Comprehensive Psychiatry 151 (2026) 152737

Contents lists available at ScienceDirect

Comprehensive Psychiatry

journal homepage: www.elsevier.com/locate/comppsych

Predictors of quality of life in OCD: Machine-derived features beyond expert consensus☆

Max Thorsson[a,*], Angela Trujillo[b], Maritza J. Romero[b], Jonathan Irreño-Sotomonte[c], Tania Borda[d,e], Raquel M. Zepeda-Burgos[f], Mayra C. Martinez Mallen[g], Beatriz Moyano[h,i,j], Jacey L. Anderberg[k,l], Josselyn S. Muñoz[k,m], Hannah C. Moore[k], Renee M. Frederick[k,n], Dayan Berrones[k], Vanessa Zavala Cruz[k], Ogechi Onyeka[k], Andrew D. Wiese[k], Kevin M. Wagner[k], Latin American Trans-ancestry INitiative for OCD genomics (LATINO), Brazilian Obsessive-Compulsive Spectrum Disorder Working Group (GTTOC), James J. Crowley[o], Eric A. Storch[k], Matti Cervin[k,a]

[a] Department of Clinical Sciences, Lund University, Lund, Sweden
[b] Facultad de Ciencias del Comportamiento, Universidad de La Sabana, Chía, Cundinamarca, Colombia
[c] Universidad del Rosario Clínica de Nuestra Señora de la Paz, Bogotá, Cundinamarca, Colombia
[d] Instituto Realize, Buenos Aires, Argentina
[e] Facultad de Psicologia, Universidad Católica Argentina, Buenos Aires, Argentina
[f] Centro Universitario de Neurociencia, Universidad Francisco Gavidia, San Salvador, El Salvador
[g] Centro Neurológico, Centro Médico ABC, Ciudad de México, Ciudad de México, Mexico
[h] Centro Interdisciplinario de Tourette, TOC, TDAH y Trastornos Asociados (CITA), Buenos Aires, Argentina
[i] Departamento de Investigación, Asociación de Psiquiatras de Argentina (APSA), Buenos Aires, Argentina
[j] Docencia, Universidad Favaloro, Buenos Aires, Argentina
[k] Psychiatry and Behavioral Sciences, Baylor College of Medicine, Houston, TX, USA
[l] Department of Psychology, University of Nebraska-Lincoln, Lincoln, NE, USA
[m] John Sealy School of Medicine, University of Texas Medical Branch, Galveston, TX, USA
[n] Department of Psychology, University of Houston, Houston, TX, USA
[o] Department of Genetics, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA

ARTICLE INFO

Keywords:
Obsessive–compulsive disorder
Quality of life
Machine learning
Depression

ABSTRACT

Background: Quality of life (QoL) varies substantially among individuals with obsessive–compulsive disorder (OCD), but the clinical and contextual factors most strongly associated with QoL remain unclear. A better understanding of these factors may improve assessment, treatment planning, and service organization.
Methods: 3152 adults with lifetime OCD completed a validated measure of QoL along with clinical and sociodemographic measures. A 22-feature expert-informed model, based on prior scientific and clinical knowledge, was compared with data-driven models using recursive feature elimination and gradient-boosted trees. Model performance was evaluated using 5-fold cross-validation.
Results: Data-driven models modestly outperformed the expert-informed model (RMSE 7.23 vs. 7.56; R^2 0.55 vs. 0.51). Although several predictors overlapped, machine-derived models identified predictors not selected by experts, including subjective social rank, anhedonia, and inequality-adjusted national human development. Across models, depression severity, subjective social rank, and anxiety severity were the strongest predictors of QoL. Reduced data-driven models using 12–18 features retained comparable predictive performance and significantly outperformed the expert model.
Conclusions: Expert knowledge captured much of the variance in QoL in OCD, but data-driven approaches improved predictive accuracy and identified novel factors. QoL was most strongly predicted by depression severity, anxiety severity, and social context rather than OCD severity alone, suggesting that clinical assessment

☆ This article is part of a Special issue entitled: 'Obsessive-compulsive and related disorders Section' published in Comprehensive Psychiatry.

* Corresponding author at: Department of Clinical Sciences, Lund Child and Adolescent Psychiatry, Sofiavägen 2D, SE-22241 Lund, Sweden.
E-mail address: max.thorsson@med.lu.se (M. Thorsson).

https://doi.org/10.1016/j.comppsych.2026.152737
Received 8 June 2026; Received in revised form 8 July 2026; Accepted 24 July 2026
Available online 28 July 2026
0010-440X/© 2026 The Author(s). Published by Elsevier Inc. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

and treatment planning should routinely consider broader affective and psychosocial difficulties alongside OCD symptoms.

1. Introduction

Obsessive-compulsive disorder (OCD) is a mental disorder characterized by the presence of obsessions and/or compulsions that are time-consuming and/or impairing [1,2]. It often onsets prior to adulthood [3] and tends to run a chronic or relapsing course in the absence of effective treatment [4–6]. Co-occurring mental disorders are the norm, affecting an estimated 60–90% of those with OCD, with depression and anxiety disorders being most common [7,8]. Moreover, obsessions and compulsions are often deeply entangled in everyday activities, such as leaving the house, going to the bathroom, or interacting with loved ones, making OCD a disorder that can result in substantial negative effects on quality of life (QoL).

QoL refers to a person's subjective evaluation of their position in life, weighed against personal goals, expectations, and cultural context, and it spans physical health, psychological states, social relationships, and daily activities [9]. Meta-analyses consistently show that individuals with OCD report poorer QoL than the general population. Among adults, especially marked deficits are seen in social, family, and emotional domains [10,11]. In children and adolescents, overall quality of life tends to be substantially lower among those with OCD, with additional impairments observed in school and social domains [12].

A variety of clinical and sociodemographic factors have been linked to reduced QoL in OCD. Naturally, a primary factor of interest has been OCD symptom severity. As expected, greater OCD severity has been linked to lower happiness and life satisfaction among individuals with OCD [13–15]. Other OCD-related factors have also been linked to poor QoL, including poor insight into symptoms, although longitudinal studies suggest that changes in insight do not necessarily translate into changes in QoL, pointing to that other factors may be involved [14]. Some studies have also shown that stigma may be linked to lower QoL in OCD [13].

Co-occurring mental disorders, especially depression and anxiety, appear to exacerbate the burden of OCD, with depressive symptoms showing a particularly strong negative effect on emotional, social, and physical aspects of QoL [15,16]. This indicates that depressive symptoms, alongside OCD, can be an important target in treatment to improve overall well-being [17]. Other factors, including sexual dysfunction [18], loneliness [15,19] and having experienced stressful life events [20] have likewise been linked to poorer QoL among individuals with OCD. While symptom remission following treatment is associated with statistically significant gains in QoL, partial treatment response often falls short of achieving similar improvements [21–23]. Moreover, substantially reduced QoL has also been observed among individuals in remission from OCD, suggesting that OCD symptoms alone do not fully account for the poor QoL seen in this population [24]. A similar finding emerged in a study on pediatric OCD, where QoL at baseline was more strongly associated with depressive symptoms and the overall burden of psychiatric symptoms than with OCD severity [25].

Most prior studies into QoL in OCD have either compared individuals with OCD to the general population, investigated different factors in isolation, or included only a limited set of features, leaving important questions unanswered about how OCD severity, co-occurring mental disorders, and a range of psychosocial and demographic factors jointly shape QoL in OCD. Given the likely interplay of these factors, there is a need for approaches capable of modeling complex, multidimensional relations using large and rich datasets.

Traditional statistical methods such as linear regression are useful for examining associations between variables under assumptions of linearity, additivity, and limited interaction. However, they may miss complex, nonlinear relationships and feature interactions that are common in real-world clinical data [26]. In contrast, machine learning methods, particularly tree-based approaches such as decision trees, random forests, and gradient boosting, are better suited to detect nonlinear patterns and interactions, and may therefore be more appropriate for this dataset and research question [27–29]. However, these models are often considered "black-box" methods because their predictions are not directly transparent.

To address this limitation, we used SHAP (Shapley Additive Explanations), which quantifies each feature's contribution to an individual prediction and supports both local and global interpretation of model output [30–32]. SHAP also indicates the direction and magnitude of each feature's effect, helping identify which variables increase or decrease the predicted outcome. In addition, SHAP can aid feature selection by highlighting the most influential predictors, which is particularly useful in clinical applications where interpretability is as important as predictive accuracy. Among tree-based models, eXtreme Gradient Boosting (XGBoost) is a strong choice because it is a regularized, highly optimized form of gradient boosting that can model complex nonlinear effects while maintaining strong predictive performance [33–35]. In this study, XGBoost was therefore selected for its suitability for structured clinical data and its compatibility with SHAP-based interpretation.

Using both theory-informed and data-driven approaches supported by XGBoost and SHAP values, this study aims to identify the most important predictors of quality of life (QoL) in OCD. The study is guided by two research questions: to what extent can an expert, consensus-based model informed by clinicians and researchers predict QoL in individuals with OCD, and how does its predictive performance compare to that of a machine learning-based feature selection approach (**RQ1**)? In addition, what are the key predictors of QoL in individuals with OCD, and how do these predictors differ between models using expert-based versus machine learning-based feature selection methods (**RQ2**)?

2. Methods

2.1. The LATINO study

The sample consisted of 3152 adults with a lifetime history of OCD from the LATINO study, which is a large-scale, trans-ancestry initiative that have recruited over 5000 individuals with OCD or clinically relevant obsessive-compulsive symptoms across the United States and Latin America. Participants were assessed using a harmonized battery of diagnostic, clinical, and psychosocial instruments to make possible cross-site comparability and integrative analysis. Inclusion criteria, for the full study, include a primary diagnosis of OCD, age $\geq$ 7 years, and self-identified Latin American ancestry. Exclusion criteria include a history of moderate to severe acute psychosis or a primary diagnosis of a neurological disorder. All assessments are conducted by trained clinicians using validated instruments translated and culturally adapted into Spanish and Portuguese. The study combines rigorous clinical phenotyping with genomic sampling and provides one of the most comprehensive datasets available for studying OCD in Latin American populations. The study, described in detail in a published protocol [36], was approved by Baylor College of Medicine Institutional Review Board (IRB) for recruitment in the United States (U.S.-based institutions entered into a reliance agreement with Baylor College of Medicine) or the local ethics committee at each foreign site. Written and/or electronic informed consent was obtained from all participants in accordance with local IRB requirements, and, where allowed, participants were compensated up to $50 USD after completing core assessments and

providing a saliva sample.

2.2. Overview of features in the LATINO study

The LATINO study includes a rich set of features capturing clinical severity, treatment history, psychosocial functioning, diagnostic status, and contextual factors relevant to OCD. Symptom severity of OCD is assessed using the Yale-Brown Obsessive Compulsive Scale – Second Edition (Y-BOCS-II), with ratings for both current and worst-ever symptom levels, and separate scores for obsessions and compulsions [37]. Treatment-related data include age at first treatment, the number and types of interventions received (such as cognitive-behavioral therapy), and perceived accessibility of care. Participants also report on the effectiveness of treatment across different domains, including therapy, care teams, and overall improvement. The dataset further captures the impact of OCD on daily functioning across multiple life areas, including work, education, family, social relationships, and general well-being.

Psychosocial and quality of life measures are assessed using validated self-report instruments. These include standardized scales for depressive symptoms, specifically the Patient Health Questionnaire-9 (PHQ-9) [38], anxiety, assessed with the Generalized Anxiety Disorder 7-item scale (GAD-7) [39], and anhedonia, measured by the Snaith–Hamilton Pleasure Scale (SHAPS) [40]. In LATINO, quality of life is measured using the Quality of Life Enjoyment and Satisfaction Questionnaire (QLES-Q) [41], a self-report measure evaluating satisfaction across domains such as physical health, mood, work, household activities, social relationships, leisure, and general life. Higher scores indicate greater overall enjoyment and satisfaction. We modeled the total raw score from the adult questionnaire QLES-Q (scores range from 14 to 70) rather than the percentage of maximum, as this preserves linearity. Children's data were not analyzed in this study, as they comprised less than 10% of the sample ($n = 211$) and were assessed with a different QoL scale (P-QLES), resulting in substantial non-overlapping questionnaires and limited comparability with the main dataset.

LATINO also includes assessments of family accommodation, perceived stigma, and experiences of discrimination. Identity-related processes – such as affirmation, resolution, and exploration – are measured to understand how individuals integrate their mental health experiences into their broader sense of self.

Self-reported contextual and life-event variables were also included among the candidate predictors. These included perceived contributors to OCD onset or worsening, such as illness, trauma, stress, bullying, family history, and biological, psychological, or environmental explanations. Additional variables captured childbirth-related factors, major life transitions, perceived control, unmet goals, and interpersonal conflict. These variables were treated as self-reported attributions or impact ratings, not as objective causal factors.

