

Eye Movement Patterns Can Distinguish Schizophrenia From the Major Affective Disorders and Healthy Control Subjects

David St Clair^{*1,2}, Graeme MacLennan^{3,◉}, Sara A. Beedie², Eva Nouzová², Helen Lemmon², Dan Rujescu⁴, Philip J. Benson⁵, Andrew McIntosh⁶, and Mintu Nath⁷

¹Division of Applied Medicine, Institute of Medical Sciences, University of Aberdeen, Aberdeen, UK; ²Clinical Research Centre, Royal Cornhill Hospital, Aberdeen, UK; ³Centre for Healthcare Randomised Trials (CHaRT), University of Aberdeen, Aberdeen, UK; ⁴Department of Psychiatry, Medical University of Vienna, Vienna, Austria; ⁵Department of Psychology, University of Aberdeen, Aberdeen, UK; ⁶Division of Psychiatry, Royal Edinburgh Hospital, Edinburgh, UK; ⁷Medical Statistics Team, Institute of Applied Health Sciences, University of Aberdeen, Aberdeen, UK

*To whom correspondence should be addressed; Clinical Research Centre, Royal Cornhill Hospital, Cornhill Road, Aberdeen, AB25 2ZH, UK; tel: 01224 557 951, fax: 01224-557400, e-mail: david.stclair@abdn.ac.uk

Background and hypothesis: No objective tests are currently available to help diagnosis of major psychiatric disorders. This study evaluates the potential of eye movement behavior patterns to predict schizophrenia subjects compared to those with major affective disorders and control groups. **Study design:** Eye movements were recorded from a training set of UK subjects with schizophrenia (SCZ; $n = 120$), bipolar affective disorder (BPAD; $n = 141$), major depressive disorder (MDD; $n = 136$), and healthy controls (CON; $n = 142$), and from a hold-out set of 133 individuals with proportional group sizes. A German cohort of SCZ ($n = 60$) and a Scottish cohort of CON subjects ($n = 184$) acted as a second semi-independent test set. All patients met DSMIV and ICD10 criteria for SCZ, BPAD, and MDD. Data from 98 eye movement features were extracted. We employed a gradient boosted (GB) decision tree multiclass classifier to develop a predictive model. We calculated the area under the curve (AUC) as the primary performance metric. **Study results:** Estimates of AUC in one-versus-all comparisons were: SCZ (0.85), BPAD (0.78), MDD (0.76), and CON (0.85). Estimates on part-external validation were SCZ (0.89) and CON (0.65). In all cases, there was good specificity but only moderate sensitivity. The best individual discriminators included free viewing, fixation duration, and smooth pursuit tasks. The findings appear robust to potential confounders such as age, sex, medication, or mental state at the time of testing. **Conclusions:** Eye movement patterns can discriminate schizophrenia from major mood disorders and control subjects with around 80% predictive accuracy.

Key words: eye movements/schizophrenia/affective disorders/biomarker/predictive modelling

Introduction

The current categorical diagnostic classification system in psychiatry relies upon face validity. There are no objective tests to support clinical diagnosis, monitor the progress of illness, or inform choice of treatment.^{1,2} Possible reasons are discussed, but no consensus has emerged.³⁻⁵ At the core of the current classification system lies the hundred-year old so-called Kraepelinian dichotomy that sets a boundary between schizophrenia and major mood disorders. Unfortunately, it remains controversial whether these disorders are qualitatively different from each other.^{6,7} Evidence from clinical and molecular genetics shows considerable genetic overlap suggesting a continuum,⁸⁻¹² whereas a recent long-term clinical outcome study comparing nonaffective with affective psychosis does not support the idea of a continuum.¹³

Biomarker discovery efforts in psychiatry have mostly focused on neuro-imaging and genomics. While some interesting results have emerged, most studies use binary case-control methodologies, replications are few and robust modelling approaches, and appropriate validations are rare.¹⁴⁻¹⁹

Atypical eye movements in unmedicated psychiatric patients were first reported in 1908.²⁰ Since then, a considerable literature has consistently reported eye movement abnormalities in schizophrenia patients compared to healthy controls, reviewed by Wolf et al.²¹ The literature for other major psychiatric disorders such as unipolar and bipolar disorders is much more modest with often inconsistent findings, small sample sizes, and few studies have compared eye movements in mood disorders with schizophrenia. The most common findings are

abnormalities of pursuit in all diagnostic groups with schizophrenia patients most impaired.²²⁻²⁶

Most recently, Clementz et al replicated earlier studies using a test battery that included neuropsychology, eye movements, EEG in a large cohort, and demonstrated excellent discrimination between psychosis and control populations, but none of the physiological biomarkers differentiated the psychosis subgroups.²⁷ In general, however, results have not shown sufficient sensitivity or specificity to be of clinical value.

We earlier reported that individuals with schizophrenia show significant eye movement differences from a mentally healthy comparison group when performing image viewing, smooth pursuit, and steady fixation tasks. Indeed using a gradient boosted (GB) decision tree machine learning algorithm, we could distinguish new schizophrenia cases from controls with around 80% predictive accuracy.²⁸ These results have now been independently replicated.²⁹ Using similar protocols, the authors derived an integrated score from the same eye movement tasks and found they could separate schizophrenia from controls with 82% accuracy.

In this study, we have expanded our schizophrenia and control cohorts and tested two new large groups of individuals with bipolar and recurrent unipolar affective disorder. We hypothesized that if eye movement viewing patterns could discriminate schizophrenia from major mood disorders as well as from mentally unaffected control subjects, they would have potential clinical utility as biomarkers to assist with the diagnosis of schizophrenia.

