

Supplementary Materials for

Classification of schizophrenia spectrum disorder using machine learning and functional connectivity: reconsidering the clinical application

Content

Methods.....	1
<i>Inclusion and exclusion criteria</i>	1
<i>Image acquisition</i>	2
<i>Machine learning using other three strategies</i>	3
<i>Statistical analysis for demographic and clinical information of chronic SSD, first episode medicated SSD and first episode unmedicated SSD in dataset 1</i>	4
<i>Correction of site and covariates</i>	4
Results.....	5
<i>Classification performances using other three strategies</i>	5
<i>Demographic and clinical information of chronic SSD, first episode medicated SSD and first episode unmedicated SSD in dataset 1</i>	6
<i>Correction of site and covariates</i>	7
Figure S4. Classification performances of model trained using chronic SSD from dataset 2, 3 and 4 to unmedicated subgroups in dataset 1 and their duration.	8
Figure S5. Classification performances of model trained using chronic SSD from dataset 2, 3 and 4 to unmedicated SSD, unmedicated schizophreniform and unmedicated SZ in dataset 1 when using other threshold to define first episode and chronic SSD.	9
Figure S6. Classification performances	10
Figure S7. Age, gender and head motion between SSD and HCs in training datasets and test dataset.	11
Figure S8. Correlations between illness duration and age.	12

Methods

Inclusion and exclusion criteria

Dataset 1: The presence or absence of Axis I psychiatric diagnoses was determined by two trained psychiatrists using the Structured Clinical Interview for Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition (DSM-IV), Axis I Disorders for participants 18 years and older, whereas the Schedule for Affective Disorders and Schizophrenia for School-Age Children–Present and Lifetime

Version (K-SADS-PL) was used for participants younger than 18 years. All patients with schizophrenia or schizophreniform were required to meet the DSM-IV diagnostic criteria for their respective disorders and to have no other Axis I disorders. The healthy controls did not have a current or lifetime history of any Axis I disorder or a history of psychotic, mood, or other Axis I disorders in first-degree relatives as determined from a detailed family history. Potential participants were excluded if they had (1) a lifetime history of substance/alcohol abuse or dependence, (2) a concomitant major medical disorder, (3) any MRI contraindications, (4) a history of head trauma with loss of consciousness ≥ 5 minutes or any neurological disorder, or (5) any abnormality identified by T1- or T2-weighted imaging.

Dataset 2: The presence or absence of Axis I psychiatric diagnoses was determined by two experienced psychiatrists according to the Structured Clinical Interview for Diagnostic and the Statistical Manual of Mental Disorders, Fourth Edition, Text Revision (DSM-IV-TR) Axis I Disorders (SCID, patient edition). Patients with any other neurological disorder, a history of severe medical illness, substance dependence, pregnancy, or treatment with electroconvulsive therapy within the past 6 months, and those with a diagnosis of any other Axis I disorder, were excluded. The healthy controls were screened using SCID (non-patient edition) and recruited if they had no history of mental and/or neurological disorder, drug or alcohol abuse, traumatic brain injury, or visible brain lesions on conventional MRI.

Image acquisition

Dataset 1: MRI data were acquired using a GE Signa HD 3.0-T scanner (General Electric, Milwaukee, WI) with a standard 8-channel head coil at the First Affiliated Hospital of China Medical University. Functional imaging was performed using a gradient-echo-planar imaging (EPI-GRE) sequence. The following parameters were used: repetition time = 2000 ms, echo time = 30 ms, flip angle = 90 °, field of view = 240 mm \times 240 mm, matrix = 64 \times 64, slice thickness = 3 mm with no gap, slices = 35. The scan lasted 6 minutes and 40 seconds, resulting in 200 volumes. The participants were instructed to rest and relax with their eyes closed but to remain awake during the scan.

Dataset 2: MRI scans were performed on a Siemens 3.0 Tesla Trio magnetic resonance scanner (Siemens Medical Systems, Erlangen, Germany) in Peking University Third Hospital. The functional imaging data were acquired with the following parameters: repetition time = 2000 ms, echo time = 30

ms, flip angle = 90 °, field of view = 220mm × 220 mm, matrix = 64 × 64, slice thickness = 4.0 mm with no gap, slices = 33 slices. The scan lasted 8 minutes, resulting in 240 volumes.. Before scanning, all participants were instructed to move as little as possible, keep their eyes closed, think of nothing in particular, and avoid falling asleep. After scanning, they were asked whether they fell asleep to reconfirm.

