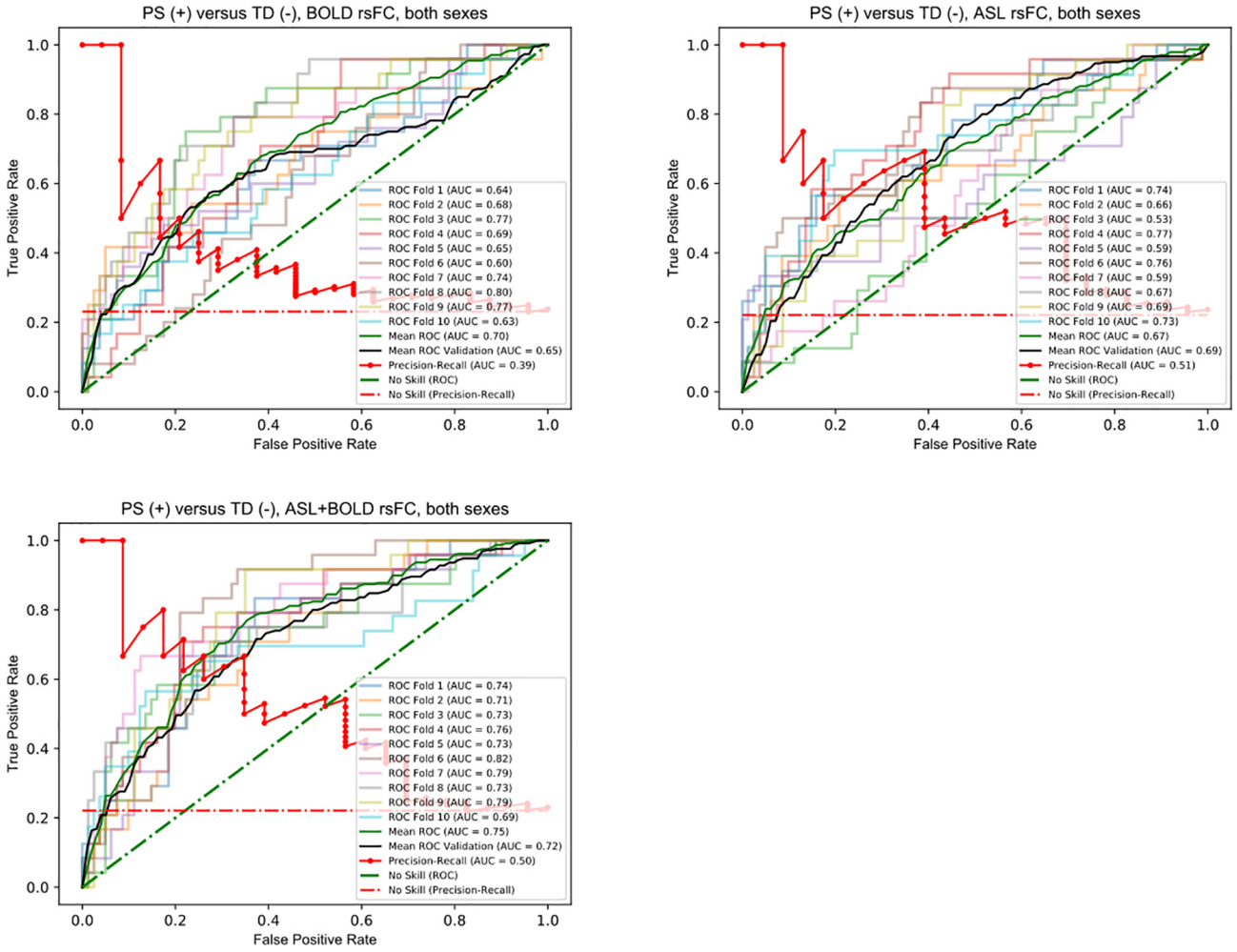

Fig. 4. ROC and precision-recall curves for 10-fold CV, linear kernel SVM classification using ASL or BOLD rsFC as the feature vector to classify between PS and typically developing (TD) youth. In each subplot, the average ROC curve across all folds against the fold's "test" data is shown in green. Additionally, the average ROC curve for performance against a completely isolated validation set is shown in black. The precision-recall curve is shown is red (and must be compared to the No Skill curve shown in dotted red). **Top left**: BOLD rsFC feature vector (mean validation AUC = 0.65). **Top right**: ASL rsFC feature vector (mean validation AUC = 0.69). **Bottom left**: Concatenated ASL and BOLD rsFC feature vector (mean validation AUC = 0.72; see also precision-recall curve). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

