

ARTICLE

Open Access

# A signature-based machine learning model for distinguishing bipolar disorder and borderline personality disorder

Imanol Perez Arribas<sup>1</sup>, Guy M. Goodwin<sup>2,3</sup>, John R. Geddes<sup>2,3,4</sup>, Terry Lyons<sup>1,5</sup> and Kate E. A. Saunders<sup>2,3,4</sup>

## Abstract

Mobile technologies offer new opportunities for prospective, high resolution monitoring of long-term health conditions. The opportunities seem of particular promise in psychiatry where diagnoses often rely on retrospective and subjective recall of mood states. However, deriving clinically meaningful information from the complex time series data these technologies present is challenging, and the current implications for patient care are uncertain. In this study, 130 participants with bipolar disorder ( $n = 48$ ) or borderline personality disorder ( $n = 31$ ) and healthy volunteers ( $n = 51$ ) completed daily mood ratings using a bespoke smartphone app for up to 1 year. A signature-based learning method was used to capture the evolving interrelationships between the different elements of mood and exploit this information to classify participants' diagnosis and to predict subsequent mood. The three participant groups could be distinguished from one another on the basis of self-reported mood using the signature methodology. The methodology classified 75% of participants into the correct diagnostic group compared with 54% using standard approaches. Subsequent mood ratings were correctly predicted with >70% accuracy. Prediction of mood was most accurate in healthy volunteers (89–98%) compared to bipolar disorder (82–90%) and borderline personality disorder (70–78%). The signature method provided an effective approach to the analysis of mood data both in terms of diagnostic classification and prediction of future mood. It also highlighted the differing predictability and the overlap inherent within disorders. The three cohorts offered internally consistent but distinct patterns of mood interaction in their reporting which have the potential to enable more efficient and accurate diagnoses and thus earlier treatment.

## Introduction

Historically the diagnosis of psychiatric disorders has been hampered by the inherent inaccuracy of retrospective narrative recall of abnormal mental states and the difficulty of defining their persistence over time. Furthermore, diagnostic classifications have changed little since the original description made by Kraepelin and are not clearly related to the underlying biology, which remains unknown<sup>1</sup>. The shortcomings of current diagnostic categories have motivated attempts (notably the NIMH Research Domain Criteria ['RDoC']) to take a

radical and 'bottom up' approach to diagnosis. In recent years, the rapid development of mobile technologies has allowed more precise measures of subjective psychopathology. However, analysis of these data streams has been challenging. Often the data generated by these measurements is sequential in nature, where the most valuable information is contained in the order at which different events occur, rather than in any individual event. These types of data need an analytic approach that exploits this distinctive feature. Signatures from rough path theory<sup>2</sup> have proven to be an effective way of analysing these streams of data since they naturally capture the order of events. As a consequence, signatures have successfully been applied to a number of machine learning problems where the object that is being studied is a

Correspondence: Kate E. A. Saunders ([kate.saunders@psych.ox.ac.uk](mailto:kate.saunders@psych.ox.ac.uk))

<sup>1</sup>Mathematical Institute, University of Oxford, Oxford, UK

<sup>2</sup>Department of Psychiatry, University of Oxford, Oxford, UK

Full list of author information is available at the end of the article.

© The Author(s) 2018



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

stream of data. For instance, using the path signature as the core feature representation for handwritten sequential data, Dr. B. Graham won the ICDAR 2013 international challenge on online Chinese character recognition<sup>3</sup>. The effectiveness of signatures in data science has led to a number of publications where a signature-based approach has achieved better performance relative to approaches that do not use signatures, in areas such as handwriting recognition, gesture recognition and action detection<sup>4–12</sup>.

Bipolar disorder (BD) and borderline personality disorder (BPD) are common mental disorders<sup>13–15</sup>. BD is a mood disorder with a strong genetic basis while BPD is a disorder of personality commonly related to abusive experiences in childhood. Although the two disorders are thought to develop through different processes and mechanisms, BD and BPD can be difficult to differentiate clinically as both are characterised by mood instability and impulsivity, and the behavioural consequences can be similar<sup>16</sup>. However, correct diagnosis is essential because of the contrasting treatment approaches to the two disorders<sup>17,18</sup>.

Here, we use a signature-based machine learning model to re-analyse data obtained from a clinical study<sup>19</sup>, which explored daily reporting of mood in participants with bipolar disorder, borderline personality disorder and healthy volunteers. Specifically, we sought to classify the diagnosis of participants on the basis of their evolving mood and also predict their mood the following day. The generality of the signature-based machine learning model allows these problems to be treated in similar ways.