Dataset 3: MR data were acquired on a 3.0 Tesla Tim Trio scanner (Siemens, Erlangen, Germany). The functional imaging data were acquired using single-shot full k-space echo-planar imaging (EPI) with ramp sampling correction using the intercomissural line (AC-PC) as a reference with the following parameters: repetition time = 2000 ms, TE = 29 ms, flip angle = 75 °, field of view = 192 mm × 192 mm, matrix = 64 × 64, slice thickness = 4 mm with no gap, slices = 32. The scan lasted 5 minutes, resulting in 150 volumes.

Dataset 4: MR data were acquired on a 3 0-T Tim Trio scanner (Siemens, Erlangen, Germany). The functional imaging data were acquired using an asymmetrical spin-echo, echo-planar sequence (T2*) with the following parameters: repetition time = 2000 ms, TE = 30 ms, flip angle = 90 °, field of view = 192 mm × 192 mm, matrix = 64 × 64, slice thickness = 4 mm with no gap, slices = 34. The scan lasted 5 minutes, resulting in 152 volumes. Subjects were instructed to rest quietly with their eyes closed.

Machine learning using other three strategies

For the each fold of the 5-fold cross-validation that pooled all datasets together, z normalization (divide by mean and subtract standard deviation) was applied to each participant in the individual level. We used principal component analysis (PCA) to reduce feature dimensionality. The top principal components with the highest eigenvalues that cumulatively explained 95% of the variance were selected. Then the selected principal components were fed into a linear support vector machine (SVM) classifier (regularization parameter $C = 1$) for training a classification model. We used the trained model to classify the unseen test data (1/5) and evaluated the classification performance of the model. After 5-fold training and testing loop, every participant will have a prediction label and a real label. According to these labels, we further evaluated the classification performance of each subgroup for dataset 1 and 2, e.g., unmedicated SSD or unmedicated schizophreniform subgroups.

For the each fold of the leave-one-site-out cross-validation, we used a given dataset as the test

data and the other three datasets as the training data. The rest of the machine learning processes were similar to that of pooling all datasets together.

For the each fold of the 5-fold cross-validation that only included first episode unmedicated subgroup in dataset 1, the machine learning processes were the same as that of pooling all datasets together.

Statistical analysis for demographic and clinical information of chronic SSD, first episode medicated SSD and first episode unmedicated SSD in dataset 1

Analysis of variance and two-sample t-test were used for the statistical analysis of continuous variables. Chi-square test was used for the statistical analysis of categorical variable. The analysis code are available here: https://github.com/easylearn-fmri/easylearn/tree/master/eslearn/statistical_analysis.

Correction of site and covariates

Site, age, gender and head motion may have a potential impact on our main findings. In this study, the illness duration was a factor of interest, because we aimed to investigate whether or not the classification model trained using chronic medicated SSD can be generalized to early stage SSD. Therefore, given that the age had strong correlation with illness duration (Figure S8), we did not regress out the effect of age. Before regression, we excluded subjects with head motion (FD) greater than 0.5 (Remained number of subjects in dataset 1=536 [SSD=268, HC=268], dataset 2=199 [SSD=99, HC=100], dataset 3=77 [SSD=27, HC=50], dataset 4=143 [SSD=34, HC=109]). Note that we set `class_weight` to 'balanced' for `svm.LinearSVC` in the scikit-learn to solve the problem of unbalance between SSD and HC.

Next, we used linear regression model to regress out effect of site, gender and head motion for all subjects. Specifically, we fitted a linear regression model taking functional connectivity as dependent variables, and taking site, gender, and head motion as independent variables. Then, we calculated the residual of the model and used it as a feature of machine learning. The process of machine learning was the same as the main text.

Furthermore, regressing out effect of site, gender and head motion for all subjects together may introduce the risk data leakage between training dataset (dataset 2, 3 and 4) and test dataset (dataset 1).

Therefore, we additionally estimated the effect of gender and head motion just limited in the training dataset, and then were applied the estimated parameters (beta values) to all subjects.

Results

Classification performances using other three strategies

For the 5-fold cross-validation that pooled all datasets, the classification model can achieve 77% (\pm 1.7%, $p < 0.0001$) accuracy, 74% (\pm 2.0%) sensitivity, 0.79 (\pm 2.9%) specificity and 0.86 (\pm 1.4%) AUC. For the leave-one-site-out cross-validation, the classification model can achieve 72% (\pm 2.5%, $p < 0.0001$) accuracy, 72% (\pm 7.6%) sensitivity, 0.74 (\pm 4.4%) specificity and 0.80 (\pm 4.1%) AUC. For the five-fold cross-validation of first episode unmedicated SSD, the classification model can achieve 71% (\pm 7.7%, $p = 0.0004$) accuracy, 71% (\pm 9.8%) sensitivity, 0.73 (\pm 18%) specificity and 0.76 (\pm 14%) AUC. The above performances are shown in Figure S1.