Diagnostic information is derived from structured clinical interviews. In the LATINO dataset, diagnoses were established using culturally adapted structured interviews, including the Mini International Neuropsychiatric Interview (MINI) and the Diagnostic Interview for Anxiety, Mood, and OCD and Related Neuropsychiatric Disorders (DIAMOND) for obsessive-compulsive and related disorders. These tools assess a broad range of psychiatric conditions, and diagnoses are coded as present if they occurred either currently or at any point in the participant's lifetime. Current depressive and anxiety symptom burden was assessed separately using PHQ-9 and GAD-7. Past diagnoses were only available for OCD-related disorders and tic disorders and were included as positive diagnoses in the diagnostic count.

Demographic and contextual features provide essential background for interpreting clinical and psychosocial data. These include age, sex, race or ethnicity, country, education, perceived income, occupational status, relationship status, reproductive history, and subjective social status. To capture country-level socioeconomic development and contextual conditions in a continuous and interpretable way, we used IHDI values from the United Nations Development Programme Human Development Report 2025, where available [42]. Puerto Rico used an UNDP-based estimate of 0.70 [43]. Where IHDI was unavailable, as for Cuba, we assigned the corresponding UNDP human-development-group IHDI value of 0.64 [42].

All variables were collected during the same LATINO assessment but differed in temporal frame. Most symptom and functioning measures assessed current or recent experiences, including QoL, depressive symptoms, anxiety symptoms, and anhedonia, whereas other variables captured worst-ever OCD severity, lifetime diagnoses, treatment history, and contextual or historical factors.

Together, these features offer a comprehensive view of OCD within the LATINO population, enabling nuanced analyses of symptomatology, treatment response, and the influence of social and cultural context.

2.3. Consensus features for QoL in OCD

The expert-informed consensus model for this study was designed within the LATINO consortium, which consists of researchers and clinicians with experience of OCD. The task was to construct a clinically interpretable feature set that would capture key determinants of QoL while remaining brief enough for comparison with data-driven feature selection. An initial pool of candidate domains was compiled from prior literature and clinical expertise. These domains included OCD severity, depressive symptoms, anxiety symptoms, sociodemographic and socioeconomic factors, illness-history variables, comorbidity, treatment response, and country-level inequality-adjusted human development. Candidate variables were discussed by the expert panel, and overlapping, sparse, or redundant variables were removed or collapsed. Disagreements were resolved through group discussion, prioritizing variables that were clinically relevant, interpretable, and available across the full sample. The shorter conceptual list therefore reflects the candidate domains considered during expert selection, whereas the final expert-informed model consisted of 22 operationalized features. For example, socioeconomic status was represented by multiple indicators, including monthly income, years of education, and occupation. Occupation was operationalized as separate dummy-coded variables for each occupational category, whereas comorbidity was represented by the total number of mental disorders for which the participant met criteria. The final 22-feature model included the following variables:

- Age at assessment
- Age at OCD onset
- Biological sex
- Country IHDI
- Current occupation (code 1–8)
- Diagnostic count
- Family history of OCD
- Monthly income
- Overall treatment improvement
- Race
- Romantic relationship status
- Total score on GAD-7
- Total score on PHQ-9
- Years of education
- Y-BOCS-II total (current)

2.4. Data preprocessing

Initial data cleaning involved removing values that were outside the allowable response range for each questionnaire item, as defined by the original instruments. Total scores were subsequently calculated after imputing missing values with each participant's own median response on that instrument, providing a conservative and individualized estimate of total symptom burden, and avoiding data leakage. Features derived from multiple-response items, where participants could select more than one option, were treated as separate binary features to preserve the full variance in responses. Features not included in instrument scores were imputed using the sample median, computed within the training folds to prevent leakage during model evaluation.

A final screening step was applied to remove features that were either (a) purely random (e.g., participant ID, row number), (b) non-interpretable or potentially misleading (e.g., clinical site identifiers), or (c) subcomponents of constructs that could introduce ambiguity. This

process yielded a final set of 94 features for processing with machine learning.

2.5. Data splitting

The dataset was split into 5 folds for nested cross-validation. In each iteration, model training and hyperparameter tuning were performed exclusively on 4 folds (the inner training set) while the remaining fold (the outer test set) was held out for unbiased evaluation. This procedure was repeated five times, so that each fold (containing 630–631 participants) served exactly once as the test set. This nested cross-validation framework provides a robust and impartial estimate of model generalizability across the entire dataset [44].

2.6. Model optimization

Although training a machine learning model (here, focusing on gradient boosted trees) can yield strong predictive performance, the choice of hyperparameters can affect final results. To optimize these parameters in a robust and reproducible way, we used the optimization framework Optuna [45] with a 5-fold cross-validation scheme and a custom objective function. The framework allowed for using the tree-structured Parzen estimator (TPE) as a sampling method for parameters. TPE sampling is considered superior to random search, due to its ability to model the distribution of good-performing parameters and prioritize promising regions of the search space.

The training pipeline was developed based on established scientific and empirical literature, preceding model evaluation. This pipeline was designed to optimize and fit the models effectively. For each inner training fold (4/5 folds at each outer iteration), key XGBoost hyperparameters were tuned using TPE, including learning rate (0.01–0.2), number of boosting rounds (500–1500), maximum tree depth (2–4), subsampling ratios (0.3–0.6), regularization strengths (lambda, alpha: 5–50, log scale), and gamma (0−10). These ranges were selected based on best practices and recent literature to encourage gradual learning through low learning rates and many trees [35] limit model complexity with shallow trees [46], and reduce overfitting via relatively aggressive subsampling and strong regularization [47]. A moderate gamma value helped avoid unnecessary splits [35]. Early stopping with a patience of 30 rounds was applied to halt training once validation performance plateaued [48,49]. Optimization consisted of 150 TPE iterations using 5-fold cross-validation on the training folds, which was sufficient as *RMSE* performance stabilized across folds (on the training data).

After optimization, the best-performing hyperparameter configuration, determined by the lowest average *RMSE* across the inner folds, was selected along with the corresponding number of boosting rounds and performance metrics. These results informed both final model selection and downstream interpretability analyses. Using the optimal hyperparameters and iteration count, a final model was then trained on the full inner training dataset.

2.7. Automated feature selection

Selecting the optimal subset of features for model training is challenging because the number of possible combinations grows exponentially, making exhaustive search computationally infeasible [50]. For example, with 94 features, there are $2^{94} \approx 1.98 \times 10^{28}$ possible subsets. Due to the vast number of possible solutions, feature selection is often formulated as an optimization problem. Yet, even advanced algorithms [51,52] can produce suboptimal results, getting stuck in local minima because of the problem's non-convexity [53]. Moreover, feature importance, whether from SHAP or other rankings, can shift depending on which features are included, as interactions and redundancies affect their value. Additionally, features may be correlated or interact non-linearly (i.e., their relevance is context-dependent), leading to the need for new feature importance estimates at each set. Thus, effective feature selection must dynamically adapt to these complexities rather than relying on static rankings or simple post-optimization [54].

One solution for feature selection based on dynamic importances is recursive feature elimination (RFE) in combination with iteratively optimizing and retraining models at each feature set, using *k*-fold cross-validation [55–57]. Recalculating SHAP values at each elimination step ensures that feature importance reflects the current subset and model parameters [55]. SHAP-RFE (SRFE) combines SHAP's model-agnostic information with RFE's stepwise elimination to create a compact, more interpretable feature set, directly connected to feature importances that deliver strong predictive performance, often outperforming static or purely optimization-based methods in both clarity and efficiency [58]. Feature elimination based on model-internal importance metrics of XGBoost may, for some datasets, yield comparable performance to SHAP-based elimination [59]. However, we used SRFE to align feature selection with the SHAP rankings used in our final model interpretation, aiming for a clinically interpretable and comparable feature set [60,61].

2.8. Model development

Before evaluation on each test fold, the following steps were applied to the corresponding training data.

For the data-driven models, 94 features were used as input. Features with pairwise Spearman's correlation greater than 0.80 were pruned, retaining only the feature most correlated with the target within each inner fold, to avoid redundancy, conflicting interpretations, and data leakage [62,63]. The features included in the analysis, along with their corresponding names, are listed in Supplementary Table S1.

For models using SRFE, we took several steps to ensure a feasible training time. Hyperparameters were first tuned with Optuna (as in section 2.6). Then, SHAP values identified features contributing to 95% of cumulative importance were kept for further analysis (cf. Kiekhaefer and Pourkamali-Anaraki [64] using 80%). Finally, a recursive feature elimination loop was performed: after each Optuna optimization (again as in section 2.6), the best model was fitted, SHAP values recalculated, and the least important feature removed, similar to [55]. This continued until only five features remained, chosen as a reasonable minimum before performance significantly declined.

The following five model types (all XGBoost) were trained and evaluated:

1. **Consensus model:** Trained using the 22 consensus features.
2. **Full model:** Trained on the full set of features, only excluding highly correlated features.
3. **SRFE optimal model:** Selected based on the lowest average *k*-fold cross-validation *RMSE*.
4. **SRFE Pareto model**: Selected as the model with the fewest features within 1% of the lowest *RMSE* (Pareto frontier).
5. **SRFE matched model**: Selected to include exactly 22 features, matching the number of consensus features.

Each of the five model types was trained with nested cross-validation and was tested on its respective held-out fold, ensuring that all available data contributed to unbiased evaluation. Consequently, each model type produced five separate model results covering the entire dataset. Note that the number of selected features may vary across folds, as feature selection was performed independently within each fold.

The expert-driven consensus feature selection process was conducted entirely independently of the machine learning model training. No tuning or feature selection was performed on the test data, which was held out and used only for final evaluation. The data preprocessing pipeline, including feature processing and selection criteria, was established a priori based on theoretical and empirical literature (See sections 2.6 and 2.7), before inspecting model outcomes.

2.9. Model comparison, feature ranking, and statistical analysis

All model comparisons, feature importance analyses, and statistical evaluations were conducted exclusively on predictions from the outer test folds, which were entirely held out from both model training and hyperparameter tuning. This ensured that all reported results reflect out-of-sample performance, free from data leakage or overfitting bias.

2.9.1. Model comparison

To evaluate model performance, we implemented a two-step procedure for each outer test fold. First, predictive accuracy was assessed using two standard regression metrics: R^2 and *RMSE*. *RMSE* was prioritized as the primary evaluation metric due to its robustness to variations in data dispersion, offering a more stable basis for model comparison. R^2 was reported alongside *RMSE* for interpretability.

For a precise comparison, we analyzed the absolute residuals – representing individual prediction errors – for each participant across all five outer test folds. To statistically compare model performance, we conducted paired *t*-tests on the absolute residuals within each fold. The resulting *p*-values were then aggregated using Fisher's method to determine overall significance across folds.

2.9.2. Feature importance

To assess the relative importance of individual features, we employed a dual approach. First, we examined the frequency with which each feature appeared among the top-ranked features across the five outer folds, as well as its average rank position. This provided insight into the consistency and stability of each feature's predictive relevance.

Second, we computed the mean absolute SHAP values for each feature, capturing their average contribution to model predictions while accounting for interactions with other features. This approach yielded a robust, model-agnostic measure of feature importance, improving interpretability and transparency in the modeling process.

3. Results

3.1. Participants

In the full LATINO adult sample ($N = 3634$), the mean age was 30.50 years ($SD = 11.80$), and 35% were men. After preprocessing, a total of 482 participants (13%) were removed due to not having adult QLES-Q scores. The final number of participants was therefore 3152 and had an average QLES raw score of 44.16 ($SD = 10.76$) points. The mean age increased slightly to 31.34 ($SD = 11.11$), and 33% were men. Current OCD severity was above the non-clinical range in 2384 participants (75.6%), while only 18 participants (0.6%) had a Y-BOCS-II score of 0, showing that OCD symptoms were very common. Using empirical benchmarks [65] among participants with available Y-BOCS-II severity data ($n = 3130$), 768 participants (24.5%) were subclinical, 1070 (34.2%) mild, 1129 (36.1%) moderate, and 163 (5.2%) severe. On average, participants missed 3.61 values (4%; $SD = 3.93$) before imputation in the machine dataset. On average, participants missed 0.78 values (4%; $SD = 0.76$) for the consensus model.

3.2. Model development

Models were trained using dual GPUs (Nvidia RTX 2080 Ti OC and RTX 3060 Ti) working in parallel per fold with a fixed random seed. The first SHAP estimation on the full dataset (after removing the highly correlated features) with the 95% cumulative cutoff, resulted in an average of 51.6 features ($SD = 2.06$, range 50–55). The inner fold *RMSE* can be seen in Fig. 1 for each number of features, resulting in an average optimal number of features at a mean of 32 features ($SD = 8.35$, range 21–41) and a Pareto at 1% of a mean of 15.4 features ($SD = 2.30$, range 12–18). Detailed results per model can be found in the Supplementary Tables S2–S6.