## Materials and methods

### Data

Prospective mood data was captured from 139 individuals who were taking part in the AMoSS study. Of the 139 recruited participants, nine participants who withdrew consent or failed to provide at least 2 months of data were excluded from further analysis, so the data set consists of 130 individuals in total. The AMoSS study was a prospective longitudinal study during which used a range of wearables in combination with a bespoke

smartphone app to better characterise mood instability and its correlates in bipolar disorder ( $n = 48$ ), borderline personality disorder ( $n = 31$ ) and healthy volunteers ( $n = 51$ ). Participants rated their mood daily across six different categories (anxiety, elation, sadness, anger, irritability and energy) using a 7-point Likert scale (with values from 1 (not at all) to 7 (very much)). These mood and symptom states were based on previous work done in the context of the OXTEXT study. Data were collected from each participant for a minimum of 3 months, although 61 of the 130 participants provided data for more than 12 months. Overall compliance was 81.2%. The number of male participants in the study was 16 for BP, 2 for BPD and 18 for healthy controls, of ages  $38 \pm 21$  (BP),  $34 \pm 15$  (BPD) and  $37 \pm 20$  (healthy volunteers). Forty-seven of the BP participants and 23 of the BPD participants were taking psychotropic medication. Healthy participants were all medication free (Table 2). An analysis of this data set has previously been published by Tsanas et al.<sup>20</sup>.

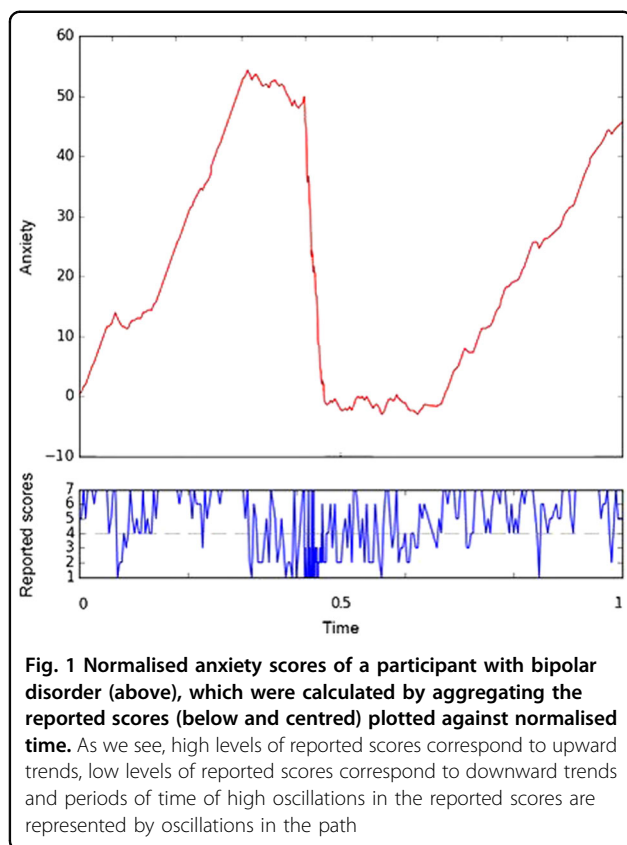
The mood data was divided into *buckets* comprised of streams of 20 consecutive self-reported mood states. Although these observations were typically recorded daily, data did not have to come from 20 consecutive days. Thus, streams could be constructed even when participants failed to record their mood on some days or when participants recorded their mood several times per day. This generated 2192 buckets of streamed data. These were randomly separated into a training set of 1534, and a

**Table 2 Demographic characteristics of the three groups. Where appropriate distributions are summarised in the form of the median +/- the interquartile range**

	Bipolar disorders (BD)	Borderline personality disorders (BPD)	Healthy controls (HC)
Originally recruited	53	33	53
Processed data from	48	31	51
Days in study	353±261	313±107	276±253
Age (years)	38±21	34±15	37±20
Gender (males)	16	2	18
Any psychotropic medication	47	23	0
Lithium	19	0	0
Anticonvulsant	19	1	0
Antipsychotic	33	6	0
Antidepressants	17	23	0
Hypnotics	3	2	0

**Table 1 Accuracy and area under the ROC curve**

	Healthy		Bipolar		Borderline	
	Accuracy	AUC	Accuracy	AUC	Accuracy	AUC
Healthy			84%	0.91	93%	0.98
Bipolar					80%	0.86
Borderline						



testing set of 658 streams of data. We then normalised the data to capture the intrinsic properties of the data independent of arbitrary units such as the scale used to record the scores or the period of time when the study was carried out.