We have also examined in detail the effects of potential confounders including age and effects of psychotropic medication. Recent large studies by Coors et al³⁰ and Takahashi et al³¹ highlight the importance of the former as a potential confounder in eye movement studies.

Methods

Participants

Data from 672 subjects constituted the main dataset for calibrating and validating the classifier (see later). The data included healthy control (CON; $n = 177$) individuals along with patients diagnosed with schizophrenia (SCZ, $n = 150$) including 64 schizophrenia patients described in the earlier study,²⁸ bipolar affective disorder (BPAD; $n = 176$), and unipolar major depressive disorder (MDD; $n = 169$). We additionally used a German cohort of SCZ ($n = 60$),²⁸ and a Scottish cohort of CON subjects ($n = 184$) as a semi-independent second validation dataset (see later). All Scottish patients were identified through the psychiatric services of local NHS Trusts. Diagnoses were initially made on ICD-10 criteria by attending clinicians and then independently confirmed by the research team through examination of psychiatric case notes, OPCRIT checklist, and a structured clinical interview for DSM-IV.³²⁻³⁴

All met DSM-IV and ICD-10 criteria for schizophrenia, bipolar disorder, or unipolar depression. The latter was defined as having at least two major depressive episodes or chronic depression lasting over 18 months. All depression cases were ascertained through psychiatric care services. Mentally healthy nonclinical controls were recruited mainly through public advertisements and also included a volunteer panel at the University of Aberdeen.

The Scottish and German studies obtained full multi-regional ethics committee (MREC) and institutional review board (IRB) approvals, respectively, and were conducted in accordance with the Declaration of Helsinki.

Phenotype Measures

Following informed consent, a brief interview and collection of demographics, eye movements were recorded on all subjects. Static image free viewing, smooth pursuit, and steady fixation tasks were administered. The protocol has been previously described.^{22,28} Minor changes were introduced in the recording protocols during the 12-year period of the study. However, we found minimal evidence of incompatibility arising from these changes and proceeded to analyze the data as a single dataset. The eye movement data were quality controlled and then scored semi-automatically offline as previously described.^{2,28} This generated a total of 98 eye movement variables (aka features); a complete list of these variables is presented in the supplement ([Supplementary Table 2](#)). These eye movement variables along with sex were considered for the development of the multiclass classifier (discussed below).

Following eye movement recording, a diagnostic interview using SCID or MINI was administered to all cases and controls together with a brief illness-agnostic neuropsychological test battery²². Neuropsychology results will be reported elsewhere.

Statistical Analysis

We partitioned the main dataset ($n = 672$) into training (Train, $n = 539$; 80% of data) and testing or hold-out (Test-1, $n = 133$; 20% of data) datasets by randomly selecting individuals from all four groups ensuring that representation in both datasets was similar to the whole cohort. We used the test dataset (Test-1) for validation-1 and the semi-independent dataset (60 SCZ and 184 CON, Test-2) for validation-2. Missing values for any features in the training dataset were imputed using the bootstrap aggregation-based approach.

We employed a gradient boosted (GB) decision tree multiclass classifier to develop a predictive model of major psychiatric disorders using eye movement features in a machine-learning (ML) framework. To develop a robust, reproducible, and accurate classifier, we implemented a

modified version of the 5×5 fold nested cross-validation (CV) scheme outlined before.^{35,36,37} Figure 1 gives a schematic representation of the full pipeline. A detailed description of machine learning outflow is provided in the Supplementary Material.

The performance of the final classifier was evaluated on both validation datasets (Test-1, Test-2) using a range

of performance metrics. We calculated the area under the curve (AUC) as the primary performance metric using three methods: the generalized overall AUC, one-versus-all (OVA) or one-versus-rest for each group, and pairwise AUCs using one-versus-one (OVO).³⁶ Other performance metrics included: sensitivity, specificity, positive predictive value (PPV), negative predictive