3.3. RQ1: model comparison

To address the first research question – namely, the extent to which a consensus-based model can predict QoL in individuals with OCD, and how its performance compares to that of a machine learning-based feature selection approach – the predictive accuracy of five distinct models was evaluated. All metrics are reported across the five folds, incorporating all available data ($n = 3152$). Statistical comparisons between models were performed using paired *t*-tests on the absolute residuals for each fold. The resulting *p*-values were then combined across folds using Fisher's method.

The *consensus model*, which included 22 human expert-selected features, demonstrated the weakest performance among the models. It yielded a mean *RMSE* of 7.56 ($SD = 0.14$) and a mean R^2 of 0.51 ($SD = 0.02$), indicating the highest prediction error and lowest explained variance. The paired *t*-tests of absolute residuals per fold and five separate Fisher's combined *p*-values revealed that the consensus model performed significantly worse than all other models ($ps < 0.001$).

The *full model*, which included 88 predictors, demonstrated moderate predictive accuracy with a mean *RMSE* of 7.21 ($SD = 0.08$) and a mean R^2 of 0.55 ($SD = 0.03$) across the five outer test folds. Reduced models generated via SRFE, containing between 12 and 41 features, achieved similar predictive accuracy overall. A repeated-measures ANOVA on the paired average differences per model revealed no significant overall effect of model type, $F(3,12) = 2.40$, $p = 0.12$.

Across all pairwise comparisons with the Full model, *p*-values ranged from 0.177 to 0.962, indicating non-significant differences. The performance of the models across folds, along with the significance levels derived from *p*-values obtained via Fisher's combination method of paired *t*-tests within each fold, is presented in Table 1. The mean absolute errors and paired differences, per model type, are shown in Table 2. The R^2 scores for each fold are shown in Fig. 2.

In sum, data-driven approaches consistently outperformed the expert-driven consensus model. However, the gains were modest, on average (across models and folds) approximately 0.35 points lower in *RMSE* (7.56 vs. 7.21) and 0.04 higher in R^2 (0.51 vs. 0.55) when comparing the consensus model to the best-performing data-driven models. This suggests that while expert-selected features capture much of the relevant variance, data-driven selection offers statistically significant, albeit practically limited, improvements in predictive accuracy.

3.4. RQ2: cross-model feature comparison

To address the second research question – identifying key predictors of QoL in individuals with OCD and comparing consensus-based versus machine learning-based feature selection, we conducted a pooled cross-model feature importance analysis. This involved averaging feature importance ranks and mean absolute SHAP values across models and cross-validation folds. This allowed us to identify features that consistently contributed to QoL prediction across different model configurations and to compare them with those selected in the consensus model.

Feature importance was evaluated using three pooled metrics: mean rank across models and folds, mean absolute SHAP values (indicating marginal effect on QoL), and presence stability, defined as the frequency of feature inclusion (maximum 20 [4 model types and 5 folds]). We first present results from the full models and the SRFE-based models, then contrast them with the consensus-based model to explore potential added value from more data-driven selection.

3.5. Feature importance ranking

Based on pooled ranks and SHAP values across cross-validation folds, the most influential predictor of QoL was depression severity (PHQ-9 total score; *total_score_phq9*). PHQ-9 was selected in 100% of models and folds (*Full, SRFE Optimal, SRFE Pareto, SRFE Matched*; folds 1–5) and ranked first overall, with a mean absolute SHAP value of 3.39 ($SD =$

0.27). It was also ranked first in the consensus model, where it exhibited an even higher absolute SHAP value of 3.84 (*SD* = 0.30).

The second most important feature was subjective socioeconomic status (Adult Socioeconomic Ladder Rung; *adult_ladder_rung*), which was included in all data-driven models and folds, but was not selected in the consensus model. It had a mean absolute SHAP value of 1.48 (*SD* = 0.14) and a mean rank of 2.00 (*SD* = 0.00), demonstrating a strong and consistent influence on QoL predictions.

Anxiety severity (GAD-7 total score; *total_score_gad7*) ranked third, being fully included across models and folds, with a mean absolute SHAP value of 1.16 (*SD* = 0.10) and a mean rank of 3.10 (*SD* = 0.30). In the consensus model, GAD-7 had a mean rank of 2.60 (*SD* = 0.89) with a SHAP value of 1.20 (*SD* = 0.21).

Country-level Human Development Index (*country_ihdi*) ranked fourth overall in the pooled cross-model comparison, with a mean rank of 4.30 (*SD* = 0.84) and a mean absolute SHAP value of 0.88 (*SD* = 0.17). Unlike subjective socioeconomic status, *country_ihdi* was also selected in the consensus model, with a mean rank of 2.80 (*SD* = 0.45) and a mean absolute SHAP value of 1.10 (*SD* = 0.05).

Two obsessive–compulsive symptom measures followed. OCD symptom severity (Y-BOCS-II total score current; *ybocsii_total_cur*) had a mean rank of 5.06 (*SD* = 0.75), a mean absolute SHAP value of 0.73 (*SD* = 0.09), and was selected in four of five folds (folds 1, 3, 4, 5). Y-BOCS-II obsessions score (current, *yb2_score_obs_cur*) had a mean rank of 5.75 (*SD* = 0.83) and a mean absolute SHAP value of 0.65 (*SD* = 0.06), but was included in only one fold, explaining its relatively large local effect. In the consensus model, *ybocsii_total_cur* had a mean rank of 3.60 (*SD* = 0.89).

Anhedonia (SHAPS score; *shaps_score*) was consistently selected across all data-driven models folds, with a mean rank of 5.85 (*SD* = 0.91) and a mean absolute SHAP value of 0.67 (*SD* = 0.09). Despite its consistent inclusion and importance, it was not selected in the consensus model.

Number of psychiatric diagnoses (*dia_count*) had a mean rank of 7.58 (*SD* = 1.35) and a mean absolute SHAP value of 0.67 (*SD* = 0.08), showing consistent inclusion across approaches. It was closely ranked in the consensus model with a mean rank of 6.00 (*SD* = 0.71) and a mean absolute SHAP value of 0.76 (*SD* = 0.06), indicating similarity across approaches.

The final two features completing the top 10 were motivation for OCD improvement (*ocd_impr_motivation*; mean rank = 8.70, *SD* = 1.82; mean absolute SHAP = 0.53, *SD* = 0.09) and access to OCD treatment (*ocd_tx_access*; mean rank = 10.60, *SD* = 2.33; mean absolute SHAP = 0.45, *SD* = 0.10), both selected across the main model variants. Notably, neither was included in the consensus model, despite their relatively high importance in the data-driven models.

Selection frequency varied across other features: some predictors (e.g., *shaps_score, country_ihdi, dia_count, ocd_impr_motivation*) were consistently selected across folds and models, whereas others (e.g., *yb2_score_obs_cur*) were included only in a minority of folds but exhibited larger local effects when present.

More detailed *pooled* cross-model metrics can be found in Supplementary Table S7, and a comparison of models' top 10 most important features is shown in Fig. 3. As shown in Fig. 4, SHAP values support the model interpretations, with the beeswarm plot of the top five features indicating that higher PHQ-9 scores, lower ladder rung, and other key predictors are associated with reduced quality of life.

We did not pursue further analysis of interaction effects due to the relatively low mean absolute SHAP values observed across interaction pairs, aggregated over models and folds (mean across types: 0.004, *SD* = 0.008). The highest mean interaction value – 0.138 (*SD* = 0.056) – was found for the pair country IHDI (*country_ihdi*) and depression severity (PHQ-9; *total_score_phq9*) in the Pareto model. Other notable interactions included *ocd_impr_new_relation* × *total_score_phq9* (mean = 0.106, *SD* = 0.033) and *dia_count* × *total_score_phq9* (mean = 0.094, *SD* = 0.032). However, these interactions were considered outside the scope of the current study and, when compared to the magnitude of the main SHAP effects, not sufficiently strong to justify deeper investigation.

In sum, some features overlapped between the consensus and data-driven selection methods, reflecting shared recognition of key predictors. However, the data-driven models uncovered additional important features, such as subjective placement in the social hierarchy and anhedonia, that were not selected by experts but showed strong empirical contributions. The most important features overall, based on mean absolute SHAP values, were depression severity, subjective placement in the social hierarchy, and anxiety severity, which consistently ranked highest across models and folds. This highlights the added value of data-driven approaches in identifying both established and novel predictors that enhance model performance.

4. Discussion

This study examined how expert-informed and data-driven models perform in predicting QoL among individuals with OCD. Specifically, we compared the predictive accuracy of an expert-based model, grounded in clinical and research expertise, to that of a purely data-driven machine learning approach, and what each method revealed about the most influential predictors of QoL in OCD.

Our results showed that data-driven models offered a modest but consistent improvement over the expert-informed model. The data-driven models, both when including all features and when using fewer features via SRFE, achieved slightly better average predictive accuracy (*RMSE* = 7.23, R^2 = 0.55) than the expert-based model (*RMSE* = 7.56, R^2 = 0.51). While this suggests that data-driven approaches can enhance prediction, the relatively small performance gap also highlights the strength and relevance of expert knowledge, particularly when it draws on well-established clinical constructs. Although statistically significant, the improvement in prediction accuracy was modest and should be interpreted as an incremental gain relevant to model refinement and risk stratification, rather than as a large standalone clinical effect.

In identifying key predictors of QoL, both approaches converged on several core features, including the severity of depression, anxiety, and OCD. These shared predictors reflect a common understanding of the psychological and social dimensions that impact QoL in OCD. However, the data-driven models also surfaced additional features – such as anhedonia and subjective placement in the social hierarchy – that were not included in the expert model but contributed meaningfully to

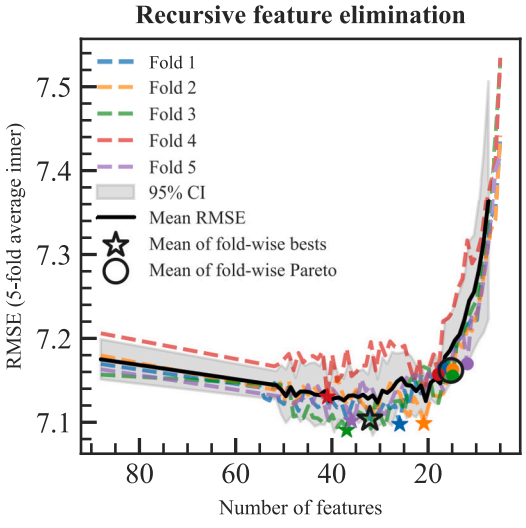

Fig. 1. Displays the *RMSE* from 5-fold cross-validation across inner training fold data for the number of features when iteratively being eliminated using SHAP-RFE. Color-matched stars display the optimal per fold and circle the Pareto for the lowest number of features within 1% higher *RMSE* than optimal.

Table 1
Performance metrics for each method and statistical significance of paired comparisons across participants.

						p-values from Fisher's combined probability test, applied to paired absolute errors per participant				
Method	*μRMSE*	*σRMSE*	*μR²*	*σR²*	*MAE*	Full	SRFE optimal	SRFE Pareto	SRFE matched	Consensus
Full	7.212	0.081	0.549	0.026	5.599	–	0.900	0.177	0.962	<0.001
SRFE optimal	7.229	0.098	0.547	0.026	5.605	0.900	–	0.203	0.476	<0.001
SRFE Pareto	7.264	0.141	0.543	0.023	5.650	0.177	0.203	–	0.213	<0.001
SRFE matched	7.205	0.093	0.550	0.025	5.605	0.962	0.476	0.213	–	<0.001
Consensus	7.556	0.140	0.506	0.021	5.912	<0.001	<0.001	<0.001	<0.001	–

Note. μ = mean, σ = standard deviation, *RMSE* = root mean square error; R^2 = coefficient of determination, *MAE* = mean absolute error. Averaged across folds and method type. Fisher's *p*-values.

Table 2
Mean absolute errors and paired differences between models.

Method	*Full*	*Optimal*	*SRFE Pareto*	*SRFE matched*	*Consensus*
Full	5.599	−0.006	−0.051	−0.006	−0.312
Optimal	0.006	5.605	−0.045	<0.001	−0.306
SRFE Pareto	0.051	0.045	5.650	0.045	−0.262
SRFE matched	0.006	<0.001	−0.045	5.605	−0.307
Consensus	0.312	0.306	0.262	0.307	5.912

Note. Mean absolute error (diagonal) and mean paired differences in absolute error between participants. Averaged across folds and method type.