The stream of mood reports was identified as a seven-dimensional path (one dimension for time, six dimensions for the scores) defined on the unit interval.

Figure 1 shows the aggregated normalised anxiety scores of a participant with bipolar disorder plotted against normalised time.

Figure 2 shows the normalised scores of each category plotted against all other categories. This way of representing the path allows interpretation of the order at which the participant changes mood. Take for example, the Angry vs. Elated plot (third row, first column). The path starts at the point (0, 0). As we see, the path moves left first. Therefore, the participant is becoming less and less elated, while the score in anger remains approximately constant. Suddenly, the period of low elation stops, and the participant starts recording low scores in anger. This low level in anger remains persistent for the rest of the path. On the other hand, if one considers the Angry vs. Irritable plot (fifth row, first column), we observe that

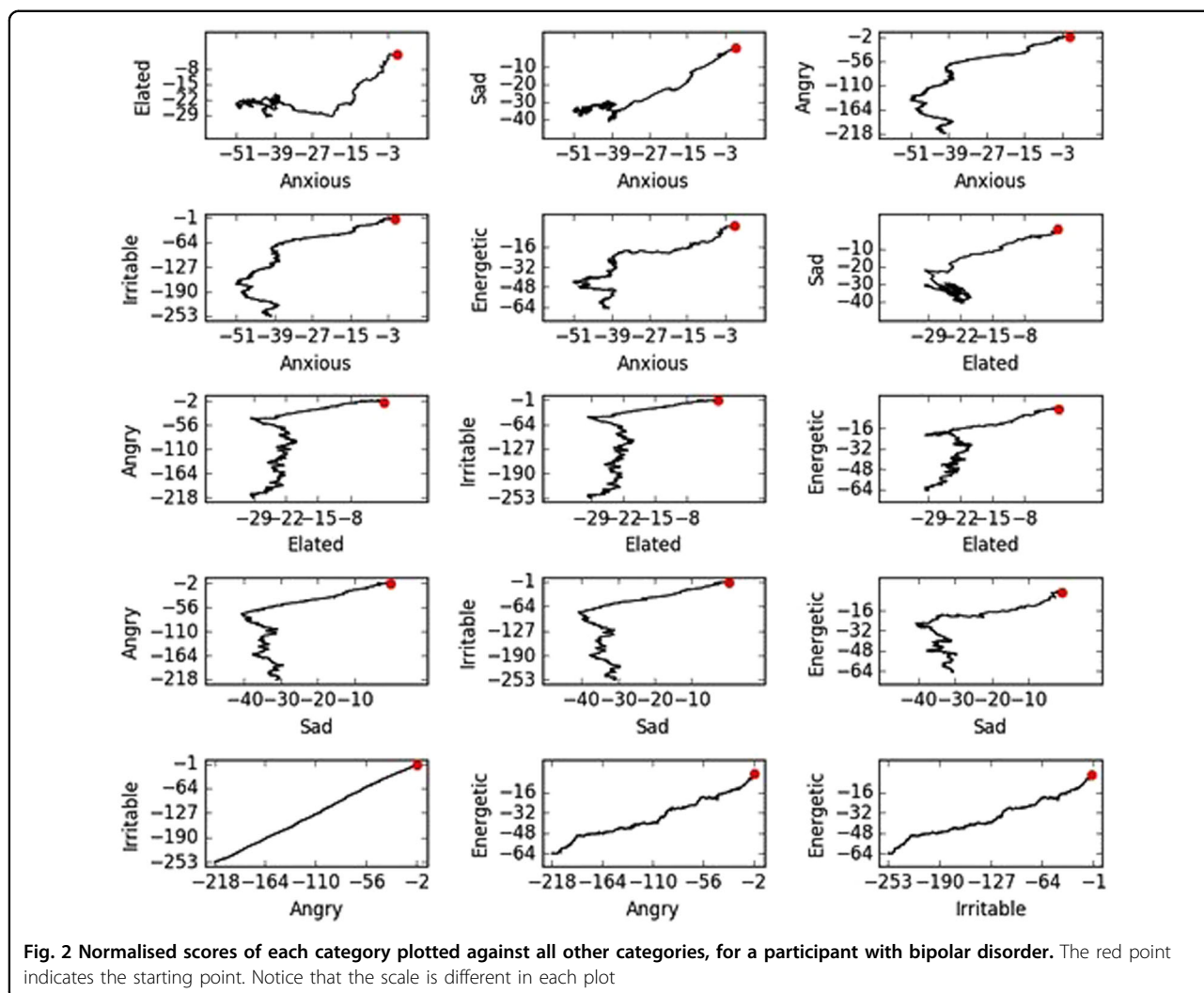
it is essentially a straight line and the levels of anger and irritability match closely for this participant.