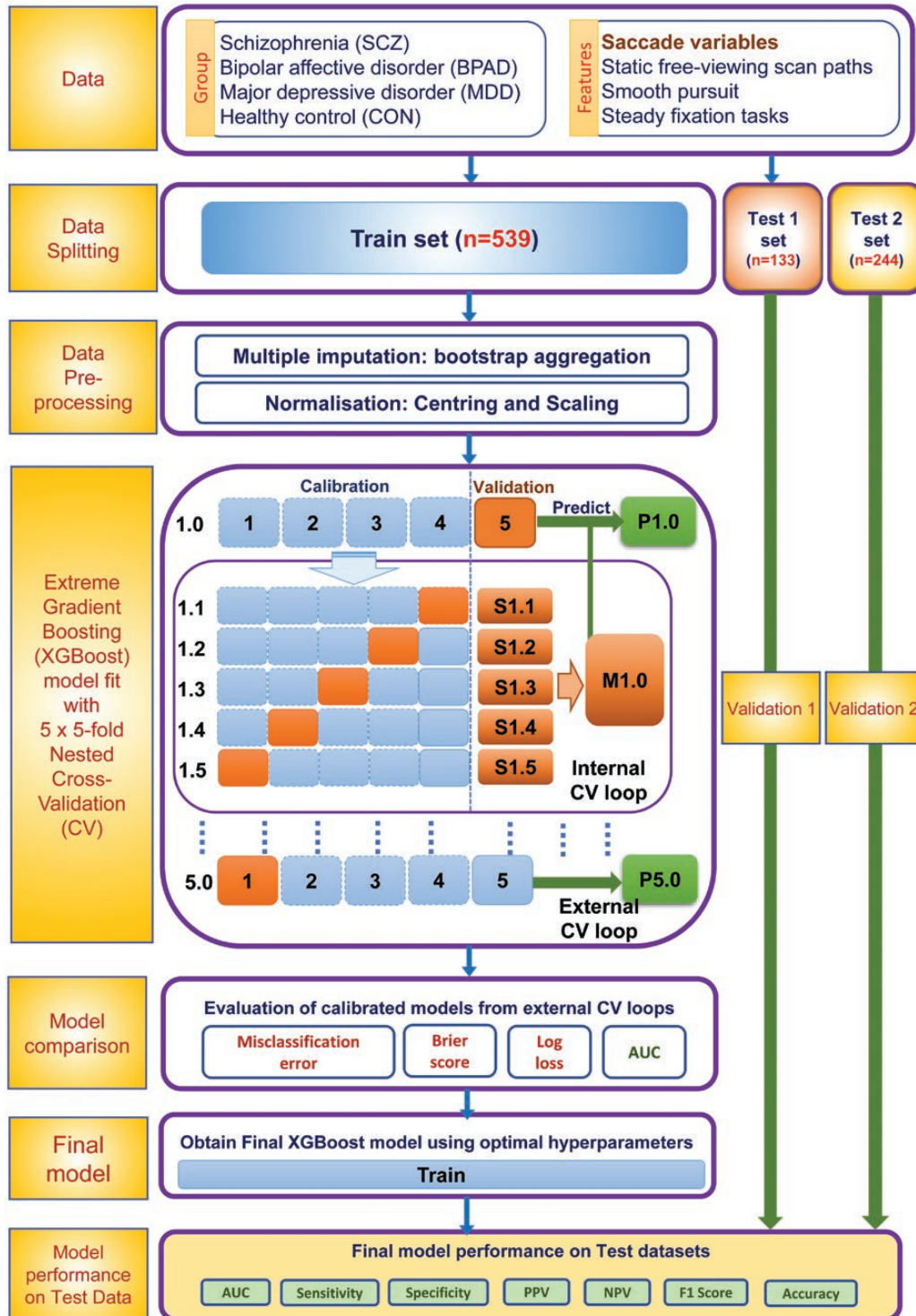


Fig. 1. Workflow pipeline for multiclass classifier to predict psychiatric disorders using eye movement data.

value (NPV), F1 score, accuracy and balanced accuracy under OVA using the 50% cutoff probability. We conducted the bootstrap sampling with 1000 replicates to calculate the 95% confidence interval of these performance metrics.

Potential Confounders

We considered demographic, behavioral, and other clinical factors that could potentially influence eye tracking performance. These included: age, caffeine consumption, nicotine use, anxiety, and depression subscores at time of testing (based on the Hamilton Anxiety and Depression Scale or HADS questionnaire) and current psychotropic medication. To assess the confounding effect of age, we developed two classifiers with eye movement data—including and excluding the age variable—and compared their performance metrics.

To further evaluate the importance of potential confounding variables, we conducted a principal component (PC) analysis of 98 eye movement features of all 672 participants and computed PC scores of each participant for the first 20 PCs, which captured approximately 80% variability of the eye movement data. We then fitted a separate single variable PC regression model by regressing the PC scores on each potential confounding variable and estimated the coefficient of determination (R^2) representing the proportion of variability explained by the confounding variable.

Results

Clinical Samples

The demographic characteristics of the groups used for training the classifier are shown in [Table 1](#). Demographics for the two test datasets are presented in [Supplementary Table 1](#). The bipolar group was similar to the schizophrenia group, but the average age was older. The patients in the recurrent unipolar depression group were older than the bipolar group. The amount and type of medication varied both within and between groups; almost all schizophrenia patients were receiving neuroleptic medication while bipolar patients received mood stabilizers and/or antidepressants, with 30% also prescribed neuroleptics. All save five recurrent unipolar subjects were prescribed psychotropic medication; these included antidepressants and in a minority of cases also mood stabilizers and/or anxiolytics.

Eye Movement Variables

We considered a total of 98 eye movement variables broadly consisting of different summarized measurements within the domains of free-viewing, fixation stability, and smooth pursuit activities. [Supplementary Table 2](#) presents the description of all the eye movement variables and [Supplementary Table 3](#) provides summary statistics (median, Q1, Q3, minimum, and maximum) based on the full (Train + Test-1) population. [Supplementary Figure](#)

Table 1. Demographics of the Patient and Control Subjects Used for Developing the Classifier. Demographics of Others Used for Validating the Classifier (Test-1 and Test-2) Can be Found in [Supplementary Table 1](#)

	Schizophrenia (SCZ)	Bipolar Affective Disorder (BPAD)	Major Depression Disorder (MDD)	Healthy Control (CON)
Training data (Train; N = 539)				
N	120	141	136	142
Sex,	35:85	81:60	86:50:00	84:58:00
Female:Male				
Age (years),	43.0	47.0	46.5	28.0
Median (Q1, Q3)	(32.8, 51.2)	(38.0, 55.0)	(36.0, 57.0)	(23.0, 45.0)
Education (years),	12.0	15.0	13.5	15.0
Median (Q1, Q3)	(9.5, 15.0)	(12.0, 15.0)	(10.0, 15.0)	(13.5, 16.0)
Illness age of onset (years),	22.0	27.0	30.0	
Median (Q1, Q3)	(19.0, 28.0)	(21.0, 35.0)	(23.0, 41.0)	
Illness duration (years),	19.0	17.0	11.0	
Median (Q1, Q3)	(13.8, 27.5)	(9.0, 26.0)	(6.0, 21.0)	
CPZ,	525.0	50.0	0.0	
Median (Q1, Q3)	(237.5, 862.5)	(0.0, 150.0)	(0.0, 33.2)	
Nicotine (cigarettes/day),	5.0	0.0	0.0	0.0
Median (Q1, Q3)	(0.0, 20.0)	(0.0, 10.0)	(0.0, 0.0)	(0.0, 0.0)
Caffeine intake (cups/day),	4.0	4.0	4.0	3.0
Median (Q1, Q3)	(2.2, 7.0)	(2.0, 5.0)	(3.0, 6.0)	(1.0, 5.0)
HADS Anxiety,	8.0	8.0	11.0	4.0
Median (Q1, Q3)	(6.0, 12.0)	(4.0, 12.0)	(7.0, 14.0)	(2.0, 6.0)
HADS Depression,	7.0	5.0	8.0	1.0
Median (Q1, Q3)	(4.0, 9.0)	(2.0, 10.0)	(5.0, 12.0)	(0.0, 3.0)