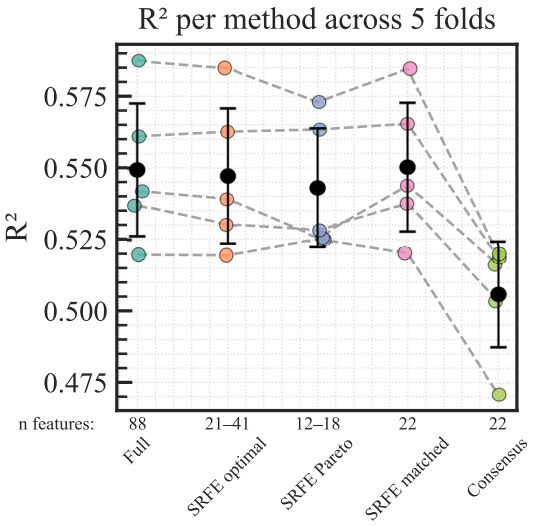

Fig. 2. Shows the scatter plots for all the models' R^2 scores per fold. Dotted lines show the pairing per fold. A black dot indicates the average, and the error bars represent the standard deviation across the folds.

predictive performance. This suggests that machine learning can complement clinical expertise by uncovering less obvious, yet empirically important, factors.

Notably, depression severity, subjective placement in the social hierarchy, and anxiety severity emerged as the most influential predictors across all models. Their prominence reinforces the central role of affective symptoms and perceived social standing in shaping QoL outcomes in OCD. At the same time, the inclusion of novel predictors by the data-driven models points to the potential of machine learning to broaden our understanding of QoL beyond traditional clinical frameworks. It should be noted that OCD severity was less important for overall QoL than depression and anxiety. This is in line with previous studies [25] and underscores the importance of a broad treatment approach when aiming to improve general life outcomes in individuals with OCD.

In addition to these primary predictors, the data-driven models highlighted OCD-specific features, such as overall OCD symptom severity and the severity of obsessions, as notable contributors to QoL. This is consistent with previous research showing that intrusive thoughts and their associated distress can be particularly impairing, sometimes exerting a stronger influence on daily functioning than compulsions [66]. The number of psychiatric diagnoses, used here as a proxy for overall psychiatric burden, also emerged as an important factor, in line with studies demonstrating the cumulative impact of comorbid conditions on QoL [25,66]. Additional variables, such as anhedonia and access to OCD treatment, align with prior work linking motivational factors, mood disturbance, and treatment availability to broader life outcomes in OCD [19]. The emergence of these predictors, some absent from expert-driven selection, shows the importance of considering both symptom-specific and contextual variables when addressing QoL in OCD, and suggests that obsessions may warrant particular clinical attention in interventions aimed at improving overall well-being.

Across models, depressive and anxiety symptoms – and perceived social standing – dominated QoL variance, with OCD severity contributing but ranking below mood/anxiety. These results echo prior syntheses showing broad QoL impairment in OCD and disproportionate burden from affective symptoms [10,12,66] and align with Brazilian cohort data in which depressive severity and comorbidity predicted QoL decrements [67].

Notably, both subjective socioeconomic status and country-level human development ranked among the most important predictors of QoL, indicating socioeconomic influences at both individual and structural levels. Subjective social status was assessed using a 10-rung ladder reflecting perceived social position relative to others in terms of income, education, and occupation; its predictive value may therefore capture psychosocial burden beyond objective socioeconomic indicators. In contrast, the IHDI reflects broader societal conditions related to health, education, and living standards, suggesting that access to resources, treatment opportunities, and social support may influence QoL beyond symptom severity alone.

The salience of subjective social status, validated in Brazil and elsewhere, suggests that perceived rank –capturing both social and economic facets – may be a tractable, contextual treatment target alongside symptom reduction. Moreover, OCD-specific phenomena such as sensory phenomena [68] and obsessional distress may impair everyday functioning beyond compulsion frequency, meriting greater emphasis when designing interventions.

Clinically, these findings suggest that QoL in OCD is linked not only to OCD symptom severity, but also to broader affective and psychosocial difficulties. Brief measures of depression and anxiety can be easily included in routine OCD assessment and should be monitored during and after treatment, as persisting affective symptoms may limit improvement in QoL even when OCD symptoms improve. The finding for subjective social status further suggests that clinicians should attend to how patients perceive their social position, value, and belonging, which may capture shame, exclusion, or reduced agency not reflected in objective SES indicators. These implications align with a broader shift from evaluating treatment only by disorder-specific symptom reduction toward also considering functioning, recovery, and quality of life [69]. However, because these findings are cross-sectional, they identify clinically relevant correlates rather than causal treatment targets.

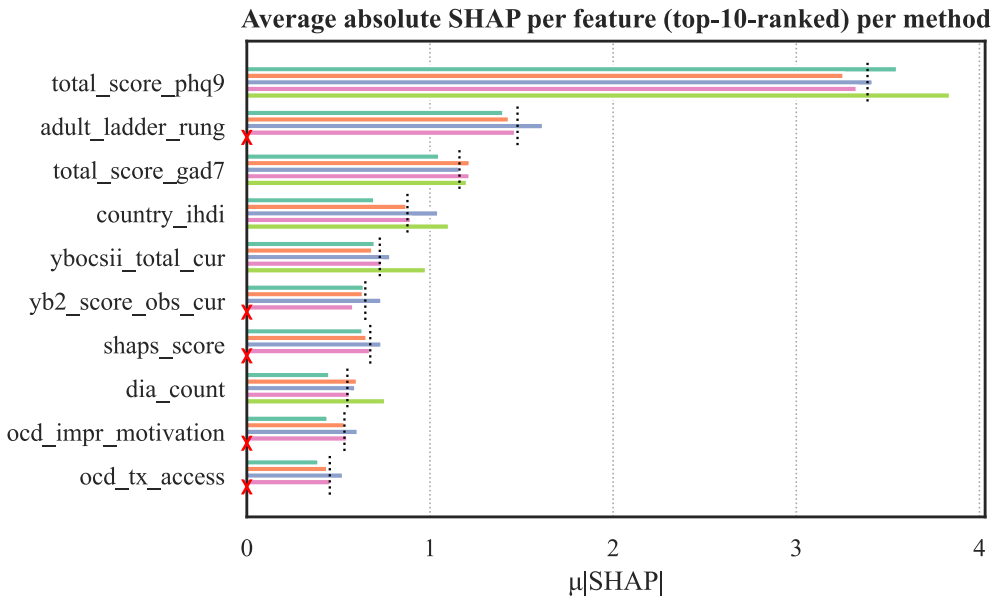

Fig. 3. Top-10-ranked features across machine-selected and consensus-selected features. The dotted line represents the pooled mean absolute SHAP value (μ |SHAP|). Features are sorted based on their pooled mean rank across all models using automated feature selection (i.e., all models except "Consensus"). A red "X" indicates that the feature was not included in the consensus feature set. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

4.1. Strengths and limitations

This study has several methodological strengths but also key limitations. The absence of an external validation dataset limits generalizability to independent populations. [70]. However, we used nested cross-validation (5 × 5 folds), a rigorous internal validation approach that reduces overfitting and yields reliable performance estimates [44]. The relatively large sample (n = 3152) further strengthens internal validity and statistical power, exceeding many comparable machine learning studies in health research [71].

A second limitation is the exclusive use of gradient boosted trees (XGBoost), selected for their balance of accuracy and interpretability [35]. While this consistency improved comparability and transparency, it limited exploration of other algorithms that might offer complementary results [72]. Future work should benchmark against alternatives such as classical random forests or neural networks.

Our models focused on the main effects, with limited investigation of interactions. Preliminary analyses suggested interactions were minor, but more advanced modeling could clarify these relationships. SHAP-based recursive feature elimination (SRFE) supports interpretable, data-driven feature selection [73]. High-cardinality features, such as race, were retained as single variables to preserve model stability across folds and avoid minority-class instability or extensive pipeline modifications beyond the scope of this study. Furthermore, using the country IHDI instead of categorical country variables promoted statistical stability and interpretability, although it reduced country-specific

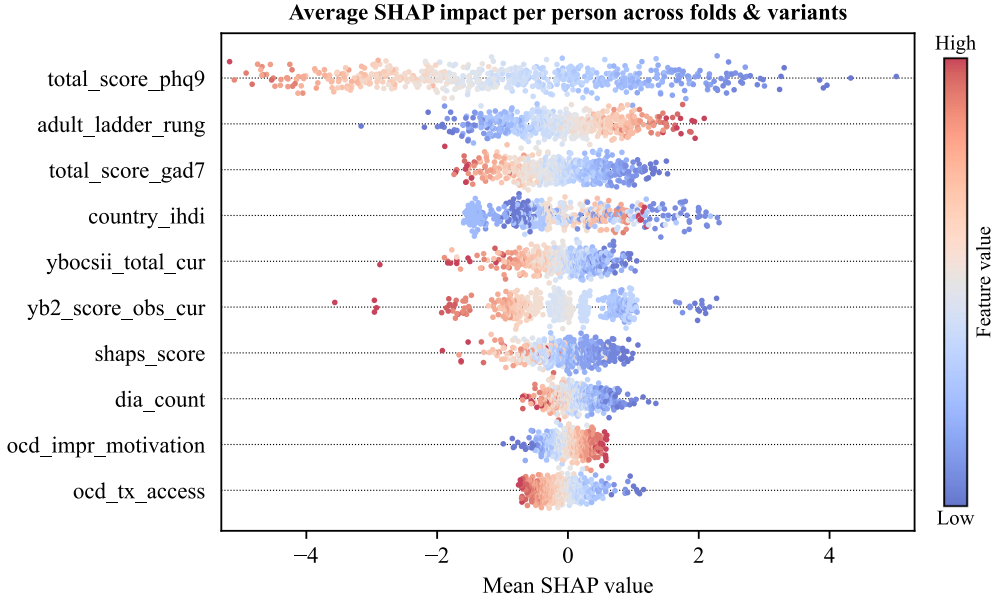

Fig. 4. Top-10-ranked features from machine-selected models. Each dot shows the mean SHAP value per individual, averaged across folds and models. The beeswarm plot depicts how feature variation influences predicted quality of life (QoL), with positive SHAP values indicating higher predicted QoL. Colors represent feature values. Extreme 1% values are clipped for visualization.

information. PHQ-9, GAD-7, and particularly SHAPS may also partly reflect psychometric proximity to QoL rather than independent clinical effects, given their conceptual overlap with QLES-Q domains such as enjoyment, satisfaction, and emotional functioning. In addition, a formal remitted-versus-active OCD status variable was not available; therefore, current Y-BOCS-II severity was used as a proxy for clinical status at the time of QoL assessment. Thus, the findings should be interpreted as conditional predictive associations rather than evidence of independent causal effects.

Our sample consisted of individuals with self-identified Latin American ancestry, which is a strength because it focuses on a historically underrepresented population in OCD research. However, this sample composition may limit generalizability to other ethnic and ancestral groups, and the findings should therefore be interpreted with caution in other contexts.

Finally, while our dataset was rich in clinical, personal, and contextual features, it did not capture all relevant variables. We prioritized variables with low missingness and minimal collinearity, but acknowledge the possibility of unmeasured confounding [74]. Nevertheless, the breadth of variables and large, diverse sample provide a strong empirical basis for future refinement.

In summary, despite limitations in external validation, algorithm diversity, and interaction modeling, this study benefits from a large sample, robust internal validation, interpretable feature selection, and comprehensive variable inclusion, collectively supporting the reliability and utility of our findings.

4.2. Conclusion

This study compared expert-informed and data-driven approaches to modeling QoL in OCD, demonstrating that while expert knowledge captures much of the relevant variance, data-driven methods can modestly enhance predictive accuracy and uncover additional influential features. Both modeling approaches identified core predictors such as depression, anxiety, and OCD severity, but the data-driven model also highlighted novel contributors, including subjective social status, anhedonia, and country IHDI. These findings underscore the value of machine learning in surfacing empirically relevant yet underrecognized factors. Importantly, the data-driven model maintained comparable performance with fewer features, suggesting potential for more efficient and scalable prediction. The consistent prominence of PHQ-9, GAD-7, and the MacArthur Scale across models reinforces the central role of affective symptoms and perceived social standing in shaping QoL. Overall, integrating machine learning with domain expertise offers a promising strategy for improving prediction and advancing our understanding of the psychological and social dimensions of mental health and well-being.

Declaration of generative AI and AI-assisted technologies in the manuscript preparation process.

During the preparation of this work, the authors used ChatGPT in order to rephrase certain sentences to improve clarity. After using this tool, the authors reviewed and edited the content as needed and take full responsibility for the content of the published article.

