Schizophrenia Bulletin vol. 38 no. 6 pp. 1234–1246, 2012
doi:10.1093/schbul/sbr145
Advance Access publication November 10, 2011

Disease Prediction in the At-Risk Mental State for Psychosis Using Neuroanatomical Biomarkers: Results From the *FePsy* Study

Nikolaos Koutsouleris[*,1], Stefan Borgwardt[2], Eva M. Meisenzahl[1], Ronald Bottlender[1], Hans-Jürgen Möller[1], and Anita Riecher-Rössler[2]

[1]Department of Psychiatry and Psychotherapy, Ludwig-Maximilian-University, Nussbaumstrasse 7, 80336 Munich, Germany;
[2]Department of Psychiatry, University of Basel, Basel, Switzerland

*To whom correspondence should be addressed; tel: +49-89-5160-5717, fax: +49-89-5160-3413, e-mail: Nikolaos.Koutsouleris@med.uni-muenchen.de

Background: **Reliable prognostic biomarkers are needed for the early recognition of psychosis. Recently, multivariate machine learning methods have demonstrated the feasibility to predict illness onset in clinically defined at-risk individuals using structural magnetic resonance imaging (MRI) data. However, it remains unclear whether these findings could be replicated in independent populations.** ***Methods:*** **We evaluated the performance of an MRI-based classification system in predicting disease conversion in at-risk individuals recruited within the prospective *FePsy* (*Früherkennung von Psychosen*) study at the University of Basel, Switzerland. Pairwise and multigroup biomarkers were constructed using the MRI data of 22 healthy volunteers, 16/21 at-risk subjects with/without a subsequent disease conversion. Diagnostic performance was measured in unseen test cases using repeated nested cross-validation.** ***Results:*** **The classification accuracies in the "healthy controls (HCs) vs converters," "HCs vs nonconverters," and "converters vs nonconverters" analyses were 92.3%, 66.9%, and 84.2%, respectively. A positive likelihood ratio of 6.5 in the converters vs nonconverters analysis indicated a 40% increase in diagnostic certainty by applying the biomarker to an at-risk population with a transition rate of 43%. The neuroanatomical decision functions underlying these results particularly involved the prefrontal perisylvian and subcortical brain structures.** ***Conclusions:*** **Our findings suggest that the early prediction of psychosis may be reliably enhanced using neuroanatomical pattern recognition operating at the single-subject level. These MRI-based biomarkers may have the potential to identify individuals at the highest risk of developing psychosis, and thus may promote informed clinical strategies aiming at preventing the full manifestation of the disease.**

Key words: early recognition/psychosis/machine learning/magnetic resonance imaging/biomarkers

Introduction

Therapeutic action in the earliest phase of schizophrenia and other psychoses may be the most beneficial strategy to modify the subsequent clinical course of the affected individuals,[1] with the potential to alleviate symptom burden and even prevent the manifestation of the frank disorder.[2–4] However, possible side effects and socioeconomic impacts of preventive treatment in the at-risk mental state for psychosis (ARMS) require therapeutic decisions to be based on solid diagnostic grounds in order to reliably target those individuals with the highest probability of developing an overt psychotic disease. Thus, valid early recognition instruments are needed that are capable of detecting subtle disease-associated signals at the single-subject level across heterogeneous subclinical populations. These instruments could provide an objective rationale for clinical decision making in the ARMS and the prodromal phase of the disorder.

Recently, multivariate disease prediction algorithms have emerged as a potential means to derive diagnostic and prognostic decisions from different sets of clinical and neurocognitive measures.[5–8] These algorithms may have the potential to increase the prediction accuracy of the established, operationalized early recognition inventories from 9% to 54%[9] to over 80%. However, these clinical algorithms typically require a thorough psychopathological assessment of subtle and thus difficult to ascertain signs and symptoms.[10] Therefore, the applicability of clinical prognostic tools largely depends on skilled personnel working within a limited number of highly specialized mental health care facilities. Hence, biomarker-based early recognition tools may complement and extend the existing early detection strategies by providing objective methods to evaluate the risk of developing overt psychosis in vulnerable individuals.

In this regard, Job et al[11] were the first to assess the feasibility of magnetic resonance imaging (MRI)-based psychosis prediction in a genetically defined ARMS population. They found that longitudinal gray matter (GM) density reductions in the inferior temporal gyrus predicted subsequent disease manifestation with a positive/negative predictive value of 60%/92%. However, the delay of preventive treatment caused by the necessity to perform repeated MRI scanning and the modest sensitivity of the underlying methodology may limit the validity of such biomarkers within a clinical real-world scenario. In this context, neuroimaging-based multivariate pattern recognition algorithms such as the support vector machine (SVM) may surmount these limitations as they have shown to reliably detect the subsequent onset of different neurodegenerative disorders during clinical and preclinical stages (see[12] for review).

Our own previous work[13] suggested that the diagnosis of the ARMS and the prediction of subsequent disease conversion could be achieved by means of nonlinear MRI-based SVMs operating *at the single-subject level*. In this regard, the most relevant clinical question is whether the utility of this early detection approach could be demonstrated in a second independent population at risk of developing psychosis.

Materials and Methods

Study Design

This imaging study was embedded in the naturalistic prospective and multidomain *FePsy* study on the prediction of psychosis development in individuals with an ARMS, covering a service area of 200 000 habitants in and around Basel, Switzerland. A more detailed description of the overall study design can be found elsewhere.[14] All aspects of the study were reviewed and approved by the institutional ethics committee of the University of Basel and written informed consent was obtained from each participant before study inclusion.

Participants

Within the prospective *FePsy* study, ARMS individuals received a structural MRI scan at study inclusion. For screening purposes, we used the Basel Screening Instrument for Psychosis, BSIP,[15] a 46-item checklist based on variables which have been shown to be risk factors or early symptoms of psychosis[14,16] such as *Diagnostic and Statistical Manual of Mental Disorders, Third Edition, Revised*—"prodromal symptoms," social decline, drug abuse, previous psychiatric disorders, or genetic liability for psychosis. The BSIP checklist facilitates a reliable identification of vulnerable individuals at risk of developing psychosis using clinical criteria that closely correspond to the well-established ultra-high-risk definitions of the Personal Assessment and Crisis Evaluation (PACE) clinic in Melbourne.[15–17] In keeping with previous MRI studies of ARMS cohorts recruited using these high-risk criteria (see[18] for review), inclusion into the present study required one or more of the following: (1) attenuated psychotic-like symptoms, (2) brief limited intermittent psychotic symptoms (BLIPS), or (3) a first- or second-degree relative with a psychotic disorder plus at least 2 further risk factors for or indicators of beginning psychosis according to the BSIP screening instrument. Inclusion because of attenuated psychotic symptoms required that change in mental state had to be present at least several times a week and for more than 1 week duration (a score of 2 or 3 on the Brief Psychiatric Rating Scale (BPRS) hallucination item or 3 or 4 on BPRS items for unusual thought content or suspiciousness). Inclusion because of BLIPS required scores of 4 or above on the hallucination item or 5 or above on the unusual thought content, suspiciousness, or conceptual disorganization items of the BPRS, with each symptom lasting less than 1 week before resolving spontaneously. A more detailed description of these ARMS criteria can be found in our previous work.[14] Additionally, (pre)psychotic and negative symptoms were assessed with the BPRS and the Scale for the Assessment of Negative Symptoms (SANS), which were used in combination with the BSIP.