assessment clinically, the baseline Global Functioning: Social and Roles scores are easily obtained in a primary care setting and, in the future, could identify candidates for further, more comprehensive risk assessment with fMRI. Building on previous work that effectively utilizes clinical and structural neuroimaging data, there is a strong possibility that combining this data with fMRI modalities for the same individual would lead to more accurate results in PS youth. A recent comprehensive study by Koutsouleris and colleagues (Koutsouleris et al., 2018) showed that by combining predictions from baseline scores on the Global Functioning: Social and Roles scales (Cornblatt et al., 2007) and grey matter volumes (GMV) in CHR youth using stacked generalization (Wolpert, 1992), functional outcome was predicted more accurately than with baseline clinical scores or GMV alone. It remains to be seen if our model could be combined with either similar clinical or structural data (or both) using the same method to improve performance, but this appears to be a promising line of research. Follow up work could validate our findings clinically, by collecting ASL and BOLD data from individuals with poor Global Functioning: Social and Roles baseline scores, determining if classifiers trained on concatenated ASL and BOLD connectivity data could improve outcome prediction compared to baseline scores alone, and if so, if the degree of improvement justifies the cost of additional testing.

5. Funding and disclosure

ANV receives funding from CIHR, NIMH, BBRF, CFI, Ontario MRI, and the CAMH Foundation. Support for the collection of the PNC data was provided by grant RC2MH089983 awarded to Raquel Gur and RC2MH089924 awarded to Hakon Hakonarson. All subjects were recruited through the Center for Applied Genomics at The Children's Hospital in Philadelphia. Database of Genotypes and Phenotypes study accession: phs000607.v2.p2. We want to thank everyone who has worked to share the PNC data. No authors have conflicts of interest to declare.

CRediT authorship contribution statement

Dawson J. Overton: Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Data curation, Writing - original draft, Writing - review & editing, Visualization. **Nikhil Bhagwat:** Conceptualization, Software, Writing - review & editing. **Joseph D. Viviano:** Conceptualization, Methodology, Software, Writing - review & editing. **Grace R. Jacobs:** Writing - review & editing. **Aristotle N. Voineskos:** Conceptualization, Resources, Writing - review & editing, Supervision, Project administration, Funding acquisition.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.nicl.2020.102304.

References

Aleman, A., Kahn, R.S., Selten, J.-P., 2003. Sex differences in the risk of schizophrenia: evidence from meta-analysis. Arch. Gen. Psychiatry 60, 565–571.

Ambrosini, P.J., 2000. Historical development and present status of the schedule for affective disorders and schizophrenia for school-age children (K-SADS). J. Am. Acad. Child Adolesc. Psychiatry 39, 49–58.

Barajas, A., Ochoa, S., Obiols, J.E., Lalucat-Jo, L., 2015. Gender differences in individuals at high-risk of psychosis: a comprehensive literature review. Sci. World J. 2015, 430735.

Birchwood, M., Todd, P., Jackson, C., 1998. Early intervention in psychosis. The critical period hypothesis. Br. J. Psychiatry Suppl. 172, 53–59.

Bluhm, R.L., et al., 2007. Spontaneous low-frequency fluctuations in the BOLD signal in schizophrenic patients: anomalies in the default network. Schizophr. Bull. 33, 1004–1012.

Breiman, L., 2001. Random forests. Mach. Learn. 45, 5–32.

Calkins, M.E., et al., 2015. The Philadelphia Neurodevelopmental Cohort: constructing a deep phenotyping collaborative. J. Child Psychol. Psychiatry 56, 1356–1369.