CRediT authorship contribution statement

Max Thorsson: Formal analysis, Conceptualization, Methodology, Writing – original draft, Writing – review & editing. **Angela Trujillo:** Investigation, Data curation, Writing – review & editing. **Maritza J. Romero:** Investigation, Data curation, Writing – review & editing. **Jonathan Irreño-Sotomonte:** Investigation, Data curation, Writing – review & editing. **Tania Borda:** Investigation, Data curation, Writing – review & editing. **Raquel M. Zepeda-Burgos:** Investigation, Data curation, Writing – review & editing. **Mayra C. Martinez Mallen:** Investigation, Data curation, Writing – review & editing. **Beatriz Moyano:** Investigation, Data curation, Writing – review & editing. **Jacey L. Anderberg:** Investigation, Data curation, Writing – review & editing. **Josselyn S. Muñoz:** Investigation, Data curation, Writing – review & editing. **Hannah C. Moore:** Investigation, Data curation, Writing – review & editing. **Renee M. Frederick:** Investigation, Data curation, Writing – review & editing. **Dayan Berrones:** Investigation, Data curation, Writing – review & editing. **Vanessa Zavala Cruz:** Investigation, Data curation, Writing – review & editing. **Ogechi Onyeka:** Investigation, Data curation, Writing – review & editing. **Andrew D. Wiese:** Investigation, Data curation, Writing – review & editing. **Kevin M. Wagner:** Investigation, Data curation, Writing – review & editing. **LATINO:** Resources. **GTTOC:** Resources. **James J. Crowley:** Resources, Methodology, Writing – review & editing. **Eric A. Storch:** Conceptualization, Methodology, Supervision, Writing – review & editing. **Matti Cervin:** Conceptualization, Methodology, Supervision, Writing – original draft, Writing – review & editing.

Ethics approval and consent to participate

All methods were carried out in accordance with relevant guidelines and regulations, including the Declaration of Helsinki. Ethical approval for data collection was obtained from the relevant institutional review boards or ethics committees at each participating site: Centre for Addiction and Mental Health Research Ethics Board (IRB# 014/2022); Centro de Investigaciones Tecnológicas, Biomédicas y Medioambientales (CITBM), Comité de Ética de la Investigación (IRB#00007–2024); Centro Interdisciplinario de Tourette, TOC, TDAH y Trastornos Asociados (CITA), Fleni Comité de Ética en Investigación (IRB# 6398); Clínica Nuestra Señora de la Paz, Comité de Ética de la Investigación (IRB# MN.DO.7.4); Corporación Universitaria Rafael Núñez, Comité de Ética de Investigación (IRB# CURN0004-CE042022); Grupo de Estudios Médicos y Familiares Carraci S.C., Comité de Ética en Investigación Biomédica para el Desarrollo de Fármacos, S.A. de C.V. (IRB# CONBIOETICA-14-CEI-006-20,170,504); Hospital das Clínicas da Faculdade de Medicina da USP, Comissão Nacional de Ética em Pesquisa (IRB# 6.568.307); Hospital Psiquiátrico de Asunción, National University of Asuncion Faculty of Medical Science Research Directorate Ethics Committee (IRB# 237/2022); Instituto Realize, Ethics Committee of the Pontificia Universidad Católica Argentina (IRB# 06/2022); Universidad de Buenos Aires (UBA), Comisión de Evaluación de Conductas Responsables en Investigación de la Facultad de Psicología (no IRB number assigned); Universidad de La Sabana, Subcomisión de Investigación y Ética (no IRB number assigned); Universidad de Valparaíso, Comité Ético-Científico (CEC-UV) (IRB# CEC 250–22); Universidad Dr. José Matías Delgado, Centro de Investigación en Ciencias y Humanidades (CICH) (no IRB number assigned); Universidad Privada del Norte (UPN), Comité de Ética en Investigación (IRB#0004–2022/ID-CIEI); Universidad San Francisco de Quito (USFQ), Comité de Ética de Investigación en Seres Humanos (IRB# 2022-006 M); University of Calgary, Conjoint Health Research Ethics Board (CHREB) (IRB# REB21–1628); and Universidad Tecnológica Privada de Santa Cruz (UTEPSA), Comité de Ética de la Investigación (IRB# 1/2022).

Informed consent was obtained from all participants and/or their legal guardians in accordance with local institutional requirements. Participants provided written and/or electronic consent prior to participation.

Ethics declaration

Written informed consent to take part in the study and to publish the article has been obtained from all participants or their legal representatives. The privacy rights of participants have been observed.

This study was performed in compliance with relevant laws, regulatory frameworks, and guidelines where the research took place, including the Declaration of Helsinki.

This study was approved by the relevant institutional review boards or ethics committees at each participating site:

Centre for Addiction and Mental Health Research Ethics Board.

(Approval No. IRB# 014/2022)

Centro de Investigaciones Tecnológicas, Biomédicas y Medioambientales (CITBM), Comité de Ética de la Investigación.

(Approval No. IRB#00007–2024)

Centro Interdisciplinario de Tourette, TOC, TDAH y Trastornos Asociados (CITA), Fleni Comité de Ética en Investigación.

(Approval No. IRB# 6398)

Clínica Nuestra Señora de la Paz, Comité de Ética de la Investigación.

(Approval No. IRB# MN.DO.7.4)

Corporación Universitaria Rafael Núñez, Comité de Ética de Investigación.

(Approval No. IRB# CURN0004-CE042022)

Grupo de Estudios Médicos y Familiares Carraci S.C., Comité de Ética en Investigación Biomédica para el Desarrollo de Fármacos, S.A. de C.V.

(Approval No. IRB# CONBIOETICA-14-CEI-006-20,170,504)

Hospital das Clínicas da Faculdade de Medicina da USP, Comissão Nacional de Ética em Pesquisa.

(Approval No. IRB# 6.568.307)

Hospital Psiquiátrico de Asunción, National University of Asuncion Faculty of Medical Science Research Directorate Ethics Committee.

(Approval No. IRB# 237/2022)

Instituto Realize, Ethics Committee of the Pontificia Universidad Católica Argentina.

(Approval No. IRB# 06/2022)

Universidad de Buenos Aires (UBA), Comisión de Evaluación de Conductas Responsables en Investigación de la Facultad de Psicología.

(No IRB number assigned)

Universidad de La Sabana, Subcomisión de Investigación y Ética.

(No IRB number assigned)

Universidad de Valparaíso, Comité Ético-Científico (CEC-UV).

(Approval No. IRB# CEC 250–22)

Universidad Dr. José Matías Delgado, Centro de Investigación en Ciencias y Humanidades (CICH).

(No IRB number assigned)

Universidad Privada del Norte (UPN), Comité de Ética en Investigación.

(Approval No. IRB#0004–2022/ID-CIEI)

Universidad San Francisco de Quito (USFQ), Comité de Ética de Investigación en Seres Humanos.

(Approval No. IRB# 2022-006 M)

University of Calgary, Conjoint Health Research Ethics Board (CHREB).

(Approval No. IRB# REB21–1628)

Universidad Tecnológica Privada de Santa Cruz (UTEPSA), Comité de Ética de la Investigación.

(Approval No. IRB# 1/2022)

Funding

Research reported in this publication was supported by the National Institutes of Health under award numbers U01MH125062, U01MH125062-03S1, and U01MH125050. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health. This research was supported by the Ancestral Population Network (APN), a National Institute of Mental Health Initiative. The APN was established by the NIMH, to support data collection and analysis of psychiatric phenotypes across ancestral populations.

Declaration of competing interest

Some authors have received research funding, consulting fees, or royalties from organizations not related to this work. All other authors declare no conflicts of interest. These relationships did not influence the design, analysis, interpretation, or reporting of the study.

Dr. Arnold receives research funding from Biohaven Inc., Canadian Institute for Health Research, Ontario Brain Institute, Genome Canada, and Alberta Children's Hospital Research Foundation.

Dr. Cervin receives research support from the Kavli Foundation, the Swedish Research Council for Health, Working Life and Welfare, the Lindhaga Foundation, Stiftelsen Clas Grochinskys Minnesfond, the Crown Princess Lovisa's Association, the Kamprad Family Foundation, Fonden för Psykisk Hälsa, and Skåne University Hospital's Foundations and Donations; and financial compensation from Springer for editorial work outside of the submitted work.

Dr. Fontenelle receives research grants not related to this work from the National Council for Scientific and Technological Development – CNPq #313162/2023–9.

Dr. Martínez-González receives funding from NIH and FDA. She is a consultant with Abbvie, Sage Pharmaceuticals and SAMHSA.

Dr. Onyeka receives research funding from the Texas Child Mental Healthcare Consortium, the International OCD Foundation, and the Texas Higher Education Coordinating Board.

Dr. Pinciotti reports receiving fees to be a consultant and workshop presenter with Jenna Overbaugh, LLC, The Knowledge Tree, and the OCD Training School.

Dr. Rodriguez (in the last three years) has been a consultant for Biohaven Inc., Osmind, and Biogen; received research grant support from Biohaven Inc.; received royalties from American Psychiatric Association Publishing; and received a stipend from APA Publishing for her role as Deputy Editor at The American Journal of Psychiatry and a stipend for her role as Deputy Editor of Neuropsychopharmacology.

Dr. Shavitt receives research grants not related to this work from the National Council for Scientific and Technological Development – CNPq #307386/2021–0; 402628/2019–5 and the Sao Paulo Research Foundation (FAPESP, #2019/27250–3).

Dr. Storch reports receiving research funding to his institution from the International OCD Foundation, Wellcome Trust, and NIH. He receives direct funding from the International OCD Foundation as well as MHNTI for providing trainings on treating obsessive-compulsive disorder with psychotherapy. Furthermore, Dr. Storch co-founded Rethinking Behavioral Health which provides training and consultation in the treatment of obsessive-compulsive disorder and related conditions. He was a consultant for Brainsway and Biohaven Pharmaceuticals in the past 36 months. He owns stock options less than $5000 in NView (for distribution of the Y-BOCS and CY-BOCS) and receives royalties from OCD Scales LLC (for distribution of the Y-BOCS and CY-BOCS). He receives book royalties from Elsevier, Wiley, Oxford, American Psychological Association, Guildford, Springer, Routledge, and Jessica Kingsley.

All other authors report no financial disclosures.

Acknowledgments

The authors would like to thank all of the participants, clinics, and colleagues who made this project possible. We would like to acknowledge Gabrielle Armstrong, Laura Barón Castano, Sanie Marie Bulsara, Shaun Bulsara, Mabel Carrera-Olivares, María Catano-Cedillo, Kyra Cheung, Jéssica Crivelari, Katia Regina da Silva, Duff Dean, José Erazo, Diego Forero, Bharathi Gadad, Katy Gathron, Miri Gitik, Letty Guerrero, Imaje Harvey, Carlos Hines, Andrea Horvath McMarques, Rachel Jocson, David Keller, Minjee Kook, Caleb Lack, Daniel Macioci, María Martínez, Guanpei Ming, Carlos Miranda, Mary Beth Morgan, Jhonny Muñoz, Emily O'Bryant, Maria Ochoa Rojas, Darpan Patel, Laura Petty, Uma Ramamurthy, Vivek Ramanathan, Catherine Rast, Bradley Riemann, Juan Rivas, Santiago Romero, Esme Rosario, Brenda Patricia Salcedo Martínez, Elda Yadira Salcedo Martínez, Leanne Scott, Natasha Sefcovic, McKenzie Sluder, Sharron Smart, Keaton Soileau, Andrew Storch, Andres Viana, and Jackie Wynn.

Recruitment for this study was facilitated by the Simons Simplex Collection (SSC). We are grateful to all of the families at the participating SSC sites, as well as the principal investigators (A. Beaudet, R. Bernier, J. Constantino, E. Cook, E. Fombonne, D. Geschwind, R. Goin-Kochel, E. Hanson, D. Grice, A. Klin, D. Ledbetter, C. Lord, C. Martin, D. Martin, R. Maxim, J. Miles, O. Ousley, K. Pelphrey, B. Peterson, J. Piggot, C. Saulnier, M. State, W. Stone, J. Sutcliffe, C. Walsh, Z. Warren, E. Wijsman).

Recruitment for the study included ResearchMatch, a national health volunteer registry that was created by several academic institutions and supported by the U.S. National Institutes of Health as part of the Clinical Translational Science Award (CTSA) program. ResearchMatch has a large population of volunteers who have consented to be contacted by researchers about health studies for which they may be eligible. Review and approval for this study and all procedures were obtained from Baylor College of Medicine.