Exclusion criteria were age below 18 years, insufficient knowledge of German, IQ <70, previous psychotic episodes treated with major tranquillizers for more than 3 weeks, a clearly diagnosed brain disease or substance dependency (except for cannabis dependency), or psychotic symptoms within a clearly diagnosed depression or borderline personality disorder. Thirty-three of 37 ARMS individuals never received antipsychotic medication prior to MRI scanning. Four participants had been administered low doses of atypical antipsychotic medication for behavioral control by the referring psychiatrist or general practitioner (3 participants olanzapine and 1 risperidone) at some time prior to study inclusion, all for less than 3 weeks.

Twenty-two healthy controls (HCs) were recruited from the same geographical area as the ARMS group through local advertisements and were matched to the ARMS sample groupwise for age, gender, handedness, and education level (table 1). These individuals had no current psychiatric disorder, no history of psychiatric illness, head trauma, neurological illness, serious medical or surgical illness, and substance dependency (except for cannabis and nicotine), and no family history of any psychiatric disorder as assessed by an experienced psychiatrist in a detailed clinical interview.

Study inclusion started in March 1, 2000 and continued until February 29, 2004. During the first year of follow-up, ARMS individuals were assessed monthly. During the second and third year, all individuals were assessed every 3 months and thereafter once a year until

Table 1. Sociodemographic, Clinical, and Global Anatomical Characteristics of the 3 Study Groups

	Study Groups			
	ARMS-T	ARMS-NT	HC	*P*
Sociodemographic variables				
N	16	21	22	
Mean age at baseline, y (SD)	26.4 (6.5)	23.4 (6.0)	23.0 (4.3)	ns
Sex (male), *n* (%)	11 (69)	11 (52)	13 (59)	ns
Handedness (mixed or left), *n* (%)	3 (19)	1 (5)	6 (29)	ns
Educational level				ns
<9 y, *n* (%)	4 (25)	7 (35)	2 (9)	
9–11 y, *n* (%)	6 (38)	8 (39)	7 (32)	
12–13 y, *n* (%)	5 (31)	3 (13)	10 (46)	
>13 y, *n* (%)	1 (6)	3 (13)	3 (14)	
Mean verbal IQ (Mehrfach-Wortschatztest-B) (SD)	109.6 (12.6)	107.6 (15.4)	—	ns
Clinical variables				
Individuals with a first degree relative with schizophrenia	3 (19%)	3 (14%)	na	ns
Mean BPRS global score at intake (SD)	41.9 (10.6)	37.2 (7.1)	na	ns
Mean SANS at intake (SD)	9.5 (5.4)	6.8 (4.4)	na	ns
Mean duration of symptoms, mo (SD)	42.6 (39.5)	43.2 (53.7)	na	ns
Mean interval between baseline MRI scan and disease transition, d (SD)	306.3 (318.3)	na	na	
Global anatomical volumes				
Mean global gray matter volume, ml (SD)	680.5 (57.5)	680.3 (67.4)	692.2 (52.6)	ns
Mean global white matter volume, ml (SD)	613.0 (79.9)	601.3 (72.3)	615.2 (68.7)	ns
Mean global cerebrospinal fluid volume, ml (SD)	212.6 (36.8)	212.0 (26.2)	204.8 (30.9)	ns

Note: ARMS-T, at-risk mental state for psychosis-converters; ARMS-NT, at-risk mental state for psychosis-nonconverters; HC, healthy control; BPRS, Brief Psychiatric Rating Scale; SANS, Scale for the Assessment of Negative Symptoms; MRI, magnetic resonance imaging.

conversion to frank psychosis or until the end of the follow-up period in 2007. All subjects were followed-up regularly and were offered supportive counseling and clinical management. Conversion to frank psychosis was monitored using the criteria described by Yung et al[17]: BPRS scores of 4 or above on the hallucination item or scores of 5 or above on the unusual thought content, suspiciousness, or conceptual disorganization items. Symptoms had to occur daily and persist for more than 1 week to be deemed a conversion to frank psychosis. Using these definitions, the ARMS group was subdivided into 21 nonconverters (ARMS-NT) and 16 converters (ARMS-T) to psychosis.

MRI Data Acquisition

Subjects were scanned using a SIEMENS (Erlangen, Germany) MAGNETOM VISION 1.5T scanner at the University Hospital Basel. Head movement was minimized by foam padding and velcrostraps across the forehead and chin. A 3-dimensional volumetric spoiled gradient recalled echo sequence generated 176 contiguous, 1 mm thick sagittal slices. Imaging parameters were time-to-echo, 4 msec; time-to-repetition, 9.7 msec; flip angle, 12; matrix size, 200 × 256; field of view, 25.6 × 25.6 cm matrix; and voxel dimensions, 1.28 × 1 × 1 mm.

MRI Data Preprocessing

After inspection for artifacts and gross abnormalities the images were segmented into GM, white matter (WM), and cerebrospinal fluid (CSF) maps in native space using the VBM5 toolbox (http://dbm.neuro.uni-jena.de), an extension of the SPM5 software package (Wellcome Department of Cognitive Neurology, London, UK). Details of this segmentation protocol have been described in our previous work.[13] Then, the estimated tissue maps of each individual were combined into a single-labeled volume (CSF: 10, GM: 150, and WM: 250) and registered to the single-subject brain template of Montreal Neurological Institute using a well-established high-dimensional elastic warping algorithm.[19] The volumetric changes occurring during this normalization process were written out to the registered tissue maps allowing for a Regional Analysis of Volumes in Normalized Space (RAVENS). Similar to the "modulation" step used in

voxel-based morphometry, RAVENS maps allow for local comparisons in standard space that are equivalent to volumetric comparisons of the original tissue maps in native space. The individual GM-RAVENS maps were proportionally scaled to the global GM volume computed from the native tissue maps and entered the susequent multivariate pattern classification analysis.

Multivariate Pattern Classification Analysis

SVM are multivariate statistical methods that have been increasingly employed for diagnostic purposes in a wide range of biomedical applications because they provide optimal decision rules for classifying individuals rather than describing statistical between-group differences. In our case, neuroanatomical features were used by the SVM to determine the best nonlinear classification model that reliably predicted the study participants' group membership. As customary in predictive analytics, the SVM models were constructed from one set of subjects (the training sample) and applied to a different set of subjects (the test sample) using cross-validation (CV). This process produced an unbiased estimate of the method's expected diagnostic accuracy on new individuals rather than merely fitting the current study population. The principles of generating and validating predictive models on separate training and testing samples have been previously described.[13] Based on the LIBSVM software (http://www.csie.ntu.edu.tw/cjlin/libsvm/), our machine-learning pipeline produced compact ensembles of SVMs that optimally separated single individuals from different groups, while avoiding the danger of overfitting to the peculiarities of the training data. It consisted mainly of 3 successive steps that were wrapped into a repeated nested CV framework (see online supplementary material for Methods)

Neuroanatomical Feature Generation. First, each training sample's GM-RAVENS maps were adjusted for age and gender effects using partial correlations and scaled voxel-wise to the range (0,1). These scaled and adjusted maps entered a recently proposed multivariate filter method,[20] which automatically determined those sets of voxels that conjointly maximize the geometric distance between the training subjects in the HC vs ARMS-T, HC vs ARMS-NT, and ARMS-T vs ARMS-NT analyses. This algorithm removed irrelevant/unreliable voxels from the high-dimensional MRI input space that did not contribute to the respective binary classification problem. Then, correlated voxels within the extracted discriminative patterns were projected to a number of uncorrelated principal components (PC) using principal component analysis (PCA).[13] This further reduced the dimensionality of the discriminative patterns to compact sets of neuroanatomical features. The optimum number of PC was determined using CV (see online supplementary material for Methods).