Calkins, M.E., et al., 2014. The psychosis spectrum in a young U.S. community sample: findings from the Philadelphia Neurodevelopmental Cohort. World Psychiatry 13, 296–305.

Cannon, T.D., et al., 2016. An individualized risk calculator for research in prodromal psychosis. Am. J. Psychiatry 173, 980–988.

Chang, C.-C., Lin, C.-J., 2011. LIBSVM: a library for support vector machines. ACM Trans. Intelligent Syst. Tech (TIST) 2, 27.

Cornblatt, B.A., et al., 2007. Preliminary findings for two new measures of social and role functioning in the prodromal phase of schizophrenia. Schizophr. Bull. 33, 688–702.

Cox, R.W., 1996. AFNI: software for analysis and visualization of functional magnetic resonance neuroimages. Comput. Biomed. Res. 29, 162–173.

Craddock, R.C., James, G.A., Holtzheimer 3rd, P.E., Hu, X.P., Mayberg, H.S., 2012. A whole brain fMRI atlas generated via spatially constrained spectral clustering. Hum. Brain Mapp. 33, 1914–1928.

Davis, T.L., Kwong, K.K., Weisskoff, R.M., Rosen, B.R., 1998. Calibrated functional MRI: mapping the dynamics of oxidative metabolism. Proc. Natl. Acad. Sci. U. S. A. 95, 1834–1839.

Detre, J.A., Wang, J., Wang, Z., Rao, H., 2009. Arterial spin-labeled perfusion MRI in basic and clinical neuroscience. Curr. Opin. Neurol. 22, 348–355.

Fusar-Poli, P., et al., 2013. The psychosis high-risk state: a comprehensive state-of-the-art review. JAMA Psychiatry 70, 107–120.

Garrity, A.G., et al., 2007. Aberrant 'Default Mode' functional connectivity in schizophrenia. AJP 164, 450–457.

Gennatas, E.D., et al., 2017. Age-related effects and sex differences in gray matter density, volume, mass, and cortical thickness from childhood to young adulthood. J. Neurosci. 37, 5065–5073.

Gur, R.E., Gur, R.C., 2016. Sex differences in brain and behavior in adolescence: findings from the Philadelphia Neurodevelopmental Cohort. Neurosci. Biobehav. Rev. 70, 159–170.

Hanley, J.A., McNeil, B.J., 1982. The meaning and use of the area under a receiver operating characteristic (ROC) curve. Radiology 143, 29–36.

Jacobs, G.R., et al., 2019. Developmentally divergent sexual dimorphism in the cortico-striatal-thalamic-cortical psychosis risk pathway. Neuropsychopharmacology. https://doi.org/10.1038/s41386-019-0408-6.

Kaczkurkin, A.N., Raznahan, A., Satterthwaite, T.D., 2018. Sex differences in the developing brain: insights from multimodal neuroimaging. Neuropsychopharmacology. https://doi.org/10.1038/s41386-018-0111-z.

Kim, J., Calhoun, V.D., Shim, E., Lee, J.-H., 2016. Deep neural network with weight sparsity control and pre-training extracts hierarchical features and enhances classification performance: Evidence from whole-brain resting-state functional connectivity patterns of schizophrenia. Neuroimage 124, 127–146.

Koutsouleris, N., et al., 2009. Use of neuroanatomical pattern classification to identify subjects in at-risk mental states of psychosis and predict disease transition. Arch. Gen. Psychiatry 66, 700–712.

Koutsouleris, N., et al., 2012. Early recognition and disease prediction in the at-risk mental states for psychosis using neurocognitive pattern classification. Schizophr. Bull. 38, 1200–1215.

Koutsouleris, N., et al., 2018. Prediction models of functional outcomes for individuals in the clinical high-risk state for psychosis or with recent-onset depression: a multimodal, multisite machine learning analysis. JAMA Psychiatry 75, 1156–1172.

Loggia, M.L., et al., 2013. Default mode network connectivity encodes clinical pain: an arterial spin labeling study. Pain 154, 24–33.