Note: CPZ, Chlorpromazine equivalents (mg/day). Q1, Q3 are first and third quartiles. Education indicates years in full time education.

1 illustrates an overview of the estimates of correlation coefficients between the variables. In general, variables within the same domain showed positive correlation while variables of different domains showed poor correlation. Variables related to fixation frequency and fixation duration within the smooth pursuit domain showed a strong negative correlation.

Performance Measures of Gradient Boosting Multiclass Classifier

Different performance metrics of the classifier on Test-1 and Test-2 datasets along with the corresponding confusion tables are presented in Tables 2 and 3. The receiver operating characteristic (ROC) plots of four groups under OVA are presented in figure 2. For Test-1, the

Table 2. Confusion Matrix and Estimates of Area Under the Curve (AUC) Based on Validations of the Fitted Multiclass Classifier on Two Test Datasets

Validation on Test-1	Reference				Group	AUC ^b			
Prediction	CON	SCZ	BPAD	MDD	CON	SCZ	BPAD	MDD	
CON	20	4	4	6	CON	0.85			
SCZ	7	18	5	3	SCZ	0.84	0.85		
BPAD	4	3	17	6	BPAD	0.85	0.82	0.78	
MDD	4	5	9	18	MDD	0.82	0.84	0.76	

Validation on Test-2 ^a	Reference				Group	AUC ^b	
Prediction	CON	SCZ	BPAD	MDD	CON	SCZ	
CON	115	17			CON	0.64	
SCZ	24	35			SCZ	0.89	
BPAD	31	1					
MDD	21	0					

Note: CON, Healthy Control; SCZ, Schizophrenia; BPAD, Bipolar Affective Disorder; MDD, Major Depressive Disorder.

^aThe validation on Test-2 dataset does not have any representation of BPAD and MDD patients.

^bAUC table represents the overall groupwise AUC by one-versus-all (OVA) (diagonal element) and pairwise AUCs by one-versus-one (OVO) methods (off-diagonal elements).

The bolds in the table on the left confusion matrix are the correctly classified cases, on the AUC right side the bold are one versus all comparisons, the non bold are One versus one comparisons.

Table 3. Performance Metrics Based on Validations of the Fitted Multiclass Classifier on Two Test Datasets

Statistics	CON	SCZ	BPAD	MDD
Validation on Test-1				
Sensitivity	0.57	0.60	0.49	0.55
Specificity	0.86	0.85	0.87	0.82
Positive Predictive Value (PPV)	0.59	0.55	0.57	0.50
Negative Predictive Value (NPV)	0.85	0.88	0.83	0.85
F1 Score	0.58	0.57	0.52	0.52
Accuracy	0.78	0.80	0.77	0.73
Balanced Accuracy	0.71	0.73	0.68	0.68
Validation on Test-2^a				
Sensitivity	0.63	0.58		
Specificity	0.60	0.91		
Positive Predictive Value (PPV)	0.83	0.67		
Neg Predictive Value (NPV)	0.34	0.87		
F1 Score	0.71	0.63		
Accuracy	0.62	0.83		
Balanced Accuracy	0.61	0.75		

Note: CON, Healthy Control; SCZ, Schizophrenia; BPAD, Bipolar Affective Disorder; MDD, Major Depressive Disorder. Further details on performance metrics are provided in Supplementary Table 4.

^aThe validation on Test-2 dataset does not have any representation of BPAD and MDD patients.

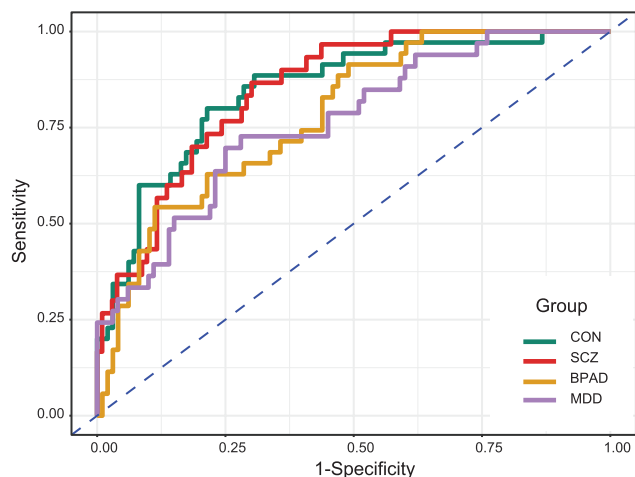


Fig. 2. Receiver operating characteristic (ROC) curves of schizophrenia (SCZ), bipolar affective disorder (BPAD), major depressive disorder (MDD), and healthy control (CON) based on validation of the fitted multiclass classifier on Test-1 dataset under One-versus-All (OVA) scenario.