SVM Training. These discriminative PC features were projected to a high-dimensional feature space using the radial basis functions in order to account for possible nonlinear relations between the training subjects' neuroanatomical features and their group membership. In this feature space, the SVM found the optimal between-group boundary by maximizing the geometric distance between the neuroanatomically most similar subjects of opposite groups (the "support vectors").[21] It has been shown that this maximum margin principle in conjunction with the nonlinear projection generates classification rules that are adaptive to subtle between-group differences and therefore generalize well to unseen individuals.[21]

Classification of Unseen Test Data. The group membership of unseen test subjects was predicted after applying all training parameters successively to their MRI data, including (1) the adjustment for age and gender effects, (2) the selection of optimally discriminative voxels, (3) the projection of these voxels to PC, and (4) the nonlinear transformation of these neuroanatomical features. Then, for each subject, the 3 trained binary SVM models (HC vs ARMS-T, HC vs ARMS-NT, and ARMS-T vs ARMS-NT) determined its geometric position relative to their learned decision boundaries, resulting in 3 decision values and group membership predictions. We used these decision values to construct a multigroup classifier (HC vs ARMS-T vs ARMS-NT), where the binary SVM model with the maximum decision value decided about the test subject's group membership (one-vs-one-max-wins method).

Feature generation, model training, and test subject prediction were wrapped into a repeated nested CV framework (see online supplementary material for Methods).[22] The main goal of this framework was to completely separate the process of estimating the SVMs' prediction performance in a large number of unseen validation samples (outer CV loop) from the process of constructing optimally discriminative SVM models from a large number of training samples (inner CV loop). On the outer CV loop, we performed 10 repetitions of the following CV cycle. First, the order of the subjects was permuted within each group, and the entire population was split into 10 nonoverlapping samples. Each of these samples was iteratively held back as validation data, while the 9 remaining samples entered the inner CV loop as the training data. At this inner loop, we used 10-fold CV with 10 repetitions to generate ensembles of SVM models. More specifically, for each validation sample at the outer CV level, 100 different training data partitions were created at the inner CV level. In each of these 100 training partitions, the most discriminative sets of neuroanatomical features were determined. Each of these sets was used to train a separate SVM model. Then, each of these models predicted the group membership of the unseen validation subjects on the

Table 2. Two-Group Classification Performance

Binary Classifiers	TP	TN	FP	FN	Sens (%)	Spec (%)	BAC (%)	FPR (%)	PPV (%)	NPV (%)	LR+	LR−
HC vs ARMS-T	20	15	1	2	93.8	90.9	92.3	6.3	95.2	88.2	10.3	0.1
HC vs ARMS-NT	30	9	12	2	42.9	90.9	66.9	57.1	62.5	81.8	4.7	0.6
ARMS-T vs ARMS-NT	14	17	4	2	81.0	87.5	84.2	19.1	77.8	89.5	6.5	0.2

Note: Sens, Sensitivity; Spec, specificity; BAC, balanced accuracy; FPR, false positive rate; PPV/NPV, positive/negative predictive values; LR+/LR−, positive/negative Likelihood Ratios; true positives (TP), false negatives (FN), true negatives (TN), and false positives (FP); SVM, support vector machine; CV, cross-validation.
Note: Sens, Spec, BAC, FPR, PPV/NPV, and LR+/LR− were calculated from the confusion matrix containing the number of TP, FN, TN, and FP.
Note: The performance of the binary SVM ensemble classifiers (group "+1" vs group "−1") was evaluated (1) by constructing a binary SVM ensemble from all SVM base learners of an inner CV partition, in which the respective outer CV test subjects had not been included, (2) by computing the average decision value in each of these binary inner CV ensembles in order to determine the group membership (average decision value > 0 or < 0) of the respective outer CV test subjects, and (3) through majority voting across those binary inner CV loop SVM ensembles, in which the outer CV test subjects had not participated in the training process (see also the Methods section for a detailed explanation of the employed ensemble learning framework).

outer loop. These predictions were averaged across all 100 training partitions to yield an ensemble decision. Finally, for each validation subject, all SVM ensemble decisions were aggregated across those outer training partitions, in which this subject had not been involved in the training process. Majority voting was used to determine the validation subject's class probability, and thus its final out-of-training group membership (tables 2 and 3).

This ensemble learning approach leads to robust classification results because it greatly reduces the risk of unfortunate selections of poorly performing single classifiers by averaging the diagnostic decisions of numerous predictive models. Furthermore, ensembles of predictive models have shown to improve classification performance particularly in small sample situations because they detect complex decision boundaries by means of training sample and training parameter variation.[22] The performance of these classifier ensembles on the unseen validation data was measured in terms of sensitivity, specificity, balanced accuracy (BAC), false positive rate, positive/negative predictive values, and positive/negative likelihood ratios.

The nonlinearity of the decision rules determining the test subjects' group membership made it difficult to directly visualize each voxel's contribution to the average SVM ensemble decision. Therefore, we first approximated the average neuroanatomical decision boundary used by the binary nonlinear SVM models as described in Koutsouleris et al[13] and then measured each voxel's probability of reliably contributing to this discriminative pattern across the entire experiment at the 95% CI. The exact visualization procedure has been detailed in the legend of figure 1. Moreover, a supplementary parcellation analysis (see online supplementary material for figure 1) was conducted in order to measure the distribution of reliably discriminative voxels across the 116 brain regions of the AAL template (Automated Anatomical Labeling[23]). Finally, the similarities and differences between the approximated neuroanatomical decision boundaries underlying the 3 binary SVM classifiers were qualitatively assessed in online supplementary material, figure 3.

Results

Sociodemographic, Clinical, and Global Anatomical Findings

The rate of conversion to psychosis was 43.2% in our ARMS sample of 37 individuals with MRI scan at baseline. The mean interval between baseline scan and disease conversion scan was 306 days (median: 263, range: 25–1137 days). HCs, subsequent converters, and nonconverters did not significantly differ with respect to age, gender, educational level, and global brain volumes (table 1). Furthermore, no significant baseline differences were found between the ARMS-NT and ARMS-T samples regarding verbal IQ, family history of psychosis, duration of symptoms prior to the MRI examination, BPRS, and SANS (table 1). A trend toward a more severe baseline psychopathology (BPRS) was detected in the conversion compared with the nonconversion sample.

SVM Classification Analysis

Classification Performance. Among the 3 binary classification analyses (table 2), the highest diagnostic performance (BAC = 92.3%) was observed in the HC vs ARMS-T comparison, where 1 ARMS-T individual was classified as HC and 2 HC subjects were assigned to the ARMS-T group (sensitivity = 93.8% and specificity = 90.9%). The lowest SVM performance was detected in the HC vs ARMS-NT analysis (BAC = 66.9%) as 12 ARMS-NT were wrongly assigned to the HC group, and 2 HC were classified as ARMS-NT (sensitivity = 42.9%

Table 3. Three-Group Classification Performance

	SVM-Predicted Classes		
	HC	ARMS-T	ARMS-NT
Clinical groups			
HC	22	0	0
ARMS-T	6	10	0
ARMS-NT	11	1	9
OOT-Performance			
TP	22	10	9
TN	20	42	38
FP	17	1	0
FN	0	6	12
Sensitivity (%)	100	62.5	42.9
Specificity (%)	54.1	97.7	100
Balanced accuracy (%)	77.1	80.1	71.4
False positive rate (%)	46.0	2.3	0
Positive predictive value (%)	56.4	90.9	100
Negative predictive value (%)	100	87.5	76.0