### Group classification

We started by establishing a set of input-output pairs  $\{(R_i, Y_i)\}_i$ , where  $R_i$  is the seven-dimensional path of aggregated normalised scores for participant  $i$ , as described above, and  $Y_i$  denotes the group it was diagnosed into at the beginning of the study. This set was transformed into a new set of input/output pairs,  $\{(S^n(R_i), Y_i)\}_i$  where  $S^n(R_i)$  denotes the truncated signature<sup>2</sup> of order  $n$  of the stream  $R_i$ . The truncated signature is a feature set describing the stream whose size depends on the truncation degree but not on the number of sample points (or indeed the parameterisation) of the stream. The signature of order  $n$  grows exponentially with  $n$ , and therefore large values of  $n$  will produce input vectors of large dimensions but far smaller than other feature sets where the sampling rate directly changes the dimension of the feature set. We have only considered feature sets based on truncated signatures up to degree  $n = 1, 2, 3, 4$ . Random forest<sup>21</sup> was used to learn the mapping between the input and the output.

Given that it has already been established that there are differences in mean mood scores between the groups overall<sup>20</sup> we also calculated the mean score in each mood category and classified streams of 20 consecutive observations on this basis as a comparison to the signature method. To assess the performance of the classification procedure we computed accuracy, sensitivity, specificity, positive predicted value (PPV). We also trained the model using pairs of clinical groups and computing the area under the receiver operating characteristic curve (ROC) to assess the performance of the classification model at different values of threshold. Since this was done for each pair of clinical groups, we obtained a matrix with different values of the AUC. Moreover, in order to have a better understanding of how robust these percentages are, we applied bootstrapping to the model with the signature order fixed to 2.

We also considered the extent to which participants were characterised as belonging to each specific clinical group. We trained the model using all participants except the one we were interested in, and then tested the model with 20-observations periods from this person. We calculated the proportion of periods of time when the participant was classified as bipolar, healthy or borderline, which allowed us to plot the participant as a point in the triangle. This process was followed for every participant, although we excluded participants that generated <5 buckets of 20 observations in order to avoid extreme values for them. 121 participants were included.



### Predicting the mood of a participant

In this case, the aim was to predict the mood of a participant, using the last 20 observations of the participant. We constructed the input-output pairs  $\{(R_i, Y_i)\}_i$  where  $R_i$  was the normalised seven-dimensional path of a particular participant, and  $Y_i \in \{1, \dots, 7\}^6$  was the mood of the participant the next observation he or she registered. We then constructed a new set of input-output pairs  $\{(S^n(R_i), Y_i)\}_i$  with  $S^n(R_i)$  the truncated signature of order  $n$ . We applied regression using random forest to these pairs of inputs and outputs, thus obtaining our model. The model was trained using data from each clinical group separately.

To measure the accuracy of the predictions, we used the mean absolute error (MAE) and the percentage of correct predictions. A prediction  $\hat{y} \in \{1, \dots, 7\}$  was considered correct if  $|\hat{y} - y| \leq 1$ , where  $y$  is the correct score.

We used the publicly available Python eSig package (version 0.6.31) to calculate signatures of streams of data, Python pandas package (version 0.20.1) for statistical

