

# Predicting Dropout From Cognitive Behavioral Therapy for Panic Disorder Using Machine Learning Algorithms

Sei Ogawa

## Abstract

**Background:** Attrition is an important problem in clinical practice and research. However, the predictors of dropping out from cognitive behavioral therapy (CBT) for panic disorder (PD) are not fully understood. In this study, we aimed to build a dropout prediction model for CBT for PD using machine learning (ML) algorithms.

**Methods:** We treated 208 patients with PD applying group CBT. From baseline data, the prediction analysis was carried out using two ML algorithms, random forest and light gradient boosting machine. The baseline data included five personality dimensions in NEO Five Factor Index, depression subscale of Symptom Checklist-90 Revised, age, sex, and Panic Disorder Severity Scale.

**Results:** Random forest identified dropout during CBT for PD showing that the accuracy of prediction was 88%. Light gradient boosting machine showed that the accuracy was 85%.

**Conclusions:** The ML algorithms could detect dropout after CBT for PD with relatively high accuracy. For the purpose of clinical decision-making, we could use this ML method. This study was conducted as a naturalistic study in a routine clinical setting. Therefore, our results in ML approach could be generalized to regular clinical settings.

**Keywords:** Panic disorder; Cognitive behavioral therapy; Machine learning; Predictor

## Introduction

Panic disorder (PD) is a common mental disorder marked by frequent and unexpected panic attacks. It has been demonstrated that cognitive behavioral therapy (CBT) is effective in treating PD [1].

In clinical practice and research, dropout during intervention is an important problem. Particularly, in the course of CBT

for PD, a lot of participants leave treatment prematurely [2]. Regarding CBT for PD, personality traits have been suggested as a predictor of treatment efficacy [3]. However, the predictors of dropping out from this intervention are not fully understood, with the exception of pre-treatment depression [2].

In recent years, machine learning (ML) is one of the areas of artificial intelligence that has been attracting attention. ML is the most widely used technique in fields ranging from medicine or psychology to business. ML is a new approach to predictive research in which data are trained by computer algorithms to build models. ML makes it possible to predict outcome with higher accuracy than traditional statistical methods [4]. In this study, we used ML to build a dropout prediction model for CBT for PD.

## Materials and Methods

### Participants

Two hundred eight Japanese patients with PD attended the group CBT program from 2001 to 2017 in our institute. Entry criteria were principal axis I diagnosis of PD according to the Diagnosis and Statistical Manual of Mental Disorders (fourth edition) [5], as evaluated by using the Structured Clinical Interview for the Diagnosis and Statistical Manual of Mental Disorders (fourth edition) [6]. Exclusion criteria for this study were current psychosis, bipolar disorder and substance use disorder. Nineteen participants dropped out prematurely from the group CBT program.

The patients provided their written informed consent after receiving a full explanation of the purpose and procedures of the study. The study was performed in accordance with the Declaration of Helsinki and the study's protocol was approved by the Ethics Committee of our institute. The clinical study registration number in the Japanese trials registry is UMIN CTR 000031147.