Note: SVM, support vector machine; ARMS-T, at-risk mental state for psychosis-converters; ARMS-NT, at-risk mental state for psychosis-nonconverters; HC, healthy control; TP, true positive; TN, true negative; FP, false positive; FN, false negative; OOT, out-of-training; CV, cross-validation.
Note: Multigroup decisions were obtained by (1) constructing a multigroup ensemble classifier for each CV2 data partition using error-correcting output codes (see "Materials and Methods" section and online supplementary material) and by (2) computing the final OOT group membership of a given CV2 test subject through majority voting of all CV2 multigroup ensemble classifiers, in which this test subject had not been part of the training data and thus had not been seen by these classifier ensembles. The OOT classification performance of the multigroup ensembles was then evaluated for one group against all other groups. For example, in the HC vs ARMS-T vs ARMS-NT analysis 22 HC subjects of 22 (sensitivity: 100%) were correctly assigned to their group, while of 20 of 37 (54.1%) ARMS subjects were correctly not labeled as HC, resulting in a balanced accuracy of (100% + 54.1%)/2 = 77.1%.

and specificity = 90.9%). In the critical ARMS-T vs ARMS-NT analysis, the BAC was BAC = 84.2%, with 4 ARMS-NT individuals being misclassified as ARMS-T and 2 ARMS-T being wrongly labeled as ARMS-NT (sensitivity = 81.0% and specificity = 87.5%). Thus, the likelihood ratio of a positive test result was LR+ = 0.81/(10.875) = 6.5, meaning that a positive prognostic test in a given ARMS subject would increase the probability of subsequent disease conversion from 43% (pretest probability: 16/37 = 0.43) to 83% (posttest probability: pretest odds × LR+ = 0.762 × 6.5 = 4.94 → 4.94/(4.94 + 1) = 0.83).

In the 3-group classification (table 3), all 22 HC individuals were correctly assigned to their group, while 6 of the 16 ARMS-T and 11 of the 21 ARMS-NT subjects were misclassified as HC (sensitivity = 100%, specificity = 54.1% and BAC = 77.1%). Of the 16 ARMS-T subjects, 10 were correctly assigned to their group, while 1 ARMS-NT individual was wrongly labeled as ARMS-T (sensitivity = 62.5%, specificity = 97.7%, and BAC = 80.1%). Nine of 21 ARMS-NT individuals were correctly identified by the pattern recognition system, and no HC or ARMS-T subject was misclassified as ARMS-NT (sensitivity = 42.9%, specificity = 100%, and BAC = 71.4%). The misclassified ARMS-T and ARMS-NT individuals did not significantly differ from the correctly labeled ARMS-T and ARMS-NT subjects with respect to the sociodemographic, clinical, and global anatomical variables (table 4).

Neuroanatomical Mapping of SVM Decision Functions. In summary, the approximation of the 3 neuroanatomical SVM decision functions (methodological descriptions in figure 1) revealed that reliable voxels were not confined to single brain regions but instead were distributed across a broad range of cortical and subcortical areas. Within these distributed patterns shown in figures 1–3, foci of high-probability voxels (>80% probability) were detected particularly in the prefrontal, parietal, temporal, thalamic, and cerebellar structures.

More specifically, the average neuroanatomical decision function of the HC vs ARMS-T ensemble classifier involved high-probability hotspots particularly in the right hemisphere (1) within the prefrontal cortex, including the right and left ventrolateral and rostral prefrontal, the right lateral orbitofrontal subregions, as well as the right Rolandic operculum; (2) the right anterior insula; (3) the medial and lateral parietal cortex; as well as (4) the basal ganglia, thalamus, and cerebellum.

Clusters of contiguous high-probability voxels involved in the average HC vs ARMS-NT ensemble decision were detected predominantly (1) in the midline structures, bilaterally covering the anterior, middle, and posterior parts of the cingulate cortex with extensions to the ventromedial and dorsomedial prefrontal cortices, the premotor and supplementary motor areas, as well as the medial parietal cortices and (2) the inferior temporal and fusiform cortices, bilaterally.

Reliable high-probability voxels contributing to the average ARMS-T vs ARMS-NT ensemble decision mainly mapped to (1) the dorsomedial, rostromedial, and cingulate cortex, bilaterally, with extensions to the medial orbitofrontal, precuneal, and premotor areas; (2) the dorsolateral prefrontal GM and WM; (3) the right parahippocampal and inferior temporal cortex; as well as (4) the thalamus, bilaterally.

Application of the Classification Method to the Munich High-Risk Cohort. A supplementary analysis (see online supplementary material for table 1 and figure 2) was carried out in order (1) to compare the performance of our pattern recognition strategy between the *FePsy* and the Munich high-risk populations and (2) to validate