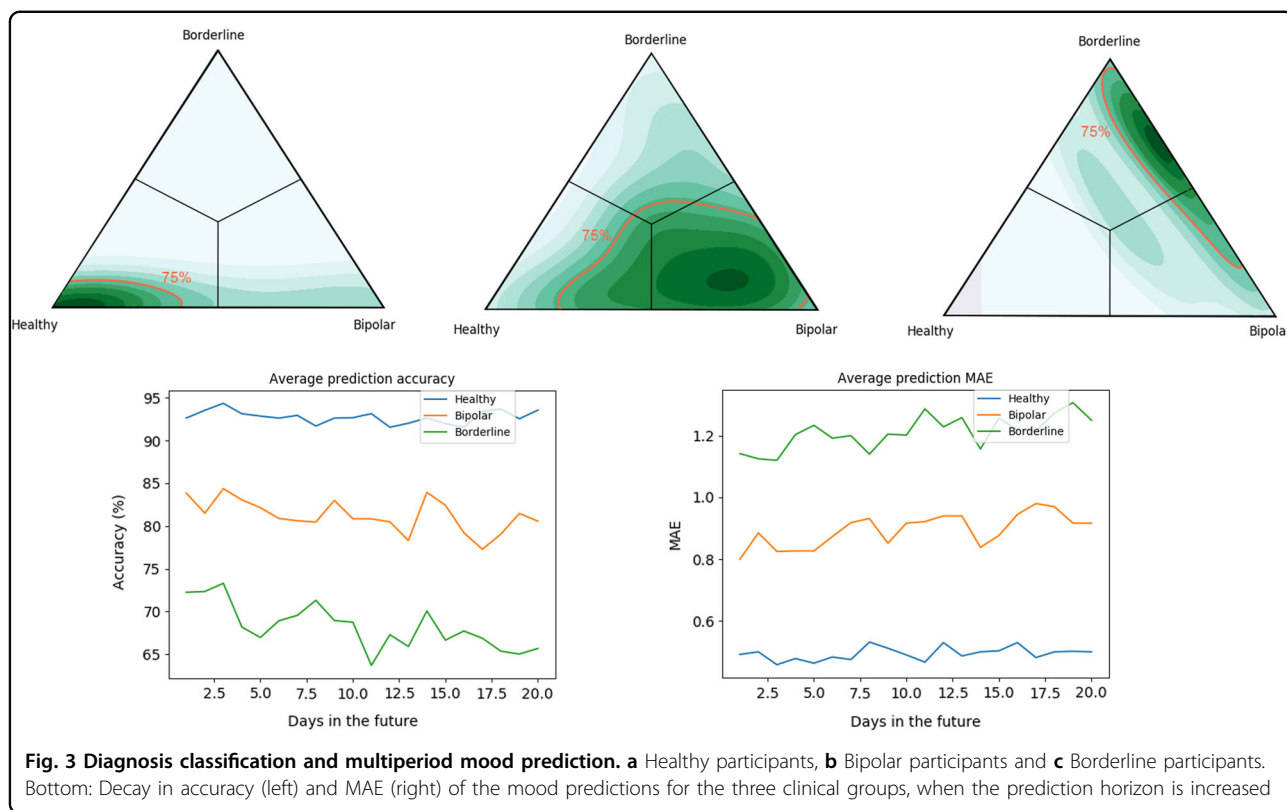
analysis, data manipulations and processing, Python scikit-learn package (version 0.18.1) for machine learning tasks and matplotlib for plotting and graphics (version 2.0.1).

The study was approved by the NRES Committee East of England—Norfolk (13/EE/0288) and the Research and Development department of Oxford Health NHS Foundation Trust.

### Results

#### Predicting group membership

The signature-based model categorised 75% of participants into the correct diagnostic group when the second-order signature was considered ( $n = 2$ , where  $n$  denotes the order of the signature). When the order was increased, the accuracy dropped slightly (70% for  $n = 3$  and 69% for  $n = 4$ ). The signature method performed significantly better than the naive model using the mean score in each category over the 20 observations, which classified just 54% of participants correctly. On bootstrapping the accuracy of the



**Fig. 3** Diagnosis classification and multiperiod mood prediction. **a** Healthy participants, **b** Bipolar participants and **c** Borderline participants. Bottom: Decay in accuracy (left) and MAE (right) of the mood predictions for the three clinical groups, when the prediction horizon is increased

second-order signature was 74.85% while the standard deviation was 2.05, suggesting that the results are stable and robust. Table 1 shows the accuracy and area under the ROC curve when only pairs of diagnosis are considered for classification, rather than all three groups.

The signature of order 2 was able to capture second-order effects on the stream of data and since the feature set has 57 dimensions it was reasonable to learn linear relations from the 517 buckets of data in the training set. To capture the full second-order effects using raw streams of data, the feature set would have required over 10,000 dimensions. A 57-dimensional feature set of randomly selected quadratic polynomials on the raw data stream was inferior to the signature model: the average accuracy was 68% for the random feature set, well below the 75% accuracy of the second-order signature model.

The proportion of a participant’s sequential mood buckets matched each specific diagnostic group and positioned that specific individual on a triangular spectrum; the distribution of that density for each initial diagnosis is shown in Fig. 3 (top). In each of the plots, the regions of highest density of participants are located in the correct corner of the triangle: the greatest consistency is with the healthy participants.