Latin American Trans-ancestry INitiative for OCD genomics (LATINO):

Victor R. Adorno[17,18], Victoria Agostini[19], William W. Aguilar[20], Cinthia Aguirre[17], Valentina Alvarado-Quiroz[21], NaEshia Ancalade[15], Maria E. Anciburo[22], Jacey L. Anderberg[11,12], Amira Y. Antich[8,23], Diego Aponte[19], Jhon E. Arango Londoño[24], Alejandro A. Arellano Espinosa[25], Luisa Arellano-Ramírez[26], Paul D. Arnold[27,28], Brygith Asenjo Carmelo[22], Elizabeth G. Atkinson[29], Tatevik Avanesyan[30], Juliana E. Avery[11,31], Andrea F. Avila-Vazquez[26], Jose N. Ayala[17,32], Tania L. Barbieri Aguirre[22], Cynthia N. Barrera[33], Julian A. Barrero Contreras[3], Kelly Barry[11,14], Amanda N. Belanger[34], Laura M. Benitez[17,35], Dayan Berrones[11], John R. Best[36], Tim B. Bigdeli[37], Hemamalini Bommiasamy[15], Tania Borda[4,5], Laura Boyajian[38], Nathan Boyle[39], Diego Briceño Muñoz[22], Christie L. Burton[40], Joseph D. Buxbaum[41], Ricardo E. Calderón Rivera[19], Jennifer L. Callahan[42], Carlos A. Camacho Gomez[43,44], Laura A. Campos Ponce[22], Ernesto G. Canchari[22], Roxana Carrasco Torres[22], María L. Catano-Cedillo[26], Matti Cervin[11,16], Alexandria Chang[39], Jessica SC. Cheng[11,45], Jessica M. Chire[46,47], Macarena Churruca Muñoz[21,48], David C. Cicero[49], Amanda P. Cifuentes-Espinoza[21], Pamela Claisse Quiroz[50], Gianna M. Colombo[11,51], Jonathan S. Comer[52], Luz Condori-Apaza[22], Jennifer Cortina[52], Carolina Costa[52], Jennifer Crosbie[40], James J. Crowley[15], Victor O. Cruz[46,53], Isabella A. Cruz[54], Allison Cunning[38], Guillermo J. Dager Perez[24], Megan M. Dailey[11,55], Brian D. Dang[11,56], Kelly L. Davalos Checco[22], Alyssa Dawson[52], Pavel del Prado[26], Daniela del Río[50], Amelia Dev[34], Sara Diaz[19], Florencia S. Estrada[8,57], Edith G. Figueroa[46], Letizia M. Flecha[17,35], Álvaro M. Flores-Garcia[22], Martin E. Franklin[58,59], Renee M. Frederick[11,14], Ricardo Fujita[60], Jami M. Furr[52], Alessa Gaeta[30], Olivia M. Gamarra[17,61], Camila Garay-Contreras[21,48], Nataly García-Bravo[21], Cynthia García-Jaimes[26], Sheldon R. Garrison[62,63], Paola Giusti-Rodriguez[64], Andrés González Galera[65], Polaris Gonzalez-Barrios[19], Wayne K. Goodman[11], Javier D. Gorosito[66], Dorothy E. Grice[41], Jerry Guintivano[67], Daniel G. Guttfreund[68], Matthew W. Halvorsen[15], Jorge AR. Haslop Miles[69,70], Francisca Henríquez-Belmar[21], Dariana G. Hernandez[30], Estefanía V. Hernández Núñez[71], John M. Hettema[72], Joseph D. Hovey[54], Bryan Huerta Ccarhuas[22], Jonathan Irreño-Sotomonte[3], Selva S. Iturralde Pedraza[73], Nisha Jagannathan[11], Matias Jensen[21], Rozmin B. Jiwani[74], Jessica S. Johnson[67], Sonia Joutteaux[75], Joali Alexandra Juárez Lujambio[25], Sabrina Kehrer[52], James L. Kennedy[76,77], Nuria Lanzagorta[26], Gabriel Lázaro-Muñoz[78,79,80], Melanie Longhurst[75], David A. Lozada Martinez[81], Elba S. Luna[46,47,53], Anna MacLellan[36], Andrea H. Marques Horvath[82], Molly Martinez[83], Galo N. Martínez[84], Mayra C. Martinez Mallen[7], Karen G. Martinez-Gonzalez[19], Maria de Los Angeles Matos[8], Manuel Mattheisen[85], Ryan J. McCarty[11], Joseph F. McGuire[39], Dean McKay[86], Ciara McOmber[15], Nancy E. Melendez de Borja[6], Charlene Minaya[86,87], Tomás Miño[21], Sara M. Mithani[74], Mariana Monasterio Zambrana[43,44], Hannah C. Moore[11], Alonso Morales-Rivero[88], Pablo R. Moya[21,48,89], Maria B. Moyano[8,9,10], Pavithra Mukunda[30], Josselyn S. Muñoz[11,13], Maria J. Muñoz Zabaleta[2], Humberto Nicolini[26,90], Ambar A. Núñez Bracho[71], Belinda E. Núñez Bracho[71], Diana Obando[2], Monica E. Ochoa[54], Marcos E. Ochoa-Panaifo[81,22], Marcelo O'Higgins[17,18], Luis A. Olavarria Castaman[81], Trinidad Olivos Balmaceda[21], María J. Olmedo Tobar[21], Ogechi C. Onyeka[11], Iliana Ortega[27,28], Vanessa S. Paredes-Linares[22], Michele T. Pato[91,92], Ivana Pavlinusich[70], Mariel Paz y Miño[93], Edilberto Peña de Leon[94], José L. Perales Orellana[43,44], Stacey Pereira[95], Tania Pérez Duarte[33], Angel Phan[75], Maritza Placencia[47,53], Lorena Ponce[33], Kendall Poovey[38], Thomas Qiao[28], Maria F. Quiroz-Padilla[2], María F. Quispe - Correa[81], Alessandra S. Ramirez Heros[22], Dayana A. Ramírez Valencia[24], Renato T. Ramos[96], Iaroslava C. Ramos[96], Kesley A. Ramsey[39], Diana Rancourt[38], Sebastian Rebatta Gomez[22], Rocío C. Reyes Espinosa[97], Miguel Reyes-Parada[98,99], Peggy Richter[96], Carolyn I. Rodriguez[30], Servando Rodriguez Barajas[11], Maritza J. Romero[2], Camilo J. Ruggero[42], Angel Ruiz-Chow[88], Melisa N. Sagarnaga[69,70], Gianella N. Salazar Montoya[22], Nadeen Salhadar[54], Russell J. Schachar[40], Lucia M. Segovia[17,35], Laura D. Seligman[54], Isaac J. Siegel[62], Samuel D. Spencer[49], S. Evelyn Stewart[36], Eric A. Storch[11], Maria J. Tames-Bravo[26], Vissente Tapia-Cuevas[21], Peter M.. Thompson[75,100], Kiara R. Timpano[34], Silvina Tonarelli[75], Beatriz Treviño-de la Garza[101], Angela Trujillo[2], Constanza Uribe Villar[21], Carolina Valdés Jardón[33], Keara Valentine[30], Alexie Vallejo-Silva[3], Peter van Roessel[30], Javier Vargas-Medrano[75,100], María I. Vásquez-Suyo[47,102], Maria V. Vergara[17,61], Guadalupe Vidal-Martinez[75,100], Mariluz V. Vilchez Huaman[22], Caroline Vollaro[52], Kevin M. Wagner[11], Andrew D. Wiese[11], Sara Wilkerson[86], Gwyneth Zai[76,103], Luz M. Zappa[8,57], Vanessa Zavala Cruz[11], Raquel M. Zepeda-Burgos[6], Anthony W. Zoghbi[11]

Brazilian Obsessive-Compulsive Spectrum Disorder Working Group (GTTOC):

Natalia G. Albuquerque[104], Gilberto S. Alves[105], Raíza Alves-Pereira[106], Larissa MS. Artiga[107], Julianna P. Azevedo[107], Izabela G. Barbosa[108], Cristiane F. Bortoncello[109], Carolina Cappi[91,92,107], Beatriz A. Carneiro[106], Renan B. Christo[110], Daniel LC. Costa[107], Fernanda Cupertino[111], Maria Alice de Mathis[107], Gabriela B. de Menezes[112,113], Luna Dib[106], Mateus JA. Diniz[114], Maria Conceição do Rosário[104], Carolina B. Dreher[115], Carina Félix-da-Silva[112,113], Ygor A. Ferrão[109], Gabriela M. Ferreira[116], Luana C. Ferreira[104], Leonardo F. Fontenelle[112,113], Gabriel A. Frota[117], Maria Luiza W. Gil[107], Izabela G. Guarinon[116], Marcelo Q. Hoexter[107], Ana G. Hounie[107], Monicke O. Lima[107], Ariane F. Martinho[104], Thays Mendes Aguiar[104], Jamile VR. Merces[110], Euripedes C. Miguel[107], Amelia Muniz[104], Beatriz SR. Oliveira[110], Jhule S. Passinho[118], Lucas C. Quarantini[106], Yana Quarantini-Alvim[106,119], Vanessa R. Ramos[107], Aline S. Sampaio[110], Leonardo C. Saraiva[107], Roseli G. Shavitt[120], Otavio AFS. Silva[105], Mirelly Soares[107], Victor S. Souza[110], Bárbara P. Stumpf[117], Hermano Tavares[107]

[11]Psychiatry and Behavioral Sciences, Baylor College of Medicine, Houston, Texas, United States

[12]Department of Psychology, University of Nebraska-Lincoln, Lincoln, Nebraska, United States

[13]John Sealy School of Medicine, University of Texas Medical Branch, Galveston, Texas, United States

[14]Department of Psychology, University of Houston, Houston, Texas, United States

[15]Department of Genetics, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina, United States

[16]Department of Clinical Sciences, Lund, Lund University, Lund, Sweden

[17]Hospital Psiquiátrico de Asunción, Asunción, Paraguay

[18]Cátedra y Servicio de Psiquiatría, Universidad Nacional de Asunción, Lorenzo, Paraguay