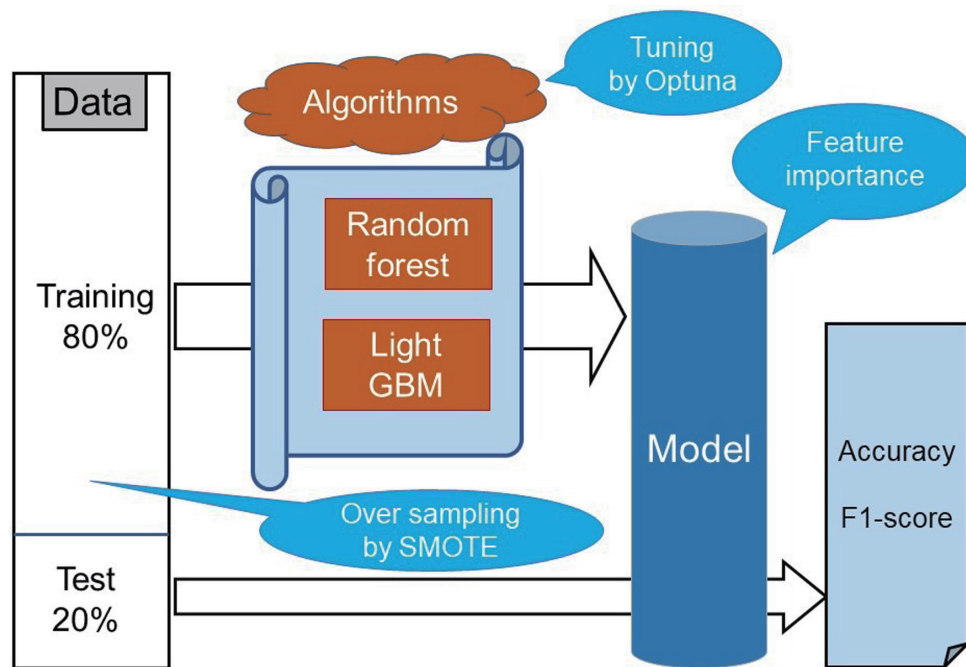
### Treatments

Our CBT program for PD was based on the treatment manual developed by Andrews et al [7]. The program was composed of psychoeducation, breath retraining, cognitive restructuring, and *in vivo* graded exposures. Two trained therapists conducted 2-h session once a week for about 10 weeks. A total of 14 therapists performed CBT.

Manuscript submitted April 2, 2024, accepted April 29, 2024  
Published online May 29, 2024

Graduate School of Humanities and Social Sciences, Nagoya City University,  
Nagoya 467-8501, Japan. Email: seiogawa1964@nifty.com

doi: <https://doi.org/10.14740/jocmr5167>



**Figure 1.** Process flow diagram for predictive models. LightGBM: light gradient boosting machine; SMOTE: synthetic minority oversampling technique.

## Measurements

At baseline, all the patients were assessed with the NEO Five Factor Index (NEO-FFI), Symptom Checklist-90 Revised (SCL-90-R) and Panic Disorder Severity Scale (PDSS).

The NEO-FFI is a 60-item self-reported instrument to measure the five personality dimensions of neuroticism, extraversion, conscientiousness, openness, and agreeableness [8]. Good reliability and validity of the Japanese version have been demonstrated [9].

The SCL-90-R is a universally used measure for general psychopathology. It is subdivided into 10 subscales (e.g. depression, anxiety) [10]. The reliability and validity of the Japanese version have been demonstrated [11]. In this study, we used the depression subscale of this scale as a candidate predictor.

The PDSS is a seven-item scale for assessing overall severity of PD in which the clinician rates the severity of PD [12]. The reliability and validity have been demonstrated for the original and the Japanese versions [13].

## Analysis

All the data were examined using Python 3.9 with scikit-learn. The statistical tests were two-tailed and we considered an alpha value of less than 0.05 to be statistically significant (Fig. 1).

As main candidate predicting variables, we used five personality dimensions in NEO-FFI. Furthermore, based on previous studies, we used age, sex, onset, the baseline score of PDSS and the depression subscale of SCL-90-R as candidate predictors.

Our prediction analysis was carried out using random forest and light gradient boosting machine (LightGBM). Random forest is a supervised ML method. This method developed several decision tree classifiers in parallel on different dataset sub-samples through bootstrap aggregation (bagging) [14]. LightGBM is an improved version of gradient boosting machine that is a supervised ML method and an ensemble classifier. LightGBM needs long running time; however, this method can potentially handle a large amount of data and attain high accuracy levels [15].

We split the dataset in an 80% training data and a 20% test data to address the risk of overfitting. Above two ML models were trained with the training data using k-fold cross-validation to ensure its generalizability.

Treatment dropout predictions are usually based on imbalanced datasets with rare class being the one of interest. Therefore, our analyses were run with synthetic minority oversampling technique (SMOTE). SMOTE is a hybrid method that equalizes class imbalance in training data. It creates new examples for each rare training observation using a nearest neighbors modeling approach [16].