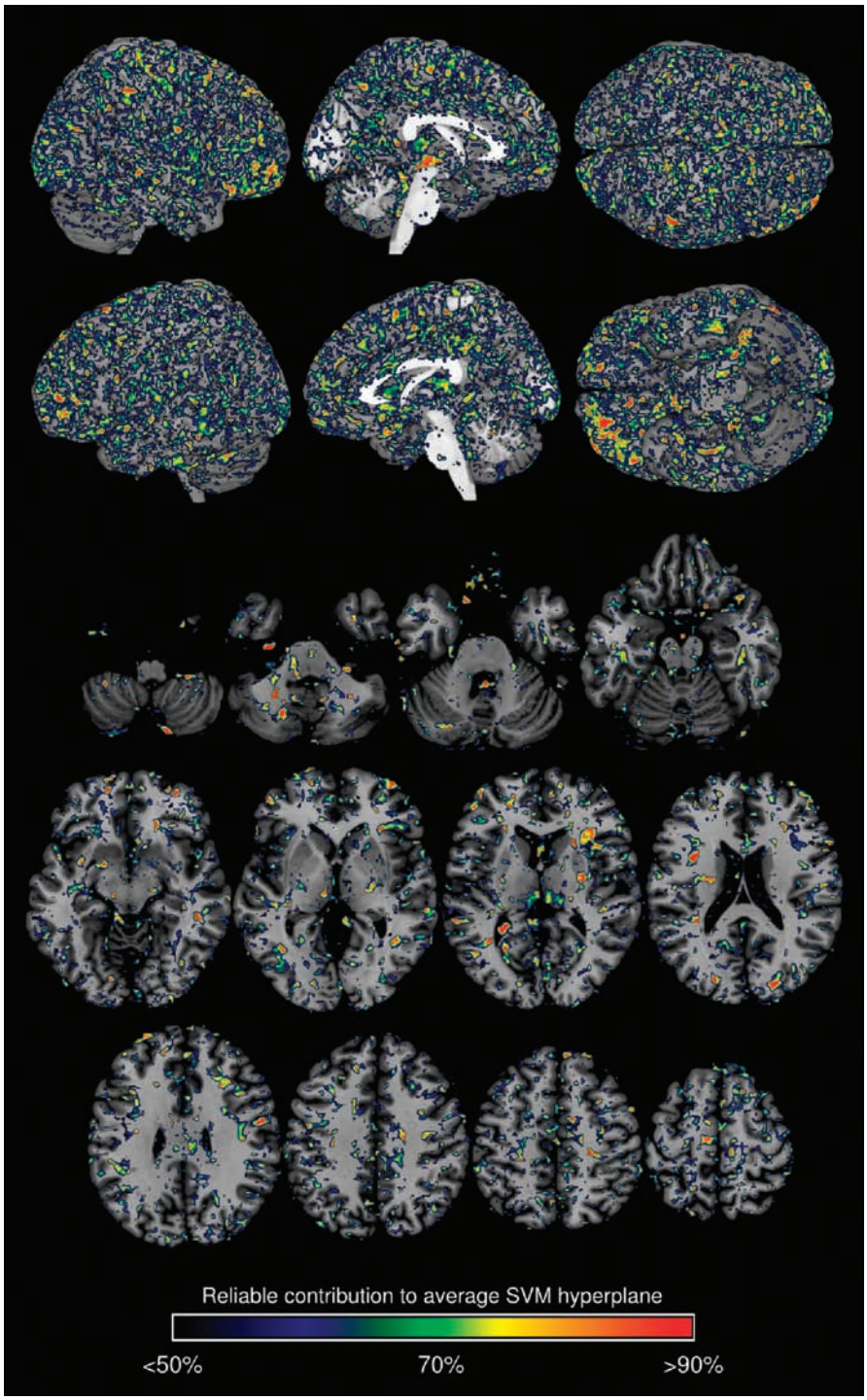

Fig. 1. Voxel probability map of reliable contributions to the healthy control vs at-risk mental state for psychosis-converters (ARMS-T) decision boundary. The approximation of each voxel's contribution to the average nonlinear classification used to separate HC from ARMS-T subjects was obtained as follows: In principal component analysis space, the average minimum difference vector ($SV_{mindiff}$) across the support vectors of a given support vector machine model was computed and projected back to voxel space as described previously.[13] This computation was performed for every training sample on the inner cross-validation (CV) loop resulting in 100 $SV_{mindiff}$ images for a given training partition on the outer CV loop. The average and SE volumes of these 100 $SV_{mindiff}$ images were computed. For every outer CV partition, the average $SV_{mindiff}$ image was binarized in that voxels with an absolute value greater than their respective standard error were set to one or to zero otherwise. This thresholding procedure extracted only those voxels that reliably contributed to the average neuroanatomical decision boundary of a given outer CV partition at the 95% CI. The obtained binary images were summed across all 100 outer CV partitions and divided by 100, thus forming a single map that specified every voxel's probability of reliably contributing to the average neuroanatomical decision boundary across the entire experiment. Voxels with a probability of >50% were overlaid on the single-subject Montreal Neurological Institute template using the MRIcron software package (http://www.sph.sc.edu/comd/rorden/mricron/).

Table 4. Misclassification Analysis

	ARMS-T → HC	ARMS-T → ARMS-T	*P*	ARMS-NT → HC	ARMS-NT → RMS-NT	*P*
Sociodemographic variables						
N	6	10		11	9	
Mean age at baseline, y (SD)	26.5 (5.6)	26.4 (7.3)	ns	23.0 (6.2)	24.4 (6.3)	ns
Sex (male), *n* (%)	6 (100)	5 (50)	ns	4 (44.4)	5 (55.6)	ns
Handedness (mixed or left), *n* (%)	2 (33.3)	1 (10)	ns	0 (0)	0 (0)	na
Educational level			ns			ns
<9 y, *n* (%)	2 (33.3)	2 (20)		5 (45.5)	1 (11.1)	
9–11 y, *n* (%)	3 (50.0)	3 (30)		4 (36.4)	4 (44.4)	
12–13 y, *n* (%)	1 (16.7)	4 (40)		na	3 (33.3)	
>13 y, *n* (%)	na	1 (10)		2 (18.2)	1 (11.1)	
Clinical variables						
Mean BPRS global score at intake (SD)	39.3 (11.2)	43.5 (10.6)	ns	36.6 (7.1)	36.1 (5.5)	ns
Mean SANS at intake (SD)	8.8 (6.2)	9.9 (5.3)	ns	5.6 (4.4)	7.8 (4.5)	ns
Mean duration of symptoms, mo (SD)	38.2 (26.8)	45.6 (47.5)	ns	55.1 (66.9)	31.2 (36.5)	ns
Mean interval between baseline MRI scan and disease transition, d (SD)	427.5 (483.6)	245.75 (215.4)	ns			
Global anatomical volumes						
Mean global gray matter volume, ml (SD)	693.0 (34.1)	672.9 (68.5)	ns	677.1 (79.5)	679.0 (56.3)	ns
Mean global white matter volume, ml (SD)	613.8 (54.3)	612.5 (94.8)	ns	599.6 (79.5)	609.4 (69.0)	ns
Mean global cerebrospinal fluid volume, ml (SD)	209.8 (31.8)	213.4 (24.0)	ns	203.3 (30.9)	227.1 (41.7)	ns

Note: Abbreviations are explained in the first footnote to table 1.
Note: Sociodemographical, clinical, and global anatomical characteristics of wrongly vs correctly classified converters and wrongly vs correctly classified nonconverters were compared using nonparametric Mann-Whitney *U*-tests.

our current approach with respect to our previous methods.[13] Therefore, we applied the identical parameter setup as employed in the analysis of the *FePsy* data in order to classify the Munich high-risk cohort, which consisted of 17 converters and 17 nonconverters to psychosis.

Our current machine learning strategy produced higher sensitivity (82.4%), specificity (94.1%), and BAC (88.2%) values compared with our previous findings[13] (see online supplementary material for table 1). These higher performance measures are mainly due to several methodological improvements that are detailed in the online supplementary Methods, including the use of a novel local learning-based feature selection strategy and the application of ensemble learning principles.

Discussion

The present investigation largely replicated our previous findings[13] in that our fully automated classification system reliably identified those individuals among a clinically defined at-risk population who subsequently developed psychosis by using only their MRI scans acquired at study inclusion. This observation agrees with recent studies outlining the good performance of MRI-based pattern recognition techniques in (1) correctly classifying patient populations with established neuropsychiatric illnesses, such as Alzheimer's Disease[24] or schizophrenia[25] and (2) predicting clinical outcome in different neuropsychiatric conditions, including dyslexia,[26] major depression,[27] and mild cognitive impairment.[28] This growing transnosological literature provides evidence that pattern recognition methods may indeed have the potential to delineate neuroanatomical intermediate phenotypes that constitute disease signatures beyond the level of coarse between-group differences.[29]

Neuroanatomical Basis of Prediction

Therefore, the main property of the SVM, which is its ability to detect subtle and distributed, but highly discriminative patterns of neuroanatomical differences, makes this method relevant for an early recognition of psychosis. Several previous imaging studies of clinically defined ARMS cohorts have demonstrated focal GM volume reductions across a wide range of brain regions with a particularly reliable involvement of the lateral prefrontal, anterior cingulate, temporoparietal, limbic, and paralimbic cortices (see ref. [18] for review). Furthermore, the recent voxel-based meta-analysis of Fusar-Poli et al[30] showed that conversion to overt psychosis may be associated with further structural alterations located primarily in the right ventrolateral prefrontal, insular,

and superior temporal cortices. Obtained from 2 completely independent ARMS populations, our present and previous[13] SVM-based results were partly consistent with these meta-analytic observations, in so far as volumetric alterations involving the prefrontal, temporal, limbic, and thalamic structures reliably contributed to the neuroanatomical separation of nonconverters from converters to psychosis (figure 3). However, the discriminative patterns underlying all 3 classification analyses may suggest that the vulnerability and prodromal state for psychosis do not relate to a circumscribed set of few highly relevant structures. Instead, our results seem to involve complex patterns of structural brain alterations that conjointly produce predictive intermediate phenotypes of the ARMS and the emerging illness. This finding agrees with the literature of morphometric changes in schizophrenia,[31] which suggests that the disease pathology is not confined to single brain regions, but rather spans distributed cortical and subcortical neural networks, in keeping with the current disconnection hypothesis of schizophrenia.[32]

Early Recognition of Psychosis Using MRI-Based Methods

The sociodemographic and clinical characteristics of our ARMS population are in keeping with other at-risk cohorts recruited at different specialized early recognition services around the world, including eg, the PACE clinic in Melbourne,[33] the TOPP clinic in Norway,[34] the FETZ service in Munich,[13,35] or the multicentric North American Prodrome Longitudinal Study.[5] In the context of these samples, our study population, albeit modest in size due to the well-known difficulties of recruiting and prospectively following at-risk individuals over time, may be regarded as being representative of a clinically defined risk for psychosis.