overall generalized estimate of AUC was 0.81. Estimates under OVA ranged from 0.76 to 0.85 while estimates under OVO ranged from 0.82 to 0.85 except for BPAD vs MDD, which was 0.69. For Test-2, estimates of OVA for CON and SCZ were 0.64 and 0.89, respectively, while the estimate of OVO between CON and SCZ was 0.77. On the Test-1 dataset, the fitted gradient boosted multiclass classifier correctly predicted 57%, 60%, 49%, and 55% patients of CON, SCZ, BPAD, and MDD while for the Test-2 dataset, the classifier correctly predicted 63% and 58% of patients of CON and SCZ group using the notional cutoff probability of 50%. The representation of four groups ranged from 23% to 26% in Test-1 dataset. The specificity of the classifier was over 82% in the Test-1 dataset, while it was over 63% in the Test-2 dataset. The overall accuracy of the classifier was 0.55 (95% confidence interval, 0.46, 0.64) for the Test-1 and 0.61 (95% confidence interval, 0.55, 0.68) for the Test-2. [Supplementary Table 4](#) provides detailed performance metrics and the 95% confidence interval (lower, upper) of the classifier on both validation datasets obtained from bootstrap sampling using 1000 replicates.

Important Eye Movement Variables

The best individual discriminators included free viewing, fixation duration, and smooth pursuit tasks. Complete information of variable importance for all 98 eye movement variables and sex is provided in [Supplementary Table 5](#).

Effects of Potential Confounders

The classifier including age showed very modest changes in the performance metrics when compared with the classifier excluding age. For example, the estimates of AUC for

SCZ vs other groups increased from 0.85 to 0.86 when age was included as a feature in the classifier ([Supplementary Table 6](#)). The complete performance measurements of the classifier including age and eye movement features are presented in [Supplementary Tables 6-8](#). [Supplementary Table 10](#) shows the age variable included in the principal component regression analyses. Results show a marginal influence of age on saccade variables (9.6% on PC1, 6.1% on PC4, and 1.7% on PC7). Age explains approximately 2.9% of the variability in saccade features. The variable importance for sex was low (ranked 84th for the classifier excluding age, and 67th for the classifier including age).

Psychotropic medication effects were the most important potential confounders. Since almost all individuals with schizophrenia were receiving neuroleptics, and among the rest, only a minority with BPAD were prescribed neuroleptics, a direct comparison across groups was not possible. By contrast, almost all affective disorder subjects were prescribed antidepressant medications. Based on principal component regression models of the first 20 PC scores that represent approximately 80% variability of eye movement data, we observed that psychotropic medications contributed 3.3% of the variability in the data. In contrast, other confounders like caffeine consumption and nicotine use, anxiety, and depression subscores from the HADS questionnaire, explained a negligible amount of variability (range from 0.3% to 0.7%) in the eye movement data. [Supplementary Table 9](#) presents additional summary statistics of different confounders. We explored the clinical and demographic features of the 12 test schizophrenia cases wrongly classified and the 15 nonschizophrenia subjects missclassified as schizophrenia on the four-way comparisons. No consistent pattern emerged among the 12 missclassified schizophrenia cases, but 3 of 5 bipolar cases missclassified as schizophrenia had severe or very severe illness eg, on depot medication or spent many years in hospital. Four of 7 control cases had minor anomalies, astigmatism, restlessness, and calibration difficulties during testing, and very odd personality. [Supplementary Table 11](#) gives the predictive probability scores of all 133 test subjects using four-way comparisons. Of 133 test cases, the average predictive probability of correctly classified cases was 0.6976 and missclassified cases 0.5882; for schizophrenia cases, average predictive probability was 0.6847 for correctly diagnosed and 0.5530 for missclassified cases.

Discussion

Our results demonstrate that eye movement behavior patterns discriminate schizophrenia from unipolar and bipolar affective disorders and unaffected controls. We validated the classifier using the hold-out (Test-1) and semi-independent (Test-2) datasets for validations. Test-1 showed an average Area Under Curve (AUC) of

0.84 when the schizophrenia group was compared to the three other groups combined. Test-2 also supported our findings with an overall AUC of 0.77. An AUC of around 0.80 is generally considered good predictive performance.^{38,39} Pairwise comparisons of schizophrenia confirmed similar predictive ability. Bipolar and MDD patients also performed well for the one-versus-all (OVA) comparisons but indicated comparatively lower discrimination on direct comparisons with each other (see [Tables 2 and 3](#) and [figure 2](#)). Based on the 50% cutoff probability, the confusion matrices presented similar levels of discrimination and accuracy on both validation sets. The overall accuracy of 61%, as observed, is considerably better than the estimates of prevalence (23% to 26%) of four groups in the Test-1 dataset or that expected by chance (ie, 25%).

Except for AUC, performance metrics like sensitivity, specificity, positive and negative predictive value etc. are the outcomes from the OVA comparison with the notional cut-off probability set at 0.50. Naturally, estimates of performance metrics are conditional on the choice of the cut-off value; this choice generally incorporates the information from the ROC data with the utility-based decision theory to identify the optimal cut-off point in a practical scenario. For example, the decision may include the disease prevalence, maximising sensitivity or specificity, maximising accuracy, cost of false positive, or negative results etc., accounting for how the model is to be used for the decision-making. On the other hand, ROC curve analysis and AUC-based estimates presented here have several advantages. AUC is an effective and combined measure of sensitivity and specificity that describes the inherent validity of the diagnostic test to discriminate between positive and negative populations. It is not influenced by decision criteria and the prevalence of the disease. AUC estimate is an ideal quantitative measurement that supports comparing different tests and combining multiple tests to improve diagnostic accuracy.^{38,39} Therefore, AUC is a reliable performance metric for the diagnostic test, and AUC estimates obtained from the model demonstrate good discriminating ability between three psychological disorders and unaffected controls.