[19]Department of Psychiatry, University of Puerto Rico, San Juan, Puerto Rico

[20]Hosptial Nacional "Hipólito Unanue", Lima, El Agustino, Perú

[21]Centro Interdisciplinario de Neurociencia de Valparaíso, Valparaíso, Chile
[22]Asociación Latinoamericana de Trastorno Obsesivo Compulsivo – ALTOC, Lima, Perú
[23]Hospital General de Agudos J. M. Ramos Mejía, Buenos Aires, Argentina
[24]Corporación Universitaria Rafael Nuñez, Cartagena, Bolívar, Colombia
[25]Talk TOC, Ciudad de México, Ciudad de México, México
[26]Departamento de Investigación Clínica, Grupo Médico Carracci, Ciudad de México, México
[27]The Mathison Centre for Mental Health Research & Education, University of Calgary, Calgary, Alberta, Canada
[28]Department of Psychiatry, University of Calgary, Calgary, Alberta, Canada
[29]Molecular and Human Genetics, Baylor College of Medicine, Houston, Texas, United States
[30]Psychiatry & Behavioral Sciences, Stanford University School of Medicine, Palo Alto, California, United States
[31]Department of Psychology, Utah State University, Logan, Utah, United States
[32]Santorio Santa Julia, Asunción, Paraguay
[33]Centro TOC México, Ciudad de México, Ciudad de México, México
[34]Psychology, University of Miami, Miami, Florida, United States
[35]Sensorium, Ciudad del Este, Paraguay
[36]University of British Columbia, Vancouver, British Columbia, Canada
[37]Institute for Genomics in Health, SUNY Downstate Health Sciences University, Brooklyn, New York, United States
[38]Psychology, University of South Florida, Tampa, Florida, United States
[39]Johns Hopkins University School of Medicine, Baltimore, Maryland, United States
[40]Hospital for Sick Children, Toronto, Ontario, Canada
[41]Psychiatry, Icahn School of Medicine at Mount Sinai, New York, New York, United States
[42]Psychology, University of Texas at Dallas, Dallas, Texas, United States
[43]Universidad Tecnológica Privada de Santa Cruz (UTEPSA), Santa Cruz, Andres Ibañez, Bolivia
[44]José Perales Clínica Privada, Santa Cruz, Andres Ibañez, Bolivia
[45]Department of Psychology, Vanderbilt University, Nashville, Tennessee, United States
[46]Instituto Nacional de Salud Mental "Honorio Delgado-Hideyo Noguchi", Lima, San Martín de Porres, Perú
[47]Universidad Nacional Mayor de San Marcos, Lima, Perú
[48]Centro de Estudios Traslacionales en Estrés y Salud Mental C-ESTRES, Valparaíso, Chile
[49]Psychology, University of North Texas, Denton, Texas, United States
[50]Centro UTOC, Ciudad de México, Ciudad de México, México
[51]Department of Psychological Sciences, Case Western Reserve University, Cleveland, Ohio, United States
[52]Center for Children and Families, Department of Psychology, Florida International University, Miami, Florida, United States
[53]Centro de Investigaciones Tecnológicas, Biomédicas y Medioambientales (CITBM), Lima, Perú
[54]Psychological Science, University of Texas Rio Grande Valley, Edinburg, Texas, United States
[55]Department of Psychological Science, University of North Carolina at Charlotte, Charlotte, North Carolina, United States
[56]Psychological, Health, and Learning Sciences, University of Houston, Houston, Texas, United States
[57]Hospital de Niños Dr. Ricardo Gutiérrez, Buenos Aires, Argentina
[58]Research Center, Rogers Behavioral Health, Philadelphia, Pennsylvania, United States
[59]School of Medicine, University of Pennsylvania, Philadelphia, Pennsylvania, United States
[60]Centro de Genetica y Biologia Molecular, Universidad de San Martin de Porres, Lima, Perú
[61]Universidad Católica Nuestra Señora de la Asunción, Asunción, Paraguay
[62]Research Center, Rogers Behavioral Health, Oconomowoc, Wisconsin, United States
[63]Biotechnology, Milwaukee Area Technical College, Milwaukee, Wisconsin, United States
[64]Psychiatry, University of Florida College of Medicine, Gainesville, Florida, United States
[65]SIN TOC, Ciudad de México, Ciudad de México, Mexico
[66]Universidad Nacional de Córdoba / Peer Support Argentina, Córdoba, Argentina
[67]Department of Psychiatry, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina, United States
[68]Private Practice, San Salvador, El Salvador
[69]Facultad de Psicología, Universidad de Buenos Aires, Buenos Aires, Argentina
[70]Peer Support Argentina, Buenos Aires, Argentina
[71]SIN TOC, Zapopan, Jalisco, Mexico
[72]Psychiatry and Behavioral Sciences, Texas A&M University, Bryan, Texas, United States
[73]Psicologia_Reclutadora, José Perales Clínica Privada, Santa Cruz, Andres Ibañez, Bolivia
[74]School of Nursing, University of Texas Health Science Center at San Antonio, San Antonio, Texas, United States
[75]Department of Psychiatry, Texas Tech University Health Sciences Center El Paso, El Paso, Texas, United States
[76]Neurogenetics Section, Molecular Science Department, Campbell Family Mental Health Research Institute, Centre for Addiction and Mental Health (CAMH), Toronto, Ontario, Canada
[77]Department of Psychiatry and Institute of Medical Science, Temerty Faculty of Medicine, University of Toronto, Toronto, Ontario, Canada
[78]Center for Bioethics, Harvard Medical School, Boston, Massachusetts, United States
[79]Department of Neurosurgery, Massachusetts General Hospital, Boston, Massachusetts, United States
[80]Department of Psychiatry, Massachusetts General Hospital, Boston, Massachusetts, United States
[81]Universidad Privada del Norte, Lima, Perú
[82]National Institute of Mental Health (NIMH), Bethesda, Maryland, United States
[83]Specialists in OCD & Anxiety Recovery (SOAR), PLLC, Richardson, Texas, United States
[84]Hospital Hermilio Valdizan, Lima, Santa Anita, Perú
[85]Department of Community Health and Epidemiology, Dalhousie University, Halifax, Nova Scotia, Canada
[86]Fordham University, Bronx, New York, United States
[87]Department of Psychiatry, Weill Cornell Medicine, Cornell University, New York, New York, United States
[88]Psiquiatría de Adultos, Centro Médico ABC, Ciudad de México, Ciudad de México, México
[89]Instituto de Fisiología, Facultad de Ciencias, Universidad de Valparaíso, Valparaíso, Chile
[90]Laboratorio de Enfermedades Psiquiátricas, Neurodegenerativas y Adicciones., Instituto Nacional de Medicina Genómica, Ciudad de México, México
[91]Department of Psychiatry, Rutgers University Robert Wood Johnson Medical School, New Brunswick, New Jersey, United States
[92]Department of Psychiatry, Rutgers University New Jersey Medical School, Newark, New Jersey, United States
[93]Psicología, Universidad San Francisco de Quito USFQ, Quito, Ecuador

[94]Centro de Investigaciones del Sistema Nervioso (CISNE), Ciudad de México, Ciudad de México, México
[95]Center for Medical Ethics and Health Policy, Baylor College of Medicine, Houston, Texas, United States
[96]Psychiatry, Sunnybrook Health Science Centre, Toronto, Ontario, Canada
[97]Investigación, Centro de Investigaciones del Sistema Nervioso (CISNE), Ciudad de México, Ciudad de México, México
[98]Centro de Investigación Biomédica y Aplicada (CIBAP), Escuela de Medicina, Facultad de Ciencias Médicas, Universidad de Santiago de Chile, Santiago, Chile
[99]Facultad de Ciencias de la Salud, Universidad Autónoma de Chile, Santiago, Chile
[100]Southwest Brain Bank, Department of Psychiatry, Texas Tech University Health Sciences Center El Paso, El Paso, Texas, United States
[101]Psicología, Universidad de Monterrey, Monterrey, Nuevo León, México
[102]Hospital Nacional Arzobispo Loayza, Lima, Perú
[103]Department of Psychiatry, Institute of Medical Science, and Department of Pharmacology and Toxicology, Temerty Faculty of Medicine, University of Toronto, Toronto, Ontario, Canada
[104]Universidade Federal de São Paulo, São Paulo, São Paulo, Brasil
[105]Universidade Federal do Maranhão (UFMA), São Luís, Maranhão, Brasil
[106]Neuropsychopharmacology Laboratory, Division of Psychiatry, Universidade Federal da Bahia, Salvador, Bahia, Brasil
[107]Institute and Department of Psychiatry, Universidade de São Paulo, São Paulo, São Paulo, Brasil
[108]Departamento de Psiquiatria, Faculdade de Medicina, Universidade Federal de Minas Gerais, Belo Horizonte, Minas Gerais, Brasil
[109]Universidade Federal de Ciências da Saúde de Porto Alegre, Porto Alegre, Rio Grande do Sul, Brasil
[110]Universidade Federal da Bahia, Salvador, Bahia, Brasil
[111]Curso de graduaçao em psicologia, Faculdade Santa Casa de Misericordia da Bahia, Salvador, Bahia, Brasil
[112]Instituto de Psiquiatria, Universidade Federal do Rio de Janeiro (IPUB/UFRJ), Rio de Janeiro, Rio de Janeiro, Brasil
[113]D'Or Institute for Research and Education (IDOR), Rio de Janeiro, Rio de Janeiro, Brasil
[114]Pax Clínica Instituto de Psiquiatria, Universidade Federal de Goiás, Goiania, Goiás, Brasil
[115]Universidade Federal do Rio Grande do Sul, Porto Alegre, Rio Grande do Sul, Brasil
[116]Universidade Federal do Paraná, Curitiba, Paraná, Brasil
[117]Universidade Federal de Minas Gerais, Belo Horizonte, Minas Gerais, Brasil
[118]Neuropsychiatry Outpatient Unit, Hospital Nina Rodrigues, São Luís, Maranhão, Brasil
[119]Faculdade Santa Casa de Misericordia da Bahia, Salvador, Bahia, Brasil
[120]Institute and Department of Psychiatry, Hospital das Clinicas HCFMUSP, Faculdade de Medicina, Universidade de São Paulo, São Paulo, São Paulo, Brasil

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.comppsych.2026.152737.

Data availability statement

The data supporting the findings of this study are available from the LATINO dataset upon reasonable request and upon approval from the dataset custodians (EAS, JC) and may require a data use agreement to ensure ethical and privacy compliance. All analyses were conducted using Python 3.8.1 with the libraries Optuna 4.4.0, SHAP 0.44.1, xgboost, 1.7.6, and scikit-learn 1.3.2. All parameters necessary for replication are provided in the Methods section, and code for data preprocessing, model training, and SHAP-based analyses is available from the corresponding author upon reasonable request.

References

[1] American Psychiatric Association. Diagnostic and statistical manual of mental disorder (DSM-5®). American Psychiatric Publisher; 2013.
[2] World Health Organization. International statistical classification of diseases and related health problems. World Health Organization; 2019.
[3] Cervin M. Obsessive-compulsive disorder: diagnosis, clinical features, nosology, and epidemiology. Psychiatric. Clinics 2023;46(1):1–16. https://doi.org/10.1016/j.psc.2022.10.006.
[4] Eisen JL, Sibrava NJ, Boisseau CL. Five-year course of obsessive-compulsive disorder: predictors of remission and relapse. J Clin Psychiatry 2013;74(3):72–86. https://doi.org/10.4088/JCP.12m07657.
[5] Liu J, Cui Y, Yu L. Long-term outcome of pediatric obsessive-compulsive disorder: a meta-analysis. J Child Adolesc Psychopharmacol 2021;31(2):95–101. https://doi.org/10.1089/cap.2020.0051.
[6] Sharma E, Thennarasu K, Reddy YJ. Long-term outcome of obsessive-compulsive disorder in adults: a meta-analysis. J Clin Psychiatry 2014;75(9):107–46. https://doi.org/10.4088/JCP.13r08849.
[7] Ruscio AM, Stein DJ, Chiu WT. The epidemiology of obsessive-compulsive disorder in the National Comorbidity Survey Replication. Mol Psychiatry 2010;15(1):53–63. https://doi.org/10.1038/mp.2008.94.
[8] Torres AR, Fontenelle LF, Shavitt RG. Comorbidity variation in patients with obsessive–compulsive disorder according to symptom dimensions: results from a large multicentre clinical sample. J Affect Disord 2016;190:508–16. https://doi.org/10.1016/j.jad.2015.10.051.
[9] Whoqol G. The World Health Organization quality of life assessment (WHOQOL): position paper from the World Health Organization. Soc Sci Med 1995;41(10):1403–9. https://doi.org/10.1016/0277-9536(95)00112-K.
[10] Coluccia A, Fagiolini A, Ferretti F. Adult obsessive–compulsive disorder and quality of life outcomes: a systematic review and meta-analysis. Asian J Psychiatr 2016;22:41–52. https://doi.org/10.1016/j.ajp.2016.02.001.
[11] Kugler BB, Lewin AB, Phares V, Geffken GR, Murphy TK, Storch EA. Quality of life in obsessive-compulsive disorder: the role of mediating variables. Psychiatry Res 2013;206(1):43–9. https://doi.org/10.1016/j.psychres.2012.10.006.
[12] Coluccia A, Ferretti F, Fagiolini A. Quality of life in children and adolescents with obsessive–compulsive disorder: a systematic review and meta-analysis. Neuropsychiatr Dis Treat 2017;13:597. https://doi.org/10.2147/NDT.S122306.
[13] Kaur R, Garg R, Raj R. Quality of life among patients with obsessive compulsive disorder: impact of stigma, severity of illness, insight, and beliefs. Ind Psychiatry J 2023;32(1):130–5. https://doi.org/10.4103/ipj.ipj_22_22.
[14] Wolf N, Du Mortier J, Van Oppen P, Hoogendoorn A, Van Balkom A, Visser H. Changes in insight throughout the natural four-year course of obsessive-compulsive disorder and its association with OCD severity and quality of life. Front Psychol 2023;14. https://doi.org/10.3389/fpsyt.2023.1231293.
[15] Żerdziński M, Burdzik M, Żmuda R, Witkowska-Berek A, Dębski P, Flajszok-Macierzyńska N, et al. Sense of happiness and other aspects of quality of life in patients with obsessive-compulsive disorder. Front Psychiatry 2022;13. https://doi.org/10.3389/fpsyt.2022.1077337.
[16] Wheaton MG, Ward HE. Intolerance of uncertainty and obsessive-compulsive personality disorder. Personal Disord 2020;11(5):357–64. https://doi.org/10.1037/per0000396.
[17] du Mortier JAM, Remmerswaal KCP, Batelaan NM, Visser HAD, Twisk JWR, van Oppen P, et al. Predictors of intensive treatment in patients with obsessive-compulsive disorder. Front Psychiatry 2021;12. https://doi.org/10.3389/fpsyt.2021.659401.
[18] Pozza A, Veale D, Marazziti D, Delgadillo J, Albert U, Grassi G, et al. Sexual dysfunction and satisfaction in obsessive compulsive disorder: protocol for a systematic review and meta-analysis. Syst Rev 2020;9(1):8. https://doi.org/10.1186/s13643-019-1262-7.
[19] Subramaniam M, Abdin E, Vaingankar J, Shafie S, Chang S, Seow E, et al. Obsessive-compulsive disorder in Singapore: prevalence, comorbidity, quality of life and social support. Ann Acad Med Singap 2020;49(1):15–25. https://doi.org/10.47102/ANNALS-ACADMEDSG.2019185.
[20] Kracker Imthon A, Antônio Caldart C, do Rosário MC, Fontenelle LF, Constantino Miguel E, Arzeno Ferrão Y. Stressful life events and the clinical expression of obsessive–compulsive disorder (OCD): an exploratory study. J Clin Med 2020;9(10). https://doi.org/10.3390/jcm9103371.
[21] Jensen S, Hybel K, Højgaard D, Nissen J, Weidle B, Ivarsson T, et al. Quality of life in pediatric patients with obsessive–compulsive disorder during and 3 years after stepped-care treatment. Eur Child Adolesc Psychiatry 2021. https://doi.org/10.1007/s00787-021-01775-w.
[22] Mataix-Cols D, Andersson E, Aspvall K, Boberg J, Crowley J, De Schipper E, et al. Operational definitions of treatment response and remission in obsessive-compulsive disorder capture meaningful improvements in everyday life. Psychother Psychosom 2022;91:424–30. https://doi.org/10.1159/000527115.
[23] Hertenstein E, Thiel N, Herbst N, Freyer T, Nissen C, Külz AK, et al. Quality of life changes following inpatient and outpatient treatment in obsessive-compulsive disorder: a study with 12 months follow-up. Ann General Psychiatry 2013;12(1):4. https://doi.org/10.1186/1744-859X-12-4.