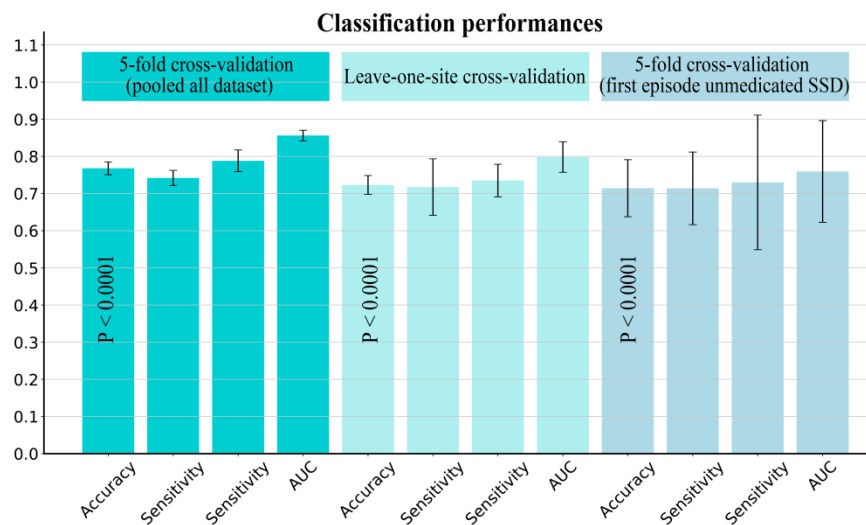


Figure S1. (A) Classification performances of 5-fold cross-validation that pooled all datasets, leave-one-site-out cross-validation as well as 5-fold cross-validation that only recruited first episode unmedicated schizophrenia spectrum disorder. Numbers in the bars were p values of accuracy. The error bar indicates the standard deviation.

Demographic and clinical information of chronic SSD, first episode medicated SSD and first episode unmedicated SSD in dataset 1