Although the overall performance metrics are encouraging, we acknowledge a model developed on larger sample sizes integrated with rigorous decision criteria are required to upgrade the model from the current levels of modest sensitivity and good specificity to the point where the model would demonstrate clinical utility. However, to our knowledge, this is the first large study to document the good model-based predictive performance of eye movement patterns of healthy controls vs schizophrenia and two other major mood disorders.

Elements of the free viewing and fixation tasks best distinguished individuals with schizophrenia from other

groups; they produced restricted viewing patterns when looking at static images and had difficulty inhibiting saccades towards a distractor during the steady fixation task. Individuals with mood problems exhibited faster saccades than other groups during free viewing, and bipolar cases also produced restricted fixation activity distinct from controls but not as pronounced as in the schizophrenia group. Subtle differences in smooth pursuit performance differentiated affective disorders from other groups. We did not try to weigh the discriminatory importance of the variables for each disorder separately in the four-way comparisons.

We observed that psychotropic medication contributed a modest 3.3% of the variability in eye movement behavior patterns. This finding suggests that medication through reverse causation is unlikely to be responsible for differences between patient groups. An extensive literature on eye movement abnormalities in medication-free schizophrenics and high-risk relatives supports this view.⁴⁰ We have previously reported eye movement patterns in a small number of schizophrenia patients medication-free at the time of testing were similar to medicated schizophrenia patients.²⁸ No new medication-free schizophrenia patients were available in the current study. We did, however, observe a minimal association between the amount of chlorpromazine equivalents and the primary eye movement patterns using principal component analysis (see [Supplementary Table 10](#)). Similar conditions applied to affective disorders. Almost all patients received antidepressants and or mood-stabilizing medications. However, the main abnormalities compared to controls were in measures of smooth pursuit, and the literature suggests relative independence of abnormal smooth pursuit eye movements from antidepressant and neuroleptic medications.^{40,41}

The affective disorders groups in this study were for comparison purposes only. The majority (ca. 65%) of bipolar cases were bipolar I, namely those that share many clinical features with schizophrenia. All MDD cases were under psychiatric supervision and represented the severer end of the MDD spectrum. Mental state at the time of testing MDDs varied across subjects, with most in partial remission of symptoms. The sample sizes of unipolar and bipolar cases are sufficient to be confident that schizophrenia cases can be distinguished from them with a high degree of accuracy. By contrast, the bipolar and unipolar cohorts showed good separability from control subjects but not from each other. Larger sample sizes are required to determine if eye movement behavior patterns can stratify affective disorders as a whole into clinically useful subgroups such as bipolar patients with and without psychosis. It remains to be seen whether cases with noncore or missing features of the disorders will reveal the same differences, thus delineating the approximate core diagnostic category.⁴² It is essential to know the time of first atypical eye movement patterns and its importance for

the early diagnosis of individuals at high risk of major mental health problems.

Strengths and Limitations

Our multiclass classifier model of eye movement behavior patterns distinguished schizophrenia from bipolar and unipolar disorder and controls with good predictive performance. All cases met DSMIV operationalized criteria. It is especially encouraging that all MDD cases were ascertained through psychiatric services. These patients with moderate to severe depression are clinically more difficult to differentiate from other forms of major mental illness. The sample sizes for all groups were reasonably large, balanced, and internally consistent. The relatively advanced age of the unipolar cohort also makes it probable that few will convert at a later date to a bipolar diagnosis. In contrast to many current predictive neuro-imaging and genetic studies in psychiatry, we validated our classifier using two validation datasets. As in our earlier article for predictive performance, we used a GB modelling framework on these much-expanded schizophrenia and control cohorts. The findings for schizophrenia vs unaffected controls groups are essentially unchanged from our earlier findings and those of Morita et al.^{28,29}

We implemented a rigorous machine learning pipeline to develop a robust, reproducible, and accurate gradient boosted multiclass classifier and validated the classifier using two validation datasets. The nested cross-validation framework separated the classifier learning task from the calibration task and accurately estimated performance metrics by averaging across folds, hence the setup allowed limiting the overfitting and yielded robust tuning of the postprocessing algorithm.^{35,36} We adopted a transparent approach to model the data and provided a detailed pipeline to account for samples, features, missing value imputation, hyperparameter search, and algorithms, therefore, enhancing the reproducibility of results.³⁵ We obtained the 95% confidence intervals of all performance metrics using a rigorous bootstrap-based approach.

The study also has some limitations. To our knowledge, we used the largest sample of this kind available and adopted rigorous strategies to limit the overfitting: however, further calibration of the model with larger training samples is necessary to try to improve sensitivity. We could not conduct full external validation of the model due to the nonavailability of all phenotypes. Test-1 and Test-2 controls were recruited from the same Scottish study and followed the same experimental protocol. We also used data from the German schizophrenia subjects exclusively for external validation. Although we previously reported the German data in a different modelling context we considered them a semi-independent group because they were recruited through Ludwig Maximilian University, but were tested using the same protocols by the Aberdeen research staff

on temporary secondment to Munich. All eye movement data for the affective disorders comparative groups were not reported earlier. The psychiatric conditions we studied are all highly heterogeneous and overlap clinically: a discrimination much higher than AUC of 0.8 would raise concerns of overfitting of the training data, and much smaller value would have minimal clinical value. There are also, as Moriarity et al⁴³ indicate, many unobserved issues of measurements of noninvariance in biological psychiatry that can complicate otherwise promising findings. In these circumstances, Onitsuka et al⁴⁴ highlight there is an urgent need for replication of promising eye movement findings in psychiatry in large multicentred studies.