Table 3 shows that the signature-based model clearly distinguishes healthy participants from the clinical groups,

**Table 3** Summary of the future mood prediction accuracy

	Bipolar		Borderline		Healthy	
	Accuracy	MAE	Accuracy	MAE	Accuracy	MAE
Anxious	82%	0.96	73%	1.17	98%	0.4
Elated	86%	0.75	78%	1.03	89%	0.57
Sad	84%	0.77	70%	1.16	93%	0.41
Angry	90%	0.60	70%	1.12	98%	0.30
Irritable	84%	0.84	70%	1.15	97%	0.39
Energetic	82%	0.90	75%	1.00	89%	0.69

obtaining an accuracy of 84% in classifying BD participants and HC, and 93% in classifying BPD participants and HC.

**Predicting the mood of an individual participant**

Using 20 consecutive observations, the signature-based model correctly predicted the subsequent mood score in healthy participants with 89–98% accuracy. In bipolar participants the mood was correctly predicted 82–90% of the time while in borderline participants this was 70–78% of the time (see Table 3). Moreover, the differences in the accuracy of predictions across the three clinical groups highlight a clear separation between the diagnostic



groups, suggesting that the three clinical groups have intrinsically different predictive relationship between reported and future mood state.

We compared our model to the naive benchmark that predicts that next day's score will be equal to the last day's score. In this case, the mood of healthy participants was correctly predicted 61–92% of the time, while the mood of bipolar and borderline participants was correctly predicted 46–67% and 44–62% of the time. Our model clearly outperformed this benchmark.

We then explored the possibility of predicting the change of mood over different time horizons. The results are shown in Fig. 3 (bottom). The predictability of the mood changes decay when the time horizon is increased. The most significant decay in predictability was found in BPD participants, consistent with the unpredictability that characterises BPD.

## Discussion

This analysis uses the signature feature set combined with random forests to study self-reported mood data. The signature-based model, with its ability to ignore parameterisation and provide canonical low dimensional sets of features, proved to be effective in distinguishing the three participant groups on the basis of their self-reported mood and was more accurate in doing so than previously used metrics, specifically mean mood scores. Furthermore, the method showed a clear overlap between the HC and BD groups suggestive of a normalisation of mood structure in some BD participants for some of the time; this confirms clinical experience. An alternative explanation could be that the six self-reported mood categories were not sensitive to the abnormal mood states associated with bipolar disorder.

By contrast there is a very clear differentiation between HC and BPD. This finding supports the concept of BPD as a distinct disorder. The distinctness from the healthy controls makes it more difficult to conceptualise BPD as an extreme variation of normal personality. There was some overlap between BD and BPD, but also marked differences. Overall the signatures of mood in the BPD group do appear to be distinct from both health and bipolar groups. The findings have implications for diagnostic practice where rates of misdiagnosis of BPD and BD are significant<sup>22,23</sup>. There is previous work published on machine learning approaches to distinguishing BD and BPD from healthy controls. For instance, some publications<sup>22,23</sup> report that MRI data can be used to classify individuals with borderline personality disorder and healthy controls, obtaining an accuracy of 80%<sup>24</sup> and 93.55%<sup>25</sup>, although the latter percentage should be taken with caution due to the small sample size. Neuroimaging has also been used to detect bipolar<sup>26</sup> disorder with an accuracy of 59%, as well as other sources of data such as

verbal fluency<sup>27</sup> with an accuracy of 79%. Therefore our approach has improved existing results, although comparison across the different papers is difficult due to the different datasets that were used.

The ability of the signature methods to predict future mood on the basis of the 20 preceding observations is significantly better than the used benchmark and again highlights the way in which the three groups can clearly be distinguished on the basis of the predictability of their mood states. The ability to predict future mood states is a key component of proactive self-management in BD and may be the key to preventing major depressive or manic relapse. While the predictability of mood in BPD was less accurate it was still in the order of 70%–78%. The unpredictable nature of mood in BPD is a characteristic feature of the disorder but the signature approach indicates that this lived experience may be more computationally predictable than has previously been thought. Further analyses and larger data sets, as well as exploration of different frequencies of measurement are required to explore whether the predictive accuracy of future mood states is maintained for more distant time points.

Our results suggest that the second-order signature is the most effective one given the available data. The signature of second-order is able to capture second-order effects across pairs of moods in a meaningful and concise way. If the same effect wanted to be obtained using the raw parameterised seven-dimensional path, we would have an initial feature set of more than 10,000 dimensions—and to train usefully would have required a much larger data set than our 733 instances. It may also be informative to enhance the mood data with automatically generated data derived from mobile phones, for example, voice<sup>28</sup> or geolocation<sup>29</sup>.