[24] Huppert JD, Simpson HB, Nissenson KJ, Liebowitz MR, Foa EB. Quality of life and functional impairment in obsessive-compulsive disorder: a comparison of patients with and without comorbidity, patients in remission, and healthy controls. Depress Anxiety 2009;26(1):39–45. https://doi.org/10.1002/da.20506.
[25] Cervin M, Højgaard DR, Jensen S, Torp NC, Skarphedinsson G, Nissen JB, et al. A general factor of psychopathology predicts treatment and long-term outcomes in children and adolescents with obsessive-compulsive disorder. J Am Acad Child Adolesc Psychiatry 2024. https://doi.org/10.1016/j.jaac.2024.06.003.
[26] James G, Witten D, Hastie T, Tibshirani R. An introduction to statistical learning: With applications in R. Springer New York; 2013.
[27] Kern C, Klausch T, Kreuter F. Tree-based machine learning methods for survey research. Surv Res Methods 2019;13(1):73–93. https://doi.org/10.18148/srm/2019.v1i1.7395.
[28] Lu Y, Fan X, Zhang Y, Wang Y, Jiang X. Machine learning models using SHapley additive exPlanation for fire risk assessment mode and effects analysis of stadiums. Journal/Source. 2023;23(4). doi:10.3390/s23042151ˆhttps://.
[29] Mo Y, Habing B, Sedransk N. Tree-based methods: a tool for modeling nonlinear complex relationships and generating new insights from data. J Data Sci 2022;20(3):359–79. https://doi.org/10.6339/22-JDS1056.
[30] Louhichi M, Nesmaoui R, Mbarek M, Lazaar M. Shapley values for explaining the black box nature of machine learning model clustering. Procedia Comput Sci 2023;220:806–11. https://doi.org/10.1016/j.procs.2023.03.107.
[31] Lundberg SM, Lee S-I, Lundberg SM, Lee S-I, Kumar I, et al. A unified approach to interpreting model predictions. Adv Neural Inform Proc Syst (NeurIPS) 2017;30(1):4765–74. https://doi.org/10.48550/arXiv.1705.07874.
[32] Ponce-Bobadilla AV, Schmitt V, Maier CS, Mensing S, Stodtmann S. Practical guide to SHAP analysis: explaining supervised machine learning model predictions in drug development. Clin Transl Sci 2024;17(11):e70056. https://doi.org/10.1111/cts.70056.
[33] Sharma A, Verbeke W. Improving diagnosis of depression with XGBOOST machine learning model and a large biomarkers Dutch dataset (n = 11,081). Front Big Data 2020;3(15). https://doi.org/10.3389/fdata.2020.00015.
[34] Sharma HH, Ogunleye B. An optimal house price prediction algorithm: XGBoost. Analytics 2024;3(1):30–45. https://doi.org/10.3390/analytics3010003.
[35] Chen T, Guestrin C. XGBoost: A scalable tree boosting system. In: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining; n.d.; 785–94. doi:https://doi.org/10.1145/2939672.2939785.
[36] Crowley JJ, Cappi C, Ochoa-Panaifo ME, Frederick RM, Kook M, Wiese AD, et al. Latin American Trans-ancestry INitiative for OCD genomics (LATINO): study protocol. Am J Med Genet B Neuropsychiatr Genet 2024;195(4):e32962. https://doi.org/10.1002/ajmg.b.32962.
[37] Storch EA, Rasmussen SA, Price LH, Larson MJ, Murphy TK, Goodman WK. Development and psychometric evaluation of the Yale–Brown Obsessive-Compulsive Scale—Second Edition. Psychol Assess 2010;22(2):223–32. https://doi.org/10.1037/a0018492.
[38] Kroenke K, Spitzer RL, Williams JBW. The PHQ-9: validity of a brief depression severity measure. J Gen Intern Med 2001;16(9):606–13. https://doi.org/10.1046/j.1525-1497.2001.016009606.x.
[39] Spitzer RL, Kroenke K, Williams JBW, Löwe B. A brief measure for assessing generalized anxiety disorder: the GAD-7. Arch Intern Med 2006;166(10):1092–7. https://doi.org/10.1001/archinte.166.10.1092.
[40] Snaith RP, Hamilton M, Morley S, Humayan A, Hargreaves D, Trigwell P. A scale for the assessment of hedonic tone: the Snaith–Hamilton Pleasure Scale. Br J Psychiatry 1995;167(1):99–103. https://doi.org/10.1192/bjp.167.1.99.
[41] Endicott J, Nee J, Harrison W, Blumenthal R. Quality of life enjoyment and satisfaction questionnaire: a new measure. Psychopharmacol Bull 1993;29(2):321–6.
[42] UNDP. .. Human Development Report 2025 UNDP (United Nations Development Programme). 2025.
[43] Fuentes-Ramirez R. An approximation of Puerto Rico's human development index. Caribb Stud 2014;42:253–8. https://doi.org/10.1353/crb.2014.0010.
[44] Varma S, Simon R. Bias in error estimation when using cross-validation for model selection. BMC Bioinform 2006;7(91). https://doi.org/10.1186/1471-2105-7-91.
[45] Akiba T, Sano S, Yanase T, Ohta T, Koyama M. Optuna: A next-generation hyperparameter optimization framework. 2019. https://doi.org/10.1145/3292500.3330701.
[46] Bishop CM. Pattern recognition and machine learning (information science and statistics). Springer-Verlag; 2006. https://doi.org/10.1007/978-0-387-45528-0.
[47] Ke G, Meng Q, Finley T, Wang T, Chen W, Ma W, et al. LightGBM: a highly efficient gradient boosting decision tree. Journal 2017:3149–57. https://doi.org/10.5555/3294996.3295074.
[48] Prechelt L. Early stopping - but when? In: Orr GB, Müller K-R, editors. Neural Networks: Tricks of the Trade. Berlin Heidelberg: Springer; 1998. p. 55–69. https://doi.org/10.1007/3-540-49430-8_3.
[49] Zhang T, Yu B. Boosting with early stopping: convergence and consistency. Ann Stat 2004;33. https://doi.org/10.1214/009053605000000255.
[50] Guyon I, Elisseeff A. An Introduction to Feature Extraction. 2008. p. 1–25. doi:10.1007/978-3-540-35488-8_1.
[51] Liao Y, Rivoir J, Latty R, Bin Y. Deep feature selection using a novel complementary feature mask (2022). doi:10.48550/arXiv.2209.12282.
[52] M. El-Kenawy E-S, Khodadadi N, Eid MM, Khodadadi E, Khodadadi E, Khafaga DS, et al. Improved cancer detection through feature selection using the binary Al Biruni Earth radius algorithm. Sci Rep (2025) 15, no. 1, 9483. doi:https://doi.org/10.1038/s41598-025-92187-2.
[53] Kamalov F, Sulieman H, Alzaatreh A, Emarly M, Chamlal H, Safaraliev M. Mathematical methods in feature selection: a review. Journal/Source 2025;13(6). https://doi.org/10.3390/math13060996ˆhttps://.
[54] Wei G, Zhao J, Feng Y, He A, Yu J. A novel hybrid feature selection method based on dynamic feature importance. Appl Soft Comput 2020;93:106337. https://doi.org/10.1016/j.asoc.2020.106337.
[55] Huang J, Peng Y, Hu L. A multilayer stacking method base on RFE-SHAP feature selection strategy for recognition of driver's mental load and emotional state. Expert Syst Appl 2024;238:121729. https://doi.org/10.1016/j.eswa.2023.121729.
[56] Rifat MAK, Kabir A, Huq A. An Explainable Machine Learning Approach to Traffic Accident Fatality Prediction. 2024. https://doi.org/10.48550/arXiv.2409.11929.
[57] Shi K, Shi R, Fu T, Lu Z, Zhang J. A novel identification approach using RFECV–Optuna–XGBoost for assessing surrounding rock grade of tunnel boring machine based on tunneling parameters. Appl Sci 2024;14:2347. https://doi.org/10.3390/app14062347.
[58] Kraev E, Koseoglu B, Traverso L, Topiwalla M. Shap-Select: Lightweight Feature Selection Using SHAP Values and Regression [[arXiv:2410.06815]. In press]. 2024.
[59] Wang H, Liang Q, Hancock JT, Khoshgoftaar TM. Feature selection strategies: a comparative analysis of SHAP-value and importance-based methods. J Big Data 2024;1. https://doi.org/10.1186/s40537-024-00905-w.
[60] Kim JY. Improving appendix cancer prediction with SHAP-based feature engineering for machine learning models: a prediction study. Ewha Med J 2025;48(2):e31. https://doi.org/10.12771/emj.2025.00297.
[61] Mohanty PK, Francis SAJ, Barik RK, Roy DS, Saikia MJ. Leveraging Shapley additive explanations for feature selection in ensemble models for diabetes prediction. Bioengineering (Basel) 2024;11(12). https://doi.org/10.3390/bioengineering11121215.
[62] Hall MA, Smith LA. Feature Selection for Machine Learning: Comparing a Correlation-Based Filter Approach to the Wrapper. In: The Florida AI Research Society; n.d. doi:https://doi.org/10.5555/646812.707499.
[63] Vaulet T, Al-Memar M, Fourie H, Bobdiwala S, Saso S, Pipi M, et al. Gradient boosted trees with individual explanations: an alternative to logistic regression for viability prediction in the first trimester of pregnancy. Comput Methods Prog Biomed 2022;213:106520. https://doi.org/10.1016/j.cmpb.2021.106520.
[64] Kiekhaefer CM, Pourkamali-Anaraki F. Advancing Precision Medicine: An Evaluative Study of Feature Selection Methods. In: 2023 International Conference on Machine Learning and Applications (ICMLA); n.d.; 1052–9. doi:https://doi.org/10.1109/ICMLA58977.2023.00156.
[65] Pinciotti CM, Avery J, Zhang C, Muñoz JS, Berrones D, Cruz VZ, et al. Benchmarking empirical severity for the Yale-Brown Obsessive Compulsive Scale-Second Edition. J Affect Disord 2025;390:119719. https://doi.org/10.1016/j.jad.2025.119719.
[66] Macy AS, Theo JN, Kaufmann SC, Ghazzaoui RB, Pawlowski PA, Fakhry HI, et al. Quality of life in obsessive compulsive disorder. CNS Spectr 2013;18(1):21–33. https://doi.org/10.1017/s1092852912000697.
[67] Velloso P, Piccinato C, Ferrão Y, Perin EA, Cesar R, Fontenelle LF, et al. Clinical predictors of quality of life in a large sample of adult obsessive-compulsive disorder outpatients. Compr Psychiatry 2018;86:82–90. https://doi.org/10.1016/j.comppsych.2018.07.007.
[68] Ferrão YA, Shavitt RG, Prado H, Fontenelle LF, Malavazzi DM, de Mathis MA, et al. Sensory phenomena associated with repetitive behaviors in obsessive-compulsive disorder: an exploratory study of 1001 patients. Psychiatry Res 2012;197(3):253–8. https://doi.org/10.1016/j.psychres.2011.09.017.
[69] Correll CU, Cortese S, Solmi M, Boldrini T, Demyttenaere K, Domschke K, et al. Beyond symptom improvement: transdiagnostic and disorder-specific ways to assess functional and quality of life outcomes across mental disorders in adults. World Psychiatry 2025;24(3):296–318. https://doi.org/10.1002/wps.21338.
[70] Collins GS, Reitsma JB, Altman DG, Moons KGM. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD statement. Ann Intern Med 2015;162(1):55–63. https://doi.org/10.7326/M14-0697.
[71] Riley RD, Ensor J, Snell KIE, Archer L, Whittle R, Dhiman P, et al. Importance of sample size on the quality and utility of AI-based prediction models for healthcare. Lancet Digit Health 2025;7(6). https://doi.org/10.1016/j.landig.2025.01.013.
[72] Bzdok D, Altman N, Krzywinski M. Statistics versus machine learning. Nat Methods 2018;15(4):233–4. https://doi.org/10.1038/nmeth.4642.
[73] Molnar C. Interpretable machine learning: A guide for making black box models explainable. Leanpub; 2022.
[74] Kuhn M, Johnson K. Applied Predictive Modeling. Springer; 2013. https://doi.org/10.1007/978-1-4614-6849-3.