Disease conversion rates in these clinically ''enriched'' at-risk samples may vary between 9% and 54%.[9,36] Therefore, the current symptom-based early recognition inventories perform well in recruiting samples with a significantly higher psychosis prevalence compared with the baseline population risk of 0.5–1%, but unfortunately, they do not provide the clinical means needed to reliably differentiate between true prodromal subjects and ''false alarms'' at the individual level. This diagnostic separation is required in order to administer preventive treatment to those at highest risk of developing psychosis, while minimizing harmful medication effects in individuals with a lower likelihood of disease conversion. In this regard, different approaches have been proposed in order to improve prognostic power within these clinical high-risk samples, including (1) multivariate clinical prediction models, which have shown to produce high levels of diagnostic performance (>80%),[5,8] (2) neurocognition-based machine learning methods correctly predicting psychosis in >85% of the cases,[22] and (3) diagnostic models combining neurocognitive and clinical data with a prognostic accuracy of 80%.[6,37]

Despite these encouraging results, several drawbacks of early recognition instruments based exclusively on clinical/behavioral signs and symptoms have to be considered: (1) their limited availability because only continuously trained personnel working at highly specialized mental health services will achieve the level of sensitivity and specificity needed to reliably detect the subtle and subclinical phenotypes of at-risk individuals, (2) the affected subjects' varying degree of motivation and insight as well as the interaction of personal and cultural backgrounds occurring during the clinical examination, which may bias the predictions of a diagnostic test. Therefore, clinical detection strategies could be further enhanced by means of objective imaging biomarkers capable of measuring the pathophysiological processes associated with emerging psychosis.[38]

In this regard, we detected high cross-validated diagnostic performances (tables 2 and 3) in the pairwise ARMS-T vs HC (BAC = 92.3%) and ARMS-T vs ARMS-NT (84.2%) classification analyses as well as in the multigroup ARMS-T vs rest comparison (80.1%). However, with respect to our previous results,[13] diagnostic performances were lower in the pairwise HC vs ARMS-NT (66.9%) and the multigroup ARMS-NT vs rest (71.4%) analyses, mainly due to 57% nonconverters in the former and 52% in the later comparison being misclassified as HC. Based on the long follow-up period of our study and the diffuse discriminative pattern observed in the ARMS-NT vs HC analysis (figure 2), this low classification performance suggests that the nonconversion sample may represent a heterogeneous help-seeking population that lacks an overarching neuroanatomical signature and hence cannot be reliably separated from the HC group. In the light of the findings obtained in our misclassification analysis (table 4), it remains to be elucidated how this neuroanatomical heterogeneity relates to a phenotypical heterogeneity within this group. Therefore, future prospective studies following larger nonconversion samples over time are needed in order to answer the question whether the coarse definition of ''nonconversion'' should be further disentangled according to the varying degree of mental and functional disturbances encountered in these subjects.[39] In this regard, the potentially heterogeneous neurobiology of these mental alterations may be better captured based on a clinical staging model of psychosis, as recently prosposed.[40]

In the light of these findings, the most promising early recognition strategy seems to consist of a 2-step diagnostic process. First, potential at-risk individuals are screened for patterns of clinical/behavioral items that meet the prodromal criteria of operationalized early recognition inventories. A positive test result at this stage would mark a significantly higher risk for psychosis

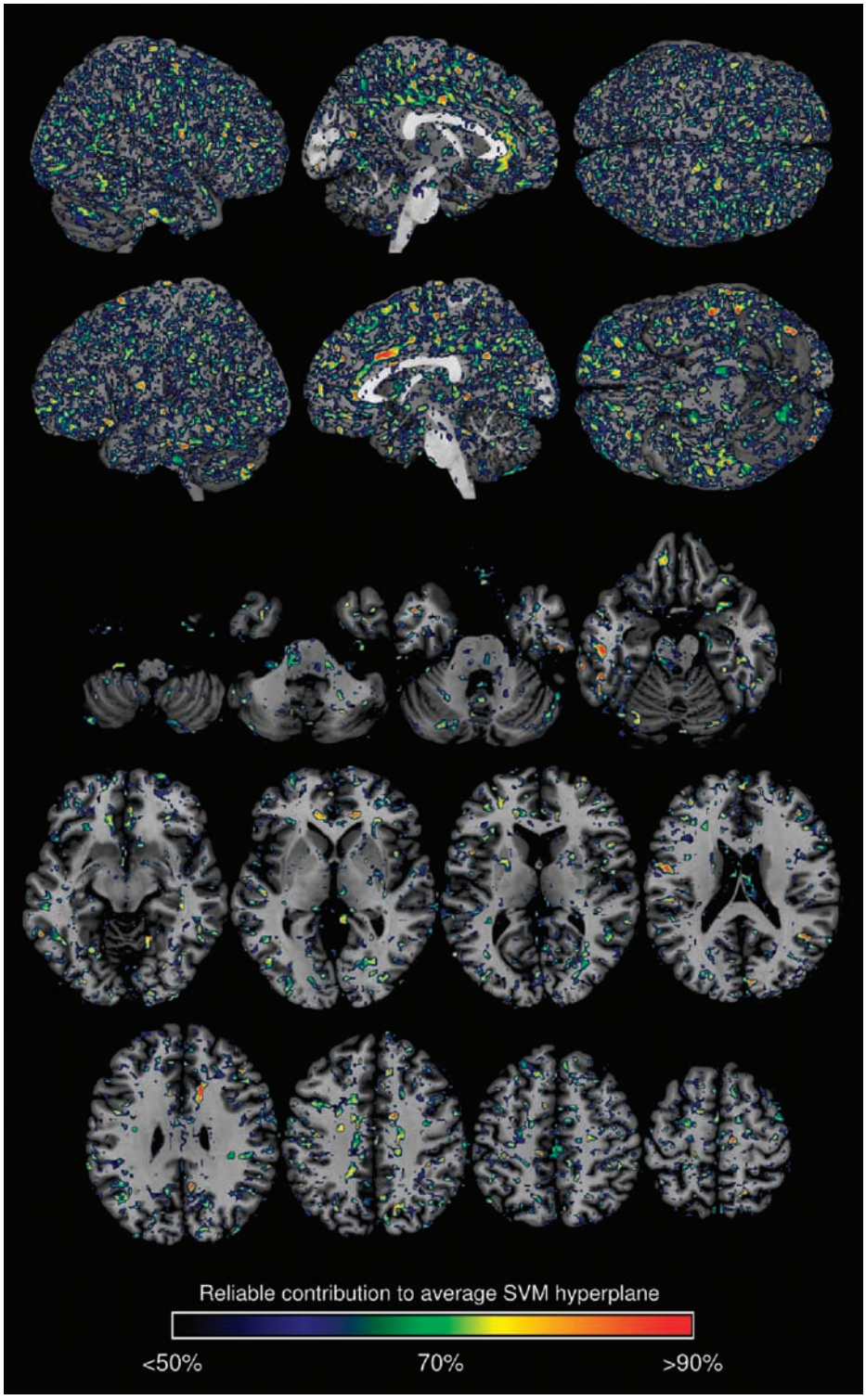

Fig. 2. Voxel probability map of reliable contributions to the healthy control vs at-risk mental state for psychosis-nonconverters (ARMS-NT) decision boundary. See legend of figure 1.

compared with the normal risk level—in our case, 43% compared with 0.5–1% in the general population. Then, the at-risk subject's probability of disease conversion would be further evaluated using a trained MRI-based pattern recognition system, which in our study achieved a clinically relevant positive likelihood ratio of >5, thus increasing prognostic certainty from 43% to 83% in case of a positive test result.