While it requires replication in other cohorts, and specifically in cohorts recruited independent of diagnosis, the signature method offers an exciting approach to the analysis of mood data. Despite the promise of time-stamped longitudinal symptom data the analytic challenge has been considerable and to date there is little consensus as to which approach to use. The signature method adds to the toolkit of analytic approaches and is particularly well suited to complex times series mood ratings because of its ability to capture order and deal with missing data.

## Limitations

The study population was derived from patient samples but did not include those who were acutely unwell or who had significant comorbidities and as such may have represented a more stable sub-population of patients. We had relatively little contact with patients during the study and only a baseline clinical assessment was carried out such that the validity of the self-reported mood scores cannot be certain. However, we have no evidence to suggest that the reports are inaccurate and

they are likely to be a good representation of the patients lived experience. This is supported by the fact that, as shown in this paper, the self-reported scores were significant enough to meaningfully distinguish the three groups. We also kept no ongoing record of medication changes through the duration of the study.

Our data set consisted of 733 examples. The data set was quite small for machine learning purposes and so we avoided testing different models in order to avoid overfitting. A bigger data set would undoubtedly allow more flexibility.

## Disclaimer

The views expressed in this publication are those of the authors and not necessarily those of the NIHR, the Medical Research Council, the National Health Service or the Department of Health.

## Acknowledgements

KE.A.S. was funded by a Wellcome Trust Strategic Award (CONBRIO: Collaborative Oxford Network for Bipolar Research to Improve Outcomes, Reference number 102616/Z) and supported by the NIHR Oxford Health Biomedical Research Centre. Terry Lyons is supported by the Alan Turing Institute under the EPSRC grant EP/N510129/1, and Imanol Perez is supported by a scholarship from La Caixa. Guy Goodwin and John Geddes are NIHR Senior Investigators.

## Author details

<sup>1</sup>Mathematical Institute, University of Oxford, Oxford, UK. <sup>2</sup>Department of Psychiatry, University of Oxford, Oxford, UK. <sup>3</sup>Oxford Health NHS Foundation Trust, Oxford, UK. <sup>4</sup>NIHR Oxford Health Biomedical Research Centre, Oxford, UK. <sup>5</sup>Alan Turing Institute, London, UK

## Conflict of interest

G.M.G. is past president of ECNP, holds a grant from the Wellcome Trust, holds shares in P1vital and has served as consultant, advisor or CME speaker for Allergan, Angelini, Compass pathways, MSD, Lundbeck (Votuska or Takeda), Medscape, Minervra, P1vital, Pfizer, Servier, Shire, Sun Pharma. J.R.G. receives grant funding from the National Institute of Health Research, Stanley Medical Research Institute, Wellcome Trust and Medical Research Council. The other authors declare that they have no conflict of interest.

## Publisher's note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Received: 25 April 2018 Revised: 18 July 2018 Accepted: 7 September 2018  
Published online: 13 December 2018