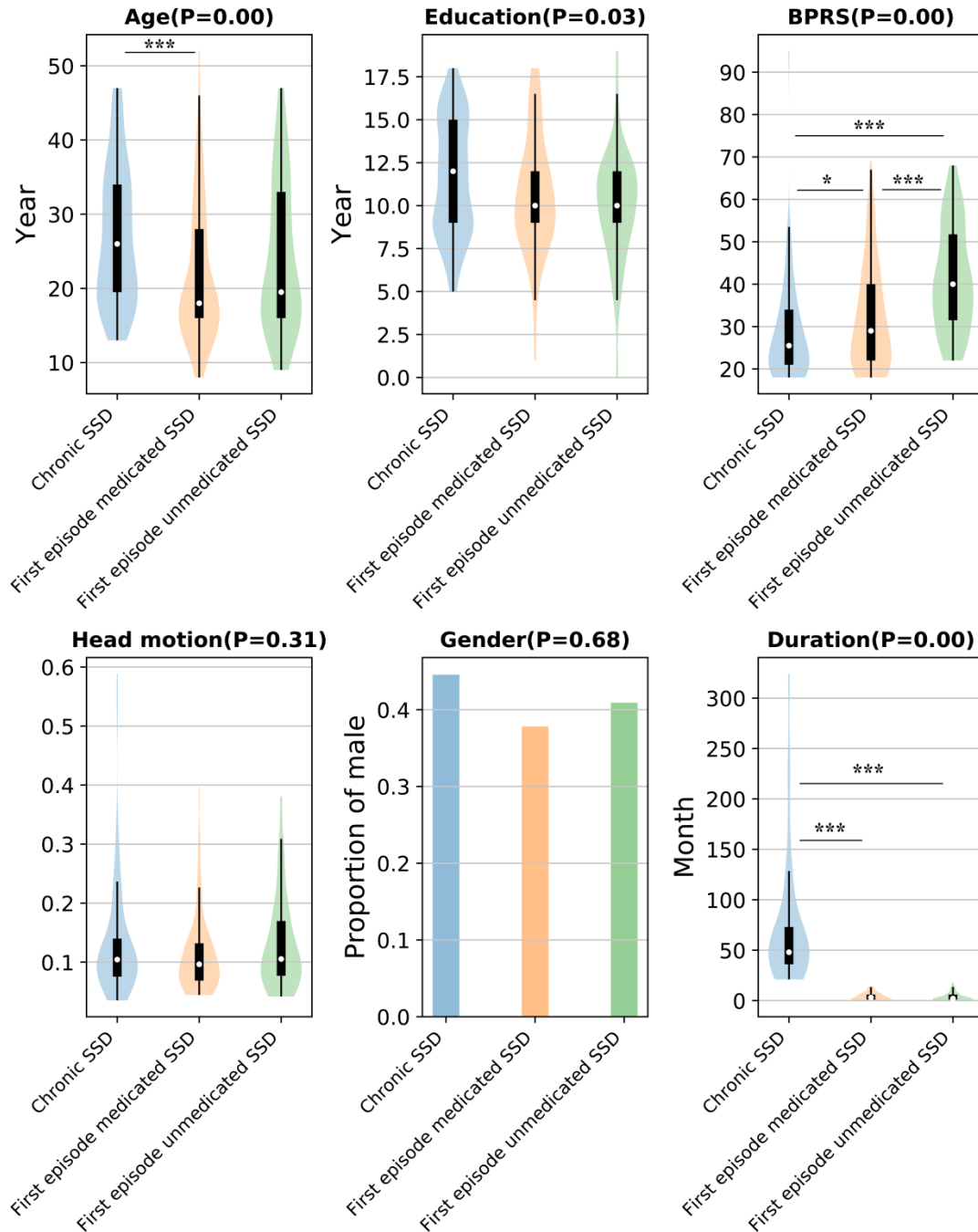


Figure S2. Demographic and clinical information of chronic SSD, first episode medicated SSD and first episode unmedicated SSD. The p value in brackets in the figure title was derived from the analysis of variance. The asterisk indicated there was statistically significance in two-sample t-test.

SSD=Schizophrenia Spectrum Disorder; BPRS=Brief Psychiatric Rating Scale. * indicate $P < 0.05$, **

indicate $P < 0.01$, *** indicate $P < 0.001$.

Correction of site and covariates

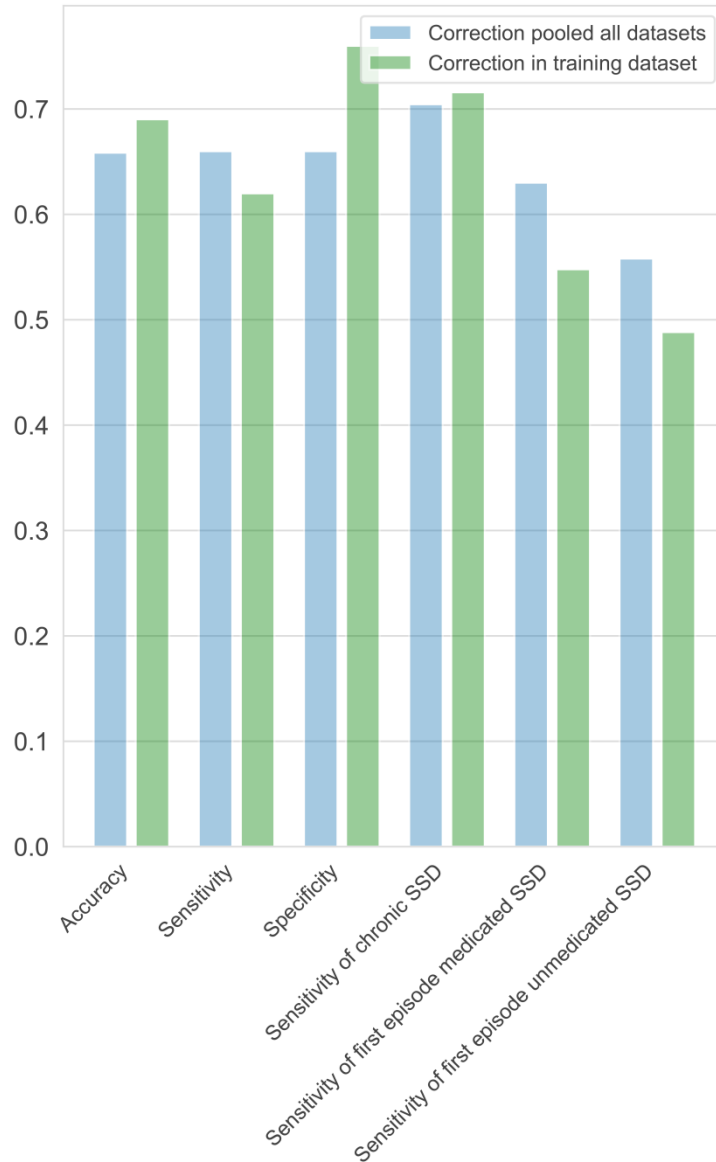


Figure S3. Classification performances of model trained using chronic SSD from dataset 2, 3 and 4 to dataset 1 after regressing out gender, age and head motion and/or site. Legend of “correction pooled all datasets” indicates performances after regressing out effect of site, gender and head motion for all subjects together. Legend of “correction in training dataset” indicates performances that effect of gender and head motion estimated just limited in the training dataset, and then were applied to all subjects. SSD=Schizophrenia Spectrum Disorder.