To improve the predictive performance of the ML models and avoid overfitting, Optuna, which is a Bayesian optimization library in Python, was used to benchmark its performance empirically and modify the hyperparameters.

We evaluated the model performance using accuracy and F1-score.

True-positive (TP), true-negative (TN), false-positive (FP), and false-negative (FN) values were calculated using the following formula into two indexes (accuracy and F1-score) to evaluate model performance. When TPs and TNs are more significant, accuracy is employed; when FNs and FPs are critical, F1-score is employed.

**Table 1.** Demographics and Baseline Characteristics

Characteristics	Completer (n = 189)	Dropout (n = 19)	P
Sex (% female)	67.2	63.6	> 0.05
Mean age	36.3 (10.9)	34.3 (12.1)	> 0.05
Onset	29.1 (10.2)	29.1 (11.8)	> 0.05
NEO-FFI			
Neuroticism	26.7 (9.0)	27.5 (7.1)	> 0.05
Extraversion	25.9 (8.1)	27.0 (8.0)	> 0.05
Openness	28.3 (6.1)	27.9 (7.4)	> 0.05
Agreeableness	32.7 (6.8)	32.0 (6.2)	> 0.05
Conscientiousness	27.4 (7.7)	27.8 (8.1)	> 0.05
PDSS	13.1 (4.8)	12.7 (5.2)	> 0.05
SCL-90-R depression subscale	1.14 (0.8)	1.15 (1.0)	> 0.05

Values are presented as mean and standard deviation (SD). NEO-FFI: NEO Five Factor Index; PDSS: Panic Disorder Severity Scale; SCL-90-R: Symptom Checklist-90 Revised.

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{FP} + \text{TN} + \text{FN}}$$

$$\text{F1-score} = 2 \times \text{Precision} \times \frac{\text{Recall}}{\text{Precision} + \text{Recall}}$$

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}}, \text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}}$$

ML algorithms often seem like black boxes [17]. However, it is important to know which baseline variables are critical in predicting responders in CBT. Feature importance in ML shows which features are the best predictors. Feature importance is obtained by permuting features for each algorithm by determining which feature yields the largest change in prediction [17].

## Results

There was no statistically significant difference in the baseline characteristics between dropouts and completers (Table 1).

Figure 2 shows the predictive performance (accuracy and F1-score) of ML models for dropout prediction of CBT for PD. Random forest exhibited 0.88 accuracy rate for prediction and LightGBM showed 0.85. In F1-score, random forest exhibited 0.83 and LightGBM showed 0.78.

Figure 3 shows the feature importance of two ML models for prediction of CBT for PD. In random forest, NEO-FFI openness and PD severity were identified as major variables with relatively high weight. In LightGBM, NEO-FFI neuroticism and depression were shown to be significant factors that carried a comparatively large weight.

## Conclusion

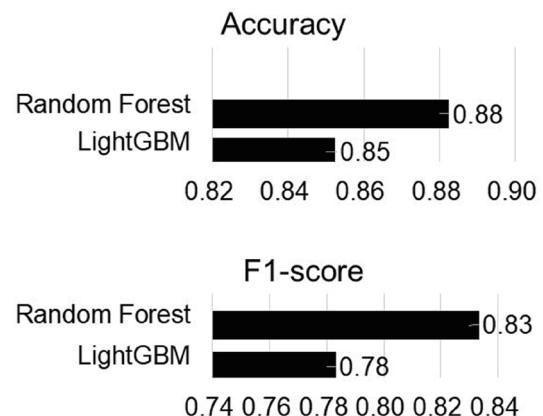
The present study showed that the ML algorithms could detect

dropout after CBT for PD. We implemented oversampling using SMOTE as a measure against imbalanced data and used Optuna for modifying the hyperparameters in order to obtain a relatively high accuracy rate and F1-score.