Lynall, M.-E., et al., 2010. Functional connectivity and brain networks in schizophrenia. J. Neurosci. 30, 9477–9487.

Mendrek, A., Mancini-Marïe, A., 2016. Sex/gender differences in the brain and cognition in schizophrenia. Neurosci. Biobehav. Rev. 67, 57–78.

Miller, T.J., et al., 2003. Prodromal assessment with the structured interview for prodromal syndromes and the scale of prodromal symptoms: predictive validity, inter-rater reliability, and training to reliability. Schizophr. Bull. 29, 703–715.

Oliveira, Í.A.F., et al., 2018. Brain functional and perfusional alterations in schizophrenia: an arterial spin labeling study. Psychiatry Res Neuroimaging 272, 71–78.

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

Pereira, F., Mitchell, T., Botvinick, M., 2009. Machine learning classifiers and fMRI: a tutorial overview. Neuroimage 45, S199–S209.

Raschka, S., 2018. MLxtend: providing machine learning and data science utilities and extensions to Python's scientific computing stack. JOSS 3, 638.

Satterthwaite, T.D., et al., 2012. Impact of in-scanner head motion on multiple measures of functional connectivity: relevance for studies of neurodevelopment in youth. Neuroimage 60, 623–632.

Satterthwaite, T.D., et al., 2014. Impact of puberty on the evolution of cerebral perfusion during adolescence. Proc. Natl. Acad. Sci. U. S. A. 111, 8643–8648.

Satterthwaite, T.D., et al., 2014. Neuroimaging of the Philadelphia neurodevelopmental cohort. Neuroimage 86, 544–553.

Satterthwaite, T.D., et al., 2015. Linked sex differences in cognition and functional connectivity in youth. Cereb. Cortex 25, 2383–2394.

Satterthwaite, T.D., et al., 2016. Structural brain abnormalities in youth with psychosis spectrum symptoms. JAMA Psychiatry 73, 515–524.

Schmithorst, V.J., et al., 2014. Optimized simultaneous ASL and BOLD functional imaging of the whole brain. J. Magn. Reson. Imaging 39, 1104–1117.

Scholkopf, B., Smola, A.J., 2001. Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond. MIT Press.

Shen, X., Tokoglu, F., Papademetris, X., Constable, R.T., 2013. Groupwise whole-brain parcellation from resting-state fMRI data for network node identification. Neuroimage 82, 403–415.

Shen, H., Wang, L., Liu, Y., Hu, D., 2010. Discriminative analysis of resting-state functional connectivity patterns of schizophrenia using low dimensional embedding of fMRI. Neuroimage 49, 3110–3121.

Smith, S.M., et al., 2004. Advances in functional and structural MR image analysis and implementation as FSL. Neuroimage 23 (Suppl 1), S208–S219.

Taki, Y., et al., 2011. Correlation between gray matter density-adjusted brain perfusion and age using brain MR images of 202 healthy children. Hum. Brain Mapp. 32, 1973–1985.

Van Dijk, K.R.A., Sabuncu, M.R., Buckner, R.L., 2012. The influence of head motion on intrinsic functional connectivity MRI. Neuroimage 59, 431–438.

Wang, J., et al., 2003. Arterial spin labeling perfusion fMRI with very low task frequency. Magn. Reson. Med. 49, 796–802.

Wolpert, D.H., 1992. Stacked Generalization.

Wu, W.-C., Fernández-Seara, M., Detre, J.A., Wehrli, F.W., Wang, J., 2007. A theoretical and experimental investigation of the tagging efficiency of pseudocontinuous arterial spin labeling. Magn. Reson. Med. 58, 1020–1027.

Yu, Q., et al., 2011. Modular organization of functional network connectivity in healthy controls and patients with schizophrenia during the resting state. Front. Syst. Neurosci. 5, 103.

Zeng, L.-L., et al., 2012. Identifying major depression using whole-brain functional connectivity: a multivariate pattern analysis. Brain 135, 1498–1507.

Zhou, Y., et al., 2008. Altered resting-state functional connectivity and anatomical connectivity of hippocampus in schizophrenia. Schizophr. Res. 100, 120–132.