Conclusions

Eye movement behavior patterns can distinguish schizophrenia from major affective disorders and unaffected controls with a reasonable degree of accuracy. Although the performance characteristics based on AUC are in the 80% range, most of this is accounted for by good specificity, and only modest sensitivity. Larger sample sizes integrated with robust decision criteria will be needed to obtain a clear picture of the biological significance of our findings and whether they may prove useful in a clinical setting.

Supplementary Material

Supplementary data are available at *Schizophrenia Bulletin Open* online.

Funding

This project was supported by the following grants: The Royal Society of London, Chief Scientist Office Scotland (CZB/4/734), NHS Grampian Tenovus Scotland (G12/31), NHS Grampian Endowment Fund, Miller MacKenzie Trust, EU-FP6 (SGENE) and Health Innovation Challenge Fund, jointly from Wellcome Trust and Department of Health (WT-103911/Z/14/Z). The funders had no role in the original study design, the ongoing data collection and analysis, interpretation, or writing of the manuscript.

Acknowledgments

We thank all who helped with clinical aspects of the study including research assistants Barbara Duff, Kate Cotton, Foteini Okonomitsiou, Elizabeth Hannaford, Zsuzsanna Nemeth and Joanna Rodzinko Paska as well as the patients and volunteers whose help was indispensable. P Benson and D St Clair are co-founders of SACCADe Diagnostics Ltd a spin out company tasked to develop eye movement technology to assist diagnosis of major mental health disorders. The University of Aberdeen has patents

pending in Europe (PCT/GB2013/050016) and USA (14/370,611). The data reported in this paper arose solely from funding by the acknowledged UK research bodies and charities none of whom have vested interests in the company. David St Clair had full access to all the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis. *Design of study and development of protocols*: D St Clair, P Benson and S Beedie. Recruitment of patients, case note review and clinical and eye movement data collection, quality control and feature extraction of eye movement variables: St Clair, Rujescu, MacIntosh, Beedie, Lemmon, Nouzova. *Drafting of the manuscript*: St Clair and Nath. *Critical revision of manuscript for important intellectual content*: all authors. *Statistical analyses*: Nath and Benson. *Interpretation of results*: Nath, Benson, MacLennan and St Clair. *Obtained funding*: St Clair, Benson, MacIntosh, Rujescu. *Supervision*: St Clair, Benson, MacIntosh, Nath, Rujescu.

Data Availability

Requests for access to anonymized study data for replication or related studies should be directed to the corresponding author. All reasonable requests will be considered positively. There are plans for the datasets to be publically posted in due course.

References

- Hyman SE. The diagnosis of mental disorders: the problem of reification. *Annu Rev Clin Psychol.* 2010;6:155–179.
- Insel T, Cuthbert B, Garvey M, et al. Research Domain Criteria (RDoC): toward a new classification framework for research on mental disorders. *Am J Psychiatry.* 2010;167:748–751.
- Kapur S, Phillips AG, Insel TR. Why has it taken so long for biological psychiatry to develop clinical tests and what to do about it? *Mol Psychiatry.* 2012;17(12):1174–1179.
- Quinlan EB, Banaschewski T, Barker GJ, et al. Identifying biological markers for improved precision medicine in psychiatry. *Mol Psychiatry.* 2020;25(2):243–253.
- Garcia-Gutierrez MS, Navarette F, Sala F, et al. Biomarkers in psychiatry: concept, definition, types and relevance to the clinical reality. *Front Psychiatry.* 2020;11:432.
- Craddock N, Owen MJ. The Kraepelinian dichotomy - going, going... but still not gone. *Br J Psychiatry.* 2010;196(2):92–95.
- Kendler KS. Psychosis within vs outside of major mood episodes: a key prognostic and diagnostic criterion. *JAMA Psychiatry.* 2013;70(12):1263–1264.
- Lichtenstein P, Yip BH, Björk C, et al. Common genetic determinants of schizophrenia and bipolar disorder in Swedish families: a population-based study. *Lancet.* 2009;373(9659):234–239.
- Cross-Disorder Group of the Psychiatric Genomics Consortium. Identification of risk loci with shared effects on five major psychiatric disorders: a genome-wide analysis. *Lancet.* 2013;381(9875):1371–1379.
- Van Snellenberg JX, de Candia T. Meta-analytic evidence for familial coaggregation of schizophrenia and bipolar disorder. *Arch Gen Psychiatry.* 2009;66(7):748–755.
- Anttila V, Bulik-Sullivan B, Finucane HK, et al. Brainstorm Consortium analysis of shared heritability in common disorders of the brain. *Science.* 2018;36(6395):eaap8757.
- Cross-Disorder Group of the Psychiatric Genomics Consortium. Genomic relationships, novel loci and pleiotropic mechanisms across eight psychiatric disorders. *Cell.* 2019;179(7):1469–1482.
- Kotov R, Leong SH, Mojtabai R, et al. Boundaries of schizoaffective disorder: revisiting Kraepelin. *JAMA Psychiatry.* 2013;70(12):1276–1286.
- Kennis M, Gerritsen L, van Dalen M, Williams A, Cuijpers P, Bocking C. Prospective biomarkers of major depressive disorder: a systematic review and meta-analysis. *Mol Psychiatry.* 2020;25(2):321–338.
- Abi-Dargham A, Horga G. The search for imaging biomarkers in psychiatric disorders. *Nat Med.* 2016;22(11):1248–1255.
- Cai XL, Xie DJ, Madsen KH, et al. Generalizability of machine learning for classification of schizophrenia based on resting-state functional MRI data. *Hum Brain Mapping.* 2020;41(1):172–184.
- Kraguljac NV, McDonald WM, Widge AS, Rodriguez C, Tohen M, Nemeroff CB. Neuroimaging biomarkers in schizophrenia. *Am J Psychiatry.* 2021;178(6):509–521.
- Chen J, Li X, Calhoun VD, et al. Sparse deep neural networks on imaging genetics for schizophrenia case-control classification. *Hum Brain Mapp.* 2021;42(8):2556–2568.
- Bracher-Smith M, Crawford K, Escott-Price V. Machine learning for genetic prediction of psychiatric disorders: a systematic review. *Mol Psychiatry.* 2021; 26(1):70–79.
- Diefendorf AR, Dodge R. An experimental study of the ocular reactions of the insane from photographic records. *Brain.* 1908;31:451–489.
- Wolf A, Ueda K, Hirano J. Recent updates of eye movement abnormalities in patients with schizophrenia: a scoping review. *Psychiatry Clin Neurosci.* 2021;75(3):82–100.
- Bestelmeyer PE, Tatler BW, Phillips LH, Fraser G, Benson PJ, St Clair D. Global visual scanning abnormalities in schizophrenia and bipolar disorder. *Schizophr Res.* 2006;87(1-3):211–222.
- Lencer R, Sprenger A, Reilly JL, et al. Pursuit eye movements as an intermediate phenotype across psychotic disorders: evidence from the B-SNIP study. *Schizophr Res.* 2015;169(1-3):326–33.
- Carvalho N, Laurent E, Noiret N, et al. Eye movement in unipolar and bipolar depression. *Front Psychol.* 2015; 6:1809.
- Takahashi J, Hirano Y, Miura K, et al. Eye movement abnormalities in major depressive disorder. *Front Psychiatry.* 2021;12:673443.
- Wang Y, Lyu HL, Tian XH, et al. The similar eye movement dysfunction between major depressive disorder, bipolar depression and bipolar mania. *World J Biol Psychiatry.* 2022;1–14. doi:10.1080/15622975.2022.2025616.
- Clementz BA, Parker DA, Trotti RL, et al. Psychosis biotypes: replication and validation from the B-SNIP Consortium. *Schiz Bull.* 2022;48(1):56–68.
- Benson PJ, Beedie SA, Shephard E, Giegling I, Rujescu D, St Clair D. Simple viewing tests can detect eye movement abnormalities that distinguish schizophrenia cases from controls with exceptional accuracy. *Biol Psychiatry.* 2012;72(9):716–724.
- Morita K, Miura K, Fujimoto M, et al. Eye movement as a biomarker of schizophrenia: Using an integrated eye movement score. *Psychiatry Clin Neurosci.* 2017;71(2):104–114.