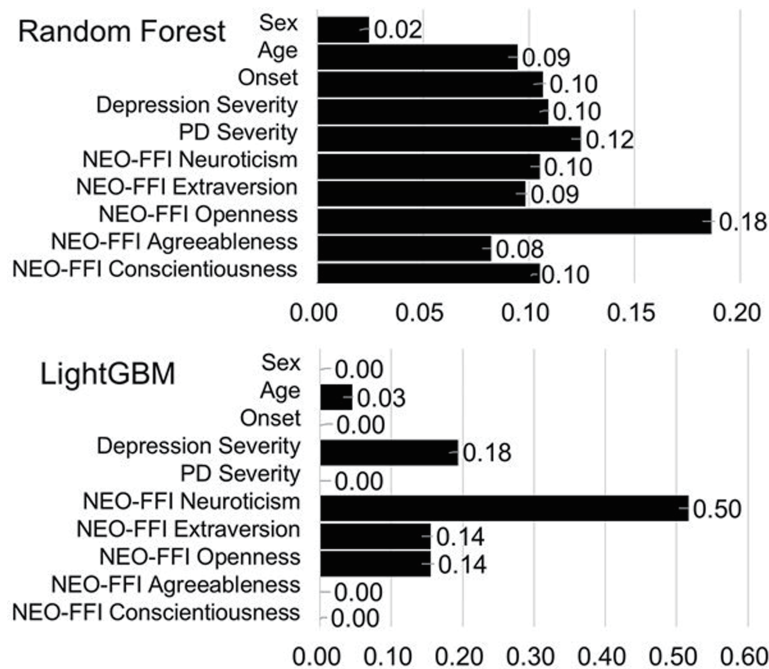
In these two ML algorithms, personal traits were identified major feature importance in predicting model. This result showed to be consistent with previous study concerning predicting efficacy in CBT.

The benefit of using ML is that this approach can show clear decision to predict attrition. For the purpose of clinical decision-making, we could use this ML method.

The present study has several limitations. First, data were collected from 2001 to 2017 and regarded to be somewhat old. Second, one of the previous studies suggested that more than 300 cases should be contained in training data for using resampling methods, therefore, our sample size might be relatively small. Third, we used only two ML algorithms, random forest and LightGBM, which are used in a lot of previous studies and several ML competitions. However, higher accuracy or F1-score could have been achieved using other ML algorithms



**Figure 2.** The predictive performance of machine learning models. LightGBM: light gradient boosting machine.



**Figure 3.** The feature importance of machine learning models. LightGBM: light gradient boosting machine; NEO-FFI: NEO Five Factor Index; PD: panic disorder.

like support vector machine.

This study was conducted as a naturalistic study in a routine clinical setting. Therefore, the demographic and clinical characteristics of this cohort did not differ much from those of general population with PD. Furthermore, we used the data that can be easily collected in clinical practice. Our results in ML approach could be generalized to regular clinical settings.

## Acknowledgments

I would like to thank the researchers of the Nagoya City University Graduate School of Medical Sciences for their assistance in data collection.

## Financial Disclosure

This study was supported by a Grant-in-Aid for Scientific Research from the Ministry of Health, Labour and Welfare, Japan.

## Conflict of Interest

There is no conflict of interest.

## Informed Consent

Informed consent was obtained from all participants included in this study.

## Author Contributions

SO was the primary investigator and engaged in intervention, analysing data using ML and writing this manuscript. Data collection was conducted in collaboration with the researchers of the Nagoya City University Graduate School of Medical Sciences.

## Data Availability

The data in this study are not publicly available. This study was started in 2001, and in accordance with the protocol at that time, the study participants were informed that the data used in this study would not be provided to third parties other than the researchers and the participants.