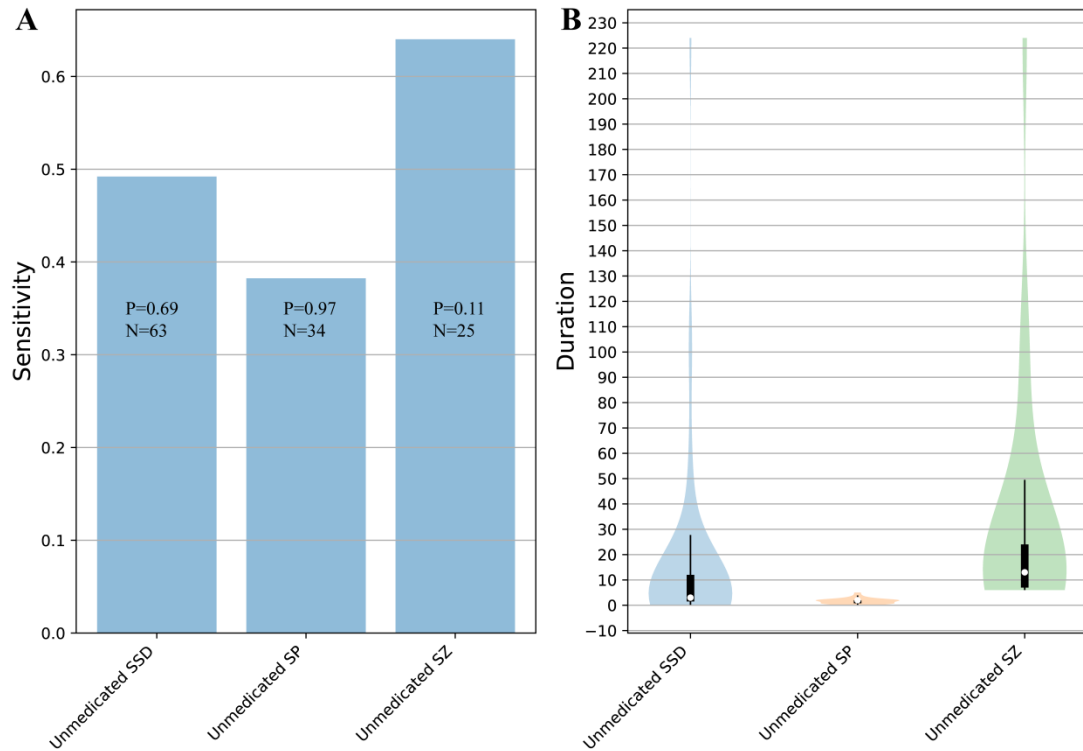


Figure S4. Classification performances of model trained using chronic SSD from dataset 2, 3 and 4 to unmedicated subgroups in dataset 1 and their duration. (A) Classification model trained using chronic SSD from dataset 2, 3 and 4 cannot be generalized to unmedicated SSD, unmedicated schizophreniform and unmedicated SZ in dataset 1. Numbers in the bars were sample size or P values of sensitivity. Four SSD missed illness duration information. (B) Illness duration (month) of unmedicated SSD, unmedicated schizophreniform and unmedicated SZ in dataset 1.

SSD=Schizophrenia Spectrum Disorder; SP=schizophreniform; SZ=Schizophrenia.