30. Coors A, Merten N, Ward DD, Schmid M, Breteler MM, Ettinger U. Strong age but weak sex effects in eye movement performance in the general adult population: evidence from the Rhineland Study. *Vision Res.* 2021;178:124–133. doi:10.1016/j.visres.2020.10.004.
31. Takahashi J, Miura K, Morita K, et al The effects of age and sex on eye movement characteristics. *Neuropsychopharmacol Rep.* 2021;41(2):152–158.
32. McGuffin P, Farmer A, Harvey I. A polydiagnostic application of operational criteria in studies of psychotic illness: development and reliability of the OPCRIT system. *Arch Gen Psychiatry.* 1991;48(8):764–770.
33. Sheehan DV, Lecrubier Y, Sheehan KH, et al The mini-international neuropsychiatric interview (M.I.N.I.): the development and validation of a structured diagnostic psychiatric interview for DSM-IV and ICD-10. *J Clin Psychiatry.* 1998;59(Suppl 20):22–33.
34. First MB, Spitzer RL, Gibbon M, et al. *Structured Clinical Interview for DSMIV Axis I Disorders-Patient Edition (SCIDIP, Version 2.0)*. New York: New York State Psychiatric Institute, Biometrics Research Department; 1997.
35. Maros ME, Capper D, Jones DT, et al. [Machine learning workflows to estimate class probabilities for precision cancer diagnostics on DNA methylation microarray data.](#) *Nature Protoc.* 2020;15(2):479–512.
36. Hand DJ, Till RJ. A simple generalisation of the area under the ROC curve for multiple classification problems. *Mach Learn.* 2001;45:171–186.
37. Hosmer DW, Lemeshow S. *Applied Logistic Regression*. 2nd ed., chapter 5. New York: John Wiley and Sons; 2000:160–164.
38. D'Agostino RB, Grundy S, Sullivan LM, et al. Validation of the Framingham coronary heart disease prediction scores: results of multiple ethnic groups investigation *JAMA.* 2001;286(2):180–187.
39. Hajian-Tilaki K. Receiver operating characteristic (ROC) curve analysis for medical diagnostic test evaluation. *Caspian J Intern Med.* 2013;4(2):627–635.
40. Reilly JL, Lencer R, Bishop JR, Keedy S, Sweeney JA. Pharmacological treatment effects on eye movement control. *Brain Cogn.* 2008;68(3):415–435.
41. Flechtner KM, Steinacher B, Sauer R, Mackert A. Smooth pursuit eye movements of patients with schizophrenia and affective disorder during clinical treatment. *Eur Arch Psychiatry Clin Neurosci.* 2002;252(2):49–53.
42. Westen D. Prototype diagnosis of psychiatric disorders. *World Psychiatry.* 2012;11(1):16–21.
43. Moriarity DP, Joyner KJ, Slavich GM, Alloy LB. Unconsidered issues of measurement of nonvariance in biological psychiatry: a focus on biological phenotypes of psychopathology. *Mol Psychiatry.* 2022. doi:10.1038/s41380-021-01414-5.
44. Onitsuka T, Hirano Y, Nemoto K, et al. Trends in big data analyses by multicenter collaborative translational research in psychiatry. *Psychiatry Clin Neurosci.* 2022;76(1): 1–14.