## References

- Harrison, P. J., Geddes, J. R. & Tunbridge, E. M. The emerging neurobiology of bipolar disorder. *Trends in Neurosciences* **41**, 18–30 (2017).
- Lyons, T. J., Caruana, M. & Lévy, T. (eds Morel, J.-M., Takens, F., Teissier, B.) in *Differential Equations Driven by Rough Paths*. pp. 81–93 (Springer, Berlin Heidelberg, 2007).
- Graham, B. Sparse arrays of signatures for online character recognition. *arXiv preprint arXiv:1308.0371* (2013).
- Xie, Z., Sun, Z., Jin, L., Ni, H. & Lyons, T. Learning spatial-semantic context with fully convolutional recurrent network for online handwritten Chinese text recognition. *IEEE Trans. Pattern Anal. Mach. Intell.* **40**, 1903–1917 (2017).
- Yang, W., Jin, L., Tao, D., Xie, Z. & Feng, Z. DropSample: A new training method to enhance deep convolutional neural networks for large-scale unconstrained handwritten Chinese character recognition. *Pattern Recognit.* **58**, 190–203 (2016).
- Yang, W., Jin, L. & Liu, M. Deepwriterid: An end-to-end online text-independent writer identification system. *IEEE Intell. Syst.* **31**, 45–53 (2016).
- Li, C., Zhang, X. & Jin, L. LPSNet: A Novel Log Path Signature Feature based Hand Gesture Recognition Framework. In *Proc. IEEE International Conference on Computer Vision Workshops (ICCVW)*, pp. 631–639 (Venice Italy, 2017).
- Lai, S., Jin, L. & Yang, W. Online signature verification using recurrent neural network and length-normalized path signature. *arXiv preprint arXiv:1705.06849* (2017).
- Liu, M., Jin, L. & Xie, Z. PS-LSTM: Capturing essential sequential online information with path signature and LSTM for writer identification. 2017 14th IAPR International Conference on Document Analysis and Recognition (ICDAR), Kyoto, 2017, pp. 664–669.
- Yang, W., et al. Leveraging the path signature for skeleton-based human action recognition. *arXiv preprint arXiv:1707.03993* (2017).
- Yang, W., Jin, L., Hao Ni and Lyons, T. Rotation-free online handwritten character recognition using dyadicpath signature features, hanging normalization, and deep neural network, 2016 23rd International Conference on Pattern Recognition (ICPR), Cancun, 2016, pp. 4083–4088.
- Ding, H. et al., "A Compact CNN-DBLSTM Based Character Model for Offline Handwriting Recognition with Tucker Decomposition," 2017 14th IAPR International Conference on Document Analysis and Recognition (ICDAR), Kyoto, 2017, pp. 507–512. <https://doi.org/10.1109/ICDAR.2017.89>.
- Coid, J., Yang, M., Tyrer, P., Roberts, A. & Ullrich, S. Prevalence and correlates of personality disorder in Great Britain. *Br. J. Psychiatry* **188**, 423–431 (2006).
- Grant, B. F. et al. Prevalence, correlates, disability, and comorbidity of DSM-IV borderline personality disorder: results from the Wave 2 National Epidemiologic Survey on Alcohol and Related Conditions. *J. Clin. Psychiatry* **69**, 533 (2008).
- Merikangas, K. R. et al. Lifetime and 12-month prevalence of bipolar spectrum disorder in the National Comorbidity Survey replication. *Arch. Gen. Psychiatry* **64**, 543–552 (2007).
- Ruggero, C. J., Zimmerman, M., Chelminski, I. & Young, D. Borderline personality disorder and the misdiagnosis of bipolar disorder. *J. Psychiatr. Res.* **44**, 405–408 (2010).
- National Collaborating Centre for Mental Health. *Borderline Personality Disorder: Treatment and Management*. (British Psychological Society, UK, 2009).
- National Collaborating Centre for Mental Health. *Bipolar Disorder: The Assessment and Management of Bipolar Disorder in Adults, Children and Young People in Primary and Secondary Care*. (The British Psychological Society, Leicester UK, 2014).
- Tsanas, A. et al. Daily longitudinal self-monitoring of mood variability in bipolar disorder and borderline personality disorder. *J. Affect Disord.* **205**, 225–233 (2016).
- Tsanas, A. et al. Daily longitudinal self-monitoring of mood variability in bipolar disorder and borderline personality disorder. *J. Affect Disord.* **205**, 225–233 (2016).
- Breiman, L. Random forests. *Mach. Learn.* **45**, 5–32 (2001).
- Ruggero, C. J., Zimmerman, M., Chelminski, I. & Young, D. Borderline personality disorder and the misdiagnosis of bipolar disorder. *J. Psychiatr. Res.* **44**, 405–408 (2010).
- Zimmerman, M., Ruggero, C. J., Chelminski, I. & Young, D. Psychiatric diagnoses in patients previously overdiagnosed with bipolar disorder. *J. Clin. Psychiatry* **71**, 26–31 (2010).
- Sato, J. R. et al. Can neuroimaging be used as a support to diagnosis of borderline personality disorder? An approach based on computational neuroanatomy and machine learning. *J. Psychiatr. Res.* **46**, 1126–1132 (2012).
- Xu, T., Cullen, K. R., Houri, A., Lim, K. O., Schulz, S. C. and Parhi, K. K. "Classification of borderline personality disorder based on spectral power of resting-state fMRI," 2014 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, Chicago, IL, 2014, pp. 5036–5039.
- Schnack, H. G. et al. Can structural MRI aid in clinical classification? A machine learning study in two independent samples of patients with schizophrenia, bipolar disorder and healthy subjects. *Neuroimage* **84**, 299–306 (2014).
- Costafreda, S. G. et al. Pattern of neural responses to verbal fluency shows diagnostic specificity for schizophrenia and bipolar disorder. *BMC Psychiatry* **11**, 18 (2011).
- Faurholt-Jepsen, M. et al. Voice analysis as an objective state marker in bipolar disorder. *Transl. Psychiatry* **6**, e856 (2016).
- Palmius, N. et al. Detecting bipolar depression from geographic location data. *IEEE Trans. Biomed. Eng.* **64**, 1761–1771 (2017).