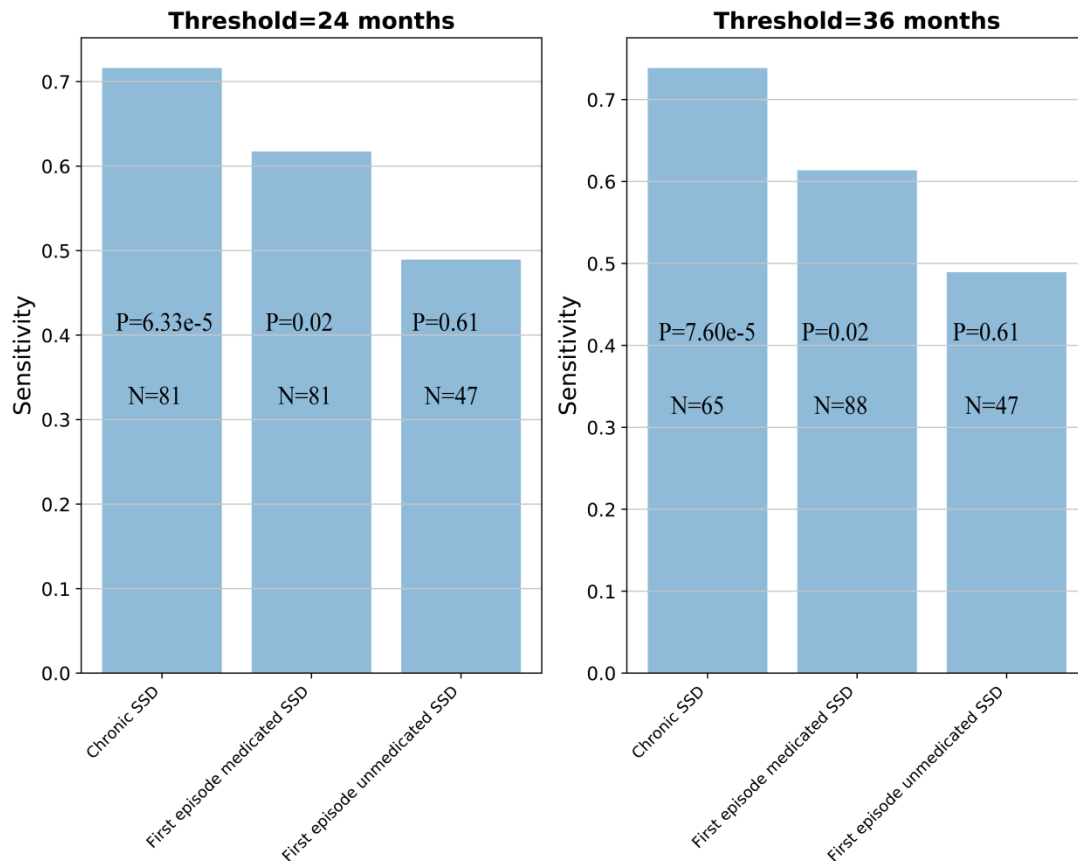


Figure S5. Classification performances of model trained using chronic SSD from dataset 2, 3 and 4 to unmedicated SSD, unmedicated schizophreniform and unmedicated SZ in dataset 1 when using other threshold to define first episode and chronic SSD. When used 24 months as threshold to define the first episode and chronic SSD, the model trained using chronic SSD from dataset 2, 3 and 4 identified unmedicated SSD, unmedicated schizophreniform and unmedicated SZ in dataset 1 with 72%, 62% and 49% sensitivity respectively. When used 36 months as threshold to define the first episode and chronic SSD, the model trained using chronic SSD from dataset 2, 3 and 4 identified unmedicated SSD, unmedicated schizophreniform and unmedicated SZ in dataset 1 with 74%, 61% and 49% sensitivity respectively. Numbers in the bars were sample size or P values of sensitivity.

SSD=Schizophrenia Spectrum Disorder.