Despite the potential clinical utility of such a 2-level early recognition instrument, we have to consider several

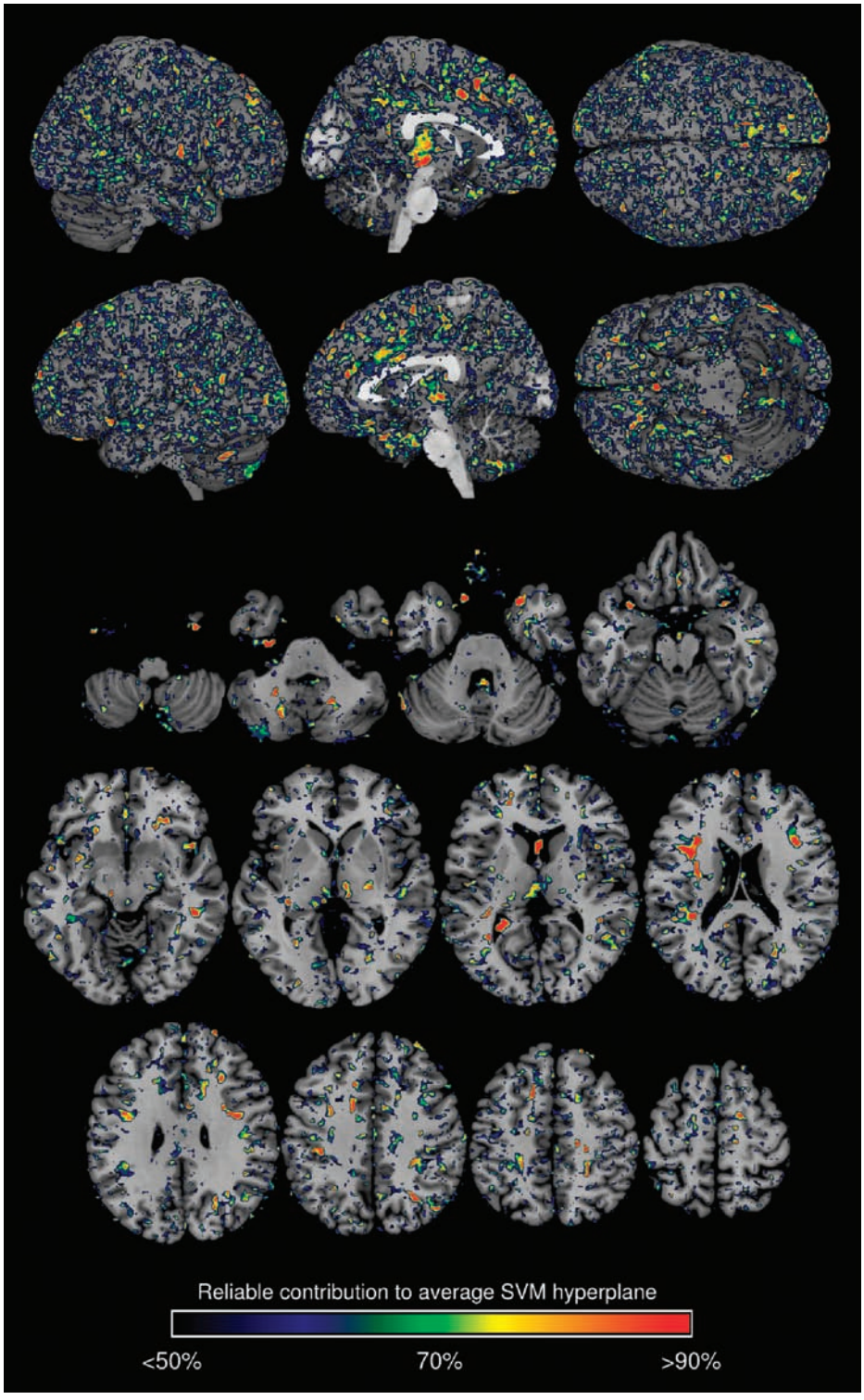

Fig. 3. Voxel probability map of reliable contributions to the at-risk mental state for psychosis-converters (ARMS-T) vs nonconverters (ARMS-NT) decision boundary. See legend of figure 1.

potential limitations of the proposed strategy. First, it remains unknown, how well our MRI-based early detection would work in clinically defined ARMS populations with a substantially lower conversion rate, as reported in recent studies from the PACE clinic.[36] Second, concomitant substance abuse, in particular cannabis, as well as the intake of antipsychotic medication increasingly encountered in at-risk individuals may likely impact on brain structure and interact with the pathophysiological process leading to the full onset of the disease.[41]

Hence, an important direction for future studies is to examine the ability of pattern recognition approaches to accurately predict disease conversion in such real-world high-risk populations. Third, as the *generalization capacity* of the proposed diagnostic method is still unclear, the next important research step is to evaluate its predictive performance in significantly larger ARMS cohorts recruited and examined across multiple centers and scanners and followed over a long period. Finally, the diagnostic specificity of the method needs to be assessed in subclinical populations at risk for developing different neuropsychiatric conditions, not only including schizophrenic psychosis but also bipolar disorder, major depression, or borderline personality disorder.

Supplementary Material

Supplementary material is available at http://schizophreniabulletin.oxfordjournals.org.

Funding

The *FePsy* project was supported by the Swiss National Science Foundation (3200-057216.99, 3200-0572216.99, PBBSB-106936, and 3232BO-119382); the Nora van Meeuwen-Haefliger Stiftung, Basel (CH), and by unconditional grants from the Novartis Foundation, Bristol-Myers Squibb, GmbH (CH), Eli Lilly SA (CH), AstraZeneca AG (CH), Janssen-Cilag AG (CH), and Sanofi-Synthelabo AG (CH).

Acknowledgments

We would like to thank patients, coworkers of the *FePsy* study, esp. Jacqueline Aston and Erich Studerus. Furthermore, we are particularly grateful for Dr Reinhold Bader's support in integrating the VBM5 and HAMMER/RAVENS and SVM algorithms into the batch system of the Linux Super-computing Cluster for the Munich and Bavarian Universities. Finally, we would like to thank Prof. Chih-Jen Lin from the National Taiwan University, Taiwan, for his help in adjusting the LIBSVM software to the needs of neuroimaging analysis. The funding sources had no involvement in the study design, the collection and analysis of the data, or the writing of the manuscript. The authors have declared that there are no conflicts of interest in relation to the subject of this study.