## References

- Otto MW, Deveney C. Cognitive-behavioral therapy and the treatment of panic disorder: efficacy and strategies. *J Clin Psychiatry*. 2005;66(Suppl 4):28-32. [pubmed](#)
- Belanger C, Courchesne C, Leduc AG, Dugal C, El-Baalbaki G, Marchand A, Godbout N, et al. Predictors of dropout from cognitive-behavioral group treatment for panic disorder with agoraphobia: an exploratory study. *Behav Modif*. 2017;41(1):113-140. [doi](#) [pubmed](#)
- Ogawa S, Kondo M, Ino K, Imai R, Ii T, Furukawa TA, Akechi T. Predictors of broad dimensions of psychopathology among patients with panic disorder after cognitive-

- behavioral therapy. *Psychiatry J.* 2018;2018:5183834. [doi pubmed pmc](#)
4. Vieira S, Liang X, Guiomar R, Mechelli A. Can we predict who will benefit from cognitive-behavioural therapy? A systematic review and meta-analysis of machine learning studies. *Clin Psychol Rev.* 2022;97:102193. [doi pubmed](#)
  5. American Psychiatric Association. Task force on DSM-IV: diagnostic and statistical manual of mental disorders: DSM-IV. 4th ed. Washington, DC, American Psychiatric Association; 1994.
  6. First MB. Structured clinical interview for DSM-IV axis I disorders: SCID - I : clinician version: administration booklet. Washington, D.C., American Psychiatric Press; 1997.
  7. Andrews G, Creamer M, Crino R, Hunt C, Lampe L, Page A. The treatment of anxiety disorders: clinician guides and patient manuals. 2nd ed. Cambridge, UK; New York, NY, Cambridge University Press; 2003.
  8. Costa PT, McCrae RR. Manual for the revised NEO personality inventory (NEO-PI-R) and NEO five-factor inventory (NEO-FFI). *Psychological Assessment Resources.* 1992.
  9. Yoshimura K, Ono Y, Nakamura K, Nathan JH, Suzuki K. Validation of the Japanese version of the NEO Five-Factor Inventory in a large community sample. *Psychol Rep.* 2001;88(2):443-449. [doi pubmed](#)
  10. Derogatis LR. SCL-90-R: administration, scoring & procedures manual-II for the (revised) version and other instruments of the psychopathology rating scale series. *Clin Psychomet Res.* 1992:1-16.
  11. Furukawa TA, Nakanishi M, Sakurai A, Suzuki A, Suzuki-Moor A, Hamanaka T. Effects of rthyl loflazepate in mood and neurosis-related disorders (ICD-10 JCM); Changes in SCL-90-R subscale scores. *Rinsho Seisinigaku (Clinical Psychiatry).* 1996;25:233-240.
  12. Shear MK, Brown TA, Barlow DH, Money R, Sholomskas DE, Woods SW, Gorman JM, et al. Multicenter collaborative panic disorder severity scale. *Am J Psychiatry.* 1997;154(11):1571-1575. [doi pubmed](#)
  13. Yamamoto I, Nakano Y, Watanabe N, Noda Y, Furukawa TA, Kanai T, Takashio O, et al. Cross-cultural evaluation of the Panic Disorder Severity Scale in Japan. *Depress Anxiety.* 2004;20(1):17-22. [doi pubmed](#)
  14. Sarker IH. Machine learning: algorithms, real-world applications and research directions. *SN Comput Sci.* 2021;2(3):160. [doi pubmed pmc](#)
  15. Ahamad MM, Aktar S, Uddin MJ, Rahman T, Alyami SA, Al-Ashhab S, Akhdar HF, et al. Early-stage detection of ovarian cancer based on clinical data using machine learning approaches. *J Pers Med.* 2022;12(8):1211. [doi pubmed pmc](#)
  16. Nakamura M, Kajiwara Y, Otsuka A, Kimura H. LVQ-SMOTE - learning vector quantization based synthetic minority over-sampling technique for biomedical data. *BioData Min.* 2013;6(1):16. [doi pubmed pmc](#)
  17. Musolf AM, Holzinger ER, Malley JD, Bailey-Wilson JE. What makes a good prediction? Feature importance and beginning to open the black box of machine learning in genetics. *Hum Genet.* 2022;141(9):1515-1528. [doi pubmed pmc](#)