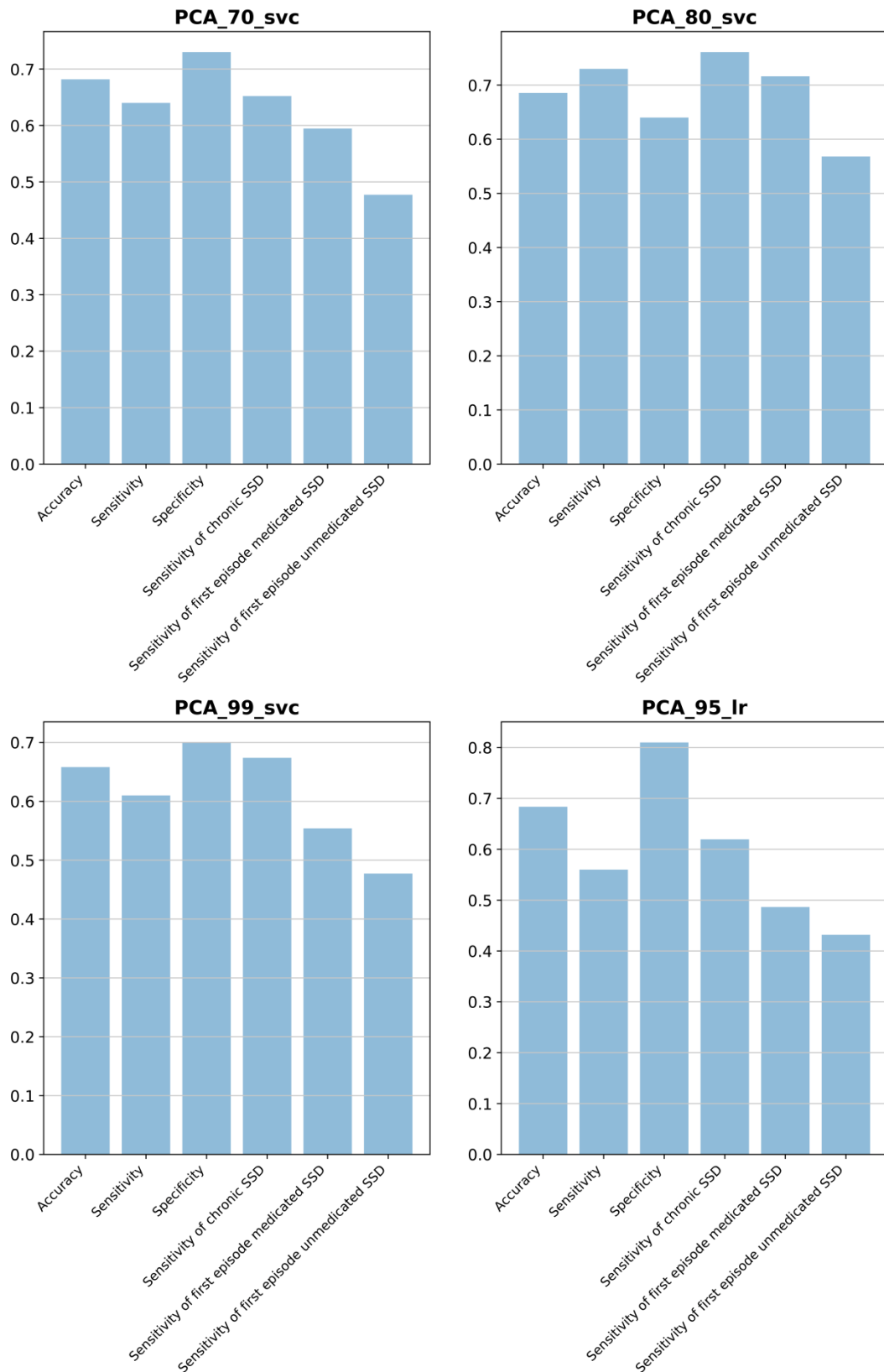


Figure S6. Classification performances of model trained using chronic SSD from dataset 2, 3 and

4 to dataset 1 when using support vector classifier combined with principal components with

70%, 80% and 99% explained variance and using logistic regression classifier combined with principal components with 95% explained variance. PCA_70_svc= Principal component analysis keeping the top 70% components and using support vector classifier. PCA_80_svc= Principal component analysis keeping the top 80% components and using support vector classifier. PCA_99_svc= Principal component analysis keeping the top 99% components and using support vector classifier. PCA_95_lr= Principal component analysis keeping the top 95% components and using logistic regression classifier. SSD=Schizophrenia Spectrum Disorder.

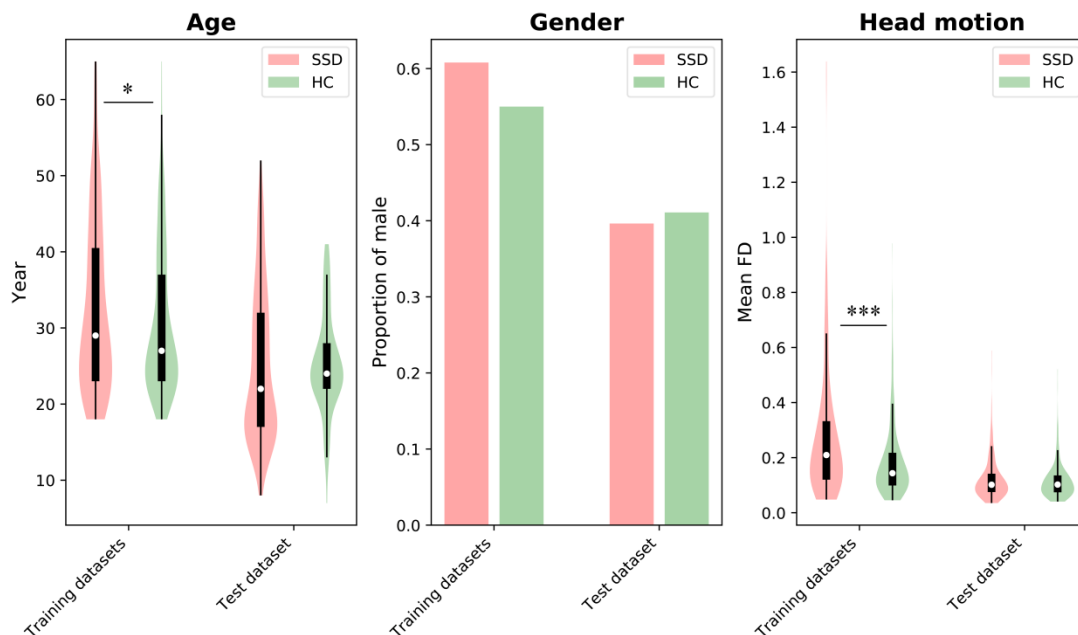


Figure S7. Age, gender and head motion between SSD and HCs in training datasets and test dataset. The asterisk indicated there was statistically significance in two-sample t-test or chi-square test. Training datasets denote dataset 2, 3 and 4; test dataset denotes dataset 1. SSD=Schizophrenia Spectrum Disorder; HC=Healthy control. * indicate $P < 0.05$, *** indicate $P < 0.001$.