References

1. Perkins DO, Gu H, Boteva K, Lieberman JA. Relationship between duration of untreated psy-chosis and outcome in first-episode schizophrenia: a critical review and meta-analysis. *Am J Psychiatry*. 2005;162:1785–1804.
2. Amminger GP, Schäfer MR, Papageorgiou K, et al. Long-chain omega-3 fatty acids for indicated prevention of psychotic disorders: a randomized, placebo-controlled trial. *Arch Gen Psychiatry*. 2010;67:146–154.
3. Woods SW, Tully EM, Walsh BC, et al. Aripiprazole in the treatment of the psychosis prodrome: an open-label pilot study. *Br J Psychiatry Suppl*. 2007;51:s96–s101.
4. Phillips LJ, McGorry PD, Yuen HP, et al. Medium term follow-up of a randomized controlled trial of interventions for young people at ultra high risk of psychosis. *Schizophr Res*. 2007;96:25–33.
5. Cannon TD, Cadenhead K, Cornblatt B, et al. Prediction of psychosis in youth at high clinical risk: a multisite longitudinal study in North America. *Arch Gen Psychiatry*. 2008;65:28–37.
6. Riecher-Rössler A, Pflueger MO, Aston J, et al. Efficacy of using cognitive status in predicting psychosis: a 7-year follow-up. *Biol Psychiatry*. 2009;66:1023–1030.
7. Seidman LJ, Giuliano AJ, Meyer EC, et al. Neuropsychology of the prodrome to psychosis in the NAPLS consortium: relationship to family history and conversion to psychosis. *Arch Gen Psychiatry*. 2010;67:578–588.
8. Ruhrmann S, Schultze-Lutter F, Salokangas RKR, et al. Prediction of psychosis in adolescents and young adults at high risk: results from the prospective European prediction of psychosis study. *Arch Gen Psychiatry*. 2010;67:241–251.
9. Haroun N, Dunn L, Haroun A, Cadenhead KS. Risk and protection in prodromal schizophrenia: ethical implications for clinical practice and future research. *Schizophr Bull*. 2006;32:166–178.
10. Johns LC, Cannon M, Singleton N, et al. Prevalence and correlates of self-reported psychotic symptoms in the British population. *Br J Psychiatry*. 2004;185:298–305.
11. Job DE, Whalley HC, McIntosh AM, Owens DGC, Johnstone EC, Lawrie SM. Grey matter changes can improve the prediction of schizophrenia in subjects at high risk. *BMC Med*. 2006;4:29.
12. Bray S, Chang C, Hoeft F. Applications of multivariate pattern classification analyses in develop-mental neuroimaging of healthy and clinical populations. *Front Hum Neurosci*. 2009;3:32.
13. Koutsouleris N, Meisenzahl EM, Davatzikos C, et al. Use of neuroanatomical pattern classification to identify subjects in at-risk mental states of psychosis and predict disease transition. *Arch Gen Psychiatry*. 2009;66:700–712.
14. Riecher-Rössler A, Gschwandtner U, Aston J, et al. The Basel early-detection-of-psychosis (FEPSY)-study–design and preliminary results. *Acta Psychiatr Scand*. 2007;115:114–125.
15. Riecher-Rössler A, Aston J, Ventura J, et al. [The Basel Screening Instrument for Psychosis (BSIP): development, structure, reliability and validity]. *Fortschr Neurol Psychiatr*. 2008;76:207–216.
16. Riecher-Rössler A, Gschwandtner U, Borgwardt S, Aston J, Pflüger M, Rössler W. Early detection and treatment of schizophrenia: how early? *Acta Psychiatr Scand Suppl*. 2006;113(s429):73–80.
17. Yung AR, Phillips LJ, McGorry PD, et al. Prediction of psychosis. A step towards indicated prevention of schizophrenia. *Br J Psychiatry Suppl*. 1998;172:14–20.
18. Smieskova R, Fusar-Poli P, Allen P, et al. Neuroimaging predictors of transition to psychosis–a systematic review and meta-analysis. *Neurosci Biobehav Rev*. 2010;34:1207–1222.

19. Shen D, Davatzikos C. Very high-resolution morphometry using mass-preserving deformations and HAMMER elastic registration. *Neuroimage*. 2003;18:28–41.
20. Sun Y, Todorovic S, Goodison S. Local learning based feature selection for high dimensional data analysis. *IEEE Trans Pattern Anal Mach Intell*. 2010;32:1610–1626.
21. Vapnik V. *Statistical Learning Theory*. New York, NY: Wiley Interscience; 1998.
22. Koutsouleris N, Davatzikos C, Bottlender R, et al. Prediction in the at-risk mental states for psychosis using neurocognitive pattern classification. *Schizophr Bull*. 2011; doi:10.1093/schbul/sbr037.
23. Tzourio-Mazoyer N, Landeau B, Papathanassiou D, et al. Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain. *Neuroimage*. 2002;15:273–289.
24. Fan Y, Resnick SM, Wu X, Davatzikos C. Structural and functional biomarkers of prodromal Alzheimer's disease: a high-dimensional pattern classification study. *Neuroimage*. 2008;41:277–285.
25. Sun D, van Erp TGM, Thompson PM, et al. Elucidating a magnetic resonance imaging-based neuroanatomic biomarker for psychosis: classification analysis using probabilistic brain atlas and machine learning algorithms. *Biol Psychiatry*. 2009;66:1055–1060.
26. Hoeft F, McCandliss BD, Black JM, et al. Neural systems predicting long-term outcome in dyslexia. *Proc Natl Acad Sci U S A*. 2011;108:361–366.
27. Gong Q, Wu Q, Scarpazza C, et al. Prognostic prediction of therapeutic response in depression using high-field MR imaging. *Neuroimage*. 2011;55:1497–1503.
28. Davatzikos C, Bhatt P, Shaw LM, Batmanghelich KN, Trojanowski JQ. Prediction of MCI to AD conversion, via MRI, CSF biomarkers, and pattern classification. *Neurobiol Aging*. 2010;32:2322.e19–2322.e27.
29. Davatzikos C. Why voxel-based morphometric analysis should be used with great caution when characterizing group differences. *Neuroimage*. 2004;23:17–20.
30. Fusar-Poli P, Borgwardt S, Crescini A, et al. Neuroanatomy of vulnerability to psychosis: a voxel-based meta-analysis. *Neurosci Biobehav Rev*. 2011;35:1175–1185.
31. Honea R, Crow TJ, Passingham D, Mackay CE. Regional deficits in brain volume in schizophre-nia: a meta-analysis of voxel-based morphometry studies. *Am J Psychiatry*. 2005;162:2233–2245.
32. Friston KJ. Schizophrenia and the disconnection hypothesis. *Acta Psychiatr Scand Suppl*. 1999;395:68–79.
33. Yung AR, Phillips LJ, Yuen HP, et al. Psychosis prediction: 12-month follow up of a high-risk ("prodromal") group. *Schizophr Res*. 2003;60:21–32.
34. Larsen TK. The transition from premorbid period to psychosis: how can it be described? *Acta Psychiatr Scand*. 2002;106:10–11.
35. Koutsouleris N, Schmitt G, Gaser C, et al. Neuroanatomical correlates of different vulnerability states of psychosis in relation to clinical outcome. *Br J Psychiatry*. 2009;195:218–226.
36. Yung AR, Yuen HP, Berger G, et al. Declining transition rate in ultra high risk (prodromal) services: dilution or reduction of risk? *Schizophr Bull*. 2007;33:673–681.
37. Lencz T, Smith CW, McLaughlin D, et al. Generalized and specific neurocognitive deficits in prodromal schizophrenia. *Biol Psychiatry*. 2006;59:863–871.
38. Lawrie SM, Olabi B, Hall J, McIntosh AM. Do we have any solid evidence of clinical utility about the pathophysiology of schizophrenia? *World Psychiatry*. 2011;10:19–31.
39. Ruhrmann S, Schultze-Lutter F, Klosterkötter J. Probably at-risk, but certainly ill–advocating the introduction of a psychosis spectrum disorder in DSM-V. *Schizophr Res*. 2010;120:23–37.
40. McGorry PD. Risk syndromes, clinical staging and DSM V: new diagnostic infrastructure for early intervention in psychiatry. *Schizophr Res*. 2010;120:49–53.
41. Stone J, Bhattacharyya S, Barker G, McGuire P. Substance use and regional gray matter volume in individuals at high risk of psychosis. *Eur Neuropsychopharmacol*. 2011; doi:10.1016/j.euroneuro.2011.06.004.