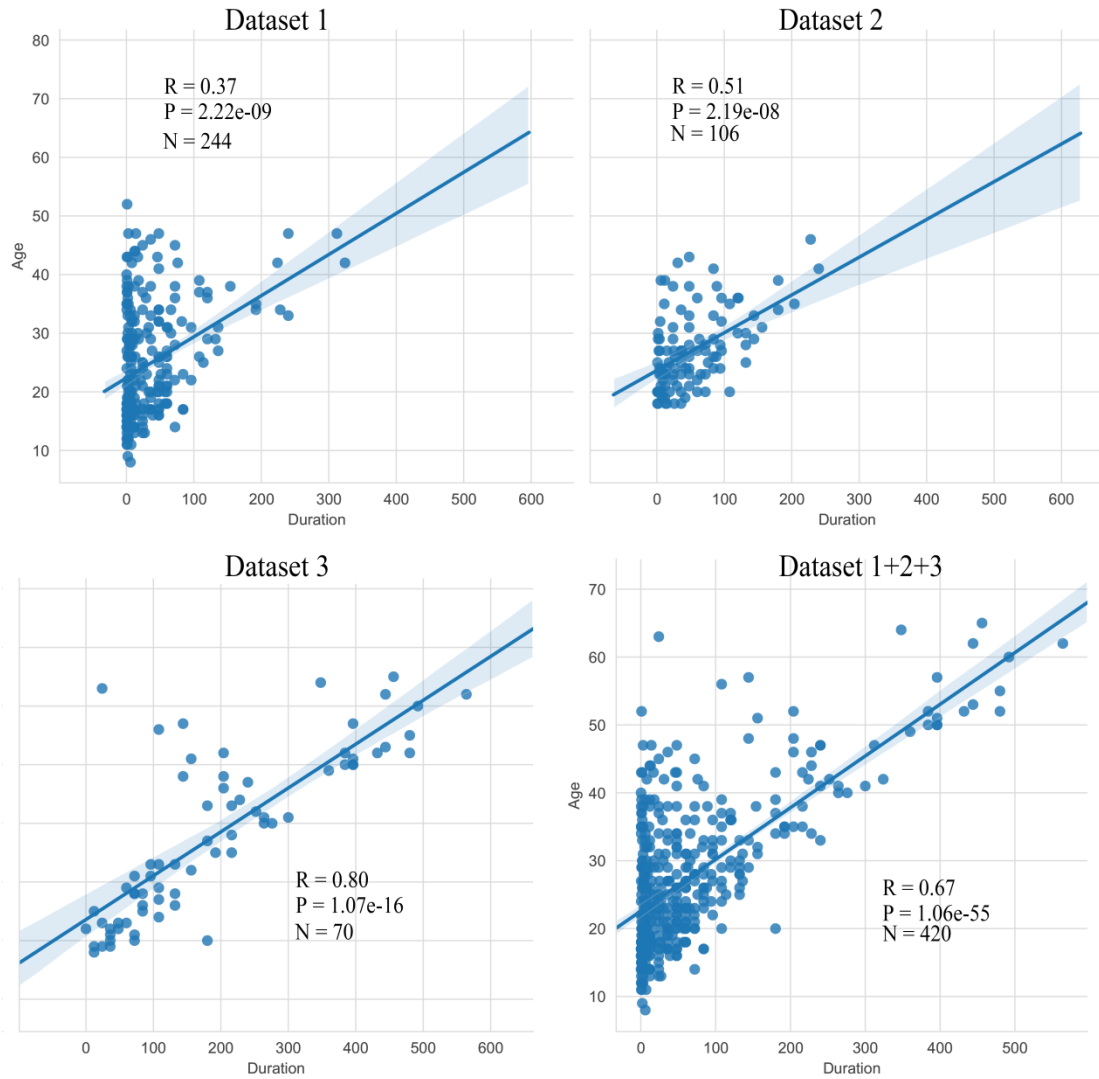


Figure S8. Correlations between illness duration and age. Correlation analyses were performed using Pearson's correlation analysis. R, P and N denote coefficient, p value as well as number of patients with schizophrenia spectrum disorder that had illness duration information respectively.